|CURIEUX|

ACADEMIC REVIEW

Economics & Business

2026 Issue Part 1

Editing Staff

Chief-In-Editor

George Hasnah

Lead Editor

Amanuel Geremew

Ashley Du

Assisting Editors

Jing Yan Huang

Nidhi Potineni

Sophia Kong

Tony Yuan

Ridit Aggarwal

Manu Sharma

Rishi Gadikota

Eric Wang

Table Of Contents

Page 6: Bitcoin as a Store of Value: Empirical Evidence from Risk-Adjusted, Crisis-Sensitive Analysis By Ethan Li

Page 21: Predicting Stock Market Prices Using Historical Opening Prices By Cahit Yankı EREZ

Page 33: The Phenomenon of Billionaires in the 21st Century: Should the Billionaires Exist? By Hanna Harchonak

Page 48: Algorithmic Modeling of Soccer Salaries By Neil Kapoor

Page 58: Culture Before Markets: Estimating the Economic Cost of Gendered Cultural Constraints in Romania from Original Survey and Interview Evidence By Jiayi Pan

Page 78: The Impact of Artificial Intelligence on Consumer Behaviour By Tracy Tan

Page 86: Limits of Arbitrage and Digital Collectible Pricing: Evidence from the CS: GO Skin Market By Jerry Chai

Page 93: Can Indian Steel Companies Remain Competitive While Transitioning Toward Sustainability? By Gautam Agrawal

Page 104: Digital & Streaming - Era Music Marketing: Strategies, Data, And Discovery in the Platform Economy By Baani Kohli

Page 114: The Viability of Diasporic Remittances as a Solution to Lebanon's Economic Crisis By Tamara Herzberg

Page 120: Instability and Innovation: Financial Uncertainty and Listed Fintech Market Performance By Qixuan Liang

Page 134: The Opportunity Cost of Corruption in Public Resource Allocation: Evidence from India's Education Sector, 2016-2022 By Vivan Balhara

Page 155: The Ethics and Effects of Nudge Theory in Business By Anun Amgalan

Page 163: IBOVESPA and Presidential Elections: A Performance-Based Forecasting Approach By Bernardo Dal Molin and Marco Pizzato

Page 177: Mega-Event Overlap and NYC Subway Ridership: A Case Study of the 2026 NBA Finals and FIFA World Cup By Zecheng Heliang

Page 188: Short-Term Returns, Long-Term Costs: How Corporate Stock Buybacks Can Hinder Innovation By Ceylan T. Urfali

Page 201: Gender-Based Price Discrimination By Yufei Guo

Page 213: Analyzing the Effectiveness of Different Monetary Policy Tools in Stabilizing Inflation During Economic Crises Using Time Series Analysis By Sahana Thothadri

Page 228: Domestic Transfer-Network Position and Emergency-Finance Strategies: Evidence from Venezuela By Héctor González, mentor By Jameson Augustin

Page 243: What Factors Influence Buyer Intention Towards Standard-Priced and Luxury Residential Apartments in Bangalore? By Vivaan Sahu

Page 261: Beyond Capital and Location FactorsInfluencingthe Survivaland GrowthofSmall Retail Businesses in Namangan, Uzbekistan By Abdikarimov Mirzoxid

Page 274: Paying with Less Friction: UPI and The Shift in Indian Consumer Spending By Saksham Bhartiya

Page 286: Financing the Future: Implementation of Value Capture in the NYC Subway System By Ray Zhang

Page 303: Cashless Economies and Consumer Spending: A Comparative Study of Kazakhstan and the United States By Islam Gumar

Page 314: Who Really Wins from The Name, Image, and Likeness (NIL)? Athletes' Gains and University Finances By Aristote Gideon Ophir

Page 332: Weak-Form Efficiency in Developed Versus Emerging Markets: A Fama-French Five-Factor Analysis By Kevin Yuan

Page 343: Scroll, Signal, Buy: An Integrated Analysis Of Identity, Psychology, Ethics, And Branding In Fashion Consumption By Ami Decelles

Page 359: Comparative Analysis of the Financial Sustainability of Three Global Pension Systems: Lessons from Sweden and Singapore for U.S. Social Security Reform By Anisha Kalaria

Page 383: Firm Scale and Advertising Efficiency in the Apparel Sector A Mixed-Methods Comparative Case Study of Paid Social Media and Out-of-Home Advertising By Ethan Moran

Page 398: The Impact of Macroeconomic and Institutional Factors on Trade By Cathy Zhang

Page 413: How does Persuasive Messaging in Chinese EV Companies' Social Media Influence Viewer Engagement? By Yan Yu

Page 432: When Spending Stops Feeling Like Spending: Digital Payments, Buy Now Pay Later, and Generation Z's Financial Behavior By Gadzhi Bilalov

Page 442: The Economics of Semiconductor Subsidies: A Comparative Review of the United States, European Union, and China By Matvei Ivashchenko

Page 460: Gender Inequality in Business Leadership: Hiring and Pay Gaps By Nakshatra Sakhamuri

Page 472: The Effect of Emotional Marketing Language on Consumer Brand Attachment By Vaishnavi Singh

Page 481: Priced for Perfection: Retail Investor Psychology and the Valuation of Unprofitable AI Companies By Ayaan Jindal

Page 491: Domestic and Global Drivers of the Brazilian Stock Market: The Role of Interest Rates, Political Stability, and International Market Conditions By Pietro Ribas

Page 502: WFH Case Study By Ajinkya Patil

Page 518: Segregated Work, Unequal Earnings: Occupation, Payment Systems, and the Gender Earnings Gap in Bombay City Cotton Mills, 1921-1937 By Anaisha Poddar

Page 529: The Impact of Artificial Intelligence on Labour Markets By Arya Khanna

Page 540: Has PMJDY Improved Financial Inclusion? Evidence from Banks in India By Ryan Bathwal

Page 551: The Impact of Perceived Authenticity in Athlete Endorsements on Consumer Trust and Purchase Intent By Rehaan Bhat

Page 563: On The Path to Financial Growth: Understanding Finance for Students By Raima Sengupta

Page 575: How External Influences Affect Individual Investor Decision-Making, and How Individual Investors Can Better Control their Emotion By Cyrus Melnyk

Bitcoin as a Store of Value: Empirical Evidence from Risk-Adjusted, Crisis-Sensitive Analysis By Ethan Li

Abstract

Bitcoin is positioned by its supporters as a decentralized store of value and inflation hedge, but its extreme price volatility calls this positioning into question. This paper examines whether Bitcoin truly possesses these characteristics by analyzing its long-term real returns, inflation sensitivity, and behavioral performance during crises. Using monthly data from Yahoo Finance and FRED covering the period from January 2014 to December 2025, this study systematically compares Bitcoin with gold, the S&P 500, and U.S. Treasury bonds within a risk-adjusted framework. The study finds that Bitcoin's real returns are driven by two narrow bull-market windows rather than consistent compounding, and its maximum real drawdown of 76% far exceeds that of gold or equities. Its inflation beta is significantly negative, and its correlation with the S&P 500 rises during the worst-performing equity months, reaching 0.336 during the bottom 5%, which is the opposite of safe-haven behaviour. These results suggest that, although Bitcoin generated substantial nominal returns during the sample period, its actual behavior more closely resembles that of a high-risk speculative asset than a reliable store of value or inflation hedge.

Introduction

Bitcoin emerged in 2009, and its original purpose was to allow people to transfer money to one another without the need for banks. However, over the past decade or so, its actual use has diverged significantly from its original design. Today, most Bitcoin holders do not use it for payments but instead hold it as a financial asset, and an increasing number of people are discussing whether it can serve as a "global store of value."

There is a major problem with this argument itself, traditional financial theory holds that a store of value must be able to maintain its purchasing power over the long term. Bitcoin's history, however, has been marked almost entirely by sharp surges and crashes; for instance, it once lost more than 80% of its value within a single year (Liu and Tsyvinski 2693). When compared to assets like gold, which have consistently maintained their value over time, Bitcoin stands out as a stark contrast.

The academic community has also engaged in endless debate on this issue, with no consensus reached. For example, some studies suggest that Bitcoin can serve as a tool for portfolio diversification under certain conditions and may even hedge against inflation, while others argue that it behaves more like a high-risk asset—rising more sharply when the market rises and falling more severely when the market falls. One reason for this divergence is the issue of research methodology. Many papers focus solely on average returns or short-term correlations, which can easily lead to overly optimistic conclusions. For instance, the major bull market from 2020 to 2021 made Bitcoin's average returns appear exceptionally strong, completely obscuring the significant downside risks investors faced during other periods. This paper contributes to the

existing literature by applying rolling real-return distributions, regime-split OLS, and conditional correlation analysis across 2014-2024.

This paper addresses three specific questions: First, has Bitcoin maintained its purchasing power relative to other traditional assets over the past decade? Second, does Bitcoin truly hedge against inflation under different economic conditions? Third, does Bitcoin behave more like a defensive asset or a high-risk asset during periods of market stress? This paper addresses these three questions from various perspectives, including risk-adjusted and crisis-sensitive analyses of Bitcoin, with a focus on long-term real returns, the characteristics of return distributions, and Bitcoin's performance during specific economic crises. Gold serves as the primary benchmark throughout the paper, as its centuries-long history as a store of value and inflation hedge makes it the most appropriate point of reference.

Literature Review

Research findings regarding Bitcoin's potential as a hedging tool, safe-haven asset, or store of value are contradictory, typically due to differences in the study's time frame, prevailing market conditions, and the specific circumstances of the country in question. Before reviewing the literature, it is necessary to clarify three related but distinct concepts. A diversifier refers to an asset that is not fully correlated with another asset and can reduce overall portfolio risk, though it does not necessarily provide protection during crises. A hedge is an asset that, on average, is uncorrelated or negatively correlated with another asset and can provide protection under normal market conditions. A safe haven is the most difficult of the three to achieve, as it requires the asset to be uncorrelated or negatively correlated with another asset during periods of extreme market volatility, precisely when protection is most needed. This set of definitions, proposed by Baur and Lucey, forms the analytical framework for this paper.

Baur and Lucey demonstrated that gold meets the criteria for both a hedging asset and a safe-haven asset relative to stocks, making it a natural benchmark for evaluating new assets such as Bitcoin. Their methodology involved distinguishing between normal market periods and crisis periods using extreme return thresholds; this approach has since been widely adopted in academia, and this paper follows suit.

Subsequently, Bouri et al. extended this framework using a dynamic conditional correlation GARCH (DCC-GARCH) model and found that the relationship between Bitcoin and stocks exhibits high temporal variability. Although Bitcoin occasionally displays characteristics of a safe-haven asset, it is better suited as a diversification asset, with overall reliability far lower than that of gold. Similarly, Urquhart and Zhang demonstrated that while Bitcoin may serve as a hedge against certain currencies, it fails to function as a safe-haven asset during periods of severe foreign exchange volatility.

An inflation hedge refers to an asset whose real value remains stable or even increases when inflation rises, thereby preserving an investor's wealth as the purchasing power of money declines. An increasing body of literature is examining whether Bitcoin possesses this characteristic. Choi et al. found that Bitcoin exhibits certain inflation-hedging characteristics

over longer time horizons, but these effects are highly volatile and dependent on market conditions; during periods of heightened uncertainty, Bitcoin generally fails to meet the criteria of a safe-haven asset. Arshad et al. compared Bitcoin with gold and silver across ASEAN countries and found that while Bitcoin demonstrated some inflation-hedging capacity in certain emerging markets, its performance was significantly less stable and reliable than that of gold. Smales reached a similar conclusion, arguing that due to extreme volatility, cryptocurrencies currently cannot serve as a substitute for gold as an inflation hedge.

A major shortcoming of the existing literature is its overemphasis on average returns and short-term correlations, which may obscure tail risks and long-term purchasing power dynamics. This study addresses this gap by focusing on the distribution of actual returns, tail risks, and behavioral patterns during crises.

Methodology

This study utilized monthly price data for five asset classes: Bitcoin price data was sourced from Yahoo Finance (ticker: BTC-USD); gold was represented by the SPDR Gold Trust ETF (GLD); and the stock market was represented by the S&P 500 ETF (SPY). For short-term government bonds, the SPDR Bloomberg 1-3 Month Treasury Bill ETF (BIL) was used as a cash proxy to represent the risk-free rate. For long-term bonds, the iShares 20+ Year Treasury Bond ETF (TLT) was used as a proxy; this ETF holds U.S. Treasury bonds with maturities exceeding 20 years and coupon rates close to zero, making it highly sensitive to changes in interest rates. All asset price series are converted to monthly total returns using month-end closing prices. Inflation data uses the Consumer Price Index (CPI) provided by the Federal Reserve Economic Data (FRED) database, and nominal returns are converted to real returns using the formula $r_{real} \approx r_{nominal} - \pi$. The sample period spans January 2014 to December 2025. Bitcoin has 135 months of data, while other assets have 143 months; this discrepancy highlights the issue of insufficient reliability in Bitcoin's price data prior to 2014. This paper uses US CPI as the inflation proxy, since all five asset return series are USD-denominated and the regression's interpretation requires currency consistency.

To address Research Question 1, this paper calculates the rolling 3-year and 5-year cumulative real returns for all assets. The distribution for each rolling window is described using the mean, median, standard deviation, 5th and 95th percentiles, and the probability of negative real returns. This ensures that the study is not dependent on any specific buying timing and accurately answers the question: "What would be the outcome of holding an investment for 3 or 5 years, regardless of when it was purchased?" To compare risk-adjusted returns across these assets, Sharpe ratios, maximum drawdowns, and Calmar ratios are computed over the full sample for comparison.

To address Research Question 2, this study divides all months into two groups—high inflation and low inflation—based on the median CPI growth rate, and compares the average real returns of various assets under different inflationary conditions, ensuring that all variables are controlled for. Additionally, to further test this hypothesis, the study groups months by inflation

quartiles, analyzing the top 25% of months with the highest inflation and the bottom 25% of months with the lowest inflation separately. Furthermore, this paper employs ordinary least squares (OLS) regression, using monthly inflation rates and exchange rate depreciation rates as explanatory variables, to estimate the sensitivity of Bitcoin's real returns to inflation. The OLS regression is repeated separately across four different interest-rate regimes: the Low Rate Era (2014-2019), the Pandemic Stimulus period (2020 - 2021), Rate-Hike Inflation (2022-2023), and Post-Hike Stabilisation (2024-2025).

Regarding Research Question 3, this paper examines several specific crisis windows, including: the COVID-19 shock from February to March 2020, the rate hike sell-off from January to June 2022, the Silicon Valley Bank crisis in March 2023, and the Turkish lira crisis from November to December 2021. For each crisis window, this paper calculates the cumulative returns and maximum drawdowns for each asset. Additionally, using the conditional correlation method proposed by Baur and Lucey (2010), this paper calculates the correlation between Bitcoin and the S&P 500 during both normal months and the bottom 10% of months in terms of stock market performance, and compares these results with the correlation to gold. If Bitcoin exhibits low or negative correlation during the worst-performing months for the stock market, it aligns with the definition of a safe-haven asset. Conversely, if the correlation increases, this indicates that it is a risky, high-beta asset rather than a defensive one.

Results and Discussion

Monthly returns: 143 months

	count	mean	std	min	25%	50%	75%	max
Bitcoin	135.0	0.0610	0.2090	-0.3777	-0.0719	0.0316	0.1744	0.6963
Gold	143.0	0.0093	0.0416	-0.0836	-0.0199	0.0036	0.0343	0.1176
S&P 500	143.0	0.0118	0.0418	-0.1249	-0.0114	0.0177	0.0367	0.1270
T-Bills	143.0	0.0015	0.0017	-0.0008	0.0000	0.0008	0.0031	0.0051
LT Bonds	143.0	0.0015	0.0385	-0.0942	-0.0229	0.0022	0.0227	0.1105

Table 1: Summary Statistics on Monthly Returns across All Assets (Jan 2014 - Dec 2024)

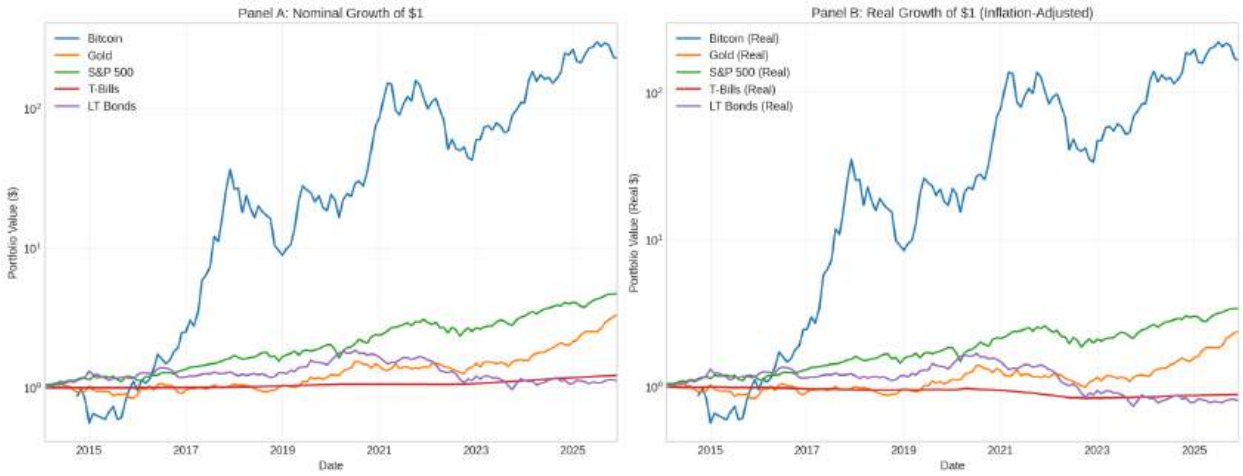


Figure 1: Nominal vs Real Growth of $1 across all assets over the investment horizon under study

As shown in Figure 1, although Bitcoin exhibits significantly higher volatility and instability, its long-term returns far exceed those of all traditional asset classes. Throughout the sample period, Bitcoin's nominal and real growth rates significantly outperformed those of gold, the S&P 500, and fixed-income assets; however, this was not the result of compound growth but was primarily concentrated in two major bull market cycles (2016–2017 and 2020–2021). In contrast, gold and the S&P 500 maintained stable and positive real growth rates, consistent with their status as traditional long-term store-of-value assets. The real growth rate of short-term U.S. Treasury bills (T-bills) was close to zero or slightly negative, indicating that cash equivalents fail to maintain their purchasing power after accounting for inflation. Finally, the real value of long-term U.S. Treasuries (represented by the iShares 20+ Year Treasury Bond ETF, or TLT — a fund holding bonds with maturities exceeding 20 years and near-zero coupon rates) plummeted during the 2021–2022 rate-hiking cycle, consistent with their high sensitivity to rising interest rates and inflation.



Figure 2: Distribution of monthly returns by asset. Bitcoin shows the largest distribution of monthly returns throughout the period,ranging from -37.8% to +69.6% monthly returns.

As shown in Figure 2, compared to all other asset classes, Bitcoin exhibits significantly higher return volatility and tail risk; therefore, it is a high-risk asset rather than a reliable store of value. Its monthly returns range from -37.8% to +69.6%, and both directions exhibit pronounced fat-tail characteristics, indicating high kurtosis and frequent occurrences of sharp price swings. Consequently, the distribution of Bitcoin's monthly returns is notably broader than that of other assets. In contrast, the return distributions of the S&P 500 and gold are symmetrical and mild, with their centers close to zero, consistent with the more stable return performance typical of store-of-value assets. The distribution of short-term U.S. Treasury bills (T-Bills) is nearly vertical, making it a highly predictable proxy for the risk-free rate. Although long-term U.S. Treasury bonds (as represented by the iShares 20+ Year Treasury Bond ETF, TLT) are widely regarded as a safe investment, their return distribution is significantly more dispersed than that of Treasury bills, indicating that they are exposed to substantial interest rate risk.

RQ1: Does Bitcoin preserve value?

=== Rolling 3-Year Real Returns ===

	Mean	Median	Std	P5 (Worst Case)	P25	P75	P95 (Best Case)	Pr(Return < 0)
Bitcoin (Real)	7.3002	3.8889	9.6500	0.1346	1.0026	8.9639	28.1277	0.02
Gold (Real)	0.1957	0.1230	0.2792	-0.0919	-0.0011	0.3078	0.6980	0.26
S&P 500 (Real)	0.3327	0.3069	0.1628	0.1234	0.2145	0.4210	0.6088	0.00
T-Bills (Real)	-0.0445	-0.0343	0.0508	-0.1257	-0.0886	-0.0129	0.0457	0.87
LT Bonds (Real)	-0.0846	-0.0443	0.2645	-0.4406	-0.3530	0.1185	0.3284	0.60

=== Rolling 5-Year Real Returns ===

	Mean	Median	Std	P5 (Worst Case)	P25	P75	P95 (Best Case)	Pr(Return < 0)
Bitcoin (Real)	23.5910	9.1227	27.8772	1.2885	4.1330	36.1758	77.1758	0.0132
Gold (Real)	0.3167	0.2897	0.1720	0.1027	0.1898	0.4244	0.6142	0.0000
S&P 500 (Real)	0.6322	0.6202	0.2070	0.3090	0.4738	0.7570	0.9942	0.0000
T-Bills (Real)	-0.0822	-0.0900	0.0337	-0.1275	-0.1084	-0.0507	-0.0257	1.0000
LT Bonds (Real)	-0.1141	-0.2012	0.2932	-0.5075	-0.3555	0.1027	0.3793	0.5921

Table 2(a): Rolling 3-year and 5-year real returns across assets. The standard deviation of Bitcoin's monthly returns is orders of magnitude greater than other assets, suggesting that its performance might be driven by a small number of extreme events as opposed to steady compounding.

Despite generating returns orders of magnitude higher than all other assets in the sample, Bitcoin exhibits the highest volatility throughout the period — a combination that fundamentally distinguishes it from traditional store-of-value assets. Although returns are extremely high, the data indicates that Bitcoin's value is not consistently stable. While its actual average returns are exceptionally high (7.3% over three years and 23.6% over five years), Bitcoin also has the

highest standard deviation of any asset (9.65 and 27.88, respectively), suggesting that its performance is driven not by steady compound growth but by a small number of extreme events. Although the probability of negative returns appears low (2% over a 3-year period and 1.3% over a 5-year period), this result should be interpreted with caution, as the sample period primarily reflects a structural bull market; under more neutral or bearish conditions, this probability could rise significantly. On the other hand, gold has a 26% probability of negative returns over a 3-year period and a 0% probability over a 5-year period, indicating greater reliability as a store of value and suggesting it can maintain purchasing power over longer time horizons. With a robust median real return of 0.62% and a 0% probability of negative returns over a 5-year period, the S&P 500 emerged as the most reliable risk-adjusted positive-return asset in this sample. In contrast, short-term U.S. Treasury bills (T-bills) consistently generated negative real returns (5-year median: -0.09%), indicating that cash-equivalent assets cannot maintain purchasing power over the long term. Finally, when evaluated against the criterion of serving as a store of value, long-term U.S. Treasuries (as represented by the iShares 20+ Year Treasury Bond ETF, or TLT) performed the worst, with a 59% probability of generating negative returns over a five-year period, highlighting their vulnerability in an environment of inflation and rising interest rates.

```
================================================================================
TABLE 1(b): Risk-Adjusted Performance Metrics — Full Sample (Real Returns)
================================================================================
                Ann. Real Return (%)  Ann. Volatility (%)  Sharpe Ratio  Max Drawdown (%)  Calmar Ratio
Asset
Bitcoin (Real)                 70.33                72.42         0.945            -76.07         0.925
Gold (Real)                     8.38                14.53         0.462            -29.08         0.288
S&P 500 (Real)                 11.42                14.48         0.668            -28.09         0.406
T-Bills (Real)                 -1.00                 1.17        -2.883            -16.45        -0.061
LT Bonds (Real)                -0.89                13.64        -0.195            -56.18        -0.016

Notes:
  Sharpe Ratio  = (mean monthly excess real return / std) × √12
  Calmar Ratio  = annualised real return / |max drawdown|
  Max Drawdown  = largest peak-to-trough decline in cumulative real wealth
```

Table 2(b): Risk-Adjusted Performance Metrics

Table 2(b) provides a more comprehensive assessment of each asset's performance from a risk-adjusted perspective. At first glance, Bitcoin's Sharpe ratio (0.945) appears to be the highest among all assets, but this figure is misleading because the Sharpe ratio penalizes upward and downward volatility equally; Bitcoin's extreme upside returns therefore artificially inflate this metric rather than reflecting sound risk management. The Calmar Ratio addresses this issue by dividing the annualized real return by the absolute value of the maximum drawdown, thereby providing a more direct measure of the relationship between returns and tail risk. Although Bitcoin's Calmar Ratio (0.925) appears equally robust on the surface, this figure is based on a maximum drawdown of -76.07%, meaning that any investor who bought at the peak of the cycle would have lost more than three-quarters of their real wealth before breaking even. The S&P 500 presents a starkly different picture: while its Sharpe ratio (0.668) is lower, its maximum drawdown is only -28.09%, and its Calmar ratio (0.406) thus reflects steady compound growth

rather than relying on extreme gains to mask massive losses. Gold's Sharpe ratio (0.462) and Calmar ratio (0.288) are the lowest among assets with positive returns, but its maximum drawdown of -29.08% is comparable to that of the S&P 500, indicating that the two are comparable in terms of downside risk control. Since both short-term and long-term Treasury bonds have negative annualized real returns, their Sharpe ratios and Calmar ratios are also negative, making them the worst performers under the store-of-value criterion.

RQ2: Does Bitcoin Hedge Inflation?

Bitcoin is not a reliable hedge against inflation. The empirical data presented in this study contradict the common perception that, as an asset with a fixed supply, it should benefit from rising inflation. When inflation rates are high, Bitcoin's returns tend to decline rather than rise, suggesting that it fails to preserve purchasing power like a true hedge. More importantly, compared to defensive stores of value such as gold, Bitcoin's performance aligns more closely with the characteristics of high-risk, high-volatility assets. This distinction is crucial because one of the primary arguments for including Bitcoin in financial portfolios over the long term is precisely its "anti-inflationary" narrative. Mean monthly returns by inflation regime across all assets. Results are presented under both a median split (Panel A) and a top-versus-bottom quartile classification (Panel B). Consequently, the findings in this section directly challenge this narrative and provide a more nuanced perspective on Bitcoin's economic function.

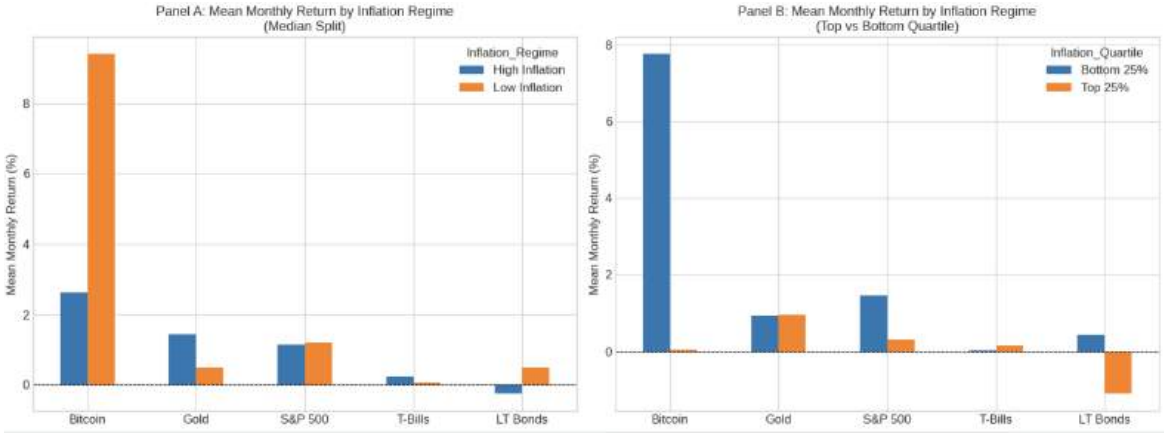


Figure 3: Mean monthly returns by inflation regime. Panel A compares high- and low-inflation periods based on a median split; Panel B compares the top and bottom quartiles of inflation months.

In Panel A, Bitcoin's average monthly return during periods of high inflation (approximately 2.6%) is significantly lower than during periods of low inflation (approximately 9.4%). Panel B illustrates this trend even more clearly: within the lowest quartile of inflation rates, Bitcoin's returns are relatively high (approximately 7.7%), but within the highest quartile of inflation rates, its average returns are nearly zero. This pattern suggests that Bitcoin performs best in low-inflation environments and loses performance as inflation increases, which is inconsistent with the expected behavior of an inflation hedge.

	Bitcoin	Gold	S&P 500	T-Bills	LT Bonds
Inflation_Regime					
High Inflation	2.629	1.437	1.150	0.233	-0.249
Low Inflation	9.413	0.584	1.177	0.079	0.423

=== Table 2b: Mean Monthly Returns (Top vs Bottom Quartile) ===

	Bitcoin	Gold	S&P 500	T-Bills	LT Bonds
Inflation_Quartile					
Bottom 25%	7.763	0.934	1.459	0.034	0.441
Top 25%	0.052	0.965	0.315	0.169	-1.099

Table 3: Mean Monthly Returns (%) by Inflation Regime

Table 3 shows again that Bitcoin's returns are always lower in high-inflation periods, no matter how you classify them. Bitcoin's average return falls sharply from 9.413% in low-inflation periods to 2.629% in high-inflation periods when the median is split. The same thing happened in the quartile analysis: Bitcoin makes 7.763% in the lowest inflation quartile but only 0.052% in the highest quartile. The fact that both methods give the same results makes the findings more reliable. It also shows that the negative link between Bitcoin returns and inflation is not caused by a certain way of defining inflation regimes. Gold, on the other hand, stays pretty stable across regimes. Long-term bonds, on the other hand, do very poorly during times of high inflation because they are sensitive to changes in interest rates.

Asset	Beta (Inflation)	p-value	R-squared	N	Interpretation
Bitcoin	-1.896339	0.012655	0.038723	132	Anti-hedge
Gold	-0.080312	0.626585	0.001763	132	No relationship
S&P 500	-0.224262	0.325821	0.012796	132	No relationship
T-Bills	0.006259	0.373849	0.006439	132	No relationship
LT Bonds	-0.300864	0.078247	0.027508	132	Anti-hedge

Table 3: Summary of ordinary least squares (OLS) regression results estimating the sensitivity of each asset's real returns to monthly inflation. A statistically significant negative beta indicates an anti-hedge relationship with inflation.

Table 3 shows a statistically significant negative correlation between Bitcoin and inflation ($\beta = -1.90$, $p < 0.05$), suggesting that Bitcoin can be classified as a “contrarian

safe-haven asset." This finding is consistent with the institutional classification analysis discussed above: Bitcoin performs worst precisely when inflation rates are highest. The most likely mechanism is monetary policy transmission: rising inflation prompts central banks to raise interest rates, which tightens liquidity across financial markets and disproportionately harms speculative, zero-yield assets such as Bitcoin. Furthermore, there is no statistically significant relationship between gold and inflation ($\beta = -0.08$, $p = 0.63$), consistent with its stable performance across economic cycles observed in Tables 2 and 2b. Long-term bonds also exhibit negative inflation sensitivity ($\beta = -0.30$, $p = 0.08$), reflecting their vulnerability to rising interest rates in high-inflation environments.

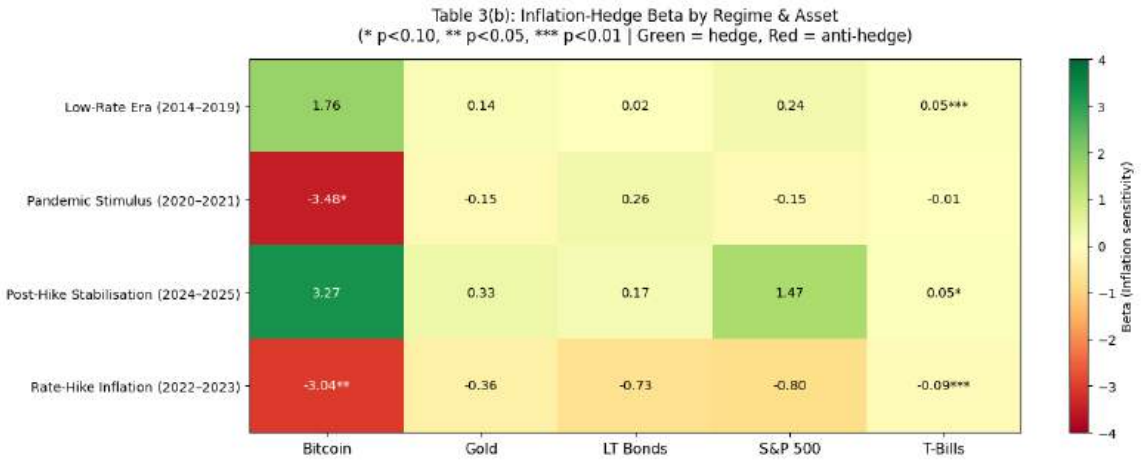


Figure 4: Inflation-hedge beta by regime and asset. Each cell displays the OLS beta coefficient estimating the sensitivity of each asset's real returns to inflation within the specified time period. Green indicates a positive (hedge-like) relationship; red indicates a negative (anti-hedge) relationship. Statistical significance is denoted as * p<0.10, ** p<0.05, *** p<0.01. Bitcoin is the only asset whose inflation beta is both large in magnitude and statistically significant, and its sign flips across regimes (positive during low-rate periods and significantly negative during high-inflation, rate-hiking environments)

Figure 4 illustrates Bitcoin's sensitivity to inflation through significant changes across different macroeconomic environments. Bitcoin's inflation beta exhibited a sign reversal across periods: it was positive during during the low interest rate period(2014-2019, $\beta = +1.76$), while it was significantly negative during the pandemic stimulus period(2020-2021, $\beta = -3.48$) and the rate hike inflation period(2022-2023, $\beta = -3.04$). This volatility suggests that Bitcoin's response to inflation depends largely on the prevailing monetary policy environment rather than reflecting a stable hedging characteristic.

Bitcoin's beta is statistically significant only when negative, reaching significant levels during both the pandemic stimulus period and the rate hike inflation period, but is not significant when positive, during the low interest rate period and the post rate hike stabilization period. This asymmetric significance further indicates that Bitcoin's anti hedge characteristics are more reliable and systematic than its potential hedge characteristics.

In comparison, gold's beta remained close to zero across all four periods, ranging from

−0.36 to 0.33, with no period reaching statistical significance. This aligns with gold's stable, neutral performance in inflationary environments, consistent with its role as a long term value store asset.

Furthermore, no asset in this sample produced a large, positive and statistically significant inflation beta during any period. This finding suggests that, over a sample period from 2014 to 2025, systematic inflation hedging capabilities were generally absent across all mainstream assets studied, rather than being a problem unique to Bitcoin.

RQ3: Does Bitcoin Behave as a Safe Haven During Crisis Periods?

Empirical evidence spanning multiple crisis periods indicates that Bitcoin consistently behaves as a high-risk asset rather than a safe-haven asset during times of market stress. Although Bitcoin advocates often position it as "digital gold," its actual behavior during crises differs markedly from that of gold.

As shown in Table 4, during the COVID-19 shock (February to March 2020), Bitcoin fell by 33.4%, nearly matching the decline of the S&P 500 (33.7%), while long-term Treasury bonds rose by 14.2% over the same period—a performance typical of a true safe-haven asset. During the 2022 rate hike sell-off, Bitcoin fell by 56.1%, far outpacing the declines in gold (+2.6%) and the S&P 500 (-23.0%), making it the worst-performing asset of all. During the Turkish lira crisis, Bitcoin similarly fell by 24.1%, while gold remained stable (+2.1%). It is worth noting that during the SVB banking crisis, Bitcoin rose by 27.9%, marking the only instance in this sample where Bitcoin exhibited safe-haven-like characteristics. This reflects investors' risk-aversion behavior driven by specific concerns about the traditional banking system, but does not indicate systemic safe-haven capabilities.

Crisis	Window	Bitcoin	Gold	S&P 500	LT Bonds
COVID Crash	2020-02-19 to 2020-03-23	-33.4%	-3.6%	-33.7%	14.2%
COVID Recovery	2020-03-23 to 2020-08-31	82.0%	26.3%	57.4%	-1.7%
2022 Rate Hike Selloff	2022-01-03 to 2022-06-16	-56.1%	2.6%	-23.0%	-22.0%
SVB/Banking Crisis	2023-03-08 to 2023-03-20	27.9%	9.1%	-0.9%	4.0%
Turkey Lira Crisis 2021	2021-11-01 to 2021-12-31	-24.1%	2.1%	3.6%	1.4%

(Negative = loss, Positive = gain during the crisis window)

Table 4: Cumulative Returns During Crisis Windows



Figure 5: Asset returns during crisis windows (%). Each cell displays the cumulative return for the respective asset during the identified crisis period. Negative values indicate losses; positive values indicate gains.

The heatmap in **Figure 5** further highlights the asymmetric high volatility of Bitcoin returns. As shown, during market downturns, Bitcoin's losses exceed those of most other assets; conversely, during market rallies, its gains do the same. This pattern of amplified market volatility is a hallmark of high-beta assets and is fundamentally at odds with the stable value preservation required of safe-haven assets.

```
=== Table 5: Conditional Correlations with S&P 500 ===
Regime                        BTC-SPY   Gold-SPY   N months
------------------------------------------------------------
All Months                      0.348      0.057        135
Normal (P10-P90)                0.203      0.092        107
Worst 10% Equity                0.296      0.257         14
Worst 5% Equity                 0.336      0.160          7
Best 10% Equity                 0.332      0.030         14
```

Table 5: Conditional correlations with S&P 500

The conditional correlation analysis in Table 5 further quantifies this conclusion: across the entire sample of 135 months, the correlation between Bitcoin and the S&P 500 was 0.348, significantly higher than that of gold at 0.057. Most importantly, during the bottom 10% of months in terms of stock market performance, the correlation between Bitcoin and the S&P 500 rose to 0.296, and further increased to 0.336 during the bottom 5% of months. In contrast, gold's correlation during the bottom 5% of months fell from 0.092 in normal periods to 0.160, moving in the opposite direction. According to the definition by Baur and Lucey (2010), safe-haven

assets are expected to exhibit low or negative correlation under extreme market stress. Therefore, the fact that Bitcoin's correlation actually increases during crises clearly refutes its potential as a safe-haven asset. Gold's conditional correlation pattern, however, aligns more closely with this definition, consistent with its historical status as a classic defensive asset.

Conclusion

This paper examines Bitcoin's characteristics as a store of value, inflation hedge, and safe-haven asset over the period from January 2014 to December 2025, benchmarking its performance against gold, the S&P 500, and U.S. Treasury bonds. The evidence from the three research questions points in the same direction. In terms of preserving purchasing power, while Bitcoin's returns appear impressive, they stem almost entirely from two brief bull market windows rather than sustained compounding. Its maximum real drawdown reached 76%, approximately three times that of gold and the S&P 500. Regarding inflation hedging, Bitcoin's full sample inflation beta is significantly negative; the higher inflation, the worse Bitcoin performs. The segmented analysis in Figure 4 further indicates that this inverse relationship is most pronounced precisely during periods of the highest inflation, while the positive coefficients observed during periods of low inflation never reached statistical significance. Regarding crisis behavior, during most crisis windows, Bitcoin declined alongside the stock market, and its correlation with the S&P 500 rose precisely when equity performance was worst, a pattern inconsistent with the safe-haven definition established by Baur and Lucey (2010).

This paper extends the existing literature in three ways. First, because most existing studies rely on point in time average returns, they are susceptible to distortion from a small number of extreme periods; this paper addresses that limitation by using rolling real return distributions, which expose tail risks that mean return analyses tend to obscure. Second, existing studies typically estimate a single inflation beta across the full sample, which fails to capture how Bitcoin's inflation sensitivity shifts across different monetary policy environments. By splitting the regression into four distinct regimes, this paper reveals structural instability that a pooled estimate would mask. Third, rather than relying on any single metric, this paper combines Sharpe ratios, Calmar ratios, maximum drawdown, and conditional correlation analysis to cross-validate the same core question from both a risk adjusted performance and a crisis-behavior perspective.

The conclusions of this study must be interpreted within certain constraints. The sample period largely overlaps with Bitcoin's structural bull market, and inflation remained exceptionally low for most of the 2010s. Therefore, some of the findings, particularly the low probability of negative returns, which may not be directly replicable in a Bitcoin bear market or a monetary policy environment dominated by high inflation. Since all asset returns are denominated in U.S. dollars, the inflation analysis relies solely on the U.S. CPI. This implies that the study does not formally test Bitcoin's hedging properties in environments of local currency depreciation. Extending the analysis to the local currency perspective in high inflation economies such as Turkey and Brazil represents a promising direction for future research. Finally, monthly

data smooths intra monthly volatility, causing some price dynamics during crises to be overlooked. Future research utilizing daily data would help more accurately identify changes in correlations during extreme events.

Works Cited

Arshad, Shaista, et al. "Are Bitcoin, Gold, and Silver Equally Effective Inflation Hedges? Evidence from ASEAN Markets." Resources Policy, vol. 84, 2023, pp. 1–14.

Baur, Dirk G., and Brian M. Lucey. "Is Gold a Hedge or a Safe Haven? An Analysis of Stocks, Bonds and Gold." The Financial Review, vol. 45, no. 2, 2010, pp. 217–229.

Baur, Dirk G., and Thomas K. McDermott. "Is Gold a Safe Haven? International Evidence." Journal of Banking & Finance, vol. 34, no. 8, 2010, pp. 1886–1898.

Bouri, Elie, et al. "On the Hedge and Safe Haven Properties of Bitcoin: Is It Really More than a Diversifier?" Finance Research Letters, vol. 20, 2017, pp. 192–198.

Choi, Sangyup, and Junhyeok Shin. "Bitcoin: An Inflation Hedge but Not a Safe Haven." Finance Research Letters, vol. 46, 2022, pp. 1–7.

Liu, Yukun, and Aleh Tsyvinski. "Risks and Returns of Cryptocurrency." The Review of Financial Studies, vol. 34, no. 6, 2021, pp. 2689–2727.

Nakamoto, Satoshi. Bitcoin: A Peer-to-Peer Electronic Cash System. 2008, bitcoin.org/bitcoin.pdf.

Smales, Lee A. "Cryptocurrency as an Alternative Inflation Hedge?" Accounting & Finance, vol. 64, no. 1, 2024, pp. 1589–1610.

Urquhart, Andrew, and Hanxiong Zhang. "Is Bitcoin a Hedge or Safe Haven for Currencies? An Intraday Analysis." International Review of Financial Analysis, vol. 63, 2019, pp. 49–57.

Predicting Stock Market Prices Using Historical Opening Prices By Cahit Yankı EREZ

Abstract

Share prices are shaped by company performance, macroeconomic conditions and investor sentiment. This makes the stock market a complex system and short term price forecasting a difficult but valuable problem. This study asked which of several standard machine learning models most accurately predicts the next day's opening price of a stock from the three preceding opening prices. It also asked whether more flexible models outperform simpler ones. Five years of daily trading history was retrieved for Google, Apple and Tesla using the Python library. The data were reshaped into 1250 examples of three consecutive opening prices paired with the following day's opening price, then split 60/20/20 into training, validation, and test sets. Linear regression, a decision tree, a random forest, and a multilayer perceptron were tuned on the validation set by grid search. They were then compared on the held-out test set using mean squared error. Linear regression produced the lowest test error for all three companies. The unpruned decision tree produced the highest. These results suggest that over a three-day window the relationship between consecutive opening prices is close to linear and that model complexity alone does not improve price forecasts. Therefore linear regression has scored the least mean squared error.

Introduction



The stock market is a marketplace where people buy and sell shares of public companies. When a company goes public, it issues shares that represent partial ownership in the firm. These shares are then traded on exchanges. Share prices change constantly based on supply and demand. Many factors shape this demand, including company performance, economic indicators, interest rates, and investor sentiment. The stock market plays several key roles in an economy. It

allows companies to raise money, helps investors build wealth, and reflects the overall health of the economy. Because so many variables affect prices, the market behaves as a complex and often unpredictable system. This complexity makes price prediction both valuable and difficult.

Whether prices can be predicted at all is a long standing question in finance. The efficient market hypothesis argues that prices already reflect all publicly available information, so future movements should be essentially unpredictable from past prices alone. Under the weakest form of this hypothesis, technical strategies built purely on historical price sequences should offer no advantage. However a large body of empirical work reports that machine learning models trained on historical price data can achieve forecasting performance better than chance. The disagreement between theory and these empirical results is part of what makes the problem worth investigating.

Machine learning approaches to this problem fall into a few broad families. Linear models assume a straight line relationship between inputs and the target and are valued for their stability and interpretability. Tree-based methods, including decision trees and random forests, partition the input space into regions and can represent non linear structure without the analyst specifying its form in advance. Neural networks approximate arbitrary functions through layers of weighted transformations and have been applied widely to financial time series. Ensemble approaches combine several models on the reasoning that their individual errors are partly independent and will partly cancel. Prior comparative studies have applied these families to index and equity prediction. However they differ in the features used, the forecast horizon, and how performance is measured.

A recurring difficulty in this literature is separating genuine predictive skill from overfitting. A model with enough capacity can reproduce its training data almost perfectly while failing on data it has not seen. Studies that report performance only on data the model was fitted to therefore risk overstating what the model has learned. Guarding against this requires holding out data that is used once, at the end, and not consulted during model selection.

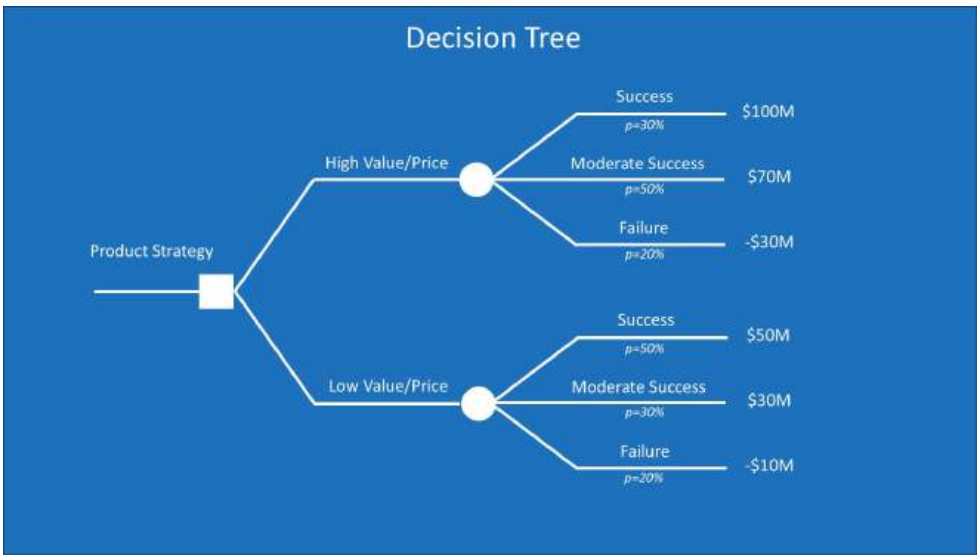
This study compares linear regression, a decision tree, a random forest, and a neural network on a deliberately minimal prediction task. That task is to forecast the next day's opening price from the three preceding opening prices. The purpose was to establish which model family performs best on this task and to test whether additional model complexity translates into better generalization.

"Machine Learning Stock Market Prediction Studies: Review and Research Directions" reviews 20 years of machine learning methods, it evaluates their limitations, finds key findings and categorizes the main approaches. "Artificial intelligence showcases methods for financial market prediction. A systematic review" research portrays machine learning methods for the financial market focusing on challenges of price forecasting. "A Comparative Study of Transformer-Based and Classical Models for Financial Time-Series Forecasting" compares classical and deep learning models for predicting stock market, showing that model performance is asset based so that there is not much difference at the accuracy between traditional and new methods.

Materials and Methods

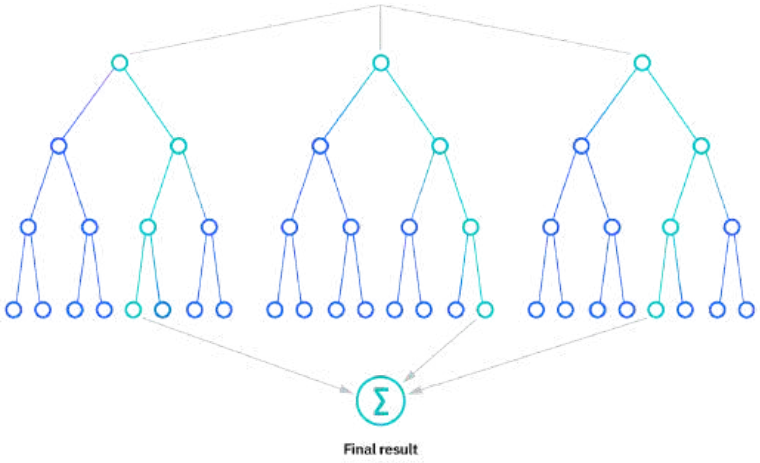
For this project, historical stock price data was collected using the finance Python library. The stock chosen was Google and five years of daily trading history was retrieved, giving 1259 data points. The same procedure was then repeated for Apple and Tesla, which were selected to give three companies from different industries. Each series covered the same five-year window. From the raw data, only the Open, Close, and High prices were kept. Open prices are the prices when the market opens at 9.30 am. Columns like Volume, Dividends, and Stock Splits were dropped since they weren't needed for this prediction task. The data was reshaped so that each input (X) consisted of three consecutive days' opening prices, and the target (Y) was the opening price of the following day.

A single dataset can not be used both to train a model and to fairly judge how good it is, because a model that has already seen the answers will look artificially accurate. To avoid this, the data was split into three parts: a training set, a validation set, and a test set. The training set, 60% of the data, is what the model actually learns from. The validation set, 20%, acts like a fake test, it is used to tune the model's settings and compare different versions of a model. Finally, the test set, the remaining 20%, is used only once, to get an unbiased measurement of how the model performs on data. This three way split prevents a common mistake in machine learning called overfitting. It occurs where a model appears accurate during development but performs poorly on real world data.

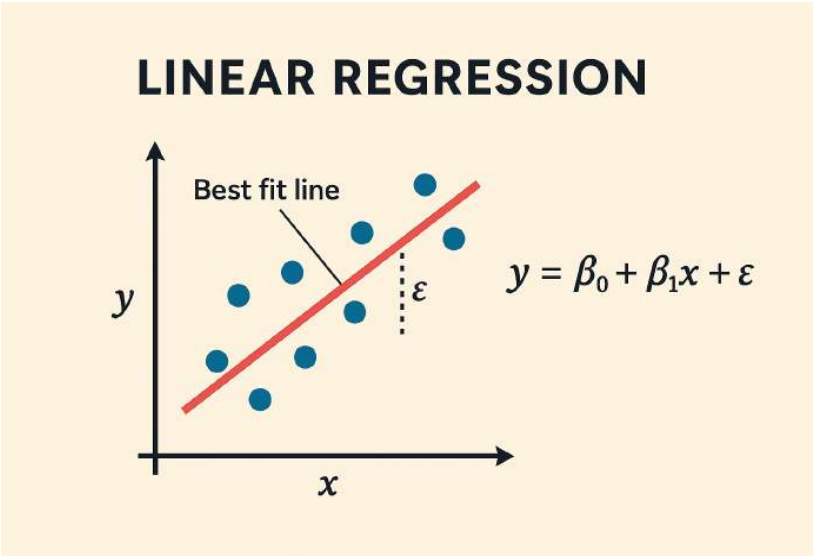


A decision tree is a machine learning method used for classification and regression. The model looks like an upside down tree, starting from a single root node that holds the whole dataset. From there, it branches into internal nodes, and each node tests one feature against a certain condition. At every split, the algorithm chooses the feature that best separates the data. To measure this, it uses criteria such as Gini impurity, information gain or variance reduction. This process repeats until a stopping rule is reached. The final branches, called leaf nodes, give the

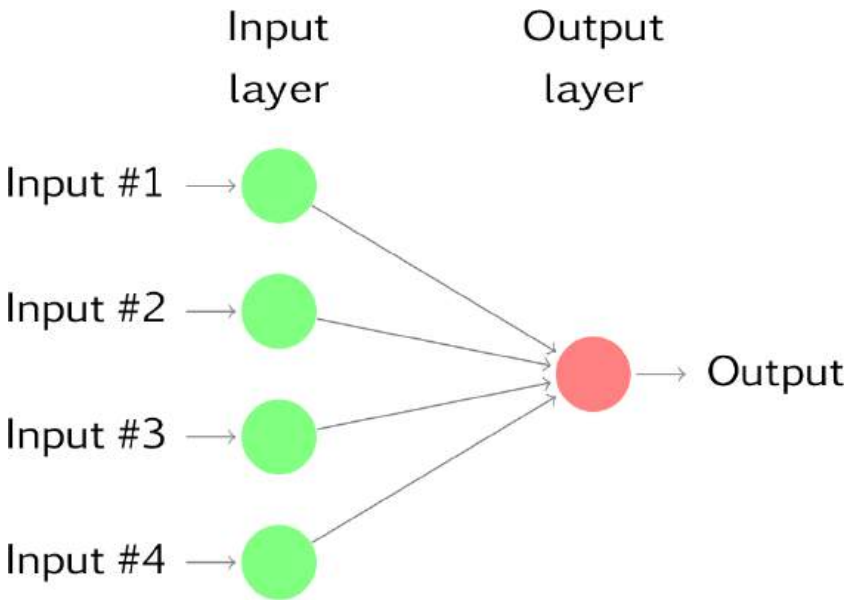
prediction. In stock market prediction, a decision tree can learn rules from past data to forecast whether a price will rise or fall.



A random forest is an ensemble method that combines many decision trees into one model. Instead of trusting a single tree, it builds a large number of trees and merges their results. For classification, it uses majority voting and for regression, it averages the outputs. Two types of randomness make this method effective. First, each tree is trained on a different random sample of the data, a technique known as bagging. Second, each split considers only a random subset of the available features. This diversity reduces the link between the trees and limits overfitting. As a result, the model becomes more stable and generalizes better to unseen data. In stock market prediction, random forests can capture complex, non linear patterns among many indicators. They also handle outliers well. In addition, they can rank feature importance, which helps researchers see which variables matter most.



Linear Regression assumes a straight-line relationship between the input prices and the predicted future price. It works by finding the line that minimizes the overall squared error between its predictions and the actual values.



Neural Network (MLPRegressor) uses layers of interconnected nodes that apply weighted mathematical transformations to the input data. Through training, it adjusts these weights to minimize prediction error. This allows it to capture more complex relationships in the data than the simpler models can.

Finally, an ensemble model was created by averaging the predictions of all five models together. It can also be based on the weighted averages of the models. It is based on the idea that combining multiple models can smooth out individual weaknesses and produce a more reliable overall prediction. Considering each model's MSE their weights are calculated for making a sum of 1. Then those weights are multiplied with the results for creating an ensemble model. However, averages are not the best because some models perform much better than others which can disturb the result.

All analysis was performed in Python using finance for data retrieval. Reshaping the 1259 day series into overlapping three-day windows produced 1250 input-target pairs. The split was carried out with train_test_split using a fixed random seed of 42. Twenty percent was separated as the test set, and the remainder was divided into 750 training and 250 validation examples.

Hyperparameters were selected by grid search with five-fold cross-validation on the training set, scoring on negative mean squared error. For the decision tree the grid covered maximum depth (3, 5, 7, 10, unlimited), minimum samples per split (2, 5, 10) and minimum samples per leaf (1, 2, 4). For the random forest it covered the number of trees (50, 100, 200) and maximum depth (5, 10, unlimited). It also covered minimum samples per split (2, 5) and

minimum samples per leaf (1, 2). For the neural network it covered hidden layer configuration ((50), (100), (100, 50)), activation function (ReLU, tanh) and initial learning rate (0.001, 0.01), with a maximum of 1000 iterations. Selected values are given in Table 3. A separate sweep of decision tree maximum depth from 2 to 20 was scored directly on the validation set. It was run to examine how tree depth relates to generalization. Every model was evaluated on the test set exactly once, after all tuning was complete.

Results

Stock price prediction involves continuous numbers, so accuracy is not a meaningful metric. A predicted price of $124.90 when the real price was $125.00 is not wrong in the same binary sense a misclassified label would be. Squaring the errors serves two purposes: it makes all errors positive and it penalizes larger errors more heavily than smaller ones. Therefore it is useful for catching models that occasionally make big, costly mistakes. A lower MSE means the model's predictions are closer to the real prices. This is making it possible to directly compare models like Linear Regression, Decision Trees, and Neural Networks against each other using a single number.

Model comparison for Google

After all the models were trained and tuned, each one was tested on the same test set. This gives a comparison of how well each model predicts new prices. The results were: Linear Regression scored a test MSE of about 15.08, the Decision Tree scored 34.82, the Random Forest scored 19.76 and the Neural Network scored 18.28. The Linear Regression, gave the lowest test error, while the Decision Tree gave the highest error. This suggests that the relationship between three days of opening prices and the next day's price is close to linear. Therefore a simple straight-line model fits it very well, while a single decision tree struggles to generalise to new data.

To understand the models better, the MSE was also measured on the training set. Here the results were: Linear Regression 16.01, Decision Tree 0.00, Random Forest 5.67, and Neural Network 18.98. The most important thing to notice is the Decision Tree. Its training MSE is exactly 0.00, which means it predicted the training data perfectly. However, its test MSE was the worst of all the models (34.82). This large gap between training error and test error is a clear example of overfitting: the tree simply memorised the training data instead of learning a general pattern, so it failed when it met new data. In contrast, Linear Regression has almost the same error on the training set (16.01) and the test set (15.08) showing it is stable and generalises well.

The paragraphs above report the Google run in full. Linear regression was most accurate at a test MSE of 15.08, followed by the neural network at 18.28 and the random forest at 19.76. The decision tree was least accurate at 34.82, roughly 2.3 times the error of linear regression.

Model comparison for Apple

Trained on the AAPL series, linear regression predicted the next day's opening price most accurately of the four models, with a test MSE of 13.10. The random forest followed at 15.30 and the neural network at 17.25. The unpruned decision tree was least accurate at 25.86, roughly double the error of linear regression. Linear regression also showed the smallest gap between its training and test errors.

Model comparison for Tesla

The same ordering held for TSLA as for the other two companies. Linear regression again gave the lowest test error at 117.33 and the decision tree the highest at 231.47. The random forest scored 127.93 and the neural network 117.14. The absolute error values for TSLA are roughly nine times those for AAPL and Google. This reflects the price scale of the two stocks rather than a difference in model quality. MSE is expressed in squared units of price, and TSLA traded at a higher and more volatile level over the sampled period. Only the ranking of models is meaningfully comparable across the three companies.

Evidence of overfitting in the unpruned tree

The decision tree reached a training MSE of exactly 0.00 on all three datasets. It recorded the worst test error in every case. A tree grown without a depth limit can continue splitting until every training example sits in its own leaf, at which point it reproduces the training targets exactly. The gap between a perfect training fit and the worst test performance is the clearest available evidence that the model memorized rather than generalized. Constraining the tree removed most of this gap. Sweeping maximum depth from 2 to 20 and selecting on validation error identified a depth of 5 as optimal for all three companies. This reduced Google validation MSE from 31.07 to 24.25, Tesla validation MSE from 18.01 to 15.54 and TSLA validation MSE from 224.97 to 150.56. A full grid search over depth, minimum samples per split, and minimum samples per leaf reduced AAPL test error further to 20.11 and TSLA test error to 129.34. It fell to 27.44 for Google, 20.11 for Apple and 29.34 for Tesla. All three remained worse than linear regression, but neither was any longer an outlier.

Ensemble Performance

Ensembles were evaluated for all three companies. Averaging the model predictions with equal weights gave a test MSE of 17.96 for Google, 14.81 for AAPL and 118.24 for TSLA. A weighted average gave 17.19, 14.15 and 117.03. No ensemble improved on linear regression alone for any company. Because the poorly generalizing decision tree contributed to the average, the combination inherited some of its error.

Figures, Tables and Captions

MSE for Google Models were tuned on a validation set and evaluated once on a held-out test set of 250 examples. Errors are in squared US dollars.

Model	Training MSE	Test MSE
Linear Regression	16.01	15.08
Decision Tree	0.00	34.82
Random Forest	5.67	19.76
Neural Network	18.98	18.28

MSE for Apple

Model	Training MSE	Test MSE
Linear Regression	12.36	13.10
Decision Tree	0.00	25.86
Random Forest	9.19	15.30
Neural Network	17.18	17.25

MSE for Tesla Absolute values are larger than in Tables 1 and 2 because TSLA traded at a higher price level. Only the relative ordering of models should be compared between tables.

Model	Training MSE	Test MSE
Linear Regression	117.82	117.33
Decision Tree	0.00	231.47
Random Forest	88.54	127.93
Neural Network	115.60	117.14

MSE for Nike Absolute values are lower compared to the first three tables as NKE is traded at lower prices. It is added to showcase the model's ability to work for non-tech companies.

Model	Training MSE	Test MSE
Linear Regression	3.50	5.11
Decision Tree	3.66	8.06
Random Forest	2.65	5.63
Neural Network	3.73	5.44

Best hyperparameters selected by grid search

Model	Google	Apple	Tesla
Decision Tree	no depth limit, min. leaf 4, min split 2	depth 5, min leaf 4, min split 2	depth 5, min leaf 4, min split 2
Random Forest	no depth limit, min. leaf 2, min split 5, 200 trees	depth 5, min leaf 4, min split 5, 100 trees	depth 5, min leaf 1, min split 5, 200 trees
Neural Network	(100, 50), ReLU, lr 0.01	(100, 50), ReLU, lr 0.01	(100, 50), ReLU, lr 0.001

Equation Formatting

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2$$

Where n is the number of examples in the evaluation set and y_i is the actual opening price of example i. The term $\widehat{y_i}$ is the price the model predicted for that example.

Discussion

The main finding of this study is that the simplest model tested performed best. Linear regression achieved the lowest test error for Google, Apple and Tesla. It finished ahead of the random forest and the neural network and well ahead of the unpruned decision tree.

Over a three day window, there is very little non linear structure available to capture. Daily opening prices are highly autocorrelated meaning tomorrow's opening price is, to a good approximation, today's price plus a small change. A linear model fitted to three consecutive prices can express this relationship almost exactly. Therefore the added flexibility of the other models has nothing useful to fit and instead fits noise. This interpretation is consistent with the observation that linear regression's training and test errors were nearly identical. That is what one expects when a model's capacity matches the complexity of the underlying relationship.

The decision tree result illustrates overfitting in an unusually clean form. A training MSE of exactly 0.00 alongside the worst test error of any model is not an unconventional case. Trees partition the input space into regions and predict a constant within each region. Without a depth limit, the algorithm continues splitting until each training example occupies its own region. The training predictions then become the training targets. Nothing has been learned about the relationship, the data has been stored. The same model then produced roughly twice the test error of linear regression confirms that a perfect fit to training data carries no guarantee about new data. Constraining the tree to depth 5 substantially closed this gap. This supports the

interpretation that unconstrained capacity, rather than the tree method itself, was the source of the problem.

There are several limitations. First, the input features were restricted to three consecutive opening prices. Trading volume, closing and high prices, technical indicators, macroeconomic variables and sentiment data were all excluded. It is entirely possible that the flexible models would outperform linear regression given richer inputs. With more features there would be non linear interactions for them to exploit. Thus, the result should be read as specific to this minimal feature set. Second and more seriously, the data were split randomly rather than chronologically. The input windows overlap and the series is strongly autocorrelated, so a randomly chosen test example may sit between two training examples in time. The model may then be interpolating within a period it has already seen rather than forecasting forward. It likely makes all reported errors optimistic. A chronological split, in which the test set consists entirely of dates after every training date, would give a more honest estimate. Third, only three companies were analyzed, all large capitalization US technology firms. The results may not extend to other sectors, smaller companies or other markets. Fourth, MSE in squared price units is not comparable across stocks at different price levels. This is why the Google, Apple and Tesla error magnitudes cannot be read against each other. Finally, a model that predicts the price level accurately is not necessarily useful, because predicting that tomorrow's price will be close to today's is both easy and uninformative. A more demanding test would be predicting the direction or magnitude of the daily change.

Despite these limitations, model complexity is not the same as model quality. The ordering of results here was simplest model best, most flexible unconstrained model worst. That is a useful counterweight to the assumption that more sophisticated methods necessarily produce better forecasts. It also demonstrates the practical value of a three-way data split, since evaluating only on training data would have ranked the decision tree first rather than last.

Several directions follow naturally. The most important is repeating the comparison with a chronological train/test split, alongside a baseline that simply predicts today's opening price as tomorrow's. If the models cannot beat that baseline, their low MSE values reflect the autocorrelation of the series rather than any predictive skill. Beyond that, the feature set could be expanded to include volume and technical indicators. The target could be redefined as the daily return rather than the price level. The input window could be varied to test whether longer histories help. Architectures designed for sequential data, such as long short-term memory networks, could be included. Finally, the analysis could be extended to companies across several sectors to test whether the ranking of models is stable or industry-specific.

Works Cited

Bergmeir, Christoph, and José M. Benítez. “On the Use of Cross-Validation for Time Series Predictor Evaluation.” Information Sciences, vol. 191, May 2012, pp. 192–213, https://doi.org/10.1016/j.ins.2011.12.028.

Breiman, Leo. “Random Forests.” Machine Learning, vol. 45, no. 1, Oct. 2001, pp. 5–32.

Dietterich, Thomas G. “Ensemble Methods in Machine Learning.” Multiple Classifier Systems, vol. 1857, 2000, pp. 1–15, https://doi.org/10.1007/3-540-45014-9_1.

Fischer, Thomas, and Christopher Krauss. “Deep Learning with Long Short-Term Memory Networks for Financial Market Predictions.” European Journal of Operational Research, vol. 270, no. 2, Oct. 2018, pp. 654–69, https://doi.org/https://doi.org/10.1016/j.ejor.2017.11.054.

Hyndman, Rob J., and Anne B. Koehler. “Another Look at Measures of Forecast Accuracy.” International Journal of Forecasting, vol. 22, no. 4, Oct. 2006, pp. 679–88, https://doi.org/10.1016/j.ijforecast.2006.03.001.

IBM. “What Is a Neural Network?” IBM, 6 Oct. 2021, https://www.ibm.com/think/topics/neural-networks.

IBM. “What Is Random Forest?” Ibm.Com, IBM, 20 Oct. 2021, https://www.ibm.com/think/topics/random-forest.

ISLAM, MD TOUHIDUL. “Linear Regression Explained with Example & Application.” Statistical Aid: A School of Statistics, 5 June 2025, https://www.statisticalaid.com/linear-regression/.

Jayanth, Samyak. 15-Year Empirical Comparison of Deep Learning, Tree-Based, and Regression Models for Stock Market Forecasting. Polygence, Oct. 2025, https://doi.org/10.58445/rars.3334. Accessed 27 July 2026.

Khaidem, Luckyson, et al. “Predicting the Direction of Stock Market Prices Using Random Forest.” arXiv.Org, 2016, https://arxiv.org/abs/1605.00003.

Liu, Ting. “A Comparative Study of Transformer-Based and Classical Models for Financial Time-Series Forecasting.” Journal of Risk and Financial Management, vol. 19, no. 3, MDPI AG, Mar. 2026, p. 203, https://doi.org/10.3390/jrfm19030203. Accessed 27 July 2026.

Malkiel, Burton G. “The Efficient Market Hypothesis and Its Critics.” Journal of Economic Perspectives, vol. 17, no. 1, Apr. 2003, pp. 59–82, https://doi.org/10.1257/089533003321164958.

Nabipour, M., et al. “Deep Learning for Stock Market Prediction.” Entropy, vol. 22, no. 8, July 2020, p. 840, https://doi.org/10.3390/e22080840.

NoorMohammadzadehMaleki, Deniz, et al. “Artificial Intelligence Methods for Financial Market Prediction: A Systematic Review.” Computers and Electrical Engineering, vol. 134, Elsevier BV, June 2026, p. 111110, https://doi.org/10.1016/j.compeleceng.2026.111110. Accessed 27 July 2026.

P, Steve. “The Stock Market Explained: NYSE, NASDAQ, S&P 500 and DJIA - Slavic401k.” Slavic401k, 22 Mar. 2024, https://slavic401k.com/the-stock-market-explained/.

Patel, Jigar, et al. “Predicting Stock and Stock Price Index Movement Using Trend Deterministic Data Preparation and Machine Learning Techniques.” Expert Systems with Applications, vol. 42, no. 1, Jan. 2015, pp. 259–68, https://doi.org/10.1016/j.eswa.2014.07.040.

Stephens, Rob. “Decision Trees: Pick the Branch with the Highest Profit.” CFO Perspective, 29 Jan. 2025, https://cfoperspective.com/decision-trees-pick-the-branch-with-the-highest-profit/.

Strader, Troy, et al. “Machine Learning Stock Market Prediction Studies: Review and Research Directions.” Journal of International Technology and Information Management Journal of International Technology and Information Management, vol. 28, no. 4, 2020, https://scholarworks.lib.csusb.edu/cgi/viewcontent.cgi?article=1435&context=jitim.

Tashman, Leonard J. “Out-of-Sample Tests of Forecasting Accuracy: An Analysis and Review.” International Journal of Forecasting, vol. 16, no. 4, Oct. 2000, pp. 437–50, https://doi.org/10.1016/s0169-2070(00)00065-0.

Vijh, Mehar, et al. “Stock Closing Price Prediction Using Machine Learning Techniques.” Procedia Computer Science, vol. 167, no. 167, 2020, pp. 599–606, https://doi.org/10.1016/j.procs.2020.03.326.

Zou, Jinan, et al. Stock Market Prediction via Deep Learning Techniques: A Survey. https://arxiv.org/pdf/2212.12717.

“11.3 Neural Network Models | Forecasting: Principles and Practice.” Otexts.Com, https://otexts.com/fpp2/nnetar.html.

The Phenomenon of Billionaires in the 21st Century: Should the Billionaires Exist?
By Hanna Harchonak

Abstract

The presence of billionaires reflects an important socio-economic paradox, appearing at the same time as recognized as the symbol of advancement and innovation and the reason for growing inequality. The presented research explores the phenomenon of the extreme concentration of wealth and its potential impact on the society if redistributed. By using various statistics and analyzing arguments against the existence of billionaires, the research shows the tremendous opportunity cost of the current distribution of wealth. It presents a practical output in the form of an interactive calculator that allows modeling of the process of reallocation of billionaires' fortunes to help in overcoming pressing global problems. The results suggest that the resources possessed by billionaires provide an enormous potential for addressing various human welfare issues.

Introduction

The existence of billionaires is one of the most debated topics in modern society. On the one hand, many people view extreme wealth as a symbol of success and innovation. They argue that the possibility of becoming immensely rich motivates people to create businesses and develop new technologies that benefit everyone. This occurs because the prospect of extraordinary financial rewards serves as a powerful incentive for entrepreneurial risk-taking and innovation, driving individuals to invest capital, time, and intellectual resources into ventures that have the potential to generate substantial returns while simultaneously advancing societal progress. Furthermore, the philanthropic efforts of the ultra-wealthy are often highlighted as a positive force for solving global issues. In total, the 400 wealthiest people in the US billionaires — worth at least $3.8 billion each — have donated an aggregate $319 billion to charitable causes over their lifetimes. Overall, however, America's richest people are still not all that generous, relatively speaking. Their collective charitable giving equals just 4.6% of their total wealth—and that's down from 5% last year. Three-quarters of them have given less than 5% of their fortunes away, including 40% who have donated less than 1% of their riches. (Malmgren, 2025) On the other hand, critics question the morality and economic logic of such vast wealth concentration in the hands of a few. They point to the growing gap between the rich and the poor as a source of social injustice and political instability. Today, the poorest half of the world's population currently owns just 2% of global wealth, whereas the richest half owns 98%. This inequality is even more concentrated at the top, with the wealthiest 1% owning 38% of total wealth and the top 0.1% owning 19%, further widening the gap between the richest and the rest of the world. (Ritter, 2024) This raises a critical question: does this level of inequality actually harm our societies? The debate forces us to examine our core values about fairness, merit, and the structure of our economy. Ultimately, we must ask ourselves if a system that allows for extreme wealth is compatible with a healthy and equitable world for all.

The debate around billionaires is so essential today because it forces us to question the very foundations of our economic and moral systems. We are living in an era of unprecedented wealth concentration, where the resources of a few individuals rival those of entire nations, making their influence impossible to ignore. This topic is fiercely debated because it touches upon a fundamental conflict between the principle of meritocratic reward and the reality of systemic inequality. Many people see extreme wealth as a justified prize for innovation and risk-taking, believing it drives progress that benefits everyone.

On the other hand, a growing number of critics argue that such vast fortunes are often built upon societal infrastructures and labor forces that are not justly compensated. This is compounded by the visible struggles of average citizens with housing, healthcare, and debt, creating a stark and unsettling contrast.

The philanthropic efforts of billionaires further fuel the debate, as they are seen by some as generous and by others as a strategic tool to maintain power and avoid taxes. This practice raises profound questions about democracy, as it allows a handful of unelected individuals to shape public priorities through their personal giving. The core of the issue is whether this concentration of power and capital is healthy for a functioning society or if it represents a failure of our collective systems.

Ultimately, the existence of billionaires is not just an economic question but a deeply philosophical one, forcing us to define what we consider fair and what kind of world we want to build for the future. The conversation is urgent because it directly challenges us to choose between a system of private, charitable solutions and one built on public, democratic accountability for the common good.

The aim of the research is to critically analyze the phenomenon of billionaire wealth and to model the potential societal benefits of its redistribution. The study seeks to provide a data-driven perspective on the debate by quantifying the opportunity cost of extreme wealth concentration.

The goals of the research are

to systematize statistical data on the dynamics and scale of global billionaire wealth;

to analyze and contrast the primary arguments in the scientific discourse for and against the existence of billionaires;

to design and develop an interactive calculator as a practical tool for modeling wealth redistribution;

to visually demonstrate alternative resource distribution.

The object of the research is the socio-economic phenomenon of extreme wealth concentration as manifested in the global billionaire class and its impact on modern society.

The subject of the research is the potential for the alternative allocation of resources, currently held by billionaires, to address a defined set of pressing global challenges.

The hypothesis: the resources concentrated in the hands of billionaires represent a significant untapped potential for global transformation, the mobilization of which is possible through mechanisms of progressive taxation.

The expected result of this work is the development of a functional, web-based interactive calculator. This tool will allow users to select a billionaire's net worth or input a custom sum and visually model its reallocation towards solving specific issues, such as eradicating hunger, providing clean water, or funding global education. Furthermore, the research is expected to yield a comprehensive analysis demonstrating that the aggregate wealth of billionaires represents a sufficient resource base to make substantial progress on multiple fronts of human development.

PART I: Theoretical and methodological foundations of the study of the phenomenon of billionaires

1.1 Definition of the main concepts

The success of any scientific research begins with a clear definition of its basic categories.

Billionaire: An individual whose net worth exceeds one billion units of a currency, typically the US dollar. In the context of this work, this concept transcends mere wealth indication and characterizes a subject possessing exceptional economic resources and influence. (Wikipedia, 2026)

Social Inequality: The uneven distribution of resources, opportunities, and burdens among individuals and groups in society. (EBSCO, 2022)

Justice: Concepts of fairness, equality, moral behavior, lawfulness, and order. (Soken-Huberty, 2026) In an economic context, two main types are distinguished:

Commutative Justice: A principle of commutative justice specifies how individuals should be treated in a given class of actions and transactions.

Distributive Justice: A principle of distributive justice specifies how things such as rights, goods and well-being should be distributed among a class of people. (The Open University, 2026)

Redistribution: Modifications of the holdings of particular persons, collective agents, or groups (as defined in terms of non-resource holding characteristics), or changes in holdings by groups (as defined by resource holdings). (The Stanford Encyclopedia of Philosophy , 2004)

It’s vital to comprehend four important ideas about money and society: billionaire, social inequality, justice, and redistribution. It's also significant to look into the core - how some people became so wealthy throughout history.

1.2 The scale of a billion

We often hear the words "million" and "billion" in the news, but their true scale is almost impossible for the human brain to grasp. Converting these abstract numbers into time and money reveals just how vast that chasm truly is. The human brain struggles to comprehend such enormous numbers as one billion, and the difference between a million and a billion is a classic example.

To put it into perspective:

- A million seconds is just 12 days. A billion seconds is approximately 31.7 years. A trillion seconds is 31,688 years. (McCarthy, 2021)
- A stack of one million US $100 bills would be approximately 1.1 metres tall — roughly to the waist of an average person, while a stack of one billion US $100 bills would be approximately 1,100 metres tall. This is taller than the Burj Khalifa (828 metres), the tallest building in the world.
- Earning $100 every single day, it would take over 27,000 years to accumulate $1 billion. This time span is longer than all of recorded human history.
- Spending $10,000 every single day without fail, it would still take 274 years to spend $1 billion.

These comparisons reveal a powerful truth: a billion is not simply one thousand times a million in a numerical sense; it represents a difference in kind. It's a timescale that spans civilizations and a sum of wealth that transcends individual human effort.

The difference between a millionaire and a billionaire is of kind. One can, through hard work, talent, and luck, become a millionaire. Becoming a billionaire, however, is an event of a different order—one that is mathematically impossible through labor income alone. It requires the systemic capture of value generated by thousands of others, the ownership of assets that appreciate exponentially, and very often, operating within a political and economic framework that permits such extreme accumulation.

1.3. How to become a billionaire

The notion of a "self-made" billionaire who accrued wealth solely through their own genius and hard work is a powerful cultural myth. However, a systematic analysis of the pathways to such extreme wealth reveals that accumulation on the scale of billions is structurally inseparable from exploitation and tax avoidance.

The accumulation of such significant fortunes became possible due to a confluence of historical and economic factors. The transition from an agrarian to an industrial economy created unprecedented opportunities for capital accumulation. This shift saw the most important, and thus highly remunerated, functions move from warriors and priests to bankers, entrepreneurs, and engineers.

In the 20th and 21st centuries, the emergence of global markets, the development of financial instruments, and, in particular, the growth of the technology sector enabled individual companies and their founders to achieve global scale and extract super-profits in record time. The fortunes of individuals like Elon Musk and Jeff Bezos are a direct consequence of these processes.

Extreme wealth doesn’t emerge from thin air—it’s extracted. From workers in mines, fields, and factories. Undocumented labor in subcontracted warehouses. Government lobbying that keeps enforcement weak. Supply chains that are complex by design, where responsibility is always someone else’s problem.

While billionaires enjoy superyachts and stock options, the human cost of their empires is hidden—subcontracted out, pushed underground, or denied outright. Most face no consequences. Many stay silent. The rest pay lawyers and PR firms to make the outrage disappear.

Some of the world's richest men are linked to companies facing lawsuits, investigations, and exposes over serious allegations of forced labor and exploitation:

Elon Musk, the world's richest man, leads Tesla — a company that continues to source cobalt from suppliers linked to forced child labor in the Democratic Republic of Congo. Though Tesla was sued by Congolese families whose children were killed or maimed in cobalt mines, the company has fought accountability at every step and refuses to disclose its full list of suppliers.

Bernard Arnault, CEO of LVMH, oversees a luxury empire built on exclusivity — and reportedly, exploitation. His label Loro Piana was placed under court oversight in Italy after authorities found undocumented workers sleeping on site and forced to work 90-hour weeks for as little as $5/hour. The brand denied responsibility, but the courts didn't.

Jeff Bezos, founder of Amazon, built a trillion-dollar empire on the backs of underpaid warehouse workers, punishing quotas, and opaque supply chains. Investigations have found that some Amazon-sold products originate from factories flagged for participation in China's state-sponsored labor transfer programs where Uyghurs are shipped around the country and forced to work— yet the company has refused to publish a full list of suppliers or take adequate steps to verify labor conditions.

Amancio Ortega, billionaire owner of Zara's parent company Inditex, has grown rich off fast fashion — an industry built on speed, volume, and outsourced labor with little oversight. Zara's parent company, Inditex, is under investigation in both France and Canada for allegedly benefiting from Uyghur forced labor in its supply chains. Neither Ortega nor Inditex has taken credible public steps to trace or clean up Zara's supply chain. (Freedom United, 2026)

While tax evasion may be a rite of passage for the uber wealthy, it undermines the ability of governments everywhere to fund education systems, provide health care, establish and take care of basic infrastructure, and fund other essential duties that people depend upon.

There's no such thing as a self-made billionaire. All billionaires became rich because of economic and social conditions that provided the resources, labor, and legal frameworks in which their wealth accrued. Billionaires took extraordinary advantage of the social contract that people within a society abide by; tax evasion thwarts that very social contract and makes it harder for other people to gain even the most basic of human rights. (McCarthy, 2021)

In theory, people with higher incomes are taxed at a higher rate, since they have such an abundance of wealth. But in our corrupt reality, the vast majority of these people are being taxed way less than they should be. According to a study from the White House, America's 400 wealthiest families paid an average of just 8.2% in federal individual income tax between 2010 and 2018. By comparison, the average American household pays 13.3% with an income that doesn't come close to the country's wealthiest families. With 37.9 million Americans living in poverty, there is no explanation or justification for the wealth that these billionaires hoard. (Iacurci, 2021)

PART II: Billionaires in the global economy

2.1 Statistical portrait: the number and total wealth of the world's billionaires, their share in global wealth.

The contemporary landscape of wealth distribution is characterized by extreme concentration. A record 3,428 entrepreneurs, investors and heirs make this year's World's Billionaires list, 400 more than in 2025. Put another way, the planet added more than one new billionaire every day over the past 12 months. Even more astonishing is the money: as a group, the billionaire class is $4 trillion richer than last year. Their combined wealth now stands at a record $20.1 trillion. The average billionaire's fortune is now $5.8 billion, up from $5.3 billion in 2025. (Peterson-Whithorn et al., 2026)

The Bloomberg Billionaires Index is a daily ranking of the world's richest people. Details about the calculations are provided in the net worth analysis on each billionaire's profile page. The figures are updated at the close of every trading day in New York. (Chappatta et al., 2026)

Rank	Name	Total net worth	Country
1.	Elon Musk	$932B	United States
2.	Sergey Brin	$312B	United States
3.	Jeff Bezos	$290B	United States
4.	Mark Zuckerberg	$268B	United States
5.	Michael Dell	$219B	United States
6.	Mark Zuckerberg	$219B	United States
7.	Larry Ellison	$197B	United States
8.	Bernard Arnault	$165B	France
9.	Jensen Huang	$163B	United States
10.	Warren Buffett	$150B	United States

2.2 Dispelling myths about creating jobs and philanthropy

Proponents of extreme wealth concentration often deploy two powerful, emotionally resonant arguments: that billionaires are "job creators" and that their philanthropy compensates for any negative impacts. A closer examination reveals that both narratives are deeply flawed and often serve to legitimize a system that primarily benefits the wealthy.

Billionaires can't be regarded as great philanthropists. Philanthropy is held up as proof that billionaires are virtuous and that the system works, as wealth "trickles down" charitably. However, much of modern philanthropy is a strategic tool for managing public perception. We see "art-washing" (funding museums to distract from labor practices), "green-washing" (environmental donations to offset a polluting core business), and "health-washing" (pharmaceutical heirs donating while their company hikes drug prices). Philanthropy allows billionaires to craft a legacy of benevolence that obscures the often exploitative practices used to acquire their wealth.

Philanthropy is undemocratic and unaccountable: when a billionaire decides to fund a specific school, disease, or park, they are making a unilateral decision about public priorities. This is a fundamentally undemocratic act.

A billionaire may be passionate about a rare disease but indifferent to lead poisoning in inner-city water pipes. The result is that funding flows to their personal hobbies rather than to the most pressing, evidence-based public needs identified by communities themselves.

If billionaires paid taxes, philanthropy would be largely unnecessary. The scale of billionaire philanthropy is a tiny fraction of the public funding that is lost due to their tax avoidance.

A billionaire might donate $100 million to a children's hospital—a headline-grabbing act of generosity. Yet, if that same billionaire and their company had paid their fair share of taxes over the years, the public treasury would have collected hundreds of millions, or even billions, more dollars. This public revenue could fund not just one hospital, but the entire public healthcare system, including dozens of hospitals, clinics, and preventative care for everyone.

Philanthropy allows billionaires to choose which part of the public good they want to support, while simultaneously lobbying to defund the rest of it through lower taxes. It transforms a public duty (taxation) into a private, voluntary act (charity) for which they receive praise and influence.

An ordinary person considers a billionaire a job creator. The reality is far more complex and often contradictory. The argument ignores the fundamental process of how large corporations, the very engines of billionaire wealth, often grow. They don't just create jobs out of thin air; they frequently consolidate and eliminate existing ones. When a massive corporation like Walmart or Amazon moves into a town, it may employ 300 people. However, it does so by drawing customers away from dozens of existing small businesses—the local hardware store, the family-owned grocery, the independent clothing boutiques. These small businesses, which often provided stable, community-oriented jobs, are forced to close. The net result is often a loss of total jobs in the community and a shift from ownership and better-paid positions to lower-wage, less secure employment.

Billionaires are not philanthropists employing people out of kindness; they are capitalists whose businesses often grow by eliminating competing jobs. The quality, stability, and community benefit of the jobs they create are often far inferior to the ecosystem they displace. These two concepts work together to create a powerful, self-justifying narrative:

1. A billionaire builds a fortune by creating a large corporation that often destroys more jobs than it creates initially.

2. They then use their wealth and influence to minimize their tax burden, starving public services of funds.

3. The resulting crises in public health, education, and infrastructure create a need for charitable intervention.

4. The billionaire then steps in as a philanthropist, using a small fraction of their untaxed wealth to address a sliver of the problem they helped create, receiving widespread praise and a reputation as a "job-creating philanthropist."

5. This enhanced reputation protects them from public scrutiny and allows the cycle to continue.

By deconstructing these myths, it is evident that the defenses of billionaires are sophisticated tools used to perpetuate the system that creates them.

2.3 Drawbacks of billionaires: environmental impact, ethical issues of resource concentration, tax evasion

Wealth begets wealth. Billionaires derive their income from returns on capital - investments, stocks, real estate, which historically grows faster than the economy. Meanwhile, the majority of people rely on income from labor, which has largely stagnated. This structural dynamic ensures that the gap between the capital-owning class and the working class widens naturally.

The concentration of wealth in the hands of a few drives up the price of real estate and stocks. The impact of billionaires extends far beyond their personal bank accounts. Their existence and actions have profound, often negative, consequences for the health of our democracies, our planetary ecosystem, and the very fabric of our societies. The concentration of resources on this scale creates a feedback loop that undermines democratic institutions, accelerates environmental decay, deepens social stratification, and poses fundamental ethical questions about justice and power.

Democracy is founded on the principle of "one person, one vote." However, extreme wealth concentration creates a parallel system of influence that operates on the principle of "one dollar, one vote," effectively transforming democracy into a plutocracy—a system ruled by the wealthy.

Billionaires and the corporations spend vast sums on lobbying to shape legislation in their favor. This includes pushing for tax cuts, deregulation of their industries, and subsidies for their businesses. For example, the fossil fuel industry, heavily backed by billionaire owners, has spent billions over decades to lobby against climate action and protect its profits.

Many billionaires own or have major influence over media outlets - Jeff Bezos owns The Washington Post, Rupert Murdoch owns Fox News and The Wall Street Journal. This allows them to set news agendas, frame public debates, and promote ideological narratives that legitimize their wealth and power while marginalizing critiques. This control over information is a direct threat to a well-informed citizenry, which is essential for a functioning democracy.

Monopoly power is escalating extreme wealth and inequality worldwide; in fact, Oxfam estimates that 18 percent of the world’s billionaire’ wealth comes from monopoly sources. Monopolistic corporations can control markets, set the rules and terms of exchange with other companies and workers, and set higher prices without losing business. These monopolies have a

huge role in shaping the lives of ordinary people around the world, from influencing how much we are paid, to the foods we eat and can afford, and the medicines we can access. (Amladi, 2025)

Billionaires have an ecological footprint that is orders of magnitude larger than that of the average person, and their business models are often intrinsically linked to environmental degradation.

The lifestyles of the ultra-wealthy are incredibly resource-intensive. A single trip on a private jet emits more CO2 in a few hours than the average person does in a year. Their mega-yachts, multiple mega-mansions, and space tourism hobbies represent a level of consumption that is ecologically obscene in a world facing climate collapse.

The primary environmental impact is not from their personal consumption, but from their investments. Many billionaire fortunes are rooted in inherently polluting industries like fossil fuels, mining, and industrial agriculture. Their drive for perpetual growth and profit discourages the long-term, sustainable investments needed for a green transition, as these may offer lower short-term returns.

As mentioned in the political context, billionaires in carbon-intensive industries use their wealth to fund climate denial think tanks, lobby against environmental regulations, and block policies that would threaten their assets. They are not passive bystanders in the climate crisis; they are active, powerful agents working to delay and dilute necessary action.

Many billionaires grew their wealth and fame for important inventions, like Elon Musk with Tesla's commitment to clean energy or Bill Gates' groundbreaking Microsoft technology. But these innovations don't excuse the lines these companies are willing to cross to maintain and expand their wealth. Especially when their inventions often fall short of their intentions — Tesla is electric but relies on lithium extraction that is harmful to the environment. (Ventura, 2023)

Many of the world's billionaires own, control, shape and financially profit from processes that generate carbon pollution. They benefit when corporations block progress on a fast and just transition to renewable energy while the people in low-income countries who produce the least climate pollution suffer the greatest consequences.

Billionaires are a powerful engine driving inequality. Their wealth grants them access to investment opportunities, tax avoidance strategies, and political influence. Billionaires and their investment funds buy up housing, they price out ordinary people, turning a basic human need into an unattainable luxury and a speculative asset for the rich.

By avoiding taxes, billionaires deprive governments of the revenue needed to fund high-quality public education, healthcare, and infrastructure. This creates a two-tier society: one where the wealthy can buy private services, and another where everyone else is left with a crumbling public system. This lack of public investment entrenches poverty and limits social mobility, locking in inequality across generations.

The existence of billionaires raises a fundamental ethical question: Is it morally justifiable for a single individual to control a quantity of resources so vast that it rivals the annual budget of a medium-sized nation? Such concentration grants a level of power that is fundamentally unaccountable. A billionaire can single-handedly decide the fate of a news outlet, fund a

movement to dismantle public education, or launch a vanity project into space. These decisions, which affect millions, are made without any democratic mandate, public debate, or oversight. This is antithetical to the core principles of a free and equal society.

Their vast financial resources allow them to live in a bubble, shielded from the consequences of their actions. The existence of billionaires, therefore, is a central node in a system that actively undermines democracy, damages the planet, deepens social divides, and challenges our most basic ethical notions of fairness.

PART III. Development of an alternative global resource distribution calculator

3.1 Concept and functional requirements for the calculator

The Money Redistribution Calculator (Harchonak, 2025) is an interactive web tool designed to vividly illustrate the scale of billionaire wealth and the potential impact of these resources on solving global problems. This tool can help people to visualize abstract sums and transform statistics into tangible, understandable scenarios, thereby enhancing users' financial and social literacy.

To function successfully the tool requires the database of the world's wealthiest people – it means integration with the current Forbes ranking, including names, photographs, and the net worth of each billionaire.

The tool has got a flexible sum selection. It offers the ability to select any billionaire from the list (with automatic insertion of their wealth) or to manually enter an arbitrary sum for experimentation. The user can "spend" the selected sum on various global initiatives and see it on "Spending" Interface. Each initiative has a clear calculation for the cost of one unit of aid.

The interface displays in real-time how many units of aid can be provided with the chosen sum, as well as the percentage of the billionaire's wealth that would be utilized. (Supplement 1)

3.2. Conflicting priorities: global needs vs. luxury expenditure

The staggering concentration of wealth among billionaires creates a morally significant duality in global resource allocation, where the capital required to solve humanity's most pressing crises is diverted towards unprecedented levels of personal luxury and consumption. This discrepancy represents a fundamental conflict of priorities between addressing foundational human needs—such as universal access to clean water, sanitation, and education, for which clear, achievable cost estimates exist—and the funding of an exclusive, hyper-consumerist lifestyle. The expenditure on private jets, multiple estates, and space tourism reflects the opportunity cost of foregone global stability, public health, and ecological sustainability, highlighting a profound misalignment between the flow of capital and the universal requirements for human dignity and survival.

Not many people can single-handedly transform the world — but billionaires can. Billionaires collectively have trillions of dollars parked in private foundations and donor-advised funds, tax havens, assets, and investment funds.

An Oxfam analysis prior to the pandemic found that billionaires own as much wealth as the poorest 4.6 billion people combined, and the richest 22 men own as much wealth as all of the women in Africa. Women and girls put in 12.5 billion hours of unpaid care work each and every day —a contribution to the global economy of at least $10.8 trillion a year, more than three times the size of the global tech industry. Getting the richest one percent to pay just 0.5 percent extra tax on their wealth over the next 10 years would equal the investment needed to create 117 million jobs in sectors such as elderly and childcare, education and health. (Oxfam, 2020)

Liquidating just their pandemic earnings would be enough to ease the daily suffering of billions of people and protect biodiversity at a time when millions of species face extinction. In fact, helping all countries comfortably adapt to climate change would cost $1.8 trillion - a clean transfer of wealth that would prevent hundreds of millions of people from dying and becoming climate refugees. In doing so, they'd be saving countries $7.1 trillion in climate costs. For a lower price tag, $6 billion could prevent 41 million people from starving this year. For less than $10 billion, billionaires could fully fund the UN's refugee programs, helping to resettle some of the most vulnerable people in the world. (McCarthy, 2021)

The divide between global need and wealthy spending cannot only be seen as a matter of size but also of ethics. The focus of the underlying controversy involves the acceptance of universal rights for all people to have enough food, receive medical assistance, and have chances to succeed and the phenomenon of turning exclusivity and extravagance into a commodity. The international society is busy looking for money to finance preparations for future pandemics, climate changes, and primary education, while one trip to space of one wealthy individual may already be enough for a long list of people to be treated against the virus and vaccinated. A super yacht can be seen as not only a boat but also hundreds of schools and hospital networks that could have been built instead of buying this boat.

Patterns of consumption among billionaires are particularly marked by an inclination towards material symbols of status and acquisition of lands. Large land ownerships, such as private islands and extensive rural estates, are regarded as investments and places of rest. The establishment of such properties often neglects the natural environment and has been shown in many cases to push vulnerable people out of their territory. The transportation choices of the wealthiest people – use of private jets, sea-going yachts, and the use of cars providing extremely high levels of emissions – are performed at a level many times higher than the average citizen. The development of those lifestyles requiring complex infrastructure, from special garages with climate control devices to pools heated throughout the year, consumes huge amounts of electricity that is mostly generated with the use of fossil fuels and thus contributes to global warming.

The way in which this spending is done is not only excessive, but also detrimental to the ecology and society as well. In terms of the carbon produced, one oligarch can create enough

emissions through travelling and maintaining his estate to equal the emissions of many dozens or even hundreds of ordinary households, yet this issue is rarely described in such terms. On the contrary, it becomes a part of the regular business statistics. If a part of this expenditure were redirected in the direction of renewable energy, implementation of new environmental practices, and other ecological activities, the results could be significant in terms of improving the environment and people's welfare.

3.3. Description of the calculator creation process

The practical component of this research involved the development of an interactive web-based application - The Money redistribution calculator. The primary objective of this tool was to translate abstract wealth data into tangible, relatable scenarios that visually demonstrate the opportunity cost of extreme capital concentration. The development process was structured around three core components: the technology stack, the algorithmic logic, and the user interface design. The code was authored and iteratively refined using Sublime Text as the primary code editor. Given the application's front-end nature, foundational code structures were adapted from open-source examples and tutorials for basic web development, which were then extensively customized to meet the specific functional requirements of this project. Upon completion, the final codebase was version-controlled and publicly hosted using the GitHub platform. Leveraging GitHub Pages, a dedicated live website was generated, providing a stable and accessible URL for user interaction and demonstration, thereby ensuring the tool's replicability and transparency for peer review and public engagement.

The application's logic was built on a client-server architecture. The front-end interface was constructed using HTML5, CSS3 for styling, and JavaScript (ES6+) to manage user interactivity. Dynamic data visualizations, such as progress bars and impact counters, were implemented to provide immediate user feedback. The back-end logic, responsible for processing user inputs and performing calculations, was handled directly by client-side JavaScript functions. The core algorithm is straightforward yet powerful: it takes a user-selected wealth value (either from a pre-populated database of billionaires or a custom input), divides it by the pre-defined cost of a selected global initiative (e.g., how many people could be fed, schools, hospitals and wells could be built), and returns the number of times that initiative could be fully funded. The data on billionaire net worth, sourced from Bloomberg.com [22] was structured into a JSON (JavaScript Object Notation) database for efficient access and manipulation by the application. (Supplement 1)

Conclusion

The interplay between wealth creation, social inequality, and justice frameworks represents one of the most significant challenges in modern societies. Understanding the historical context of wealth accumulation and the mechanisms of redistribution helps us engage in informed discussions about creating economic systems that balance innovation, fairness, and

social welfare. These conversations are essential for building societies that both reward achievement and ensure dignity and opportunity for all members.

Ultimately, the question is not only one of economic feasibility but also of the ethical choice of society. Justice requires that social and economic inequalities be to the greatest benefit of the least-advantaged. In this light, the resources of billionaires are a test of humanity's capacity for solidarity and the rational management of common resources to create a more equitable future.

To make this abstract wealth tangible, the creation of an interactive "Money Redistribution Calculator" was applied. This tool allows anyone to see the life-changing potential of these concentrated resources. The calculator shows that the cost of another billionaire's superyacht or a suborbital joyride could add up to the cost of solving global problems. It visualizes the breathtaking opportunity cost of extreme luxury.

However, the very existence of this calculator demonstrates one undeniable truth: a system that allows a single individual to amass this much power and wealth is a broken system. It is a system that prioritizes the whims of the few over the needs of the billions. The problem is that we're not operating in a regulated capitalist environment. We're living in a rigged capitalistic environment where a handful of companies do not have to follow the rules. But ordinary people have to follow the rules. Therefore, we must embrace the principles of limitarianism—the ethical belief that no individual should possess wealth beyond a certain limit. This can be achieved through robust, global policies like a significant wealth tax, the closure of tax havens, and strict anti-monopoly regulations. We must build a system that does not allow a modern Marie Antoinette to suggest cake, but one that ensures everyone has bread.

Work Cited

Malmgren, Ella. America's Richest Billionaires Keep Getting Richer. They're Not Donating Much More, 9 Sept. 2025, www.forbes.com/sites/ellamalmgren/2025/09/09/americas-richest-billionaires-keep-getting-richer-theyre-not-donating-much-more/.

Ritter, Ilka, et al. UN DESA Policy Brief No. 168: Net Wealth Taxes: How They Can Help Fight Inequality and Fund Sustainable Development, Financing for Sustainable Development Office, UN DESA, 4 Dec. 2024, desapublications.un.org/policy-briefs/un-desa-policy-brief-no-168-net-wealth-taxes-how-they-can-help-fight-inequality-and.

Wikipedia. (2026). Billionaire - Wikipedia. Billionaire. https://en.wikipedia.org/wiki/Billionaire

EBSCO. (2022). Social inequality | law | research starters | ebsco research. Social inequality. https://www.ebsco.com/research-starters/law/inequalitysocial-inequality

Soken-Huberty, E. (2026). What is justice? | human rights careers. What Is Justice? https://www.humanrightscareers.com/issues/what-is-justice/

The Open University. (2026). Rights and justice in international relations. https://www.open.edu/openlearn/mod/oucontent/view.php?id=4743&section=4.1

The Stanford Encyclopedia of Philosophy. (2004). Redistribution (Stanford Encyclopedia of Philosophy). Redistribution. https://plato.stanford.edu/entries/redistribution/index.html

Freedom United. (2026). Stop billionaire-backed forced Labor. Stop billionaires profiting from exploitation. https://www.freedomunited.org/advocate/stop-billionaire-backed-forced-labor/

Iacurci, G. (2021). America's richest 400 families pay a lower tax rate than average taxpayer. America's richest 400 families pay a lower tax rate than average taxpayer. https://www.cnbc.com/2021/09/23/americas-richest-400-families-pay-a-lower-tax-rate-than-average-taxpayer.html

Peterson-Whithorn, C., Chung, G., & Durott, M. (2026). Forbes 2026 billionaires list - the richest people in the world ranked. World's Billionaires List. https://www.forbes.com/billionaires/

Chappatta, B., Heathcote, A., Mak, P. Y., Witzig, J., & Maloney, T. (2026, July 8). How are billionaire and corporate power intensifying global inequality?. Bloomberg Billionaires Index. https://www.bloomberg.com/billionaires/

Amladi, D. (2025). How are billionaire and corporate power intensifying global inequality? https://www.oxfamamerica.org/explore/stories/how-are-billionaire-and-corporate-power-intensifying-global-inequality/

Ventura, I. (2023). There's no such thing as an ethical billionaire - san francisco foghorn. There's no such thing as an ethical billionaire . https://sffoghorn.com/theres-no-such-thing-as-an-ethical-billionaire/

Harchonak, H. (2025). Money Redistribution Calculator. https://harchonakhanna-ux.github.io/money-calculator2.0/

Oxfam. (2020). World's billionaires have more wealth than 4.6 billion people. https://oxfam.org.ph/worlds-billionaires-have-more-wealth-than-4-6-billion-people/

McCarthy, J. (2021, July 26). 6 things billionaires keep spending money on instead of helping solve the world's biggest problems. 6 Things Billionaires Keep Spending Money on Instead of Helping Solve the World's Biggest Problems. https://www.globalcitizen.org/en/content/what-billionaires-spend-money-on-instead-charity/

Algorithmic Modeling of Soccer Salaries By Neil Kapoor

Abstract

This project aimed to predict soccer players' salaries based on multiple factors. It utilized machine learning to do so, assigning a training case so that the models could be trained for a testing set. MLS player data was used and imported from Kaggle, that included the input of first name, last name, the club a player plays for, their position, and their guaranteed compensation. Guaranteed compensation was then removed as a feature ablation, as the models were relying heavily on this part of the input to make predictions. The results of five different models, linear regression, decision tree, random forest, neural network, and KNN, were averaged together in a single ensemble model. Mean squared error was calculated to determine the accuracy of the predictions. Machine learning allowed for a dense analysis of the salaries without introducing inherent biases.

Introduction

Figure 1



This study focused on the intersection between metrics used to identify and evaluate a given soccer player, and the wage that soccer player is paid. In soccer, a player is paid based on their performance and what their apparent worth is on a specific team (Li et al.; Yaldo and Shamir). A player's guaranteed compensation is what they will be paid regardless of their performance and regardless of any performance-based bonuses they may have on their contract. There are several other factors that may have an affect on how good a player may seem to someone looking in from the outside. One such factor is the club a player is playing for. There are dozens of clubs in every continent, and each club varies in its respective competition and skill level. The level of a club depends largely on the league that club plays in. For example, the

Premier League, the highest level league in England, if not in the world, is home to twenty different clubs. Clubs competing in the premier league will naturally be some of the best in the world, such as Manchester City, Manchester United, Arsenal, and Chelsea, to name a few. There are also second and third division leagues in England, which are still professional level leagues, but just attract a lower level of play and players in general. Therefore, the club a soccer player is at is a huge indicator of their capabilities and skill (Aljohani and Alnahdi; Idris and Ng). Naturally, how much a player is paid is a large indicator of how good they are (Li et al.; Yaldo and Shamir). The position a player plays can also dictate how much they are paid (Aljohani and Alnahdi; Li et al.). For example, attackers are paid more on average than defenders.

A question would arise after analyzing the multiple factors that determine how "good" a player is: If wage is an accurate and proportional indicator of how well a player performs, how can factors such as the club a player plays for and the position they play in affect their salary? Patterns arise after some thought and analysis into this inquiry - European clubs tend to pay more on average than clubs elsewhere such as South America or Africa. Additionally, like mentioned before, attackers and midfielders get paid more than defenders. However, since this logic cannot be extrapolated in the sense that all attackers and midfielders are better players than defenders (because there are multiple ways to measure performance), salary is not always a direct link to how good a player is (Li et al.). An objective analysis of salary takes the prior out of the question and causes a new one to arise: how can the salary of a player be predicted given their first name, last name, their club, their position, and their guaranteed compensation?

Artificial intelligence, particularly machine learning, can be utilized in answering this question. Since the correlation between such factors and salary is so ambiguous and somewhat unknown, ML would be necessary to find patterns, if any, between variables to conduct an objective analysis. Machine learning would be useful in this case as there is no apparent link between the independent and dependent variables that could lead to an obvious regression problem, for instance ("Predicting Soccer Player Salaries"; Yaldo and Shamir). Additionally, soccer has so much publicly available data online regarding such statistics. In this study specifically, data from the American soccer league, MLS (Major League Soccer), was used. The input included the columns of the clubs, first name, last name, position, guaranteed compensation, and base salary, and the base salary column was removed after data cleansing. After training, testing and validation, ML models provided the output of the base salary of any given player, given those specific inputs. The results of several different trained models were compared and combined into an ensemble average for final predictions.

A recent 2025 study conducted a similar study - it predicted a player's salary with both traditional and automated machine learning approaches ("Predicting Soccer Player Salaries"). The name of the study was "Predicting Soccer Player Salaries with Both Traditional and Automated Machine Learning Approaches," and it largely used player performance as a predictor. They found that adding engineered features improved the performance of the predictions ("Predicting Soccer Player Salaries"). This study largely differed in that rather than combining the results of multiple different models, it focused on performance-based features and

automated model selection. Another study titled “Computational Estimation of Football Player Wages” discussed estimating player wages based on player characteristics like scoring performance, player behavior, physical abilities, and technical skills. It found a statistically significant relationship between predicted and actual salaries after using pattern-recognition algorithms (Yaldo and Shamir).

Results

Figure 2

Model	Training MSE	Testing MSE	Average Percent Error
Linear Regression	3,124,694,002	5,572,174,963	—
Decision Tree	171,443,062	1,947,833,480	—
Random Forest	336,692,077	2,287,354,559	**7.18%**
KNN	5,267,529,911	7,574,599,088	**36.60%**
Neural Network	2,102,191,531	5,216,918,055	**29.55%**
Ensemble	**1,143,886,520**	**2,654,293,999**	**15.83%**

Model Version	Average Percent Error
Original Ensemble	**15.83%**
Without Guaranteed Compensation	**122.50%**
Difference	**+106.67 percentage points**

Discussion

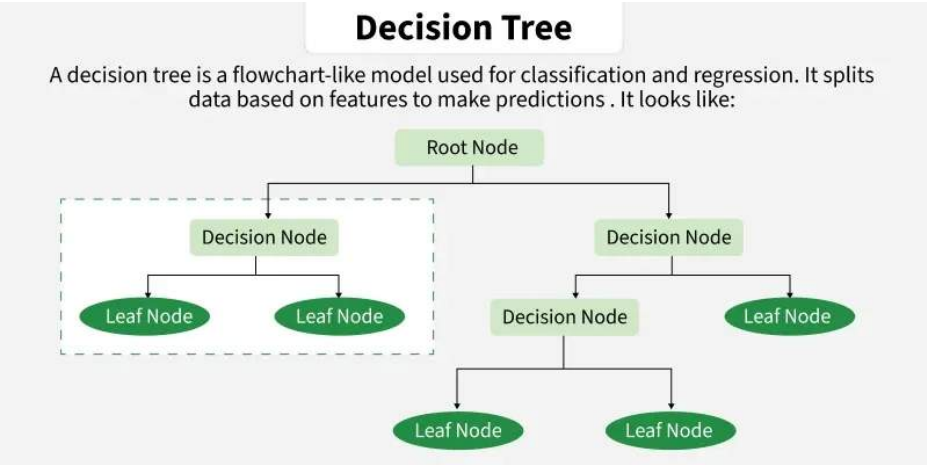
The decision tree model resulted in the lowest mean squared error, while the KNN showed the highest mean squared error. Interestingly, the random forest model had a higher testing MSE than the decision tree, but a much lower average percent error at 7.18% (figure 2). This could be because MSE penalizes for large errors, even if there are just a few of them, but percent error treats error equally relative to the true salary. Even though the random forest model was relatively accurate for most players, a few small misses increased its MSE substantially. Random forest would overall be well equipped for a salary prediction as the multiple decision tree pathways presented in this model leads to a greater ability to account for each input factor that would affect the output. One decision tree could train on the club of a player while the other would take guaranteed compensation into account, which could not be done with the simple

if-else branches present in a single decision tree. Moreover, a linear regression model would not be well equipped for this prediction as there is no linear correlation observed. A possible skewer of the results would be the guaranteed compensation as it is very closely related to base salary, which is why this input column was removed in a further iteration.

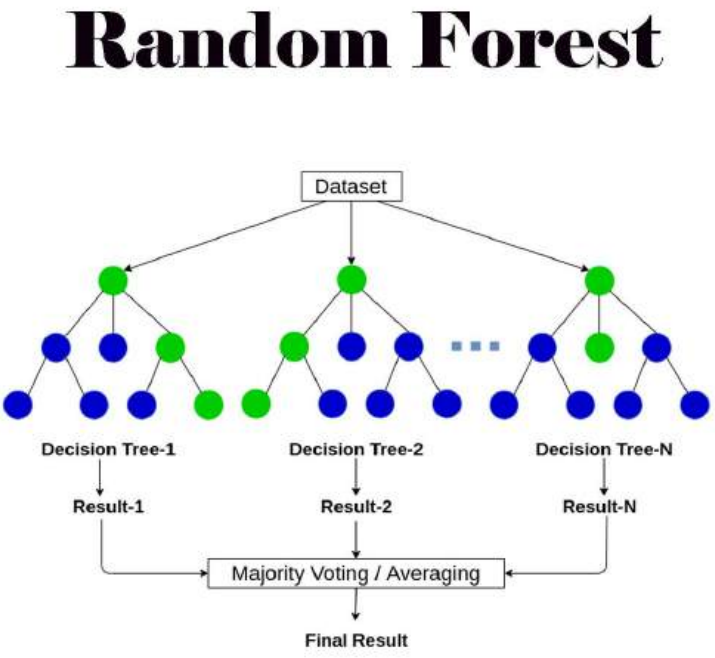
If this study was continued, a possible avenue of exploration could be predicting a given player's transfer value when they change clubs. This value is also a numerical indicator of how good a player is. Both the starting and ending clubs of a given player's transfer pathway would be passed as inputs to determine if their value would increase or decrease. This would also measure a likely amount a club would be willing to pay for the player. In this study, other factors such as a player's form at the moment they are leaving a club would have to be taken into account. If a given player is normally reliable and is the star player of their current team, but they are out of form come the transfer window, their value would decrease. A decision tree model would be useful in this analysis because the starting and ending club of a player represent multiple different branches that a prediction needs to account for.

Materials and Methods

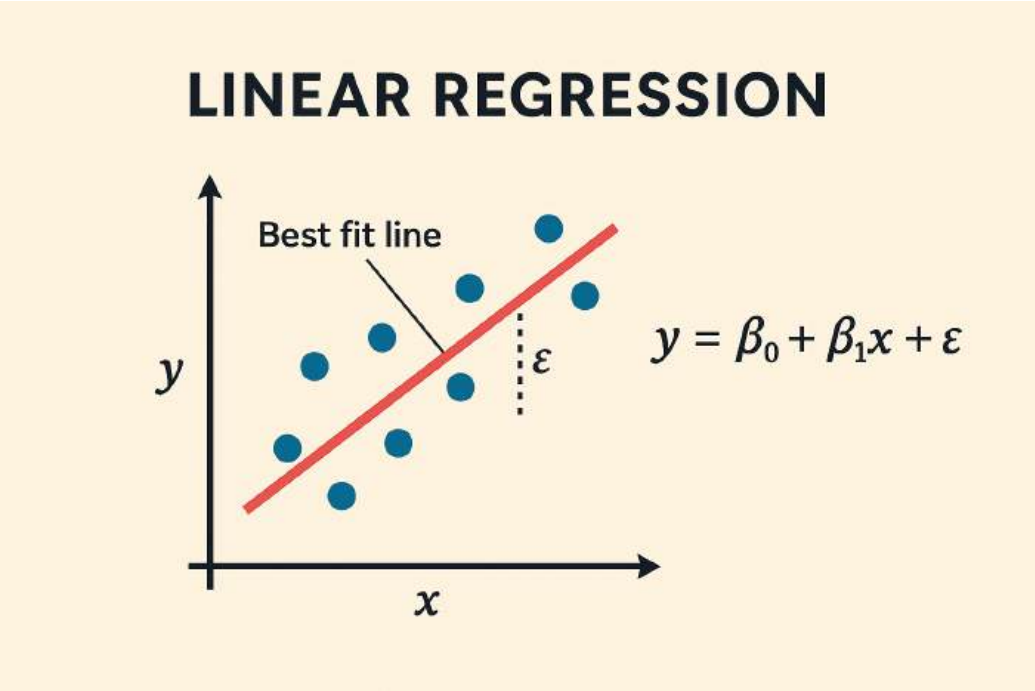
The following is the link to the dataset used:
https://www.kaggle.com/datasets/crawford/us-major-league-soccer-salaries.
The dataset included the columns of first name, last name, club, position, guaranteed compensation, and base salary. The base salary column was removed when the models were being trained, as this is what they are trying to predict. The program was coded in Python, with pandas being utilized to combine the data from the different seasons in the dataset. Missing values were also removed by reading every CSV file into one DataFrame. The x and y training values were targeted, and model.fit was used to train the models with these values. This was done for each model, and percent error was calculated at the end. Due to distribution shift, or the test cases being different from the training cases due to variance in the MLS dataset, some models such as the decision tree overfit. Dropout was then used to remove a part of the dataset randomly so that the model was forced to use a new pattern. This is when columns were removed at random from the input, so that the model did not focus too much on a specific feature (ex. guaranteed compensation). This study aimed to train five different models with an 80/20 training/testing split.



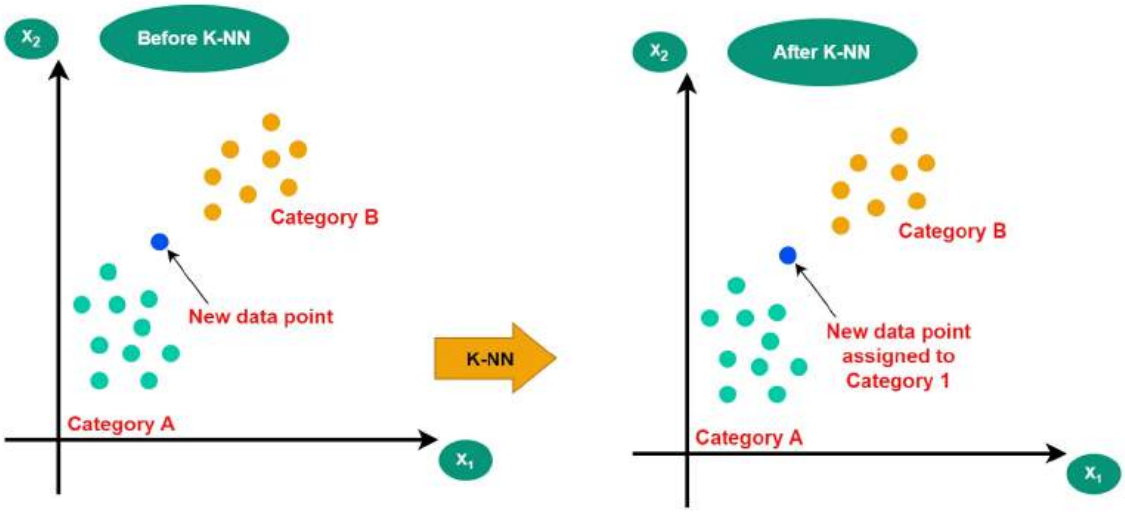
Decision trees are used for classification and regression with the goal of predicting a value for a target variable. One hyper-parameter of the decision tree is the max depth. After data is inputted, the model learns simple decision rules from the data features to help with prediction. Decision trees are also non-parametric. They use a set of if-else decision rules to narrow down on a result, specifying different characteristics of the data as it moves down on the upside-down tree. Not much preparation is needed for the data in this model, and it is able to handle multi-output problems. The different parts of the hierarchical tree structure within this model are the root node, branches, internal nodes, and leaf nodes. Internal nodes represent attribute tests, branches represent attribute values, and leaf nodes represent final decisions or predictions.



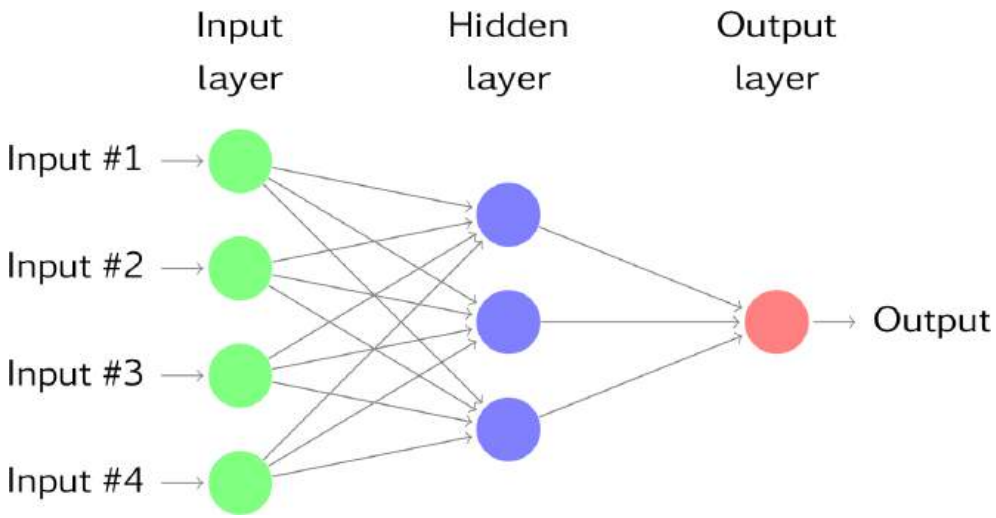
The random forest algorithm implements multiple decision trees to make predictions. The hyper parameters for random forest models are the number of decision trees used and the max depth of each. Each tree takes in data and provides an output, and the results of each tree are averaged together for the final prediction. Each tree looks at random and different parts of the data and averaging them together improves accuracy and reduces errors. Random forest can handle missing data, so preprocessing occurs before the algorithm is run. It can also handle large datasets with many features without losing accuracy or speed as the different trees can efficiently handle different parts of the dataset. Each tree makes its own predictions without relying on other trees and random data is used for each to reduce errors.



Linear regression is used to model the relationship between a dependent variable and one or more independent variables. It multiplies each input by a scale factor for a predicted outcome. Continuous values can be predicted as a straight line or line of best fit is fitted to existing data that then extends to predicted values. Error can be minimized using the least squares method which minimizes distances between actual and predicted values, called residuals. The line of best fit should best represent the data where the sum of the squared differences between predicted and actual values is as small as possible.



K-Nearest Neighbor (KNN) takes in no parameters and essentially memorizes a dataset to provide new data points. The hyperparameter for KNN models is the number of points, or neighbors, it takes into account. It identifies the K closest data points to a given input and makes predictions based on average values of these data points. It basically maps these average values of the neighbors onto the data point. It is classified as a lazy learner algorithm because it computes at the time of classification of the data and does not immediately learn from the training set. The value of k in KNN decides how many neighbors the algorithm looks at before making the prediction. If there is a lot of noise or outliers in the dataset, using a larger k can make predictions more accurate, but if k is too large, important patterns can be missed. There are multiple different distance metrics used in KNN such as euclidean distance, manhattan distance, and minkowski distance. They all have their own upsides and downsides; for example, minkowski distance is most flexible, while manhattan can be the most accurate with little outliers.



A neural network machine learning model stacks “neurons” in layers and learns biases from data based on patterns in order to map inputs to outputs. The hyper parameters for neural network models are the number of layers and the number of neurons per layer. For example, a neural network is used in detecting spam email. If an email is fed into the network, features like words or phrases would be used as inputs. Early neurons classify each signal based on importance and later layers of neurons combine information into higher-level cues such as those that capture context and tone. Finally, a last layer computes the probability of the email being spam, and if the probability is high enough, the email is flagged. Weights are one key concept of neural networks, and they act like dials that determine how strongly each input influences a decision. Biases are another key concept, and they are basically built-in values that shift a decision threshold. They allow a neuron to activate even if inputs are weak.

After training and testing each model, the predictions of all five models were combined into a single ensemble model. The five models were linear regression, decision tree, random forest, KNN, and neural network, each with their respective strengths and weaknesses. Combining the predictions of the five models allows for a greater range of input as well as a combination of the upsides of each model’s predictions. After one iteration, it was found that guaranteed compensation was easily the largest influencer in the models’ predictions, so removing it was tried in an attempt to test this hypothesis. When this data column was removed, average percent error increased by over 100%, confirming the fact that it is closely correlated with the salary of a player. The essence of a time series, something that changes over time in the field of AI, can also be used in evaluating results of the models. It takes into account how far a prediction is from being accurate, and assigns negative connotation to those predictions. A more effective system that would not place a cap on a very effective model could be to determine the expert model as more data points are gathered. Each time a model is accurate, it gets a point, so this straightforward system allows for each model to be accurately represented. If the data points were numerical values, the original rule would be to just take the average of the results from the different models rather than a majority vote. Going forward, a weighted average could just be taken based on the performance of the different models. This allows for each result to be properly considered while taking every model into account. Systems that eliminate the results from less accurate models would not be effective as it is not taking a combined prediction at the end, but rather that of the last model standing.The outputs of the models was evaluated using mean squared error (MSE). The equation used to calculate MSE was $\frac{1}{n}\sum_{i=1}^{n}(y - \widehat{y})^2$. This essentially takes the summation of the squared differences between the actual and predicted values (salaries, in this case), and divides them by n terms, giving the average.

The accuracy of different machine learning models are determined by their performance in the testing cases. To combine predictions of different models, it is important to take the performance and accuracy of each model into account so that their results can be weighted accordingly. For example, if there were 5 models which cannot be equally trusted, but it is unknown which model has the highest accuracy, their performance must first be tested. In the

first round of results, a majority vote can be used as there is no information to go off of. After a few more data points are gathered, a voting system can be implemented. Essentially, the models that are consistently accurate will be given a higher influence and will earn more points when they perform correctly going forward. While this system does place a cap on effective models, meaning they are shown only marginally better than models that may perform significantly worse, it gives weightage to each model's influence based on its accuracy.

A possible question regarding this study could be about the possible upsides and downsides of each model. To answer this, the decision tree model can be used as an example. It learns non-linear rules and interactions and memorizes player names in an attempt to make predictions, but as a result, it often overfits. Overfitting occurs when a model adapts too heavily to the training case and therefore shows better results when ran against the training case as opposed to the testing case, which is not ideal. Averaging the results of the linear regression model with the decision tree model can solve this as linear regression assumes linear relationships between the different features and salary. The multiplicative weights update algorithm was also used, in which the weight of the models that perform better in testing stages are increased and the weight of the models that perform worse in testing stages are decreased. A possible inquiry would be if each column of the input weighs equally towards the predictions. Data augmentation was used to address this, where weight was added to the minority class of the dataset and subtracted from the majority class.

Works Cited

Aljohani, A., and M. Alnahdi. “Exploring Football Player Salary Prediction Using Random Forest: Leveraging Player Demographics and Team Associations.” International Journal of Artificial Intelligence and Machine Learning, 2025, https://ijaim.net/journal/index.php/ijaim/article/view/115.

Dey, S. “Pricing Football Players Using Neural Networks.” arXiv, 2017, https://arxiv.org/abs/1711.05865.

Idris, M. I., and K. H. Ng. “Developing a Predictive Model for Football Players’ Market Value Using Machine Learning.” Journal of Informatics and Web Engineering, 2025, https://journals.mmupress.com/index.php/jiwe/article/view/1760.

Li, Y., S. Kampakis, and P. Treleaven. “Machine Learning Modeling to Evaluate the Value of Football Players.” arXiv, 2022, https://arxiv.org/abs/2207.11361.

“Machine Learning for NBA Salary Prediction: Integrating Human Capital and Compensation Theories.” Journal of Sports Analytics, 2025, https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artiId=ART003272779.

Papadaki, A., and M. Tsagris. “Estimating NBA Players Salary Share According to Their Performance on Court: A Machine Learning Approach.” arXiv, 2020, https://arxiv.org/abs/2007.14694.

“Predicting Soccer Player Salaries with Both Traditional and Automated Machine Learning Approaches.” Applied Sciences, vol. 15, no. 14, 2025, p. 8108, https://doi.org/10.3390/app15148108.

“Quantifying Football Player Value: A Novel Combined Rating and ML Ensemble Approach for Predicting Top Players.” Journal of Big Data, 2026, https://doi.org/10.1186/s40537-026-01369-w.

“When Interpretable Machine Learning Meets the Beautiful Game.” International Journal of Sports Science & Coaching, 2026, https://www.sciencedirect.com/org/science/article/pii/S2042678X25000290.

Yaldo, D., and L. Shamir. “Computational Estimation of Football Player Wages.” International Journal of Computer Science and Information Systems, vol. 13, no. 1, 2017, pp. 1–11, https://doi.org/10.1515/ijcss-2017-0002.

Yigit, T., et al. “An XGBoost-Lasso Ensemble Modeling Approach to Football Player Value Assessment.” Journal of Intelligent & Fuzzy Systems, 2020, https://doi.org/10.3233/JIFS-189098.

Appendix

Linktocode:https://colab.research.google.com/drive/13zIEiG0MnrtBtw4LFGRYQD73aCOa76Rz?usp=sharing

Culture Before Markets: Estimating the Economic Cost of Gendered Cultural Constraints in Romania from Original Survey and Interview Evidence By Jiayi Pan

Abstract

This paper argues that cultural expectations are not a secondary overlay on economic life but one of the mechanisms through which economic possibility is allocated in the first place. Building on an original Romanian-language questionnaire conducted in 2026 and follow-up interviews with Romanian women, the paper examines how family expectations, respectability norms, unequal care obligations, mobility constraints, and gendered assumptions about leadership and earnings shape educational, occupational, and income trajectories. The study is culture-first in emphasis: economics enters not as the primary object of inquiry but as a way of tracing the material consequences of cultural restriction. This framing is consistent with a wider literature that treats beliefs, norms, and institutions as persistent determinants of economic behavior and prosperity (Guiso, Sapienza, and Zingales; Fernández and Fogli; Nobel Prize Outreach, "Press Release").

The original survey suggests that the strongest losses are likely to occur before conventional economic indicators register inequality. Respondents frequently reported that women are steered toward caution, family-first choices, support roles, and reputational self-policing, while men face breadwinner pressure, higher-risk work expectations, and stigma around reduced earnings or flexible care roles. In the sample, respondents estimated that, on average, 2.70 out of 10 women they knew had reduced education, work, or business plans because of family or social expectations, compared with 0.85 out of 10 men; they also estimated that 4.77 out of 10 women and 2.96 out of 10 men could probably earn more under different family and social expectations (Jiayi, Romanian Culture and Economic Opportunity Questionnaire). The interview transcripts make those numbers concrete: one woman described losing a possible singing path because public performance was not seen as respectable for girls; another described household care as a second job; and a Bucharest CEO described women as judged for family outcomes even when responsibility was shared or unclear (Jiayi, Gendered Cultural Constraints in Romania: Interview Transcripts).

Using a simple scenario-based accounting model that assumes the sample reflects the broader Romanian population, this paper estimates annual worker-adjusted national losses of approximately RON 13.2 billion in a conservative scenario, RON 52.6 billion in a central scenario, and RON 123.7 billion in an upper-bound scenario. A broader potential-population specification yields RON 20.7 billion to RON 186.4 billion. These ranges should not be mistaken for national accounts. They are illustrative, theory-informed extrapolations intended to render culturally produced waste legible in economic terms. The main finding is conceptual as much as numerical: if culture narrows ambition, mobility, bargaining, care time, and network access before the market begins measuring wages, then the most expensive inequalities may be the ones that happen before the spreadsheet begins.

Keywords: culture, gender inequality, Romania, unpaid care, labor markets, opportunity cost, social norms

Introduction

Economists increasingly recognize that markets do not operate in a cultural vacuum. The 2024 Nobel Prize in Economic Sciences was awarded to Daron Acemoglu, Simon Johnson, and James A. Robinson "for studies of how institutions are formed and affect prosperity," with the Nobel materials emphasizing that persistent differences in societal institutions help explain durable differences in national prosperity (Nobel Prize Outreach, "The Prize in Economic Sciences 2024"). Yet institutions, especially at the everyday level, are often lived as culture: shared expectations about who should move, who should care, who should lead, who should earn, and who should remain cautious.

The 2023 Nobel Prize in Economic Sciences, awarded to Claudia Goldin, moved the same conversation into the labor market and the household. Goldin's work showed that women's labor-market outcomes are shaped by expectations, marriage, parenthood, care responsibilities, labor-market structure, and norms around career continuity. The Nobel background materials are explicit that unequal opportunity wastes labor and expertise, and Goldin's own research stresses that decisions made early in life can affect an entire career (Nobel Prize Outreach, "The Prize in Economic Sciences 2023"; Goldin, "A Grand Gender Convergence").

This paper takes those insights into the Romanian case and poses a narrower but deeper question: how much income may be lost when culture constrains economic choice before the market even records a wage? That question matters because many standard indicators capture downstream outcomes but not upstream narrowing. Eurostat defines the unadjusted gender pay gap as a difference in gross hourly earnings among employees and warns that such gaps must be interpreted cautiously because composition effects matter (Eurostat, "Gender Pay Gap Statistics"). If a society channels many women or men away from particular occupations, hours, locations, promotions, or networks before wage comparison begins, standard wage measures can understate culturally generated loss.

Romania is a particularly important setting for this question. Recent labor statistics indicate a marked gender difference in employment among those aged twenty to sixty-four. Eurostat reported Romania's 2025 employment rate for people aged twenty to sixty-four at 69.0%, below the EU average, and identified Romania as one of the EU countries with the largest gender employment gaps, at 18.7 percentage points (Eurostat, "Employment - Annual Statistics"). These macro facts do not by themselves reveal the underlying mechanisms. They tell us that inequality exists, but not whether it is produced primarily through wages, hours, educational sorting, care burdens, mobility restrictions, family expectations, or reputational discipline.

The central claim of this paper is therefore twofold. First, culture is economically consequential because it structures legitimate aspiration, mobility, bargaining behavior, and time use. Second, the most important losses in Romania may lie not only in unequal pay but in missed pathways: abandoned degrees, unasked raises, reduced paid hours, avoided late shifts, foregone

relocation, weakened networking access, and invisible unpaid labor. The paper makes this argument through original survey evidence, integrates the most revealing parts of three qualitative interviews, and then translates those mechanisms into an explicit loss model.

Literature Review

A large literature now treats culture not as a vague residual but as a set of beliefs and preferences that can be studied empirically. Guiso, Sapienza, and Zingales argued that economists had long been reluctant to use culture because of definitional breadth, but that improved data and methods made it possible to identify systematic differences in preferences and beliefs and connect them to economic phenomena. Their synthesis remains foundational because it reframed culture as testable rather than ornamental.

Subsequent work strengthened the causal claim. Fernández and Fogli showed that for second-generation American women, beliefs and role expectations transmitted from countries of ancestry retained explanatory power for work and fertility behavior even after controlling for education and spousal characteristics. Culture in this account is not reducible to current prices or institutions; it persists through expectations and learned norms about women's roles (Fernández and Fogli). Tabellini, meanwhile, found that historically rooted values such as trust, respect for others, and confidence in self-determination were strongly associated with current regional development in Europe, using historical instruments to isolate the exogenous component of culture.

The gender-and-work literature deepens this picture. Goldin's Nobel-recognized research shows that women's labor-market outcomes are shaped by expectations formed before labor-market entry, by marriage and childbearing, and by labor-market structures that disproportionately reward temporal inflexibility and constant availability. Critically, Goldin's work implies that observed pay gaps may reflect earlier expectation-formation and family-role sorting rather than only contemporaneous wage discrimination. This is especially important for a culture-first analysis because it suggests that economic outcomes are downstream from cultural role assignment (Goldin, "A Grand Gender Convergence").

Recent experimental evidence also shows that norms matter not merely through values but through what people believe others believe. Bursztyn, González, and Yanagizawa-Drott demonstrated that in Saudi Arabia, many married men privately supported women working outside the home but underestimated support among peers; correcting that misperception increased husbands' willingness to help wives search for work and increased women's applications for higher-paying outside-the-home jobs. The broader implication is that labor supply and occupational choice may be constrained by socially enforced scripts even when private preferences are more flexible.

This literature clarifies why a Romania-focused cultural study matters. Eurostat reports that the average unadjusted gender pay gap in the EU was 11.1% in 2024 and explains that the indicator is unadjusted, meaning that it captures a broad earnings difference before correcting for worker and job characteristics (Eurostat, "Gender Pay Gap Statistics"). Goldin likewise

emphasizes that unequal opportunity may arise because women's choices are shaped by expectations, care structures, and job design before wages are measured. In other words, a country can appear less unequal on one downstream metric while still wasting large amounts of human potential upstream.

Study Design, Data, and Method

The empirical core of the paper is an original forty-four-response Romanian-language questionnaire collected in May 2026 (Jiayi, Romanian Culture and Economic Opportunity Questionnaire). The sample is not nationally representative, but it is socially informative. Among respondents with non-missing values, the average age was 44.1 years; thirty-six of forty-three respondents were women; thirty-nine of forty-three reported full-time work; and thirty-one of forty-three reported children or regular care responsibilities. Reported unpaid care time averaged 15.2 hours per week among those who answered, with a median of 6 hours and a maximum of 80 hours.

The questionnaire asked about gendered family expectations, educational encouragement, field stereotypes, work ambition, promotion and bargaining, marriage and children, mobility and safety, networking access, unpaid household labor, and perceived income loss. Some questions were descriptive, such as whether respondents had heard particular gendered assumptions; others were interpretive, asking respondents to estimate how many out of ten men or women they knew had reduced education, work, or business plans because of family responsibilities or social expectations.

As part of the study, three semi-structured interviews were conducted with Romanian women aged forty-seven, forty-eight, and sixty. The interviewees were a saleswoman from a small city, an educator from a big city, and a CEO from Bucharest. The interviews were not designed to replace the survey or to claim national representativeness. They are used as qualitative mechanism evidence: they show how broad survey categories such as reduced plans, unpaid care, reputational pressure, and family-status penalties are expressed in concrete life histories.

The transcripts were coded around five mechanisms that also appear in the questionnaire: respectability discipline, field sorting, leadership and earnings assumptions, unpaid care burdens, and family-status penalties. The article uses short excerpts when a respondent's own wording reveals a mechanism more clearly than summary could. This matters because cultural constraint is often ordinary in its language. It can appear as advice about being a good daughter, a good wife, a responsible mother, or a real man long before it becomes an observable difference in income.

Interview	Profile	Main analytical contribution
Interviewee 1	47; small city; saleswoman; child and elder-care responsibilities	Respectability, lost aspiration, unpaid care burden
Interviewee 2	48; big city; education worker; child, elder-care, and housework responsibilities	Intergenerational conflict, field stereotypes, household labor as second job
Interviewee 3	60; Bucharest; CEO; three children and nephews	Protective family exception, provider norm for men, persistent judgment of women

Table 1. Interview evidence used as qualitative mechanism evidence.

Survey Evidence: Culture as Pre-Market Sorting

The original contribution of this study lies in the type of inequality it makes visible. Existing research robustly shows that culture and institutions affect economic choices. What this project adds is a more concrete mapping of how those cultural pressures are experienced in an Eastern European setting and where standard economic metrics are likely to miss them. The underlying research dossier explicitly framed the project around Romanian cultural stereotypes, gender inequality, education inequality, and the possibility that common inequality measures may understate lived disadvantage.

Three findings are especially important. First, the survey suggests that a major share of economic loss is generated through mechanisms that standard wage or inequality metrics do not directly observe: reputational caution, family permission structures, mobility discipline, networking asymmetry, and anticipation of judgment. This extends Goldin's framework by showing that not only parenthood and job flexibility but also respectability and social timing shape opportunity in a Romania-centered context.

Second, the original evidence suggests an asymmetric but not one-directional system. Women appear more constrained overall, but men are not culturally unconstrained. Open-ended responses repeatedly described men as expected to be breadwinners, accept physically or psychologically riskier work, migrate for income, and avoid visibly flexible or care-oriented roles (Jiayi, Romanian Culture and Economic Opportunity Questionnaire). This matters because it shifts the discussion from a narrow "women's issue" frame to a broader cultural-economy account in which both genders are disciplined, though unequally.

Third, the survey suggests a contradiction with any simplistic reading of low measured wage gaps as evidence of low gendered economic loss. In the original data, respondents reported strong cultural effects on educational choice, career scaling-back, childcare burdens, late-hour mobility, and networking access. These losses often occur before hourly wages among employees are compared. That interpretation is consistent with Eurostat's warning about the limits of unadjusted pay-gap measures and with Goldin's argument that expectations and family structure shape paths long before pay is observed (Eurostat, "Gender Pay Gap Statistics"; Goldin, "A Grand Gender Convergence").

The descriptive results are striking. Among respondents, 51.2% said girls are usually expected to help more with cooking, cleaning, siblings, or older relatives than boys; 67.4% said

boys are more strongly encouraged to become financially independent; and 60.5% had seen a girl or woman reduce education, work, or career plans because of family expectations, compared with 44.2% who had seen the same for a boy or man. These patterns align with Goldin's emphasis on expectation formation, but the survey adds more explicit evidence on family scripts and domestic labor in a Romanian context.

Schooling and aspiration results reinforce the point. Although thirty-seven of forty-three respondents said boys and girls were encouraged equally to speak up in youth settings, twenty-two of forty-three had heard that mathematics, technology, economics, business, or leadership are more suitable for boys, and twenty-five of forty-three had heard that care, communication, teaching, support, or appearance-related domains are more suitable for girls. At work, thirty-seven of forty-three had heard men described as more suitable for leadership, thirty-three of forty-three had heard women described as more suitable for support or administrative roles, twenty-five of forty-three said men were usually more comfortable asking for raises or promotions, and twenty-four of forty-three said women could be judged negatively for directly asking for more money (Jiayi, Romanian Culture and Economic Opportunity Questionnaire).

Mobility, safety, and network access emerged as especially important and under-measured. Thirty of forty-two respondents had seen someone avoid an education, work, or networking opportunity because transport, late hours, safety, or reputation made participation difficult; twenty-eight of forty-two said women were more likely than men to change schedules because of safety or reputation; and twenty-four of forty-one said men were freer to socialize for career or networking purposes. This is a crucial finding because it identifies a channel of economic loss that traditional earnings statistics almost never capture directly.

The perceptions summarized in the "out of ten" questions are even more revealing. Respondents estimated that, on average, 2.70 out of 10 women they knew had reduced education, work, or business plans because of family responsibilities or social expectations, while the figure for men was 0.85 out of 10. They further estimated that 4.77 out of 10 women and 2.96 out of 10 men would probably earn more if family and social expectations were different. In ratio terms, the perceived plan-reduction burden on women is roughly 3.2 times that on men, while the perceived earnings upside for women is about 1.6 times larger. These are not trivial perceptual gaps; they imply a strongly gendered distribution of foregone economic possibility.

Survey pattern	**Result**
Girls expected to help more with cooking, cleaning, siblings, or older relatives	51.2%
Boys more strongly encouraged to become financially independent	67.4%
Had seen a girl/woman reduce education, work, or career plans because of family expectations	60.5%
Had seen a boy/man reduce education, work, or career plans because of family expectations	44.2%
Had heard math, technology, economics, business, or leadership described as more suitable for boys	22 of 43

Had heard care, communication, teaching, support, or appearance-related domains described as more suitable for girls	25 of 43
Had seen someone avoid opportunity because transport, late hours, safety, or reputation made participation difficult	30 of 42
Said men were freer to socialize for career or networking purposes	24 of 41

Table 2. Selected survey findings from the original questionnaire.

Interview Evidence: The Mechanisms Behind the Numbers

Interview transcripts of conducted interviews show how economic loss can begin before a formal economic decision is recorded. In the survey, respondents reported seeing women reduce education, work, or business plans. In the interviews, that pattern appears through everyday language: girls are told to be quiet and respectable, women describe care as a second job, employers are described as wary of mothers, and men are told that they must solve financial problems because they are men. These details shift the analysis from inequality as an outcome to inequality as a process.

The first mechanism is respectability discipline. Interviewee 1 remembered adults telling girls to **“be responsible, be nice, don’t talk too much”** and not to **“have more opinions.”** The economic significance of such language is not that politeness itself is costly, but that repeated lessons in self-containment can make directness feel socially dangerous. A girl who learns that speaking too directly is unfeminine may later hesitate before negotiating salary, entering leadership, presenting a public idea, or choosing a visible career.

The clearest example of foregone aspiration came from the same interview. Interviewee 1 said she wanted to become a singer and had won a competition, but her parents rejected the path because girls were expected to be **“respectable and responsible”** and singers were not seen as **“good girls.”** The loss occurred before any employer, university, or wage record could register it. Talent had already appeared; cultural permission did not follow. This is exactly the kind of pre-market loss that a wage comparison cannot capture.

The second mechanism is field sorting. Interviewee 2 described boys as historically associated with mathematics, physics, and sport, while girls were associated with **“languages, arts and crafts.”** The effect of such messages is cumulative. They do not need to formally ban anyone from a field; they change the social cost of entering it. A girl entering a technical competition may feel she must prove not only competence but belonging, while a boy entering care work or teaching may feel he must justify choosing a lower-status or less masculine path.

The third mechanism is unpaid work as a hidden second job. Asked who usually does more household labor, Interviewee 1 answered **“me of course,”** adding that she has a paid job and then works at home as well. She estimated her unpaid work at four to five hours per day and said that, if treated as paid labor, it would be worth **“half of my salary.”** Interviewee 2 described household care as **“more than a part time job”** and valued it at at least $1,000 per month. These estimates are informal, but the mechanism is clear: time that is culturally treated as natural female duty is also time not available for training, paid hours, rest, networking, or business risk. The interviews also show how care can erase choice without looking like a formal prohibition.

Interviewee 1 described her sister, whose autistic child required constant attention, as carrying responsibility **“24 hours per day, 7 days per week.”** Interviewee 2, asked what family pressure could cost over five years, answered: **“The whole career of that girl.”** The phrasing is blunt, but it captures a central problem: an education or career loss may be caused not by a single discriminatory decision, but by a sequence of ordinary family demands that make advancement progressively less realistic.These statements, while not without possible limitations, strengthen the paper’s economic model because they show why care is not merely a moral or family issue. Four hours per day is twenty-eight hours per week. Five hours per day is thirty-five hours per week. At that level, unpaid work competes directly with training, rest, overtime, promotion preparation, networking, and business-building. Even when a woman remains formally employed, her opportunity set may be reduced. She may not leave the labor force, but she may lose the surplus hours through which careers normally grow.

The fourth mechanism is family-status risk. Interviewee 2 said some Romanian employers prefer young women without children or family because children imply the woman will stay home when a child is sick. Interviewee 3 made the same point from a leadership position: when problems happen at home, people assume **“the woman has to solve”** them, so women with children can be treated as less serious workers. In this mechanism, the household does not remain private. It travels into hiring, promotion, scheduling, and trust.

The fifth mechanism is the burden of judgment. Interviewee 3 is especially revealing because she did not describe herself as raised in a weak female environment. She spoke of powerful women in her family and of growing up believing she could do what she needed in life. Yet she still said women are judged **“all the time”** when family outcomes appear imperfect. Her example was specific: after her divorce, people blamed her when her son gained weight, even though he was eating outside every day. The example matters because it shows how women can be made responsible not only for labor they perform, but for the moral interpretation of family life itself.

The transcripts also resist a simple story of old norms versus modern freedom. Interviewee 3 taught her sons to do all household chores; Interviewee 2 contrasted a grandmother who told her to cook for marriage with a mother who told her to study because career mattered more. These details show that culture is not a single command issued by society. It is a contested field, with older and newer scripts coexisting inside families. That is why change is possible, but also why change is uneven.

Mechanism	Interview evidence	Economic interpretation
Respectability discipline	Girls told to be responsible, nice, quiet, and less opinionated	Lower voice, reduced bargaining, weaker leadership confidence
Foregone aspiration	A singing path was rejected because “good girls” should be respectable	Loss before labor-market measurement begins
Field sorting	Math, physics, and sport coded masculine; languages, arts, and crafts coded feminine	Higher social cost for crossing fields

Unpaid second shift	Household work described as more than part-time, worth half a salary or at least $1,000/month	Reduced time for paid work, training, rest, and promotion
Family-status penalty	Women with children seen as less available because home problems are assumed to be theirs	Motherhood penalty and reduced employer trust
Gendered accountability	Women judged for family outcomes, including children's behavior or appearance	Persistent reputational pressure even among high-achieving women

Table 3. Qualitative mechanisms drawn from the interview transcripts.

Economic Model and National Loss Estimates

To translate these cultural mechanisms into economic terms, the study uses a transparent scenario model rather than a structural national macro-model. The purpose is not to generate a definitive national account, but to estimate the plausible order of magnitude of culturally generated loss if the survey's patterns scale to Romania more broadly. The interviews support the model by clarifying what the monetary loss parameter represents: not just lower wages, but lost time, reduced ambition, avoided promotion, unchosen training, and constrained bargaining.

The core equation is: $L = \Sigma g \in \{f,m\}\ Ng \times pg \times \ell$. In this equation, L is estimated annual economic loss; g refers to gender group; Ng is the relevant population base for each gender group; pg is the share whose earnings are constrained by cultural expectations; and ℓ is annual income loss per affected person.

Symbol	Definition
L	Estimated annual economic loss
g	Gender group: female or male
Ng	Relevant population base for each gender group
pg	Share whose earnings are constrained by cultural expectations
ℓ	Annual income loss per affected person

Table 4. Model variables.

Two population specifications are used. The conservative worker-adjusted specification multiplies Romania's twenty-to-sixty-four population by observed employment rates. The model uses an approximate twenty-to-sixty-four population of 11.44 million, with about 5.63 million women and 5.81 million men; the employment assumptions in the model are 59.1% for women and 77.7% for men. These assumptions are consistent with Eurostat's 2025 description of Romania as having one of the EU's largest gender employment gaps, with women below 60% employment and men above 77% among those aged twenty to sixty-four (Eurostat, "Employment - Annual Statistics"). The broader potential-population specification uses the full twenty-to-sixty-four population, on the argument that culturally produced loss often takes the form of non-entry, reduced labor supply, or foregone entrepreneurship rather than merely lower wages among the employed.

The affected-share parameters are derived from the original survey. For the conservative scenario, the model uses the mean reported shares of women and men who had reduced education, work, or business plans because of family or social expectations: 27.0% for women and 8.5% for men. For the upper scenario, it uses the mean reported shares who would probably earn more if family and social expectations were different: 47.7% for women and 29.6% for men. The central scenario uses the average of those two figures for each gender: 37.4% for women and 19.1% for men (Jiayi, Romanian Culture and Economic Opportunity Questionnaire).

The annual loss parameter is derived from the survey's monetary questions after annualizing the monthly care-related response and excluding a clear 1 RON placeholder entry from the upper-bound mean. The resulting positive-response distribution yields three anchors: RON 10,250 at the 25th percentile, RON 25,000 at the median, and RON 42,250 as the mean positive annualized loss (Jiayi, Romanian Culture and Economic Opportunity Questionnaire). These amounts are plausible in relation to Romania's recent earnings data. Romania's National Institute of Statistics reported average net monthly earnings of RON 5,914 in December 2025, and RON 5,843 in April 2026, meaning that the model's median annual loss parameter is below half a year of average net earnings (National Institute of Statistics; "Monthly Average Earning").

The resulting national loss estimates are shown below.

Scenario	Female affected share	Male affected share	Annual loss per affected person	Worker-adjusted national loss	Potential-population national loss
Conservative	27.0%	8.5%	RON 10,250	RON 13.2 bn	RON 20.7 bn
Central	37.4%	19.1%	RON 25,000	RON 52.6 bn	RON 80.3 bn
Upper-bound	47.7%	29.6%	RON 42,250	RON 123.7 bn	RON 186.4 bn

Table 5. Scenario assumptions and estimated national economic loss.

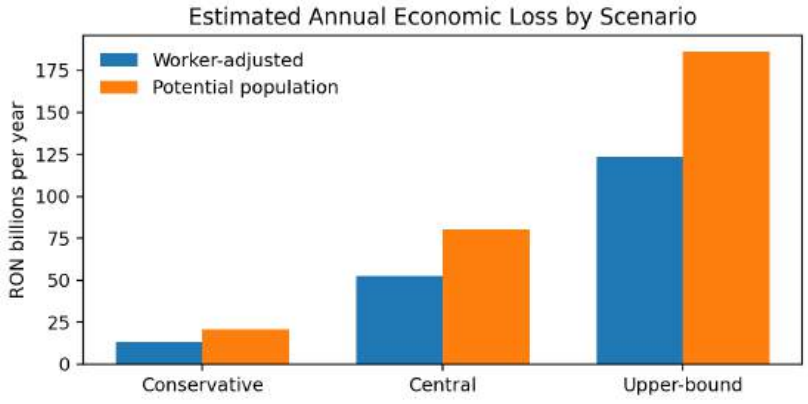


Figure 1. Estimated annual economic loss by scenario.

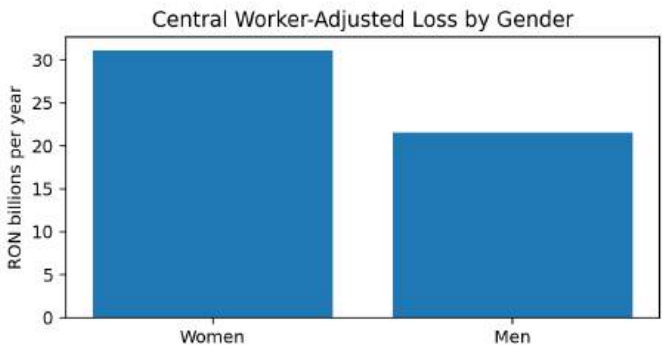


Figure 2. Central worker-adjusted loss by gender.

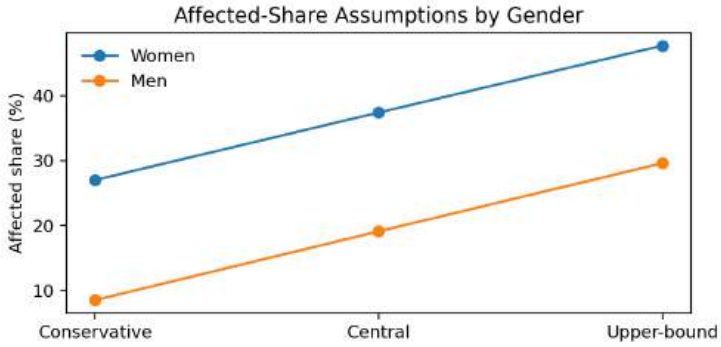


Figure 3. Affected-share assumptions by gender.

Set against Romania's recent GDP level and labor-market scale, these are large values rather than rounding errors. Using the World Bank's current-LCU GDP series as the GDP benchmark, the worker-adjusted range is roughly 0.8% to 7.5% of GDP, while the broader potential-population range rises to roughly 1.3% to 11.3% of GDP (World Bank, "GDP Current LCU - Romania"). Against that benchmark, the national loss range is uncertain, but its order of magnitude is economically meaningful rather than symbolic.

These estimates should be read carefully. They do not mean that Romania literally loses the exact amount shown each year in cash terms. They mean that if the survey's perceived incidence of culturally constrained educational, occupational, and bargaining decisions approximates broader reality, then the implied forgone income is plausibly large. Even the conservative scenario suggests that the aggregate cost of cultural narrowing is too large to be dismissed as anecdotal or purely sociological.

The model also helps clarify why a culture-first analysis changes the economic story. If inequality were understood only as a wage differential among current employees, many losses would disappear from view. But once one counts reduced paid hours due to unpaid care, unasked raises, avoided networking, unchosen majors, restricted relocation, and fear-based self-limitation, culture becomes economically visible as a system of allocation. That conclusion is consistent with Goldin's account of expectation formation and labor-market structure, yet the original data

add concrete Romanian channels, especially reputation, timing, and mobility, that standard metrics capture poorly.

Discussion: What the Interviews Add to the Paper

The principal interpretive finding of this paper is that culture is functioning here as a pre-market sorting device. It does not simply explain who gets paid more after entering the labor market. It helps determine who studies what, who relocates, who stays late, who asks for advancement, who absorbs care demands, who networks without reputational risk, and who imagines themselves as economically legitimate. On this reading, culture is not a soft variable. It is a mechanism of economic selection.

The interview evidence sharpens this argument because it shows cultural restriction arriving as ordinary advice rather than open coercion. **"Learn to cook," "be nice," "don't talk too much,"** and **"he's a man, he has to solve it"** do not look like economic policy. Yet such phrases assign girls to care, boys to provision, women to respectability, and men to problem-solving. Over time, they become a social curriculum for economic behavior.

The transcripts also show that loss can occur even when school appears formally equal. Interviewee 1 and Interviewee 3 both said boys and girls were encouraged equally to speak in school, yet both described family and work expectations that narrowed women's options. Formal encouragement therefore does not exhaust opportunity. If school says "you can do anything" while family, employers, and reputational communities say "not like that," opportunity remains conditional.

Unpaid care must therefore be treated as an economic variable, not merely a private habit. Four or five hours of daily household work changes the realistic space available for study, overtime, business development, job search, and recovery. Love does not erase opportunity cost. Interviewee 1's comment that the interview made her realize women should not be expected to work this much is analytically important: a burden can be deeply normalized and still economically damaging.

The transcripts also show that culture harms men, though not symmetrically. Interviewee 3 remembered the phrase **"he's a man, he has to solve it."** The survey likewise found breadwinner pressure on boys and men. This does not dilute the larger burden on women; it clarifies the system. Men can be pushed into risk, migration, overwork, and emotional silence, while women are pushed into care, caution, and reputation management. In both directions, the economy loses when people are matched to inherited scripts rather than ability and preference.

A culture-first account also avoids a false opposition between agency and structure. The interviews contain agency: mothers who encourage study, women who become CEOs, parents who teach sons to clean, families that judge husbands for not washing dishes. But the agency does not cancel the structure. The presence of exceptions shows that norms can be resisted; it does not prove that the norms are absent. In fact, the exceptions help identify the mechanism more clearly. When Interviewee 3 describes her unusually empowering family and still describes broader judgment of women, the evidence suggests that individual strength can reduce cultural cost but not fully remove it.

Limitations and Future Research

The paper has important limitations. The sample is small, non-random, and heavily female. It likely overrepresents adults with work and care experience and may underrepresent the poorest, the least educated, men who do not discuss care expectations openly, and some rural populations. The survey measures self-reports and perceptions rather than administrative earnings records. Not all monetary questions had many valid responses, and some entries were noisy. Any national extrapolation therefore has to be expressed as a scenario rather than a point estimate.

The interviews also have limitations. Three interviews cannot represent Romania. They clarify mechanisms rather than estimate prevalence. The structured question-and-answer format makes comparison easier but can limit narrative depth, and some answers are brief, translated, or linguistically imperfect. The analysis therefore treats wording carefully. Even so, the interviews are valuable because they reveal mechanisms that quantitative surveys often flatten: a career not pursued, a field made socially awkward, a mother blamed for household outcomes, and domestic duty treated as normal rather than work.

The model risks overlap between channels. A person whose income fell because of unpaid care may also have missed a promotion or networking opportunity. The scenario approach partly manages this by collapsing effects into a per-affected-person annual loss anchor, but it cannot remove all double-counting risk. The model also assumes that survey perceptions can be cautiously extrapolated to a broader population, which should be tested rather than assumed in future work.

A larger nationally representative survey could estimate separate effects by region, class, parental status, education level, and urban-rural background. Linked administrative and survey data could test whether Romania's relatively low measured gender pay gap conceals larger pre-market sorting losses. Time-use data would also allow a better conversion of unpaid care hours into forgone earnings. Finally, richer qualitative interviewing could distinguish between internalized and externally imposed constraint, a question the current survey and interviews can only partially address.

These limitations do not erase the value of the evidence. On the contrary, they point to the likely direction of future research. The project's original data are important precisely because they reveal mechanisms that broader administrative systems often fail to ask about. If the cost of culture is largely hidden before wages are measured, then research must move upstream, into households, expectations, schedules, safety, respectability, and the social meaning of ambition.

Using the Evidence for Social Good

The value of this research is not only descriptive. If the study is correct that many losses happen before wages are measured, then social action should not wait until inequality becomes visible in salary data. The findings can be used as a diagnostic tool for schools, employers, public agencies, and civil-society organizations. A practical intervention should ask five questions: who is expected to care, who is expected to earn, who can move safely and without

reputational cost, who is believed to be leadership material, and who is allowed to make a risky or ambitious choice without being treated as selfish.

Romania already has institutions that could use this evidence. ANES, the National Agency for Equal Opportunities between Women and Men, defines its mission around promoting equal opportunity and treatment between women and men, eliminating sex-based discrimination, and preventing and combating domestic violence (Agenția Națională pentru Egalitatea de Șanse între Femei și Bărbați). Its existing strategy and monitoring role make it a logical first contact for turning the survey into a larger national or regional diagnostic. A short policy brief for ANES could translate the paper's mechanisms into measurable indicators: unpaid-care hours, mobility restrictions, parental-status penalties, reluctance to negotiate, and expectations that girls should protect respectability while boys should become providers.

A second contact point is CNCD, the National Council for Combating Discrimination. CNCD is listed by Equinet as Romania's equality body and is responsible for non-discrimination work, including discrimination on grounds of sex (Equinet). The research could help CNCD identify forms of discrimination that are not always framed as direct exclusion: mothers treated as risky employees, women judged as too ambitious, men pressured away from care roles, and women losing networking access because late travel or reputation makes participation difficult. The point is not to convert every cultural pattern into a legal complaint, but to help equality bodies see where informal norms generate formal exclusion.

Employment institutions are another practical route. ANOFM materials state that Bucharest and each of Romania's forty-one counties has a government employment agency, AJOFM, that can support job seekers (National Agency for Employment). This creates a ready-made structure for regional pilots. A low-cost pilot could ask AJOFM offices in selected counties to add a short optional module to career counseling: whether family care, transport safety, late hours, reputation, or negotiation discomfort is limiting job search. The resulting data would show whether cultural barriers are concentrated in particular regions, age groups, or family situations.

The first pilot locations should combine national influence with local variation. Bucharest is necessary because it concentrates ministries, equality bodies, universities, national NGOs, and large employers. Cluj-Napoca is useful because Babeș-Bolyai University has official gender-equality guidance and an academic environment capable of testing measurement tools (Babeș-Bolyai University). Sibiu is useful because A.L.E.G. is based there and has worked since 2004 on preventing violence against women and promoting gender equality (A.L.E.G.). Iași and other Moldavian counties would be important because a culture-first theory should not be tested only in the capital or wealthy university cities. Timișoara, Brașov, and Constanța could add employer and mobility variation, especially for questions about commuting, late schedules, service-sector work, and regional labor markets.

Civil society could turn the findings into public language. FILIA describes itself as a feminist organization working through research, advocacy, and activism for gender equality and nondiscrimination (SALTO-YOUTH). A.L.E.G. describes its work as preventing and combating

violence against women and promoting gender equality (A.L.E.G.). The Coalition for Gender Equality platform also lists organizations such as Centrul Parteneriat pentru Egalitate and A.L.E.G. as members or contacts (ONGEN). These groups could use the paper not as a finished national truth, but as a conversation tool: a way to ask women and men whether opportunity is being lost through care, reputation, safety, and family scripts before any formal discrimination is visible.

Universities are the best route for research contribution. The University of Bucharest's Gender Equality Plan and ICUB Gender Studies Platform make it a natural home for improving the survey, adding interviews, and testing the model with stronger sampling (University of Bucharest; ICUB). Professor Ionela Băluță's Equal Opportunities Policies program at the University of Bucharest is especially relevant because the university describes it as one of the few gender-studies programs in Romania and the first and only program of its type at the institution (University of Bucharest, "UB Researcher at the Microphone"). SNSPA is another potential partner because its political science faculty lists research experience in social inequalities, labor relations, and gender equality (SNSPA). Babeș-Bolyai University in Cluj-Napoca could contribute through regional research, student surveys, and time-use analysis.

Place	**Possible partners**	**Why this site matters**	**How the research could be used**
Bucharest	ANES; CNCD; University of Bucharest; FILIA	National policy, equality enforcement, research design, and public advocacy.	Send a short policy brief; propose a larger survey; request feedback on indicators.
Cluj-Napoca	Babeș-Bolyai University	Regional research capacity and existing institutional gender-equality framework.	Pilot student and working-adult surveys; add time-use and care questions.
Sibiu	A.L.E.G.	Civil-society experience in gender equality and violence prevention.	Use interview phrases for workshops on care, respectability, safety, and autonomy.
Iași / Moldavian counties	AJOFM county employment agencies; local universities and NGOs	Important for testing whether cultural constraints differ outside wealthier university and capital-city settings.	Add cultural-barrier screening to career counseling and job-search support.
Timișoara, Brașov, Constanța	Employers, AJOFM offices, universities, and women's NGOs	Useful for comparing service, industrial, university, tourism, and commuting contexts.	Run employer audits on promotion, scheduling, networking, and parental-status assumptions.
All counties	AJOFM / AMOFM employment network	Existing infrastructure in Bucharest and each of Romania's 41 counties.	Collect comparable data on care, mobility, family expectations, and negotiation barriers.

The most useful next product would be a two-page Romanian-language policy brief with three parts: the economic estimate, the five cultural mechanisms, and a contact map for action. The brief should avoid abstract language and use the interview phrases directly: “good girls,” “me of course,” “more than a part-time job,” “the whole career of that girl,” and “he’s a man, he has to solve it.” Those phrases make the research legible to people outside academia because they show that culture does not operate only through laws or slogans. It operates through sentences people have heard all their lives.

The research could also support three practical interventions. First, schools and universities could run ambition-and-care workshops that ask students to identify how family expectations shape subject choice, study time, and leadership confidence. Second, employers could audit promotion, scheduling, and networking practices for parental-status penalties and informal after-hours exclusion. Third, county employment agencies and NGOs could create referral pathways for women whose job search is blocked by childcare, eldercare, mobility, or domestic expectations. None of these interventions requires treating culture as destiny. They require treating culture as a measurable constraint that can be named, reduced, and redesigned.

A contact strategy should remain modest and specific. The strongest approach would not be to send the full article first, but to send a short summary asking whether the organization would be willing to comment on three questions: whether the mechanisms match what they see in Romania, which region or population should be studied next, and whether any existing program already collects data on care, mobility, reputation, or bargaining. This protects the research from becoming a one-way claim about society. It turns the study into a request for collaboration, correction, and local knowledge.

For ANES and CNCD, the most useful output would be a measurement appendix rather than a moral appeal. The appendix could propose survey questions such as: Have you ever avoided a course, job, training, promotion, or networking opportunity because family responsibilities made it difficult? Have you ever avoided such an opportunity because transport, late hours, safety, or reputation made it difficult? Have you ever felt unable to ask for higher pay because directness would be judged negatively? Have you ever been treated as less serious at work because of marriage, children, or expected care duties? These questions translate the interview evidence into indicators that public bodies can debate, revise, and test.

For employers, the research suggests an audit of informal opportunity. A company may have equal-pay rules and still reproduce inequality if advancement depends on staying late, traveling without notice, drinking or socializing after work, taking weekend calls, or appearing constantly available. The issue is not only whether women and men are paid the same for the same task, but whether the path toward better tasks is equally reachable. An employer pilot could therefore compare who receives stretch assignments, who attends networking events, who turns down opportunities because of care or reputation, and whether managers interpret the same family responsibility differently when it belongs to a mother or a father.

For schools and universities, the findings could become an early-warning tool. Students can be asked not only what subjects they like, but what subjects they feel allowed to like. They

can be asked whose disappointment they fear, what career would be considered embarrassing, whether family duties reduce study time, and whether ambition feels socially acceptable. This would make the paper useful before the labor-market stage. If respectability, field sorting, and care expectations begin early, then interventions should begin while students are still choosing courses, competitions, internships, and identities.

The research should also be shared in a way that includes men without recentring them. The interviews and survey both suggest that men can be harmed by breadwinner pressure, financial-risk expectations, and the phrase "he's a man, he has to solve it." A useful public message would therefore be: gender equality is not a project to make women work more while men are ignored. It is a project to let women refuse invisible overwork and let men refuse compulsory invulnerability. That framing may make the evidence easier to use in families, schools, and workplaces where the language of gender equality is sometimes heard as accusation rather than redesign.

The broader contribution is therefore translational. The paper turns private sentences into public evidence, public evidence into economic estimates, and economic estimates into possible institutional action. Its usefulness depends on keeping all three levels connected. If the research remains only a model, it risks abstraction. If it remains only testimony, it may be dismissed as anecdote. If it connects testimony, survey data, and institutional partners, it can help Romania ask a more useful question: not only whether women and men are formally equal, but where opportunity is quietly redirected before equality can be measured.

Conclusion

This paper has argued that the economic cost of inequality in Romania cannot be understood adequately if culture is treated as background noise. The original survey evidence indicates that cultural expectations shape educational ambition, paid work, bargaining behavior, mobility, timing, and network access. These pressures are asymmetrical: women bear more of the measurable burden, but men are also structured through breadwinner and risk-bearing expectations. The economic consequence is not only unequal pay. It is the systematic misallocation of human capability, and that misallocation can be studied, named, and reduced.

The interviews make that conclusion more concrete. They show a girl discouraged from a public artistic career because respectability mattered more than talent; a woman describing household work as a second job; an educator estimating that family expectations can cost "the whole career of that girl"; and a CEO explaining that women are judged for family outcomes even when the facts are more complicated. These are not side stories. They are the living mechanisms through which culture becomes economics.

This matters for publication because the article does not ask readers to accept a purely emotional claim or a purely numerical one. The interview evidence gives the argument texture, the questionnaire gives it pattern, and the scenario model gives it scale. Each layer is incomplete on its own. Together, they make visible a kind of economic loss that is easy to miss precisely because it is absorbed into ordinary family life and ordinary advice.

The strongest implication is that equality policy should look earlier than the labor contract. A woman who never applies for a leadership role because directness has been coded as unfeminine is not visible in a promotion-denial statistic. A student who avoids a field because it feels socially inappropriate is not visible in a wage-gap table. A mother who loses study time to care is not always visible as unemployed or discriminated against. The paper's central contribution is to bring those pre-measurement losses into the language of economics without reducing them to economics alone.

The same logic applies to men. The breadwinner script can look like privilege from one angle and constraint from another. It can push men toward overwork, risk, migration, and shame around dependence or care. Recognizing this does not make the system symmetrical. Women still carry the larger unpaid-care and respectability burden in the evidence presented here. But it does make the reform agenda broader: a fairer economy would reduce the pressure on women to be silently useful and reduce the pressure on men to be permanently invulnerable.

Placed in dialogue with Nobel-recognized work by Acemoglu, Johnson, Robinson, and Goldin, the Romanian evidence supports a broader conclusion: prosperity depends not merely on markets, and not even merely on formal institutions, but on the lived cultural rules that determine who may fully inhabit economic possibility. If the original sample is taken as broadly indicative, the annual loss associated with culturally constrained economic choice is plausibly in the range of RON 13.2 billion to RON 123.7 billion on a worker-adjusted basis, and RON 20.7 billion to RON 186.4 billion on a broader potential-population basis. Those numbers are uncertain, but they are not small. Their main significance is conceptual as much as quantitative: they show that the most expensive inequalities may be the ones that occur before the spreadsheet begins.

Works Cited

A.L.E.G. “About Us.” Association for Liberty and Equality of Gender, aleg-romania.eu/en/about-us/. Accessed 6 July 2026.

Agenția Națională pentru Egalitatea de Șanse între Femei și Bărbați. “Agenția Națională pentru Egalitatea de Șanse între femei și bărbați.” ANES, 2026, anes.gov.ro/. Accessed 6 July 2026.

Agenția Națională pentru Egalitatea de Șanse între Femei și Bărbați. “Contact.” ANES, 2026, anes.gov.ro/contact/. Accessed 6 July 2026.

Jiayi. Gendered Cultural Constraints in Romania: Interview Transcripts. Unpublished interview transcripts, 2026.

Jiayi. Romanian Culture and Economic Opportunity Questionnaire. Unpublished dataset, 2026.

Babeș-Bolyai University. Guidelines on Gender Equality. 2021, www.ubbcluj.ro/en/despre/organizare/files/EN_Ghid_egalitate_de_gen.pdf. Accessed 6 July 2026.

Bursztyn, Leonardo, Alessandra L. González, and David Yanagizawa-Drott. “Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia.” American Economic Review, vol. 110, no. 10, 2020, pp. 2997-3029. American Economic Association, doi:10.1257/aer.20180975.

Equinet. “National Council for Combating Discrimination.” European Network of Equality Bodies, equineteurope.org/author/romania_nccd/. Accessed 6 July 2026.

Eurostat. “Employment - Annual Statistics.” Statistics Explained, European Commission, 2026, ec.europa.eu/eurostat/statistics-explained/index.php?title=Employment_-_annual_statistics. Accessed 6 July 2026.

Eurostat. “Gender Pay Gap Statistics.” Statistics Explained, European Commission, 2026, ec.europa.eu/eurostat/statistics-explained/index.php?title=Gender_pay_gap_statistics. Accessed 6 July 2026.

Fernández, Raquel, and Alessandra Fogli. “Culture: An Empirical Investigation of Beliefs, Work, and Fertility.” American Economic Journal: Macroeconomics, vol. 1, no. 1, 2009, pp. 146-77. American Economic Association, doi:10.1257/mac.1.1.146.

Goldin, Claudia. “A Grand Gender Convergence: Its Last Chapter.” American Economic Review, vol. 104, no. 4, 2014, pp. 1091-1119. American Economic Association, doi:10.1257/aer.104.4.1091.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. “Does Culture Affect Economic Outcomes?” Journal of Economic Perspectives, vol. 20, no. 2, 2006, pp. 23-48. American Economic Association, doi:10.1257/jep.20.2.23.

ICUB. “The Gender Studies Platform.” University of Bucharest, icub.unibuc.ro/the-gender-studies-platform/. Accessed 6 July 2026.

National Agency for Employment. Employment Agency Guide. ANOFM, 2023, www.anofm.ro/upload2/ANOFM/28529/ANOFM_Guidance_Tool_ENG.pdf. Accessed 6 July 2026.

National Institute of Statistics. “Monthly Average Earning.” INSSE, 2026, insse.ro/cms/en/tags/press-release-monthly-average-earning. Accessed 6 July 2026.
Nobel Prize Outreach. “The Prize in Economic Sciences 2023 - Popular Information.” NobelPrize.org, 2023, www.nobelprize.org/prizes/economic-sciences/2023/popular-information/. Accessed 6 July 2026.
Nobel Prize Outreach. “The Prize in Economic Sciences 2023 - Press Release.” NobelPrize.org, 9 Oct. 2023, www.nobelprize.org/prizes/economic-sciences/2023/press-release/. Accessed 6 July 2026.
Nobel Prize Outreach. “The Prize in Economic Sciences 2024 - Press Release.” NobelPrize.org, 14 Oct. 2024, www.nobelprize.org/prizes/economic-sciences/2024/press-release/. Accessed 6 July 2026.
ONGEN. “About Us.” Coalition for Gender Equality, ongen.ro/en/about-us/. Accessed 6 July 2026.
SALTO-YOUTH. “FILIA Centre.” Otlas Partner-Finding, www.salto-youth.net/tools/otlas-partner-finding/organisation/filia-centre.11017/. Accessed 6 July 2026.
SNSPA. “Faculty of Political Science.” National University of Political Studies and Public Administration, www.politice.ro/en/FSP. Accessed 6 July 2026.
Tabellini, Guido. “Culture and Institutions: Economic Development in the Regions of Europe.” Journal of the European Economic Association, vol. 8, no. 4, 2010, pp. 677-716, doi:10.1111/j.1542-4774.2010.tb00537.x.
University of Bucharest. “Gender Equality.” University of Bucharest, unibuc.ro/despre-ub/egalitate-de-gen/?lang=en. Accessed 6 July 2026.
University of Bucharest. “UB Researcher at the Microphone: Gender Studies as an Antidote to Intolerance.” University of Bucharest, 13 Feb. 2024, unibuc.ro/despre-studiile-de-gen-ca-antidot-pentru-intoleranta-cu-profesoara-ionela-baluta-invitata-unei-noi-editii-a-seriei-microscop-cercetator-ub-la-microfon/?lang=en. Accessed 6 July 2026.
World Bank. “GDP Current LCU - Romania.” World Development Indicators, 2026, data.worldbank.org/indicator/NY.GDP.MKTP.CN?locations=RO. Accessed 6 July 2026.
World Bank. “Population Ages 15-64, Total - Romania.” World Development Indicators, 2026, data.worldbank.org/indicator/SP.POP.1564.TO?locations=RO. Accessed 6 July 2026.

The Impact of Artificial Intelligence on Consumer Behaviour By Tracy Tan

Abstract

Artificial intelligence now mediates much of the consumer journey, from the recommendations shoppers see to the prices they are offered and the assistants they speak to. This literature review draws together recent scholarship on how these systems affect what consumers do — what they buy, whom they trust, what they disclose, and how they feel. A consistent theme emerges: AI rarely helps or harms in the abstract. The same personalisation that feels convenient can feel intrusive; the same human-like chatbot that builds rapport can unsettle; the same pricing algorithm that offers a good deal can read as betrayal. Value, trust, and a sense of control matter more than raw capability, and the review closes by mapping where the evidence is still thin.

1. Introduction

Artificial intelligence has moved from the back office of marketing into the everyday texture of consumption. Recommendation engines shape what shoppers notice, chatbots and voice assistants field their questions, targeted advertising follows them across sites, and algorithms increasingly set prices and even act on their behalf. AI here is not one technology but a family of them — systems that predict, personalise, converse, and decide (Huang and Rust) — and their reach now spans the whole customer journey rather than the moment of purchase alone. As these applications have spread, so has the research examining them, drawing in marketing, consumer research, psychology, and economics.

That research did not begin with AI as such. Its roots lie in the technology-acceptance tradition of the 1980s and 1990s, which asked why people take up new information systems at all (Davis; Venkatesh et al., "User Acceptance"). Through the 2000s, interest shifted to e-commerce personalisation and recommender systems, framing the tension between relevance and privacy that still runs through the field. A more distinctly AI-focused literature took shape in the late 2010s, as service robots, chatbots, and voice assistants entered everyday commerce and scholars began treating consumer–AI interaction as a subject in its own right rather than a special case of technology adoption (Mende et al.; Puntoni et al.). Alongside it, work on algorithm aversion and appreciation is asked when people defer to machines (Dietvorst et al.; Logg et al.). Since 2022, generative AI has driven a fast-growing new wave (Cillo and Rubera; Arce-Urriza et al.), and systematic reviews have begun to consolidate what is now a sprawling, interdisciplinary body of research (Mariani et al.).

This review asks how that shift is changing consumer behaviour, understood broadly to take in purchase decisions, adoption, trust, disclosure, loyalty, perceived autonomy, and emotion. A useful starting point is Puntoni et al.'s argument that people do not experience AI as a neutral efficiency tool but through data capture, classification, delegation, and social interaction, each carrying its own psychological weight. The thread running through the evidence below is that AI's effects are conditional. It can cut search costs, sharpen relevance, and deepen engagement,

yet those gains erode the moment a system feels intrusive, unfair, opaque, or presumptuous. Outcomes depend less on how capable the technology is than on how consumers read it.

2. Personalisation and the Shape of Choice

Personalisation is AI's most familiar effect on consumers. By mining behavioural and contextual data, recommender systems rank products, anticipate preferences, and narrow a vast catalogue to a manageable few — lowering search effort and, in several studies, lifting both the usefulness and the enjoyment of shopping (Yin and Qiu). Newer work suggests these systems persuade through experience as much as information: relevant, well-timed recommendations raise clicking and purchase intentions by way of immersion and ease, but the effect collapses once consumers sense their data is being over-harvested or the suggestions are simply poor (Yin et al.).

That is the personalisation–privacy paradox in miniature. Consumers like recommendations that save time and reflect their tastes, yet dislike the surveillance that produces them. Aguirre et al. showed personalised advertising works when the underlying data exchange feels trustworthy and backfires when it feels covert, while Karwatzki et al. found privacy-conscious consumers were not won over by transparency features alone. Canhoto, Keegan, and Ryzhikh recast the paradox as an ongoing negotiation rather than a one-off calculation, with consumers continually re-weighing benefit against risk as circumstances change. Personalisation, then, is never quite neutral: it reshapes which options a person sees and how hard they must work to choose, and it can quietly narrow the field to what an algorithm already expects them to like — with real consequences for autonomy (Bonicalzi et al.).

None of which makes personalisation inherently manipulative. A meta-analysis of 53 experiments found personalised advertising more persuasive than generic advertising largely because it felt more relevant, not more intrusive (Yeo et al.). But relevance can itself be the catch: an ad lands precisely because the system has inferred something the consumer never chose to reveal.

3. Trust, Adoption, and the Human-Like Machine

Whether consumers accept an AI service comes down mostly to trust. Classic technology-acceptance research explains adoption through perceived usefulness and ease of use (Davis), and both still matter — but with AI they are rarely enough. Consumers also weigh reliability, risk, fairness, and how much independence they are willing to grant the system. Studies of recommendation tools and virtual assistants find trust encouraging adoption and fear or perceived risk suppressing it (Cabrera-Sánchez et al.; Xiong et al.), and they draw a telling line between trusting the company and trusting its AI: brand trust gets people to try a service, but a system that behaves too autonomously makes them hesitate even when they trust the firm behind it (Frank et al.).

Design blurs the picture further. Voice assistants and chatbots are built to feel social — names, voices, warmth, conversational give-and-take — and consumers respond in kind,

applying human social rules to them (Pitardi and Marriott). Perceived empathy and friendliness build trust and lower resistance, though the effect bends with task complexity and whether the bot admits to being one (Cheng et al.). Anthropomorphism can lift attitudes and purchase intentions (Balakrishnan and Dwivedi), yet the same cues that make a system persuasive do not necessarily make it competent, and they steer disclosure and future patronage in ways firms do not always intend (Pizzi et al.). Generative AI raises the stakes again: more fluent chatbots can deepen trust in some settings, such as online grocery shopping (Chakraborty et al.), while sharpening privacy and authenticity worries in others.

Emotion runs alongside cognition. More sophisticated "thinking" and "feeling" AI can stir stronger positive emotions and, through them, more spending and loyalty (Schepers et al.). But human-likeness has a floor as well as a ceiling — Mende et al. found humanoid service robots could feel eerie and even threaten people's sense of human distinctiveness, prompting compensatory consumption. Where AI does earn loyalty it tends to do so indirectly, by building trust and satisfaction and leaning on good personalisation (Hassan et al.), and it forfeits that loyalty quickly once it starts to feel like surveillance or a downgrade from human service.

4. Privacy, Manipulation, and Pricing

The ethical questions sharpen because AI does more than react to choices — it classifies people, infers what they want or fear, decides what they see, and sometimes sets their price. That capacity edges persuasion toward manipulation, which Susser, Roessler, and Nissenbaum define as hidden influence that quietly erodes a person's decision-making power. Dark patterns are its bluntest form: manipulative interface design has been shown to push people into subscriptions they would otherwise have skipped, at times swaying choice more than price does (Luguri and Strahilevitz).

Pricing is where it becomes personal. When algorithms charge different customers different prices, perceived fairness falls and a sense of betrayal rises (Wu et al.); consumers read the price not merely as a number but as a statement about how the firm regards them. A good personalised deal may be welcome, but a price that feels secretly discriminatory corrodes trust. The concern is not confined to how individuals are priced, either: Calvano et al. show that autonomous pricing algorithms, left to learn on their own, can converge on tacitly collusive, supra-competitive prices without any instruction or communication to do so — a market-level harm that emerges from the technology itself rather than from any single firm's intent.

Nor are consumers simple accuracy-maximisers. People can lose faith in an algorithm the instant they watch it err, even when it outperforms humans (Dietvorst et al.), yet elsewhere they favour algorithmic advice over human judgement (Logg et al.). Some of the contradictions resolved by task — algorithms are trusted more for objective judgements than subjective ones (Castelo et al.) — but the aversion is no fixed reflex: hold benefits and control steady, and the supposed preference for humans largely disappears (Schaap et al.). The live question is not whether consumers trust AI, but when, and control sits at the centre of the answer.

5. Theoretical Frameworks

Several frameworks organise this evidence, each catching part of the picture. The Technology Acceptance Model and its successors, UTAUT and UTAUT2, explain adoption through usefulness, ease, social influence and — in consumer versions — hedonic value and habit (Davis; Venkatesh et al., "User Acceptance"; Venkatesh et al., "Consumer Acceptance"). They travel well across applications but lean heavily on stated intention, which can drift from actual behaviour, and they say little when the sticking point is trust or fairness rather than usability. Trust-and-risk models and privacy-calculus accounts fill that gap, explaining why a genuinely useful system still gets refused, though they can shrink privacy to an individual sum rather than a question of power and surveillance.

Process models add movement. The stimulus–organism–response framework links AI features such as accuracy, interactivity, and anthropomorphism to internal states like value, trust, and emotion, and then to behaviour, while self-determination theory foregrounds autonomy — apt, given evidence that consumer response to algorithmic decision autonomy follows an inverted U, souring when the system does too little or too much (Fan and Liu). Puntoni et al.'s experiential lens complements these by asking how AI feels: how data capture, classification, delegation, and social contact bear on control and identity, which helps explain why an objectively helpful system can still repel. Dual-process ideas hover behind much of this work, since AI can nudge both deliberate and automatic responses, but they tend to be assumed rather than tested.

6. Gaps and Future Directions

Real limits remain. Much of the evidence rests on cross-sectional surveys and stated intentions, leaving open whether intention becomes purchase, habit, or lasting loyalty; longitudinal and field studies are overdue. The technologies are often lumped together, too, when a recommender, a generative chatbot, a pricing engine, and a targeting system differ sharply in autonomy, transparency, and the cost of being wrong.

Generative AI is the clearest frontier, moving faster than the research tracking it. Reviews still describe the area as barely mapped (Cillo and Rubera) and press for ethical guardrails around autonomy, accountability, and non-discrimination (Hermann and Puntoni). Early findings are mixed in instructive ways: folding generative AI into retail chatbots raised perceived usefulness, human-likeness, and familiarity but left trust roughly unchanged and lifted privacy concern (Arce-Urriza et al.); disclosing that content was AI-made cut perceived brand authenticity (Brüns and Meißner); and in sensitive settings such as mental health, reliability and accountability become safety issues rather than conveniences (De Freitas et al.). Capability, in short, is outrunning trust.

Finally, the field still leans on convenient samples and says too little about difference — across age, digital literacy, privacy sensitivity, economic vulnerability, culture, and the stakes of the decision at hand. Vulnerable consumers, high-stakes purchases, algorithmic discrimination,

and the cumulative effect of years of AI-shaped persuasion all deserve far more attention than they have so far received.

7. Conclusion

Across a decade of research, one conclusion holds up better than any other: AI's influence on consumers is conditional, not automatic. Recommendation and advertising systems tend to work when they read as accurate and relevant; chatbots and assistants win people over when competence is paired with warmth and a trustworthy brand; and emotional response can break either way depending on how human — and how presumptuous — a system feels. Personalisation, anthropomorphism, and algorithmic pricing each cut both ways, capable of building engagement or eroding trust from the very same features.

Beyond drawing these findings together, this review makes a modest contribution of its own. Where much of the literature examines personalisation, trust, emotion, pricing, and manipulation in isolation, it gathers them into a single frame and advances one organising claim: that AI's influence on consumers is conditional, turning on perceived value, trust, and control rather than on technological capability. That lens helps reconcile results that otherwise look contradictory — why the same anthropomorphic cue can attract or repel, why the same algorithm can be embraced or resisted, and it reframes the ethical stakes around autonomy, fairness, and vulnerability that the newest generative systems make urgent. The aim is less to add another finding than to offer a way of reading the field as a whole.

What separates the two outcomes is not the sophistication of the technology but whether it delivers real value while leaving consumers feeling informed, fairly treated, and in control. That reframes the task for researchers and firms alike. The interesting question is no longer whether people will adopt AI — increasingly they already have — but whether AI-mediated markets will support genuine consumer choice or slowly substitute ever more refined influence for it.

Works Cited

Aguirre, E., et al. “Unraveling the Personalization Paradox: The Effect of Information Collection and Trust-Building Strategies on Online Advertisement Effectiveness.” Journal of Retailing, vol. 91, no. 1, 2015, pp. 34–49. https://doi.org/10.1016/j.jretai.2014.09.005.

Arce-Urriza, M., et al. “From Familiarity to Acceptance: The Impact of Generative Artificial Intelligence on Consumer Adoption of Retail Chatbots.” Journal of Retailing and Consumer Services, vol. 84, 2025, article 104234. https://doi.org/10.1016/j.jretconser.2025.104234.

Balakrishnan, J., and Y. K. Dwivedi. “Conversational Commerce: Entering the Next Stage of AI-Powered Digital Assistants.” Annals of Operations Research, 2021. https://doi.org/10.1007/s10479-021-04049-5.

Bonicalzi, S., et al. “Artificial Intelligence and Autonomy: On the Ethical Dimension of Recommender Systems.” Topoi, vol. 42, 2023, pp. 819–832. https://doi.org/10.1007/s11245-023-09922-5

Brüns, J. D., and M. Meißner. “Do You Create Your Content Yourself? Using Generative Artificial Intelligence for Social Media Content Creation Diminishes Perceived Brand Authenticity.” Journal of Retailing and Consumer Services, vol. 79, 2024, article 103790. https://doi.org/10.1016/j.jretconser.2024.103790.

Cabrera-Sánchez, J.-P., et al. “Online Recommendation Systems: Factors Influencing Use in E-Commerce.” Sustainability, vol. 12, no. 21, 2020, article 8888. https://doi.org/10.3390/su12218888.

Calvano, E., et al. “Artificial Intelligence, Algorithmic Pricing, and Collusion.” American Economic Review, vol. 110, no. 10, 2020, pp. 3267–3297. https://doi.org/10.1257/aer.20190623.

Canhoto, A. I., et al. “Snakes and Ladders: Unpacking the Personalisation–Privacy Paradox in the Context of AI-Enabled Personalisation in the Physical Retail Environment.” Information Systems Frontiers, 2023. https://doi.org/10.1007/s10796-023-10369-7.

Castelo, N., et al. “Task-Dependent Algorithm Aversion.” Journal of Marketing Research, vol. 56, no. 5, 2019, pp. 809–825. https://doi.org/10.1177/0022243719851788.

Chakraborty, D., et al. “Enhancing Trust in Online Grocery Shopping through Generative AI Chatbots.” Journal of Business Research, vol. 180, 2024, article 114737. https://doi.org/10.1016/j.jbusres.2024.114737.

Cheng, X., et al. “Exploring Consumers’ Response to Text-Based Chatbots in E-Commerce: The Moderating Role of Task Complexity and Chatbot Disclosure.” Internet Research, vol. 32, no. 2, 2021, pp. 496–517. https://doi.org/10.1108/INTR-08-2020-0460.

Cillo, P., and G. Rubera. “Generative AI in Innovation and Marketing Processes: A Roadmap of Research Opportunities.” Journal of the Academy of Marketing Science, 2024. https://doi.org/10.1007/s11747-024-01044-7.

Davis, F. D. “Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology.” MIS Quarterly, vol. 13, no. 3, 1989, pp. 319–340. https://doi.org/10.2307/249008.

De Freitas, J., et al. “Chatbots and Mental Health: Insights into the Safety of Generative AI.” Journal of Consumer Psychology, vol. 34, no. 3, 2024, pp. 481–491. https://doi.org/10.1002/jcpy.1393.

Dietvorst, B. J., et al. “Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err.” Journal of Experimental Psychology: General, vol. 144, no. 1, 2015, pp. 114–126. https://doi.org/10.1037/xge0000033.

Fan, Y., and X. Liu. “Exploring the Role of AI Algorithmic Agents: The Impact of Algorithmic Decision Autonomy on Consumer Purchase Decisions.” Frontiers in Psychology, vol. 13, 2022, article 1009173. https://doi.org/10.3389/fpsyg.2022.1009173.

Frank, D.-A., et al. “In Companies We Trust: Consumer Adoption of Artificial Intelligence Services and the Role of Trust in Companies and AI Autonomy.” Information Technology & People, vol. 36, no. 8, 2023, pp. 155–173. https://doi.org/10.1108/ITP-09-2022-0721.

Hassan, N., et al. “The Moderating Role of Personalized Recommendations in the Trust–Satisfaction–Loyalty Relationship: An Empirical Study of AI-Driven E-Commerce.” Future Business Journal, vol. 11, 2025, article 76. https://doi.org/10.1186/s43093-025-00476-z.

Hermann, E., and S. Puntoni. “Generative AI in Marketing and Principles for Ethical Design and Deployment.” Journal of Public Policy & Marketing, 2024. https://doi.org/10.1177/07439156241309874.

Huang, M.-H., and R. T. Rust. “A Strategic Framework for Artificial Intelligence in Marketing.” Journal of the Academy of Marketing Science, vol. 49, 2021, pp. 30–50. https://doi.org/10.1007/s11747-020-00749-9.

Karwatzki, S., et al. “Beyond the Personalization–Privacy Paradox: Privacy Valuation, Transparency Features, and Service Personalization.” Journal of Management Information Systems, vol. 34, no. 2, 2017, pp. 369–400. https://doi.org/10.1080/07421222.2017.1334467.

Logg, J. M., et al. “Algorithm Appreciation: People Prefer Algorithmic to Human Judgment.” Organizational Behavior and Human Decision Processes, vol. 151, 2019, pp. 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005.

Luguri, J., and L. J. Strahilevitz. “Shining a Light on Dark Patterns.” Journal of Legal Analysis, vol. 13, no. 1, 2021, pp. 43–109. https://doi.org/10.1093/jla/laaa006.

Mariani, M. M., et al. “AI in Marketing, Consumer Research and Psychology: A Systematic Literature Review and Research Agenda.” Psychology & Marketing, vol. 39, no. 4, 2022, pp. 755–776. https://doi.org/10.1002/mar.21619.

Mende, M., et al. “Service Robots Rising: How Humanoid Robots Influence Service Experiences and Elicit Compensatory Consumer Responses.” Journal of Marketing Research, vol. 56, no. 4, 2019, pp. 535–556. https://doi.org/10.1177/0022243718822827.

Pitardi, V., and H. R. Marriott. “Alexa, She’s Not Human but… Unveiling the Drivers of Consumers’ Trust in Voice-Based Artificial Intelligence.” Psychology & Marketing, vol. 38, no. 4, 2021, pp. 626–642. https://doi.org/10.1002/mar.21457.

Pizzi, G., et al. “I, Chatbot! The Impact of Anthropomorphism and Gaze Direction on Willingness to Disclose Personal Information and Behavioral Intentions.” Psychology & Marketing, vol. 40, no. 7, 2023, pp. 1372–1387. https://doi.org/10.1002/mar.21813.

Puntoni, S., et al. “Consumers and Artificial Intelligence: An Experiential Perspective.” Journal of Marketing, vol. 85, no. 1, 2021, pp. 131–151. https://doi.org/10.1177/0022242920953847.

Schaap, G., et al. “The ABC of Algorithmic Aversion: Not Agent, but Benefits and Control Determine the Acceptance of Automated Decision-Making.” AI & Society, 2023. https://doi.org/10.1007/s00146-023-01649-6.

Schepers, J., et al. “How Smart Should a Service Robot Be?” Journal of Service Research, vol. 25, no. 4, 2022, pp. 565–582. https://doi.org/10.1177/10946705221107704.

Susser, D., et al. “Online Manipulation: Hidden Influences in a Digital World.” Georgetown Law Technology Review, vol. 4, no. 1, 2019, pp. 1–45.

Venkatesh, V., et al. “Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology.” MIS Quarterly, vol. 36, no. 1, 2012, pp. 157–178. https://doi.org/10.2307/41410412.

Venkatesh, V., et al. “User Acceptance of Information Technology: Toward a Unified View.” MIS Quarterly, vol. 27, no. 3, 2003, pp. 425–478. https://doi.org/10.2307/30036540.

Wu, Z., et al. “The Impact of Algorithmic Price Discrimination on Consumers’ Perceived Betrayal.” Frontiers in Psychology, vol. 13, 2022, article 825420. https://doi.org/10.3389/fpsyg.2022.825420.

Xiong, Y., et al. “More Trust or More Risk? User Acceptance of Artificial Intelligence Virtual Assistant.” Human Factors and Ergonomics in Manufacturing & Service Industries, vol. 34, no. 3, 2023, pp. 190–205. https://doi.org/10.1002/hfm.21020.

Yeo, T. E. D., et al. “How Persuasive Is Personalized Advertising? A Meta-Analytic Review of Experimental Evidence of the Effects of Personalization on Ad Effectiveness.” Journal of Advertising Research, 2025. https://doi.org/10.1080/00218499.2025.2467763.

Yin, J., and X. Qiu. “AI Technology and Online Purchase Intention: Structural Equation Model Based on Perceived Value.” Sustainability, vol. 13, no. 10, 2021, article 5671. https://doi.org/10.3390/su13105671.

Yin, J., et al. “The Impact of AI-Personalized Recommendations on Clicking Intentions: Evidence from Chinese E-Commerce.” Journal of Theoretical and Applied Electronic Commerce Research, vol. 20, no. 1, 2025, article 21. https://doi.org/10.3390/jtaer20010021.

Limits of Arbitrage and Digital Collectible Pricing: Evidence from the CS: GO Skin Market By Jerry Chai

Abstract

This paper examines the limits to arbitrage in a digital collectible market. A popular video game, Counter-Strike 2, owns the largest and most valuable in-game cosmetic skin economy. An unexpected reform in October 2025 enabled an arbitrage opportunity that introduced a new creation mechanism that allowed players to exchange five weapon skins for a knife or gloves from the same collection that used to trade at fundamentally different prices. Using publicly available price data collected from SteamDT, this study analyzes the price movements of representative Covert weapon skins and glove skins before and after the shock. The results show that input skin prices increased sharply due to higher derived demand, while expected output values initially declined before partially recovering as market participants adjusted their expectations of future supply. The implied profit from trade-up conversions rapidly disappeared and eventually became negative, suggesting that the initial arbitrage opportunity was short-lived. These findings provide evidence that even in transparent digital marketplaces, arbitrage is constrained by behavioral responses, speculative demand, and market frictions.

Introduction

A central benchmark in finance is that competitive markets incorporate publicly available information into prices, leaving little scope for predictable abnormal returns (Fama, 1970). Yet a large body of evidence documents systematic departures from this benchmark, including short-run underreaction and longer-run reversals that resemble overreaction (De Bondt and Thaler, 1985; Barberis, Shleifer, and Vishny, 1998; Daniel, Hirshleifer, and Subrahmanyam, 1998; Hong and Stein, 1999).

A leading explanation emphasizes that mispricing can persist not only because beliefs are distorted, but also because arbitrage is risky and resource-constrained. Canonical "limits of arbitrage" mechanisms include noise-trader risk, funding and horizon constraints, and synchronization frictions that deter traders from aggressively correcting mispricing (De Long et al., 1990; Shleifer and Vishny, 1997; Abreu and Brunnermeier, 2002; Gromb and Vayanos, 2010). Importantly, these forces are predicted to be especially salient in settings with subjective valuations and restricted shorting, where disagreement and sentiment can have larger price effects (Miller, 1977; Baker and Wurgler, 2006).

We study these themes in a high-frequency retail marketplace with transparent trading rules: the Steam Community Market for Counter-Strike weapon "skins." These digital items have no cash flows, but they provide consumption and status value inside the game, and they are actively traded as investable collectibles. Recent work has begun to characterize the risk–return properties of large cross-sections of such virtual assets, suggesting they can behave like speculative "alternative" assets.

Our empirical setting is a sharp, platform-driven shock that changed the "production technology" of the rarest cosmetic items. In October 2025, Counter-Strike 2 expanded Trade-Up Contracts so that submitting five Covert-grade ("Red") weapon skins yields a guaranteed knife or gloves from an associated collection. Consistent with a derived-demand channel, Covert input skins appreciated while many knives and gloves declined as market participants revised expected future supply upward. The official patch notes reported this new knife/glove trade-up functionality for five Covert items, making the rule change common knowledge.

Three patterns emerge. First, the update produced immediate and large cross-item repricing consistent with a shock to both derived demand for inputs and expected supply of outputs. Second, the anticipated post–trade-lock crash did not occur on the focal date; instead, prices reached their local minimum earlier, as traders preemptively sold ahead of the expected event. Third, the implied conversion P&L deteriorated rapidly—transitioning from strongly positive to negative—indicating that any initial arbitrage opportunity was short-lived and ultimately reversed as input prices rose and output prices recovered.

These findings point to both behavioral and friction-based mechanisms: the early "bottom" is consistent with a self-defeating, focal narrative (Abreu and Brunnermeier, 2002), while resale locks and limited trading instruments impose clear limits to arbitrage that can sustain expectation-driven price moves (Shleifer and Vishny, 1997; Gromb and Vayanos, 2010). Our contribution is a clean, rule-based shock with publicly known conversion rules and mechanically delayed supply, allowing retail coordination to be inferred from price dynamics around the expected event date.

Background

The CSGO Skin Market

Counter-Strike: Global Offensive (CSGO) is a popular first-person shooter game developed by Valve. In the game, players compete in teams, most commonly as terrorists and counter-terrorists, with the objective of completing missions such as planting or defusing a bomb. Since its release, CSGO has grown into one of the most influential e-sports titles in the world, with millions of active players and a large online community.

One of the most popular features of CSGO is its skin system. Skins are cosmetic designs that change the appearance of weapons, knives, or gloves in the game. They do not affect gameplay or give players any competitive advantage; instead, they allow players to personalize their weapons and express their identity. Skins vary in rarity, design, and condition, which makes some of them highly valuable.

Players use skins in several ways. Many players equip skins simply for visual enjoyment during matches, while others collect rare skins as a hobby. Some players treat skins as digital assets, holding onto valuable items or showcasing them in their inventories. Because certain skins are limited or difficult to obtain, they can become symbols of status within the community.

CSGO skins can also be traded, forming a large in-game economy. Players trade skins through Steam's official trading system, where items can be exchanged directly between accounts. Skins can also be bought and sold on the Steam Community Market using wallet funds. Over time, trading has become an important part of the CSGO ecosystem, with skin values influenced by rarity, demand, and game updates. This trading system eventually led to major policy changes that reshaped how skins, especially high-value items like knives, are exchanged.

A Shock of New Arbitrage Opportunities

In October 2025, Counter-Strike 2 (CS2) implemented a major change to the production technology of high-end cosmetic items. Before the update, knives and gloves were commonly viewed as the scarcest and most valuable cosmetics because acquisition was largely limited to stochastic "case" openings or costly purchases on the Steam Community Market.

The update introduced a new, more deterministic pathway: players could submit five Covert-grade ("Red") weapon skins—the highest firearm-skin rarity—into a Trade-Up Contract to obtain a knife or gloves, with the outcome determined by the input skins' case collection. Economically, this reform reduced both uncertainty and effective marginal cost by converting what had been a lottery-like procurement process into an input-based production process. Prices adjusted rapidly to the implied shifts in derived demand and expected supply. Covert skins appreciated sharply because they became required inputs for crafting, while many knives, gloves, and other items declined as traders revised their expectations of future supply upward. A small set of niche items, such as certain rare souvenirs plausibly held by concentrated groups, appeared more insulated from the broad repricing.

Crucially, institutional frictions shaped the timing of supply. Newly crafted items were subject to a 7-day trade lock before they could be resold, so the market did not immediately receive a surge of new knives and gloves. Traders therefore coordinated on a focal prediction: roughly seven days after launch (October 30), trade locks would expire en masse, triggering a discrete supply wave and a corresponding price drop.

The realized dynamics deviated from this forecast because expectations became self-defeating. Anticipation of an October 30 crash induced preemptive selling on October 28–29, which generated the actual local minimum before the expected event date. When October 30 arrived, the predicted collapse did not occur; prices stabilized and many skins rebounded, consistent with the market absorbing incremental supply more smoothly than expected.

Data Analysis

Price data are obtained from SteamDT, a website that provides public information on transaction-based listing prices that reflect real market supply and demand among retail participants. Because all trades occur within the same rule-based system, prices incorporate shared information about game updates, trade restrictions, and crafting mechanics. We hand-collected data from October to November 2025 including the prices of six different gloves and guns in the same family before and after the reform.

Results

Input costs (Covert "gun" skins). The input series exhibits a strong and persistent upward drift over the sample window, consistent with a sustained increase in derived demand for Covert skins as trade-up ingredients. Prices begin in late October at roughly ¥9,000 RMB, rise quickly to approximately ¥12,000, and continue trending upward to about ¥14,000 by early November. The most salient feature is a discrete step-up around November 9, when input prices jump sharply to above ¥21,000, indicating either a major revision in beliefs about future profitability/scarcity or a rapid tightening of available inventory at prevailing quotes. After this discontinuity, the series continues to increase at a slower pace, reaching roughly ¥25,000–¥26,000 by November 23. Taken together, the pattern is not mean-reverting within the window; instead it resembles a regime shift followed by continued appreciation, implying that the cost side of the trade-up technology became progressively more expensive over time.

Expected output value (gloves). In contrast, the expected value of the output (gloves) displays a pronounced V-shaped trajectory. It starts at a relatively high level—around ¥24,000 RMB—and then declines steeply to approximately ¥15,500 by late October. This initial drop is consistent with an immediate negative valuation effect from the update: market participants likely priced in a substantial expansion of future glove supply once crafting became feasible. Following the trough, the series stabilizes and gradually recovers through November, rising toward ¥18,000 by mid-month. After a brief and modest pullback, the recovery resumes, and the series ends near ¥19,000 by late November. Relative to the input series, the output series shows weaker trend growth and much smaller late-period acceleration, implying that any recovery in glove valuations was limited compared with the rapid inflation in the cost of required inputs.

Implied conversion P&L. The profit-and-loss series—the implied surplus from converting inputs into expected glove outcomes—captures the interaction of these two movements and reveals a rapid erosion of apparent arbitrage profitability. Early in the sample, P&L is strongly positive at roughly ¥15,000 RMB, indicating that, at prevailing prices, the conversion technology appeared highly lucrative. That profitability collapsed quickly: by late October, P&L fell to near zero, suggesting that competitive entry (or anticipatory bidding for inputs) rapidly arbitraged away the initial surplus. From early November onward, the series turns negative and continues declining, meaning the conversion becomes loss-making at quoted prices. Losses deepen steadily through mid-to-late November, reaching roughly –¥8,000 RMB by the end of the window. Conceptually, the series indicates that the market moved from an apparent "free lunch" to sustained negative expected returns—driven primarily by the sharp escalation in input prices outpacing the partial recovery in expected glove values. This divergence is consistent with a short-lived arbitrage window followed by over-competition for inputs and/or frictions (e.g., trade locks, inventory constraints, and limited ability to short) that prevent the usual no-arbitrage linkage from holding tightly at every point in time.

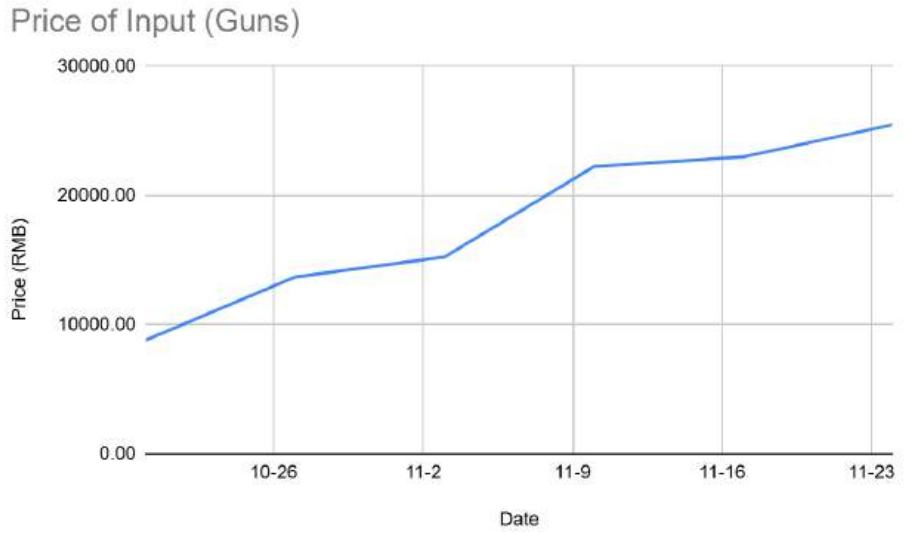


Figure 1. Input prices of a potential arbitrage

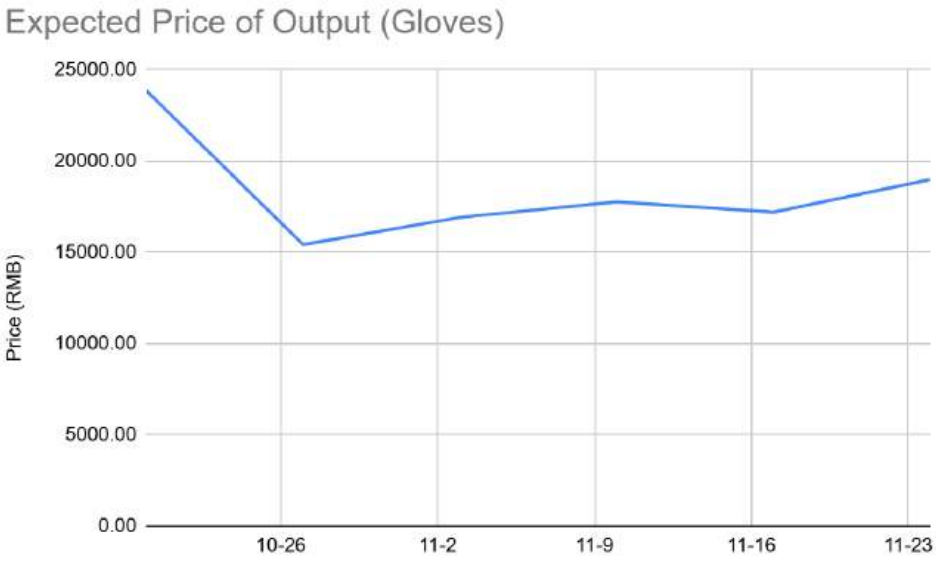


Figure 2. Output prices of a potential arbitrage

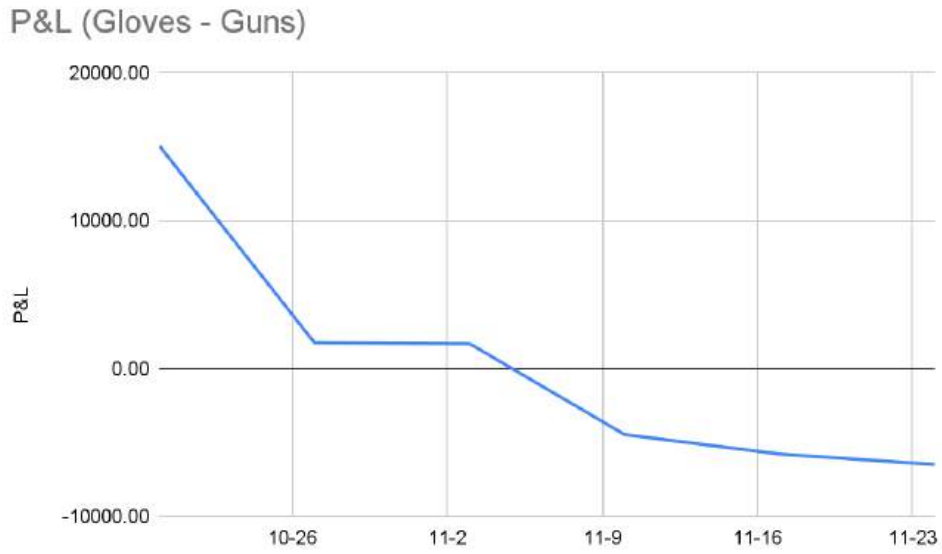


Figure 3: Implied profit & loss from an arbitrage

Conclusion

This paper examines how the October 2025 trade-up reform changed price dynamics in the CS: GO skin market. The update created a new production pathway, which initially generated an apparent arbitrage opportunity between input skins and output items.

We find three main results. First, prices adjusted immediately: required input skins rose sharply while many outputs fell. Second, input prices continued trending upward instead of stabilizing, suggesting sustained speculative demand and inventory constraints. Third, the implied conversion profit quickly collapsed and turned negative as traders competed for inputs and market frictions limited effective arbitrage.

Overall, the evidence shows that even in a transparent digital marketplace, prices can deviate from frictionless predictions. Behavioral forces and institutional constraints interact to produce short-lived arbitrage, trend amplification, and persistent mispricing — patterns similar to those seen in traditional financial markets.

Works Cited

Abreu, D., & Brunnermeier, M. K. (2002). Synchronization risk and delayed arbitrage. Journal of Financial Economics. ScienceDirect

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. Journal of Finance.

Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. Journal of Financial Economics. ScienceDirect

Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under- and overreactions. Journal of Finance. Wiley Online Library

De Bondt, W. F. M., & Thaler, R. (1985). Does the stock market overreact? Journal of Finance. Wiley Online Library

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. (1990). Noise trader risk in financial markets. Journal of Political Economy. JSTOR

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. Journal of Finance. JSTOR

Gromb, D., & Vayanos, D. (2010). Limits of arbitrage: The state of the theory. Annual Review of Financial Economics. LSE Personal Web Pages

Hong, H., & Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. Journal of Finance. Wiley Online Library

Miller, E. M. (1977). Risk, uncertainty, and divergence of opinion. Journal of Finance. Wiley Online Library

Shleifer, A., & Vishny, R. W. (1997). The limits of arbitrage. Journal of Finance. Wiley Online Library

Reichenbach, F. (2025). Skin in the game: The returns of digital assets from Counter-Strike skins. Finance Research Letters, Volume 83, October 2025, 107670

Can Indian Steel Companies Remain Competitive While Transitioning Toward Sustainability? By Gautam Agrawal

Abstract

India’s steel sector makes the country more industrial. It faces pressure to curb pollution, use resources more efficiently, and follow new environmental rules. However, being sustainable is costly, and waiting too long to change would cause bigger problems with finances, regulations, and product sales. The present study examines whether Indian steel companies can remain competitive as they move toward greener steelmaking. It explores how companies manage the costs of making their factories cleaner and what they gain in return. Factors such as energy savings, investor financing, regulatory readiness, and product differentiation are key to making companies sustainable. The study used information from reports by Tata Steel, JSW Steel, Steel Authority of India Limited, Jindal Steel and Power Limited, and AM/NS India from 2021 to 2023. It was found that going green initially raised costs due to new machinery, improved energy efficiency, and changes to older factories. However, companies that made green plans part of their main strategy performed better in the long run. This shift can help with compliance, attracting investment, building resilience, and selling into better markets.

Keywords

Sustainable, Steel Industry, India, TATA Steel and Greener steel making

Introduction

India’s steel story reflects the country's broader economic journey. The sector has grown from a relatively small contributor to global output to become the world’s second-largest steel producer, overtaking Japan in 2018 (“India’s Steel Industry”). This change is crucial to India's path to modernization, self-sufficiency, and sustainable industrialization. The industry has grown significantly between 2019 and 2023. During this period, India’s steel production grew at 6% CAGR, significantly ahead of China’s 1%. In the past five years, the world has generated nearly 62 million tonnes of steel capacity, of which India accounts for 6%. Notably, India and ASEAN are projected to account for nearly 89% of Asia’s steelmaking capacity additions ("India's Steel Industry”). This demonstrates the country’s strategic position in the global market, backed by domestic manufacturing initiatives like ‘Atmanirbhar Bharat’. India is not only producing steel but is evolving a model to reduce import dependence and increase export competitiveness ("India's Steel Industry”).

The National Steel Policy (NSP) 2017 has been crucial to this success, setting ambitious targets to achieve a crude steel capacity of 300 million tonnes by 2030-31, with expected crude steel production of 255 million tonnes and finished steel production of 230 million tonnes. The policy aims to make India self-sufficient in steel production while catering to domestic and international markets, emphasizing capacity expansion, quality improvement, and the adoption of sustainable, energy-efficient practices ("India's Steel Industry”). Crude steel production

increased to 144.299 MT in 2023-24 from 109.137 MT in 2019-20, registering a strong growth of 13.4% over the previous year (127.197 MT in 2022-23). An expansion in the domestic steel industry's capacity from 142.299 MT per annum in 2019-20 to 179.515 MT per annum in 2023-24 supported production growth. Capacity utilization also increased to 81% during the same period ("India's Steel Industry").

However, India's steel growth is not only about capacity expansion but also about enhancing domestic consumption. The total consumption of finished steel has increased from 100.171 MT in 2019-20 to 136.291 MT in 2023-24, indicating a strong domestic demand with a growth rate of 13.7% over the previous year. Industry estimates, including a conservative 6% CAGR in steel demand through FY27, indicate the Indian steel sector is likely to meet, if not exceed, demand. This growth can be credited to the strategic policies and industry-led initiatives focused on capacity expansion and the development of high-quality steel products ("India's Steel Industry").

Steel is a deregulated sector, and the government acts as a facilitator by creating a conducive policy environment for its development. The biggest challenge in steel production is to produce steel in the greenest possible way, without damaging the environment (Birat, 2016). Industrialists have to be mindful of compliance with evolving environmental regulations and of keeping freshwater consumption in check while continuing to produce large amounts of steel. All these factors have created pressure on the steel industry worldwide to reduce its water consumption and improve wastewater quality through new technologies (Conejo et al.). Indian steel production is expected to grow by at least 100 million tonnes (Hall et al., Federation of Indian Chambers of Commerce and Industry). The iron and steel industry is the most emission-intensive heavy industrial sector, accounting for 12% of the country's emissions (International Energy Agency, India Energy Outlook 2021). Increased production in this industry could lead to increased CO2 emissions.

The Indian domestic industrial emissions intensity is currently about 2.5 tonnes of CO2 per tonne of crude steel (t CO2/tcs), higher than the global average of about 1.83 t CO2/tcs (Hall et al, World Steel Association). Unlike many other countries, India is a major coal consumer in iron and steelmaking, with coal accounting for 90% of the sector's energy consumption (International Energy Agency, Iron and Steel Technology Roadmap). To address emissions, India's Ministry of Steel has tentatively proposed a nationally determined contribution for the industry under both business-as-usual and mitigation scenarios (Ministry of Steel). According to the Ministry's report, the emissions in 2030 without decarbonization could be 200 Mt higher than the steel sector scenario with decarbonization strategies (Ministry of Steel).

Carbon is the main reducing agent in iron manufacture. There are efforts to mitigate the CO2 emissions from this carbon use, either by replacing it with hydrogen or capturing and storing it using carbon capture and sequestration (CCS) technologies (Boldrini et al., 2024). Bahzad et al. (2019) investigated an iron-based chemical looping process for hydrogen co-generation and iron product reduction. Pimm et al. (2021) evaluated hydrogen generation and storage technologies for the steel industry. Fan and Friedmann (2021) studied techno-economic aspects of hydrogen use, biomass use, grid decarbonization, and carbon capture and storage in

the decarbonization of the I&S industries. Filho et al. (2022) proposed combining solid-state direct reduction and hydrogen plasma reduction to improve process efficiency. Devlin et al. (2023) evaluated opportunities for green hydrogen-based steel production.

Increased demand for the product causes environmental pollution by releasing fuel gases directly into the atmosphere and by improperly disposing of solid waste. This affects the health of human beings and other organisms due to the direct discharge of wastewater into water bodies (Das and Mishra). It was only in the 1990s that industrial sectors, including the construction industry, started to understand the environmental costs of their work (Andrade and Bragança, 2011). It is an enormous task for industries to manage their waste and by-products in environmentally friendly ways. The only way to solve this is through sustainable development. The term "sustainable" is Latin and was first used in the publication of the Brundtland report "Our Common Future" in 1987 (Andrade and Bragança, 2011). Sustainability is an independent word itself, which can be described as "to sustain either with time, efficiency, durability, less pollution, cost-effectiveness, etc." A wider concept of sustainability can be addressed through the lens of the environment. This addresses the ecological stability of human habitation as a basic requirement, as well as the risks to public services posed by investments in ecosystem services. There is increasing concern regarding the resilience, productivity and economic diversity of the natural world. Environmental factors must also be addressed as economic assets, also referred to as natural and social capital.

The sustainability transition in India's steel sector is widely seen as costly and potentially competitiveness-reducing in the short term, especially because low-carbon routes require high capital spending, expensive clean hydrogen, and large-scale technology upgrades. At the same time, several sources argue that delaying this transition could create greater competitiveness risks later through carbon border costs, tighter disclosure rules, and lost access to green finance and premium markets. Existing research on the steel industry has mainly focused on environmental aspects, such as carbon emissions and resource utilization, or on technical solutions for cleaner production. However, literature on the influence of the transition towards sustainability on the business competitiveness of Indian steel companies is limited. This is especially true when it comes to managing the costs of sustainable practices whilst still maintaining profitability, market position and operational efficiency. This study addresses this gap by examining how Indian steel companies thrive while remaining sustainable, and whether they can remain competitive as they move towards more sustainable practices.

Literature Review

The steel industry is among the most resource-intensive sectors of the global economy, responsible for approximately 7–9% of total global CO2 emissions (World Steel Association, 2023). As environmental concerns grow and regulations tighten, steelmakers around the world are making sustainability a major strategic priority. In this regard, Environmental, Social, and Governance (ESG) frameworks have gained significant traction as tools for measuring and communicating corporate sustainability performance (World Steel Association).

ESG issues in the steel industry include carbon emissions and energy use, water consumption and wastewater treatment, waste production and circular economy approaches, labor safety, community involvement, and corporate governance at the board level (Schramade, 2016). Institutional investors are increasingly incorporating ESG metrics into their capital allocation decisions, and strong ESG performance is becoming a prerequisite for access to international finance (Friede, Busch & Bassen, 2015). As a result, steel companies are under increasing pressure from shareholders, regulators and civil society to put sustainability at the core of their business strategies.

On the subject of sustainability practices, the world's leading steel producers have taken up several decarbonisation strategies. They include the transition from blast furnace–basic oxygen furnace (BF-BOF) routes to electric arc furnaces (EAF) powered by renewable energy, the adoption of hydrogen-based direct reduced iron (H-DRI) technologies and investments in carbon capture, utilisation and storage (CCUS) infrastructure (Hasanbeigi et al., 2016). Beyond climate-related measures, sustainability practices increasingly encompass supply chain transparency, responsible sourcing of raw materials, and robust social impact programmes aligned with the United Nations Sustainable Development Goals (SDGs) (World Steel Association; Tata Steel Limited) .

- **Competitiveness in Business**

To remain competitive in the market, companies strive to outperform their competitors. This term is most commonly used in business and refers to a company's ability to achieve the highest possible results within its environment (Porter, 1990). Companies within an industry are evaluated on four different criteria of efficiency: cost, operational, innovation, and market. D'Costa (1999), in his work on the global steel industry, emphasizes that commodity producers compete on cost and that price differentials among producers can be a deciding factor for market share and survival. Several traditional parameters, such as access to low-cost raw materials, energy, and labor, as well as the potential to achieve economies of scale, have underpinned competitive advantage. However, as markets change, technological capability and innovation have become increasingly important determinants of competitiveness (Porter; D'Costa; Lall). The steel companies that invest in cutting-edge process technology, digitization, and product differentiation, such as high-strength and special steels, can capture value-added markets and reduce exposure to commodity price volatility (Lall, 2001). Furthermore, market competition in the global steel industry has been reshaped by the rise of large emerging-market producers, trade policy interventions, and shifting demand patterns driven by infrastructure development, automotive manufacturing, and the green energy transition.

- **Sustainability vs Competitiveness**

One of the major issues in current business literature is the interrelation between sustainability and competitiveness, including whether and how environmental and social responsibilities can be implemented without reducing a company's profitability. The classic point

of view in this discussion was expressed by Friedman (1970), who stated that social responsibility is not cost-effective for shareholder value. Within this framework, firms in cost-competitive industries such as steel face a structural challenge. To improve sustainability, their operations will temporarily increase costs, thereby reducing their current profitability and their potential to compete in the market.

However, contradictory views have received empirical support. Porter and van der Linde (1995) put forward the "Porter Hypothesis" that well-designed environmental regulation could stimulate innovation, increase resource efficiency and ultimately strengthen competitive position. More recent work suggests that sustainability and competitiveness are not mutually exclusive. Eccles, Ioannou & Serafeim (2014) find that companies that actively manage ESG risks tend to have better long-term financial performance, lower regulatory risk, and greater resilience to market shocks. However, balancing the need for short-term profits with the promise of long-term sustainability remains a difficult act, especially in capital-intensive heavy industries with long time horizons for green investment paybacks.

- **Indian Steel Industry**

India is the second-largest producer of steel in the world, producing over 125 million tonnes (MT) of crude steel per annum as of 2023, and the sector is a key component of the country's overall industrial and infrastructure development agenda (Ministry of Steel, Government of India, 2023). The Indian steel landscape comprises large integrated producers and secondary producers operating EAFs and induction furnaces, with a growing private-sector presence.

Among the major players, Tata Steel is India's largest private-sector steel company, with an integrated production capacity of more than 21 million tonnes per annum (MTPA) across its domestic and international operations. Tata Steel demonstrated an early commitment to sustainability by establishing ambitious net-zero carbon targets and investing in hydrogen steelmaking pilots in Europe (Tata Steel, 2023). JSW Steel, another major private producer, has prioritized aggressive capacity expansion alongside efficiency improvements and the integration of renewable energy, with an installed capacity of nearly 28 MTPA (JSW Steel, 2023). The Steel Authority of India Limited (SAIL), a public sector undertaking, remains the largest state-owned steelmaker, operating five integrated steel plants, despite ongoing challenges in modernizing aging infrastructure and improving operational efficiency (SAIL Annual Report, 2022-23). Other notable producers include Jindal Steel & Power (JSPL) and ArcelorMittal Nippon Steel India (AM/NS India), both of which have implemented sustainability and modernization initiatives to varying degrees. These firms, taken together, form the central empirical terrain against which this study investigates questions of sustainability, competitiveness, and their interaction in India.

Methodology

Research Design

The study is a qualitative secondary research design. It does not collect primary data through surveys or field research; rather, it analyses existing published materials to examine how major Indian steel companies balance their sustainability commitments with the imperatives of business competitiveness. This approach is suitable for the complexity and contextual nature of the research questions, which require interpretive analysis of corporate strategy, policy discourse, and industry dynamics rather than statistical measurement (Bryman, 2016). The structure of secondary research enables the holistic and comparative study of many firms across a shared set of thematic dimensions.

Data Sources

The analysis is based on four types of secondary sources. Firstly, the annual reports published by Tata Steel, JSW Steel, SAIL, JSPL and AM/NS India for the year 2021-23 provide detailed information on the financial performance, operational strategy and capital investment decisions of the companies. Second, sustainability and ESG reports from these same companies provide systematic disclosure of environmental performance data, decarbonization targets, energy consumption, and social impact data. Third, the analytical framework draws on theoretical and empirical grounds from peer-reviewed journal articles in disciplines such as business management, industrial ecology, and environmental economics. Fourth, Industry reports from organizations such as the World Steel Association, the Climate Policy Initiative, and the Ministry of Steel, Government of India, provide industry-wide data on production trends, technology adoption, and policy context.

Analytical approach

Thematic analysis was carried out on data from these sources using the framework developed by Braun and Clarke (2006) to identify common patterns, differences, and tensions across the five companies. The analysis was based on three main themes: sustainability strategies and environmental performance, business competitiveness and financial resilience, and the relationship between sustainability and competitive outcomes. Where possible, comparisons were made across companies to assess how they differentiated their responses to the shared challenge of decarbonization in a cost-sensitive industry environment.

Discussion

Among the five companies studied, there is considerable variation in the ambition and pace of sustainability initiatives. Tata Steel has the highest level of strategic commitment and implementation. The company has a group-wide net-zero emissions target for 2045 and a formal decarbonization governance framework which includes a dedicated Center of Excellence for GHG Emissions (Tata Steel, 2023). Its most tangible effort is to replace a blast furnace at its Netherlands facility with a DRI-EAF unit that will be partially powered by green hydrogen, aiming for a 40% reduction in CO_2 emissions from that site. Tata Steel is also the first Indian

steel company to join the Leadership Group for Industry Transition (LeadIT), and to sign up to the group's active alignment with global decarbonization coalitions.

JSW Steel has made significant progress despite not being technologically advanced. The company has vowed to achieve net-zero carbon by 2050. It has signed agreements to purchase about 996 MW of wind and solar power in Phase I, increasing to 1,470 MW in Phase II, via its subsidiary JSW Energy (JSW Steel, 2023). In 2022, its Carbon Disclosure Project (CDP) rating was raised to 'A', which indicates enhanced transparency and performance in climate-related disclosures. JSW Steel also reports on Scope 3 emissions and has piloted circular-economy measures, such as converting industrial slag by-products into civil-engineering aggregates. Digitalization and AI-driven process optimization have been identified as further levers to reduce energy intensity.

SAIL is a public sector undertaking and has legacy infrastructure and public procurement obligations, so it has a more incremental sustainability trajectory. The company has laid out a three-pronged decarbonization roadmap that aims for 2.19 tCO_2 per tonne of crude steel in the 2030-31 period and a net-zero target in line with India's national commitment of 2070. It has set up 12.58 MW of solar power and plans to commission 384 MW by 2028-29 (SAIL, 2023). Energy-efficient technologies like Coke Dry Cooling Plants and Waste Heat Recovery systems have been implemented, and specific water consumption has decreased by about 14% over five years. However, SAIL's green transition remains constrained by the capital-intensive nature of its modernization program and the challenges of operating within a public-sector governance framework.

Business Strategies and Competitiveness Challenges

The five companies face similar competitiveness challenges, but have differing strategic priorities. The biggest problem is the price pressure. Green steel technologies – hydrogen-DRI, EAF transitions, CCUS – have significantly higher capital and operating costs than traditional BF-BOF routes. The price premium for green steel in export markets has not yet been consistently achieved in India's largely price-sensitive domestic market (Climate Policy Initiative, 2023). This creates a structural tension: investing in sustainability entails higher short-term costs, but not investing carries the risk of competitive exclusion from premium global markets in the medium term, especially as the EU's Carbon Border Adjustment Mechanism (CBAM) begins to impose carbon costs on steel imports.

The second major challenge is technology adaptation. Innovation in the steel sector is characterized by high capital intensity, long development cycles and significant technology risk, which do not easily align with conventional investor risk-return expectations (Climate Policy Initiative, 2023). One such example is JSW Steel's strategy to tackle this twin challenge through backward integration for cost optimization and sustainability, efficiency improvements through AI, and the phased procurement of renewable energy. At least in Europe, Tata Steel's willingness to invest in demonstration-scale hydrogen steelmaking suggests a more segmented global portfolio plan. Another dimension of asymmetric competitive pressure comes from global

competition from Chinese overcapacity and state-backed European green steel producers, particularly for SAIL in its core domestic markets.

Competitiveness Impact

The link between sustainability investment and competitive performance in the Indian steel context is complex and non-linear. The incremental capex required to adopt green technology will hurt profitability in the near term, especially for SAIL, which has lower margins than its private peers. However, for JSW Steel and Tata Steel, sustainability investments are starting to yield competitive dividends. The CDP 'A' rating of JSW Steel also improves its access to international capital markets and ESG credentials with multinational customers (JSW Steel, 2023). Tata Steel's ESG rankings with Sustainalytics and MSCI have helped it become a preferred supplier in European markets where sustainability credentials are becoming procurement requirements (Tata Steel, 2023).

Companies spending more aggressively on green transitions appear better positioned in the long run. The EU CBAM will extend steel imports from 2026, putting high-carbon producers at a direct cost disadvantage. Proven ESG performance is becoming a prerequisite for access to premium market segments – automotive, aerospace, specialty steel. Moreover, the cost trajectory of green technologies, particularly renewables and green hydrogen, indicates that early movers will have learning curves that late adopters will not (Climate Policy Initiative, 2023). Therefore, innovation appears to be the key mediating variable between sustainability investment and competitive advantage.

The evidence from this comparative analysis suggests that sustainability presents both challenges and opportunities for competitiveness in the Indian steel sector, with the balance shifting over time. In the short term, sustainability investment is a real cost burden due to the high upfront costs of green technology, limited domestic carbon pricing, and the absence of a green steel premium in the Indian market. However, over the medium-to-long term, companies that have proactively embedded sustainability into their core business strategy are creating durable competitive advantages: differentiated product portfolios, preferential access to ESG-linked finance, reduced regulatory risk exposure, and improved brand equity in premium export markets. This trajectory is supported by the broader literature, which suggests that firms that integrate environmental and business strategy achieve better outcomes than those that treat sustainability as a compliance obligation (Eccles, Ioannou & Serafeim, 2014; Porter & van der Linde, 1995).

Conclusion

The study explored the connection between sustainability and competitiveness among five leading Indian steel companies: Tata Steel, JSW Steel, SAIL, JSPL, and AM/NS India. The research was based on company annual reports, ESG disclosures, academic literature, and industry analyses. The findings show that sustainability creates both opportunities and challenges

for competitiveness and also that the balance between them is shaped primarily by the quality of strategic integration and the degree of innovation investment.

Companies that have embedded sustainability into their core business strategy – especially Tata Steel and JSW Steel – are better positioned to capture the long-term value of the green transition: improved ESG ratings, preferential access to capital, compliance readiness for international carbon regulation, and access to premium market segments. SAIL's legacy assets and the public sector's governance dynamics make its transition more difficult. However, with its phased decarbonization plan and renewable expansion plans, it demonstrates that it is on the path to change. The evidence from all companies supports Porter and van der Linde's (1995) argument that well-designed sustainability strategies need not detract from, but can actually improve, competitive positioning, provided they are coupled with real investment in innovation and efficiency.

Recommendations

Based on the findings of this study, three strategic recommendations are put forward for Indian steel producers to balance the imperatives of sustainability and competitiveness. Most importantly, boost investment in green technology. Companies need to prioritize capital allocation to low-carbon steelmaking technologies, especially EAF transitions, renewable energy procurement, and hydrogen-DRI pilots, to build early-mover advantages before international carbon regulation changes market access conditions. Public-private partnerships and blended finance instruments should be used to share the cost burden of deploying technology, especially for capital-intensive breakthrough technologies and where traditional investors remain risk-averse.

Second, improve ESG integration throughout the value chain. This should not be limited to production processes but should also include procurement, logistics, product design, and supply chain governance. Strong ESG disclosure, aligned with international standards such as the GRI, ISSB, and World Steel Association guidelines, enhances investor confidence, improves premium market positioning, and prepares companies for mandatory reporting requirements that are becoming more common in key export markets.

Third, use sustainable innovation as a strategic differentiator. Companies should invest in research and development partnerships with academic institutions, industry consortia, and government bodies to develop next-generation steelmaking solutions and build the absorptive capacity required to commercialize emerging technologies as they mature. Innovation in digital processes, circular material flows and product sustainability credentials will increasingly shape competitive positioning in domestic and global steel markets.

Works Cited

Bhardwaj, Nishant, et al. "Carbon Abatement Options for Large Iron and Steel Plants in India." Journal of Cleaner Production, vol. 486, 2025, article 144505. https://doi.org/10.1016/j.jclepro.2024.144505.

Birat, Jean-Pierre. "Steel Cleanliness and Environmental Metallurgy." Metallurgical Research & Technology, vol. 113, no. 2, 2016, article 201. https://doi.org/10.1051/metal/2015050.

Braun, Virginia, and Victoria Clarke. "Using Thematic Analysis in Psychology." Qualitative Research in Psychology, vol. 3, no. 2, 2006, pp. 77–101. https://doi.org/10.1191/1478088706qp063oa.

Bryman, Alan. Social Research Methods. 5th ed., Oxford UP, 2016.

Conejo, Alberto N., et al. "A Review of the Current Environmental Challenges of the Steel Industry and Its Value Chain." Journal of Environmental Management, vol. 259, 2020, article 109782. https://doi.org/10.1016/j.jenvman.2019.109782.

D'Costa, Anthony P. The Global Restructuring of the Steel Industry: Innovations, Institutions and Industrial Change. Routledge, 1999.

Das, Lipsa, and Sarita Mishra. "Environmental Sustainability and Steel Manufacturing Industries of Odisha, India." International Journal of Environment and Climate Change, vol. 15, no. 7, 2025, pp. 234–251. https://doi.org/10.9734/ijecc/2025/v15i74927.

Eccles, Robert G., et al. "The Impact of Corporate Sustainability on Organizational Processes and Performance." Management Science, vol. 60, no. 11, 2014, pp. 2835–2857. https://doi.org/10.1287/mnsc.2014.1984.

Federation of Indian Chambers of Commerce and Industry. Indian Secondary Steel Industry: Opportunities and Challenges. Oct. 2016, https://ficci.in/study_details/20782.

Friede, Gunnar, et al. "ESG and Financial Performance: Aggregated Evidence from More than 2000 Empirical Studies." Journal of Sustainable Finance & Investment, vol. 5, no. 4, 2015, pp. 210–233. https://doi.org/10.1080/20430795.2015.1118917

Friedman, Milton. "A Friedman Doctrine: The Social Responsibility of Business Is to Increase Its Profits." The New York Times Magazine, 13 Sept. 1970, https://www.nytimes.com/1970/09/13/archives/a-friedman-doctrine-the-social-responsibility-of-business-is-to.html.

Hall, Will, et al. Towards a Low Carbon Steel Sector: Overview of the Changing Market, Technology and Policy Context for Indian Steel. The Energy and Resources Institute, 2020, https://shaktifoundation.in/wp-content/uploads/2020/01/Towards-a-Low-Carbon-Steel-Sector-Report.pdf.

Hasanbeigi, Ali, et al. "Comparison of Carbon Dioxide Emissions Intensity of Steel Production in China, Germany, Mexico, and the United States." Resources, Conservation and Recycling, vol. 113, 2016, pp. 127–139. https://doi.org/10.1016/j.resconrec.2016.06.008.

International Energy Agency. India Energy Outlook 2021. 2021, https://www.iea.org/reports/india-energy-outlook-2021.

International Energy Agency. Iron and Steel Technology Roadmap: Towards More Sustainable Steelmaking. 2020, https://www.iea.org/reports/iron-and-steel-technology-roadmap.

Johnson, Sydney, et al. “Environmental and Economic Evaluation of Decarbonization Strategies for the Indian Steel Industry.” Energy Conversion and Management, vol. 293, 2023, article 117511. https://doi.org/10.1016/j.enconman.2023.117511.

JSW Steel Limited. Integrated Report 2022–23. 2023, https://www.jswsteel.in/jsw-steel-annual-report-2022-23/pdf/JSW-Steel-IR-22-23.pdf.

Kashyap, Yash, and Saarthak Khurana. “Taking Stock of Steel: India’s Domestic Production Outlook and Global Investments in Green Steel Production.” Climate Policy Initiative, 6 Sept. 2023, https://www.climatepolicyinitiative.org/taking-stock-of-steel-indias-domestic-production-outlook-and-global-investments-in-green-steel-production/.

Lall, Sanjaya. Competitiveness, Technology and Skills. Edward Elgar, 2001.

Ministry of Steel, Government of India. Annual Report 2023–24. 2024, https://steel.gov.in/sites/default/files/Annual%20Report%202023-24%20Final_0.pdf.

Ministry of Steel, Government of India. Template on INDC: Mitigation. https://steel.gov.in/sites/default/files/TEMPLATES-%20MITIGATION_0.pdf. Accessed 26 July 2026.

Porter, Michael E. The Competitive Advantage of Nations. Free Press, 1990.

Porter, Michael E., and Claas van der Linde. “Toward a New Conception of the Environment-Competitiveness Relationship.” Journal of Economic Perspectives, vol. 9, no. 4, 1995, pp. 97–118. https://doi.org/10.1257/jep.9.4.97.

Schramade, Willem. “Integrating ESG into Valuation Models and Investment Decisions: The Value-Driver Adjustment Approach.” Journal of Sustainable Finance & Investment, vol. 6, no. 2, 2016, pp. 95–111. https://doi.org/10.1080/20430795.2016.1176425.

Steel Authority of India Limited. Annual Report 2022–23. 2023, https://sail.co.in/sites/default/files/2023-09/SAIL_Annual-Report-2022-2023.pdf.

Steel Authority of India Limited. Corporate Sustainability Report 2022–23. 2023, https://www.sail.co.in/sites/default/files/2024-05/SAIL%20Sustainability%20Report%202022-23.pdf.

Tata Steel Limited. Business Responsibility and Sustainability Report 2022–23. 2023, https://www.tatasteel.com/investors/integrated-report-2022-23/pdfs/TataSteel-IR23_BRSR.pdf.

Tata Steel Limited. Integrated Report and Annual Accounts 2022–23. 2023, https://www.tatasteel.com/media/18370/tata-steel-ir-2022-23.pdf.

Vorontsov, Ihor. “How India Will Decarbonize Its Steel Industry.” GMK Center, 7 July 2025, https://gmk.center/en/posts/how-india-will-decarbonize-its-steel-industry/.

World Steel Association. “Sustainability Indicators.” https://worldsteel.org/wider-sustainability/sustainability-indicators/. Accessed 26 July 2026.

World Steel Association. Sustainability Indicators 2023 Report: Sustainability Performance of the Steel Industry, 2004–2022. Nov. 2023, https://worldsteel.org/wp-content/uploads/Sustainability-indicators-report-2023.pdf.

Digital & Streaming - Era Music Marketing: Strategies, Data, And Discovery in the Platform Economy By Baani Kohli

Introduction

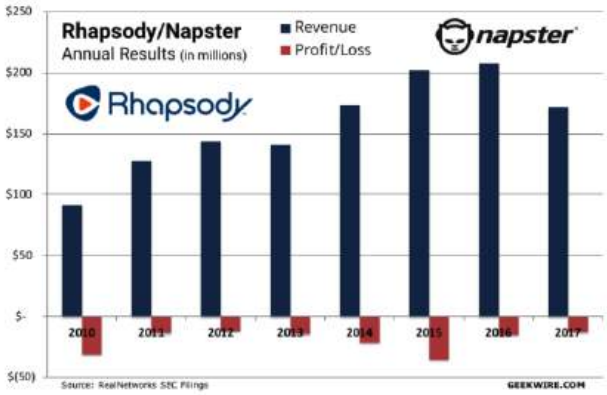
The music industry, like many sectors, has shifted from a physical, ownership model to an access model centered on digital delivery. This transition has reshaped revenue streams, marketing practices, and the relationship between artists and audiences. An industry once defined by record stores, CDs, and cassette tapes from the 1950s to 1980s is now in the 2000s, reaching global listeners through digital platforms. Specifically, streaming services such as Spotify in 2008 and Apple Music in 2015 function as the primary channels of consumption. As streaming becomes the default listening format, platforms compete through brand positioning, privacy policies, and algorithmic personalization. More broadly, the move from physical sales to digital access in the late 1900s has transformed the industry's marketing infrastructure. Data and performance metrics now guide how music is promoted, how audiences are targeted, and how major events are commercialized. In this environment, digital marketing and analytics have become central drivers of growth and profitability across the modern music economy.

The Evolution of Music Consumption: From Records to Streaming

In early human history, music was embedded in cultural practice and live performance, shaped by traditional instruments used for rhythm and storytelling. Over time, technological innovation streamlined music reproduction and distribution. A major shift occurred in the late 1800s when Thomas Edison introduced the phonograph. According to PBS, "He noticed a speech-like noise as he accidentally ran indented tin foil under the telegraph stylus" (Taintor). This approach used a phonograph used a cylinder to capture and replay sound onto a surface. Soon after, inventors challenged the cylinder format. Emile Berliner developed the flat disc: a seven-inch disc that held roughly two minutes of sound with a wax and zinc process so that "grooves in exposed zinc become a negative stamp that allows Berliner to mass produce discs in vulcanized rubber" (Taintor). Edison's cylinder was difficult to duplicate at scale, since songs often had to be recorded repeatedly to create a small number of copies. Berliner's "negative stamp," by contrast, enabled efficient mass production. Companies such as the Victor Talking Machine Company helped popularize the disc format, lowering production costs, simplifying storage, and contributing to the rise of vinyl records in the 20th century.

As recorded music expanded, convenience and portability became increasingly important. Vinyl records, while widely adopted, were large, fragile, and dependent on relatively expensive playback equipment. Cassette tapes later made listening more personal and mobile, supporting everyday use outside the home. The arrival of compact discs accelerated the shift toward digital storage: "Billy Joel's 52nd Street, released in 1982 in Japan, became the first CD released in the world. By 1988, the CD surpassed the LP in sales" (Kelly). Although CDs remained physical media, their digital encoding laid the groundwork for later distribution via the internet.

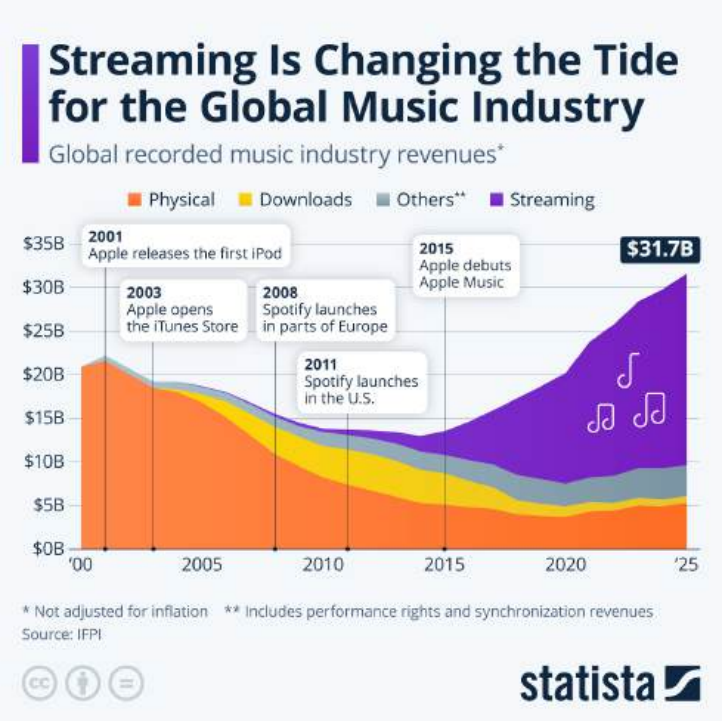
The most dramatic transformation arrived with widespread internet access. The MP3 format popularized lossy compression, reducing file sizes by roughly 90–95% while preserving acceptable listening quality. This made it practical to store and transfer large music libraries, and it fundamentally changed distribution because now music was available on files stored in iPods at the time. Platforms such as Napster (now Rhapsody), a platform that allowed music to be downloaded and shared across users for free, demonstrated how quickly music could circulate online, forcing the industry to adapt.



RealNetworks illustrates the rapid rise in Napster's revenue, a trend that shows an increase in public use of downloads.

At the same time, the internet lowered barriers to music production. "The rise of the internet allowed artists to use affordable (or even free) digital audio workstations (DAWs) and virtual instruments (VSTs)" (Wetware Admin). Music creation was no longer limited to those with extensive resources or professional studio access. "Programs like FL Studio, Ableton Live, and Logic Pro X have made it possible for virtually anyone with a computer to make music" (Wetware Admin). This democratization of tools contributed to an expanding creator base; "more than 40 million active music creators uploading and engaging with content" (XtendedView) illustrates the scale of participation.

Streaming services ultimately made distribution even more accessible. In many cases, independent artists can now release music with minimal barriers; virtually anyone can release music now, especially without the need of a record label, contributing to rapid growth in available content and audience reach. These same developments also accelerated the decline of physical formats: "Sales of CDs in the UK in 2018 amounted to 32 million units, that's some 100 million fewer discs than were selling just 10 years ago in 2008, and a drop of 9.6 million year-on-year, according to a report from BBC" (Sparrow). Ultimately, the internet ushered in an era in which the industry could monetize music through multiple channels beyond ownership. As Forbes summarizes, "Data shows that listeners get more music, and the industry makes more money from it, through 'access' models than through 'ownership'" (Rosenblatt). This encompasses the success the music industry was able to gain from such a shift from ownership to access.



Statista's chart indicates the impactful rise of streaming starting in 2015, demonstrating how streaming revenue surpasses revenue generated from physical music sales and digital downloads.

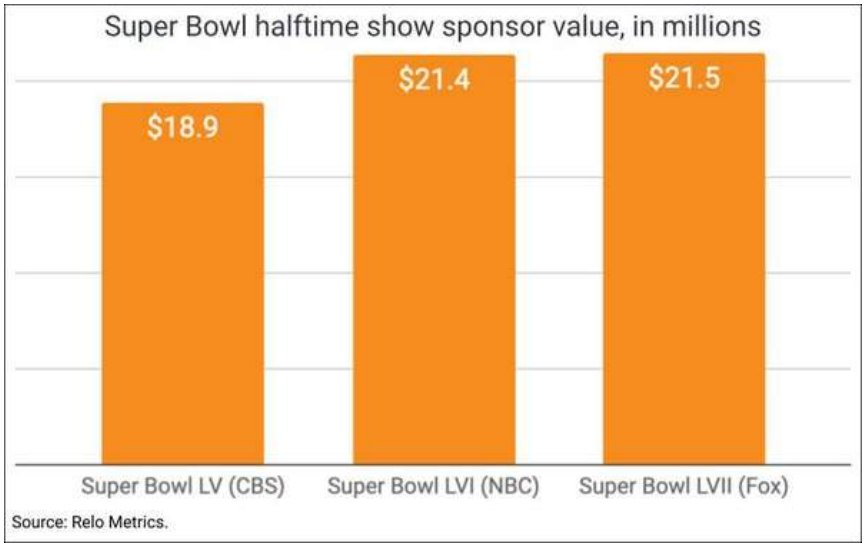
How Streaming Services Work: Access, Reach, and Revenue

Streaming services such as Spotify, Apple Music, Amazon Music Unlimited, and TIDAL provide on demand access to streaming, which allows users to select specific songs, albums, and playlists instantly. In contrast to ownership based listening, streaming offers "access to millions of songs without purchasing physical records or downloading files" (AIMM Blog staff). This model aligns with contemporary expectations of convenience by removing barriers associated with buying, storing, and transporting physical media.

The accessibility of music with an internet connection has inevitably led to streaming expansions and how music is used in everyday settings. The same access model broadens distribution for creators: streaming "allows major labels and independent artists …to reach broader audiences" (AIMM Blog staff). In practice, this means that music can move from creation tools (such as DAWs) to release platforms through distributors that deliver content to Spotify, Apple Music, Amazon Music, and others.

From the platform perspective, streaming creates multiple revenue opportunities. Services can monetize each song, podcast episode, or other piece of content through subscriptions, advertising, or both. They also benefit from a constant flow of new material; for example, "106,000 new tracks are uploaded to streaming music services every single day" (Cross). In addition, major platforms have expanded beyond music into podcasts, audiobooks, short length videos, and exclusive concert content, widening their audiences and diversifying monetization.

Many platforms also increase visibility through sponsorships and partnerships. A prominent example is Apple Music's sponsorship of the Super Bowl Halftime Show, which functions both as brand positioning and advertising. However, sustaining growth requires more than successful partnerships. To retain users and attract new listeners, streaming services increasingly rely on performance metrics and recommendation algorithms.

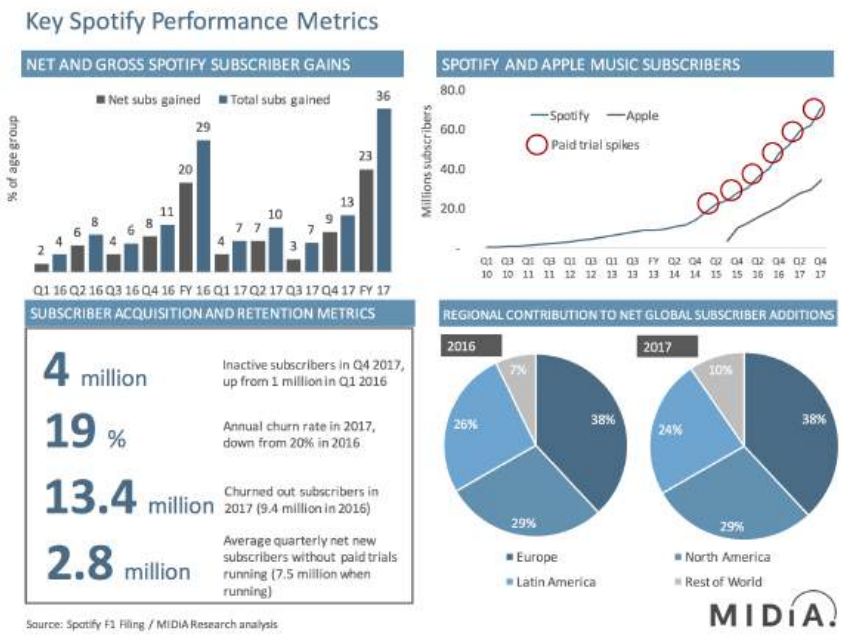


Relo Metrics shows the sponsor value Apple Music generates from sponsoring the Super Bowl Halftime show. With an average of 20 million from these three channels, this is one of the many profitable ways streaming services are able to increase visibility.

Metrics and Music Analytics: Measuring Success in the Streaming Era

In the streaming era, commercial success is increasingly shaped by how performance is measured and translated into strategy. As competition among platforms intensifies, data has become central to marketing, audience targeting, and monetization. "Universal Music Group has leaned heavily into data and analytics to guide everything from artist support to campaign design" (Bean). This reflects "A broader industry truth: intuition alone isn't enough anymore" (Neu). Instead, platforms and labels use analytics to align music with audience behavior, evaluate performance, and plan future campaigns.

Metrics help identify which artists, genres, and sounds are gaining traction, as well as how listeners engage over time. Common indicators include total streams, repeat listens, saves, playlist adds and removals, and other measures of sustained interest. Demographic and geographic insights, such as age ranges and location, further support targeted promotion by revealing who is listening and where momentum is building.



Midiaresearch shows an example of a metrics streaming service, like Spotify, that analyzes targeted age groups, the location of subscribers, and retention rates of subscribers.

When properly utilized, analytics can support prolonged planning rather than reactions informed by recency biases. As Repertoire Analytics notes, "Music analytics is most powerful when used to inform long-term planning, especially in artist development" (Neu). For labels and managers, metric trends can influence decisions about touring, marketing spend, and release timing. For example, releasing music when an artist's audience growth is accelerating can improve initial performance and increase the likelihood of sustained engagement. This planning can only occur based on the success in streams that is measured through metrics.

At the same time, interpreting data at scale can be challenging. Multiple platforms maintain separate analytics environments, and fragmented reporting can make it difficult to develop a unified view of audience behavior. As a result, "music professionals are increasingly turning to centralized tools that compile, clean, and visualize performance data in one place" (Neu).

Overall, platform strategy can no longer rely on intuition alone and must be informed by measurable listener behavior. As the industry becomes more digitally integrated, analytics will increasingly shape how streaming services design user experiences, allocate marketing resources, and compete for attention.

Algorithmic Discovery: How Metrics Power Music Recommendations

Recommendation systems are built on the metrics that platforms collect. Once services analyze performance data to identify patterns, such as what users replay, save, or skip, automated

systems can increase the visibility of content that is likely to retain attention. Personalization is driven by listening behavior. For instance, frequent skipping and playlist removals may signal that a user is dissatisfied with a given style, while repeated listening and saves typically signal a stronger preference.

Patterns are processed computationally, and "Machine learning plays a crucial role in analyzing and predicting user preferences, enhancing the accuracy" (Bergmann). Greater predictive accuracy supports tailored recommendations, automated playlists, and more responsive discovery experiences. At the same time, personalization is not entirely automated. In fact, "Human curators complement the algorithm by adding a nuanced touch to playlist curation" (Randwick). This ensures a combination of technology and human insight when music resources are especially curated for users. Together, human and automated inputs shape a listening environment that amplifies positive signals (such as saves and repeat plays) and reduces negative signals (such as skips), producing a more individualized experience.

"Algorithmic discovery makes music listening more personal, helping listeners create a relationship and reliance with streaming platforms" (Hesmondhalgh). This dynamic can encourage long term loyalty and aligns with a UK government study finding that algorithmic recommendations drive the majority of music consumption on streaming platforms. Recommendation systems can also benefit artists who understand how platforms surface content. For instance, "Artists can improve their chances of being featured on Apple Music by understanding the algorithm's workings and optimizing their music accordingly" (Randwick), suggesting that release strategy and platform optimization can influence visibility. Specifically, at a point in time when rock music is successful algorithmically opens up an advantageous opportunity for rock artists to resell music during this time period.

Streaming Strategies and Impact: Spotify vs. Apple Music

Spotify is often cited as a leading example of personalization in streaming. The platform uses listener metrics to generate customized features; for instance, its data supports tools that "create tailored experiences such as Discover Weekly and Daily Mix playlists" (Inspirepreneur Team). Discover Weekly draws on signals such as saved tracks, likes, replays, and playlist additions, then clusters those preferences by artist and genre to recommend related music. In doing so, it expands listener exposure to new artists and songs aligned with established interests. Daily Mix applies similar logic by assembling rotating playlists based on frequent listening patterns, which encourages repeated engagement through regularly refreshed recommendations.

Spotify's personalization capabilities expanded further with its acquisition of "The Echo Nest," which strengthened the platform's ability to interpret both musical characteristics and user behavior. The system analyzed audio features such as key, mood, and tempo and combined them with behavioral and cultural metadata to generate more detailed representations of songs and listening patterns. This approach moves beyond surface-level engagement metrics by incorporating attributes of the music itself.

Spotify's recommendation engine is often described as a hybrid model that integrates

natural language processing with collaborative filtering and content-based filtering. “This hybrid approach ensured recommendations were both relevant and surprising” (The Product Brief). Supporting product features such as “Discover Weekly” and “Daily Mix.” As these systems mature, discovery can become increasingly more reliant on the platform than on the user. Spotify also provides contextual cues, for example, “Because you listen to…” that explain why certain recommendations appear, which can strengthen transparency and user trust. Overall, Spotify illustrates how analytics and machine learning can be deployed as competitive advantages in the contemporary streaming market.

Platform success in streaming depends on both technology and business strategy. A useful comparison is Spotify and Apple Music, two dominant services that compete through distinct models. Spotify emphasizes algorithmic discovery and a freemium structure, while Apple Music positions itself around paid access, brand identity, and a more editorial approach to curation.

Spotify’s freemium model allows users to listen at no cost in exchange for advertising, while offering a paid tier for ad-free listening and additional features. This structure lowers the barrier to entry, expands the user base, and creates opportunities for conversion as some listeners upgrade after sustained use or through limited trials.

Financially, the model supports multiple revenue streams. Advertising monetizes the free tier, and subscriptions generate recurring income from premium users. As one source explains, “This way Spotify can expand the growth of its ads business, enabling brands to show up where consumers spend a significant portion of their online lives, and at a lower cost of entry” (Accenture).

Apple Music competes at the top of the market with only paid subscriptions in its model and removes advertising to achieve a premium brand position. The service offers broad catalog access and expands into adjacent formats such as podcasts and audiobooks. Unlike freemium platforms, Apple Music’s revenue is driven primarily by paid subscriptions. As one source notes, “A monthly subscription to Apple Music is priced at $9.99, offering a free 3-month trial for potential subscribers. Apple is also extending a family plan for a monthly fee of $14.99” (Mitchell). The family plan, in particular, signals an attempt to increase perceived value by supporting multiple users under one payment.

Apple Music also incorporates personalization. It has been described as “A dynamic multimedia streaming service, focused on empowering artists and music lovers to create curated experiences” (Mitchell). Relative to Spotify’s heavily automated discovery, Apple Music is often associated with a more editorial emphasis in playlisting while still using listening data to inform recommendations. Together, these approaches illustrate how competing platforms can succeed through different combinations of pricing strategy, brand messaging, curation, and algorithmic design.

These platform strategies also shape outcomes for artists, audiences, and the broader economy. Historically, when “an artist would drop a new single, which would climb the chart but often fade into obscurity” (AIMM), competition was often concentrated around short-lived chart

performance. Streaming, however, can extend the lifecycle of successful songs from weeks to months or longer, especially when tracks re-emerge through social platforms such as TikTok, Instagram, and YouTube. Streaming services also respond to these attention cycles through curated placements; for example, Spotify playlists such as “TikTok Hits” are frequently updated to reflect trending audio.

In addition, the convenient availability of streaming encourages deeper catalog consumption. When albums remain continuously accessible, listeners can move across multiple tracks rather than focusing only on singles. In this environment, binge listening can increase total streams and visibility within ranking systems. Successful releases that place multiple tracks into prominent chart positions illustrate how streaming can amplify album-level engagement.

The streaming economy also distributes revenue across a wider set of stakeholders than traditional retail models. Rights holders, including record labels, music publishers, and distributors, receive a significant share; for example, “Around 70% of a streaming platform's revenue goes to rights holders such as record labels, music publishers, and distributors”(AIMM). Platforms also generate income through advertising, and “ad-supported revenue…continues to tick up, increasing by 7% Y/Y” (Spotify Editorial Team). Beyond direct listening, streaming services increasingly participate in digital commerce and partnerships that produce value both for the platform and for external businesses. Taken together, sustained engagement, access, and data promotion support growth for artists, audiences, and the wider music economy.

Conclusion: Convenience, Algorithms, and the New Music Economy

The contemporary music industry differs fundamentally from earlier eras defined by physical media and ownership. What started as a phonograph evolved through vinyl, cassettes, and CDs and then culminated in on-demand large-scale streaming. Portability and instant access have made music a routine part of daily life, while digital distribution has expanded monetization beyond album sales to include subscriptions, advertising, and platform promotion. Within this landscape, services such as Spotify and Apple Music illustrate how pricing models and personalization strategies shape consumer choice, engagement, and perceived value. The commercial lifespan of music has been gratefully extended, creating sustained engagement.

More broadly, algorithms and performance metrics function as operational tools: they help platforms identify what is resonating with listeners, refine recommendation systems, and improve retention. Algorithms embody the idea of depersonalization for users. The traditional intuition that once guided industry decisions is now guided by data-based facts. The result is an ecosystem characterized by rapid content growth, global creator participation, and an industry capable of generating billions of dollars annually. The commercial infrastructure of music has also been transformed through democratization. Through DAWs, subscription, advertising, promotions, and partnerships like the Apple Music-sponsored Super Bowl, streaming has restructured connections and creative infrastructure, making it a prominent growth factor for the modern music industry.

Works Cited

Academy of Interactive Music & Media. "How AI Will Empower Musicians to Become One-Person Production Teams." AIMM, aimm.edu/blog/how-ai-will-empower-musicians-to-become-one-person-production-teams

Academy of Interactive Music & Media. "How Streaming Changed the Music Industry." AIMM Blog, aimm.edu/blog/how-streaming-changed-the-music-industry/.

Accenture. "Spotify Advertising Revolution." Accenture, www.accenture.com/us-en/case-studies/communications-media/spotify-advertising-revolution.

Affinco. "Apple Music Statistics." Affinco, affinco.com/apple-music-statistics/.

Desrocher, Nick. "Apple Music's Growth Marketing Case Study." LinkedIn, www.linkedin.com/pulse/apple-musics-growth-marketing-case-study-nick-desrocher-mba-rlync/.

Dope Music Promotions. "How the Apple Music Algorithm Works." Dope Music Promotions, dopemusicpromotions.com/how-apple-music-algorithm-works/.

Eliya. "Data Privacy." Eliya Blog, www.eliya.io/blog/marketing-measurement/data-privacy.

Financial Post. "Apple Inc. Puts Human Curation at Centre of Apple Music Overhaul (WWDC 2016)." Financial Post, financialpost.com/fp-tech-desk/personal-tech/wwdc2016-apple-inc-puts-human-curation-at-centre-of-apple-music-overhaul.

GoBrandVerge. "Navigating Data Privacy Regulations in Digital Marketing: A Marketer's Guide to Compliance and Protection." GoBrandVerge, www.gobrandverge.com/navigating-data-privacy-regulations-in-digital-marketing-a-marketers-guide-to-compliance-and-protection/.

Inspirepreneur Magazine. "Personalisation Power: How Spotify Transformed Music Discovery." Inspirepreneur Magazine, inspirepreneurmagazine.com/case-studies/personalisation-power-how-spotify-transformed-music-discovery/.

McKinsey & Company. "The Consumer Data Opportunity and the Privacy Imperative." McKinsey & Company, www.mckinsey.com/capabilities/risk-and-resilience/our-insights/the-consumer-data-opportunity-and-the-privacy-imperative.

Mitchell, Julian. "WWDC 2015: The Hidden Strategy That Gives Apple Music the Edge Over Its Competitors." Forbes, 9 June 2015, www.forbes.com/sites/julianmitchell/2015/06/09/wwdc-2015-the-hidden-strategy-that-gives-apple-music-the-edge-over-its-competitors/.

Product Brief. "Spotify's Algorithmic Alchemy: How AI-Powered Playlists Built a 100 Billion Empire." Medium, medium.com/@productbrief/spotifys-algorithmic-alchemy-how-ai-powered-playlists-built-a-100-billion-empire-eac5a0d0bceb.

Recording Connection. "Top DAWs." Recording Connection, recordingconnection.com/top-daws/.
Reprtoir. "Analytics Music Industry / Music Streaming Trends." Reprtoir Blog, www.reprtoir.com/blog/analytics-music-industry-https:/www.dataart.com/blog/music-streaming-trends.
Rosenblatt, Bill. "New Data Shows That More People Want to Access Music Than Own It." Forbes, 17 Apr. 2017, www.forbes.com/sites/billrosenblatt/2017/04/17/new-data-shows-that-more-people-want-to-access-music-than-own-it/.
ScienceDirect. "Article S0378720625001910." ScienceDirect, www.sciencedirect.com/science/article/abs/pii/S0378720625001910.
Soundcharts. "How to Get Recommended by Streaming Algorithms." Soundcharts, soundcharts.com/en/blog/how-to-get-recommended-by-streaming-algorithms.
Sparrow, Mark. "Sales of Physical Music Media Slump as Consumers Move to Streaming Services." Forbes, 3 Jan. 2019, www.forbes.com/sites/marksparrow/2019/01/03/sales-of-physical-music-media-slump-as-consumers-move-to-streaming-services/.
Spotify Advertising. "Spotify Fourth Quarter Earnings 2024." Spotify Advertising, 2024, ads.spotify.com/en-US/news-and-insights/spotify-fourth-quarter-earnings-2024/.
Statista. "Spotify Ad Revenue." Statista, www.statista.com/statistics/1305705/spotify-ad-revenue/.
Syncopated Times. "A History of the Music Business and Where It’s Headed Now." Syncopated Times, syncopatedtimes.com/a-history-of-the-music-business-and-where-its-headed-now/.
UK Government. "The Impact of Algorithmically Driven Recommendation Systems on Music Consumption and Production: A Literature Review." GOV.UK, www.gov.uk/government/publications/research-into-the-impact-of-streaming-services-algorithms-on-music-consumption/the-impact-of-algorithmically-driven-recommendation-systems-on-music-consumption-and-production-a-literature-review.

The Viability of Diasporic Remittances as a Solution to Lebanon's Economic Crisis
By Tamara Herzberg

Introduction

The Lebanese financial crisis is one marked by a consistent lack of responsible governmental response. A prolonged civil war from 1975 to 1990 accumulated significant debt that has continued to haunt the Lebanese government well into the 21st century. Relying on precarious foreign aid from the Gulf States, as well as earnings from tourism and the financial sector, the government managed to stay afloat until 2019, when a lack of policy changes combined with domestic unrest held back millions in foreign funds, forcing Lebanese banks to shut down (Blair). Since 2019, the Lebanese currency has lost more than 98% of its value, and in 2021, Lebanon's income poverty rate rose to a staggering 74% (US Department of State). In the face of this crippling economic instability, Lebanon's diaspora has acted as a financial lifeline for much of the population. Since Lebanese mass migration began in the late 19th century, the Lebanese diaspora has maintained uniquely close ties with their home country through philanthropy and remittances (Fersan). Hourani, director of the Lebanese Emigration Research Center at Notre Dame University-Louaize, credits the diaspora's contributive nature to a nationalistic desire to reincarnate some idyllic past, philanthropy fueled by a burning vision of the homeland's former glory.

However, it may also be that the reasons for these close international ties are more personal. In his narrative essay about a visit back to a home ravaged by natural disaster, Japanese author Haruki Murakami describes the reasoning behind his return: "I wanted to see with my own eyes what transformations had taken place – what this awful violence had stolen from the town, and what it had left behind. There had to be at least some connection, I felt, with who I am now" (Murakami). Like Murakami, perhaps the diaspora's desire to reconnect with its home stems from an individual search for identity, one made even more urgent amid the turbulence of the modern Arab world. To members of the diaspora, it may seem that in the homeland, there remain traces of a self untouched by the transformative powers of immigration, war, natural disaster, and poverty. While Hourani demands increased governmental engagement with the diaspora to maintain their critical support, this sense of personal connection with the home, unbothered by political reality, stands quite powerfully on its own. As Ann Killion describes in a personal essay about the relationship between a long-gone ballpark and her deceased father, what a place represents, rather than its physicality, can be enough to keep us returning again and again (Killion).

The diaspora's strong connection to their home, however, does more than just satisfy their craving for selfhood: it has tangible economic implications for the Lebanese who remain in Lebanon, most notably through remittance flows. Remittances, defined by the United Nations as "money earned or acquired by migrants that are transmitted back to their country of origin", comprised 54% of the Lebanese gross domestic product in 2024 (Emilsson and Nordin; Omar), demonstrating their importance as a source of income for many Lebanese families. In the

absence of tangible governmental reform, the overwhelming reliance of the Lebanese populace on diasporic remittances begs the question: How viable are these remittances as a solution to the Lebanese economic crisis?

This paper argues that remittances are not a viable economic solution. The first paragraph acknowledges the short-term value of remittances, but goes on to argue that these same benefits erode workers' incentives and permit government complacency, thereby preventing meaningful systemic change. The second paragraph pivots to explore the long-term implications of remittances, arguing that the permanent nature of Lebanese migration means that remittances are merely temporary financial support and cease as entire families settle abroad. The paper concludes by reaffirming that remittances are not a sustainable solution to the Lebanese financial crisis and calls for further efforts from the Lebanese diaspora to find a lasting solution.

Short-term implications of remittance flows

In the short-term, a reliance on remittances has significant value in alleviating immediate financial burdens as well as contributing to community development. In her paper examining the factors affecting the financial and cultural engagement of the Lebanese diaspora with Lebanon, Omar argues that because of their valuable monetary contributions through remittances, the diaspora has the potential to significantly remedy the current economic crisis. This is supported by data from Amery, who analyzed the impact of remittances on the household income and community economies of two villages in Southeast Lebanon, one of which received significant annual remittances and the other which did not. A questionnaire was distributed to the family heads of each village. Questions explored the number of emigrant family members abroad, the monetary value of annual remittances, and the areas into which these remittances were funneled. For households in the village receiving remittances, it was found that remittances accounted for almost 50% of their total annual income (Amery). This overwhelming financial dependence on remittances indicates their value in undertaking daily transactions. Specifically, Amery found that remittances contributed to both tangible and intangible financial transactions. Results showed that 52% of remittance-receiving households used the money to improve their land, while 56% used it to buy fertilizer (Amery). Thus, remittances are reinvested in capital that enhances recipients' trades, bolstering their agriculture and thereby stabilizing their flow of income. In addition, remittances can contribute to more abstract transactions: 71% of households use remittances to pay debts, and 58% to educate children (Amery). Thus, remittances offer more than just material aid: they bolster local social systems through debt management and the intellectual development of the Lebanese youth. However, it is very possible that since the publication of Amery's study in 1992, the role of remittances in Lebanon has changed. This paper assumes that the presence of a Lebanese diaspora with a desire to contribute to the homeland has remained alongside the country's conflict and economic instability, and argues that the general theme of Amery's study - that remittances have short-term financial value - still holds. That said, the age of this study does limit its relevance to Lebanon's current economic climate.

Furthermore, building on the individual financial benefits demonstrated by Amery, Hourani explains how remittances contribute to the development of recipient communities through their reinvestment into institutions. In her discussion paper on the various political and economic interactions between Lebanon and its diaspora, Hourani argues that diasporic contributions could "play a vital role in the economic recovery of the country"; she attributes the construction of numerous Lebanese hospitals to migrant donations, including the St. Louis Hospital in Jounieh and the Kfar Falous Hospital, as well as educational insitutions like the Yafeth Library at the American University of Beirut (Hourani). Thus, remittances hold not only the power to bolster individual transactions but also to accelerate communal development, benefiting third-party citizens not necessarily recipients of remittances themselves.

However, while the short-term value of remittances is significant, this paper counters this argument by proposing that these same benefits can erode workers' incentives and contribute to government complacency, preventing long-term economic sovereignty. Firstly, remittances necessitate labor only by the breadwinner abroad and not necessarily by recipients, which weakens incentives to work themselves. A study by Emilsson and Nordin explored the relationship between receiving remittances and supplying labor to the workforce in Jordan. Data on household characteristics, relationships with remittances, and income earned from work, all collected from Jordan's Department of Statistics, were entered into multiple regression models. While this study analyzed Jordanian data, Jordan's geographical proximity and cultural similarity to Lebanon suggest that these findings are applicable. Results revealed a negative correlation between receiving remittances and labor supply, meaning that the more remittances a household received, the less labor it contributed to the workforce. Thus, the short-term ease that remittances afford can erode incentives to work, which reduces the labor supply. This is significant because an inefficient local workforce complicates any efforts at expanding and profiting from Lebanese industry, preventing Lebanon from ever achieving long-term economic self-sufficiency.

Moreover, the simplicity and short-term value of the remittance process could contribute to government complacency in the face of societal issues. Ralph Chami is an economist with over 20 years of experience at the International Monetary Fund (IMF), who currently serves as its Assistant Director in the Institute for Capacity Development. He compiled data from the IMF's Balance of Payment Statistics and the World Bank's World Governance Indicators to create a scale measuring government effectiveness, defined as the responsiveness of a government to the needs of its citizens, to economic dependence on remittances. Scores ranged from -2.5 (very ineffective) to 2.5 (very effective). Results showed that countries in which remittances made up more than 10% of their GDP averaged a score of nearly -0.6, compared to countries in which remittances comprised between 1 and 2% of their GDP, which scored slightly less than -0.2 (Chami et al.). Thus, a reliance on remittances is correlated with reduced government effectiveness. Remittances offer a quick fix to urgent economic burdens, but their convenience makes it easy for politicians to avoid addressing the root issues of Lebanon's economic instability. A country that relies primarily on periodic foreign allowances cannot be

expected to develop a sovereign, self-sustaining, and responsive government, which is necessary for any lasting change to occur in Lebanese society.

Long-term limitations

Furthermore, the practical implications of remittances in a country where temporary migration often becomes permanent exile make it likely that remittances are eventually cut off from the home country. Gevorkyan, Assistant Professor of Economics at the Peter J. Tobin College of Business in St. John's University, published a literature review exploring the value of diasporas to developing economies. He posits that as families eventually follow migrant breadwinners abroad, the incentive to send remittances to the home country is weakened and remittances eventually cease. Thus, family immigration threatens the sustainability of remittance flows as it elicits their inevitable cutoff date. This idea becomes particularly significant in the case of Lebanese migration: De Bel-Air finds in her migration profile for the European University Institute that, from 2007 to 2015, Lebanon's net migration pattern was consistently negative, meaning that throughout that period more Lebanese migrants departed the country than arrived. This demonstrates that Gevorkyan's idea of permanent leave is extremely relevant to Lebanon: temporary leave often turns into long-term exile, making it likely that families follow each other abroad. Therefore, although remittances may offer immediate financial relief, their inevitable termination means that they cannot be counted on by communities as reliable, long-term economic support systems.

This is further supported anecdotally by Murakami's "A Walk to Kobe", where he describes his sense of alienation from his former home: "the axis of my memories is faintly, but audibly, creaking within me" (Murakami). And yet, this profound feeling of loss for his home does not keep Murakami stubbornly anchored there. Despite his deep nostalgia, he acknowledges that more important than meaningless pondering is the ability to move on: "Ultimately I'm the kind of person who can only make progress through moving my legs, moving my body, through a step-by-step, halting, physical process" (Murakami). Murakami's story may be geographically distant from Lebanon, but the narrative of immigration and fragile connection is universal. While the Lebanese diaspora may maintain a special bond with the home they once knew, Murakami's anecdote demonstrates that native affinity does not overcome the urge to migrate in search of better opportunities, supplementing the net negative migration patterns demonstrated by De Bel-Air with lucid psychological context. Thus, Gevorkyan's proposed relationship between family migration and the end of remittances, in combination with the Lebanon-specific migration data of De Bel-Air and the anecdotal evidence provided by Murakami, indicates that the permanence of Lebanese migration limits the time frame in which remittance flows operate, thereby rendering them unsustainable.

Conclusion

It should be noted that this paper does not call for the termination of remittance networks. Rather, it concludes that remittances cannot be relied on by the Lebanese populace and

government as a viable long-term solution to the Lebanese economic crisis. As Hourani and Omar argue, remittances are critical to keeping households functional amid the daily struggles of life in an impoverished nation. However, their arguments fail to consider the question of longevity, which is critical in the search for a sustainable solution to an economic problem as perennial as Lebanon's. The short-term benefits of remittances entail long-term liabilities by complicating efforts to achieve Lebanese economic sovereignty. Moreover, the permanent nature of Lebanese migration suggests that remittance networks inevitably cease to operate. Both points indicate that the value of remittances is unsustainable. However, remittances are but one of the many interactions that make up the complex relationship between Lebanon and its diaspora. Hourani mentions the diaspora's potential to enact political change through "long-distance nationalism" and lobbying international organisations. More research is needed on the intricacies of diasporic political engagement, for it may only be through codified policy reforms that lasting systemic change is achieved.

Works Cited

Amery, H. A.-M. The Effects of Migration and Remittances on Two Lebanese Villages. McMaster University, 1992. http://hdl.handle.net/11375/8583.

Blair, Edmund. "Explainer: Lebanon's Financial Crisis and How It Happened." Reuters, 23 Jan. 2022, https://www.reuters.com/markets/rates-bonds/lebanons-financial-crisis-how-it-happened-2022-01-23/.

Chami, R., E. Ernst, C. Fullenkamp, and A. Oeking. "Is There a Remittance Trap?" International Monetary Fund, Sept. 2018, https://www.imf.org/en/publications/fandd/issues/2018/09/is-there-a-remittance-trap-chami.

De Bel-Air, F. Migration Profile: Lebanon. European University Institute, 2017, https://cadmus.eui.eu/entities/publication/5de2aa29-6f72-5a55-9c2e-88490cc52147.

Emilsson, A., and M. Nordin. The Impact of Remittances on Labor Supply: The Case of Jordan. Lund University, 2011.

Fersan, Eliane. "Syro-Lebanese Migration (1880–Present): 'Push' and 'Pull' Factors." Middle East Institute, Apr. 2010, https://mei.edu/publication/syro-lebanese-migration-1880-present-push-and-pull-factors/.

Gevorkyan, A. V. "Diaspora and Economic Development: A Systemic View." The European Journal of Development Research, vol. 34, no. 3, 2022, https://doi.org/10.1057/s41287-021-00432-x.

Hourani, G. "Lebanese Diaspora and Homeland Relations." Migration and Refugee Movements in the Middle East and North Africa, 2007, pp. 1–21.

Killion, Ann. "Dad, Ballpark Will Live On in Shared Memories." SFGate, 2013, https://www.sfgate.com/49ers/article/dad-ballpark-will-live-on-in-shared-memories-5087001.php.

Murakami, Haruki. "A Walk to Kobe." Granta, 2013, https://granta.com/a-walk-to-kobe/.

Omar, M. H. "Engaging the Lebanese Diaspora: Socioeconomic Determinants, Barriers, and Opportunities for Sustainable Development." Edelweiss Applied Science and Technology, vol. 9, no. 2, 2025, pp. 1784–1793, https://doi.org/10.55214/25768484.v9i2.4915.

US Department of State. 2024 Investment Climate Statements: Lebanon. 2024, https://www.state.gov/reports/2024-investment-climate-statements/lebanon/.

Instability and Innovation: Financial Uncertainty and Listed Fintech Market Performance
By Qixuan Liang

Abstract

This study examines how moderate and sudden increases in financial uncertainty affect listed fintech market performance and how this relationship compares with other innovation-linked sectors. It also investigates whether different forms of instability have divergent impacts on innovation. It adopts a comparative quantitative empirical approach by analysing exchange-traded funds (ETF) and indices in terms of monthly returns, correlations, scatterplots with simple ordinary least squares (OLS) linear regression, VIX-level regime tests, and R^2 values to evaluate the relationship between instability and innovation-related market performance. This paper uses the Global X FinTech ETF (FINX) as the primary proxy for innovation-linked financial technology exposure, with the CBOE Volatility Index (VIX) as the primary proxy for financial instability. Furthermore, this study compares FINX with other innovation-adjacent benchmarks and contrasts VIX with the Global Economic Policy Uncertainty Index (GEPU) as alternative measures of instability to examine whether the observed relationship is fintech-specific and whether innovation-linked market performance responds differently to market-based and policy-based uncertainty. Overall, the results suggest that sharp increases in market volatility are associated with weaker performance among listed fintech firms. Higher VIX levels are also associated with weaker FINX returns, although this relationship is substantially weaker than the relationship with sudden VIX changes. In addition, the uncertainty effect reflects a broader risk-off pattern across innovation-linked growth sectors, while listed fintech performance is more closely associated with market-based volatility than with policy-related uncertainty.

Keywords: financial uncertainty, innovation, economics, fintech, market volatility.

1. Introduction

The U.S. stock market made a sharp recovery after the COVID-19 financial market trough. The S&P 500 rose from 2237.40 on 23 March 2020 to 7609.78 by 2 June 2026, an increase of approximately 240% in only a few years ("Table Data – S&P 500"). However, strong aggregate market performance does not necessarily imply that innovation-linked growth sectors are equally robust against financial uncertainty. Growth sectors such as fintech and innovation-linked firms may be vulnerable when financial uncertainty rises because growth-oriented firms are particularly sensitive to changes in investor risk appetite and financing conditions (Baker et al.; Paule-Vianez et al.). Based on this background, this paper investigates whether sudden increases in financial uncertainty are associated with weaker listed fintech market performance and how this relationship compares with other innovation-linked sectors. \

This paper uses the Global X FinTech ETF (FINX) as the primary proxy for listed fintech performance and changes in the CBOE Volatility Index (VIX) as the main proxy for financial uncertainty. The study also adds benchmark innovation sectors, prevailing VIX levels, and

another measure of uncertainty, the Global Economic Policy Uncertainty Index (GEPU), to examine whether listed fintech market performance is especially sensitive to market volatility shocks by comparing FINX with other innovation-linked benchmarks and whether fintech market performance responds differently to market-based and policy-based uncertainty.

The results indicate that FINX returns are negatively associated with VIX changes. However, similar negative relationships among the benchmark assets suggest that this result is not purely fintech-specific but is a broader risk-off pattern across innovation-linked sectors.

2. Literature Review

An earlier study by Baker et al. suggests that financial uncertainty can affect market performance through changes in investor risk appetite, required returns, funding conditions, and willingness to hold riskier growth-oriented assets. When uncertainty rises, investors may become more willing to shift toward more liquid or safer assets rather than hold riskier ones such as stocks. This decline in investors' willingness to bear risk may raise the discount rate applied to expected future cash flows and reduce current equity values. Investors may sell their stocks even before firms experience a decline in profitability or productivity. These valuation effects are strongly associated with innovation-linked companies, where a large proportion of their market value depends on expected future profits. Financial uncertainty may also tighten funding conditions if banks become less willing to provide capital or require higher compensation for risk. Thus, innovation-linked firms may be especially affected by uncertainty shocks because they often require large amounts of external funds to support technological development, commercialisation, and expansion. More broadly, financial uncertainty may weaken economic activity and future cash flow, reducing incentives for investing and hiring.

To distinguish fintech performance under different measures of uncertainty, this paper compares the VIX and the GEPU to FINX. Both indices capture uncertainty, but they represent different sources. The VIX provides a market-based indicator of financial uncertainty because it represents the market's expectation of S&P 500 volatility over approximately the next 30 days, derived from real-time prices of S&P 500 Index options. A high VIX value indicates that investors expect larger fluctuations in the S&P 500 over the next 30 days (Cboe Global Markets). In contrast, Baker et al. measure Economic Policy Uncertainty (EPU) indices from the relative frequency of newspaper articles containing terms related to the economy, uncertainty, and policy ("Monthly EPU Indices for 22 Countries"). Davis constructs the GEPU Index as a GDP-weighted average of these national EPU indices. A high GEPU value indicates that the masses are more sensitive to public discussion of uncertainty, such as future policy decisions or their economic consequences. GEPU may affect real-economy decisions first, such as firms delaying investment and households reducing spending due to pessimism about the future economy, and may subsequently influence financing conditions and economic growth over time. The VIX is a market-based indicator, whereas GEPU is a policy-based indicator. Although the two measures may rise together during major shocks, they are related rather than

interchangeable. This distinction supports the use of VIX change as the primary market-based uncertainty variable and GEPU change as a secondary policy-related comparison.

It is hard to directly measure the exact value of listed fintech market performance. A study conducted by Chen et al. analysed the fintech sector by using FINX returns, supporting that exchange-traded fund (ETF) data can be used as a proxy to investigate the market behaviour of listed fintech firms. The FINX returns measure changes in the market value of a portfolio of listed fintech-related firms and reduce the dependence of the analysis on the characteristics of individual firms. The continuously traded price also generates a consistent return series, which has a strong association with market-wide uncertainty. Chen et al. further concluded that FINX is connected to traditional financial sectors through return and volatility spillovers, suggesting that it may act as a net receiver of shocks. However, changes in FINX returns are associated with interest rates, discount rates, investor sentiment, firm-specific news, and other factors. Therefore, a decline in FINX returns should be interpreted as weaker listed fintech market performance, rather than direct evidence that fintech innovation or technological development has declined.

However, evidence based on FINX alone would only suggest that listed fintechs are associated with volatility shocks, but it is ineffective to test whether sensitivity to volatility shocks is specific to fintech or reflects a wider pattern among innovation-linked assets. In this study, Global X Robotics & Artificial Intelligence ETF (BOTZ) is used as a broader, non-financial innovation benchmark because it provides exposure to robotics and artificial intelligence companies across multiple sectors ("Robotics & Artificial Intelligence ETF"). Another fintech-adjacent innovation benchmark is Amplify Blockchain Technology ETF (BLOK). This ETF invests in companies associated with blockchain technology and crypto-related benchmarks ("Amplify ETFs - BLOK"). The decentralised finance (DeFi) literature explains the financial relevance of blockchain by showing how distributed ledger technology and smart contracts can support financial services such as lending, investment, payments, and cryptoasset exchange. However, BLOK is not treated as a direct measure of fintech innovation activity or DeFi innovation activity (Auer et al.).

In sum, the existing literature provides a foundation for understanding financial uncertainty through market-based and policy-based indicators, using fintech ETFs as a proxy for the fintech sector and clarifying the financial relevance with other ETF benchmarks. However, it does not fully examine whether the listed fintech firms are more sensitive to volatility shocks than other innovation-linked sectors. This paper contributes by comparing fintech with AI, robotics, blockchain, crypto, and digital payment benchmarks. The comparison assesses whether listed fintech appears unusually sensitive to market volatility shocks or whether its behaviour is partly consistent with a broader innovation-sector risk-off pattern.

3 Data and Methodology

As mentioned in the Introduction, this study hypothesises that sudden increases in financial uncertainty are associated with weaker listed fintech market performance and that this pattern does not only appear in the fintech sector. To empirically examine this hypothesis, a

comparative, association-based study using monthly returns, Pearson's product-moment correlation coefficient (r), scatterplots with simple ordinary least squares (OLS) linear regression, a VIX-level regime test and an R^2 value was conducted.

The FINX return data is used as the main dependent variable in this study for listed fintech market performance since it tracks publicly traded firms connected to financial technology ("FinTech ETF (FINX)"). ARK Blockchain & Fintech Innovation ETF (ARKF) is included as a secondary fintech-related ETF, while Nasdaq CTA Global Digital Payments Index (WALLET) is used as a digital payments index because it is more directly linked to payment technology ("ARKF - ARK Blockchain & Fintech Innovation ETF"; "History for WALLET"). However, WALLET has a shorter sample period, so it is treated as a supporting comparison rather than the main model. BOTZ is included as an AI and robotics benchmark, representing the broader innovation sector ("Robotics & Artificial Intelligence ETF"). BLOK is included as a blockchain and crypto-related financial innovation benchmark and is also used as a fintech-adjacent benchmark ("Amplify ETFs - BLOK"). For uncertainty proxies, changes in the VIX are used as the main independent variable to test the extent of market volatility shocks, while the VIX level is included as a comparison variable representing the static level of expected volatility. The study also uses another uncertainty variable, GEPU change, as a secondary measure of uncertainty to test whether policy-related uncertainty has a similar relationship with listed fintech returns. The VIX and GEPU indices began in January 1990 and February 1997, respectively ("CBOE Volatility Index"; "Global Economic Policy Uncertainty Index"). However, this period of data extends much further back than the majority of ETF series, which are all approximately concentrated from 2016. Therefore, the study does not impose a single start date across all variables. Instead, each pairwise comparison uses the maximum overlapping period for the majority of variables. As a result, the effective sample period for the paper is from September 2016 to April 2026, with 115 monthly observations.

To ensure all variables are analysed in the same structure, all raw data are aligned to a monthly frequency. Monthly returns were calculated for FINX, ARKF, BOTZ, and BLOK using consecutive current monthly closing values divided by the previous monthly closing value minus one ("Global X FinTech ETF"; "ARK Blockchain & Fintech Innovation ETF"; "Global X Robotics & Artificial Intelligence ETF"; "Amplify Blockchain Technology ETF"). Monthly average values were calculated for VIX, the current-price GDP-weighted GEPU index, and WALLET using the equation below because those indices are provided at a daily frequency ("CBOE Volatility Index"; "Global Economic Policy Uncertainty Index"; "History for WALLET"). Where $X_{\square}$ is the monthly value of variables, $X_{D,M}$ is the daily observation on day D within month M, and $n_{\square}$ is the number of available daily observations in that month. These monthly average values were then converted into monthly changes by comparing each month's average value with that of the previous month.

$$X_m = \frac{1}{n_m} \sum_{d=1}^{n_m} X_{D,M}$$

After converting all data into monthly returns and monthly percentage changes, this paper uses the Pearson correlation coefficients, simple ordinary least squares (OLS) linear regression, and R^2 values to examine the association between listed fintech market performance and financial uncertainty.

The Pearson correlation value r is used to test whether two variables tend to move together and whether the relationship is positive or negative. A positive r suggests that as uncertainty rises, the listed fintech market also increases, and vice versa. A negative r indicates that those two variables are moving in opposite directions. Values farther from zero indicate a stronger linear association. The R^2 value measures how much of the variation in the dependent variable is explained by the model. It ranges between 0 and 1. When R^2 is closer to 0, it indicates that only a small proportion of the variation can be explained by the model, whereas, when it is closer to 1, a large proportion of the variation can be explained by the model. Scatterplots with OLS linear regression are used to provide a visual representation of each pairwise relationship and to estimate the direction and magnitude of the fitted linear association. Each point in the OLS linear regression diagram represents one monthly observation, and OLS estimates the best-fitting line by minimising the sum of the squared differences between the observed values. The OLS linear regression model is expressed as the equation below, where Y represents the percentage change in the innovation proxy, X denotes the relevant uncertainty variable change, α is the intercept, β is the slope coefficient indicating the marginal effect in the uncertainty proxy on the financial innovation sector, and ε is the residual.

$$Y = \alpha + \beta X + \varepsilon$$

This study not only tests the relationship between FINX returns and VIX change, but it also compares FINX returns and VIX change with other benchmarks to test whether this phenomenon is fintech-specific. FINX returns are compared with ARKF, BOTZ, BLOK and WALLET returns using Pearson correlation analysis to examine how closely each benchmark co-moves with listed fintech performance. In addition, these innovation-linked benchmarks are compared with VIX change to test whether a sudden increase in financial uncertainty reflects a broader growth-sector risk-off pattern or is a purely fintech-specific phenomenon. The benchmark design also compares different uncertainty measures. The study also compares changes in VIX and GEPU as market-based and policy-based uncertainty may affect the fintech market differently and tests which uncertainty fintech market performance is more closely associated with.

To distinguish sudden volatility shocks from broader conditions in which volatility is already high or low, this VIX-level regime comparison test is established. VIX change measures whether volatility rises sharply from one month to the next and VIX level measures the general level of market uncertainty within a given month. In this study, the sample is divided into volatility regimes based on the VIX level percentile method shown below in Table 1, where P20

and P80 refer to the 20th and 80th percentile thresholds of the VIX level distribution. In this sample, P20 = 13.37 and P80 = 22.94.

Table 1. VIX-level regime classification

VIX regime	Percentile rule	Operational definition	Interpretation
Low VIX regime	Bottom 20%	VIX level < P20	Low-volatility condition
Moderate VIX regime	Middle 60%	P20 ≤ VIX level ≤ P80	Ordinary/moderate volatility condition
Extreme VIX regime	Top 20%	VIX level > P80	High-volatility condition

After each month is classified into a VIX-level regime, this paper calculates the percentage of months with average FINX return, median FINX return, and negative FINX returns that occur in the overall sample. These statistics are used to test whether listed fintech market performances are consistently weaker during high financial uncertainty periods and whether the stronger association appears mainly through sudden changes in volatility. The average FINX return reflects the mean fintech performance, while the median FINX return captures the monthly return that is less sensitive to extreme observations.

4 Empirical Results and Analysis

Figure 1 reports the results of the relationship between FINX returns and VIX and EPU changes. The Pearson correlation coefficient between FINX and VIX returns is approximately -0.5064, which indicates that months with larger increases in VIX tend to have weaker FINX monthly returns. The scatterplot further illustrates the relationship. As the FINX return is treated as the dependent variable and VIX change as the independent variable, the downward-sloping fitted trendline slope (β) of -0.1541 indicates that a 10-percentage-point increase in VIX change is associated with an approximately 1.541-percentage-point reduction in the fitted FINX monthly return. The R^2 value is approximately 0.2564, indicating that VIX change may explain about 25.6% of the variation in monthly FINX returns. These empirical findings align with theoretical expectations that sudden increases in market volatility (VIX) are associated with weaker listed fintech market performance. However, in the scatterplot, observations are moderately dispersed, suggesting that the relationship should be interpreted as partial rather than complete. A decrease in FINX returns may also be influenced by interest rate movements rather than financial uncertainty, as a central bank announcing an increase in the interest rate may suppress investment motivation, causing a decline in FINX returns. Therefore, the relation remains correlational but should not be interpreted as proving a direct causal effect. Moreover, the FINX and VIX results alone do not establish whether listed fintech performance responds similarly to other forms of uncertainty, making comparison with GEPU change necessary.

Figure 1. (OLS) Linear scatter regression graph for FINX returns against VIX and GEPU changes

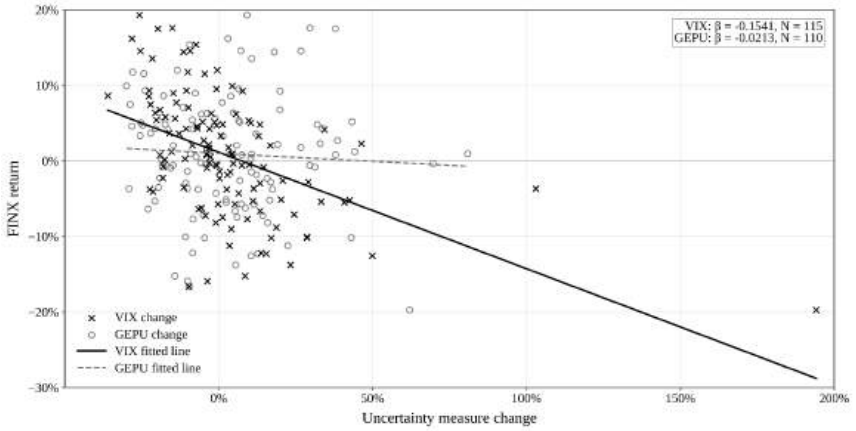


Table 2 compares the relationship between those two uncertainty variables and FINX return. FINX return has a Pearson correlation of −0.0581 with GEPU change compared with −0.5064 with VIX change. The difference is also reflected in their R^2 values, where VIX explains approximately 25.6% of the variation in FINX monthly returns in the simple linear model, while GEPU change explains only approximately 0.34%. Figure 1 provides a consistent visual result for the two uncertainty variables and FINX, where the fitted line slope (β) for FINX returns against GEPU change is -0.0213, which is much flatter than the slope for VIX against FINX (-0.1541). This comparison suggests GEPU explains almost none of the variation in monthly FINX returns. This might be due to GEPU affecting fintech performance through slower channels, such as government policies and future regulation. Its effects may emerge with a time lag that the current same-month analysis does not capture. By contrast, VIX reflects a rapid shift in investment expectations and risk appetite. Therefore, these empirical results align with theoretical expectations that VIX is the more appropriate primary uncertainty measurement for fintech market performance.

Table 2. Pearson Correlation and R^2 results for VIX and GEPU comparison

Relationship	**Correlation (r)**	**R^2**
FINX return and VIX change	-0.5064	0.2564
FINX return and GEPU change	-0.0581	0.0034
VIX change and GEPU change	0.3712	0.1378

The paper also tests whether two uncertainty measures are related but not interchangeable. Their Pearson correlation is approximately 0.3712, with an R^2 value of approximately 0.1378. This indicates that market-volatility shocks and policy-uncertainty shocks sometimes move together, but most of the variation in one measure is not explained by the other.

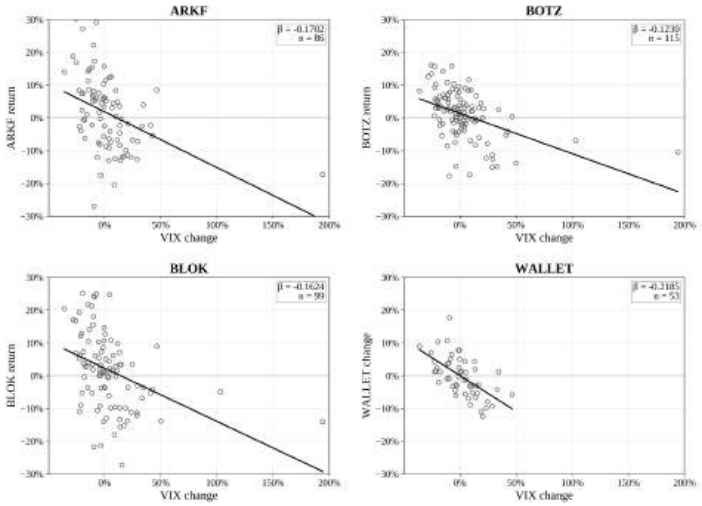
Therefore, the result suggests that GEPU change does not capture the same short-term market movement as VIX change within this monthly ETF return framework. However, GEPU is still meaningful as a secondary uncertainty proxy to compare with the listed fintech market performance.

As mentioned above in the methodology section, the benchmark comparison was designed to determine whether the negative FINX and VIX relationship is fintech-specific or broader innovation risk-off behaviour. Table 3 reports the estimated Pearson correlation value and R^2 for FINX return and VIX change along with benchmarks. In addition, Figure 2 reports the slope coefficient between VIX changes and benchmarks.

Table 3. Correlation between benchmark with VIX changes and FINX returns

Performance measure	Correlation with VIX change (r)	R^2	Correlation with FINX return (r)
ARKF return	−0.4115	0.1693	0.9319
BOTZ return	−0.4544	0.2065	0.7829
BLOK return	−0.3826	0.1464	0.8299
WALLET change	−0.6001	0.3601	0.6541

Figure 2. (OLS) Linear scatter regression graph between VIX changes and benchmarks



ARKF provides the closest fintech-related comparison with a larger r value of 0.9319, indicating that ARKF and FINX move very closely over the overlapping sample. ARKF also shows a negative association with VIX changes, with an r of approximately −0.4115 and an R^2 of approximately 0.1693. The OLS regression graph reports an estimated slope of $\beta = -0.1702$, meaning that a 10-percentagepoint increase in VIX change is associated with an approximately 1.702-percentage-point reduction in the fitted ARKF return. The ARKF result supports the view

that the negative VIX relationship is not limited to FINX alone and that listed fintech-related assets share some sensitivity to market volatility shocks. However, ARKF is still closely tied to fintech and digital finance itself, so it cannot show whether the result is part of a broader non-fintech innovation pattern.

BOTZ provides an important non-fintech comparison because it represents a broader AI and robotics innovation benchmark. It has a positive r of approximately 0.7829 with FINX return, which suggests that it shares some market movement with FINX. More importantly, BOTZ return is also negatively associated with VIX change, with an r of approximately −0.4544 and an R^2 of approximately 0.2065. Its fitted slope is $\beta = -0.1230$, meaning that a 10-percentage-point increase in VIX change is associated with an approximately 1.230-percentage-point reduction in the fitted BOTZ return. Although BOTZ has a smaller absolute β than FINX, its relatively strong correlation and R^2 indicate that the negative relationship remains meaningful. Then BLOK provides a finance-adjacent benchmark as it relates to blockchain and crypto assets. It has a strong positive r with FINX returns at approximately 0.8299. Its relationship with VIX change is also negative, with an r of approximately −0.3826 and an R^2 of approximately 0.1464. Its fitted slope is $\beta = -0.1624$, indicating that a 10-percentage-point increase in VIX change is associated with an approximately 1.624-percentage-point reduction in the fitted BLOK return. However, BLOK has a slightly larger absolute β than FINX, but its weaker correlation and R^2 mean that it should not be interpreted as definitively more sensitive to volatility shocks. Taken together, the empirical results for BOTZ and BLOK suggest that a sudden increase in financial uncertainty is a broader growth-sector risk-off pattern, as non-financial innovation benchmark performance also weakens when VIX change rises.

WALLET provides digital payment asset performance and has the strongest association with VIX change among the benchmarks. Its r with VIX change is approximately −0.6001, with an R^2 of approximately 0.3601. Its correlation with FINX returns is the smallest compared with other benchmarks at approximately 0.6541. Furthermore, WALLET reports the steepest fitted negative slope, with $\beta = -0.2185$, indicating that a 10-percentage-point increase in VIX change is associated with an approximately 2.185-percentage-point reduction in the fitted WALLET monthly change. The results suggest that WALLET is strongly associated with volatility shocks in the available sample period. However, WALLET has the shortest sample period, N=53 (November 2021 to April 2026), whereas BOTZ and FINX contain 115 observations. Therefore, it is only considered as an appendix comparison, and this result serves as a short-period robustness check.

In sum, the benchmark results suggest that the FINX and VIX relationship should not be interpreted as purely fintech-specific. The empirical results suggest that weaker listed fintech performance during volatility shocks is partly connected to a broader innovation-linked market pattern. However, the strength of the relationship differs across benchmarks.

Table 4. VIX-level regime comparison results

VIX regime	Definition	Number of months	Average FINX return	Median FINX return	% negative FINX months
Low VIX	VIX < P20	23	3.85%	3.66%	13.0%
Moderate VIX	P20 ≤ VIX ≤ P80	69	0.48%	−0.49%	52.2%
Extreme VIX	VIX > P80	23	−1.56%	−3.66%	60.9%

Table 4 reports the descriptive statistics showing the average FINX returns, median FINX returns, and negative FINX returns in the overall sample. The VIX-level regime comparison result suggests that listed fintech performance becomes weaker as the VIX level increases. During low-VIX months, FINX recorded an average return of 3.85% and a median return of 3.66%, while only 13.0% of observations produced negative returns. Moderate-VIX months showed weaker and more mixed fintech performance, with an average return of 0.48%, a median return of −0.49%, and negative returns in 52.2% of observations. Extreme-VIX months resulted in the weakest fintech performance, where FINX recorded a negative average return of −1.56% and a negative median return of −3.66%, while 60.9% of observations were negative. The decline in both mean and median returns suggests that the result is not driven entirely by a small number of extreme monthly observations. The increasing proportion of negative-return months also indicates that weaker FINX performance becomes more common under higher-volatility conditions. These findings suggest that higher prevailing VIX levels are associated with weaker listed fintech market performance.

Table 5. Comparison of FINX return relationships with VIX change and VIX level

Uncertainty measure	Correlation with FINX return (r)	R^2
VIX change	-0.5064	0.2564
VIX level	-0.2368	0.0561

In conclusion, the VIX-regime test and the analysis between FINX returns and VIX change both suggest that higher VIX levels and sudden increases in the VIX are associated with weaker FINX performance. However, the strength of the relationship differs substantially. According to data in Table 5, the correlation between FINX return and VIX level is approximately −0.2368, with an R^2 value of approximately 0.0561. By comparison, the correlation between FINX return and VIX change is approximately −0.5064, with an R^2 value of approximately 0.2564. This distinction suggests that listed fintech performance may respond more strongly to volatility shocks than to an already extreme level of uncertainty. The VIX level

captures the general market environment, which may remain high after investors have already adjusted their expectations and asset valuations. A sudden increase in the VIX, by contrast, represents a rapid deterioration in market sentiment and may trigger more immediate repricing, reduced risk appetite, and selling pressure on growth-orientated fintech assets.

5 Limitations

There are several limitations to this study. First, the study uses a fintech ETF and the VIX as proxies rather than direct measures of the fintech sector and uncertainty. This means changes in ETF and index returns are not directly measuring fintech innovation or uncertainty. FINX, ARKF, BLOK, BOTZ and WALLET reflect the investor's reaction to changes in financial market values, but it is influenced by market sentiment, macroeconomic conditions, and firm-specific developments. Also, future growth expectations may already be priced into current market valuations. Therefore, market values do not necessarily reflect the real value of the fintech sector. Thus, a weaker FINX return may be interpreted as weaker listed fintech market performance but not necessarily as direct evidence of weaker fintech innovation.

Second, the empirical methods indicate association rather than causation. Pearson correlation, OLS linear regression, and R^2 do not directly prove that changes in VIX and GEPU cause fintech innovation to improve or worsen, they only show whether two variables are moving in the same direction. Omitted factors, including interest-rate changes, broader equity-market movements, and sector-specific news may also influence the estimated relationships. As a result, the negative relationship between FINX return and VIX change should be interpreted as evidence of association, not as proof of a causal mechanism.

Third, the sample design and regime analysis also limit the strength of the variables. The variables do not share the same sample period. The study only uses a suitable data period for comparison, but this construction of variables may undermine the accuracy of the methodology used by this paper.

6 Conclusion

This study aimed to investigate the association between financial uncertainty and listed fintech performance and also examine whether the association is fintech-specific. It used FINX returns as a proxy for fintech performance and VIX change as a proxy for financial uncertainty. It found that sudden increases in financial uncertainty are negatively associated with listed fintech market performance. The main empirical result indicates that months with larger increases in market volatility are associated with weaker FINX returns. By contrast, another type of uncertainty, GEPU change, shows a much weaker association with fintech performance in the ETF return framework. The VIX-level regime comparison also suggests that higher-volatility conditions are associated with weaker FINX returns, not only sudden increases in uncertainty. However, the relationship between FINX returns and the VIX level is substantially weaker than the relationship between FINX returns and VIX change, indicating that sudden increases in the volatility are associated with a greater weakening of listed fintech performance. The benchmark

comparison suggests that fintech is not uniquely exposed to volatility shocks. ARKF, BOTZ, BLOK, and WALLET also show negative relationships with VIX change, indicating that listed fintech weakness during volatility shocks is partly connected to a broader risk-off innovation-linked market pattern. However, the strength of the relationship differs across benchmarks. FINX remains among the more sensitive benchmarks. These findings should be interpreted with caution. The results are associational rather than causal, and ETF returns should be understood as market performance proxies rather than direct measures of fintech innovation. This means that listed fintech market performance is negatively associated with VIX change, but it cannot determine whether uncertainty directly affects fintech firms' innovation activity. Future studies could also add interest-rate controls to help separate volatility shocks from monetary and discount-rate effects.

This research can be further enhanced and strengthened by using firm-level fintech data, moving from a broad ETF level to examine whether the associations remain across individual fintech firms or are driven mainly by particular business models. Moreover, including market index controls would make it possible to test whether FINX remains sensitive to VIX change after accounting for broader equity market movements. Additionally, formal significance testing would further assess whether the negative association is statistically reliable rather than only descriptively supported by Pearson correlation and OLS linear regression.

Acknowledgments

I would like to thank Dr Jose Valer Davila for his guidance throughout the data collection and empirical analysis of this study, also grateful to Maria Jose Villalba Giubi for her constructive feedback and support in revising and refining the manuscript.

Works Cited

"Amplify Blockchain Technology ETF (BLOK) Stock Historical Prices & Data." Yahoo Finance, https://finance.yahoo.com/quote/BLOK/history/?period1=1516890600&period2=1783844652&frequency=1mo.

"Amplify ETFs - BLOK." Amplify ETFs, https://amplifyetfs.com/blok/.

"ARK Blockchain & Fintech Innovation ETF (ARKF) Stock Historical Prices & Data." Yahoo Finance, https://finance.yahoo.com/quote/ARKF/history/?frequency=1mo&period1=1549290600&period2=1783844279.

"ARKF - ARK Blockchain & Fintech Innovation ETF." ARK Invest, https://www.ark-funds.com/funds/arkf.

Auer, Raphael, et al. The Technology of Decentralized Finance (DeFi). Bank for International Settlements, Jan. 2023. BIS Working Papers, no. 1066, https://www.bis.org/publ/work1066.pdf.

Baker, Scott R., et al. "Measuring Economic Policy Uncertainty." The Quarterly Journal of Economics, vol. 131, no. 4, 2016, pp. 1593–1636, https://doi.org/10.1093/qje/qjw024.

Cboe Global Markets. Volatility Index Methodology: Cboe Volatility Index. https://cdn.cboe.com/resources/indices/Volatility_Index_Methodology_Cboe_Volatility_Index.pdf.

"CBOE Volatility Index: VIX." FRED, Federal Reserve Bank of St. Louis, https://fred.stlouisfed.org/series/VIXCLS.

Chen, Yuxuan, et al. "Givers or Receivers? Return and Volatility Spillovers between Fintech and the Traditional Financial Industry." Finance Research Letters, vol. 46, part B, 2022, article 102458, https://doi.org/10.1016/j.frl.2021.102458.

Davis, Steven J. An Index of Global Economic Policy Uncertainty. National Bureau of Economic Research, Oct. 2016, NBER Working Paper 22740, https://doi.org/10.3386/w22740.

"FinTech ETF (FINX)." Global X ETFs, https://www.globalxetfs.com/funds/finx. "Global Economic Policy Uncertainty Index." Economic Policy Uncertainty, https://www.policyuncertainty.com/global_monthly.html.

"Global X FinTech ETF (FINX) Stock Historical Prices & Data." Yahoo Finance, https://finance.yahoo.com/quote/FINX/history/?frequency=1mo&period1=1473773400&period2=1783843919&filter=history.

"Global X Robotics & Artificial Intelligence ETF (BOTZ) Stock Historical Prices & Data." Yahoo Finance, https://finance.yahoo.com/quote/BOTZ/history/?frequency=1mo&period1=1473773400&period2=1783844432. "History for WALLET." Nasdaq Global Index Watch, https://indexes.nasdaq.com/Index/History/WALLET.

"Monthly EPU Indices for 22 Countries." Economic Policy Uncertainty, https://www.policyuncertainty.com/all_country_data.html.

Paule-Vianez, Jessica, et al. "The Impact of Economic Policy Uncertainty on Stock Types While Considering the Economic Cycle. A Quantile Regression Approach." European Journal of Management and Business Economics, vol. 34, no. 1, 2025, pp. 88–102, https://doi.org/10.1108/EJMBE-12-2022-0365.

"Robotics & Artificial Intelligence ETF (BOTZ)." Global X ETFs, https://www.globalxetfs.com/funds/botz.

"Table Data – S&P 500." FRED, Federal Reserve Bank of St. Louis, https://fred.stlouisfed.org/data/SP500.

The Opportunity Cost of Corruption in Public Resource Allocation: Evidence from India's Education Sector, 2016-2022 By Vivan Balhara

Abstract

India allocates close to 3 percent of GDP annually to education, yet foundational learning deficits remain widespread: the 2022 ASER survey found 42.8 percent of rural Class 5 students could not read a Class 2-level text. This paper asks whether state-level prosecutorial impunity under the Prevention of Corruption Act explains part of that gap, using a panel of 15 major Indian states over 2016-2022. Impunity is defined as one minus the ratio of PCA convictions to registered cases, from NCRB Crime in India reports. Two-way fixed effects regressions, controlling for per capita income and pupil-teacher ratio, associate a one-unit increase in impunity with 6.3 additional points of reading deficiency ($p = 0.107$), with directionally consistent results for enrolment. An opportunity cost exercise using verified UDISE+ enrollment figures implies over 1 million students across three high-impunity states could reach grade-level literacy at Kerala's accountability level. Identification is not causal; weak significance reflects the conservative design with a small panel and a single corruption proxy.

Keywords: corruption; education outcomes; fixed effects; NCRB; India; impunity; public resource allocation

1. Introduction

In 2022, Pratham's Annual Status of Education Report found that 42.8 percent of children enrolled in Class 5 across rural India could not read a text designed for Class 2, down from 50.5 percent in 2018, a decline driven substantially by pandemic-related school closures and the disruption to learning they caused. India's school education system enrols more than 265 million children from the pre-primary through higher secondary stages, of which elementary education (Classes 1 to 8) alone accounts for close to 189 million, and the country draws on combined central and state public expenditure equivalent to between 2.7 and 2.9 percent of GDP each year on education, according to successive editions of the Economic Survey of India. That level of spending is not trivial, which makes the persistence of poor learning outcomes harder to explain away.

A natural question follows: why does substantial public spending on education fail to produce proportionate improvements in learning? The macroeconomic literature identifies several candidate answers: teacher supply constraints, household income, social inequality, and linguistic diversity. This paper investigates one that receives considerable qualitative attention but limited panel scrutiny in the Indian context: corruption in public resource allocation.

The intuition is relatively direct. Elementary education in India is primarily state-funded and state-administered. Resources flow from central government transfers and state budgets through district and block administrative offices to individual schools, and at each node of that chain there is opportunity for diversion. Teachers who are absent but remain on payroll consume salary

expenditure without delivering instruction; infrastructure contracts that are inflated or left unexecuted absorb capital without improving classrooms; and underprovided textbooks and mid-day meals amount to budget allocations that reach an administrator rather than a student. Where the institutional machinery for detecting and penalising these diversions is weak, they tend to persist, and where it functions, they are at least partially constrained.

This paper uses a state-level corruption impunity score, constructed from Prevention of Corruption Act conviction and registration data in NCRB Crime in India reports, to assess whether within-state variation in prosecutorial accountability over 2016-2022 predicts elementary education outcomes. Three outcome variables are examined: reading deficiency rates from ASER surveys, school dropout rates, and gross enrolment ratios from UDISE+ Flash Statistics.

The results below are reported as they came out of the regressions, without dressing up the interpretation. The two-way fixed effects estimator, while appropriate for this setting, is conservative by construction: it absorbs all time-invariant cross-state variation through state fixed effects, leaving only within-state variation in impunity over time to explain within-state variation in outcomes. With fifteen states and seven years, degrees of freedom are limited. The headline reading deficiency result, a coefficient of 6.3 with $p = 0.107$, does not clear the conventional 5 percent threshold. The dropout coefficient has the wrong sign, a consequence of COVID-related data artifacts examined in detail below. The GER result is directionally correct but also statistically insignificant.

What comes out of this exercise is three things, best read together rather than in isolation. There is a reproducible panel dataset that combines administrative sources not previously used together in this way. There is regression evidence that is directionally consistent, alongside a cross-sectional pattern stark enough to be informative even without a causal interpretation. And there is an opportunity cost framework that turns the coefficient into a headcount of students, with the caveats stated plainly rather than left in a footnote.

The rest of the paper proceeds as follows. Section 2 reviews the relevant literature. Section 3 develops the theoretical mechanisms. Section 4 describes the data. Section 5 presents the methodology. Section 6 reports results. Section 7 estimates the opportunity cost. Section 8 presents robustness checks. Section 9 states limitations. Section 10 concludes.

2. Literature Review

The relationship between corruption and economic performance has been studied rigorously since Mauro's (1995) cross-country analysis, which showed that higher perceived corruption is associated with lower investment ratios and slower GDP growth. The paper used ethno-linguistic fractionalization as an instrument for perceived corruption, a creative if controversial identification strategy, and its directional finding has held across a range of subsequent methodological approaches. For the purposes of this paper, the relevant extension runs from corruption's aggregate effect on growth to its specific effect on public service delivery outcomes.

Gupta, Davoodi, and Tiongson (2000), in an IMF Working Paper examining a broad cross-country sample, found that higher corruption is associated not only with lower public expenditure on education as a share of GDP but also with worse outcomes conditional on that spending: higher dropout rates, lower literacy, worse health indicators. The key contribution of their analysis is the distinction between quantity and quality of public expenditure: corrupt governments do tend to spend less on education, but even the resources they do allocate deliver lower educational returns. This distinction matters for the Indian case because India's elementary education expenditure is not obviously deficient in quantity; the question is what those resources actually purchase at the school level.

The most direct empirical predecessor of this paper is Reinikka and Svensson's (2004) study of Uganda's primary school capitation grant programme. Using a newspaper public information campaign as quasi-experimental variation, they demonstrated that schools in districts with better access to information about their grant entitlements captured a substantially larger share of those grants. Their headline finding, that Ugandan schools received on average only 13 percent of the capitation grant to which they were entitled, became one of the most cited pieces of evidence for the magnitude of corruption's effect on public education funding. The design is not replicable in the Indian context studied here, since no analogous information shock occurs within the 2016-2022 study window, but the theoretical mechanism it validates is directly applicable: funds allocated at the centre do not reliably reach the schools they are intended for.

Teacher absenteeism is a related but distinct channel, with its own empirical literature. Chaudhury, Hammer, Kremer, Muralidharan, and Rogers (2006) conducted unannounced visits to primary schools across six developing countries and found an average teacher absence rate of approximately 19 percent across the sample, with India recording one of the higher country-level rates at roughly one-quarter of government primary teachers absent on a given day. The India-specific evidence was extended and deepened by Muralidharan, Das, Holla, and Mohpal (2017), who used a later nationally representative panel and estimated the salary cost of unauthorized teacher absence in India at approximately US$1.5 billion annually at 2010 exchange rates. They document that high-absence schools are disproportionately concentrated in districts with weak administrative accountability structures. Their work provides the empirical foundation for the teacher absenteeism mechanism this paper invokes in Section 3.

For India specifically, the existing quantitative literature on corruption and education is largely composed of single-state studies or relies on perceived corruption indices with insufficient panel coverage. The Comptroller and Auditor General produces education department audit reports that document specific irregularities: ghost beneficiaries in mid-day meal schemes, construction payments for work not completed, and teacher recruitment irregularities. These reports are state-specific and inconsistently formatted across states and years, which is why the CAG audit trail could not be operationalized as a panel variable here. They serve in this paper as qualitative case evidence for the mechanisms invoked theoretically. The contribution of this paper to the existing literature is narrow but specific. Cross-country studies lose the institutional specificity of India's federal structure: the variation in how state

governments administer anti-corruption institutions, how actively they register and prosecute PCA cases, and how that variation has evolved over time. Single-state studies capture institutional detail but cannot isolate the effect of corruption from the many other ways Indian states differ from each other. A within-India panel exploiting variation across 15 states over 7 years, with state and year fixed effects, addresses the cross-state omitted variable problem that cross-sectional analysis cannot handle and generalises in a way that single-state work cannot. That is this paper's methodological contribution, and a modest one at that.

3. Theoretical Framework

Three mechanisms plausibly link prosecutorial impunity under the Prevention of Corruption Act to elementary education outcomes. They operate at different points in the resource allocation chain and are not mutually exclusive.

Teacher absenteeism. Teacher salaries account for the large majority of recurring elementary education expenditure in India. Ghost teachers, meaning those on payroll who do not exist, and chronically absent teachers amount to a diversion of salary funds that produces zero instructional output. Maintaining this arrangement over time requires either active connivance by administrative supervisors or at minimum their unwillingness to investigate. In states where the institutional apparatus for detecting and penalizing corruption is weak, as proxied here by consistent failure to secure PCA convictions, the expected cost of maintaining such arrangements is low. One would therefore expect states with higher prosecutorial impunity to have higher effective teacher absence, lower instructional time, and as a consequence, worse learning outcomes. The reading deficiency outcome variable captures this most directly: it measures whether students have achieved the most basic threshold of literacy despite several years of enrollment.

Fund diversion at the transfer chain. A substantial portion of elementary education funding flows through scheme-based transfers: the Samagra Shiksha Abhiyan, the mid-day meal programme, free textbook distribution, and school infrastructure grants under the Rashtriya Madhyamik Shiksha Abhiyan. Each transfer involves procurement decisions and funds moving through multiple administrative layers. Corruption at the block or district level can manifest as inflated procurement bills, payment for goods not delivered, or straightforward misappropriation of scheme funds. The educational impact depends on what is diverted: diversion of mid-day meal funds affects both nutrition and the incentive to attend school, diversion from infrastructure grants leaves schools without classrooms, toilets, or drinking water, and diversion of textbook funds simply deprives students of materials. These mechanisms affect both enrollment (captured in GER) and learning (captured in reading deficiency). The 2004 capitation grant evidence from Uganda (Reinikka and Svensson, 2004) provides the most rigorous cross-national evidence for how large these diversions can be; CAG audit reports for Indian states document that similar patterns occur domestically, even if not in a form that can be systematically quantified across states.

Procurement rent extraction in school construction. Capital expenditure in elementary education, covering classrooms, boundary walls, and separate toilet facilities for girls, is administered through state public works departments or dedicated construction agencies. Construction contracts are subject to the familiar problems of procurement corruption: bid rigging, contractor collusion, payments for substandard work, and certification of incomplete projects. The connection to educational outcomes is somewhat indirect: poor infrastructure does not by itself prevent teaching, but it affects student attendance, particularly girls' attendance given evidence that adequate toilet provision significantly increases female enrollment and retention in Indian schools. Gross Enrolment Ratio is most likely sensitive to this channel. What makes all three mechanisms jointly susceptible to a single institutional measure is that each requires the persistent failure of investigative and prosecutorial accountability. A functional enforcement apparatus that actively registers PCA cases and secures convictions raises the expected cost of all three practices. The NCRB impunity score captures the long-run failure of this apparatus at the state level. In states where cases are registered but convictions rarely follow, the deterrent effect of anti-corruption law is minimal regardless of how many cases enter the pipeline.

4. Data and Variables

4.1 Sample and Study Period

The study covers 15 major Indian states over 2016-2022, yielding a maximum of 105 state-year observations. That maximum is not achieved in practice; missing and excluded observations documented in Section 4.2 reduce the usable panel. Summing Census 2011 population figures for the 15 sampled states against the all-India total of approximately 1.21 billion (Office of the Registrar General & Census Commissioner, 2011) gives a combined share of roughly 86 percent, calculated using Telangana's post-bifurcation population as derived from undivided Andhra Pradesh's 2011 figures, consistent with the approximation already flagged for Telangana's urbanisation share in Section 9. The sample spans a wide range of institutional quality, geographic size, and economic development, making it a reasonable if not exhaustive basis for within-India inference.

The states are grouped into three prior tiers based on their classification in the corruption literature and their NCRB impunity scores. The high-corruption tier consists of Uttar Pradesh, Bihar, Rajasthan, Odisha, Jharkhand, and Assam. The middle tier consists of West Bengal, Chhattisgarh, Telangana, and Maharashtra. The low-corruption tier consists of Kerala, Tamil Nadu, Gujarat, Karnataka, and Madhya Pradesh. Madhya Pradesh's placement in the low tier requires an immediate note: its mean impunity score of 0.466 is the lowest in the sample, but this reflects anomalously high Anti-Corruption Bureau conviction rates in 2017-2019, likely an artifact of courts clearing a pre-existing backlog, rather than institutionally robust enforcement. This classification anomaly is discussed as a limitation in Section 9 and flagged in all tables.

The study period begins in 2016 for a specific data reason. NCRB Crime in India Chapter 8C, which presents state-wise Prevention of Corruption Act conviction and registration data, adopts a consistent and comparable table format from 2016 onward. The 2014 and 2015 reports used a different chapter structure that makes longitudinal comparison problematic. 2016 is therefore the first year of clean, usable data.

4.2 Corruption Measure: NCRB Impunity Score

The primary corruption measure is constructed from NCRB Crime in India, Volume II, Tables 8C.1 and 8C.4, for 2016-2022. The impunity score for state i in year t is:

ncrb(i,t) = 1 - [Convictions(i,t) / Cases Registered(i,t)]

A score approaching 1.0 indicates near-complete prosecutorial failure: cases are registered but almost none result in conviction. A score approaching 0 indicates convictions keep pace with registrations. The theoretical range is [0, 1] for positive case volumes; negative values occur when courts convict on a backlog from prior years, and these years are excluded from the panel (see below).

Table 1 presents the full state-year impunity matrix.

Table 1: NCRB Impunity Scores by State and Year, 2016-2022

State	2016	2017	2018	2019	2020	2021	2022	Mean
Uttar Pradesh	0.833	0.948	0.857	0.918	0.984	0.866	0.832	0.891
Bihar	0.890	0.750	0.899	0.944	1.000	0.975	0.954	0.916
Madhya Pradesh	0.659	0.129	0.125	0.233	0.854	0.636	0.626	0.466
Rajasthan	0.806	0.748	0.659	0.658	0.846	0.890	0.787	0.770
Odisha	0.854	0.759	0.759	0.810	0.935	0.860	0.686	0.809
Jharkhand	1.000	0.977	0.789	0.701	0.762	0.982	0.776	0.855
West Bengal	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Chhattisgarh	0.459	—	—	—	—	1.000	0.692	0.717*
Assam	0.870	0.926	0.960	0.889	0.875	0.875	0.965	0.909
Telangana	0.753	0.455	0.698	0.944	0.917	0.928	0.925	0.803
Kerala	—	0.803	0.803	0.587	—	0.484	0.523	0.640*
Tamil Nadu	0.559	0.611	0.583	0.778	0.796	0.853	0.671	0.693
Gujarat	0.872	0.547	0.721	0.922	0.889	0.838	0.801	0.799
Maharashtra	0.910	0.936	0.940	0.938	0.985	0.977	0.957	0.949
Karnataka	—	0.820	0.892	0.868	0.865	0.894	0.846	0.864*

Source: NCRB Crime in India, Vol. II, Tables 8C.1 and 8C.4, 2016-2022. Notes: — indicates year excluded due to data anomaly (see text). * Mean computed over available years only.

Three states required year-specific exclusions. Karnataka's 2016 value is excluded because only 25 PCA cases were registered that year, against 401 in 2014 and 289 in 2017, indicating a source data error in the published report. Kerala's 2016 and 2020 values are excluded: the 2016 figure reflects anomalously high case registrations (430 in a single year) that distort the score, while the 2020 figure is missing from the published NCRB report with no explanation provided. Chhattisgarh's 2017-2020 values are excluded because court convictions exceeded registered cases in those years, producing negative nominal impunity values that reflect courts clearing pre-2017 backlogs rather than current enforcement activity.

West Bengal requires a separate note, because a score of 1.000 that holds unchanged across all seven years might look like a data error rather than a real finding. The West Bengal Lokayukta Act, 2003, established the office in name, but the West Bengal Lokayukta (Amendment) Act, 2018, narrowed its practical reach further: it excluded the Chief Minister from the institution's investigative purview and required prior state government approval before any complaint against a public servant could be investigated, restrictions that limit the office's independence relative to Lokayuktas in states such as Kerala or Karnataka. Against this institutional backdrop, annual PCA case registrations in West Bengal were minimal, between 8 and 25 cases per year across the full study window, with zero convictions recorded in each year. The 1.000 score accurately captures a state of sustained prosecutorial inaction and is retained in the panel as an extreme case of institutional weakness. Because West Bengal's impunity score never changes, it contributes no within-state variation to the fixed effects estimator and its inclusion does not affect the regression coefficients; it does, however, anchor the cross-sectional descriptive comparisons.

4.3 Education Outcome Variables

Three outcome variables are used, each capturing a distinct dimension of elementary education performance.

Reading deficiency is defined as the percentage of Class 5 students unable to read a Class 2-level text. Data come from ASER (Annual Status of Education Report) surveys published by Pratham. ASER uses household-based sampling methodology, which means it captures out-of-school children as well as enrolled ones, an advantage over school-based surveys that systematically miss the most disadvantaged students. Within the 2016-2022 study period, ASER conducted full national rural surveys in 2018 and 2022 only. The reading variable therefore has at most two observations per state, a severe constraint on statistical power discussed explicitly in the limitations.

School dropout is the percentage of enrolled elementary students (Classes 1-8) who leave before completing the elementary cycle, sourced from UDISE+ Flash Statistics for 2018-19 through 2022-23. UDISE+ provides annual state-level data with consistent variable definitions across the study period. A critical data quality problem affects 2020-21: dropout rates for that year are zero or near-zero for most states because schools were closed and administrative data collection was disrupted under COVID. These zeros are administrative placeholders rather than

genuine measurements and create a downward artifact in the within-state correlation between impunity and dropout. The problem is tested in robustness checks and discussed as a limitation. Gross Enrolment Ratio at the elementary level is also drawn from UDISE+ Flash Statistics, 2018-19 to 2022-23. GER measures total enrollment as a proportion of the age-eligible population. Five states in the sample report mean GER exceeding 100, namely West Bengal, Assam, Telangana, Karnataka, and Maharashtra, reflecting over-age student retention and cross-border enrollment flows. These values are genuine features of the administrative data and are not data errors, but they complicate regression interpretation and are discussed as a limitation. Net Enrolment Ratio, which is bounded between 0 and 100 and available in UDISE+, would be a cleaner alternative in future extensions of this work.

4.4 Control Variables

Per capita NSDP at constant 2011-12 prices is drawn from the RBI Handbook of Statistics on Indian Economy. This controls income effects on education outcomes: wealthier states can fund better teachers, more materials, and better infrastructure independent of whatever corruption-related dynamics are also at work.

Urbanisation is measured as the proportion of state population residing in urban areas, using Census of India 2011 figures (Office of the Registrar General & Census Commissioner, 2011) applied statically across all years. This variable is time-invariant within the dataset. In the two-way fixed effects specification, it is automatically absorbed by state fixed effects and does not enter the estimating equation directly; it functions as a control only in pooled OLS.

Pupil-teacher ratio at the primary level (Classes 1-5) comes from UDISE+ Flash Statistics. Including PTR ensures that the NCRB coefficient does not pick up the effect of teacher shortages on learning outcomes, a channel distinct from corruption, namely insufficient teacher supply, as opposed to misallocation of the teachers who are nominally present.

4.5 Variables Evaluated and Dropped

Several variables considered in the original design were evaluated and dropped before estimation.

CAG irregularity data: Comptroller and Auditor General compliance audit reports document specific financial irregularities in state education departments but are not published in a consistent numerical format across all 15 states in every year. They cannot be transformed into a panel variable without imposing arbitrary discretion on what counts as an “irregularity unit.” CAG findings are used qualitatively in Section 3 but not quantitatively.

RTI rejection rates: Right to Information compliance data for state School Education Departments is maintained by State Information Commissions rather than the Central Information Commission. Coverage and format vary significantly across states, and several states, including Tamil Nadu, Andhra Pradesh, and Jharkhand, had overlapping or fragmented SIC jurisdictions during the study period. A consistent cross-state panel series was not constructable.

Lokayukta complaint backlog: Annual pending complaint and resolution rate data is published by Lokayukta institutions in inconsistent formats. Kerala, Karnataka, and Madhya Pradesh have reasonably active and documented Lokayukta offices. Bihar and Uttar Pradesh lacked functional Lokayukta institutions or did not publish systematic data during large portions of the study period, while West Bengal's Lokayukta, though nominally established, operated under statutory restrictions, including the 2018 amendment narrowing its jurisdiction over complaints against senior officials, that left it without a meaningful enforcement record to draw on.

Transparency International sub-national scores: TI's state-level corruption surveys are not annual instruments. At most one or two observations per state exist within the 2016-2022 window, making panel inclusion impossible.

The failure to build a composite corruption index from multiple measures reflects the data constraints documented above rather than an oversight in the research design. NCRB alone is used as a single noisy proxy, with the understanding that measurement error in a single instrument attenuates coefficient estimates toward zero and biases against finding a statistically significant effect.

5. Methodology

5.1 Baseline OLS Specification

As a baseline, a pooled OLS regression is estimated separately for each outcome variable:

$$Y(i,t) = a + b * ncrb(i,t) + c * ndsp(i,t) + d * ptr(i,t) + e * urban(i) + u(i,t)$$

where Y(i,t) is the outcome for state i in year t, ncrb(i,t) is the impunity score, ndsp(i,t) is the log of per capita NSDP, ptr(i,t) is the primary pupil-teacher ratio, and urban(i) is the time-invariant 2011 urbanisation share. OLS pools all variation, both within-state over time and across states, into a single coefficient. The limitation is well-known: states differ on many dimensions that correlate with both corruption and education outcomes. Richer states systematically tend toward lower impunity scores and better outcomes. The OLS coefficient for ncrb therefore conflates the income effect with the corruption effect and is reported only as a comparator to the fixed effects estimates.

5.2 Two-Way Fixed Effects Specification

The main specification is a two-way fixed effects model:

$$Y(i,t) = a(i) + l(t) + b * ncrb(i,t) + c * ndsp(i,t) + d * ptr(i,t) + e(i,t)$$

State fixed effects a(i) absorb all time-invariant differences between states: historical institutional quality, geographic endowments, cultural factors, and any stable features of each state's political economy that might simultaneously predict both corruption and education outcomes. Year fixed effects l(t) absorb shocks common to all states in a given year, such as national policy changes, macroeconomic cycles, and central government education scheme

modifications. What remains in the error term is idiosyncratic variation, and what the coefficient b identifies is the within-state relationship: when a state's impunity score rises or falls over 2016-2022, does its education outcome move in the predicted direction, after removing the national year trend?

This estimator is conservative precisely because it removes so much variation. For states with relatively stable impunity scores across the seven years, which describes most states in the sample since institutional change is slow, the FE estimator has limited variation to work with. West Bengal's invariant score of 1.000 contributes nothing to within-state identification. This is the trade-off of the design: it sacrifices statistical power in exchange for protection against cross-state omitted variable bias.

Estimation is carried out in R 4.6.0 using the plm package for panel models. The key command is:

```
plm(outcome ~ ncrb + ndsp + ptr, data=pdata, model='within', effect='twoways')
```

5.3 Standard Errors

Standard errors are clustered at the state level using the HC1 heteroskedasticity-consistent estimator, implemented via coeftest() with vcovHC(type='HC1', cluster='group'). Clustering at the state level accounts for serial correlation within states across years, which is to be expected when the same institutional and political features affect outcomes in successive years. Cameron and Miller (2015) note that cluster-robust inference with fewer than 30 clusters may be liberal, meaning confidence intervals may be too narrow, and 15 states fall well below that threshold. The standard errors reported here should therefore be treated as approximate rather than exact; if anything, they likely understate the true uncertainty.

5.4 Why Difference-in-Differences Was Not Pursued

Difference-in-Differences was the planned identification strategy, intended to exploit Lokayukta institutional reform timing as quasi-experimental variation. The design requires a well-defined reform event, such as passage of legislation, restoration of powers, or appointment of an active Lokayukta, occurring at different times in different states so that treated states can be compared against controls. No qualifying events occurred within the 2016-2022 study window. Karnataka, the most active Lokayukta state in India, had its institution's powers significantly curtailed in 2016 and only partially restored by 2019. Madhya Pradesh had a Lokayukta but its operational effectiveness fluctuated without a clean break. Bihar lacked a functional Lokayukta institution throughout, and West Bengal's Lokayukta operated under jurisdictional restrictions, formalized further by the 2018 amendment discussed in Section 4.2, that kept it from generating a usable reform event. The DiD design is kept in this section because it was the original plan, and the reason it could not be executed says something substantive about the institutional landscape of the period: no state in the sample underwent a clean, dateable reform of its anti-corruption apparatus during 2016-2022.

6. Results

6.1 Descriptive Evidence

The cross-sectional pattern in the data deserves attention before regression results are presented. Table 2 reports mean impunity scores and mean education outcomes for each state.

Table 2: State-Level Averages, 2016-2022

State	Tier	Mean NCRB	Mean Dropout (%)	Mean GER
Uttar Pradesh	High	0.891	7.22	92.72
Jharkhand	High	0.855	5.70	95.00
Assam	High	0.909	5.04	107.64
Bihar	High	0.916	4.16	90.08
West Bengal	Middle	1.000	4.16	104.88
Rajasthan	High	0.770	3.42	100.12
Madhya Pradesh	Low*	0.466	3.06	89.08
Chhattisgarh	Middle	0.717	2.28	94.78
Maharashtra	Middle	0.949†	1.42	102.60
Tamil Nadu	Low	0.693	1.04	98.12
Gujarat	Low	0.799	0.86	93.74
Odisha	High	0.809	0.84	94.94
Telangana	Middle	0.803	0.80	107.14
Kerala	Low	0.640	0.66	99.04
Karnataka	Low	0.864†	0.46	105.92

Source: NCRB Crime in India 2016-2022; UDISE+ Flash Statistics 2018-19 to 2022-23. * Madhya Pradesh NCRB score reflects ACB conviction artifact; see text. † High impunity scores in institutionally stronger states are not necessarily evidence of weak enforcement. Maharashtra registers more PCA cases nationally than any other state in the sample, which mechanically raises the denominator of the impunity ratio even where conviction activity is comparatively normal; this is documented directly in Section 9 under the NCRB registration paradox. Karnataka's elevated score (0.864) plausibly reflects a related dynamic, since Karnataka also maintains one of the more active anti-corruption bureaus in the sample, but this study does not have state-level case registration volumes to confirm that mechanism for Karnataka specifically, and the comparison should be read as suggestive rather than established.

The dropout gradient is the most immediate finding from Table 2. Uttar Pradesh records a mean dropout rate of 7.22 percent against a mean impunity score of 0.891. Kerala records 0.66

percent dropout against an impunity score of 0.640. Jharkhand (5.70 percent, 0.855) and Bihar (4.16 percent, 0.916) sit at the high end, while Karnataka (0.46 percent, 0.864) and Tamil Nadu (1.04 percent, 0.693) sit at the low end. The rank correlation is positive and visually apparent, though Maharashtra (0.949 impunity, 1.42 percent dropout) is a notable exception: a high-impunity state with relatively low dropout, suggesting that income effects and Maharashtra's urbanisation level partially offset the corruption channel for this outcome.

GER shows a noisier pattern. Several high-impunity states record GER above 100 (West Bengal, Assam), which reflects over-age student retention rather than genuine universal enrollment. Madhya Pradesh, with the lowest impunity score, has one of the lower GER values in the sample. The cross-sectional correlation between impunity and GER is weaker than for dropout, consistent with GER's sensitivity to demographic and infrastructure factors beyond the corruption channel.

6.2 Regression Results

Table 3 presents the two-way fixed effects regression results for all three outcome variables. The coefficient reported is for the NCRB impunity score; control variable coefficients are estimated but not reproduced in the main table.

Table 3: Two-Way Fixed Effects Estimates, NCRB Impunity Score Coefficients

Outcome Variable	Coefficient	Std. Error	p-value	Direction
Reading deficiency (%)	+6.301	3.413	0.107	Correct (positive)
Gross Enrolment Ratio	-4.131	2.855	0.155	Correct (negative)
Dropout rate (%)	-3.542	1.819	0.058	Wrong (negative)

Notes: All regressions include state and year fixed effects and control for log per capita NSDP and primary pupil-teacher ratio. Standard errors clustered at state level (HC1). N = 71 for dropout and GER models (out of a balanced maximum of 75, with Chhattisgarh contributing 2 of 5 years and Kerala 4 of 5 years due to NCRB exclusions in those years); N = 29 for the reading model (out of a balanced maximum of 30, with Chhattisgarh contributing 1 of 2 years for the same reason).

6.3 Reading Deficiency

The reading deficiency coefficient is 6.301 percentage points (p = 0.107). A one-unit increase in the NCRB impunity score, moving from the theoretical limit of full conviction to complete prosecutorial failure, is associated with 6.3 additional percentage points of Class 5 students unable to read at grade level, controlling for per capita income, pupil-teacher ratio, and state and year fixed effects. The direction is consistent with the teacher absenteeism and fund diversion mechanisms in Section 3: states where anti-corruption enforcement produces fewer convictions tend, within their own history over 2016-2022, to show worse learning outcomes. The coefficient does not achieve statistical significance at the 5 percent or 10 percent threshold. The low precision has a structural explanation beyond small N. The reading model uses ASER data from 2018 and 2022 only, producing 29 usable observations once Chhattisgarh's missing

2018 NCRB value is accounted for. Two-way fixed effects then absorb 15 state means and 2 year means, leaving only the change from 2018 to 2022 within each state, after removing the national trend, as the identifying variation. Estimating a coefficient from this residual with any precision requires that the within-state change in impunity between 2018 and 2022 correlates strongly with the within-state change in reading deficiency after the year fixed effect is removed. For states where impunity barely moved between 2018 and 2022, the reading model cannot detect a relationship even if one exists. The coefficient should be read as an indication of direction and order of magnitude rather than a precise estimate.

6.4 Gross Enrolment Ratio

The GER coefficient is -4.131 ($p = 0.155$). The negative sign means higher impunity is associated with lower gross enrolment, which is directionally correct. The coefficient is not statistically significant. The GER outcome is further complicated by the presence of states where GER exceeds 100, which creates heterogeneity in the outcome's interpretation that the linear FE model cannot cleanly handle. The GER result provides weak corroborating evidence for the direction of the hypothesis but is not the primary empirical contribution of this paper.

6.5 Dropout Rate

The dropout rate coefficient is -3.542 ($p = 0.058$), the most problematic result, since the sign is wrong: a negative coefficient implies higher impunity predicts lower dropout, the opposite of the theoretical prediction.

The primary explanation is the COVID data artifact described in Section 4.3. For 2020-21, most states reported zero or near-zero dropout because schools were closed and administrative recording was suspended. High-corruption states, including Bihar (0 percent in 2020), Odisha (0 percent in 2020 and 2021), and UP (0 percent in 2020), are among the most affected. The presence of these zeros creates a negative within-state association: in 2020, when impunity was at or near its highest for several states, dropout was artificially recorded at zero. The FE model interprets this pattern as: when impunity was high, dropout was low, an artifact of data collection failure rather than a substantive relationship.

This interpretation is examined in Section 8.1. Excluding 2020 and 2021 from the sample and reestimating produces a coefficient of -5.644 ($p = 0.146$), which is more negative and more imprecise rather than correct. The persistence of the wrong sign after COVID-year exclusion points to a secondary issue: floor effects in the dropout series for several otherwise well-performing states (Karnataka mean 0.46 percent, Gujarat mean 0.86 percent) leave very little within-state variation for identification. Dropout is concluded to be a weak outcome variable for two-way FE estimation in this setting.

7. Opportunity Cost Estimation

7.1 Methodology

The reading deficiency coefficient is used to estimate the number of additional elementary students who would read at grade level if high-impunity states reduced their impunity scores to the level of Kerala, the lowest-impunity state in the sample (mean ncrb = 0.640). This is an associational counterfactual, not a causal prediction. The exercise asks: based on the within-state correlation estimated in the regression, what does the coefficient imply for the scale of the learning gap attributable to the impunity differential?

Four steps are applied for each of three target states: Uttar Pradesh, Bihar, and Rajasthan.

1. Compute the impunity gap between the target state and Kerala.

2. Multiply the gap by the FE coefficient (6.301) to obtain the predicted reduction in reading deficiency in percentage points.

3. Apply this percentage point reduction to the state's elementary enrollment, taken directly from Table 5.1 of the UDISE+ Flash Statistics 2021-22 report (Classes 1 to 8, all school management types), to obtain a student headcount.

4. State the assumptions and interpret the result as an order of magnitude rather than a point estimate.

7.2 State-Level Calculations

Uttar Pradesh

Impunity gap: 0.891 - 0.640 = 0.251

Predicted reduction in reading deficiency: 6.301 x 0.251 = 1.58 percentage points

Elementary enrollment (UDISE+ Flash Statistics 2021-22, Table 5.1): 35,772,729 students

Implied student headcount: 0.015816 x 35,772,729 = approximately 566,000 students

Bihar

Impunity gap: 0.916 - 0.640 = 0.276

Predicted reduction in reading deficiency: 6.301 x 0.276 = 1.74 percentage points

Elementary enrollment (UDISE+ Flash Statistics 2021-22, Table 5.1): 21,433,075 students

Implied student headcount: 0.017391 x 21,433,075 = approximately 373,000 students

Rajasthan

Impunity gap: 0.770 - 0.640 = 0.130

Predicted reduction in reading deficiency: 6.301 x 0.130 = 0.82 percentage points

Elementary enrollment (UDISE+ Flash Statistics 2021-22, Table 5.1): 12,564,969 students

Implied student headcount: 0.008191 x 12,564,969 = approximately 103,000 students

Combined total across three states: approximately 1,041,000 additional students at grade-level reading.

7.3 Interpretation and Caveats

Several qualifications are essential. The regression coefficient carries substantial uncertainty (p = 0.107); the headcount therefore reflects that same uncertainty and should be interpreted as an order of magnitude, not a point estimate. The counterfactual assumes the impunity-outcome relationship is locally linear and approximately stable, neither of which is

directly testable from this data. Kerala is used as the comparator because it is the lowest observed impunity value in the sample, but it is not a theoretical minimum: reducing impunity beyond Kerala's level would presumably yield further gains if the relationship holds.

The exercise does not translate these headcounts into rupee terms in the main text, because doing so requires wage return assumptions, employment projections, and discount rates that this study cannot independently estimate. The labour economics literature consistently finds positive returns to basic literacy, and over a million additional literate students would imply non-trivial productivity gains over their working lives, but the precision of any rupee figure would be driven by assumptions rather than evidence. The headcount alone conveys the human scale of what the regression coefficient implies.

The purpose of this exercise is to give empirical content to a coefficient that otherwise sits as an abstract statistical quantity. A finding that higher prosecutorial impunity is associated with worse reading outcomes matters more when the within-state coefficient, translated to verified enrollment figures, implies that over a million children in three states alone may be functionally illiterate who would not be in a counterfactual where state prosecution of corruption functioned at Kerala's observed level.

8. Robustness Checks

Three robustness checks are reported, targeted at the main threats to the baseline results.

8.1 COVID Year Exclusion

The dropout model is most exposed to the COVID artifact in 2020-21 data. Excluding 2020 and 2021 and reestimating the fixed effects model produces a coefficient of -5.644 ($p = 0.146$) for dropout. The sign is unchanged and the magnitude is larger. This confirms that the COVID artifact is not the sole driver of the wrong sign: even without those years, within-state variation in dropout does not correlate positively with within-state variation in impunity. The floor effects discussed in Section 6.5 appear to be the residual problem, since states like Karnataka with consistently near-zero dropout leave no upward movement to be explained. The reading and GER results are not affected by COVID-year exclusion because ASER data for 2020 and 2021 does not exist in any form, and the GER COVID anomaly was handled separately.

8.2 Leave-One-Out Analysis

To assess whether any single state drives the reading deficiency result, the FE model is re-estimated fifteen times, each time dropping one state and recording the ncrb coefficient. Stability of the coefficient across these sub-samples would indicate that the result is not dependent on any one influential observation. West Bengal, which contributes no within-state variation (constant score of 1.000 across all seven years), should have no effect on the reading coefficient when dropped; this serves as an internal validity check. Maharashtra, with a mean impunity score of 0.949 and mean dropout of 1.42 percent, is the most likely candidate for individual influence on the reading result given its combination of high impunity and relatively

moderate deficiency. Full leave-one-out results are available from the replication code in Appendix A.

8.3 Lagged Corruption

A one-year lag of the impunity score (ncrb(i, t-1)) replaces the contemporaneous score in an alternative specification. It is reasonable to think that education outcomes in year t are influenced by the corruption environment of the preceding year, since teacher absenteeism patterns and fund diversion arrangements develop over time. If the lagged coefficient is directionally similar to the main result, this provides partial protection against short-run reverse causality. The specification loses one year of data per state (2016 observations are used as lags for 2017 outcomes but cannot themselves have a lagged predictor). Results are available from the replication code.

9. Limitations

Each of the following limitations applies to the analysis and must be weighed by any reader evaluating its conclusions.

NCRB registration paradox. States with more active anti-corruption bureaus register more PCA cases, which mechanically inflates the denominator of the impunity score. Maharashtra registers the most PCA cases nationally each year and records a mean impunity score of 0.949, higher than Bihar and comparable to Jharkhand. This likely reflects genuine institutional activity in which cases are filed faster than courts can process them, rather than corruption of the same character as in states that register almost no cases. The impunity score conflates these two institutional situations, a problem that a composite index with multiple measures would partially address.

Conviction lag. Courts convict on cases registered in prior years, sometimes many years prior. Using same-year registrations and convictions introduces noise in the impunity measure, particularly for states with small annual case volumes where the ratio is sensitive to the timing of individual verdicts. Chhattisgarh's anomalous 2017-2020 scores directly reflect this problem. Chhattisgarh thin data. Only three years are usable for Chhattisgarh (2016, 2021, 2022). Its mean impunity score of 0.717 should be treated with caution. A leave-one-out check excluding Chhattisgarh is recommended alongside the main results.

ASER not published annually. The reading deficiency model uses ASER data from 2018 and 2022 only. The 2016, 2017, 2019, 2020, and 2021 cells are blank and are not imputed. The FE model for reading effectively identifies a single within-state change across a four-year gap, after removing the year fixed effect. This is the most significant structural limitation on the reading result.

COVID dropout zeros. Dropout data for 2020-21 are zero or near-zero for most states due to school closures and administrative disruption. These are not genuine zero dropout rates. The robustness check excluding these years does not resolve the wrong-sign result for dropout, indicating that floor effects in the dropout series are a secondary contributor.

GER above 100. Five states report a GER exceeding 100. These values are not data errors but complicate regression interpretation. Net Enrolment Ratio would be a cleaner outcome variable for future work.

Reverse causality. Poor education outcomes may contribute to corruption by producing populations with weaker capacity to monitor and sanction public officials. The FE design partially addresses this by exploiting within-state variation over time, but full resolution requires an instrumental variable that is not developed here.

Static urbanisation. Census 2011 urbanisation figures are applied across all seven study years. India's urbanisation shifted during this period. In OLS specifications, this introduces measurement error; in the FE specification, the time-invariant urbanisation variable is absorbed by state fixed effects and does not affect the main estimates.

Telangana urbanisation approximation. Telangana's urbanisation figure (38.9%) is approximated from Andhra Pradesh Census 2011 data adjusted for the districts constituting Telangana at the time of bifurcation in 2014. This is standard practice for post-bifurcation analysis but introduces some imprecision.

Composite index not constructable. The original design intended a composite corruption index combining NCRB, Lokayukta, CAG, RTI, and TI data. Four of five intended components were unavailable as consistent panel variables. A single noisy proxy attenuates coefficient estimates toward zero relative to what a well-measured composite would produce; the weak statistical significance of the results is therefore likely a lower bound on the true relationship rather than an upper bound.

Difference-in-Differences not executed. The planned quasi-experimental design was infeasible within the study window. This is stated as a limitation rather than omitted from the paper.

Small N. Fifteen states is a small cross-section for a two-way fixed effects estimator. Approximately 20 degrees of freedom, 14 from state fixed effects and 6 from year fixed effects, are consumed relative to pooled OLS, from a panel that begins at 105 observations and loses more to missing data. Inferences from this sample should not be generalised without further validation on a larger state panel.

10. Conclusion

This paper examined whether prosecutorial impunity under the Prevention of Corruption Act predicts elementary education outcomes across 15 Indian states from 2016 to 2022. The evidence points in the expected direction, but the signal is weak and one result has the wrong sign entirely. Both findings are reported honestly rather than one of them being buried.

The reading deficiency result, 6.3 percentage points of additional Class 5 students unable to read at grade level per unit increase in the impunity score, is the most informative finding. The direction is consistent with the theoretical channels of teacher absenteeism and fund diversion. The coefficient is not statistically significant at conventional thresholds. The opportunity cost exercise gives this coefficient empirical content: applied to verified UDISE+ 2021-22 elementary

enrollment figures for Uttar Pradesh, Bihar, and Rajasthan, it implies an order of magnitude of just over 1 million additional students who could be at grade-level reading if those states matched Kerala's prosecutorial accountability. That figure is the associational implication of the regression estimate, translated into headcount terms, and not a causal prediction.

The dropout result is the least informative. A wrong-sign coefficient that persists after excluding COVID-contaminated years reflects a data quality problem that UDISE+ dropout series shares in the 2018-2023 window: floor effects in well-performing states and administrative zeros in COVID years leave insufficient within-state variation for the FE estimator to recover a positive coefficient even if a genuine positive relationship exists. Reporting this honestly is a service to researchers who might otherwise treat the dropout series as a reliable panel outcome variable in similar designs.

The descriptive evidence is, in some respects, the most legible element of the paper. The rank ordering of states by dropout rate corresponds closely enough to the rank ordering by impunity score, with UP and Jharkhand at the top and Kerala and Karnataka at the bottom, to be informative on its own, even without a causal interpretation. Cross-sectional patterns are not evidence of causation, but they establish that the hypothesis is at minimum consistent with the data rather than contradicted by it.

Three policy implications follow from the evidence, though each is stated at the level of confidence the data actually support. The most direct is institutional strengthening of state Lokayuktas. States like Bihar, which lack a functional anti-corruption oversight institution, and West Bengal, where the Lokayukta exists in statute but operates under significant jurisdictional restrictions, sit at the high end of the impunity distribution, and the study period contains no Lokayukta reform events simply because none occurred. Whether reform of these institutions would causally reduce corruption is a question that needs a longer time series and an identification strategy this paper could not execute, but the pattern in the data still makes it a reasonable policy priority. A second lever is extending the GeM (Government e-Marketplace) procurement platform to cover school construction and materials supply contracts, which would shrink the discretionary administrative space where procurement corruption operates and address the fund diversion channel directly. Third, moving ASER to annual rather than biennial surveys would substantially improve the statistical power of any reading deficiency model built on within-state variation, and would make the design used here considerably more useful for researchers who take it further.

Research that is honest about what it cannot show is more useful than research that projects false certainty, especially when the stakes involve the educational trajectories of hundreds of millions of children, and although the findings reported here are modest, the pattern underlying them looks large enough to be worth taking seriously.

Works Cited

Bardhan, P. (1997). Corruption and development: A review of issues. Journal of Economic Literature, 35(3), 1320-1346.

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. Journal of Human Resources, 50(2), 317-372. https://doi.org/10.3368/jhr.50.2.317

Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K., & Rogers, F. H. (2006). Missing in action: Teacher and health worker absence in developing countries. Journal of Economic Perspectives, 20(1), 91-116. https://doi.org/10.1257/089533006776526058

Gupta, S., Davoodi, H., & Tiongson, E. (2000). Corruption and the provision of health care and education services (IMF Working Paper No. WP/00/116). International Monetary Fund.

Mauro, P. (1995). Corruption and growth. Quarterly Journal of Economics, 110(3), 681-712. https://doi.org/10.2307/2946696

Ministry of Finance, Government of India. (2017-2023). Economic Survey of India (Annual editions; education expenditure as a percentage of GDP, Social Services chapter). Department of Economic Affairs. https://www.indiabudget.gov.in

Muralidharan, K., Das, J., Holla, A., & Mohpal, A. (2017). The fiscal cost of weak governance: Evidence from teacher absence in India. Journal of Public Economics, 145, 116-135. https://doi.org/10.1016/j.jpubeco.2016.11.005

National Crime Records Bureau. (2016-2022). Crime in India, Volume II (Annual reports for data years 2016-2022). Ministry of Home Affairs, Government of India. https://ncrb.gov.in

Office of the Registrar General & Census Commissioner, India. (2011). Census of India 2011: Primary census abstract data highlights (DDW-PCA0000). Ministry of Home Affairs, Government of India. https://censusindia.gov.in

Pratham Education Foundation. (2018). Annual status of education report (rural) 2018. ASER Centre.

Pratham Education Foundation. (2022). Annual status of education report (rural) 2022. ASER Centre.

Reinikka, R., & Svensson, J. (2004). Local capture: Evidence from a central government transfer program in Uganda. Quarterly Journal of Economics, 119(2), 679-705. https://doi.org/10.1162/0033553041382120

Reserve Bank of India. (2016-2022). Handbook of statistics on Indian economy (Annual editions, state-wise NSDP tables). Reserve Bank of India. https://rbi.org.in

Unified District Information System for Education Plus. (2018-2022). Flash statistics (Annual reports, 2018-19 to 2022-23). Ministry of Education, Government of India. https://udiseplus.gov.in

Appendix A: Replication Code

The following R code replicates the main two-way fixed effects estimates. Data must be loaded from my_data.csv using the column names specified in the text. The working directory should be

set to the folder containing the data file. The script renames the pupil-teacher ratio column to a clean identifier on load and enters per capita NSDP in log form to match the specification described in Section 5.1, where ndsp(i,t) is defined as the log of per capita NSDP.

```
library(plm)
library(lmtest)
library(sandwich)

# Load data and create panel frame
my_data <- read.csv("my_data.csv")

# Clean the pupil-teacher ratio column name (R mangles spaces and brackets
# from the original spreadsheet header into dots on import)
names(my_data)[names(my_data) == "ptr..Primary.only."] <- "ptr"
pdata <- pdata.frame(my_data, index = c("State", "Year"))

# Main FE model: Reading deficiency
fe_reading <- plm(reading ~ ncrb + log(ndsp) + ptr,
  data = pdata, model = "within", effect = "twoways")
coeftest(fe_reading, vcov = vcovHC(fe_reading, type = "HC1", cluster = "group"))

# Main FE model: Dropout rate
fe_dropout <- plm(dropout ~ ncrb + log(ndsp) + ptr,
  data = pdata, model = "within", effect = "twoways")
coeftest(fe_dropout, vcov = vcovHC(fe_dropout, type = "HC1", cluster = "group"))

# Main FE model: Gross Enrolment Ratio
fe_ger <- plm(ger ~ ncrb + log(ndsp) + ptr,
  data = pdata, model = "within", effect = "twoways")
coeftest(fe_ger, vcov = vcovHC(fe_ger, type = "HC1", cluster = "group"))

# Robustness: COVID year exclusion (dropout)
data_nc <- my_data[!(my_data$Year %in% c(2020, 2021)), ]
pdata_nc <- pdata.frame(data_nc, index = c("State", "Year"))
fe_drop_nc <- plm(dropout ~ ncrb + log(ndsp) + ptr,
  data = pdata_nc, model = "within", effect = "twoways")
coeftest(fe_drop_nc, vcov = vcovHC(fe_drop_nc, type = "HC1", cluster = "group"))

# Robustness: Leave-one-out for reading
states <- unique(my_data$State)
loo_coefficients <- sapply(states, function(s) {
  temp <- pdata.frame(my_data[my_data$State != s, ], index = c("State", "Year"))
  m <- plm(reading ~ ncrb + log(ndsp) + ptr,
    data = temp, model = "within", effect = "twoways")
  coef(m)["ncrb"]
})
print(data.frame(State_dropped = states, ncrb_coefficient = round(loo_coefficients,
3)))

# Robustness: Lagged impunity score
# plm:: is specified explicitly because base R's lag() and dplyr's lag()
# are not panel-aware; only plm's method respects the State-Year index
# structure of the pdata.frame, so an explicit namespace call protects
# this line if dplyr or another masking package is loaded elsewhere in
# the same session.
fe_reading_lag <- plm(reading ~ plm::lag(ncrb, 1) + log(ndsp) + ptr,
  data = pdata, model = "within", effect = "twoways")
coeftest(fe_reading_lag, vcov = vcovHC(fe_reading_lag, type = "HC1", cluster =
"group"))
```

Appendix B: Notes on Figure Construction

Figure 1 should be a scatter plot of mean NCRB impunity score (x-axis, range 0.45 to 1.00) against mean dropout rate (y-axis, range 0 to 8), with state abbreviations as point labels. State points should be coloured by tier (red for high, grey for middle, blue for low). A least-squares fitted line through all points will show the positive cross-sectional association. Madhya Pradesh should be labelled explicitly as the outlier below the fitted line, with a note explaining the ACB artifact. West Bengal sits precisely at (1.000, 4.16) and should also be labelled as the extreme case.

Figure 2 should be a scatter plot of mean NCRB impunity score (x-axis) against mean GER (y-axis, range 85 to 110), with state labels. Points with GER above 100 should be marked with a hollow circle to signal the measurement concern. The fitted line will show a weakly negative association, though the relationship is noisier than for dropout.

Figure 3 should use the 30 state-year observations available for the reading model (15 states, ASER 2018 and 2022). The x-axis is the NCRB impunity score for the corresponding year; the y-axis is the state reading deficiency rate. Points should be coloured by year (one colour for 2018, another for 2022). A single fitted line through all 30 pooled observations will show the positive raw association before fixed effects are applied.

Figure 4 should be a horizontal bar chart, ordered from lowest to highest mean NCRB score, showing all 15 states. Bars should be shaded by tier. This figure makes the institutional variation in the sample immediately visible and allows readers to situate any state they are familiar with within the broader distribution. Note that Karnataka (0.864) and Maharashtra (0.949) appear in the upper half of the chart despite their low and middle tier classifications respectively, a point worth flagging in any presentation of this figure.

The Ethics and Effects of Nudge Theory in Business By Anun Amgalan

Abstract

From setting pre-selected tip amounts on digital payment screens to automatically enrolling new employees in retirement plans, businesses constantly shape choices to influence daily decisions without taking away freedom. This design strategy relies on choice architecture, defined in nudge theory as subtle changes to an environment that alter behavior predictably without forcing any options or changing financial rewards. As companies use nudges more often, the line between helping people and manipulating them creates serious ethical questions. While research often measures whether nudges work, few studies connect ethical principles with real-world results. This paper addresses that gap by reviewing secondary literature across marketing, retail, agriculture, and workplace management. To evaluate existing research, the paper categorizes five key conditions for an ethical nudge established in prior studies: interest alignment, transparency, accessible opt-out options, autonomy preservation, and equitable effect distribution. Ultimately, nudges are not inherently unethical; their morality depends on the business's intent, transparency, and the power dynamic between the nudger and nudgee.

Introduction

In everyday life, subtle environmental cues constantly direct human choices without explicit commands. A grocery store positions low-cost candies next to the register to prompt impulse spending at checkout, while a government sets organ donation registration as an automatic "opt-out" default to increase donor participation rates. These intentional arrangements illustrate choice architecture in action. Co-developed by Richard Thaler and Cass Sunstein, nudge theory formalizes this approach. A nudge is defined as "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives" (Thaler and Sunstein). Within this framework, the entity designing the choice environment is the nudger, while the individual whose decision is being influenced is the nudgee.

This literature review explores the use of nudge theory in business and how it can be ethically justified when considering its effects on consumer autonomy, trust, and well-being. However, this raises ethical and philosophical questions, such as whether nudges respect individual autonomy or manipulate people’s cognitive biases for profit, and how people should judge the morality of nudges when business interests may not align with the consumer’s well-being. The ethics of nudge theory is a point of contention, and while some researchers conclude that nudge theory is not inherently problematic, its usage remains controversial (Clavien).

Existing literature highlights a debate over whether influencing behavior without explicit consent can ever be morally justified. Scholars suggest that nudges are ethically acceptable when they pursue shared goals between the nudger and nudgee, align with consumer welfare, and account for human cognitive biases (Clavien; Weimer). Conversely, critics argue that many

"libertarian paternalist" nudges act as covert mandates that undermine true autonomy by taking advantage of habit through pre-set defaults and mandatory disclosures (Bubb and Pildes). As nudging expands into digital platforms and workplace management, researchers warn that power imbalances and automated systems risk creating "organized immaturity" or covert coercion (Häyry et al.; Kretzer and Maedche; Ruehle). Furthermore, empirical studies demonstrate that nudges yield inconsistent real-world outcomes, showing that depending on their design, nudges can either empower vulnerable populations or exploit cognitive biases for corporate profit (Medina; Mrkva et al.; Wang and Jiao).

Nudge theory is a powerful tool for influencing human behavior, but has complex ethical and philosophical implications because of its ability to both harm and benefit nudgees. The ethical acceptability of nudges ultimately depends on their intent, how much they align with consumer interests, and how much they preserve autonomy. Although many studies analyze the ethics or effectiveness of nudges separately, there is a lack of research combining ethical justification with evaluation to determine how nudges can be designed to promote autonomy and fairness, as well as a lack of research on how individuals respond to specific nudges, especially in business contexts. The next steps would be to do more research for more sources that could close this gap, such as collecting more information from peer-reviewed journal articles and online databases.

To address the research question, the paper is organized into four main sections. The Research Methodologies section details the selection of secondary peer-reviewed literature across retail, marketing, agriculture, and workplace management. The Results, Findings, Products section then evaluates the effectiveness of various nudges and introduces a five-condition framework to distinguish ethical nudges from manipulative ones. Next, the Conclusions and Analysis section interprets these findings through the lens of consumer autonomy and choice architecture. Finally, the Implications and Next Steps section highlights practical takeaways and identifies key areas for future research.

Research Methodologies

This study uses secondary, credible research articles from databases EBSCO and JSTOR. Furthermore, this study explores how nudges are used in the retail industry, such as through product placement, pricing cues, and default options that influence consumer purchasing behavior, and in the agriculture industry, where nudges may be used to guide decisions related to sustainability and resource use. In addition, the study also examines the use of nudge theory within management, focusing on how organizations influence employee behavior and decision-making, and within marketing, where nudges are commonly employed to shape consumer preferences and brand loyalty. This study looks for techniques, effects on nudgees, as well as the conditions that determine whether a nudge is ethically justified or not. This was done by examining the choice architecture of nudges, the responses it creates from nudgees, and ethical considerations such as transparency and respect for autonomy. Finally, the goal for this process was for the information to be more effective in identifying common themes and changes

in perspectives in how businesses and other beings use nudge theory and the effects on nudgees. The procedure for collecting evidence was searching for key terms such as “nudge theory”, “experiment”, “behavioral economics”, and “choice architecture” to find relevant peer-reviewed journal articles.

Results, Findings, Products

Most of the ethical debate surrounding nudge theory focuses on whether influencing behavior without explicit consent can ever be morally justified. Christine Clavien argues that nudges are not inherently unethical so long as they meet specific moral conditions to be considered acceptable (Clavien). According to her framework, a nudge is ethically justified when it pursues goals shared between the nudger and the nudgee, and when it aligns with important moral principles such as autonomy and consumer welfare. Similarly, David Weimer, a professor of political economy, states that evaluating nudges requires acknowledging that consumers often make inconsistent or irrational choices, which complicates how policymakers can determine the true benefits of using nudges (Weimer). In contrast, Ryan Bubb and Richard H. Pildes are more critical, suggesting that many “libertarian paternalist” nudges only maintain the appearance of freedom while functioning as covert mandates (Bubb and Pildes). Thaler defines libertarian paternalism as the idea that it is possible for institutions to shape behavior while preserving the freedom of choice (Thaler and Sunstein). Bubb and Pildes argue that defaults, automatic pre-selected settings, and disclosures, mandatory information releases, can limit real choice by taking advantage of people’s habits and biases, which undermines true autonomy and does not preserve it. These three perspectives reveal that the ethical justification for nudging depends on original intent and the preservation of genuine autonomy, not manipulation disguised as guidance.

As nudging expands into new environments, such as digital ecosystems and the workplace, ethical concerns expand into questions of power and control. Rebecca Ruehle explores how digital nudges in the workplace, such as automated reminders and prompts, can unintentionally create “organized immaturity”, the idea that systematic nudging may lower employee capacity for independent decision-making (Ruehle). Her framework of justification proposes that organizations should not only justify the purpose of nudges but also allow employees and nudgees to challenge them, creating what she calls meta-autonomy. Similar concerns are raised by Martin Kretzer and Alexander Maedche, who examine social nudges in enterprises that subtly move employees through AI-driven recommendation agents, often without their knowing (Kretzer and Maedche). These interventions, while efficient, blur the line between guidance and manipulation and raise ethical concerns. Matti Häyry extends this critique to public contexts, stating that even well-intentioned nudges can become coercion when individuals are subject to unaccountable and nontransparent decision-making systems (Häyry et al.). Across these studies is the common theme that differences in power worsen ethical risks. When people are unable to detect or resist the nudges influencing their behavior, the ethical justification of nudging is more necessary.

Research also reveals that the effects of nudges are not the same across populations. Kellen Mrkva demonstrated that low-socioeconomic-status (SES) consumers are more strongly influenced by nudges than those with more knowledge or resources (Mrkva et al.). When designed well, these "good nudges" can help reduce decision-making disparities by improving outcomes for the disadvantaged and SES consumers. However, "bad nudges" that prioritize corporate profit tend to worsen inequality. Other researchers, such as Shu Wang and Catherine K. Jiao, explore how AI-based "smart nudges" can be made fairer and more transparent by adapting to individual contexts such as tipping in restaurants (Shu and Jiao). In their study examining point-of-sale payment screens at a bubble tea store, Wang and Jiao tested an intelligent reasoning model that adjusted suggested tipping percentages based on factors like order size and customer interaction. They found that custom, context-aware tip suggestions significantly increased tipping amounts while maintaining positive customer sentiment. In the context of ethics, this demonstrates that when digital nudges align with social norms rather than applying one-size-fits-all defaults, they can boost business revenues while preserving consumer trust. Still, studies like Medina's study on credit card interventions show that even well-intentioned nudges can produce unintended side effects when they disrupt consumer behaviors in unpredictable ways. Medina investigated this by analyzing a field experiment where credit card users received automated text and email notifications prompting them to pay their credit card balances. While the reminders successfully prompted consumers to pay off credit card debts, Medina discovered that many cardholders offset these payments by drawing down their checking account liquidity, ultimately incurring higher overdraft fees elsewhere. In context, these findings reveal that isolated nudges can backfire. By shifting attention to one financial task, choice architecture can inadvertently create other financial harm. These disparate real-world outcomes emphasize that evaluating a nudge requires looking beyond its success to assess who it benefits and who it potentially harms.

Table 1: Effectiveness and Impact of Nudges

Research Study	Nudging method	Group	Results
Calabro	Packaging	Vending machine users	No impact
Wang, Jiao	Smart-tipping	Bubble tea store consumers	Positive impact
Mrkva	Consumer decision	Low SES consumers	Positive impact
Medina	Notification messages	Credit card users	Negative impact
Chabé-Ferret	Informational flyers	Farmers	Negative impact

Analyzed sources show a variety of effectiveness in nudges

An examination of studies across retail, digital marketing, and workplace settings reveals a wide spectrum of nudge effectiveness, ranging from highly impactful interventions to complete failures. The first row of Table 1 evaluates whether altering visual architecture by adding healthy imagery on vending machine backgrounds shifts consumer choices, and the "no impact" result

demonstrates that the passive physical nudge failed to disrupt consumer habits (Calabro et al.). Ryan Calabro and his colleagues evaluated whether altering visual backgrounds on vending machines to feature healthy imagery would shift consumer purchases toward healthier beverages. Their study found that passive visual cues had no statistically significant impact on beverage selection, suggesting that subtle environmental nudges are often not enough to disrupt deeply ingrained consumer purchasing habits. The second row examines AI-based "smart tipping" screens that adjust tip suggestions, and the "positive impact" result shows that the digital nudge was effective (Wang and Jiao). The third row highlights the interaction between choice architecture and consumer knowledge levels, and "positive impact" means that nudges affect lower socioeconomic status consumers more strongly (Mrkva et al.). The fourth row outlines a controlled intervention using text and email notifications to prompt credit card payments, and "negative impact" is shown because the prompt altered immediate attention but failed to change overall financial situations (Medina). The fifth row measures the impact of informational flyers designed to persuade farmers to adopt biological pest control, and the "negative impact" is shown when the nudge backfired and led to counterproductive losses in overall production (Chabé-Ferret et al.). A field experiment by Sylvain Chabé-Ferret examined the effect of sending informational flyers to farmers to encourage the adoption of biological pest control. Instead of boosting adoption, the informational nudge backfired, causing farmers to misinterpret the risks and leading to counterproductive decisions that resulted in crop losses. Recognizing this variety in behavioral impact implies that behavioral efficacy alone cannot determine whether an intervention is acceptable, making a structured framework necessary to evaluate its moral validity.

Analysis of peer-reviewed literature shows that the use of nudge theory in business produces measurable behavioral changes, but whether it is acceptable ethically depends on intent, transparency, power balance, and outcomes. Across industries, nudges were most ethically defensible when they improved consumer welfare without significantly undermining autonomy, and most ethically controversial when they primarily served to increase corporate profit or operated without consumer awareness.

Table 2: Conditions Associated with Ethical vs. Unethical Nudges

Condition	Ethically Justified Nudges	Ethically Problematic Nudges
Alignment of interests	Benefits both business and consumer	Primarily benefits business
Transparency	User aware of influence	Hidden or deceptive
Ease of opting out	Simple, low cost	Difficult or costly
Impact on autonomy	Supports informed choice	Exploits biases or inertia
Distribution of effects	Reduces inequality	Worsens inequality

Nudges must follow all five conditions to be ethically justified

To see how the five ethical conditions listed in Table 2 apply, they can be mapped against reviewed studies. First, when looking at interest alignment, a nudge is only ethical if it benefits both the business and the customer. For instance, the smart-tipping screens evaluated by Wang and Jiao show a fragile balance, as while they help employees make more money, they risk exploiting the customer's social pressure to conform (Wang and Jiao). A complete failure of interest alignment is shown in the Chabé-Ferret study, where informational flyers ignored the financial situations of the farmers, leading to crop losses instead of helping them (Chabé-Ferret et al.). Second, transparency means that users must easily see when and how their choices are being guided. Calabro's healthy vending machine images were entirely open and honest, meaning they had no ethical risk of tricking people, even though the nudge did not successfully change consumer habits (Calabro et al.). In contrast, the automated business recommendation tools explored by Kretzer and Maedche run the risk of functioning as hidden agents that guide behavior without the user's conscious awareness (Kretzer and Maedche). Third, an ethical nudge must always offer a completely free and easy way to opt out. In Medina's credit card study, the opt-out cost is minimal because the customer can simply ignore a text message reminder (Medina). This is different from other deceptive corporate defaults, such as high pre-checked tip amounts or automatic subscription renewals, which rely on consumer laziness or forgetfulness. Fourth, a nudge must protect a person's true autonomy instead of exploiting human weaknesses. Bubb and Pildes argue that many corporate nudges only offer an illusion of freedom while secretly taking advantage of human habits (Bubb and Pildes). Finally, the distribution of effects looks at whether a nudge closes societal gaps or makes them worse. The data from Mrkva's study serves as an example, as because choice architecture has a stronger impact on low SES consumers, a helpful nudge can act as an equalizer (Mrkva et al.). However, if a business uses nudges to maximize profits, the financial harm falls on the same vulnerable groups.

These five conditions were repeatedly observed as indicators of ethicality in nudges. Nudges must be aligned with the interests of both parties, be transparent, be opt-out accessible, preserve autonomy, and distribute effects evenly across nudgees. Overall, the research indicates that nudge theory is not inherently unethical in business contexts. Its morality depends on how, why, and on whom it is used. Businesses that ignore these factors risk undermining consumer trust even when nudges achieve their intended outcomes.

Conclusions and Analysis

These findings indicate that nudge theory is not inherently unethical, but rather, its morality is determined by the context of its application, specifically regarding the intent of the nudger and the transparency of the nudge. Ethical justification comes from five core conditions: the alignment of interests, transparency, the ease of opting out, the preservation of autonomy, and the fair distribution of effects. However, nudges are not a universal solution. Their impact ranges from significant (AI smart-tipping) to negligible (vending machine packaging) or even negative (informational flyers for farmers). AI-driven "smart-tipping" interventions showed a significant positive impact on consumer behavior (Shu and Jiao). Simple visual cues, such as vending

machine packaging or credit card notification messages, often yielded minimal to no behavioral change (Calabro et al.; Medina). In specific contexts, such as informational flyers for farmers, nudges can even "backfire," leading to negative outcomes such as the loss of crop production (Chabé-Ferret et al.). While nudges can improve outcomes for disadvantaged consumers by reducing decision-making disparities, "bad nudges" that prioritize corporate profit over consumer welfare tend to exploit biases and worsen inequality. Evidence supports the critique that many nudges, particularly nudges using default options, may function as "covert mandates" that maintain an appearance of freedom while taking advantage of human habits (Bubb and Pildes). Because nudges take advantage of habits and cognitive biases, they can undermine true autonomy even when the "choice" remains available. Furthermore, the data suggests that "good nudges" can act as an equalizer. For example, well-designed choice architecture can compensate for low consumer knowledge among low-socioeconomic-status individuals, helping to close decision-making gaps (Mrkva et al.). Ultimately, what was discovered is that for a nudge to be truly ethical, it must provide a "shared preference" goal where the nudgee's welfare is a primary consideration, not a byproduct of a profit-driven architecture (Clavien).

Implications and Next Steps

Based on the reviewed literature, the findings of this study place nudge theory into a broader framework of libertarian paternalism, highlighting the balance between guiding behavior and preserving freedom. These findings are important for business leaders, digital designers, and policymakers because they demonstrate that unethical nudging risks undermining consumer trust, even if the nudge achieves its well-intended goal. In the workplace, the shift toward AI-driven "social nudges" implies a risk of organized immaturity, where employees lose the capacity for independent decision-making due to constant, unaccountable automated prompts. This research aligns with the foundational definitions provided by Thaler regarding choice architecture, but leans toward the critical perspectives of other researchers, such as Clavien and Ruehle, regarding the risks of manipulation. It disagrees with the idea that all nontransparent nudges are harmless, suggesting instead that power imbalances worsen ethical risks. Additionally, this study noted a current lack of research that combines ethical justification with direct evaluation of how individuals respond to specific nudges in a retail or business context. To close this gap, future research should conduct primary field experiments that measure both the behavioral outcome and the nudgee's perceived level of autonomy, investigate meta-autonomy, allowing nudgees (such as employees) to challenge and participate in the design of the nudges they are subjected to, and explore more deeply how digital and "smart" nudges can be made transparent to ensure fairness.

Works Cited

Bubb, Ryan, and Richard H. Pildes. "How Behavioral Economics Trims its Sails and Why." Harvard Law Review, vol. 127, no. 6, 2014, pp. 1593–678. JSTOR, http://www.jstor.org/stable/23741986

Calabro, Ryan, et al. "Vending Machine Backgrounds: Nudging Healthier Beverage Choices." Current Psychology, vol. 43, no. 2, Feb. 2023, pp. 1733–42, https://doi.org/10.1007/s12144-023-04420-8.

Chabé-Ferret, et al. "When Nudges Backfire : Evidence from a Randomised Field Experiment to Boost Biological Pest Control." Annals of Economics and Statistics, no. 156, 2024, p. 9, https://doi.org/10.2307/48804180.

Clavien, Christine. "Ethics of Nudges: A General Framework with a Focus on Shared Preference Justifications." Journal of Moral Education, vol. 47, no. 3, Jan. 2018, pp. 366–82, https://doi.org/10.1080/03057240.2017.1408577.

Häyry, Matti, et al. "Let the Chips Fall! Public Nudging Arrangements, Coercion, and the Role of Independent Shopkeepers." Society, vol. 60, no. 4, Aug. 2023, pp. 525–38, https://doi.org/10.1007/s12115-023-00844-x.

Kretzer, Martin, and Alexander Maedche. "Designing Social Nudges for Enterprise Recommendation Agents: An Investigation in the Business Intelligence Systems Context." Journal of the Association for Information Systems, vol. 19, no. 12, Dec. 2018, pp. 1145–86, https://doi.org/10.17705/1jais.00523.

Medina, Paolina C. "Side Effects of Nudging: Evidence from a Randomized Intervention in the Credit Card Market." Review of Financial Studies, vol. 34, no. 5, May 2021, pp. 2580–607, https://doi.org/10.1093/rfs/hhaa108.

Mrkva, Kellen, et al. "Do Nudges Reduce Disparities? Choice Architecture Compensates for Low Consumer Knowledge." Journal of Marketing, vol. 85, no. 4, July 2021, pp. 67–84, https://doi.org/10.1177/0022242921993186.

Ruehle, Rebecca C. "The Moral Permissibility of Digital Nudging in the Workplace: Reconciling Justification and Legitimation." Business Ethics Quarterly, vol. 33, no. 3, July 2023, pp. 502–31, https://doi.org/10.1017/beq.2023.4.

Thaler, Richard H., and Cass R. Sunstein. Nudge: Improving Decisions About Health, Wealth, and Happiness. Yale University Press, 2008.

Wang, Shu, and Catherine K. Jiao. "Leveraging Behavioural Economics in Smart Nudge Design through Data-Driven Prospect-Theoretic Modelling and Context-Aware Intelligent Reasoning: Application to Smart Tip Nudging." Journal of Engineering Design, vol. 33, no. 11, Nov. 2022, pp. 896–918, https://doi.org/10.1080/09544828.2022.2150448.

Weimer, David L. "When Are Nudges Desirable? Benefit Validity When Preferences Are Not Consistently Revealed." Public Administration Review, vol. 80, no. 1, Jan. 2020, pp. 118–26, https://doi.org/10.1111/puar.13114.

IBOVESPA and Presidential Elections: A Performance-Based Forecasting Approach
By Bernardo Dal Molin and Marco Pizzato

Abstract

Economic voting theory holds that national economic performance shapes support for incumbent presidents, and forecasting models built on this premise have been tested extensively in the United States and other established democracies. Whether such models transfer to volatile emerging-market democracies remains unclear. This study asks to what extent the IBOVESPA, the benchmark index of Brazil's B3 exchange, can forecast Brazilian presidential election outcomes. Using a purposive sample of the four elections between 1998 and 2022 in which a sitting president sought reelection, the study measured the index's cumulative percentage change across two three-month pre-election windows (30 June to 30 September, and 31 July to 31 October) and coded each contest as an incumbent win or loss. The hypothesis—that positive index growth would coincide with incumbent reelection—was supported in only one of four cases in each window. Pearson correlation coefficients were strongly negative (−0.98 for Period 1 and −0.85 for Period 2), the opposite of the predicted direction, although a sample of four elections is far too small to support statistical inference. The findings suggest that investor confidence does not translate directly into voter confidence in Brazil, and that forecasting models imported from other national contexts require country-specific adaptation and the inclusion of non-economic variables.

Introduction

"The vote is determined almost entirely by the amount of prosperity the incumbent party delivers" (Erikson and Wlezien). For each presidential election worldwide, extensive polling aims to predict outcomes before the actual voting and counting of ballots. Various models have been developed and tested across different contexts to forecast elections and bring greater social and economic certainty; however, despite extensive polling, the uncertainty of presidential election outcomes remains an inherent characteristic of democracies. For example, the 2024 United States presidential election between Kamala Harris and Donald Trump exemplified how this electoral period generates uncertainty, impacting the country's stability and increasing economic volatility. Yet, while multiple factors are associated with constituent votes, one of the primary indicators of voter choice is seen to have ties to the national economic performance of the incumbent if they seek reelection (Markus). This problem extends to contexts like Brazil, where uncertainty worsens already great economic volatility, corruption, and high poverty rates. Hence, addressing and evaluating the applicability of the United States' performance-based election theories in Brazil could provide a stronger conclusion of the effectiveness of the method's application in foreign countries, adding to the existing body of papers evaluating this relationship in other national contexts (Wink). Doing so may increase presidential election forecasting ability in Brazil, adding to the country's limited economic research body and potentially improving economic measurements during election periods.

This research paper analyzes IBOVESPA, the benchmark stock index of the São Paulo Stock Exchange that reflects the performance of the most actively traded stocks in Brazil, and its ability to forecast Brazilian presidential election outcomes. This question is relevant to the fields of political science and economics. Considering Brazil's high market volatility, corruption, and unique macroeconomic conditions, this research is significant because it addresses the degree to which this relationship is present and analyzes its efficacy as an election outcome predictor. Additionally, its findings could potentially reduce election uncertainty and stabilize economic conditions, especially during the months before upcoming elections.

Literature Review

When analyzing national economic performance and its impact on voting behavior, evidence suggests that positive economic conditions increase incumbent support, while economic downturns do the opposite. Becher, a political scientist at IE University, and Donnelly, an associate professor at the University of Toronto, provide a detailed analysis of how voters' perspectives of the economy translate into electoral support. They argue that voters associate economic conditions with the current government, and their analysis highlights how the people's economic perspectives are highly influential to voting, even outweighing other areas of consideration (Becher and Donnelly). They also note that, in multi-party systems, economic evaluations can drive voters from incumbents to alternative options, as supported by their findings that economic voting is a tool by which voters hold governments accountable. Yet the relationship between economic conditions and electoral outcomes is not uniformly strong, since political context may weaken the predictive power of the economic indicators available to us.

Guntermann et al., led by political scientist and data scientist Eric Guntermann of the University of California, Berkeley, investigated this relationship, analyzing different systems to explore accountability in coalition governments and one-party systems. In coalition governments, for instance, shared power between parties complicates voters' ability to view economic performance as specifically the incumbent's success, causing less straightforward economic voting. Meanwhile, in one-party systems, which have limited political competition, the correlation between economic performance and voter behavior tends to be weak. Because of this, Guntermann et al. suggest that election forecasting models should consider political strategies, as they may fluctuate accordingly. Still, it is found that, generally, economic performance impacts voter support for incumbents.

Examining the effect of economic indicators on election outcomes, research supports the idea that specific economic conditions, including voters' personal finances, should significantly shape voting behavior. Markus, a political scientist from the University of Michigan, says individuals who obtained personal economic gains are more likely to vote for the incumbent, according to the connection between the incumbent's policies and their income. This "pocketbook voting" suggests that voters value their financial well-being more and give credit to the incumbent for any positive changes during these periods. This behavior demonstrates the relationship between the economic context that generates major satisfaction and, thus, support

for the incumbent. Still, broader economic conditions are more likely to shape voting behavior as elections approach. Erikson, a specialist in American political behavior and elections, and Wlezien, a professor of government at the University of Texas at Austin, suggest that voters tend to heavily consider economic trends near the upcoming election months, even prioritizing that over their financial well-being (Erikson and Wlezien). These findings showcase a nuanced connection between specific and broad economic indicators of a country and its constituents. Even though personal finance gains can increase support, Erikson and Wlezien's analysis supports the idea that national economic health is the more important factor when evaluating an incumbent's likelihood of reelection.

The evolution of forecasting models has grown, aiming to predict outcomes with greater efficacy each time by incorporating economic and non-economic variables to predict voters' complex choices. For a long time, these forecasting models have used primary economic indicators like GDP growth and unemployment rates, assuming that positive economic performance would benefit the incumbent and vice versa. This approach, however, presents limitations in modern elections, as changes in voters' feelings and ideological division may make the model seem less effective in today's politics. As Wink highlights, economic forecasting models tend to ignore quick changes in constituents' feelings towards a candidate and biases that affect people's understanding of economic data. So, researchers adapted these models to include non-economic variables like political polarization, approval ratings, and crisis events, which tend to affect opinion as much as economic factors. These adjustments allow models to capture various variables that could affect voter decisions and attempt to forecast outcomes more effectively, rather than only considering economic factors. This reliance on economic indicators alone has also been challenged, specifically because voters may ignore economic conditions to match their personal ideology (Bisgaard). Heidemanns et al. assess how modern models minimize these errors by considering multiple variables, including media coverage and demographic shifts. This well-rounded approach generates a more accurate and complete prediction; however, it is still unsuccessful in representing short-term shifts in belief. In this way, even though modern models attempt to forecast election outcomes as accurately as possible, traditional economic forecasting models still provide the basis for election forecasting, as they are more straightforward but still highly effective.

Gap in Research

Although there is existing research analyzing the ability of macroeconomic trends to forecast election outcomes in countries such as the United States, there is no specific research addressing such trends in Brazil. Given the Brazilian stock market's higher volatility and different macroeconomic scenarios (i.e., interest rates), it is valuable to see whether this relation applies to a country with a different social and economic background. Addressing this gap in research will also bring to light potential extraneous variables that may prove such a model untrue within its new context. Addressing these aspects, this study examines to what extent the IBOVESPA can serve as a predictive indicator for presidential election outcomes in Brazil.

Methodology

The aim of this research was to respond to the question of the extent to which IBOVESPA could effectively forecast presidential election outcomes in Brazil. It was hypothesized that if the index had a positive percent change three months before the election, the incumbent would succeed and be reelected. Meanwhile, if the index's growth were negative, the incumbent would lose reelection.

Approach and Design

The research used quantitative data to investigate the correlation between IBOVESPA's performance and election outcomes. While the research emphasized numerical analysis, the study also acknowledged the human phenomenon of election uncertainty and stock market volatility, which were relevant in influencing voter sentiment and investor behavior. The research utilized a mix of frameworks, including quantitative statistical analysis and qualitative investigation. While statistical methods identified correlations between the dependent and independent variables, the study's phenomenological approach led the discussion toward considering economic optimism, political stability, and market reactions. This ensured human behavior was considered alongside measurable outcomes, preventing potential bias while addressing the limitations of solely focusing on numerical data. Treating all data points with equal importance was crucial for the accuracy of this analysis, and by using quantitative findings with qualitative considerations, the study provided an insightful understanding of the relationship between stock market performance and electoral outcomes in Brazil.

Delimitations

As for delimitations, a time period delimitation was set to guarantee recent, effective, and viable data gathering and analysis. Because of societal and technological evolution, data from over three decades before this research was published might have provided unreliable conclusions. Furthermore, the inaccuracy and difficulty of accessing data earlier than the 1990s further justified this delimitation. Even though multiple economic indicators might have more closely appealed to individuals rather than the general population, the IBOVESPA was Brazil's primary stock market index and was widely referenced as a reflection of investor confidence, making it an appropriate measure for economic performance. To follow the literature analyzed and to effectively determine the method's efficacy or inefficacy, IBOVESPA's analysis was conducted only during the three months before each election. According to research and background information, this timeframe also captured the critical period when economic sentiment might have influenced voter behavior and election outcomes. Considering the research's boundaries, the study only included elections where the incumbent president was running for reelection; while instances where the incumbent did not seek reelection were not directly analyzed, they were considered in the paper's discussion, as they could indicate a link with economic performance. This was primarily due to the hypothesis specifically examining the relationship between economic performance and voter support for incumbents. To ensure its

clarity and measurability, all data were translated into binary results, being gathered from reliable sources such as B3's IBOVESPA performance statistics, the Biblioteca da Presidência da República for incumbent terms, and the Superior Electoral Court (TSE) for election results. Finally, the study only focused on secondary data from Brazil, excluding any non-economic factors, such as social movements, political scandals, or campaign strategies.

Sampling

This research utilized a purposive sampling strategy focusing on presidential elections from the last three decades in which the incumbent ran for reelection. This approach ensured the sampling was relevant to the research question, particularly the contexts in which the performance of the stock market index during the three months before elections could potentially predict election outcomes. As the analysis restricted data to an approximately thirty-year frame, the research included accurate information, also due to the stable period Brazil faced during this period. Even though this strategy limited the general applicability to elections outside of these specific conditions, it allowed the research to clearly obtain results that either supported or refuted the hypothesis while also providing four sample election cases to make meaningful analysis easier.

Data Collection Method

The first step in executing the content analysis was to acknowledge its different variables clearly. Specifically, they consisted of the explanatory variable (IBOVESPA's performance, measured by its percentage change in points from the last day of June to the end of September and from the last day of July to the end of October before each election), the outcome variable (election result, recorded as a win or loss for the incumbent), and the control variable (whether the incumbent ran for reelection). The data-gathering process started by accessing historical IBOVESPA monthly growth data via the B3 website and downloading data for the last day of June to September and July to October of election years 1998, 2002, 2006, 2010, 2014, 2018, and 2022. In a spreadsheet, data were organized into columns for year, month, percent growth for Period 1 (from the end of June through the end of September), percent growth for Period 2 (from the end of July through the end of October), and IBOVESPA's growth (binary: 1 = positive, 0 = negative). Meanwhile, candidate data were gathered by visiting the Biblioteca da Presidência da República to identify incumbents for each term and whether they ran for reelection, while cross-checking with the TSE to verify incumbents' participation in elections and record their outcomes. It was then inputted into a spreadsheet with columns for year, incumbent, incumbent seeking reelection (1 = yes, 0 = no), and election outcome (1 = incumbent elected, 0 = incumbent not elected, X = no incumbent running).

Academic sources used to justify the procedure mentioned above included "The Impact of the Economy on Presidential Elections throughout US History" (Guntermann et al.), written in part by Eric Guntermann, a political and data scientist at the University of California, Berkeley. The study specifically applied a categorical approach to analyzing the influence of economic

indicators on electoral outcomes, similar to this research's application of IBOVESPA's performance as an independent variable and election results as the dependent variable. Additionally, "The Impact of Personal and National Economic Conditions on Presidential Voting, 1956–1988" (Markus), written by Gregory Markus, professor in the Department of Political Science at the University of Michigan, demonstrated the use of a binary approach to labeling economic conditions—either favorable or unfavorable—and their correlation with incumbent reelection. Finally, Robert Erikson's "Economic Conditions and the Presidential Vote" offered insight into the role of a logistic regression model in analyzing the results of behavior in elections.

Pilot Study

In the pilot study, IBOVESPA's percent change from June to September and July to October, measured in positive or negative results, was analyzed, in addition to whether the incumbent was running for reelection and whether he or she succeeded in being reelected. The pilot study exhibited a positive correlation between IBOVESPA's positive growth during the period analyzed and the president's success in being reelected. As for data collection, the procedure went as planned: by analyzing the total IBOVESPA points between the 30th of June and the 30th of September, and the 30th of July to the 30th of October, a percentage growth was calculated. After inputting the percent growth, the binary values of positive or negative growth, an incumbent running for reelection or not, and the incumbent being reelected, a correlation was identified. Although there were a few issues, the initial problem arose because B3, the secondary source primarily being analyzed, does not calculate IBOVESPA's monthly yield. However, since B3 provided its points for each day of the month historically, its percent change could be easily calculated. Additionally, elections were held on different dates every four years, so it was difficult to collect data from a fixed date within the months specified. Therefore, although the dates of the period varied slightly, the periods were the same length to ensure accurate consideration.

Limitations

It was important to consider the study's limitations, beginning with its small sample size, limited to elections spread across a timeframe of thirty years. Particularly, from the seven election samples obtained, three could not be used because the incumbent did not run for reelection in that specific year, which meant only four samples were analyzed. This was a limitation because, although it aimed to guarantee accurate results that could be applied to today's modern society, it might not have been robust or predictive for future elections due to the limited dataset. There were also significant confounding variables where IBOVESPA's performance might have correlated with factors other than investor sentiment, including inflation, unemployment, GDP growth, political events, and international market trends, which were not considered in the analysis. While the study hypothesis also claimed a causal relationship between its dependent and independent variables, it was important to consider the

complex mix of economic, social, and political factors, as the stock market might have worked as an indicator but not a determining factor. Finally, as the study focused on IBOVESPA's performance in the last three months before elections, it might not have captured longer-term trends or sudden events influencing both the market and voter sentiment—an aspect that, even though it was a delimitation of the research, could also have been overlooked.

Ethical Considerations

Multiple ethical considerations were taken into account, particularly addressing concerns regarding data accuracy, the value of conclusions that positively contributed to the current body of research, and justice when accurately representing findings. Respect for sources in this research was of great importance, as the secondary data, the basis of the study, originated from extremely credible sources, including the Brazilian government's electoral commission (TSE) and the IBOVESPA index archives. Transparency in data selection and analysis was also handled carefully, focusing on publicly available data to ensure it was fair and accessible to all. By examining research that consisted of verified historical data and outlining the methods, this approach ensured accountability and minimized ethical or misinterpretation risks. Overall, this research provided a benefit to society, both at the individual and national level, as increased stability might improve economic, social, and governmental concerns that might otherwise worsen the population's quality of life.

Results and Analysis

Results

The results of this study provide an analysis of IBOVESPA's performance during the periods before presidential elections and their correlation with election outcomes. Viewing the descriptive statistics, the mean IBOVESPA growth for Period 1, from June 30 to September 30, was 3.1585%, while for Period 2, from July 31 to October 31, the mean growth was slightly higher at 3.9699%. The standard deviation for Period 1 was 5.7646, a moderate level of variability in stock market performance, while in Period 2 it was slightly greater at 6.6368, indicating increased shifts. When viewing the range of IBOVESPA growth values (Fig. 1), Period 1 varied from −0.4948% to 11.6649%, while Period 2 varied from −2.1509% to 12.4775%, showing that the stock market had movements upwards and downwards during the months before elections. From the cases analyzed, the incumbent was reelected three out of four times, meaning there was a 75% reelection rate.

Year	Cumulative IBOVESPA Growth (%) (Period 1: 30/06–30/09)	IBOVESPA Growth Period 1 (Binary)	Cumulative IBOVESPA Growth (%) (Period 2: 31/07–31/10)	IBOVESPA Growth Period 2 (Binary)	Incumbent Ran for Reelection	Election Outcome
1998	−0.3188	0	−0.3418	0	1	1
2006	−0.4948	0	5.8949	1	1	1
2014	1.7826	1	−2.1509	0	1	1
2022	11.6649	1	12.4775	1	1	0

Fig 1: Cumulative IBOVESPA growth across both pre-election periods, incumbent candidacy, and election outcome for the four elections analyzed

The analysis of IBOVESPA growth during Period 1 (Fig. 2) exhibited inconsistencies, seen in 1998, when IBOVESPA experienced negative growth at −0.3188% yet the incumbent was reelected, and in 2006, when IBOVESPA's growth remained negative at −0.4948% but the incumbent was again reelected. Both findings contradict the hypothesis that negative market performance would lead to an incumbent loss. The 2014 election was the only instance in this period where IBOVESPA's performance aligned with the hypothesis, as the stock market exhibited a positive growth of 1.7826% and the incumbent was reelected. However, this exception was later also challenged by the year 2022, when, despite IBOVESPA reaching its highest growth of 11.6649%, the incumbent lost reelection, contradicting the hypothesis.

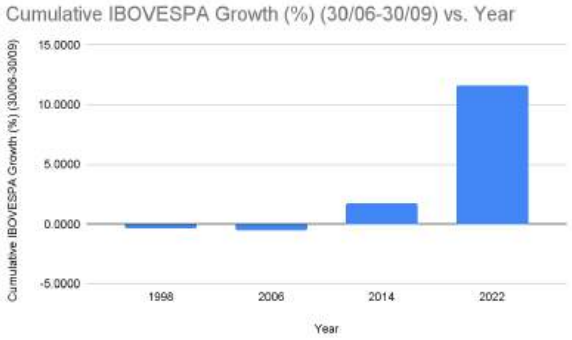


Fig 2: Cumulative IBOVESPA growth (%) for Period 1 (30/06–30/09) by election year

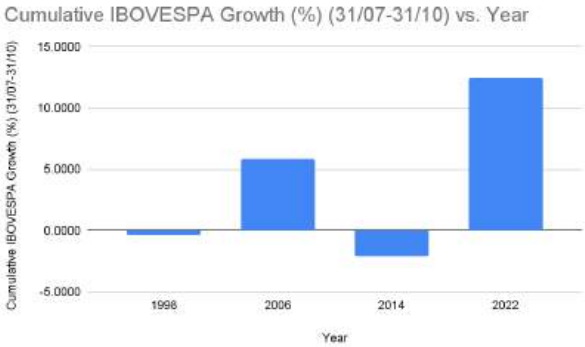


Fig 3: Cumulative IBOVESPA growth (%) for Period 2 (31/07–31/10) by election year

Similar trends are seen in Period 2 (Fig. 3). In 1998, IBOVESPA's growth remained negative at −0.3418%, yet the incumbent was reelected. The year 2006 was the exception, supporting the research's hypothesis, as IBOVESPA grew by 5.8949% and the incumbent won reelection. In 2014, however, despite IBOVESPA's negative growth at −2.1509%, the incumbent was still reelected, while in 2022, despite its most significant growth of 12.4775%, the incumbent still lost reelection. Again, the hypothesis was confirmed in only one of the four cases examined, suggesting that IBOVESPA's performance during this period does not provide a consistent indicator of election results.

Outliers in these findings mainly include the years 2022 and 2014, as they vary significantly from the relationship the hypothesis suggested. Specifically, in 2022, despite record growth rates in Period 1 and Period 2, the incumbent did not get reelected, possibly due to external political factors, including polarization, strong anti-incumbency sentiment, and other socio-political influences. Similarly, 2014 demonstrated another irregularity according to the hypothesis, where despite IBOVESPA's growth in Period 2 being negative at −2.1509%, the incumbent was reelected. It is important to note that Period 1 exhibited a 1.7826% growth, which could potentially suggest that economic optimism in earlier months of the year may have played a more important role in affecting voter sentiment, but the sample size is too small, and other results do not support this claim. Overall, these factors highlight the limitations of using IBOVESPA as the only predictor of election results and indicate that a broader set of economic, social, and political variables should be utilized to obtain a more effective forecasting model that generates more precise results.

Results Analysis and Interpretation

The research question of this study was, "To what extent can IBOVESPA effectively forecast presidential election outcomes in Brazil?" The hypothesis stated that if IBOVESPA experienced positive growth in the three months before the Brazilian presidential elections, the incumbent would be reelected, yet if the index showed negative growth, the incumbent would lose reelection. However, the results from the data collected do not consistently support the hypothesis.

Analyzing four Brazilian presidential elections in which the incumbent ran for reelection, Period 1 (June 30 to September 30) showcased only one example (2014) of alignment with the hypothesis, where there was positive IBOVESPA growth and an incumbent victory; meanwhile, Period 2 (July 31 to October 31) also exhibited only one case (2006) of alignment with the hypothesis (Fig. 4). It is important to highlight how, despite the same election outcome for both years, the periods that demonstrated positive and negative growth before the election differed. On the other hand, 1998 recorded negative growth for both periods and the incumbent still won reelection, while in 2022 the index recorded its highest positive growth in both periods, yet the incumbent was not reelected. These findings indicate that IBOVESPA's performance does not consistently predict election outcomes, and despite certain elections exhibiting correlations that align with the hypothesis, it did so inconsistently across the two periods.

The inconsistencies observed suggest that the use of IBOVESPA alone is not an accurate predictor of Brazilian presidential election outcomes and support the idea that external factors might be influencing areas outside stock market performance, as exhibited in 2022, for instance. These findings also align with research that mentions potential limitations when utilizing economic indicators as the only variable indicating voter behavior and sentiment, especially in contexts with many confounding variables.

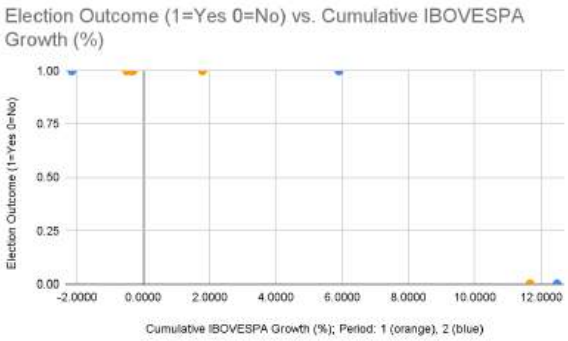


Fig 4: Election outcome (1 = incumbent reelected, 0 = incumbent not reelected) plotted against cumulative IBOVESPA growth (%) for Period 1 (orange) and Period 2 (blue)

A Pearson correlation coefficient was calculated for both periods to quantify the relationship between IBOVESPA growth and electoral outcomes. Both exhibit a strong negative correlation, as viewed by Period 1, June 30 to September 30, at −0.98 (Fig. 5), and Period 2, July 31 to October 31, at −0.85 (Fig. 6).

These coefficients suggest that positive IBOVESPA growth was associated with a higher likelihood of the incumbent losing reelection, rather than the opposite, initially suggested by the research's hypothesis. This is because a correlation coefficient closer to −1 indicates an inverse relationship between the variables analyzed. It must be stressed, however, that with only four observations these coefficients are highly unstable and should be read as descriptive of the sample rather than as evidence of a generalizable inverse relationship.

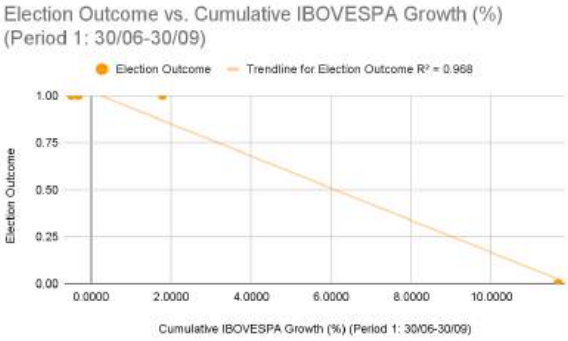


Fig 5: Election outcome versus cumulative IBOVESPA growth (%) for Period 1 (30/06–30/09), with trendline (R^2 = 0.968)

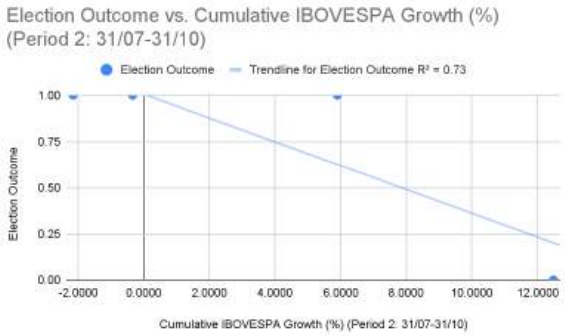


Fig 6: Election outcome versus cumulative IBOVESPA growth (%) for Period 2 (31/07–31/10), with trendline (R^2 = 0.730)

The findings contribute to the broader approach some papers may suggest by avoiding the application of stock market performance as a single predictor due to its limitations. Despite previous research, including in countries such as the United States, that indicates a stronger correlation between economic performance and electoral outcomes, Brazil's macroeconomic and political scenario may reduce the level of predictive efficacy and potentially indicate the need for the accounting of non-economic variables, such as inflation, unemployment, political approval ratings, and public sentiment analysis, to develop a more accurate forecasting model.

Implications

The research's findings directly address the gap identified in the literature review by applying economic election forecasting methods—widely studied in the United States—to the Brazilian context, assuming stock market performance, reflected by Brazil's IBOVESPA, would positively correlate with an incumbent victory. A wide range of studies have indeed suggested that economic optimism is one of the primary influences on voter behavior, yet the results obtained challenge this assumption. Nonetheless, this paper adds to the limited body of literature on economic forecasting models in Brazilian elections and underscores the need for country-specific models rather than the application of theories proven efficient for other nations.

As the results indicate, IBOVESPA growth alone is not a strong predictor of election outcomes in Brazil, despite certain elections supporting the hypothesis. External factors, such as political polarization, anti-incumbency sentiment, and non-economic voter priorities, should be measured to consider a potentially improved model.

Even though the sample is limited to four elections in which the incumbent sought reelection, the broader electoral trends observed can still provide valuable insights, such as the importance of considering economic performance alongside political, social, and historical contexts to make accurate predictions. Additionally, it also provides input into specific relationships, like the fact that investor confidence, reflected by IBOVESPA's growth, does not necessarily translate into voter confidence, again reinforcing the importance of testing a more holistic approach to making more accurate predictions.

These findings are also significant because they challenge the effectiveness of traditional economic voting theories in Brazil and highlight how electoral outcomes are not isolated from political and social dynamics. This insight is valuable for policymakers, economists, and political analysts who seek to make election forecasting models more accurate for personal or societal benefit, by developing country-specific forecasting models rather than relying on models built for economically developed countries. Finally, this research may encourage further studies that analyze this connection between multiple variables, which, if successfully done, could enhance the accuracy and applicability of electoral forecasting methods in Brazil and potentially other political and economic contexts.

Limitations

The research's limitations include the absence of clear patterns that support the hypothesis, as in some cases negative IBOVESPA growth corresponded with an incumbent win (e.g., 1998 and 2006), while in another case strong positive growth did not prevent an incumbent loss (2022). Additionally, the methodology assumes that IBOVESPA growth three months before an election directly reflects voter sentiment at the time of voting, which is not necessarily the case, as the stock market may often react to anticipated policy changes rather than to conditions in the moment. The study also does not account for confounding variables such as campaign strategies, corruption scandals, ideological polarization, or external factors like commodity prices and central bank decisions, present in cases like 2022, for instance, where COVID-19 had been a key factor influencing the incumbent's decisions during the term. Finally, its small sample size may not provide results as accurate as those from a larger sample, while also preventing statistical models such as a logistic regression model from being used due to perfect separation, which happens when the independent variable aligns too strongly with the dependent variable in a small dataset.

Conclusion

This study sought to investigate whether IBOVESPA can be a predictor of Brazilian presidential election outcomes by measuring its effectiveness against performance-based theories used in other contexts. Analyzing historical data on stock market performance and electoral results, the research sought to address a gap in the understanding of the applicability of economic forecasting models used in the United States to Brazil. The results suggest that even though IBOVESPA demonstrates some correlation with presidential election outcomes, it fails to predict outcomes consistently, which suggests that economic indicators alone may not be enough to correctly anticipate election outcomes in Brazil, potentially due to political and social factors that may also influence results.

This report has highlighted economic performance and voter sentiment, and its findings suggest that although market fluctuation is a measure of investor confidence, it fails to fully represent voter behavior. This exhibits the limitations of utilizing a purely economic approach to

election forecasting and also raises questions about the role of political stability in developing election dynamics.

Key to this research's contribution is the fact that IBOVESPA works alongside a wider range of influences when it comes to election forecasting, unlike in the United States, where stock market performance has been more closely correlated with the incumbent party's success in getting reelected. This suggests that economic performance-based theories should consider Brazil's unique variables in addition to its economic index.

Its implications also go beyond electoral forecasting and may be valuable to political analysts, economists, and policymakers. IBOVESPA's use in predictive models should complement other metrics, potentially including inflation rates, unemployment figures, and approval ratings. In addition, future research should conduct studies that determine whether similar trends are constant for different emerging economies or political contexts, and should focus on growing the model's robustness by considering variables like policy shifts, which would strengthen the forecasting model.

Yet this study raises an important question to consider: can any single economic indicator truly capture the complexities of democratic elections? While IBOVESPA provides valuable insights into investor confidence levels, it is still only one piece of the puzzle, and as Brazil's economic and political landscape continues to evolve, approaches to understanding the influence of other variables should also do so.

Appendix

Appendix A presents the raw data gathered for all Brazilian presidential elections from 1998 to 2022, regardless of whether the incumbent ran for reelection. The three years in which no incumbent sought reelection (2002, 2010, and 2018) were excluded from the analysis reported above.

Year	Cumulative Growth (%) (31/07–31/10)	Growth (Binary)	Cumulative Growth (%) (30/06–30/09)	Growth (Binary)	Incumbent Ran for Reelection	Election Outcome	Notes
1998	−0.3418	0	−0.3188	0	1	1	Fernando Henrique Cardoso ran for reelection
2002	−0.1497	1	−0.2299	0	0	0	Lula elected; no incumbent in the race
2006	5.8949	1	−0.4948	0	1	1	Lula reelected
2010	−8.8772	1	13.9390	1	0	0	Lula did not run; Dilma Rousseff elected
2014	−2.1509	0	1.7826	1	1	1	Dilma Rousseff elected for a second term
2018	10.3568	1	9.0430	1	0	0	Dilma Rousseff had been impeached; Bolsonaro elected
2022	12.4775	1	11.6649	1	1	0	Bolsonaro ran for reelection but lost

Appendix A: Raw data gathered from 1998 to 2022, regardless of whether the incumbent ran for reelection

Works Cited

B3. “Índice Ibovespa (Ibovespa) — Estatísticas Históricas.” B3, www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/indice-ibovespa-ibovespa-estatisticas-historicas.htm. Accessed 14 Apr. 2025.

Becher, Michael, and Michael Donnelly. “Economic Performance, Individual Evaluations, and the Vote: Investigating the Causal Mechanism.” The Journal of Politics, vol. 75, no. 4, 2013, pp. 968–79, https://doi.org/10.1017/S0022381613000959.

Biblioteca da Presidência da República. “Ex-Presidentes.” Biblioteca da Presidência da República, www.biblioteca.presidencia.gov.br/presidencia/ex-presidentes. Accessed 14 Apr. 2025.

Bisgaard, Martin. “How Getting the Facts Right Can Fuel Partisan-Motivated Reasoning.” American Journal of Political Science, vol. 63, no. 4, 2019, pp. 824–39, https://doi.org/10.1111/ajps.12432.

Erikson, Robert S. “Economic Conditions and the Presidential Vote.” American Political Science Review, vol. 83, no. 2, 1989, pp. 567–73, https://doi.org/10.2307/1962406.

Erikson, Robert S., and Christopher Wlezien. “Forecasting US Presidential Elections Using Economic and Noneconomic Fundamentals.” PS: Political Science & Politics, vol. 47, no. 2, 2014, pp. 313–16, https://doi.org/10.1017/S1049096514000092.

Guntermann, Eric, et al. “The Impact of the Economy on Presidential Elections throughout US History.” Political Behavior, vol. 43, 2021, https://doi.org/10.1007/s11109-021-09677-y.

Heidemanns, Merlin, et al. “An Updated Dynamic Bayesian Forecasting Model for the US Presidential Election.” Harvard Data Science Review, vol. 2, no. 4, 2020, https://doi.org/10.1162/99608f92.fc62f1e1.

Markus, Gregory B. “The Impact of Personal and National Economic Conditions on Presidential Voting, 1956–1988.” American Journal of Political Science, vol. 36, no. 3, 1992, pp. 829–34, https://doi.org/10.2307/2111594.

Tribunal Superior Eleitoral. “Justiça Eleitoral.” Tribunal Superior Eleitoral, www.tse.jus.br. Accessed 14 Apr. 2025.

Wink, Kenneth A. “Forecasting Models and the Presidential Vote.” Political Science Quarterly, vol. 134, no. 2, 2019, pp. 193–216, https://doi.org/10.1002/polq.12898.

Mega-Event Overlap and NYC Subway Ridership: A Case Study of the 2026 NBA Finals and FIFA World Cup By Zecheng Heliang

Abstract

Research suggests that mega-events like the Olympics can increase local public transit ridership, but little research has investigated the difference between the ridership volume of overlapping mega-events and isolated events. So, we investigated this using MTA turnstile data from four major Manhattan stations, which are 34th Street–Penn Station, Times Square–42nd Street/Port Authority Bus Terminal, Grand Central–42nd Street, and 14th Street–Union Square, during the unique overlap of the 2026 NBA Finals and 2026 FIFA World Cup on June 13, 2026. First, we hypothesized that overlap day would generate greater ridership than a standard Saturday baseline, the average isolated NBA Finals game days, and the average Isolated World Cup game days. Second, we hypothesized that 34th Street–Penn Station would show the largest percentage increase in ridership volume because it serves a more important role in facilitating the NBA Finals and World Cup compared to other stations. We analyzed the data and discovered that the overlap day did exceed the standard baseline at three of the four stations, but it was unexpectedly outnumbered by both isolated NBA Finals and World Cup days averages at all four stations. And Grand Central–42nd Street station turned out to show the largest percentage increase over baseline instead of 34th Street-Penn station. These results suggest that overlapping mega-events may not simply produce additional ridership surges, possibly due to decentralization, differences between weekday and weekend baselines, and decentralized watch-party attendance. Because MTA turnstile data captures only entries, and our overlap day sample was limited to a single date, further research with entry-and-exit data and larger samples is needed to confirm these dynamics.

INTRODUCTION

Mega-events and major sport events draw millions of visitors and large-scale gatherings (Monk). These large-scale gatherings place immense stress on public transportation and require extensive preparation for potentially transporting millions while also maintaining daily operations for regular commuters (Federal Transit Administration; Federal Transit Administration, "Public Transportation During Major Events"). Because public transit operators cannot easily restrict or regulate passenger flows at turnstiles during high volumes, these concentrated influxes often push local transit networks and station capacities to their physical limits (Xu et al. 130). For example, the Paris Olympics of 2024 had 9.5 million tickets sold, leading to immense stress on preparation, which included the closure of 185,000 roads and an increase in demand for public transit services (RATP Group). Driven by the Olympics, metro ridership increased by 13.9% to 1,514 journeys while the RER network increased by 4.9% to 494 million journeys (RATP Group).

While many investigations suggest that isolated mega-events like the 2024 Paris Olympics can significantly increase ridership volume (RATP Group), there are little investigation specifically to what extent the effect of overlapping events, specifically the overlapping of the 2026 Knicks NBA Finals and the 2026 FIFA World Cup on June 13 (FIFA; Quinn), on ridership volume in major stations. New York City (NYC) relies heavily on its Metropolitan Transportation Authority (MTA) subway system to support massive crowds during sporting events. According to a news article, MTA President Demetrius Crichlow said there was a 15% increase in ridership compared to a normal Saturday and ridership rose up to 3.16 million on June 13 (SILive.com). However, these city wide percentage increases only highlight the effect of overlapping events on ridership bluntly. Analyzing major station complexes like the 34th Street-Penn Station and Times Square-42 Street/Port Authority Bus Terminal can draw a clearer picture when identifying the effect of overlapping events on ridership because they directly serve as the entrance or a key pathway to stadiums and sport venues. For example, for the 2026 Knicks NBA Finals at Madison Square Garden, the primary transit gateway is the 34th Street–Penn Station complex, as it brings fans from boroughs of NYC, Long Island, and New Jersey (Madison Square Garden, "MSG Directions"). Simultaneously, World Cup fans traveling from NYC to MetLife Stadium at New Jersey are directed to use major transfer hubs, Times Square-42 Street/Port Authority Bus Terminal, as it serves the Port Authority Bus Terminal, a direct route to New Jersey (Port Authority of New York and New Jersey; Metropolitan Transportation Authority, "FIFA World Cup 2026 Transit Guide"). So, we compared the ridership volume of the overlap of NBA Finals and FIFA World Cup game on June 13, 2026, with the isolated NBA Finals games, isolated FIFA World Cup games, and a standard Saturday baseline at major stations to maximize seeing the effect of the unique overlap on ridership volume.

First, we hypothesize that the overlapping of the 2026 NBA Finals and the 2026 FIFA World Cup on June 13, 2026, will generate a greater ridership volume at the major stations than the isolated NBA Finals, the isolated World Cup, and the standard baseline (Hypothesis 1). We hypothesize this because studies have shown that mega-events will increase the demand of public transit, resulting in a surge of ridership volume. So, the overlap of the NBA Finals and the FIFA World Cup would certainly cause a surge in ridership volume. Second, we hypothesize that 34th Street-Penn Station will have a higher percentage of ridership increase than other major Manhattan hubs on the day when the NBA Final game overlaps with the FIFA World Cup game (Hypothesis 2). We believe 34th Street-Penn Station serves the biggest role in facilitating both mega-events because it serves the New Jersey Transit—which allows NYC fans to commute to MetLife Stadium—and it is a few blocks away from Madison Square Garden, and it serves numerous subway lines.

RESULTS

We collected turnstile swipes across four selected station complexes, which are 34th Street–Penn Station, Times Square–42 Street/Port Authority Bus Terminal (PABT), Grand Central–42Street, and 14th Street–Union Square. Then, we categorized the data into four

comparison conditions, which are the standard Saturday baseline, average isolated NBA Finals games, average isolated FIFA World Cup games, and overlap day, for each selected station.

We observed that the June 13, 2026 NBA Finals & World Cup overlap day outnumbered the turnstile swipes of the standard Saturday baseline across 3 selected stations: 34 St-Penn Station (98,026 Overlap vs. 80,819 Baseline; +21.3%), Grand Central-42 St (78,570 Overlap vs. 60,386 Baseline; +30.1%), and Times Sq-42 St/Port Authority Bus Terminal (130,423 Overlap vs. 113,564 Baseline; +14.8%) (Figure 1, Table 1, Figure 3). However, it is also observed that the Standard Saturday Baseline with 61,516 turnstile swipes actually slightly outnumbered the overlap day with 61,432 turnstile swipes for the 14th St-Union Sq station (Figure 1, Table 1) In overall, most stations show that the overlap day exceeds standard Saturday baseline in turnstile swipes besides 14th St-Union Sq, which shows an negligible difference in turnstile swipes as overlap day only decreased by 0.1% compared to standard Saturday baseline (Figure 3).

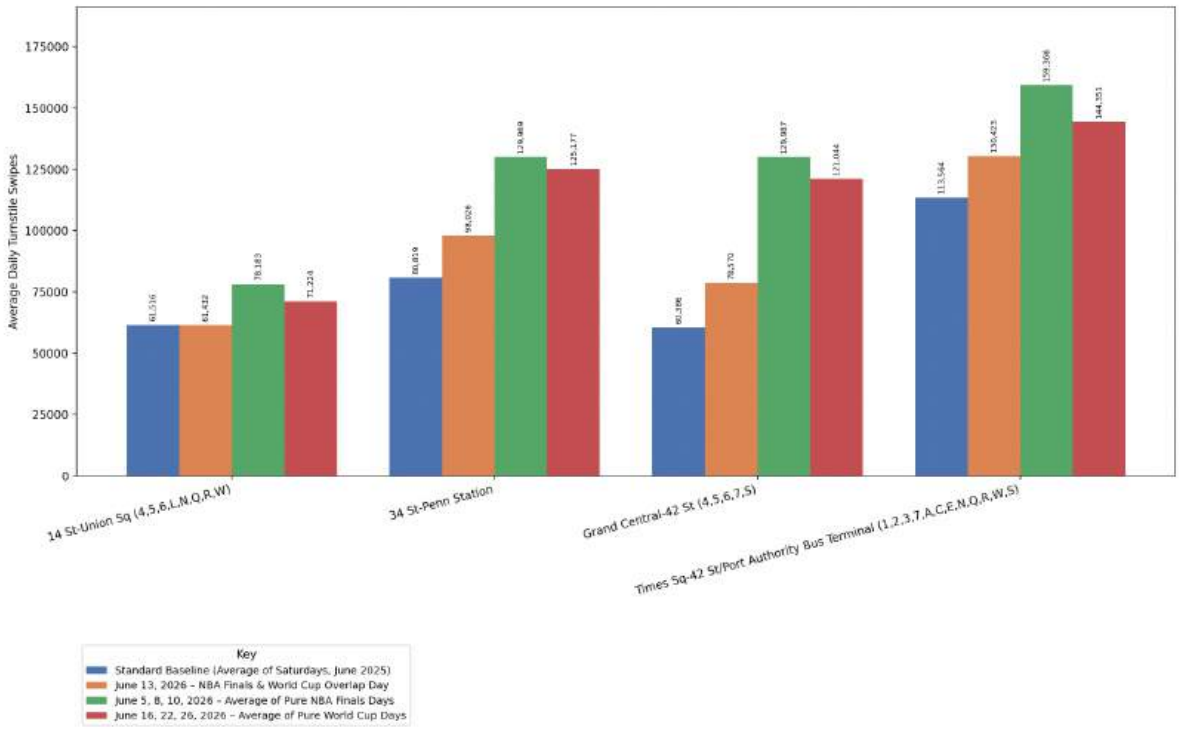


Figure 1: Comparison of average daily turnstile swipes across station complexes under four comparison groups. Comparison of average daily turnstile entries across four NYC subway complexes (14th St–Union Square, 34th St–Penn Station, Grand Central–42 St, and Times Sq–42St/PABT) across standard Saturday baselines (n = 4), the June 13, 2026 overlap day (n = 1), average isolated NBA Finals dates (n = 3), and average isolated FIFA World Cup dates (n = 3).

Station	June 2025 Saturdays	13-Jun-26	June 5, 8, 10 (2026)	June 16, 22, 26 (2026)
14 St-Union Sq (4,5,6,L,N,Q,R,W)	61516	61432	78183	71224
34 St-Penn Station	80819	98026	129969	125177
Grand Central-42 St (4,5,6,7,S)	60386	78570	129987	121044
Times Sq-42 St/Port Authority Bus Terminal (1,2,3,7,A,C,E,N,Q,R,W,S)	113564	130423	159366	144351

Table 1: Results of average daily turnstile swipes across station complexes and comparison conditions. Average daily MTA turnstile entries across four major Manhattan station complexes compiled for the standard June 2025 Saturday baseline (n = 4), the June 13, 2026, overlap day (n = 1), isolated NBA Finals dates (June 5,8, and 10; n = 3), and isolated FIFA World Cup dates (June 16, 22, and 26; n = 3).

The overlap day turnstile swipes is, in contrast, outnumbered by both the average isolated NBA Finals Days and average isolated World Cup Days across all 4 selected stations: 14th St-Union Sq (61,432 Overlap vs. 78,183 NBA vs. 71,224 World Cup), 34th St-Penn Station (98,026 Overlap vs. 129,969 NBA vs. 125,177 World Cup), Grand Central-42 St (78,570 Overlap vs. 129,987 NBA vs. 121,044 World Cup), Times Sq-42 St/Port Authority Bus Terminal (130,423 Overlap vs. 159,366 NBA vs. 144,351 World Cup) (Figure 1, Table 1). Using the data from Figure 1 and Table 1 is not sufficient to prove Hypothesis 1. By applying the Crawford & Howell (1998) single case t-test (Crawford and Howell 482), we can distinguish if the surge or drop in turnstile swipes between the overlap day and other conditions are statistically meaningful and not out of randomness. When compared against the standard Saturday baseline, 34th Street–Penn Station ($t(3)=3.47$, $p=0.0404$), Grand Central–42nd Street ($t(3)=4.29$, $p=0.0233$), and Times Square–42nd Street/PABT ($t(3)=4.91$, $p=0.0161$) all demonstrated a statistically significant ridership increase while 14th Street–Union Square showed no significant change ($t(3)=-0.02$, $p=0.9866$)(Figure 2, Table 2).

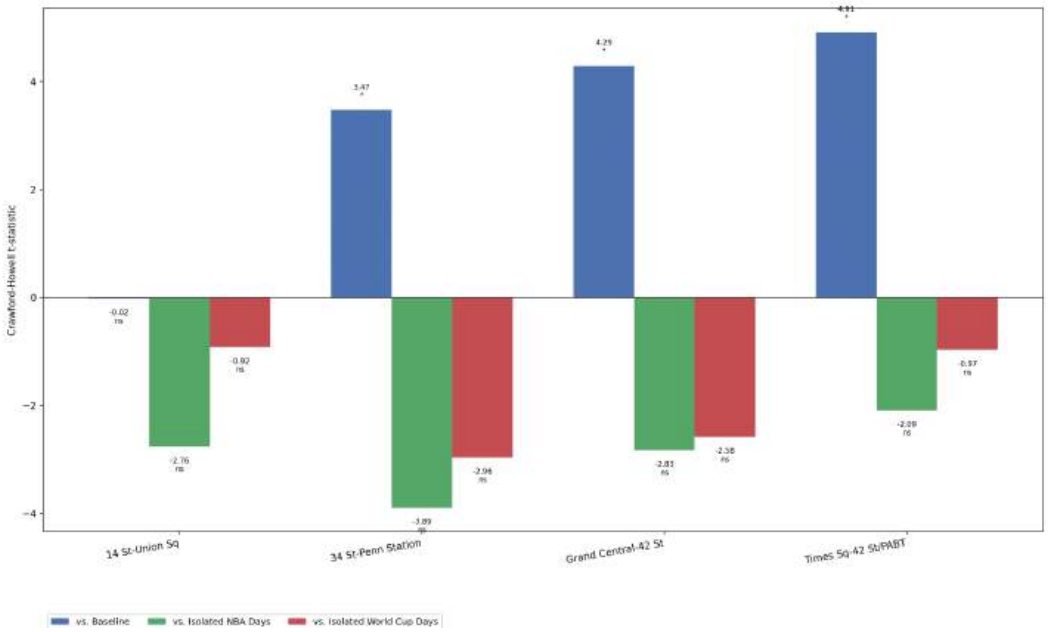


Figure 2: Crawford–Howell single case t-test comparisons between overlap day and other conditions across stations. Crawford & Howell (1998) single case t-test evaluated ridership statistical significance on the June 13, 2026, overlap day (n = 1) compared against the standard Saturday baseline (n = 4), isolated NBA Finals dates (n = 3), and isolated FIFA World Cup dates (n = 3) across four NYC subway station complexes. Significance was determined from the p-value that is calculated for each comparison and is indicated above or below each bar ($p < 0.05$; ns = not significant).

Station	t (vs Baseline)	df	p (vs Baseline)	t (vs NBA)	p (vs NBA)	t (vs World Cup)	p (vs World Cup)
14 St-Union Sq	-0.02	3	0.9866 ns	-2.76	0.1099 ns	-0.92	0.4530 ns
34 St-Penn Station	3.47	3	0.0404 *	-3.89	0.0601 ns	-2.96	0.0976 ns
Grand Central-42 St	4.29	3	0.0233 *	-2.83	0.1053 ns	-2.58	0.1228 ns
Times Sq-42 St/PABT	4.91	3	0.0161 *	-2.09	0.1723 ns	-0.97	0.4350 ns

Table 2: Crawford–Howell single case t-test results between overlap day and other conditions across stations. Crawford–Howell *t*-test results comparing ridership on the June 13, 2026,overlap day (n = 1) against the standard Saturday baseline (n = 4), isolated NBA Finals dates (n = 3), and isolated FIFA World Cup dates (n = 3), across four NYC subway station complexes. *t*= Crawford–Howell *t*-statistic; df = degrees of freedom; p-values are two-tailed ($p< 0.05$; ns = not significant).

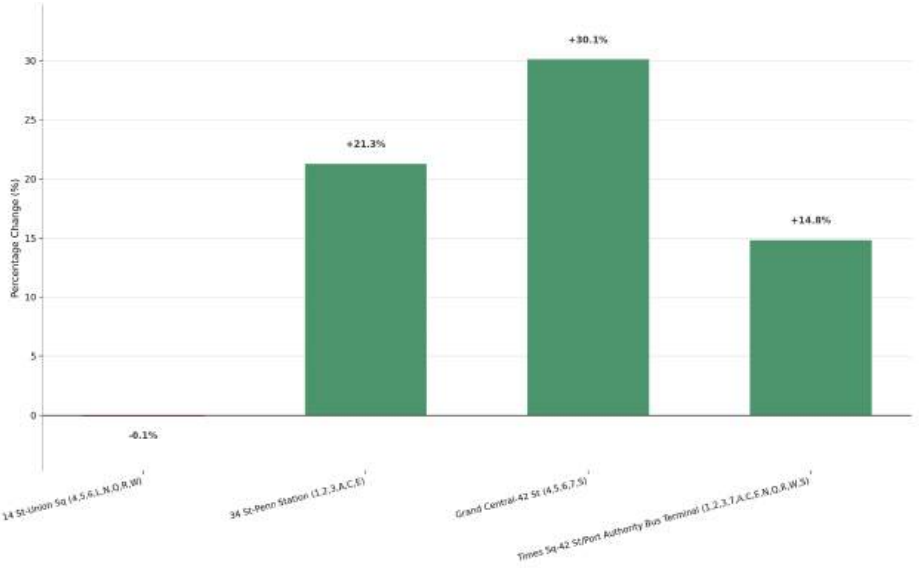


Figure 3: Percentage change in daily turnstile swipes on overlap day vs. Saturday baseline. Bar chart displaying the relative percentage change in total turnstile entries on June 13, 2026, overlap day (n = 1) compared to the standard Saturday baseline average (n = 4) across four major NYC subway station complexes. Values above each bar indicate the calculated percentage increase or decrease relative to baseline levels.

In contrast, when comparing the overlap day to both isolated event conditions, no station complexes exhibited statistically significant differences ($p>0.05$ for all comparisons; Table 2). For example, 34th Street-Penn Station's lower volume on the overlap day compared to isolated NBA days was not significant ($t(2)=-3.89$, $p=0.0601$). With the data collected shown in Figure 1 and Table 1, we used Python to calculate the percentage increase in turnstile swipes on overlap day compared to standard Saturday baseline through Equation 2.

DISCUSSION

In this study, we analyzed the impacts of the overlap of the NBA Finals and World Cup on MTA ridership volume. To understand the impact, we tested two hypotheses: first, the overlap day will surpass the standard Saturday baseline and the average of isolated event days in ridership volume; second, 34 St-Penn Station will have the greatest percentage increase in ridership volume compared to other selected stations on the overlap day. Based on our analysis, the overlap day's turnstile swipes consistently exceed the standard Saturday baseline across three out of four selected stations (Figure 1, Table 1, Figure 3). Additionally, applying the Crawford & Howell (1998) single-case t-test (Crawford and Howell 482) confirmed that these surges were statistically significant ($p<0.05$ across all three hubs). When comparing overlap day against the average of isolated mega-events dates, however, we discovered that turnstile entries across all four stations were consistently lower on the overlap date than during the average isolated events days across all four selected stations. Additionally, the Crawford-Howell test revealed that these turnstile swipes reduction were not statistically significant ($p>0.05$ across all stations), illustrating that the lower raw counts on overlap day are statistically insignificant. In sum, the

empirical data does not support Hypothesis 1 because the overlap day does not demonstrate a statistically significant positive surge ($p<0.05$) above both average isolated events days' levels while also having to outnumber the total turnstile swipes.

The finding that total turnstile swipes were lower on overlap day than average isolated sports event days may likely be caused by behavioral and structural reasons. First, overlap day might have caused a decrease in large scale gathering or decentralization. Regular commuters and local residents may purposely avoid the use of public transit due to anticipated congestion. Commuters may especially avoid stations like 34 St-Penn Station and Times Square–42nd Street/Port Authority Bus Terminal (PABT) since they serve a significant role for fans' commute. Second, due to the fact that both isolated events conditions included weekdays whereas the overlap day was on a Saturday, the turnstile swipes of both isolated conditions naturally integrated the combination of the workforce and fans commuting, while the overlap day lacked as many workforces commutes. Third, because the June 13 NBA Finals game took place in San Antonio instead of Madison Square Garden, fans may choose to attend watch parties or viewing events around their neighborhoods and remain more decentralized. Consequently, the use of stations like 34th St-Penn Station may drop.

The unexpected finding that Grand Central–42nd St (+30.1%, Figure 3) rather than 34th St–Penn Station (+21.3%, Figure 3) can be explained by the existence of only turnstile entries in the MTA turnstile datasets. 34th St-Penn Station may be a destination for fans as it is a few minutes from Madison Square Garden, so turnstile exits may outnumber entries. Conversely, fans may begin at Grand Central–42nd St, so turnstile entries may outnumber exits.

Several methodological limitations should be considered when interpreting these findings. First, MTA turnstile data only capture entry swipes, which is mentioned before, rather than passenger exits or internal line-to-line transfers. As a result, spectators transfer between subway lines within complex hubs without exiting through turnstiles, and exits are not captured in individual station swipe counts. Second, our sample size for the overlap day condition was limited to only June 13, 2026, while the isolated sports events parameters were calculated across three-day averages. A small sample size, as such, is too blunt to conclude that the overlap day does not increase public transit ridership volume more than isolated event days as well as the baseline. Third, although we carefully selected dates that are similar to June 13, the selected dates for each comparison group may not exactly match June 13 in terms of temperature, humidity, precipitation, or other weather conditions that could influence subway usage.

Material and Methods

We picked the station complexes that are most sensitive to the NBA Finals and World Cup and most accessible regionally, as they all serve many subway lines. 14th Street-Union Square serves the 4, 5, 6, L, N, Q, R, and W subway lines, while Grand Central-42 Street serves the 4, 5, 6, 7, and S subway lines (Metropolitan Transportation Authority, "Meatpacking District... Neighborhood Map"; Metropolitan Transportation Authority, "Turtle Bay... Neighborhood Map"). 34th Street–Penn Station sits near Madison Square Garden and connects

to the 1, 2, 3, A, C, E subways, LIRR, and NJ Transit (Metropolitan Transportation Authority, "Turtle Bay... Neighborhood Map"). Times Square–42nd Street/PABT serves 1, 2, 3, 7, N, Q, R, W, S, A, C, and E subway lines, as well as the Port Authority Bus Terminal (Metropolitan Transportation Authority, "Turtle Bay... Neighborhood Map"). Importantly, both station complexes offer direct transit links to New Jersey for New York City fans to commute to MetLife Stadium and New Jersey fans to commute to Madison Square Garden.

The dates for four comparison groups, which are the standard Saturday baseline (June 7, 14, 21, and 28 of 2025), overlap day (June 13, 2026), isolated NBA Finals game days (June 5, 8, and 10 of 2026), and isolated World Cup game days (June 16, 22, and 26 of 2026), are selected because they share similar seasonal and weather conditions. Additionally, no mega-events were hosted on any of the Baseline's dates, making it a perfect baseline for comparison.

To evaluate ridership fluctuations across selected stations and test Hypothesis 1, we used Python 3.9.6 to retrieve, filter, calculate, and generate visualizations by using the Metropolitan Transportation Authority (MTA) ridership data obtained through the New York State Open Data API. The Requests package was first used to query the API by filtering selected stations and dates. Python then used the data to prepare four comparative groups by computing the data into Equation1. Notably, the overlap day is not involved in the equation because it is a singular day, so it doesn't require averaging. In the equation, M represents the mean daily turnstile ridership volume for a specific station is under comparison parameter c — which refers to either standard baseline, Isolated NBA Finals parameter, or Isolated FIFA World Cup parameter; V represents the ridership value for a specific day d, station s, and comparison condition; and n represents the total number of observed days in comparison condition c.

$$M_{s,c} = \frac{\sum_{d} V_{d,s,c}}{n_c}$$

(Equation 1)

In order to determine whether the differences in ridership volume between the overlap day parameter and the other three parameters are truly events-driven instead of being a coincidence, we applied the Crawford & Howell (1998) single case t-test (Crawford and Howell 482), as it is suited for investigations that involves small sample like ours (n=1). We retrieved the turnstile data Requests package and then calculated the mean and standard deviation using the NumPy package. The overlap day's ridership value was then compared independently against the standard baseline and the two average isolated events days comparison groups using Equation 2. The SciPy package helped do so by computing the Crawford and Howell (1998) single-case *t*-test and calculating the two-tailed *p*-values from the *t*-distribution. In the equation, t represents the t-statistics of station s, O represents the total turnstile swipes on the overlap day in station s, M represents the average turnstile swipes of condition c in station s, S represents the standard deviation of condition c in station s, and is the number of days in the condition (n = 4 for

Baseline; n = 3 for isolated events). Two-tailed p-values were calculated from a standard t-distribution using equation 3, which is the degrees of freedom equation (df = 3 for Baseline; df = 2 for isolated events). Additionally, statistical significance was set at $p<0.05$.

$$t_s = \frac{O_s - M_{s,c}}{S_{s,c} \times \sqrt{\frac{n_c+1}{n_c}}}$$

(Equation 2)

$$df = n_c - 1$$

(Equation 3)

To answer Hypothesis 2, we compared the overlap day's total turnstile swipes with the average turnstile swipes of Saturdays of June 2025. We first used the Requests package to retrieve turnstile data from the two comparison groups. The data was then computed into Equation 1 to find the average turnstile swipes of the Saturdays in June 2025. With the average turnstile swipes of Saturdays obtained, we compared them with the overlap day by computing Equation 4. In the equation, P represents the percentage surge in a specific station s, O represents the overlap day's ridership volume for station s, and M represents the standard Saturday baseline or the mean turnstile swipes for Saturdays of June 2025 for stations.

$$\Delta P_s = \left(\frac{O_s - M_{s,baseline}}{M_{s,baseline}}\right) \times 100$$

(Equation 4)

Finally, visualization and numerical tables of all data collection, analysis, and results were generated using the matplotlib package. All Python scripts used for data cleaning, equation calculations, and visual charts are publicly available on GitHub at https://github.com/zechengheliang57-ux/overlap-events-research.git.

Works Cited

Crawford, J. R., and David C. Howell. “Comparing an Individual’s Test Score Against Norms Derived from Small Samples.” The Clinical Neuropsychologist, vol. 12, no. 4, 1998, pp. 482–486. https://doi.org/10.1076/clin.12.4.482.7241.

Monk, Paul. “How Cities Can Prepare Their Transportation Networks for Major Events.” Cubic, 8 Sept. 2025, https://www.cubic.com/transportation/news/blog/how-cities-can-prepare-their-transportation-networks-for-major-events.

Federal Transit Administration. “FTA Major Event Playbook.” U.S. Department of Transportation, 4 Dec. 2025, https://www.transit.dot.gov/sites/fta.dot.gov/files/2025-12/FTA-Major-Event-Playbook.pdf.

---. “Public Transportation During Major Events.” U.S. Department of Transportation, 15 Apr. 2026, https://www.transit.dot.gov/about/public-transportation-during-major-events.

FIFA. “FIFA World Cup 26™ Match Schedule Revealed.” FIFA, 4 Feb. 2024, https://www.fifa.com/en/tournaments/mens/worldcup/canadamexicousa2026/articles/fifa-world-cup-26-match-schedule-revealed.

Madison Square Garden. “Finals, Game 2 Watch Party.” Madison Square Garden, 5 June 2026, https://www.msg.com/events-tickets/new-york-knicks-road-game-watch-party-madison-square-garden-june-2026/3B0064BC946A2668.

---. “MSG Directions.” Madison Square Garden, https://www.msg.com/madison-square-garden/directions/.

Metropolitan Transportation Authority. “FIFA World Cup 2026 Transit Guide.” Metropolitan Transportation Authority, https://www.mta.info/world-cup-2026-new-york-new-jersey.

---. “Meatpacking District, West Village, Flatiron District, Greenwich Village, Union Square, Gramercy, Noho, Chelsea Neighborhood Map.” Metropolitan Transportation Authority, https://www.mta.info/document/726.

---. “Turtle Bay, Garment District, Hell’s Kitchen, Times Square, Chelsea, Murray Hill Neighborhood Map.” Metropolitan Transportation Authority, https://www.mta.info/document/721.

Port Authority of New York and New Jersey. “FIFA World Cup Travel Advisory for Monday, June 22.” Port Authority Press Room, June 2026, https://www.panynj.gov/content/port-authority/en/press-room/press-release-archives/2026-press-releases/fifa-world-cup-travel-advisory-for-monday--june-22.html.

Quinn, Sam. “2026 NBA Finals Schedule: Knicks vs. Spurs Odds, Times with Game 5 on Saturday.” CBS Sports, 10 June 2026, https://www.cbssports.com/nba/news/2026-nba-finals-schedule-odds/.

RATP Group. “Performances.” RATP Group 2024 Activity and Sustainability Report, https://ratpgroup.com/en/2024-asdr/performances/.

SILive.com. “MTA Ridership during Knicks Championship Win: What the Numbers Show.” SILive.com, Advance Local Media, June 2026, https://www.silive.com/nyc/2026/06/mta-ridership-during-knicks-championship-win-what-the-numbers-show.html.

Xu, Xin-yue, et al. “Capacity-Oriented Passenger Flow Control under Uncertain Demand: Algorithm Development and Real-World Case Study.” Transportation Research Part E: Logistics and Transportation Review, vol. 87, 2016, pp. 130–148. https://doi.org/10.1016/j.tre.2016.01.004.

Xu, Yanyan, and Marta C. González. “Collective Benefits in Traffic during Mega Events via the Use of Information Technologies.” Journal of the Royal Society Interface, vol. 14, no. 129, 2017, article 20161041. https://doi.org/10.1098/rsif.2016.1041.

Short-Term Returns, Long-Term Costs: How Corporate Stock Buybacks Can Hinder Innovation By Ceylan T. Urfali

Abstract

In 2025, S&P 500 companies spent more than one trillion dollars on stock buybacks, setting a new record and creating arguments over how large U.S. firms use company funds on spendings that give short-term income. This paper examines whether high corporate stock buyback activity hinders long-term investment in Research and Development (R&D), arguing that stricter regulation of buybacks through reforms to SEC Rule 10b-18 would enable funds to get spent on innovation. To investigate this relationship, a meta-analytic approach was used, synthesizing evidence from peer-reviewed empirical studies coded across four factors: buyback activity, R&D investment, investor time horizon, and macroeconomic effects. Across all the studies that have been analyzed, firms with higher repurchase intensity showed lower R&D spending, reduced patent output, and a consistent pattern of underinvestment in long-horizon productive capabilities. The repurchase event window data further confirmed that R&D intensity declines during active buyback periods compared to both before and after. These results support the central thesis: the current permissive structure of SEC Rule 10b-18, and counting the short-term shareholder pressure, systematically disadvantages long-term innovation investment. Redesigning this framework is a critical step toward building a more innovation-driven American economy

Literature Review

Introduction

By September 2025, S&P 500 buyback spending over the course of 12 months reached a record high of $1.020 trillion, highlighting the enormous scale of stock buyback usage in the modern U.S. market (S&P Dow Jones Indices, 2025). This trend has intensified arguments whether buybacks primarily show efficient capital distribution or whether they increase short-term incentives that prioritizes near-term share value contribution and earnings-per-share targets instead of long-term investment. The consequences of this debate extend beyond financial markets because corporate funds are limited, and large repurchase programs can directly reduce internal financing for Research and Development, workforce building, and other innovation-related investments that require stable multi-year support (Lazonick, 2014; Palladino, 2019). As a result, the regulatory design that governs how repurchases are executed has become a key factor in evaluating whether payout practices influence firms' capacity to sustain innovative activity over time.

This question is closely connected to Securities and Exchange Commission Rule 10b-18, which provides a non-exclusive safe harbor that reduces manipulation-liability risk when repurchases satisfy conditions related to timing, pricing, daily volume, and manner of purchase

(Securities and Exchange Commission, 2003). Because the safe harbor focuses on execution rules rather than cumulative scale, firms may comply on a day-by-day basis while still conducting large programs over extended periods. Critics argue that this structure can enable repurchases to become routine, while reform proposals reflect continuing concern about whether existing transparency rules truly address the incentives surrounding buyback timing, intensity, and opportunity costs (Lipson & Fernstrom, 2019). Empirical research further suggests that repurchases associated with meeting earnings benchmarks can be linked to reductions in company investment, including innovation-related spending, strengthening the need to examine buybacks through the lens of incentives and capital allocation rather than treating them as neutral payout events (Almedia et al., 2013). Stricter regulation of corporate stock buybacks in large publicly traded U.S. firms strengthens long-term innovation by reducing short-term incentives and redirecting funds toward sustained Research & Development investment. The purpose of this paper is to examine how stricter regulation of corporate stock buybacks, particularly through reforms to SEC Rule 10b-18 affects payout incentives and the use of company funds, focusing on the trade-off between stock buybacks, and long-term investment in Research and Development.

Background

Stock buybacks have become a dominant corporate payout mechanism in U.S. markets, they now change how large public firms return cash to shareholders and how investors interpret corporate financial decisions. A stock buyback occurs when a firm uses corporate funds to repurchase its own shares from the market in order to reduce the number of shares outstanding. Because earnings per share (EPS) is calculated by dividing profits by shares outstanding, reducing the denominator can increase EPS even if total profits remain unchanged, which helps explain why buybacks can appear attractive as a tool for improving per-share performance (Almedia, Fos, & Kronlund, 2016). This mechanical feature matters because modern financial markets often treat EPS as a signal of corporate success even though EPS can rise for reasons that do not reflect stronger long-run productivity.

Buybacks also matter because they represent a major capital allocation choice, not just the financial "distribution." Since the adoption of Securities and Exchange Commission Rule 10b-18 in 1982, open market repurchases have seen a growth, this is partly because the regulatory structure saw reduced constraints and increased flexibility in how companies execute repurchase programs. As a result, repurchases now represent a large share of corporate cash outflows, raising a key question in corporate finance: when firms have excess cash, will they prioritize shareholder payouts, or will they direct those funds toward long-horizon investments that build future value (Fried & Wang, 2018)? Corporate funds are limited, and money used to repurchase shares cannot simultaneously fund Research & Development projects, long-term skill development, or other investments whose payoffs come gradually.

This short-term orientation is strengthened by how public firms are evaluated. Quarterly earnings cycles, analyst forecasts, and EPS targets mostly dominate corporate narratives, and

stock prices frequently respond sharply to whether firms meet or miss expectations. Survey evidence suggests many executives treat meeting short-term benchmarks as a top priority, even when the actions taken to satisfy those benchmarks reduce long-term company value (Graham, Harvey & Rajgopal, 2005). In that environment, buybacks and dividends can become "prefered" choices because they deliver clear and immediate signals to markets, while investments in innovation may not show measurable results within quarterly windows. The outcome is a financial system that can unintentionally reward speed over long-horizon value creation.

How a Stricter SEC Rule 10b-18 Could Constrain Large-Scale Buybacks

SEC Rule 10b-18 is a non-exclusive safe harbor that decides how open-market buybacks are executed by lowering a U.S. public company's risk of manipulation liability only if the company satisfies four conditions while doing buybacks on a given day: timing, pricing, daily volume, and manner of purchase (Securities and Exchange Commission, 2003; Securities and Exchange Commission, 2017). Rule 10b-18 was adopted in 1982 to provide this limited protection while still giving the idea that compliance does not immunize firms from other anti-fraud problems, like repurchasing while possessing material nonpublic information (Securities and Exchange Commission, 1982). Because the rule is designed around how trading occurs, instead of if the buybacks are "too large" in total, its impact depends on whether the execution constraints meaningfully raise the difficulty or risk of sustaining large programs across a period of time (Securities and Exchange Commission, 2003).

At the same time, the current design can still allow firms to conduct large repurchase programs through repetition by spreading them across many trading days. The safe harbor applies day-by-day, meaning that a firm can stay within the daily conditions while still spending enormous amounts over months or a full year (Securities and Exchange Commission, 2003). This helps explain why some observers argue that having execution conditions does not necessarily have effects on "scale" in practice, especially when firms can distribute stock buyback actions across a long horizon to keep each day within the rule's structure (Palladino, 2019). Even Jesse Fried, testifying to the U.S. Congress, suggested that abolishing the safe harbor "would likely have little effect" unless firms believed they faced significantly higher manipulation-liability risk, pointing directly to the idea that execution-based constraints matter only when they change a company's perceived legal risk (Fried, 2019).

Because of this gap between execution limits and scale, reform proposals often focus on tightening the conditions in ways that make high-frequency stock buyback transactions harder to run continuously, especially around information-sensitive time periods. For example, petitions submitted to the SEC have argued that Rule 10b-18 has not kept pace with free market structure and that its protections can be too permissive, motivating calls to revise the framework rather than leaving the safe-harbor unchanged (Lipson et al., 2018). Similarly, the SEC itself has repeatedly revisited reforms through developed disclosure alternatives, most notably the "Share Repurchase Disclosure Modernization" effort (Securities and Exchange Commission, 2021). Considered together, these reform methods reflect a common logic: if Rule 10b-18 is tightened

through smaller timing periods, stricter volume limits, or stronger disclosure hooks, firms may find it harder to rely on buybacks as a routine (Palladino, 2019).

High repurchase activity can conflict with long-horizon investment because both are taken from the same pool of corporate funds. Almedia, Fos, and Kronlund find that repurchases, especially those linked to meeting earnings benchmarks, are associated with declines in investment and R&D-related spending (Almedia et al., 2016). Lazonick similarly argues that large-scale open-market buybacks can weaken "productive capabilities" by shifting resources away from innovation and workforce development (Lazonick, 2014). If reforms to Rule 10b-18 make large, continuous repurchasing more costly or risky, firms may reduce payout intensity and commit more consistently to long-horizon projects such as exploratory Research & Development and talent development, which would ensure companies to focus on long-term growth while cutting funds on short-term incentive spendings such as buybacks. (Almedia et al., 2016; Lazonick, 2014).

Shareholder time horizons and the Long-Term Value of R&D

Shareholders don't only depend on what they own, they also depend on how long they intend to own the stock, and time horizons establish what they demand from corporate managers. Research in accounting and finance has been categorizing investors by horizon and shows that firms dominated by short-term oriented owners behave differently than firms with more long-term oriented owners because different investors reward different outcomes (Bushee, 1998).

In markets where ownership changes quickly, the most "valuable" corporate actions are often those that move near-term metrics and valuation, even if they do not strengthen the firm's productive capacity. Shleifer and Vishny's work (1990) on market incentives and pressures helps explain why: stock markets can reward actions that lift short-term market valuation, creating pressure for managerial choices that may harm long-term value creation (Shleifer & Vishny, 1990). This distinction matters for payout policy because shareholders are not a single united audience. Short-horizon traders tend to benefit from near-term price movements, while long-horizon shareholders benefit when the firm becomes more capable, competitive, and innovative over years.

The main consequence of horizon differences is that market pressure for short-term performance can motivate firms to take actions that boost near-term gains at the expense of long-term investment. That said, evidence from executive surveys indicates that many managers feel strong pressure to meet short-term earnings benchmarks, and that this pressure can influence real decisions by encouraging the cutting or delaying of long-term investments to avoid missing expectations (Graham, Harvey & Rajgopal, 2005). Market pressure for short-term performance does not operate only within managerial decision-making; it is also reflected in observable investment patterns. Asker, Farre-Mensa, and Ljungqvist (2015) find that public-market pressures are connected with reduced investment relative to private firms, consistent with the idea that incentives taken in public equity markets can discourage long-horizon spending (Asker, Farre-Mensa, & Ljungqvist, 2015). When management is evaluated through quarterly outcomes

and immediate stock reactions, the corporate incentive naturally leans toward decisions that can deliver fast financial results. In that environment, buybacks become tempting because they can raise earnings per share mechanically by reducing shares outstanding and can support price trends quickly, even when underlying productive skills have not improved. By contrast, Research & Development typically has delayed and uncertain payoffs and rarely produces immediate, visible earnings gains in the same quarter it is funded. As a result, even when R&D is strategically great for long-term growth, it can lose internal budget competition to payout actions that satisfy short-term valuation demands.

The time horizon also determines who gets the benefits of innovation. Long-horizon investment requires patience for uncertainty and a willingness to fund projects whose results may not be fully visible for years. Theory on incentives for innovation emphasizes that "patient capital" can be important for supporting exploratory projects, efforts that carry higher uncertainty, require longer development cycles, and pay off later rather than immediately (Manso, 2010). This logic clarifies why shareholders who stay invested over time are more closely linked to R&D value: they remain exposed to the long-run trajectory of the firm and therefore benefit when innovation eventually raises productivity, competitive advantage, or market power. In contrast, investors who exit quickly, and don't benefit the company, can profit from short-term price movements without bearing the waiting time or uncertainty that innovation demands. Incentive design inside the firm can increase this tension, because compensation and systems that tie rewards to near-term stock outcomes can move decision-making toward actions that deliver immediate market approval instead of long-horizon value creation (Hall, 2003).

These incentive dynamics create a governance misalignment: managers can become overly responsive to the investors who influence near-term pricing, even though innovation depends on multi-year investment and tolerance for uncertainty. In that environment, payout mechanisms that produce quick, clear market signals, especially buybacks, often gain an advantage over R&D, whose payoffs are delayed and harder to verify within quarterly windows. This is why shareholder time horizon has an impact on capital allocation: long-horizon shareholders are the group most exposed to the eventual outcomes of innovation and therefore have the strongest stake in preserving long-term productive investment. From this perspective, the buyback debate is partly about whether governance pressures and market incentives push firms to satisfy short-term traders or to protect long-term investors who rely on innovation paying off. The time horizon shapes what gets rewarded, and that difference helps explain why firms may choose buybacks over long-term investment even when capability-building is the better path for sustainable growth (Bushee, 1998; Graham et al., 2005).

Reducing Dividends to Finance Long-Term Innovation

Reducing dividends can potentially be an efficient way to expand a firm's capacity to be able to use corporate funds toward long-horizon innovation. Unlike one-time financing decisions, dividends are a recurring commitment that conditions how much cash remains available after a firm meets its payout expectations. When dividends are high or treated as

“sticky”, managers face an ongoing constraint: a big portion of free cash flow is marked for distribution before the firm can fund multi-year R&D programs, scale-up investments, or capability- building initiatives (Lintner, 1956). That said, dividend policy is not just a shareholder-relations choice; it is a capital allocation rule that decides whether the firm retains patient internal funding for projects whose benefits arrive later and are uncertain in the early stages (Myers & Majluf, 1984; Hall & Lerner, 2010).

The key mechanism is very straightforward. Lowering dividends increases retained earnings and preserves liquidity, giving management more freedom to fund investments that cannot easily be justified by immediate accounting returns. This matters because many innovation projects, especially exploratory R&D and the technology behind new products, require sustained funding across several years, with learning and failure built into the process (Manso, 2010). External financing can cover some of these needs, but it can also be costly, cyclical, and sensitive to Market sentiment (Hall & Lerner, 2010). Internal funds by contrast connect as “patient capital” inside the firm: they allow managers to maintain community in R&D even when quarterly conditions fluctuate or when early-stage outcomes are ambiguous (Myers & Majluf, 1984). By reducing the automatic outflow of cash through dividends firms can protect the steady investment, which long-horizon innovation typically requires.

Dividend reductions also reshape managerial incentives by changing what good performance looks like to investors and boards. When a firm commits to generous dividends, maintaining that payout often becomes a visible benchmark of success, and missing it can trigger negative market reactions or reputational harms (Brav, Graham, Harvey, & Michaely, 2005). That dynamic can bias managers to preserve payout capacity, sometimes by cutting investment budgets that do not turn quickly into visible earnings (Brav, Graham, Harvey, & Michaely, 2005). Cutting dividends can be politically difficult for management precisely because it forces a reset of shareholder expectations, but that reset can be strategically valuable if the firm is serious about long-term innovation (Brav et al., 2005). A lower dividend commitment can reduce the pressure to prioritize distributable cash today, opening room for investment strategies whose payoff is measured in capability growth, competitive positioning, and eventual productivity gains rather than just short-term financial earnings (Manso, 2010).

Importantly, the case for reducing dividends is strongest when it is paired with credible commitment and a clear investment rationale. Dividend cuts can be interpreted as a signal of weakness if the firm provides no explanation or if retained cash is later absorbed by low return projects or value neutral financial maneuvers. But when management links the dividend reduction to a disciplined, transparent long-horizon investment plan, such as multi-year R&D targets, capability-building milestones or strategic scaling goals the policy can be framed as reallocating cash from distribution to value creation. This aligns with the logic of horizon base shareholder value: long-term shareholders benefit most when the firm sustains investment through uncertainty and builds the productive base that generates future growth.

Overall reducing dividends can be understood as a practical finance and strategy for innovation. It increases the ability to use corporate funds more freely, supports continuity in

long-horizon projects, and shifts performance attention away from payout mechanisms toward capability building (Graham et al., 2005). For firms operating in innovation intensive industries or firms attempting to transition to a more innovation-driven strategy, dividend policy is therefore a meaningful lever: lowering dividends is not simply "paying shareholders less," but it is choosing to retain capital so the firm can invest in the long-term assets that ultimately determine durable competitiveness.

Conclusion

Over the course of this paper, the evidence demonstrates that corporate payout choices, particularly stock buybacks conducted under SEC Rule 10b-18, can systematically favor short-term incentives over long-horizon innovation. Because Rule 10b-18 primarily regulates how repurchases are executed rather than how large they may become in total, firms can sustain substantial buyback programs while remaining within the safe-harbor framework. This structure strengthens managerial and market pressures to prioritize near-term earnings outcomes. At the same time, shareholder time horizons are not uniform: short-term traders may benefit from immediate price and earnings-per-share effects, whereas long-term shareholders benefit more from sustained investment in exploratory R&D that strengthens a firm's productive capacity over many years. Dividend policy further complicates capital allocation because dividends represent sticky commitments that are costly to reduce, limiting flexibility specifically when firms require steady internal funding for long-horizon innovation and workforce development.

As a whole, these findings support the central claim of this paper: stricter regulation of stock buybacks, particularly reforms to rule 10b-18 that reduce short-term incentive payoffs and constrain large-scale repurchase programs, can strengthen long-term innovation by preserving internal funds for sustained research and development projects. The policy implications extend beyond corporate finance. Regulators should modernize rule 10b-18 through tighter volume and timing limits, stronger disclosure requirements, and enhanced restrictions during information-sensitive trading windows. At the same time, corporate boards and long-horizon shareholders should adopt payout policies that explicitly protect multi-year investment strategies.

If the United States seeks durable economic growth, firms must be encouraged to invest consistently in innovation, workforce upgrading, and productive capability. Exploratory R&D supports job creation, skill development, and global competitiveness in strategically important industries. Without reforms that realign incentives away from short-term payout strategies, companies may continue to trade long-run innovation and job-creating investments for short-term market rewards, weakening not only individual firms but also the long-term strength of the American economy.

Methods

This study examines how financial decision-making in large publicly traded companies influences long-term innovation projects, with a focus on the use of corporate funds between stock buybacks and Research & Development (R&D). The main question of this paper is how

companies can shift their priorities toward innovation projects rather than short-term shareholder payouts. Specifically, this research asks how large publicly traded companies can focus on Research & Development instead of investing more money each year in stock buybacks in order to preserve long-term growth. The study hypothesizes that implementing stricter policies that would limit the amount as a whole that is being invested in stock buybacks would allow firms to strategically spend more money and company resources toward R&D, therefore protecting long-horizon projects and strengthening overall firm performance.

To explore this relationship, the research utilizes a meta-analytical design, synthesizing findings from five key academic studies: Lazonick (2014), Asker et al. (2011), Tibbs and Zantout (2016), Nguyen et al. (2020), Swift (2018) that examine corporate financial behaviour, innovation gains, and investor incentives. This approach relies on evaluation of previously published academic research in order to identify consistent findings across different methodologies, rather than just collecting primary data. The primary research tool used in this process is a structured comparative dataset, developed in a spreadsheet format, where each selected study was analyzed according to the main questions that are being answered in this paper and as well as standardized categories such as research focus, data collection methods, key findings, and relevance to research question. This method helps to make a direct study comparison and highlights trends, like the allocation of corporate funds in stock buybacks and R&D investment, as well as the effects of short-term investor pressure in shaping corporate decision making.

The data used in this study includes both quantitative and qualitative analysis. Quantitative data comes from firm-level financial trends such as stock buyback volumes, R&D intensity, and different ways of innovation output measures like patent amount and citations. These numerical data gives evidence of the relationship between financial strategies of firms and innovation performance. By comparison, qualitative data makes up literature analysis, such as discussions of executive incentives, shareholder expectations from firms, and industrial trends that influence corporate financial behaviour. By using these two forms of data, the study creates a wider understanding of the factors contributing to found financial distribution within the firm.

Overall, this methodological approach contributes to making an evidence-based evaluation of how stock buyback activity may impact long-term innovation capability of a firm. Through this analytical research combined with different methods of analysis, the study aims to give a detailed perspective on the balance between short-term financial strategies and long-term economic growth of American corporations.

Results

The results of this meta-analysis draw from several peer-reviewed empirical studies, each examining the relationship between stock buyback activity and R&D investment from a different method. The pattern across all the studies is consistent: higher repurchase intensity is negatively associated with long-term innovation investment.

The most direct quantitative evidence comes from Tibbs and Zantout (2016), who measured R&D intensity as a percentage of total assets across firms grouped by repurchase intensity. Their findings show that underinvestment in R&D reaches -1.04% of total assets in the highest repurchase timeframe, compared to only -0.03% in the lowest. This difference is statistically significant. Moreover, Figure 1 shows that event window data reinforces the underinvestment in R&D happening during repurchase event windows, showing that R&D intensity decreases during the repurchase event window compared to both the pre- and post-repurchase periods, this trend confirms that funds are being actively redirected away from innovation during buyback programs.

Figure 1

Appendix C
Example to show logic behind using adjusted Total Assets for repurchases in time-series ratios

		Pre-Repurchase Event Window								Repurchase Event Window			Post-Repurchase Event Window							
	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	+1	+2	+3	+4	+5	+6	+7	+8
R&D	8	8	8	8	8	8	8	8	8	7	6	7	8	8	8	8	8	8	8	8
Repurchase Amount (Single Qtr)	0	0	0	0	0	0	0	0	0	2	1	1	0	0	0	0	0	0	0	0
Repurchase Amount (Cumulative)	0	0	0	0	0	0	0	0	0	2	3	4	4	4	4	4	4	4	4	4
Total Assets (Unadjusted)	100	100	100	100	100	100	100	100	100	98	97	96	96	96	96	96	96	96	96	96
Total Assets (Adjusted)	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
R&D-Intensity % (Unadjusted)	N/A	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	7.00	6.12	7.22	8.33	8.33	8.33	8.33	8.33	8.33	8.33	8.33
R&D-Intensity % (Adjusted)	N/A	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	7.00	6.00	7.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00

Nguyen et al. (2020) provide a more supporting evidence by using a sample of 9,970 U.S. firms from years 1996 to 2009, they document a negative relationship between share repurchases and firm innovation, finding that firms with higher buyback activity consistently showed lower investment in R&D, patents, and other innovation-related assets. The breadth of this sample makes the pattern difficult to deny by just seeing as industry-specific or coincidental. Nguyen et al. (2020) also identify short-term investor pressure as a key driver. They argue that firms respond to shareholder demands by deprioritizing long-horizon spending, which aligns with the time-horizon argument central to this paper.

Lazonick (2014) gives an analysis of top-spending firms between 2003 and 2012, he found out that Pfizer spent 71% of its net income into stock buybacks during this period, even while publicly stating that high U.S. drug prices were necessary to fund R&D. Across the firms he examined, companies spent almost triple on buybacks that they invested in R&D. This finding is significant because it shows that the trade-off is not just an accident but shows a deliberate prioritization trend within the firm level.

Swift (2018) measures the relationship through patent-based variables to show firms' innovative output. By using a sample of publicly traded manufacturing firms from 1996 to 2005, Swift finds that both patent counts and patent citation rates are negatively associated with high buyback activity. This means the effect extends beyond spending patterns to the actual production of innovation. Swift also proves that many firms conduct repurchase programs when they face investor pressure, connecting short-term shareholder demands to real capital allocation decisions.

Asker, Farre-Mensa, and Ljungqvist (2011) add a structural dimension by comparing publicly traded and privately held U.S. firms. Using a private firm database developed by Sageworks Inc., they find that public firms invest less than their private counterparts, attributing this to executive incentives and what they call managerial short-termism. Because private firms face less pressure from quarterly earnings cycles, they are better positioned to sustain long-horizon investment. Their study references evidence of a negative correlation between buybacks and innovative output, and their structural argument strengthens the thesis of this paper.

In conclusion, the five studies turns out to give consistent results. Across different time periods, firm samples, industries, and methodological approaches, each study finds evidence that high buyback activity is associated with lower R&D investment or reduced innovation output. The quantitative findings from Tibbs and Zantout (2016) and the event window data confirm a statistically significant, real-time reduction in R&D intensity during a repurchase activity. The broader samples from Nguyen et al. (2020), the sectoral data from Lazonick (2014), the patent measures from Swift (2018), and the structural comparison from Asker et al. (2011) collectively demonstrate the robustness of this relationship. These results provide strong support for the hypothesis that corporate stock buybacks, at the scale enabled by the current structure of SEC Rule 10b-18, hinder the long-term innovation investment that firms and the broader American economy depend on in order to have sustainable growth.

Discussion

The findings from the meta-analysis support the hypothesis that stricter policies that put on limitations on stock buybacks could help redirect more company funds and resources toward Research & Development projects that protect company long-term growth. Across studies that have been analyzed, there is a consistent pattern: firms that execute big amounts of stock buybacks show weaker commitment and lower spending on innovation related investment, especially when company executives and shareholders are focused on short-term financial profits. Several studies in the dataset find a relationship between high buyback activity with reduced R&D intensity, lower innovative outcomes, or bad technological developments. This does not prove that every stock buyback directly causes a reduction in R&D, but it does suggest that buybacks can become part of a larger corporate trend where short-term shareholder returns are given more importance over long-term productive investment that would actually benefit the company.

The main conclusion drawn from this research is that stock buybacks are not necessarily harmful in every financial situation, but large-scale and repeated buyback programs can create a trade-off for companies that would be dangerous for company life and popularity, as well as companies that need to keep up with innovation. The evidence reviewed shows that Research & Development requires continuous funding, patience, and tolerance for uncertainty, while stock buybacks provide faster and certain financial gains, such as higher earnings per share or short-term stock price support, which increases companies' popularity and establishes them as

respectful in the eyes of investors. This contrast helps explain why executives may choose repurchases over R&D, especially when they are responding to investor pressure, short-term shareholders, or incentives that are connected to stock performance. So, the issue is not just that companies buy back shares, but that current financial incentives may encourage firms to overuse buybacks at the expense of innovation.

Another limitation is the possibility of researcher selection bias within this study. Since the studies chosen focused on buybacks, innovation, executive incentives, and investor time horizons, the dataset naturally emphasizes research that is already connected to the central argument. Even though this helped keep the study focused, it also means that future research should include more studies that defend buybacks or argue that repurchases can efficiently return unused capital to shareholders. Including these opposing perspectives would make the analysis more balanced and would help understand when buybacks are economically efficient or when they become harmful by limiting innovation spendings.

Based on these findings, the main recommendation is not to ban stock buybacks entirely, but to reform the policies that allow large publicly traded firms to conduct repurchases with limited restrictions. A stronger version of SEC Rule 10b-18 could require greater disclosure of buyback intentions, clearer descriptions of how repurchases affect the investment plan of companies, and stronger limits on buybacks when firms are reducing R&D or using repurchases mainly to support short-term stock performance. Future research should also examine firm-level data over time to compare companies before and after major repurchase programs, especially in R&D-intensive industries such as pharmaceuticals, technology, energy, and manufacturing. This would help determine whether buybacks are actually hindering innovation or whether the relationship depends on industry, firm size, executive incentives, and investor pressure.

Works Cited

Almeida, H., Fos, V., & Kronlund, M. (2016). The real effects of share repurchases. Journal of Financial Economics, 119(1), 168–185. https://doi.org/10.1016/j.jfineco.2015.08.008

Asker, J., Farre-Mensa, J., & Ljungqvist, A. (2011). Comparing the investment behavior of public and private firms (No. w17394). National Bureau of Economic Research.

Brav, A., Graham, J. R., Harvey, C. R., & Michaely, R. (2005). Payout policy in the 21st century. Journal of Financial Economics, 77(3), 483–527. https://doi.org/10.1016/j.jfineco.2004.07.004

Bushee, B. J. (1998). The influence of institutional investors on myopic R&D investment behavior. The Accounting Review, 73(3), 305–333.

Fried, J. M. (2019, October 17). Testimony of Jesse M. Fried on stock buybacks before the U.S. House of Representatives Subcommittee on Investor Protection, Entrepreneurship, and Capital Markets [Testimony]. SSRN. https://ssrn.com/abstract=3474175

Graham, J. R., Harvey, C. R., & Rajgopal, S. (2005). The economic implications of corporate financial reporting. Journal of Accounting and Economics, 40(1–3), 3–73. https://doi.org/10.1016/j.jacceco.2005.01.002

Hall, B. H., & Lerner, J. (2010). The financing of R&D and innovation. In B. H. Hall & N. Rosenberg (Eds.), Handbook of the economics of innovation (Vol. 1, pp. 609–639). Elsevier. https://doi.org/10.1016/S0169-7218(10)01014-2

Lazonick, W. (2014). Profits without prosperity: Stock buybacks manipulate the market and leave most Americans worse off. Harvard Business Review, 92(9), 46–55.

Lintner, J. (1956). Distribution of incomes of corporations among dividends, retained earnings, and taxes. The American Economic Review, 46(2), 97–113.

Lipson, M. L., & Fernstrom, A. (2019). IEX Group, Inc. [Case]. Darden Business Publishing, University of Virginia.

Manso, G. (2011). Motivating innovation. The Journal of Finance, 66(5), 1823–1860. https://doi.org/10.1111/j.1540-6261.2011.01688.x

Myers, S. C., & Majluf, N. S. (1984). Corporate financing and investment decisions when firms have information that investors do not have. Journal of Financial Economics, 13(2), 187–221. https://doi.org/10.1016/0304-405X(84)90023-0

Nguyen, L., Vu, L., & Yin, X. (2021). Share repurchases and firm innovation. Accounting & Finance, 61, 1665-1695.

Palladino, L. (2019). Ending shareholder primacy in corporate governance [Report]. Roosevelt Institute.

S&P Dow Jones Indices. (2025, December 18). S&P 500 buybacks (Q3 2025) [Report].

Securities and Exchange Commission. (1982). Purchases of certain equity securities by the issuer and others (17 C.F.R. § 240.10b-18). eCFR. https://www.ecfr.gov

Securities and Exchange Commission. (2003). Purchases of certain equity securities by the issuer and others (Release Nos. 33-8335; 34-48768; File No. S7-35-02). U.S. Government Printing Office. https://www.sec.gov

Securities and Exchange Commission. (2017). Purchases of certain equity securities by the issuer and others (17 C.F.R. § 240.10b-18). eCFR. https://www.ecfr.gov

Securities and Exchange Commission. (2021, December 15). Share repurchase disclosure modernization (Proposed Rule, Release Nos. 34-93783; IC-34440; File No. S7-21-21) [Proposed rule]. https://www.sec.gov

Swift, T. (2022). Do stock buybacks suppress corporate innovation?. International Journal of Innovation Management, 26(06), 2250037

Gender-Based Price Discrimination By Yufei Guo

Abstract

Women are widely observed to pay more than men for comparable products, a pattern known as the "pink tax." Prior research has offered two competing explanations — higher production costs for female-targeted goods, and differences in price elasticity of demand (PED) between genders — but has rarely examined both together. This study investigates gender-based pricing on a Chinese e-commerce platform by jointly estimating cost, PED, and profit margin across personal care and clothing products. Daily price, cost, and sales-volume data were collected for matched male- and female-targeted products across four categories (shampoo, lotion, cleansing foam, and clothing) over a two-month window. PED was estimated using a log-log regression, a hedonic ordinary-least-squares (OLS) model tested whether a gender price gap persisted after controlling for cost and sales volume, and regression trees were used to explore nonlinear cost structure. Female-targeted demand was found to be more price elastic than male-targeted demand (PED of -2.07 vs. -1.89), contradicting the assumption that lower elasticity explains higher prices for women. In personal care, women paid a higher markup even after accounting for higher production costs. In clothing, the pattern reversed: both cost and price were lower for female-targeted products, with comparable profit margins across gender. Overall, gendered pricing was highly heterogeneous across product types, and cost differences alone could not explain the observed price gaps.

Introduction

In the consumer market, it has long been observed that women pay more than men for comparable products, a phenomenon widely known as the "pink tax." The New York City Department of Consumer Affairs found that, across the industries surveyed, products targeted at women cost an average of 7% more than men, and that female items carried a higher price 42% of the time, compared with only 18% for male items (New York City Department of Consumer Affairs). This pattern is commonly referred to as "gender-based price discrimination." While the public often attributes it directly to discrimination, researchers have isolated two competing explanations for why it might arise. On the supply side, Moshary et al. used a national retail scanner dataset together with ingredient data to show that much of the price gap in personal care products can be traced to differences in the production cost of male- and female-targeted goods. On the demand side, Zhang and Zhang argued that gendered marketing creates a perception of necessity through messaging aligned with cultural gender expressions, which lowers price elasticity of demand (PED) and allows firms to charge more without altering production costs at all. Prior literature as a whole, however, has rarely examined these two explanations — cost and elasticity — together. This paper therefore aims to shed light on the cause of gendered pricing differences by integrating both the supply-side factor (cost of production) and the demand-side factor (PED).

Yet while Zhang and Zhang's account would predict a lower PED for women — allowing firms to raise revenue simply by charging a higher price — Barnes and Brounstein found the opposite: PED for women is actually 11% to 26% higher than for men. This finding contradicts both the demand-side account above and the inverse-elasticity prediction typically assumed in pricing theory, and it is precisely the gap this paper sets out to address: if women are more price-sensitive rather than less, why would a price gap persist at all? Building on this work, the present paper analyzes PED in relation to gender-based price discrimination. To investigate the true cause of the disparity in price, it first examines differences in production cost across categories of consumer goods, including personal care products (e.g., lotion, shampoo, cleansing foam) and clothing. It then estimates how PED differs across these same products, exploring how product differentiation helps alter consumer price sensitivity across gender.

Literature Review

Barnes and Brounstein reach a related but distinct conclusion using retail scanner and Nielsen panel data: they find that PED for females is in fact higher than for males, by 11% to 26%, meaning that women are more sensitive to price changes than men — a finding that directly contradicts both Zhang and Zhang and the inverse-elasticity prediction typically assumed in pricing theory. If female consumers are indeed more price-sensitive, standard pricing logic would predict that female-targeted products should carry a lower price, not a higher one. Barnes and Brounstein instead attribute the observed price difference to a "sorting effect": women tend to select higher-quality products within a category, and these higher-quality products carry correspondingly higher marginal costs, so the price premium women pay reflects what they buy rather than a discriminatory markup charged for who they are.

Pricing strategies have also evolved considerably with the growing availability of consumer data. Some firms have integrated machine learning methods into their pricing strategies, utilizing feature extraction to achieve personalized pricing (Perumallapalli). Others have argued that algorithm-driven pricing further enhances firms' ability to identify consumer heterogeneity and implement more precise price discrimination (Ezrachi and Stucke). While such technologies can improve market efficiency by matching prices more closely to consumer preferences, they may equally allow firms to exploit less observable characteristics and reduce transparency in pricing decisions. Machine learning therefore provides a potential mechanism through which existing gender-based price differences could be reinforced rather than corrected. Given the extensive use of recommendation systems and data-driven marketing by Taobao, the Chinese e-commerce platform analyzed in this study, algorithmic pricing offers a plausible mechanism through which gender-based price differences may emerge or be amplified, making it an important context for empirical investigation.

Methods

This study used a leading Chinese e-commerce platform (Taobao) as its data source. Product listing and transaction data were collected through automated web-scraping methods,

producing a structured extraction of public operational metrics, product listing details, and market feedback data. The finalized dataset spans the period from May 6 to July 4, 2026, capturing daily sales volume, daily average price, and daily average cost across four product categories: shampoo, lotion, cleansing foam, and clothing.

Within each category, we tracked matched pairs of gender-targeted products, choosing items that offered directly comparable male- and female-targeted versions of a similar product from the same brand, and requiring that both the male- and female-targeted listing have sufficient daily sales volume and price/cost data across the full study window to support the regression analysis. The personal care products (shampoo, lotion, and cleansing foam) were all drawn from a single brand, L'Oréal, so that brand, formulation line, and product type were held constant within this category while only gender varied. Clothing products, by contrast, were drawn from three separate brands — Champion (polo shirts), Gap (T-shirts), and Bosideng (down jackets) — with each brand supplying its own matched male and female pair; brand was therefore held constant within each clothing sub-type but varied across the sub-types themselves. This means that clothing-category comparisons combine results from three different brands, and any clothing-level pattern should be interpreted with this in mind, since cross-brand differences in positioning or pricing strategy could contribute to the observed results alongside gender.

Gender segmentation was performed using text-matching algorithms applied to product titles: products containing "Men" or related masculine identifiers were categorized as "Male," while products targeted at women were grouped under "Female." This classification provided the baseline gendered-marketing label used throughout the analysis.

To capture elasticities and stabilize variance, the key continuous variables — Daily Sales Volume, Daily Average Price, and Daily Average Cost — were transformed into their natural logarithms for the demand estimation. PED was then estimated for each product category using the following log-log specification:

$$log(Daily\ Sales\ Volume_i) = \beta_0 + \beta_1 \cdot log(Daily\ Average\ Price_i)$$

Here, β_1, the coefficient on log price, represents an estimate of PED. Comparing β_1 across genders allowed us to test whether female-targeted marketing successfully lowers price elasticity, making women less price-sensitive as a result of perceived necessity.

To determine whether gendered price differences persisted after fully accounting for production costs and consumer purchasing behavior, we implemented a multivariate OLS hedonic-style regression model across the pooled dataset:

$$Daily\_Average\_Price_i = \alpha_0 + \alpha_1 \cdot Gender_i + \alpha_2 \cdot Daily\_Average\_Cost_i + \alpha_3 \cdot Daily\_Sales\_Volume_i + \Sigma\, \gamma \cdot Category + \varepsilon_i$$

This specification allowed us to account for cost, market demand, and category-level baseline differences simultaneously. Including Daily Average Cost (α_2) filtered out legitimate

premium variation caused by raw materials, packaging, or supply-chain differences. Controlling for Daily Sales Volume (α_3) and product category fixed effects ($\Sigma \gamma_i \cdot \text{Category}_i$) accounted for baseline differences inherent to product categories and to scale economies. The key parameter of interest was α_1, the coefficient on Gender: a statistically significant α_1, after these controls, indicates a residual price gap that neither cost nor demand differences alone can explain.

We also calculated the daily gross profit margin for each gendered product using a separate set of OLS regression models:

$$Margin_t = \beta_0 + \beta_1 \cdot Volume_t + \varepsilon_t$$

Here, $Margin_t$ denotes the gross profit margin of the product for a given gender group on day t, and $Volume_t$ represents the daily sales volume of the product on day t. β_0 is the intercept term, representing the baseline gross profit margin of the product once sales-volume fluctuations are excluded; β_1 captures the marginal effect of sales-volume variation on the gross profit margin, reflecting the dilution of profitability caused by promotional activities such as large-scale discounts; and ε_t is the stochastic error term.

Finally, to investigate whether firms' pricing follows a more flexible structure than the linear specification above assumes — consistent with the machine-learning-based personalized pricing approaches documented by Dubé and Misra — we additionally fit a regression tree using category, gender, and sales volume as predictors, following the same feature-extraction logic that Dubé and Misra describe firms using to uncover segment-level cost and demand structure. In the resulting tree diagrams, each node reports the predicted cost and the share of observations it covers; the labeled condition sends an observation down the left branch if the condition holds and down the right branch otherwise, continuing until the terminal leaves, which gives the tree's final predicted cost for that subgroup.

Results



Fig 1: Combined Comparison of Price and Cost Across Categories. Differences between Price (amount charged, orange) and Cost (amount required to manufacture, blue) amongst females and males are measured across four categories: cleansing foam, clothing, moisturizing lotion, and shampoo.

Table1_Descriptive_Summary

Category	Gender	Observations_Days	Mean_Price	Mean_Cost	Total_Sales_Volume
Cleansing Foam	**Female**	60	101.68	29.97	91555
Cleansing Foam	**Male**	60	95.89	27.99	75711
Clothing	**Female**	120	179.57	61.55	64435
Clothing	**Male**	120	252.08	87.05	26399
Moisturizing Lotion	**Female**	60	236.46	79.89	13386
Moisturizing Lotion	**Male**	60	98.90	34.95	167843
Shampoo	**Female**	120	119.49	35.01	548189
Shampoo	**Male**	60	74.76	24.98	179095

Fig 2: Descriptive summary of mean price, mean cost, and total sales volume by product category and gender.

The mean price for female-targeted lotions was 236.46 RMB, approximately 2.4 times the mean price of male-targeted alternatives (98.90 RMB). Although the production cost for female-targeted lotions (79.89 RMB) was higher than that for male-targeted lotions (34.95 RMB), firms charged a substantially higher markup on female-targeted lotion than on male-targeted lotion even after accounting for this cost difference, resulting in a much larger profit margin (see Fig. 1, Fig. 2). For cleansing foam and shampoo, both price and cost were higher for female-targeted products, with shampoo carrying a markup approximately 32.7% higher than that of cleansing foam. The opposite pattern appeared in the clothing category, where

both price and production cost were higher for male-targeted products than for female-targeted products.

To further explore the role of PED in the price disparity, aggregate PED was estimated separately by gender across all categories.

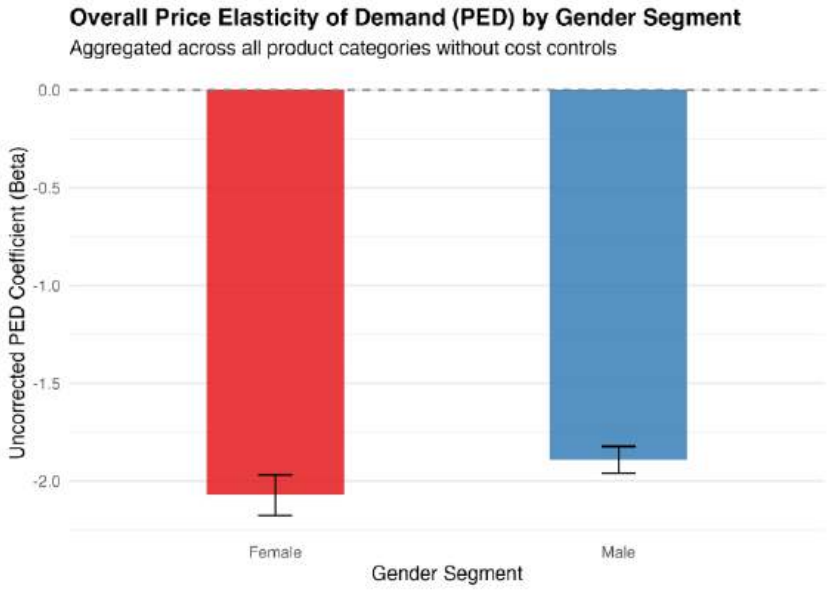


Fig 3: Overall Price Elasticity of Demand (PED) by gender, aggregated across all product categories, estimated without cost controls.

Table2_PED_Estimates_Overall

Gender	PED	SE	p_value
Male	-1.891	0.035	< 0.001
Female	-2.072	0.053	< 0.001

Fig 4: PED estimates (coefficient, standard error, and p-value) by gender, aggregated across all categories.

The female segment displayed a PED of approximately -2.07, meaning that a 1% increase in price was associated with a 2.07% drop in quantity demanded, while the male segment displayed a less negative PED of approximately -1.89, meaning that a 1% increase in price was associated with only a 1.89% drop in quantity demanded (see Fig. 3, Fig. 4).

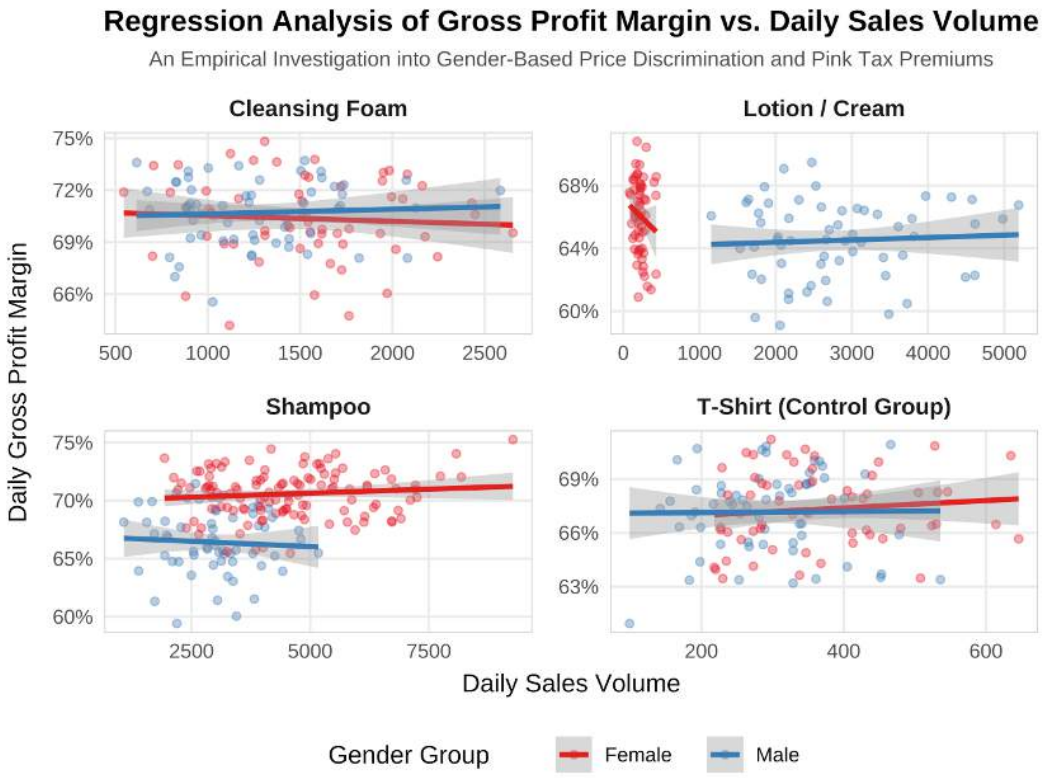


Fig 5: Regression of gross profit margin on daily sales volume, by product category and gender.

Table3_Final_Regression_Coefficients

term	estimate	std.error	statistic	p.value
(Intercept)	22.451	1.822	12.323	< 0.001
Daily_Average_Cost	2.744	0.028	99.404	< 0.001
Daily_Sales_Volume	-0.0001	0.0005	-0.199	0.842
GenderMale	-6.098	1.062	-5.742	< 0.001
CategoryClothing	-7.439	1.939	-3.837	< 0.001
CategoryMoisturizing Lotion	-9.136	1.907	-4.790	< 0.001
CategoryShampoo	-2.315	2.108	-1.098	0.273

Fig 6: Final regression coefficients from the hedonic OLS price regression, including the estimate, standard error, test statistic, and p-value for each predictor.

simplified_margin_table

Category	Gender	Average_Gross_Profit_Margin
Cleansing Foam	Female	70.52%
Cleansing Foam	Male	70.81%
Clothing (Control Group)	Female	65.67%
Clothing (Control Group)	Male	65.11%
Lotion / Cream	Female	66.22%
Lotion / Cream	Male	64.66%
Shampoo	Female	70.7%
Shampoo	Male	66.59%

Fig 7: Mean gross profit margin by product category and gender.

With respect to markup, the disparity in profit margin was significant in the personal care market, while the gender disparity in profit margin in the clothing sector was negligible. This gap persisted even after accounting for the higher production cost of female-targeted goods, as evidenced by a higher regression intercept (β_0) for female-targeted products. The baseline gross profit margins for female-targeted shampoo (70.7%) and female-targeted essence cream (66.33%) were significantly higher than those of the male control groups, by roughly 2 to 5 percentage points, while cleansing foam and clothing showed an approximately equal profit margin across genders (see Fig. 5, Fig. 6, Fig. 7).

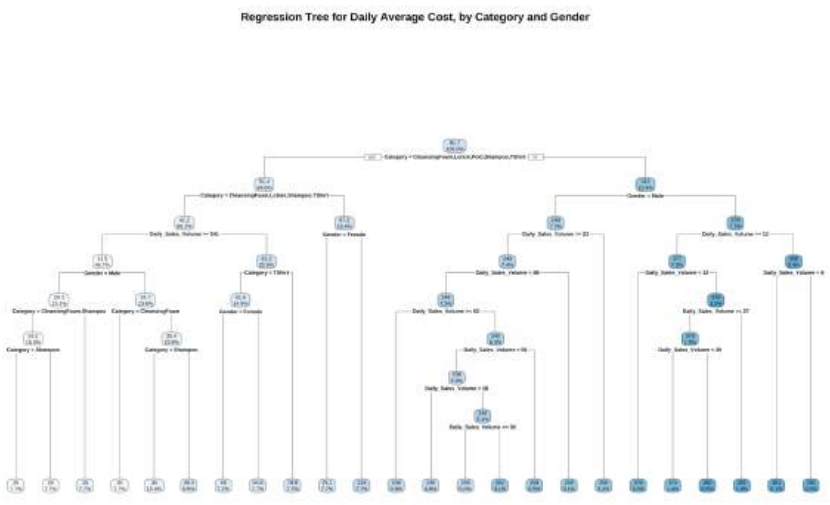


Fig 8: Regression tree for daily average cost, using category, gender, and sales volume as predictors across all six product categories (CleansingFoam, Lotion, Polo, Shampoo, TShirt, Jacket).

In the regression tree for cost (Fig. 8), Jacket was isolated at the very first split (¥91.7 root → ¥313 Jacket branch, 15.4% of the sample); within that branch, the next split was Gender = Male, separating a ¥249 male-cost node from a ¥378 female-cost node. Polo showed the same pattern on the other side of the tree: after being split off from the other four categories (cost ≈ 97.2), the next split was again Gender = Female, producing a ¥75.1 female leaf versus a ¥119 male leaf. (Polo is never printed by name on the tree; it is simply the leftover group after the tree splits off "CleansingFoam, Lotion, Shampoo, TShirt," a group we confirmed against the raw data to be 100% Polo.) For the remaining four categories (CleansingFoam, Lotion, Shampoo, TShirt), the first split in the tree was instead on Daily Sales Volume (≥541), with gender and category splits appearing only afterward.

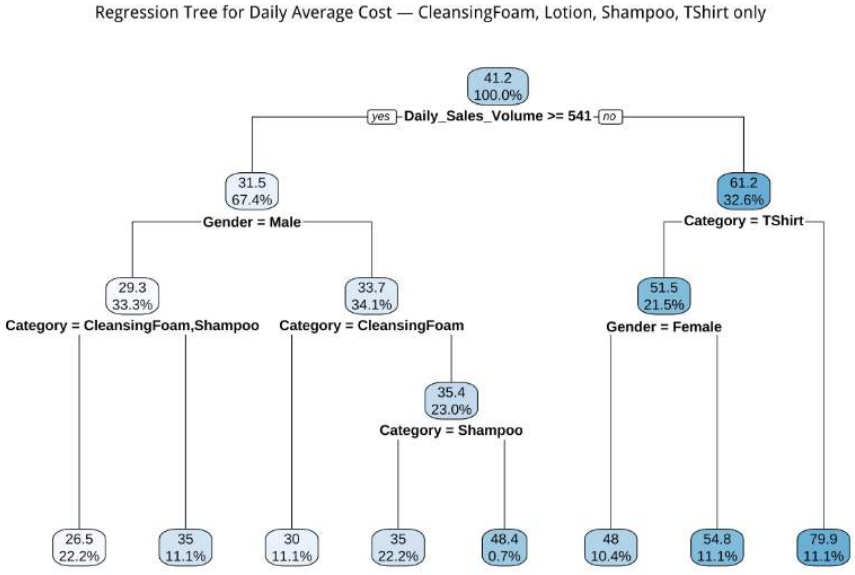


Fig 9: Regression tree for daily average cost, restricted to the four categories with mixed-direction gender gaps (CleansingFoam, Lotion, Shampoo, TShirt).

Restricting the tree to just these four categories (Fig. 9) showed that this first, volume-based split isolated Lotion-Female as a pure leaf (cost = 79.9, matching its true mean exactly); within the remaining mixed branch, gender became the very next split (Male 29.3 vs. Female 33.7), and further splits recovered the mean cost for CleansingFoam, Shampoo, and Lotion-Male individually. On the low-volume side, once Lotion-Female was set aside, TShirt was isolated and likewise split by gender (Female 48 vs. Male 54.8), reflecting its reversed gap.

Discussion

The divergence between the personal care and clothing categories shows that gendered pricing is highly heterogeneous across product types: cost alone cannot explain the difference in price across gender, and on its own does not establish the presence of gender discrimination in pricing.

Contrary to the finding by Zhang and Zhang, these results show that female-targeted products have a higher PED overall, a pattern consistent with the result reported by Barnes and Brounstein.

In the personal care market (cleansing foam, moisturizing lotion, and shampoo), one possible explanation is that female consumers often pursue higher-quality products with specific ingredients and correspondingly higher marginal costs, consistent with the "sorting effect" proposed by Barnes and Brounstein. Under this explanation, the higher price of women's personal care products would be driven primarily by higher cost. However, even after accounting for this higher cost, the markup on lotion and shampoo remained higher for women than for men. This residual gap could be attributed to gender discrimination inherent in society: firms may charge a higher price for women because they draw on underlying social norms that lead women to perceive price as roughly proportional to quality, while in practice charging a higher markup simply to increase profit. This interpretation is best understood as one plausible explanation for the residual markup gap rather than a conclusion this dataset can establish on its own, and it would benefit from direct evidence — such as survey or experimental data on consumer quality perceptions — in future work.

In the clothing market, this effect may be explained from two different angles. First, if demand for clothing from women is higher than from men — potentially due to a higher purchasing frequency for replacing existing garments — firms could achieve economies of scale, in which a larger scale of production leads to lower cost. This could be driven primarily by bulk buying, where greater purchasing volume leads to greater bargaining power with suppliers and a greater ability to negotiate down the cost of raw materials; the resulting savings would then be passed on to consumers in the form of a lower market price. Second, if more firms compete in the market for female clothing than for male clothing, the result would be a more competitive market in which each firm has less pricing power, compressing prices closer to marginal cost. We present both explanations as hypotheses for future work rather than as established findings, since this dataset does not include direct measures of purchase frequency, firm counts, or bargaining power by category.

The clothing-category results from the regression tree suggest that gendered pricing within clothing is more complex than within personal care: the divergences run in opposite directions for Jacket and Polo, and both present a large price gap between genders,despite Jacket pulling in the opposite direction, the aggregate still skews male-higher because Polo's gap is larger. As with the tree built on all six categories, the four-category tree indicates that the gender effect within CleansingFoam, Lotion, Shampoo, and TShirt is smaller relative to cost: the specific leaf values within the volume-based branch reflect a volume-filtered subset of days rather than each category's overall average, so this part of the tree is not the most reliable place to read the cost difference off directly; the category-level means reported in the Results section remain the more reliable reference for these four categories. Overall, every leaf in the four-category tree matched the real category-by-gender cost means closely, suggesting that

gender's underlying structure was recovered by the tree even in places where it was not directly visible as a labeled split.

As digital platforms have developed, machine learning has lowered the cost for firms of predicting consumer behavior. In gendered product markets, firms can use browsing behavior, purchasing history, and brand-preference interaction data to identify demand characteristics, including consumers' quality preferences and brand loyalty, and adjust pricing accordingly. Machine learning itself, however, is not the root cause of price discrimination; rather, it may amplify existing market characteristics. For instance, if female consumers form a preference for higher-quality products that is itself shaped by social norms, firms could use algorithms to identify this spending pattern and charge a higher price on that basis. As Ezrachi and Stucke argue, algorithmic pricing raises concerns precisely because it may allow firms to exploit consumer characteristics that are difficult for individuals to observe or challenge. Future research should investigate whether machine-learning-based pricing amplifies existing gender disparities.

Conclusion

After investigating production cost, PED, and profit margin for gendered products, this paper finds that different market segments present markedly different pictures for men and women. Aligning with the result reported by Barnes and Brounstein, a higher PED for women suggests that women were more price-sensitive than men within this sample. In the personal care market, a higher production cost together with a higher PED and a higher markup suggests that women tend to choose personal care products of higher quality, and that firms may operate with a degree of gender-based pricing in the shampoo and lotion markets, charging female-targeted products a higher markup despite their higher PED. In the clothing market, this paper's results contradict previous studies, as both cost and price for female-targeted products were lower, with an approximately equal profit margin for male- and female-targeted products. This pattern could be attributed to economies of scale driven by greater demand and a higher frequency of replacement, which may in essence stem from bulk buying, with the resulting savings passed on to consumers in the form of a lower price. This study is limited by its two-month window on a single platform, its combination of three different clothing brands within one category, and its lack of a direct control for a major promotional period that overlapped the study window; future research with a larger, multi-brand, multi-platform sample would help establish how general these patterns are.

Works Cited

Barnes, K., and J. Brounstein. “The Pink Tax: Why Do Women Pay More?” University of California, Berkeley, 2023.

Dubé, Jean-Pierre, and Sanjog Misra. “Personalized Pricing and Consumer Welfare.” Journal of Political Economy, vol. 131, no. 1, 2023, pp. 131-89.

Ezrachi, Ariel, and Maurice E. Stucke. Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy. Harvard UP, 2016.

Moshary, S., et al. “Gender-Based Pricing in Consumer Packaged Goods: A Pink Tax?” Marketing Science, vol. 45, no. 1, 2023, pp. 1-14.

New York City Department of Consumer Affairs. “From Cradle to Cane: The Cost of Being a Female Consumer.” 2015, www.nyc.gov/assets/dca/downloads/pdf/partners/Study-of-Gender-Pricing-in-NYC.pdf.

Perumallapalli, R. K. “Machine Learning Algorithms for Dynamic Pricing Optimization in Retail.” International Journal of Research and Analytical Reviews, vol. 1, no. 4, 2014, pp. 642-49.

Zhang, R., and Y. Zhang. “The Pink Tax: Gender-Based Pricing Strategy and Underlying Mechanism.” 2025 AAEA & WAEA Joint Annual Meeting, 2025, Denver, CO, ageconsearch.umn.edu.

Analyzing the Effectiveness of Different Monetary Policy Tools in Stabilizing Inflation During Economic Crises Using Time Series Analysis By Sahana Thothadri

1. Abstract

The main purpose of this review is to analyze and evaluate the effectiveness of monetary policy in controlling inflation during various types of economic crises. It also takes a look at how time-series methods influence empirical findings and conclusions. The review examines how monetary policy affects inflation in many different crisis situations and how time-series methods impact the measurement of these effects, analyzing whether or not methodological choices change conclusions about the effectiveness of policy. The review includes conventional tools such as interest rate adjustments, along with unconventional tools such as quantitative easing, reserve requirements, negative interest rates, forward guidance, and yield curve control, and compares the effectiveness of both conventional and unconventional monetary policy tools throughout different crisis conditions. Financial crises, pandemics, recessions, and supply-shock crises are all considered; these crises differ in what causes them and in how inflation responds to monetary policy.

The evidence supports the idea that monetary policy is generally effective, but how successful it is depends on the source of inflation, institutional conditions, and the severity of the crisis. Policy outcomes improve once central banks adapt their tools based on how economic circumstances are changing. Interest rate adjustments remain the most effective tool when controlling inflation that is demand driven. During severe crises, forward guidance and quantitative easing can provide additional support when conventional policy becomes restrained. Monetary policy is less effective when inflation is driven by supply shocks or when there have been major disruptions within the financial markets. It also becomes less effective when conventional policy is restricted by the zero lower bound and during periods of severe financial instability.

Advanced economies generally experience stronger monetary policy outcomes because they have better developed financial systems and more credible central banks, while growing economies often deal with greater challenges due to higher economic volatility and weaker institutions. Countries with stronger central bank credibility and institutional quality generally achieve better inflation stabilization compared to those with weaker institutions. Different time series methods depend on different assumptions and can produce various estimates of monetary policy effectiveness, so choosing an appropriate method enhances the reliability of policy conclusions. Nonlinear and structural break models often capture crisis dynamics more accurately than traditional linear models, resulting in more credible and reliable estimates of monetary policy effects.

The review identifies limited research on unconventional monetary policy tools, crisis specific comparisons, and the joint use of many policy tools, and it emphasizes the need for more studies using advanced time series methods. Existing research provides very minimal focus on unconventional monetary tools, comparison across different crisis types, and how multiple policy

tools work together. The review suggests that central banks should adapt monetary policy to the primary causes of inflation rather than depending on one single policy approach; flexible and targeted strategies help stabilization during crises. During the early stages of a crisis, central banks should focus on liquidity support, clear communication, and emergency lending to help stabilize markets. As inflationary pressures rise, they should move toward interest rate adjustments and other appropriate tools when managing fiscal or supply side policies.

2. Introduction

Inflation control is uniquely challenging during economic crises because policymakers must balance stabilizing prices while also supporting economic growth and employment. Raising interest rates or cutting spending are procedures that could reduce inflation; however, they can heighten unemployment and hinder recovery during an economic crisis as well. Simultaneously, crisis conditions like supply disruptions or panic buying could increase prices even higher during a time period when the economy can't handle that. This creates a complicated situation where certain actions may help one problem while also creating or intensifying another ("Energy Security and Economic Stability").

Crises disrupt the normal transmission of monetary policy by weakening the pathways through which policy changes affect the economy, like investment and consumer spending. Financial instability can cause households to try to save more and banks to start restricting credit, which reduces the effects of lower interest rates. Businesses tend to become less responsive to policy signals when there is uncertainty, causing hiring and investment delays. Due to this, banks may need to utilize tools such as emergency lending programs to influence economic activity (Federal Reserve Bank of New York).

Time series analysis is well suited to studying monetary policy because it analyzes how economic variables change over time. This allows researchers to follow the cause and effect relationship between policy measures and its results such as inflation and employment. Because monetary policy usually takes months or even years to truly influence the economy, time series helps to identify lags. By examining trends and shocks, economists can differentiate fluctuations from long term results (Tableau).

Current reviews of monetary policy effectiveness face limitations such as trouble separating policy effects from other economic factors like global shocks, supply disruptions, and fiscal policy. Measurement faults and data restrictions can also add to the difficulty of accurately assessing long term outcomes. The effects of policy frequently vary across time periods, financial situations, and countries, which can lessen how widely conclusions can be applied (Center for American Progress).

This review differs by strongly emphasizing methodology, as it closely examines how various systematic approaches shape conclusions about the effectiveness of monetary policy. It is also crisis specific and focuses on how policy tools work during periods of economic stress versus regular conditions. It also contrasts different policy mechanisms, such as interest rate changes, within the same layout in order to gauge their strengths and weaknesses.

3. Conceptual Framework: Monetary Policy and Inflation Dynamics

Monetary policy theoretically influences inflation through various transmission channels. Through the expectations channel, credible policy signals form how households and firms expect future inflation, affecting pricing and current wage settings. Through the interest rate channel, central banks lower or raise policy rates in order to influence borrowing costs. The credit channel occurs when tighter policy makes banks lend less, which limits access to loans and reduces economic activity. The exchange rate channel works because higher interest rates strengthen the currency, making imports cheaper and lowering costs overall (International Monetary Fund, "Monetary Policy").

Crisis conditions can weaken or amplify these channels. The interest rate channel may weaken if rates are already close to zero, which leaves little room for central banks to stimulate the economy any more. The credit channel could weaken as banks may become risk averse and limit lending even when policy supports borrowing. Simultaneously, the expectations channel can become stronger and more powerful, since clear central bank communication can greatly shape confidence and inflation expectations during times of uncertainty. The exchange rate channel may expand if investors quickly move capital between countries at times of crisis, which causes larger currency fluctuations that affect prices (Firstup).

During crises, the assumption of rational expectations can break down, as people's actions are based on fear, uncertainty, or incomplete information, rather than hypothesizing future economic conditions. Investors and consumers may overreact to bad news, which can result in panic selling or unexpected drops in spending. The idea of market efficiency also weakens, because prices may no longer reflect all available information. Instead, they can be driven by liquidity shortages, herd behavior, or guesswork. Consequently, markets can become tense and mispriced, leading to the effects of monetary policy being less predictable (CAINZ).

Lag structures complicate inflation responses because monetary policy does not affect the economy right away. There are recognition lags, or the time to realize inflation is rising; decision and execution lags, or the time for policymakers to decide on and apply policy; and impact lags, or the time for interest rate changes to affect spending, prices, and investment. These effects come to light very slowly, so inflation may continue rising even after the policy securing process has begun. This delay also makes it difficult for central banks to align their actions precisely, causing them to over or under correct (Federal Reserve Bank of St. Louis).

4. Overview of Monetary Policy Tools

The literature examines both conventional and unconventional monetary policy tools. Conventional tools include policy interest rates, reserve requirements, and open market operations, while unconventional tools include quantitative easing (QE), forward guidance, negative interest rates, and yield curve control.

4.1 Conventional Tools

Conventional monetary policy typically affects inflation gradually in normal conditions, as effects appear only after about one to two years due to the gradual transmission through interest rates, credit markets, and expenditures. Financial markets and banks function normally, so because of this, policy signals go through the economy in a somewhat predictable manner. During periods of crisis, transmission can tend to become weaker or disturbed as banks limit lending and financial markets malfunction; however, exchange rates and expectations may react faster, causing inflation rates to be more volatile (Federal Reserve Bank of San Francisco).

Interest rates are oftentimes not symmetric between tightening and easing cycles. When central banks raise rates, or tighten, borrowing becomes more expensive, which can rapidly slow down investment and spending. Lower rates, or easing, may be slower due to households and businesses still being careful about borrowing, especially during times of economic crisis. Financial discord, expectations, and debt can make easing have less immediate effects than tightening (Washington Center for Equitable Growth).

Evidence of diminished effectiveness at the zero lower bound comes from times when central banks lowered interest rates to close to 0%, however economic growth and inflation were still weak. For example, after the 2008 financial crisis, the Federal Reserve kept interest rates near zero for years, but despite this, recovery was slow and inflation stayed below the aim. This shows that when rates are very low, further cuts don't do much to facilitate spending and borrowing. Because of this, central banks must rely on a nontraditional strategy, quantitative easing, to support the economy (National Bureau of Economic Research).

4.2 Unconventional Tools

Unconventional monetary policy tools, such as quantitative easing, primarily influence inflation expectations by showing the central bank's long term commitment to restoring the economy. These tools affect how households, businesses, and investors expect the behavior of future inflation, even if current inflation doesn't change right away. Headline inflation has a slower responding rate most of the time due to its dependence on actual economic activity and prices in the real economy. Because of this, unconventional policies tend to shift expectations at first, which can influence real inflation later through spending, investment, and other financial decisions ("Inflation Expectations during the Great Recession").

Statistically significant effects from unconventional monetary policy on inflation expectations typically appear over medium to long term horizons. Studies have found that market based measures, such as bond breakeven inflation rates, respond within days or even weeks after policy announcements, revealing an instant shift in expectations. Measurable effects on actual headline inflation, however, take several quarters to years because of transmission lags within the economy. These results show that expectations are easily adjustable, while real inflation happens at a slower pace ("Unconventional Monetary Policy and Inflation Expectations").

Unconventional monetary policy tools tend to be more effective in higher leveled economies than in emerging markets. Advanced economies usually have more reliable central banks, more secure institutional frameworks, and deeper financial markets. These make policy

signals such as forward guidance more trusted by households and investors. In rising markets, higher inflation volatility, weaker financial systems, and concerns about currency depreciation can lower the efficiency of these tools. Consequently, unconventional policies may influence expectations less strongly and create capital outflows or financial instability instead of balancing expectations and inflation (International Monetary Fund, Working Paper WP/09/226).

5. Economic Crisis Typologies in the Literature

Inflation responses differ because the way each type of crisis affects the economy varies depending on the crisis. During financial crises and recessions, demand typically tends to fall, causing central banks to lower interest rates and increase spending to prevent deflation and promote growth. During pandemics and supply shock crises, inflation is a result not of weak demand, but more so of disrupted supply chains or shortages; therefore raising interest rates may reduce inflation more slowly, but also risks lower economic growth. Supply shock inflation can be difficult, because policymakers have to balance controlling prices along with avoiding unemployment and recession (Congressional Research Service).

Monetary tools are generally more effective against demand driven inflation, because higher interest rates directly reduce consumer borrowing, spending, and investment, lowering the overall demand in the economy. In supply driven inflation, prices rise because of either shortages or higher production costs, so simply raising interest rates won't fix the underlying supply problem. Stricter monetary policy, however, could slow economic growth while inflation is still high, creating a stagflation risk. Because of this, supply driven inflation tends to require a mix of monetary policy and supply side solutions, like improving production or lowering supply chain disturbances (Bank for International Settlements, Quarterly Review).

Structural breaks in inflation behavior are most commonly seen during pandemics, major supply shock crises, and detrimental financial crises that permanently change economic conditions. For example, the 1970s oil shocks changed inflation dynamics by establishing prolonged cost-push inflation, while the COVID-19 pandemic interrupted labor markets and supply chains, causing a shift in inflation patterns. Financial crises such as the Global Financial Crisis can also lead to structural breaks by altering lending, spending, and monetary policy behavior for years afterward. These events tend to cause inflation relationships that held before to weaken or change drastically, needing new economic models and policy approaches ("Structural Breaks in Inflation Dynamics").

Central bank credibility and institutional strength help anchor inflation expectations and increase confidence from the public in economic policies. When people trust that a central bank will control inflation, businesses along with their workers are less likely to raise prices and wages significantly, making inflation easier to take care of. Powerful institutions also allow policymakers to respond more efficiently and frequently during crises. Consequently, countries with strong institutions and credible central banks often experience lower and shorter periods of inflation than those with weaker or less powerful institutions ("Central Bank Credibility and Institutional Resilience").

6. Time Series Methodologies Used

The literature relies on several families of econometric approaches: traditional time series models such as ARIMA and distributed lag models; multivariate models such as Vector Autoregression (VAR) and Structural VAR (SVAR); regime-switching and nonlinear models, including Markov-switching models and threshold VARs; and cointegration and long-run analysis, including Johansen cointegration and error correction models.

6.1 Traditional Time Series Models

Linear models often struggle with crisis data because they assume that relationships between variables remain constant over time. During crises, economic behavior could potentially change suddenly, leading to nonlinear effects, structural breaks, and severe outcomes that linear models may not grasp accurately. They also tend to underestimate feedback loops and threshold effects, where small changes can evoke huge economic responses. Because of this, linear models may create unreliable predictions and policy suggestions during periods of great economic disturbance ("Limitations of Linear Models Applied to Crisis Data").

Results can be very sensitive to stationarity assumptions and lag selection, because different lag lengths may capture or miss vital delayed effects between economic variables. Choosing too few lags can leave out relevant information, while too many lags can reduce the efficiency of the model and increase noise. Likewise, if non-stationary data are incorrectly treated as stationary, the model may create inaccurate or misleading relationships. Researchers typically do robustness checks using many lag structures and stationarity tests to make sure that their findings are correct and reliable ("Lag Selection and Stationarity in Time Series Models").

6.2 Multivariate Models

Studies identify monetary policy shocks by isolating unforeseen changes in central bank actions that are not just responses to current economic situations. Common methods include analyzing unexpected changes in policy interest rates, using econometric models like Vector Autoregressions, and looking at market reactions shortly after central bank announcements. Some studies use high frequency data to separate anticipated policy moves from unexpected ones. These methods allow researchers to estimate the effects of monetary policy on factors such as employment, output, and inflation ("Deciphering Monetary Policy Shocks").

The most defensible identification strategies are those that isolate unexpected monetary policy changes and reduce bias from other economic factors. High frequency identification using market responses during central bank announcements is considered one of the most effective approaches, considering it grasps policy surprises rather than expected choices. Researchers also tend to use structural VAR models with reasonable restrictions and external instruments to find causal effects. Merging various detection methods and carrying out robustness checks generally provides the most reliable outcomes ("Identification of Monetary Policy Shocks").

Impulse response functions often vary across crisis periods in aspects such as size and duration. During financial crises and recessions, monetary policy shocks tend to have weaker or

slower effects due to the fact that credit markets are disturbed and interest rates may already be nearing zero. On the other hand, during supply shock crises, inflation typically responds more consistently to shocks, while output may decline simultaneously, indicating stagflationary pressures. These differences reveal that the economy's reaction to policy is state dependent, suggesting that the same monetary policy will produce varying outcomes depending on the type of crisis ("State-Dependent Impulse Responses to Monetary Policy Shocks").

6.3 Regime-Switching and Nonlinear Models

Evidence for nonlinear inflation responses comes from studies showing that inflation doesn't react proportionally to economic shocks during crises. Major disturbances, like pandemics or large supply shocks, often result in stronger and more prolonged inflation effects than minor shocks. Researchers using nonlinear models, such as threshold and regime switching models, notice that inflation acts differently depending on how severe the crisis is and the state of the economy during the crisis. This shows that traditional linear models may underestimate inflation relationships during times of economic stress ("The Nonlinear Phillips Curve Helps Explain the Inflation Overshoot").

Regime shifts alter policy effectiveness by changing how the economy responds to monetary policy. For example, during periods of financial instability, high inflation, or supply disruptions, the same policy action could potentially have stronger, weaker, or delayed effects than under regular conditions. These shifts can change consumer and business expectations, making economic activity along with inflation respond in a different way. Due to this, policymakers sometimes need to adjust their strategies to meet the needs of the new economic environment ("Regime Shifts and Policy Effectiveness").

Inflation dynamics can remain persistent even after a crisis regime ends, though the extent varies depending on the type of crisis. Wage growth, inflation expectations, and supply chain disruptions may continue to influence prices for a long period after the initial shock has subsided. However, reliable monetary policy and improving economic conditions can slowly bring inflation back toward its goal. As a result, post crisis inflation often goes away over time, but may remain high after intense or supply driven crises ("Why Inflation Sticks Around").

6.4 Cointegration and Long-Run Analysis

Monetary tools can have long run equilibrium relationships with inflation because they influence borrowing, spending, and inflation expectations over time. While the short term effects of monetary policy may differ, many studies have found that interest rates and the money supply are connected to inflation in the long run. Oftentimes, researchers use cointegration analysis to gauge whether these variables share a stable long term relationship. The strength of this relationship, however, relies on factors such as economic structure, type of crisis, and central bank credibility (Bernanke).

The speed at which inflation returns to equilibrium after shocks depends on what kind of shock it was, the efficiency of monetary policy, and the overall economic state. Demand driven

shocks tend to resolve quicker because central banks can influence spending through interest rate changes. On the other hand, supply driven shocks, such as energy shortages, often take longer to settle because monetary policy cannot directly confront the primary causes. Consequently, inflation may return to equilibrium within months or sometimes years, depending on the shock severity ("Why Inflation Sticks Around").

7. Empirical Findings Across Studies

Interest rate adjustments are the monetary tool regularly associated with reducing inflation volatility because they help stabilize inflation expectations and manage demand. Central bank communication and forward guidance also reduce volatility by making policy more predictable in the future. In some instances, balance sheet policies, such as quantitative tightening, can bolster price consistency when conventional tools are limited. These tools are the most effective when they are being implemented by credible central banks with reliable policy frameworks (Investopedia).

During severe crises, traditional monetary tools, such as modest interest rate adjustments, are typically inadequate to balance inflation, especially when inflation is motivated by financial disturbances or supply shocks. Raising or lowering interest rates can't directly solve supply chain disruptions, shortages, or rises in food or energy expenses. In these circumstances, monetary policy itself isn't as effective and may need to be accompanied by supply side policies and fiscal measures. Stabilizing inflation during times of crisis often requires a more extensive policy response (Chicago Booth Review).

Results can vary depending on country income level, because higher income countries usually have stronger institutions, better designed financial markets, and more credible central banks. These factors make monetary policy more successful at managing inflation and stabilizing the economy. Conversely, low and middle income countries experience larger and more constant inflation due to weaker institutions, lack of policy tools, and greater vulnerability to external shocks. As a result, inflation tends to be harder to manage in lower income countries (International Monetary Fund, "Monetary Policy and Inflation across Income Levels").

Expectations and central bank credibility play a major role in empirical outcomes, because they influence how businesses and households respond to monetary policy. When a central bank is labeled as credible, inflation expectations remain secure and stable, making it easier to control inflation with only small policy changes. However, if credibility is weak, people may anticipate higher inflation and adjust wages and prices appropriately, making inflation more relentless. Empirical studies have regularly found that stronger credibility increases the effectiveness of monetary policy and leads to better, more stable inflation results (BSI Economics).

8. Cross-Country and Institutional Comparisons

Independent central banks are better at keeping inflation under control, because they can make policy decisions based on economic conditions rather than political pressures. They are

more likely to keep inflation expectations well secured and maintain stable prices, resulting in lower and more balanced inflation over time. Politically influenced central banks, on the other hand, may postpone required policy actions or focus on short term electoral or financial goals, resulting in higher and more tenacious inflation. Studies conclude that greater central bank independence is linked to higher economic stability and stronger inflation performance (European Central Bank).

Inflation outcomes are better in inflation-targeting regimes, because they offer a direct policy goal and help anchor inflation expectations. Countries with inflation-targeting systems tend to face lower and more stable inflation, particularly when backed by a credible and independent central bank. These regimes also make monetary policy more foreseeable and enhance transparency, expanding popular support and confidence. However, their effectiveness is highly dependent on strong institutions and may be lessened in the face of major supply shocks ("Inflation Targeting and Inflation Outcomes").

Exchange rate regimes affect policy transmission by guiding how monetary policy influences inflation and economic activity. In flexible exchange rate regimes, central banks have greater autonomy when it comes to adjusting interest rates, which as a result makes monetary policy more effective. Conversely, in fixed exchange rate regimes, policy decisions are typically confined by the necessity to maintain the currency peg, lowering the central bank's capability to respond to domestic inflation. Monetary policy tends to have more direct and stronger effects in countries with flexible exchange rates, rather than countries with fixed ones (Reserve Bank of Australia).

9. Methodological Limitations in Existing Literature

Endogeneity and reverse causality can make it difficult to find out whether monetary policy causes changes in inflation or just responds to them. For example, central banks regularly adjust interest rates in response to increasing inflation, making the relationship between the two factors bidirectional. If these issues are not acknowledged, study results could potentially over or understate the true outcomes of monetary policy. Researchers, therefore, make it a practice to use identification strategies, such as structural models or instrumental variables, to strengthen the credibility of their verdicts ("Endogeneity and Reverse Causality in Monetary Policy Analysis").

Policy shocks are not always truly exogenous in crisis studies because central banks tend to change policies as a result of changing economic circumstances. During crises, policy decisions react to unemployment, financial stability, and inflation, making it harder to separate the results of the policy from the actual crisis itself. To handle the issue, researchers use identification methods or structural models to help isolate sudden policy changes. These strategies improve causal analysis, but can't get rid of worries about endogeneity completely ("Exogeneity of Monetary Policy Shocks during Crises").

Short crisis sample periods can introduce bias by offering insufficient observations to correctly predict long-term relationships. They may exaggerate temporary or unusual events, making results less representative of more comprehensive economic patterns. Small samples can

also lower statistical power, which increases the probability of unstable or inaccurate predictions. Conclusions drawn from short crisis periods may not apply well to other crises or even normal economic situations ("Small-Sample Bias in Macroeconomic Estimation").

Structural breaks are frequently not fully taken care of in empirical studies, especially when researchers depend on linear models that assume consistent relationships over time. Ignoring structural breaks can contribute to biased estimates and false conclusions because economic relationships are prone to change during times of crisis. Many recent studies consider these changes using structural break tests, regime switching models, or time varying parameter models. The degree to which structural breaks are properly handled, however, changes across studies ("Forecasting Structural Breaks in R").

10. Research Gaps and Future Directions

Numerous monetary tools continue to be under-studied in crisis contexts, especially unconventional policies such as negative interest rates, forward guidance, and yield curve control. While these tools have gathered more attention since COVID-19 and the Global Financial Crisis, evidence on their long term effectiveness during various types of crisis situations remains limited. Researchers have also given less regard to how combinations of monetary tools perform across changing crisis conditions. Additional research is required to gain a better understanding of which policy tools are the most effective during periods of major economic disturbance and why (Bank for International Settlements, "A Monetary Lifeline").

High-frequency data improve identification by acquiring financial market reactions right before and right after monetary policy announcements. This enables researchers to separate unexpected policy shocks from other economic situations taking place over longer periods. Due to this, high-frequency data lower endogeneity concerns and provide more accurate predictions of the policy effects. This leads to more trustworthy conclusions about how monetary policy impacts inflation and other economic results ("High-Frequency Identification of Monetary Policy Effects").

Machine learning-augmented time series models can improve analysis of inflation by identifying intricate, nonlinear relationships that conventional econometric models may skim past. They can better identify structural breaks, interactions among economic factors, and regime switches, resulting in more reliable forecasts during crises. These models can also acclimatize better to changing economic states as new data becomes accessible. As a result, they have the potential to improve inflation predicting and monetary policy analysis in environments of uncertainty (Comet).

Future studies should incorporate expectations data by including measures of inflation expectations from households, financial markets, businesses, and forecasting professionals. These data can allow researchers to better understand how expectations affect inflation behaviors and how effective monetary policy really is in times of crisis. Integrating expectations data with traditional macroeconomic indicators can boost model precision and make causal analysis

stronger. This approach allows for a more extensive understanding of how expectations influence inflation outcomes (Board of Governors of the Federal Reserve System).

11. Policy Implications

During the early stages of a crisis, central banks should prioritize emergency lending establishments, liquidity support, and forward guidance to help financial markets remain stable and confident. As the crisis intensifies and inflationary pressures start to rise, they should redirect focus to tools such as quantitative tightening, interest rate adjustments, and balance sheet normalization to help keep inflation under control and uphold long term economic stability. The proper mix of tools depends on whether the crisis is a result of demand or supply shocks. Using the right tools for each stage is crucial for improving the effectiveness of monetary policy (Committee on the Global Financial System).

Communication strategies should complement quantitative tools by clearly explaining policy decisions, goals, and future direction to the public. Open communication helps reduce uncertainty, support inflation expectations, and strengthen the efficiency of monetary policy efforts. It also increases public confidence in the central bank, making businesses and households more inclined to respond as desired. Together, transparent communication and quantitative policy tools improve economic stability during crises and the recovery periods following (Economic Statistics Centre of Excellence).

Current inflationary environments emphasize the importance of adjusting monetary policy to the root causes of inflation rather than depending on one singular approach. When inflation is motivated by demand, interest rate increases are typically effective, while supply driven inflation may need accompanying fiscal or supply-side policies. Upholding central bank credibility and clear communication is vital for supporting inflation expectations and reducing uncertainty. These lessons can aid policymakers in responding better to future circumstances of high inflation (Center on Budget and Policy Priorities).

12. Conclusion

The evidence collectively suggests that monetary policy is effective for the most part; however, its impact depends on the severity of the crisis, the type of crisis, the strength of financial establishments, and the source of inflation. Interest rate changes typically work the best when inflation is driven by elevated demand, while they are less effective during supply disturbances. Central bank credibility, well-backed expectations, and clear communication can bolster policy outcomes. Monetary policy is most successful when tools are altered to economic conditions and anchored by supply-side or fiscal measures (International Monetary Fund, “Monetary Policy”).

Monetary policy is less effective at stabilizing inflation when inflation is driven by supply-side shocks. It may also be less effective during intense financial crises or when interest rates are near the zero lower bound, restricting the central bank’s capability to support or slow the economy through traditional monetary policy. Additionally, unsupported inflation

expectations and weak central bank credibility can diminish the impact of policy decisions. Under these conditions, supply-side or complementary fiscal measures must be used in order to revive inflation control (Bank for International Settlements, Working Paper No. 1349).

Methodological choice matters because various econometric models rely on different assumptions and can generate different estimates of the effectiveness of monetary policy. Some methods have better results when addressing structural breaks, endogeneity, nonlinear relationships, and crisis-specific conditions, while others may fail to account for these factors and produce biased results. The choice of methodology is important because it can influence policy recommendations and the findings policymakers draw from empirical evidence. Selecting the best method for each crisis leads to more reliable findings and better supported monetary policy decisions (International Monetary Fund, Working Paper WP/06/80).

Works Cited

Bank for International Settlements. "A Monetary Lifeline: Central Banks' Crisis Response." BIS Annual Economic Report 2020, June 2020, www.bis.org/publ/arpdf/ar2020e2.htm.

Bank for International Settlements. BIS Quarterly Review, Dec. 2024, www.bis.org/publ/qtrpdf/r_qt2412d.htm.

Bank for International Settlements. BIS Working Papers No. 1349, www.bis.org/publ/work1349.htm.

Bernanke, Ben S. "Inflation Expectations and Inflation Forecasting." Speech at the NBER Summer Institute, Board of Governors of the Federal Reserve System, 10 July 2007, www.federalreserve.gov/newsevents/speech/bernanke20070710a.htm.

Board of Governors of the Federal Reserve System. "Lead Article." Federal Reserve Bulletin, Apr. 1997, www.federalreserve.gov/pubs/bulletin/1997/199704lead.pdf.

BSI Economics. "Self-Fulfilling Prophecies, Bias and Economics: Individual Decisions and Their Impact on Economic Activity." bsi-economics.org/self-fulfilling-prophecies-bias-and-economics-individual-decisions-and-their-impact-on-economic-activity-note/.

CAINZ. "Herd Behaviour and Market Overreaction." CAINZ Digest, www.cainz.org/13758/.

Center for American Progress. "The Limitations of Monetary Policy as a Financial Stability Tool." www.americanprogress.org/article/the-limitations-of-monetary-policy-as-a-financial-stability-tool

Center on Budget and Policy Priorities. "Inflation: Key Measures to Understand." www.cbpp.org/research/economy/inflation-key-measures-to-understand.

"Central Bank Credibility and Institutional Resilience." International Journal of Central Banking, vol. 21, no. 4, www.ijcb.org/journal/v21n4/central-bank-credibility-and-institutional-resilience.

Chicago Booth Review. "What Makes It Hard to Control Inflation." www.chicagobooth.edu/review/what-makes-it-hard-control-inflation.

Comet. "Understanding the Vast Applications of Time Series Analysis in Machine Learning." www.comet.com/site/blog/understanding-the-vast-applications-of-time-series-analysis-in-machine-learning-with-comet/.

Committee on the Global Financial System. Unconventional Monetary Policy Tools: A Cross-Country Analysis. CGFS Papers No. 63, Bank for International Settlements, 2019, www.bis.org/publ/cgfs63.pdf.

Congressional Research Service. Inflation in the U.S. Economy: Causes and Policy Options. CRS Report R47273, www.congress.gov/crs-product/R47273.

"Deciphering Monetary Policy Shocks." VoxEU, Centre for Economic Policy Research, cepr.org/voxeu/columns/deciphering-monetary-policy-shocks.

Economic Statistics Centre of Excellence. "Improving the Communication of Economics and Economic Statistics." www.escoe.ac.uk/improving-the-communication-of-economics-and-economic-statistics/.

"Endogeneity and Reverse Causality in Monetary Policy Analysis." shinkangecon.github.io/KP1.pdf.

"Energy Security and Economic Stability in the Age of Surging Inflation." Energy Economics, 2023, ScienceDirect, www.sciencedirect.com/science/article/pii/S0140988323004474.
European Central Bank. "Central Bank Independence and Inflation Performance." The ECB Blog, 23 Dec. 2025, www.ecb.europa.eu/press/blog/date/2025/html/ecb.blog.20251223~aad70ce537.en.html.
"Exogeneity of Monetary Policy Shocks during Crises." Journal of Macroeconomics, 2022, ScienceDirect, www.sciencedirect.com/science/article/abs/pii/S0164070422000246.
Federal Reserve Bank of New York. "Monetary Policy Transmission Before and After the Crisis." Liberty Street Economics, June 2016, libertystreeteconomics.newyorkfed.org/2016/06/monetary-policy-transmission-before-and-after-the-crisis/.
Federal Reserve Bank of San Francisco. "How Quickly Do Prices Respond to Monetary Policy?" FRBSF Economic Letter, Apr. 2024, www.frbsf.org/research-and-insights/publications/economic-letter/2024/04/how-quickly-do-prices-respond-to-monetary-policy/.
Federal Reserve Bank of St. Louis. "Examining Long and Variable Lags in Monetary Policy." The Regional Economist, May 2023, www.stlouisfed.org/publications/regional-economist/2023/may/examining-long-variable-lags-monetary-policy.
Firstup. "Common Challenges of Crisis Communications and How to Overcome Them." firstup.io/blog/common-challenges-of-crisis-communications-how-to-overcome-them/.
"Forecasting Structural Breaks in R." r-statistics.co, r-statistics.co/Forecasting-Structural-Breaks-in-R.html.
"High-Frequency Identification of Monetary Policy Effects." Research in International Business and Finance, 2023, ScienceDirect, www.sciencedirect.com/science/article/pii/S0275531923002040.
"Identification of Monetary Policy Shocks." Economics Letters, 2025, ScienceDirect, www.sciencedirect.com/science/article/pii/S0165176525005233.
"Inflation Expectations during the Great Recession." Journal of International Economics, 2020, ScienceDirect, www.sciencedirect.com/science/article/pii/S0022199620300167.
"Inflation Targeting and Inflation Outcomes." Journal of International Money and Finance, 2012, ScienceDirect, www.sciencedirect.com/science/article/abs/pii/S0261560612001556.
Institute for New Economic Thinking. "Why Inflation Sticks Around: The Social Roots of Price Persistence." www.ineteconomics.org/perspectives/blog/why-inflation-sticks-around-the-social-roots-of-price-persistence.
International Monetary Fund. "Monetary Policy and Inflation across Income Levels." IMF eLibrary, 2025, www.elibrary.imf.org/view/journals/006/2025/001/article-A001-en.xml.

International Monetary Fund. “Monetary Policy: Stabilizing Prices and Output.” Finance & Development, www.imf.org/en/publications/fandd/issues/series/back-to-basics/monetary-policy.
International Monetary Fund. IMF Working Paper WP/06/80, 2006, www.imf.org/external/pubs/ft/wp/2006/wp0680.pdf.
International Monetary Fund. IMF Working Paper WP/09/226, 2009, www.imf.org/external/pubs/ft/wp/2009/wp09226.pdf.
Investopedia. “What Methods Can the Government Use to Control Inflation?” www.investopedia.com/ask/answers/111314/what-methods-can-government-use-control-inflation.asp.
“Lag Selection and Stationarity in Time Series Models.” ScienceDirect, www.sciencedirect.com/science/article/pii/S1093968726008224.
“Limitations of Linear Models Applied to Crisis Data.” PubMed Central, National Library of Medicine, pmc.ncbi.nlm.nih.gov/articles/PMC12186762/.
National Bureau of Economic Research. NBER Working Paper No. 22351, 2016, www.nber.org/papers/w22351.
“The Nonlinear Phillips Curve Helps Explain the Inflation Overshoot.” VoxEU, Centre for Economic Policy Research, cepr.org/voxeu/columns/nonlinear-phillips-curve-helps-explain-inflation-overshoot.
“Regime Shifts and Policy Effectiveness.” Journal of Environmental Economics and Management, 2014, ScienceDirect, www.sciencedirect.com/science/article/abs/pii/S0095069614000199.
Reserve Bank of Australia. “The Transmission of Monetary Policy.” Explainer, www.rba.gov.au/education/resources/explainers/the-transmission-of-monetary-policy.htm
“Small-Sample Bias in Macroeconomic Estimation.” Economic Modelling, 2016, ScienceDirect, www.sciencedirect.com/science/article/abs/pii/S0264999316304035.
“State-Dependent Impulse Responses to Monetary Policy Shocks.” Journal of Economic Behavior and Organization, 2024, ScienceDirect, www.sciencedirect.com/science/article/pii/S0167268124004256.
“Structural Breaks in Inflation Dynamics.” Economics Letters, 2023, ScienceDirect, www.sciencedirect.com/science/article/abs/pii/S0165176523004202.
Tableau. “Time Series Analysis: Definition, Types, and Techniques.” www.tableau.com/analytics/what-is-time-series-analysis.
“Unconventional Monetary Policy and Inflation Expectations.” Omega, ScienceDirect, www.sciencedirect.com/science/article/abs/pii/S0305048317309246.
Washington Center for Equitable Growth. “The Unequal Impact of the Federal Reserve’s Monetary Tightening and Easing on Firms in the United States.” equitablegrowth.org/the-unequal-impact-of-the-federal-reserves-monetary-tightening-and-easing-on-firms-in-the-united-states/.

Domestic Transfer-Network Position and Emergency-Finance Strategies: Evidence from Venezuela By Héctor González, mentor By Jameson Augustin

Abstract

This paper examines how adults' positions in domestic interpersonal-transfer networks are associated with emergency-finance strategies in Venezuela. Using survey-weighted individual data from the 2021 and 2024 World Bank Global Findex surveys, it distinguishes four roles: neither sending nor receiving domestic remittances, receiving only, sending only, and both sending and receiving. The analysis corrects two measurement problems that can materially affect inference. First, "don't know" and refused responses are excluded rather than treated as substantive outcomes. Second, the ordinal emergency-fund item is coded in the direction specified by the World Bank: higher valid values indicate greater ease, not greater difficulty. The primary ease outcome therefore equals one only when respondents report that obtaining the required emergency amount within 30 days would be "not difficult at all." Relative to adults who neither send nor receive, recipient-only respondents are 16.7 percentage points more likely to name family, relatives, or friends as their main emergency source; sender-only respondents are 8.5 points less likely; and two-way participants are 8.7 points more likely. None of the three roles is significantly associated with using savings as the main emergency source. Sender-only respondents are 7.6 points more likely to report that obtaining emergency funds would not be difficult, although this pooled association is concentrated in 2021 and is not statistically significant in 2024. Weighted-logit models preserve the direction and significance pattern of the pooled estimates. The results are descriptive conditional associations, not causal effects, and do not show that remittances create dependence. They instead document heterogeneous emergency-finance profiles across observable positions in Venezuela's domestic transfer system.

Introduction

Remittances are an important source of household support in economies affected by unstable employment, inflation, and weak public safety nets. They can help families maintain consumption, purchase food and medicine, and respond to unexpected expenses. In such settings, however, the ability to receive money from relatives or friends is not the same as possessing savings that can be used independently. A household may be able to manage a shock because its network responds when needed even though the household has accumulated little liquid wealth of its own.

Venezuela provides a particularly useful setting in which to study this distinction. The country has experienced severe macroeconomic instability, partial dollarization, exchange controls, and repeated gaps between official and parallel exchange rates. At the same time, large-scale migration and persistent economic hardship have increased the importance of transfers between households. These conditions make remittances potentially valuable while also limiting households' ability to convert temporary inflows into durable savings.

This paper asks how an individual's position within Venezuela's domestic transfer system is associated with savings-based and family-based emergency strategies. It uses individual-level data from the 2021 and 2024 World Bank Global Findex surveys. The surveys record domestic

remittance sending and receipt in both waves, together with ease of obtaining emergency funds and the respondent's main expected source of those funds.

The focus on transfer-network position corrects a weakness in a simple comparison between recipients and non-recipients. Finding that remittance recipients are more likely to rely on family or friends for emergency money is not, by itself, strong evidence that remittances reinforce family dependence: a remittance is already an interpersonal transfer, often involving family. The same observed association could reflect selection into remittance receipt, the wording of the emergency-source question, or a common underlying family connection. The present analysis therefore does not treat family reliance as an outcome caused by remittance receipt. Instead, it asks whether different combinations of sending and receiving correspond to different emergency-finance profiles.

Four mutually exclusive roles are constructed: adults who neither send nor receive domestic remittances, recipient-only adults, sender-only adults, and adults who both send and receive. Sending is not a perfect measure of family closeness or of having relatives abroad, and it is not an exogenous control. It nevertheless provides an observed indicator of active participation on the giving side of the domestic transfer system. Comparing these roles is more informative than treating every recipient as equivalent and every non-recipient as outside a transfer network.

Three results stand out. First, domestic transfer participation is widespread: 47.7 percent of the weighted pooled sample either sends, receives, or does both. Second, no network role is significantly associated with greater reliance on savings as the main source of emergency funds. Third, emergency-finance profiles differ across roles. Controlling for sex, age, education, income quintile, workforce status, account ownership, and survey year, recipient-only respondents are 16.7 percentage points more likely than non-participants to report family, relatives, or friends as their main emergency source. Sender-only respondents are 8.5 points less likely to do so, while two-way participants are 8.7 points more likely. Sender-only respondents are also 7.6 points more likely to report that obtaining the specified emergency amount would be "not difficult at all." The latter association is not stable across waves and is therefore interpreted cautiously.

These patterns should not be interpreted as showing that remittances cause dependence or financial ease. The data are repeated cross-sections, do not observe remittance amounts or uses, and contain no independent measure of family closeness or relatives abroad. A defensible interpretation is narrower: observable transfer roles identify different conditional emergency-finance profiles. Recipient-only adults are more likely to name incoming interpersonal support, whereas sender-only adults are less likely to do so and, in the pooled sample, more likely to report that obtaining emergency money would not be difficult. These differences illustrate why "financial resilience" should not be reduced to a single scale or assigned a causal interpretation in these data.

The paper uses the term *individually accumulated resilience* for the capacity to meet shocks from one's own liquid savings and *network-based resilience* for the capacity to obtain resources through interpersonal support. The latter can be valuable and adaptive without necessarily producing financial independence. This distinction avoids treating family support as

inherently harmful while recognizing that private networks may be insufficient during economy-wide shocks.

Remittance channels remain relevant context, but they are not retained as a separate research question. The channel items in the two Findex waves are not sufficiently comparable to identify a modernization trend: the 2021 survey records financial-institution and money-transfer-operator receipt, whereas the harmonized 2024 measure captures digital receipt more directly. The paper therefore reports channel use descriptively and makes no claim about an increase or decrease over time.

The rest of the paper proceeds as follows. Section 2 reviews the literature and develops the distinction between individual and network-based resilience. Section 3 describes the data and variables. Section 4 presents the empirical strategy. Section 5 reports the results. Section 6 discusses interpretation, policy implications, and limitations. Section 7 concludes.

Literature Review and Conceptual Framework

Remittances, consumption, and private insurance

Remittances can improve welfare by relaxing short-run budget constraints. They may allow households to smooth consumption, purchase essential goods, and respond to health or income shocks. Evidence from Venezuela indicates that remittance-receiving households experienced better food-security outcomes, including higher caloric intake and more diverse diets (Stampini et al.). This is an important benefit even if remittances do not generate savings.

The economic literature also treats migration and remittances as part of household strategies rather than as isolated transactions. Lucas and Stark distinguish altruistic, exchange, and contractual motives for remitting and show that observed transfers can reflect an implicit arrangement between migrants and their families. Stark and Lucas further describe migration, remittances, and the family as connected components of household risk management. Private transfers may therefore behave like informal insurance: family members diversify locations and income sources, while resources move toward members facing greater need.

Research on private inter-household transfers likewise cautions against assuming a single motive. Cox, Eser, and Jimenez examine altruistic and exchange explanations using evidence from Peru and emphasize the role of capital-market imperfections. The implication for the present paper is that receipt alone cannot identify why a transfer occurs or what it does to the recipient's longer-term resilience. A transfer can simultaneously relieve hardship, express reciprocity, and maintain a relationship through which later support is expected.

Individual and network-based resilience

The Global Findex measures used here distinguish between the difficulty of raising emergency money and the expected source of that money. This makes it possible to separate two conceptually different forms of resilience.

First, *individually accumulated resilience* refers to the ability to finance a shock using liquid assets already controlled by the respondent, especially savings. This is close to conventional self-insurance. Second, *network-based resilience* refers to access to funds through

family or friends. It depends on relationships and reciprocal capacity rather than solely on assets held by the individual.

Neither form is automatically superior in every context. Family support can respond flexibly to need, cross borders, and function when formal credit or public assistance is unavailable. At the same time, network-based resilience has limitations. The amount and timing of help are uncertain; providers may also be under financial pressure; and covariate shocks can affect many network members at once. Individual savings provide greater autonomy but may be extremely difficult to accumulate when inflation, low real incomes, and recurring essential expenses absorb available cash.

This framework changes the interpretation of remittance receipt. If a transfer is spent on food or medicine, the absence of a savings association does not mean that it had no welfare value. It means only that the transfer is not observed as stronger savings-based emergency capacity. Likewise, reporting family as an emergency source is not proof of harmful dependence. It is evidence that the respondent's coping strategy is network-based.

Why transfers may not become savings

Several mechanisms could explain why remittances support current welfare without appearing as accumulated emergency savings. First, transfers may arrive in response to existing needs. Recipients are selected partly because their own resources are insufficient, so a cross-sectional comparison may show equal or worse resilience even when transfers help. Second, ongoing spending on food, medicine, transportation, housing, and debt can absorb transfers quickly. Third, high inflation can reduce the attractiveness of holding domestic-currency balances. Fourth, exchange-rate spreads, fees, and conversion constraints can reduce the value that reaches the recipient.

The present data cannot distinguish these mechanisms directly. They do not report remittance amounts, frequency, purpose, consumption, fees, or the exchange rate obtained by an individual household. The mechanisms therefore organize interpretation rather than constitute tested causal channels.

Exchange-rate distortions and the Venezuelan setting

The macroeconomic effects and household value of remittances depend partly on the exchange-rate regime. In economies with distorted official rates, a transfer's purchasing power can depend on where and how it is converted. Multiple exchange-rate systems may impose implicit taxes, generate arbitrage, and direct activity toward parallel markets (Kiguel and O'Connell; Gray). Evidence specific to Venezuela shows how inflation and the parallel exchange rate can interact in a repressed currency market (Cerra and Saxena).

At the macroeconomic level, remittance inflows can affect real exchange rates and domestic prices (Lopez, Molina, and Bussolo; Barajas et al.; Ball, Lopez, and Reyes). At the household level, however, these aggregate relationships do not reveal the amount or conversion rate received by a particular respondent. Exchange-rate distortion is therefore treated in this paper as an important contextual condition, not as an individually identified treatment.

Contribution and expectations

The paper makes a limited, descriptive contribution. It replaces a recipient–non-recipient contrast with a four-role classification that captures both sides of domestic transfers. This does not eliminate selection, but it reveals heterogeneity hidden by a binary receipt indicator.

Three expectations guide the analysis. First, if transfer receipt translates into greater individually accumulated resilience, recipient roles should be associated with more frequent reliance on savings and lower emergency difficulty. Second, if transfer roles reflect positions within a private risk-sharing system, the probability of naming family or friends should differ across recipients, senders, and two-way participants. Third, because a sender may face obligations to others as well as personal liquidity constraints, sending need not imply greater financial security.

Data

Survey design and sample

The analysis uses the Venezuela samples from the 2021 and 2024 World Bank Global Findex surveys (Demirgüç-Kunt et al.; Klapper et al.). Each wave contains 1,000 adult respondents, producing a pooled repeated cross-section of 2,000 observations. Survey weights are used in descriptive statistics and regression models.

The 2021 round was fielded from October 4 through November 24, 2021, using landline and mobile telephone interviews. The 2024 round was fielded from May 1 through December 31, 2024. Because the same people are not followed over time and the questionnaires are not identical, the data are treated as harmonized repeated cross-sections rather than a panel.

Weighted sample characteristics

Variable	2021	2024	Pooled
Unweighted observations	1,000	1,000	2,000
Female (%)	50.3	49.8	50.0
Age, mean	38.71	40.25	39.48
Education: low (%)	23.2	28.5	25.9
Education: secondary (%)	57.6	51.8	54.7
Education: tertiary (%)	19.2	19.8	19.5
In workforce (%)	78.9	62.4	70.7
Owns an account (%)	84.4	87.3	85.9

Notes: Entries are weighted percentages except age and unweighted observations.

Table 1 shows broadly similar sex and education distributions across waves, a modest increase in mean age, and high account ownership in both years. Workforce participation differs

more sharply, falling from 78.9 percent in 2021 to 62.4 percent in 2024. Regression models therefore include a year indicator and individual controls.

Transfer-network roles

Domestic remittance sending and receipt are observed in both survey rounds. These variables are combined into four mutually exclusive roles:

1. Neither: does not report sending or receiving domestic remittances;
2. Recipient only: receives but does not send;
3. Sender only: sends but does not receive; and
4. Both: both sends and receives.

The original Findex question refers specifically to money given or sent to, or received from, relatives or friends living in another city within Venezuela during the preceding 12 months. The categories therefore describe domestic interpersonal transfers, not international remittances and not complete family networks. A respondent classified as "neither" may still have close relatives, receive non-monetary support, or participate in transfers not captured by the question. Moreover, sending may be influenced by resources, obligations, or the same unobserved factors that affect emergency finance. The measure is consequently treated as an observed *domestic transfer position*, not as an exogenous measure of family connectedness.

Domestic transfer-network roles and family-based emergency finance

Network role	Unweighted N	Weighted share (%)	Family/friends source (%)
Neither sends nor receives	978	52.3	47.3
Recipient only	344	19.7	64.4
Sender only	256	10.5	33.0
Both sends and receives	422	17.4	48.1

Notes: Weighted pooled estimates from the 2021 and 2024 Venezuela Global Findex samples. The final column excludes "don't know" and refuses responses to the emergency-source item.

Table 2 shows why the role classification is informative. Recipient-only adults have the highest unadjusted probability of naming family or friends, while sender-only adults have the lowest. Adults who both send and receive are close to non-participants in the raw descriptive comparison.

Emergency-finance outcomes and coding audit

Three outcomes are consistently observed across the two waves. The first is derived from fin24a, which asks how difficult it would be to obtain an economy-specific emergency amount—defined by Findex as one-twentieth of local gross national income per capita—within 30 days. The World Bank coding is 1 for "very difficult," 2 for "somewhat difficult," and 3 for

"not difficult at all." Codes 4 and 5 are "don't know" and refused. To avoid treating an ordinal scale as cardinal and to make direction unambiguous, the primary outcome equals one for "not difficult at all" and zero for the two difficult categories. Codes 4/5 and system-missing observations are excluded.

The other outcomes are derived from fin24, which asks respondents to identify their *main* source of the emergency amount. Separate indicators equal one for savings and for family, relatives, or friends. Codes 1–7 are substantive sources, including "could not come up with the money." Codes 8/9—"don't know" and refused—are excluded rather than coded as zero. This correction changes both the valid sample and the descriptive shares relative to versions that treated non-substantive responses as outcomes.

Weighted financial-resilience indicators

Variable	2021	2024	Pooled
Obtaining funds "not difficult at all" (%)	13.0	15.9	14.4
Emergency funds mainly from savings (%)	5.5	10.4	8.0
Emergency funds mainly from family/friends (%)	51.5	47.1	49.3

Notes: Survey-weighted percentages among valid responses. The ease outcome excludes fin24a codes 4/5 and missing observations; the source outcomes exclude fin24 codes 8/9.

Savings are an uncommon main source in both years, despite account ownership exceeding 84 percent. Only 14.4 percent of valid pooled responses state that obtaining the emergency amount would be "not difficult at all." These variables measure anticipated strategies for a hypothetical emergency, not realized transfers, total household resources, or comprehensive welfare.

Remittance channels as descriptive context

Channel variables are not symmetrical across waves. In 2021, 72.8 percent of the observed recipient subsample reported receipt through a financial institution, while 40.1 percent reported a money transfer operator in a smaller item-specific sample. In 2024, 71.4 percent of recipients with observed channel information reported digital receipt. These figures show that formal or digital intermediation is common. They cannot demonstrate a trend because "financial institution," "money transfer operator," and "digital receipt" are not equivalent survey measures. Consequently, channel comparisons are not used as a research design or as evidence of modernization.

Methods

Primary specification

The primary specification relates each binary emergency-finance outcome to the four-category domestic transfer position:

$$Y_i = \alpha + \Sigma\{r \in \{R,S,B\}\} \beta_r \, Role^r_i + X_i'\gamma + \lambda t + \varepsilon_i .$$

The omitted group neither sends nor receives. R denotes recipient only, S sender only, and B both sending and receiving. The vector X_i contains sex, age, education category, income quintile, workforce status, and account ownership; λt is a 2024 indicator. All three outcomes are binary, so the primary models are survey-weighted linear probability models (LPMs). Coefficients are percentage-point differences relative to the omitted group. HC1 heteroskedasticity-robust standard errors are reported.

The survey documentation supplies an individual weight but no public-use primary-sampling-unit or stratum identifier in the analysis file. The estimates therefore apply the available Findex weights but cannot reproduce a full design-based variance estimator. Robust standard errors address heteroskedasticity, not all possible features of the original sample design.

Robustness and interpretation

Weighted binomial generalized linear models with a logit link are estimated as a functional-form check. Because logit coefficients are not directly comparable with LPM percentage-point effects, Appendix Table 5 reports odds ratios. Models are also estimated separately by survey year to test whether pooled associations are qualitatively stable.

The analysis is descriptive and associational. Transfer position is not randomly assigned, and the model cannot determine whether transfers precede or respond to emergency capacity. The family-source outcome is also conceptually close to the transfer-role variables because both refer to relatives or friends. The role comparison reduces the tautology of a simple receipt indicator by separating sending from receiving, but it does not remove common-method measurement or identify a causal network mechanism.

Results

Pooled conditional associations

Table 4 presents the primary pooled estimates.

Domestic transfer position and emergency-finance outcomes

	Not difficult at all	**Savings source**	**Family/friends source**
Recipient only	0.024	-0.002	0.167***
	(0.024)	(0.018)	(0.037)
Sender only	0.076**	0.008	-0.085**
	(0.033)	(0.025)	(0.042)
Both sends and receives	0.040	0.020	0.087**

	(0.024)	(0.021)	(0.037)
Reference group	Neither	Neither	Neither
Controls	Yes	Yes	Yes
Year indicator	Yes	Yes	Yes
Observations	1,931	1,975	1,975
Weighted dependent mean	0.145	0.080	0.494
Weighted R^2	0.114	0.044	0.084

Notes: Survey-weighted linear probability models with HC1-robust standard errors. The omitted group neither sends nor receives domestic remittances. Controls are sex, age, education, income quintile, workforce status, account ownership, and survey year. DK/refused responses are excluded. ** $p < 0.05$; *** $p < 0.01$.

Recipient-only respondents are not significantly different from non-participants in reporting that emergency funds would be "not difficult at all" or in naming savings. They are 16.7 percentage points more likely to name family, relatives, or friends as their main source. This difference is precisely estimated, but it should not be interpreted as an effect of receipt: both measures capture aspects of access to interpersonal transfers.

Sender-only respondents are 7.6 percentage points more likely to state that obtaining the emergency amount would not be difficult and 8.5 points less likely to name family or friends as their main source. The first result describes greater reported ease, not greater difficulty. It may reflect stronger underlying resources among senders, selection into a provider role, or other unobserved differences.

Adults who both send and receive are 8.7 percentage points more likely to name family or friends. Their 4.0-point difference in the ease outcome is positive but not statistically significant at the 5 percent level.

None of the three transfer roles is significantly associated with savings as the main emergency source. The estimates are small relative to their uncertainty. This is the most consistent null result in the analysis.

Robustness across functional form and survey year

The weighted-logit models in Appendix Table 5 preserve the pooled LPM pattern. For the "not difficult at all" outcome, the sender-only odds ratio is 1.81 ($p = 0.005$); recipient-only and two-way estimates are not significant at 5 percent. For the family-source outcome, the odds ratios are 2.07 for recipient only, 0.68 for sender only, and 1.46 for both. No savings-source odds ratio is statistically significant.

Wave-specific estimates in Appendix Table 6 show that not every pooled association is temporally stable. The recipient-only difference in naming families is positive in both 2021 (19.7 points, $p < 0.001$) and 2024 (12.6 points, $p = 0.017$). By contrast, the sender-only ease

association is 12.5 points in 2021 ($p = 0.019$) but 2.8 points in 2024 ($p = 0.494$). The pooled ease result should therefore be reported as sample-specific rather than as a stable structural relationship.

What the results establish

The results establish three descriptive facts in these samples. First, observed domestic transfer positions are associated with different reported emergency sources. Second, recipient-only, sender-only, and two-way participants should not automatically be combined into one category. Third, no transfer role is significantly associated with naming savings as the main emergency source.

The results do not establish that remittances reduce savings, cause vulnerability, or lock households into family dependence. A transfer may substantially improve food consumption or prevent hardship while leaving the respondent's main emergency source unchanged. The greater reported ease observed among sender-only adults in the pooled sample may precede transfer behavior, help determine who becomes a sender, or reflect other unobserved differences.

Discussion

From a tautology to a network-position interpretation

A finding that remittance recipients rely on family is vulnerable to the objection that it restates the definition of a remittance. The revised analysis addresses this problem in two ways.

First, it narrows the substantive claim. The paper does not claim that receipt causes family reliance. It claims that observed positions within the transfer system correspond to different emergency strategies. Second, it uses information on both sending and receiving. The contrast between recipient-only and sender-only adults shows that active transfer participation does not uniformly predict greater family reliance. Direction within the network matters.

This distinction makes the empirical pattern more informative, but it does not fully isolate family connectedness. Sending is not an exogenous control for close ties, and the survey does not identify relatives abroad. The evidence should therefore be understood as a role-based description of private-transfer behavior.

A mechanism consistent with the evidence

The results are consistent with—but do not prove—a mechanism in which repeated financial pressure directs transfers toward current needs rather than asset accumulation. In Venezuela, remittances may finance food, medicine, transportation, housing, debt repayment, or other immediate expenses. Stampini et al. provide direct evidence that remittances can improve food security in Venezuelan households. Such improvements are meaningful even if the money is spent rather than saved.

When another shock occurs, a recipient may again seek support from the transfer network. This produces network-based resilience: the household manages adversity through

relationships rather than solely through assets accumulated beforehand. Lucas and Stark's account of remittances as part of family arrangements and Stark and Lucas's household risk-sharing perspective provide a theoretical foundation for this interpretation.

The sender-only pattern adds an important qualification. In the pooled sample, sender-only adults are more likely to report that obtaining the emergency amount would be "not difficult at all" and less likely to identify family help as their main source. This profile is consistent with positive selection of more financially capable adults into a provider role, but that interpretation is not identified by the design. Moreover, the ease association appears in 2021 but not 2024, so it should not be generalized beyond the pooled sample.

Financial inclusion is not the same as liquidity

Account ownership exceeds 84 percent in both waves, while only 8.0 percent of valid pooled responses name savings as the main emergency source. Formal access therefore does not necessarily imply the presence of a usable emergency balance. This gap matters for policy. Expanding accounts can reduce transaction barriers, but account access alone cannot overcome low and unstable income, inflation risk, or recurring expenditure needs.

The channel evidence reinforces the need for careful language. Formal and digital methods are common in the observed recipient samples, but the survey items do not support a valid cross-wave modernization comparison. Digital receipt may make a transfer faster or more convenient; it does not automatically convert the transfer into savings or greater financial security.

Policy implications

Because the design does not estimate policy effects, the following implications are hypotheses for policy design and evaluation rather than recommendations demonstrated by the regressions. The appropriate goal is not to replace family support. Private networks may provide valuable assistance, while public and financial institutions can reduce transfer frictions and help households build optional buffers.

First, regulators and financial institutions could examine whether transparent fees, interoperability, and reliable settlement preserve more value in transfer transactions. The present survey does not measure fees or the effective exchange rate, so these potential benefits must be evaluated with transaction-level evidence.

Second, remittance services could offer voluntary emergency-savings features. A recipient might choose to direct a small percentage of each transfer into a separate, liquid balance. The design should be optional and easily reversible because many households need the full transfer immediately. Where legally and operationally feasible, products that protect balances from inflation or currency depreciation would be more useful than nominal domestic-currency accounts alone.

Third, transfer platforms could connect users to low-cost emergency insurance, bill payment, or targeted financial counseling. These services should be evaluated carefully and should not make receipt conditional on purchasing another product.

Fourth, governments should not assume that observed private transfers substitute fully for public social protection. Economy-wide crises can weaken senders and recipients simultaneously. Whether targeted assistance complements or displaces private transfers is an empirical question beyond the scope of this study.

Finally, policy evaluation should measure more than account ownership or digital adoption. Useful indicators include transfer fees, effective exchange rates, receipt reliability, voluntary emergency balances, time needed to raise funds, and whether households can meet shocks without sacrificing food, medicine, or other essential consumption.

Limitations

The study has several important limitations. First, the data do not provide an independent measure of family connection. There is no consistent variable for having relatives abroad, closeness to family, or the number and resources of potential supporters. The role classification improves description but does not isolate a causal network mechanism.

Second, the surveys do not report remittance amounts, frequency, sender-recipient relationships, or uses. A small occasional transfer and a large regular transfer receive the same binary coding. The analysis cannot test whether larger transfers are more or less likely to be saved.

Third, the data are repeated cross-sections. They cannot reveal whether entering a recipient or sender role changes resilience over time. Fourth, remittance sending and receipt are not randomly assigned. Income shocks, health, migration histories, household composition, and social norms may influence both transfer roles and emergency strategies.

Fifth, the emergency-source variables record a respondent's main anticipated source rather than the full set of available resources or a realized response to an observed shock. The similarity between the transfer and emergency-source questions also creates common-method overlap.

Sixth, the public analysis file provides weights but not the full sampling-design variables needed to reproduce design-based standard errors. HC1 errors do not account for every possible feature of the original design. Seventh, multiple coefficients are examined across outcomes and roles; the reported p-values are not adjusted for family-wise error and should be read together with effect sizes, confidence intervals, and the preponderance of evidence. Finally, the exchange rate actually received by each household is not observed, and channel questions differ across waves. Exchange-rate distortions and channel conditions can therefore be discussed only as context.

These limitations restrict the paper's claim but do not eliminate its contribution. The evidence shows meaningful heterogeneity across observable transfer roles and demonstrates why remittance receipt should not be equated with either financial security or harmful dependence.

Conclusion

This paper examined the relationship between domestic transfer-network position and financial resilience in Venezuela using the 2021 and 2024 Global Findex surveys. Its central question was not whether remittances mechanically predict reliance on family, but how the roles of recipient, sender, two-way participant, and non-participant correspond to different emergency-finance strategies.

The results do not show stronger savings-based self-insurance among transfer participants. Instead, they reveal distinct reported profiles. Recipient-only adults are more likely to name family, relatives, or friends as their main emergency source. Sender-only adults are less likely to name that source and, in the pooled sample, more likely to report that obtaining emergency money would not be difficult. Two-way participants show a smaller positive association with family-based emergency finance. These patterns are obscured when the analysis uses only a binary receipt indicator.

The findings support a distinction between individually accumulated and network-based resilience. Family support can improve welfare and help households manage shocks without producing large personal savings. This does not make remittances ineffective, nor does it prove dependence. It means that resilience can reside partly in relationships rather than exclusively in financial assets.

The policy implication is correspondingly limited. Low-friction transfers, optional emergency-saving tools, and public protection are plausible areas for evaluation, but their effects are not estimated here. Future research should collect transfer amounts, uses, fees, effective exchange rates, sender-recipient relationships, independent measures of family connection, and longitudinal outcomes. With such data, researchers could test whether transfers change savings and resilience over time. The present evidence supports a narrower conclusion: observable domestic transfer positions are associated with how adults say they would finance an emergency.

Data and Code Availability

The analysis uses anonymized public-use microdata from the World Bank Global Findex Database for Venezuela, survey years 2021 and 2024. The harmonized analysis file is DATA/data/analysis_ready/findex_venezuela_harmonized_final_corrected_v2.xlsx. All revised coding rules, validation checks, primary models, and robustness models are reproduced by DATA/05_revised_network_analysis.py. Machine-readable outputs are written to RESULTS/revised_network_analysis/. The script explicitly excludes non-substantive response codes before estimation.

Appendix: Robustness Results

Weighted-logit robustness models: odds ratios

	Not difficult at all	Savings source	Family/friends source
Recipient only	1.31	0.97	2.07***

Sender only	1.81***	1.12	0.68**
Both sends and receives	1.37*	1.27	1.46***
Observations	1,931	1,975	1,975
Controls and year indicator	Yes	Yes	Yes

Notes: Exponentiated coefficients from survey-weighted binomial generalized linear models with a logit link and HC1-robust standard errors. The omitted group neither sends nor receives. * p < 0.10; ** p < 0.05; *** p < 0.01.

Wave-specific linear probability models

	Not difficult at all		**Family/friends source**	
	2021	2024	2021	2024
Recipient only	0.069*	-0.015	0.197***	0.126**
	(0.038)	(0.028)	(0.052)	(0.053)
Sender only	0.125**	0.028	-0.107	-0.067
	(0.053)	(0.041)	(0.065)	(0.054)
Both sends and receives	0.032	0.042	0.076	0.099**
	(0.035)	(0.033)	(0.055)	(0.049)
Observations	963	968	991	984
Controls	Yes	Yes	Yes	Yes

Notes: Survey-weighted linear probability models with HC1-robust standard errors. Each column is estimated separately. The omitted group neither sends nor receives. Controls are sex, age, education, income quintile, workforce status, and account ownership. * p < 0.10; ** p < 0.05; *** p < 0.01.

Appendix: Coding Decisions

Audited outcome coding

Findex item	**Substantive coding**	**Exclusions and analysis variable**
fin24a	1 very difficult; 2 somewhat difficult; 3 not difficult at all	Exclude 4 don't know, 5 refused, and missing. Primary indicator equals one only for code 3.
fin24	1 savings; 2 family/relatives/friends; 3–7 other substantive outcomes	Exclude 8 don't know and 9 refused. Construct separate indicators for codes 1 and 2.
Domestic transfer sending/receipt	Binary indicators derived from the relevant wave-specific sending and receipt questions	Four mutually exclusive positions: neither, recipient only, sender only, and both.

Works Cited

Ball, Christopher P., C. Lopez, and Javier Reyes. "Remittances, Inflation and Exchange Rate Regimes in Small Open Economies." The World Economy, vol. 36, no. 4, 2013, pp. 487–507. https://doi.org/10.1111/twec.12042.

Barajas, Adolfo, et al. "Workers' Remittances and the Equilibrium Real Exchange Rate: Theory and Evidence." IMF Working Paper no. 10/287, International Monetary Fund, 2010.

Cerra, Valerie, and Sweta Chaman Saxena. "Inflation and the Black Market Exchange Rate in a Repressed Market: A Model of Venezuela." IMF Working Paper no. 16/159, International Monetary Fund, 2016.

Cox, Donald, Zekeriya Eser, and Emmanuel Jimenez. "Motives for Private Transfers over the Life Cycle: An Analytical Framework and Evidence for Peru." Journal of Development Economics, vol. 55, no. 1, 1998, pp. 57–80. https://doi.org/10.1016/S0304-3878(97)00056-4.

Demirgüç-Kunt, Asli, et al. The Global Findex Database 2021: Financial Inclusion, Digital Payments, and Resilience in the Age of COVID-19. World Bank, 2022. https://doi.org/10.1596/978-1-4648-1897-4.

Gray, Simon T. "Recognizing Reality: Unification of Official and Parallel Market Exchange Rates." IMF Working Paper no. 2021/025, International Monetary Fund, 2021. https://doi.org/10.5089/9781513568638.001.

Kiguel, Miguel A., and Stephen A. O'Connell. "Parallel Exchange Rates in Developing Countries." The World Bank Research Observer, vol. 10, no. 1, 1995, pp. 21–52.

Klapper, Leora, et al. The Global Findex Database 2025: Connectivity and Financial Inclusion in the Digital Economy. World Bank, 2025. https://doi.org/10.1596/978-1-4648-2204-9.

Lopez, Humberto, Luis Molina, and Maurizio Bussolo. "Remittances and the Real Exchange Rate." World Bank Policy Research Working Paper no. 4213, World Bank, 2007. https://doi.org/10.1596/1813-9450-4213.

Lucas, Robert E. B., and Oded Stark. "Motivations to Remit: Evidence from Botswana." Journal of Political Economy, vol. 93, no. 5, 1985, pp. 901–918. https://doi.org/10.1086/261341.

Stampini, Marco, et al. Effect of Remittances on Food Security in Venezuelan Households. Inter-American Development Bank, 2021. https://doi.org/10.18235/0003346.

Stark, Oded, and Robert E. B. Lucas. "Migration, Remittances, and the Family." Economic Development and Cultural Change, vol. 36, no. 3, 1988, pp. 465–481. https://doi.org/10.1086/451670.

What Factors Influence Buyer Intention Towards Standard-Priced and Luxury Residential Apartments in Bangalore? By Vivaan Sahu

Abstract

The purpose of the study is to assess the elements affecting the buyer's intention concerning conventional and luxury apartments in Bangalore. The study investigates the relationship between elements including price sensitivity, perceived luxury, and value justification; in particular, when it comes to the market where luxury real estate is gaining popularity. The research included a quantitative cross-sectional study with the help of a survey. In the end, the study obtained over 30 anonymous responses from different people living in Bangalore. The survey measured such variables as apartment preference, price sensitivity, luxury perception, willingness to pay more, and several specific features including location, building quality, amenities, design, and others. The results indicate that buyer inclination to buy luxury apartments is conditional and not automatic. Most respondents stated that they would only go for an option based on the price and value received, while over 90% of the participants considered that they would turn to a luxury apartment. Price was still highly significant; however, they were not looking for the cheapest option. They were trying to account for tangible values, including construction quality that had the highest mean rank of 4.67, location, and apartment design. Respondents were also careful in assuming that high prices identify higher quality or status automatically. Overall, the research effort proved that luxury apartment intention to purchase in Bangalore is influenced more by apparent functional and experiential value, rather than just the presence of branding, exclusivity or status. Future studies might involve larger samples, comparison of various Indian cities, and interviews with luxury property buyers.

Keywords Buyer intention, Luxury apartments, Price sensitivity, Perceived value, Bangalore housing market

Introduction

Residential real estate plays a crucial role in shaping urban living patterns, household wealth, and long-term financial decision-making. In rapidly growing metropolitan cities such as Bangalore, what people choose to buy is increasingly pushed around by rising land costs, lifestyle expectations that shift year by year, and the growing visibility of luxury apartment developments. As these cities expand and disposable incomes climb for certain groups, buyers get stuck between a standard-priced home and a premium residential choice which is comparatively more upscale. So understanding what really steers these buying decisions matters not just for developers or policymakers, but also if you want to map wider consumer behaviour patterns within urban real estate markets.

One main factor that shapes residential purchase decisions is price sensitivity. While price traditionally functions as a constraint, research in luxury consumption suggests that higher prices can also act as signals of quality, exclusivity, and social status. In the context of housing, this creates an interesting tension: luxury apartments are often significantly more expensive than

standard housing, yet they continue to attract demand. That leads to the question that whether buyers see price as just a wall, or whether price also helps in building perceived value, and then purchase intention. Still, other elements like construction craftsmanship, the property's amenities, its location, the developer's reputation, and even long term investment potential can help buyers rationalize those higher prices.

The current research in the field of luxury consumption indicates that the process of making purchases is not dictated by economics alone. The psychological aspects of status perception, fear of paying too much, influence of social environment, and experience are vital in forming the intentions of consumers. Research indicates that consumers may perceive luxury items as a way to enhance their social status, taking into account their fears of paying too much and losing valuable opportunities. Although a lot of this research is devoted to luxury items such as clothes or cars, the same behavioral patterns are becoming more significant in the real estate market.

Despite growing research in the luxury market, there is relatively limited research examining buyer intention in the residential apartment market, particularly within the Indian urban or suburban context. Bangalore, as one of India's fastest-growing cities with a diverse population of renters, tech, homeowners, and investors, provides a relevant environment for exploring how individuals perceive luxury housing and evaluate price versus value. The city of Bangalore, being one of the fastest-growing cities in India with a varied mix of renters, IT professionals, homeowners, and investors, can be an interesting setting to study the perception of luxury houses and how buyers consider price in relation to value.

This study aims to investigate the factors that affect the decision making process of residential apartment purchases in the city of Bangalore, focusing specifically on price sensitivity, luxury perception, and buyer intention. This study uses primary data gathered using a structured questionnaire to analyse how demographic factors and housing situation affect buyer preference towards standard-priced apartments and luxury ones. This paper is organised in the following manner: first, there is a review of the literature on luxury consumer behaviour and housing decision making; then there is a description of the survey design and data gathering methods; further, the results of the survey are presented; and finally, the discussion and conclusions are presented.

1. Literature review

Previous research on luxury consumption provides useful insights into the evaluation of expensive products by consumers and the influence of price, value perceptions, and psychological factors on purchase decisions. Even though a large part of this body of literature is concerned with luxury products such as fashion, cars, or branded items, many of the fundamental mechanisms are also highly applicable to residential real estate, since purchases in this area involve both a functional necessity and a symbolic meaning.

A key issue in the literature concerning luxury consumption is the dual role of price. Indeed, it is a financial constraint; price also functions as a sign of status and exclusivity. As

Zhang et al. (2024) demonstrate, higher prices for luxury items can enhance consumer's 'face perception', that is, the social image they perceive themselves to have, which in turn has a positive impact on their decision to make the purchase. The study also looks at the role of expected regret, distinguishing between the fear of overpaying and the fear of missing out. Even though the fear of overpaying can reduce people's intention to buy, the regret experienced for not buying a product that one actually wants may increase the intention to make the purchase. This model is particularly relevant in the case of luxury housing, since buyers have to balance the financial risk against the perceived long-term value and prestige of owning the property.

Besides price signalling, several studies have shown the significance of various value factors when it comes to consumers' decisions to buy luxury products. Hung et al. (2011) identify three main value areas (functional, experiential, and symbolic value) and show that both the functional quality and the experiential enjoyment have a strong positive effect on consumers' intention to purchase a luxury brand. Interestingly, though, symbolic value on its own does not always result in a higher intention to buy, which suggests that current consumers may place greater emphasis on tangible quality and on the real experience of using a product than on pure status symbolism. This finding is especially applicable to residential apartments, as aspects such as the standard of construction, the layout, and the facilities directly affect daily living.

The fact that luxury has an experiential dimension is also shown by cross-cultural research that looks at luxury consumption in different areas. As Ko and Megehee (2021) point out, the hedonic value continuously forecasts customers' intention to buy regardless of the cultural environment, for example in China and in the United Kingdom. Social influence has an effect as well, but the degree of this varies from one market to another. These findings show that the demand for luxury housing cannot be fully accounted for by people's economic ability or by their desire for prestige and it is also necessary to consider lifestyle aspirations and the experiential benefits.

Recent research also demonstrates that many factors influence how much people are willing to pay such as the rarity and sustainability of a product and the way in which people believe it has meaning as a viable product. A 2024 study, published in the Journal of International Consumer Marketing, found that if people think a product is both rare and sustainable, then their desire for it to be luxurious increases and this leads to a higher willingness to pay for luxury items. In the case of real estate, this can be seen in features such as a limited supply, a prime location, or an environmentally friendly design, all of which enhance the perceived value and enable premium pricing.

Branding on its own may not be sufficient to increase buyer intention. Research carried out in the area of luxury branding in the housing market shows that luxury labels do not always result in a higher perception of value where there is no real quality and differentiation. This lends support to the idea that luxury real estate must combine branding with genuine functional and experiential benefits if it is to counter price sensitivity.

Even though these insights have already been obtained, there is still a shortage in the current literature concerning the application of these luxury consumption models to decisions

about purchasing residential apartments in Indian cities. The present study examines this gap by examining buyer intention in Bangalore and how price sensitivity, perceived luxury, and value drivers differ in the case of standard-priced and luxury apartments.

2. Methodology

2.1 Aims

This study aims to investigate the factors influencing buyer intention towards standard-priced and luxury residential apartments in Bangalore. Specifically, the research examines how price sensitivity, perceived luxury, construction quality, amenities, developer reputation, investment value, and buyer psychology influence residential property purchase decisions.

The objectives of this research are the following:

- To effectively identify factors that buyers most importantly consider when purchasing a new residential apartment.
- To examine whether price acts as a barrier or a signal of quality in residential property purchases.
- To compare the perceptions of respondents who have previously purchased residential property versus those who have not.
- To investigate whether luxury apartments are perceived primarily through lifestyle, investment potential, social status, or other value drivers.
- To analyse and study buyer intention using quantitative survey responses collected from residents of Bangalore.

2.2 Research Design

The quantitative cross-sectional research design was used in this study. The use of a structured questionnaire was effective because this guarantees the same manner of attitudes, opinions, and purchase motivation measurement for different respondents, unlike a case-study method which is used for evaluating certain properties or developers. The questionnaire analysis allows relevant primary data to be collected straight from potential buyers of residential properties.

The design was based on evaluating existing literature in the area of luxury, its consumption, buyer psychology, purchase intentions, and value perception. Thus, the concepts discovered in literature were the basis for the development of survey questions.

The independent variables include:

- Price sensitivity
- Construction quality
- Apartment design
- Amenities

- Location
- Investment value
- Developer reputation
- Proximity to workplace
- Buyer psychology
- Perception of luxury

Demographic variables such as age, occupation, housing status, previous property ownership, and residential location within Bangalore were also collected to provide context for the findings.

2.3 Participants

The study consisted of 33 respondents residing in Bangalore.

The research involved participants from various backgrounds in terms of age, job, housing position, previous ownership experience, and place of residence. The goal is to collect opinions from a wider audience instead of targeting only luxury home buyers.

The respondents were selected from the participant pool using convenience sampling. The link to the survey was shared electronically via emails or instant messages by the researchers among their family members, friends, colleagues, and acquaintances. The participation was completely voluntary and anonymous.

At the same time, since the research focused on consumers' opinion instead of their behaviour, it was not critical for the participants to be actively buying homes at the time of the survey.

Variable	**Categories**
Sample size	33 respondents
Location	Bangalore
Sampling method	Convenience Sampling
Survey mode	Google Forms
Participation	Voluntary
Personal information collected	None (Anonymous survey)

Table 1 Respondent Characteristics (N=33)

2.4 Data collection

Primary data were collected using a structured questionnaire created through Google Forms. The questionnaire consisted of **14 questions**, organised into five sections:

Section	**Purpose**

Demographics	Collect respondent characteristics
General Buying Intention	Understand purchase preferences and housing history
Value & Justification	Identify important purchase factors
Luxury Perception	Measure perceptions towards luxury apartments
Buyer Psychology	Examine behavioural influences on purchase intention

Table 2: survey questionnaire structure (N=33)

Questions consisted of multiple-choice responses, Likert-scale ratings, ranking questions, and agreement scales. The final questionnaire was distributed electronically, and responses were automatically recorded in Google Sheets. Participation remained anonymous throughout the study. No names, email addresses, or personally identifiable information were collected.

2.5 Survey questionnaire

The questionnaire consisted of 14 structured questions, organised into five sections. The questions were designed to examine respondents' demographic characteristics, residential purchase preferences, perceptions of luxury apartments, and psychological factors influencing buying intention.

Section 1: Demographics

This section collected background information about each respondent to understand the demographic composition of the sample.

The following questions were included:

1. What is your age group?
2. What is your current residential status?
3. What area of Bangalore are you situated in?
4. What is your occupation?

Section 2: General Buying Intention

This section explored respondents' previous experience with residential property, apartment preferences, and the role of price in purchase decisions.

The following questions were asked:

1. What is your current housing status?
2. Have you purchased a residential property before?
3. If you were to buy an apartment, which option would you most prefer?
4. How important is price in your decision to buy an apartment?

5. At what point does apartment pricing begin to feel difficult to justify?

Section 3: Value and Justification

This section investigated which apartment characteristics respondents considered most important when evaluating residential properties.

1. Rate the importance of the following factors when choosing an apartment.
   (Linear scale: 1 = Least Important, 5 = Most Important)

- Quality of construction
- Apartment design and layout
- Amenities (gym, security, clubhouse, swimming pool, recreational areas, etc.)
- Location and neighbourhood
- Investment / resale value
- Reputation of the developer
- Proximity to workplace
- Apartment size and layout

2. Which of the following factors would make you consider paying more for an apartment?
   Rank them according to your preference, where 1 = most important and 6 = least important.

- better amenities
- higher construction quality
- strong developer reputation
- prime location
- better long-term investment value
- exclusivity (limited units / premium feel)

Section 4: Luxury versus Standard-Priced Apartments

This section examined respondents' perceptions of luxury apartments and whether higher prices influence perceived quality.

1. Compared to a standard-priced apartment, what do you associate most with a luxury apartment?

- Better lifestyle and comfort
- Higher social status
- Better investment
- Mainly higher price
- Security and privacy
- No major difference

2. Would you consider buying a luxury apartment if it clearly offered more value than a standard-priced apartment?

3. To what extent do you agree with the following statement? "A higher price can indicate better quality and status."

Section 5: Buyer Psychology

The final section explored behavioural factors influencing residential property purchase decisions.

1. Which statement best describes your thinking?
- I prefer safer, reasonably priced options.
- I am willing to pay more if the value is clear.
- I prefer premium options even if they cost more.
2. Have you ever not bought something because you felt it was overpriced?
3. Which situation worries you more?
- Paying too much for a property
- Missing out on a good property
- Neither

2.6 Data analysis

The results of the survey were exported directly to Google Sheet system and processed using Microsoft Excel software. The responses were cleaned and prepared before analysis.

Descriptive statistics like frequencies, means, and other indicators were provided for each question in the survey. Sum scores were calculated for all possible variables like price sensitivity, construction quality, availability of amenities, reputation of the developer.
Comparative analysis was also conducted by separating respondents into two groups:

- Individuals who had previously purchased residential property
- Individuals who had not previously purchased residential property

Mean scores for both groups were compared across several variables (price sensitivity, construction quality, amenities, investment value, and developer reputation) and then their comparisons formed the basis of the Results and Discussions.

2.7 Ethics

This research conducted the study following essential ethical principles. The study's participants had the option of responding completely voluntarily to the survey questions made available, and could opt to respond without being incentivized or pressured to do so. Participants were informed of the fact that the survey itself was being conducted solely for academic research purposes. Data were collected anonymously so that no manner of personal information such as names, emails, contacts and addresses could be traced back to the respondents. The data obtained in the course of the research did not leave the researchers' analytical framework of this study.

Overall findings of the present research reflect the privacy and confidentiality of the participants taken into consideration in the study.

3. Results and Discussion

This section presents the findings from the survey responses collected from 33 respondents in Bangalore. The results are analysed using descriptive statistics, including frequencies, percentages, mean scores, and comparative tables. The discussion connects the findings to existing research on residential purchase behaviour, luxury consumption, and Bangalore's housing market.

3.1 Overall Demography

Variable	Category	Frequency	Percentage
Age group	25–34	9	27.27%
	35–44	7	21.21%
	45–54	14	42.42%
	55+	3	9.09%
Current residential status	Renting an apartment/house	21	63.64%
	Own a residential property	12	36.36%
Occupation	Employed, private sector	31	93.94%
	Self-employed	2	6.06%
Purchased residential property before	Yes	19	57.58%
	No	14	42.42%

Table 3: Demographic profile of respondents (N = 33)

According to the demographics as shown in Table 3, the majority of respondents to the survey were people aged 45-54, making up 42.42% of the responses gathered. Most of the survey participants were people who were renting accommodation (63.64%), while only 36.36% of the respondents had properties of their own. The study included respondents who reported previously having purchased real estate (57.58% of people) as well as the ones who owned none (42.42%), thus allowing comparisons to be made between the views of the experienced and inexperienced buyers of real estate.

However, since the sample was heavily situated among private-sector employees (93.94%), it means that these findings may be reflecting the preferences of working professionals more than those of other occupational groups. This should be treated as a *limitation* when analysing the results.

3.2 Perceptions of Luxury Apartments

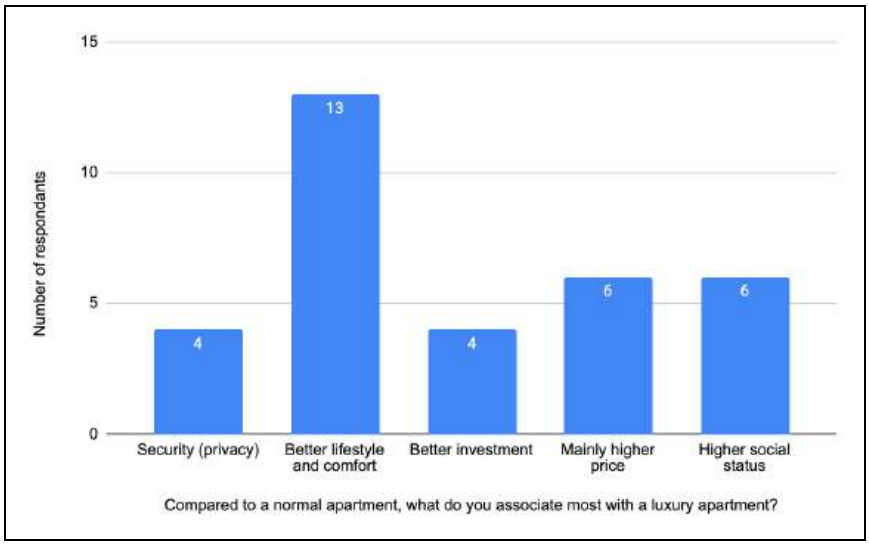


Figure 1: Association with a luxury apartment in comparison to a normal apartment (N=33)

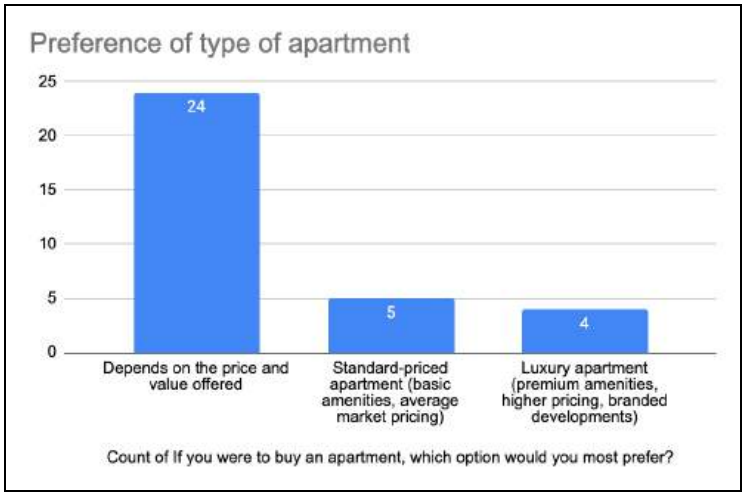
According to the information presented in Figure 1, the majority of participants (39.39%) said that the best aspect of a luxury apartment is that it is associated with a better lifestyle and greater comfort. This means that in the given case, luxurious apartments are viewed by the respondents first of all through the lens of lifestyle improvement, rather than through the perspective of wealth display or return on investment. The second most popular viewpoints among the participants were high social status and high price (18.18%). As for satisfaction with personal investment and security and privacy, their share was only 12.12%.

These observations support the idea that luxury real estate is created not only with consideration of its monetary value but also its experiential characteristics. This confirms the findings from research in the field of luxury consumption that was presented in the Literature review section, according to which the enhancement of life's pleasures is identified among the leading motivations to purchase luxury goods (JLL).

3.3 Apartment Preference and Price Sensitivity

Figure 2: apartment preference by type (N=33)

The key conclusion of Figure 2 is that 24 out of 33 respondents (72.73%) indicated their choice as "depends on the price and offered value." This tells us that most respondents are not categorically against luxury apartments, but at the same time, we can see that respondents do not always prefer luxury apartments. Instead, their intention to purchase may depend on whether the higher price is really worth it due to additional value.



Only 12.12% indicated their preference for luxury apartments and 15.15% preferred standard-priced apartments, showing that the main logic of individuals in terms of making a choice is not based on luxury and means of standard, instead it is dependent on price and the perceived value.

This conclusion is supported by Sundrani's study of apartment buyers in India, who have shown that price and product-related factors and location are the major reasons for making purchase decisions. In case we consider larger apartments, it appears that product and location become especially crucial, thus showing that buyers evaluate various value factors, rather than only price (DOI).

Price Sensitivity Question	Result
Mean importance of price on 1–5 scale	4.30
Respondents rating price as 5/5	19
Respondents rating price as 4/5	7
Respondents who felt pricing becomes difficult to justify when significantly higher than average	18
Respondents who felt pricing becomes difficult to justify when moderately higher than average	11
Respondents who felt pricing becomes difficult to justify when slightly above average market price	4

Table 4: Price sensitivity indicators (N = 33)

The average price value has a mean score of 4.30 out of 5 and indicates that price is valuable in making purchase decisions. However, the fact that most respondents preferred the option of ‘depends on price and value’, shows that price sensitivity does not mean buyers only

want the cheapest option. Instead, respondents appear willing to consider higher prices if the additional value is clear.

This is further verified by JLL's Q1 2026 report on the Indian residential market. The report indicates that the sales volume in the housing segment priced above INR 10 million showed a year-on-year increase of 30 percent, while the sales numbers in the segment priced below INR 10 million decreased. JLL identified many reasons for this change, among which higher salaries, preference for larger homes, good locations, advanced technologies in houses, and less sensitivity of rich buyers towards price changes are worth mentioning (JLL).

3.4 Factors Considered Important When Choosing an Apartment

Factor	Mean Score
Quality of construction	4.67
Location and neighbourhood	4.42
Apartment design and layout	4.36
Size and layout	4.27
Reputation of the developer	4.24
Investment / resale value	4.15
Proximity from workplace	4.15
Amenities	3.73

Table 5: Mean importance of apartment purchase factors on a scale of 1–5 (N = 33)

According to Table 5, the highest score was assigned to the quality of construction, which scored 4.67 points, whereas the second position belongs to location/neighbourhood (4.42); the third one is occupied by apartment design/layout (4.36). This proves that respondents prioritize more of the practical aspects more than just aesthetic considerations.

As for amenities, this is the lowest score of 3.73 among all the selected indicators, but still higher than the average. Thus, amenities appear to be a relatively important factor, but less significant than construction quality, location, and apartment design.

This finding is strongly supported by existing residential real estate research. Gaikwad and Bhalerao identify location, price and affordability, amenities and infrastructure, builder reputation, property size and layout, and lifestyle aspirations as key factors in residential purchase decision-making. Similarly, JLL's Bengaluru residential report states that buyer preference for larger, higher-quality homes helped drive the city's strong residential performance in Q1 2026 (JLL).

3.5 Factors That Justify Paying More for an Apartment

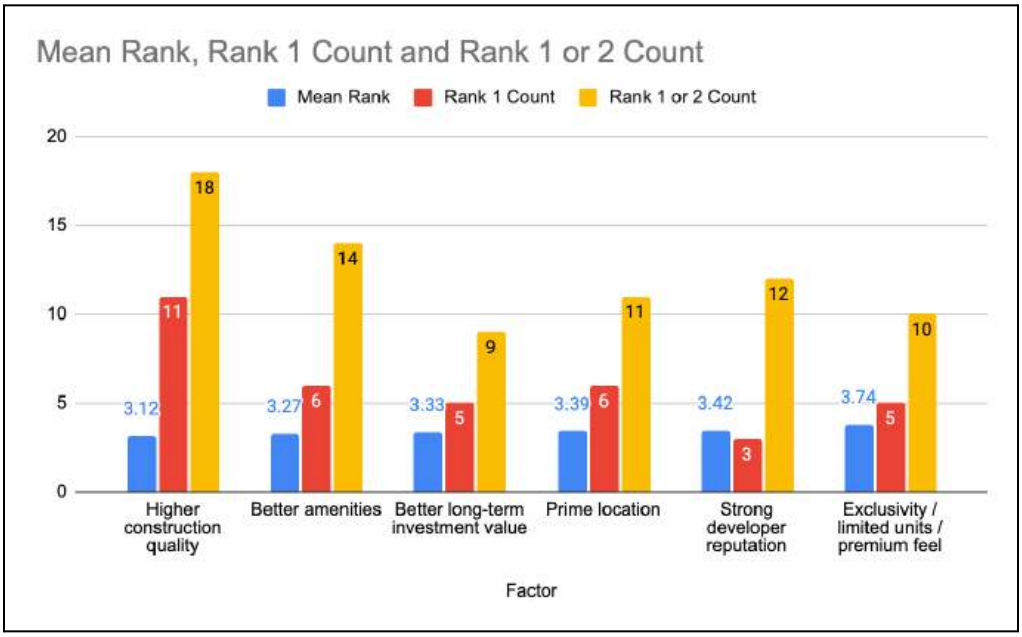


Figure 3: factors that justify paying more for an apartment (N=33; lower mean rank = stronger preference)

Important: this ranking question was messy because respondents were able to assign the same rank to multiple factors. So we should not present this as a perfect forced ranking. The fairest method is to treat it as an indicative average rank. Where respondents selected two ranks in one cell, the average of those ranks was used.

Based on the information provided in Figure 3, it can be stated that increased apartment construction quality was the leading factor that respondents will want to pay a higher apartment price. The mean rank was 3.12. The second factor had a better mean rank of 3.27, namely better amenities.

However, the exclusivity factor recorded the lowest average score of 3.74, which means that the participants may not be very much convinced by the exclusivity of the apartment to make the price higher. This is particularly surprising because luxury real estate usually exploits the ideas of exclusivity and prestige in its marketing strategies and complains of low demand but the results of this study indicate that what consumers are interested in is the actual value of the apartment, improved construction quality and amenities.

This supports the argument from Jennifer Deerman (2022) that branding or luxury signalling alone is insufficient unless supported by real product quality. Her research reaches a similar conclusion: luxury branding in real estate does not automatically increase perceived home value unless backed by substantive quality or differentiation (Deermen, 2022).

3.6 Willingness to Consider Luxury Apartments

Figure 4: willingness to consider luxury apartments if more value is offered (N=33)



One of the significant outcomes from this study is that 30 out of 33 participants (90.91%) said “Yes” or “Maybe” when they were asked if they would consider investing in a luxury apartment given it offers much more than a low-end one. This demonstrates that the respondents had no issues with luxury housing itself, but their decision depended entirely on the financial aspects involved.

Hence, we see that the above-mentioned outcome of the research is in line with the fundamental concept of this paper: the intention of the luxury buyers to buy this property is based on their wishes and expectations. That is, they are ready to pay extra for the apartment if its quality of construction, comfort and design meet their expectations.

This is something that is currently happening in the real estate market of India. According to JLL, during the first few months of 2025 the premium segment accounted for 71% of the total sales in the seven largest metropolises of the country instead of 59% earlier on. This proves that there is no reason for the buyers to ignore the price but they simply hope that they will receive value equal or higher than the price paid (JLL).

3.7 Price as a Signal of Quality and Status

Score	**Interpretation**	**Frequency**	**Percentage**
1	Strongly disagree	5	15.15%
2	Disagree	9	27.27%
3	Neutral	13	39.39%
4	Agree	5	15.15%
5	Strongly agree	1	3.03%
Mean score	**2.64**		

Table 6: Agreement with the statement “A higher price can indicate better quality and status” (N = 33)

The mean score for the statement was only 2.64 out of 5, showing that respondents were relatively cautious about assuming that a higher price automatically indicates better quality and status. Most respondents were either neutral or disagreed with the statement.

This contrast with luxury consumption theory, where price can sometimes act as a signal of exclusivity, prestige, and quality is important to note as proven by this sample, respondents did not strongly accept price as a status signal by itself. This suggests that in Bangalore's residential apartment market, high pricing is supported by visible value such as construction quality, location, design, and developer credibility.

This finding links back to the earlier result that construction quality had the highest importance score (**4.67**) and the strongest average rank as a justification for paying more. Therefore, the data suggests that respondents are more value-conscious luxury evaluators than status driven.

3.8 Buyer Psychology: Overpricing, Risk, and Regret

Question	Response	Frequency	Percentage
Which statement best describes your thinking?	I'm willing to pay more if the value is clear	21	63.64%
	I prefer safer, reasonably priced options	12	36.36%
	I prefer premium options even if they cost more	0	0.00%
Have you ever not bought something because you felt it was overpriced?	Yes	30	90.91%
	No	3	9.09%
Which situation worries you more?	Paying too much for a property	14	42.42%
	Missing out on a good property	8	24.24%
	Neither	11	33.33%

Table 7: buyer psychology indicators (N=33)

The buyer psychology results further support the importance of value justification. A majority of respondents (63.64%) stated that they were willing to pay more if the value was clear. At the same time, 90.91% had avoided buying something before because they felt it was overpriced. This shows that respondents are open to premium pricing but remain highly sensitive to perceived overpricing. The worry-based question also shows that 42.42% of respondents were more worried about paying too much for a property, compared to 24.24% who were more worried about missing out on a good property. This suggests that fear of overpayment is stronger than fear of missing out in this sample. This connects closely with the expected-regret framework from the literature review. Luxury purchase decisions can involve both fear of overpaying and fear of missing out, but in this dataset, overpayment concern appears stronger. A

study on the national library of medicine by Jiahua Wei and colleagues also identifies this distinction between regret from buying and regret from not buying (Wei & colleagues).

Yes	4.36
No	4.21

Table 8: Average response on a scale of 5 - How important is price in your decision to buy an apartment? (N = 33)

In table 8, it is reported that people who have bought an apartment before keep price as a decision maker, however it is important to say that even people who didn't buy before kept it at a high priority marginally less than Yes responses.

3.9 Comparison Between Previous Property Buyers and Non-Buyers

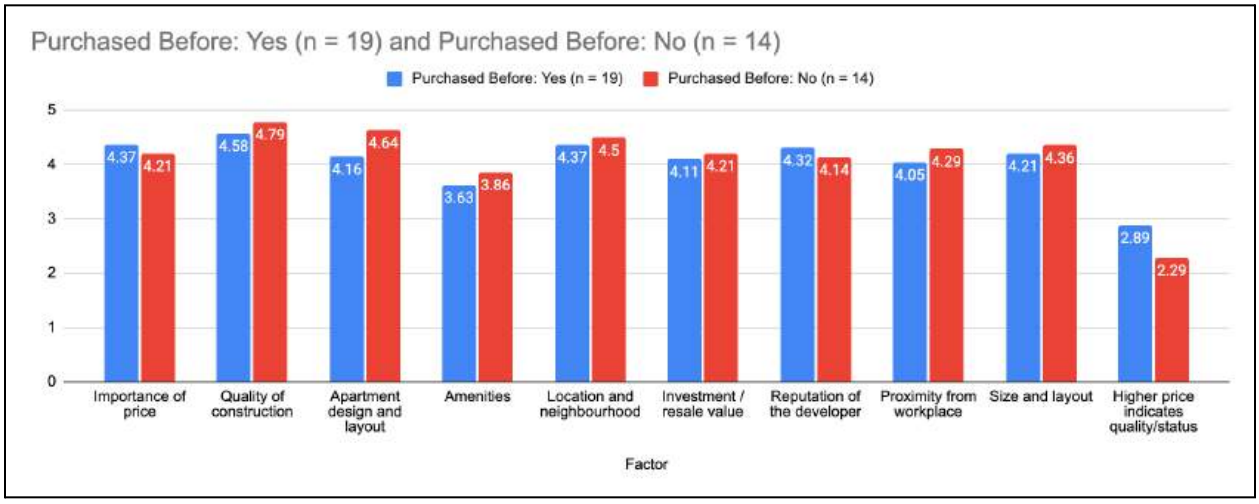


Figure 5: Comparison of mean responses between respondents who have and have not purchased residential property before (N = 33)

Figure 5 compares respondents who had previously purchased residential property with those who had not. Respondents with prior property purchase experience gave price a slightly higher mean score (4.37) than those without prior experience (4.21). This suggests that experienced buyers may be slightly more sensitive to price as a decision-making factor, possibly because they have previously gone through the financial process of buying property. For most product-related factors, however, respondents who had not purchased property before gave slightly higher scores. For example, non-buyers rated quality of construction at 4.79, compared to 4.58 among previous buyers. They also rated apartment design and layout higher (4.64 compared to 4.16) and location higher (4.50 compared to 4.37). This may indicate that respondents without purchase experience hold stronger ideal expectations regarding apartment features.

The only product-related factor where previous buyers scored higher was developer reputation, with a mean score of 4.32 compared to 4.14 among non-buyers. This suggests that people who have already purchased property may place slightly greater emphasis on developer credibility and reliability. This is consistent with JLL's Q1 2026 India residential market analysis, which states that established developers with strong execution and brand strength are expected to gain market share, especially in quality developments located in prime areas (JLL).

3.10 Summary of Key Findings

Overall, the results prove that buyer intention towards luxury housing in Bangalore is shaped by a combination of both price sensitivity as well as perceived value. Most of the individuals surveyed did not solely choose between regular apartments with typical prices and luxury ones. The overwhelming majority, or exactly 72.73% of them, specified that their preferences are determined by both price and the value they receive. The discussion indicates that the respondents are open to using luxury apartments, but the precondition is that the premium price should be warranted. The importance of the construction quality, location, design, size, and reputation of the developer turned out to be greater than the importance of the exclusivity or high status of the apartment alone. Luxury apartments are most frequently associated with lifestyle and comfort improvements. Thus, it can be concluded that an individual does not consider luxury housing to be a mere status indicator but stresses upon the experiential and functional aspects of living there.

Conclusion

In this paper, the factors that influence buyers' choices when buying either standard-priced or luxury apartments in Bangalore were investigated. The study evaluated whether people regard luxury apartments as a prestige item or whether buyers' decisions are affected more by practical aspects of apartments, such as construction quality, location, amenities, design, investment value, developer reputation, etc. The results demonstrate that buyers' attitude towards luxury apartments is very conditional. Most of the respondents did not show a strong preference for either standard or luxury apartments, with 72.73% stating that their choice would depend on both price and value. Price remained an important factor for potential buyers deciding on the apartment. The average value of the importance of price was 4.30 out of 5, which shows that price is considered important. However, the findings demonstrate that price sensitivity does not mean that consumers are looking solely for the cheapest option. Construction quality scored the highest value of 4.67, followed by location and neighborhood with 4.42.This research also showed that participants did not assume expensive goods guaranteed high quality and prestige, as indicated by the low mean score of 2.64. Buyer attitudes towards luxury apartments are based on their willingness to pay and their perception of the value of luxury apartments. Further studies can involve a bigger and more diverse sample from various cities in India and can incorporate interview schedules with real luxury apartment buyers and developers.

Works Cited

Banerjee, S. (2026, June 17). Residential Dynamics Report Q1 2026. https://www.jll.com/en-in/insights/market-dynamics/india-residential?utm_source=chatgpt.com

Bengaluru | JLL. (2026, May 20). JLL. https://research.jllapsites.com/appd-market-report/q1-2026-residential-bengaluru/?utm_source=chatgpt.com

Checking your browser - reCAPTCHA. (n.d.). https://pmc.ncbi.nlm.nih.gov/articles/PMC7847243/?utm_source=chatgpt.com

Deerman, J. M. (2022). Does luxury real estate branding increase buyer perceptions of home value? An empirical study of stable, affluent communities [Doctoral dissertation, Temple University]. Temple University ScholarShare. https://scholarshare.temple.edu/server/api/core/bitstreams/c0167221-c809-4d0d-908a-0f5a994dbd8a/content

Gaikwad, D. D., & Bhalerao, A. (2025). Factors influencing homebuyers' decision-making in the residential real estate market. International Journal of Research -GRANTHAALAYAH, 13(5), 321–327. https://doi.org/10.29121/granthaalayah.v13.i5.2025.6485

Hung, K., Chen, A. H., Department of Business Administration, Ming Chuan University, Taipei City, Taiwan, Peng, N., Westminster Business School, University of Westminster, London, UK, Hackley, C., School of Management, Royal Holloway, University of London, London, UK, Tiwsakul, R. A., School of Management, University of Surrey, Guildford, UK, Chou, C., & Department of Business Administration, Ming Chuan University, Taipei City, Taiwan. (2011). Antecedents of luxury brand purchase intention. Journal of Product & Brand Management, 457–467. https://thehometrust.com/wp-content/uploads/2024/03/Antecedents_of_luxury_brand_purchase_int.pdf?utm_source=chatgpt.com

Sharma, R. (2026, May 26). Bengaluru Mass Residential Market Dynamics Q1 2026. https://www.jll.com/en-in/insights/market-dynamics/bengaluru-residential?utm_source=chatgpt.com

Sundrani, D. M. (2018). Factors influencing home-purchase decision of buyers of different types of apartments in India. ideas.repec.org. https://ideas.repec.org/a/eme/ijhmap/ijhma-06-2017-0062.html?utm_source=chatgpt.com

Sundrani, D. M. (2018). Factors influencing home-purchase decision of buyers of different types of apartments in India. International Journal of Housing Markets and Analysis, 11(4), 609–631. https://doi.org/10.1108/IJHMA-06-2017-0062

Wang, X., Sung, B., & Phau, I. (2024). Examining the mediating role of consumer desire for luxury: Can perceived sustainability and natural rarity evoke willingness to pay more? Journal of International Consumer Marketing, 36(4), 301–319. https://doi.org/10.1080/08961530.2024.2333756

Wei, J., Shen, H., & Lu, Z. (2024). A study on the purchase intention of luxury goods from the perspective of face perception and expected regret. PLoS ONE, 19(3), e0297050. https://doi.org/10.1371/journal.pone.0297050

Beyond Capital and Location
FactorsInfluencingthe Survivaland GrowthofSmall Retail Businesses in Namangan, Uzbekistan By Abdikarimov Mirzoxid

Abstract

Small retail businesses play an important role in local economies by providing employment, products, and services while creating opportunities for individual entrepreneurship. However, many small retailers face substantial challenges, including limited startup capital, high rent, strong competition, changing consumer demand, and pressure from family members to abandon unprofitable businesses. This study examines the factors associated with the survival and growth of small retail businesses in Namangan, Uzbekistan, with particular attention to entrepreneurial mindset, customer relationships, communication skills, previous retail experience, location, rent, product selection, and access to capital.

The study draws on field research conducted among approximately 90 to 100 retail businesses located primarily in Sardoba Shopping Center and smaller shopping centers in the surrounding area. Respondents included approximately 50% shop managers, 30% business owners, and 20% sales employees. The research also considers differences between entrepreneurs who started businesses independently, those who continued family businesses, and those who entered entrepreneurship after gaining retail experience as employees.

Preliminary findings indicate that business survival cannot be explained by financial resources alone. Respondents repeatedly emphasized customer communication, experience, customer loyalty, adaptive product selection, and persistence during periods of low profitability. Approximately 40% of surveyed businesses reported having seriously considered closing at some point. Among those who continued, respondents frequently identified accumulated experience, expectations of future profitability, and opportunities created by competitors leaving the market as reasons for continuing. The findings suggest that entrepreneurial mindset may influence how business owners respond to financial and competitive pressures, although further quantitative analysis is required to establish the strength of these relationships.

Keywords: small business, entrepreneurship, retail, business survival, business growth, customer relationships, entrepreneurial mindset, Uzbekistan, Namangan

1. Introduction

Small businesses provide an important source of employment and economic activity in developing economies. Retail businesses are particularly significant because they connect producers and wholesalers with consumers and often require less initial capital than many other forms of entrepreneurship.

Despite their relatively accessible entry requirements, small retail businesses face substantial risks. Entrepreneurs may begin with limited capital, insufficient experience, uncertain customer demand, high rental costs, and strong competition. These challenges can become

particularly significant when businesses operate in concentrated commercial environments where many sellers offer similar products at similar prices.

Namangan provides an interesting setting for studying these issues. The local retail environment includes large commercial centers such as Sardoba Shopping Center as well as numerous smaller neighborhood shops and shopping centers. Businesses operating in these environments differ considerably in their starting conditions. Some entrepreneurs start independently with very limited capital and no previous retail experience. Others enter businesses that were established by parents or relatives. A third group develops retail skills as employees before establishing their own businesses. This study focuses on an important question:

Which factors influence the survival and growth of small retail businesses in Namangan, Uzbekistan?

The study pays particular attention to factors that are often difficult to measure using financial information alone. These include customer relationships, communication skills, previous experience, entrepreneurial mindset, product selection, location decisions, and responses to competition.

The study is based on field research conducted among approximately 90 to 100 retail businesses. Because many business owners were unwilling to disclose their profits, the research uses alternative indicators of business performance, including business survival, expansion, customer traffic, customer loyalty, employee numbers, ability to continue operating, and perceived stability.

2. Research Questions

1. Whichfactorsare associatedwiththe survival of small retail businesses in Namangan?
2. How does previous retail experience influence entrepreneurial decision-making and business development?
3. How do communication skills and customer relationships contribute to customer retention?
4. How do rent, location, competition, and capital affect business survival?
5. How does entrepreneurial mindset influence the way business owners respond to financial and competitive difficulties?

3. Research Hypotheses

H1

Previous retail experience is positively associated with business survival and growth. The fieldwork suggests that experienced sellers develop knowledge about suppliers, customers, inventory, pricing, employees, competitors, and locations before becoming business owners.

H2

Stronger customer relationships are positively associated with customer retention. Respondents repeatedly connected returning customers with trust, communication, recommendations, and personalized service.

H3

High rental costs are negatively associated with business stability when customer traffic is insufficient. Several respondents identified rent as one of the major pressures affecting their businesses.

H4

Customer-oriented communication is positively associated with customer loyalty. Respondents described successful sellers as individuals who identify customer needs, explain products, compare alternatives, and provide recommendations.

H5

Adaptive product selection is positively associated with business survival. Respondents described monitoring customer requests and testing new products through small initial purchases.

H6

Entrepreneurial persistence and adaptive mindset are associated with the ability to continue operating during periods of low profitability. This is particularly relevant because approximately 40% of respondents reported having seriously considered closing.

4. Methodology

4.1 Research Design

This study uses a field-based mixed-method exploratory design. The research combines structured information collected from retail businesses, respondent descriptions of their business experiences, observations of business environments, and qualitative information about survival, competition, customers, and decision-making. The study focuses specifically on small retail businesses in Namangan, Uzbekistan.

4.2 Research Location

The primary research location was Sardoba Shopping Center, supplemented by smaller shopping centers and retail businesses located nearby. Sardoba provides a particularly relevant environment because businesses operate in close proximity to one another and frequently sell similar products. Respondents described nearby businesses as direct competitors.

4.3 Sample

Approximately 90 to 100 businesses and retail workers participated in the research. The respondent composition was approximately 50% shop managers, 30% business owners, and 20% sales employees. Most businesses were retail businesses involved in selling consumer products. Most businesses had approximately three employees. Approximately 20% of the owners surveyed operated more than one business.

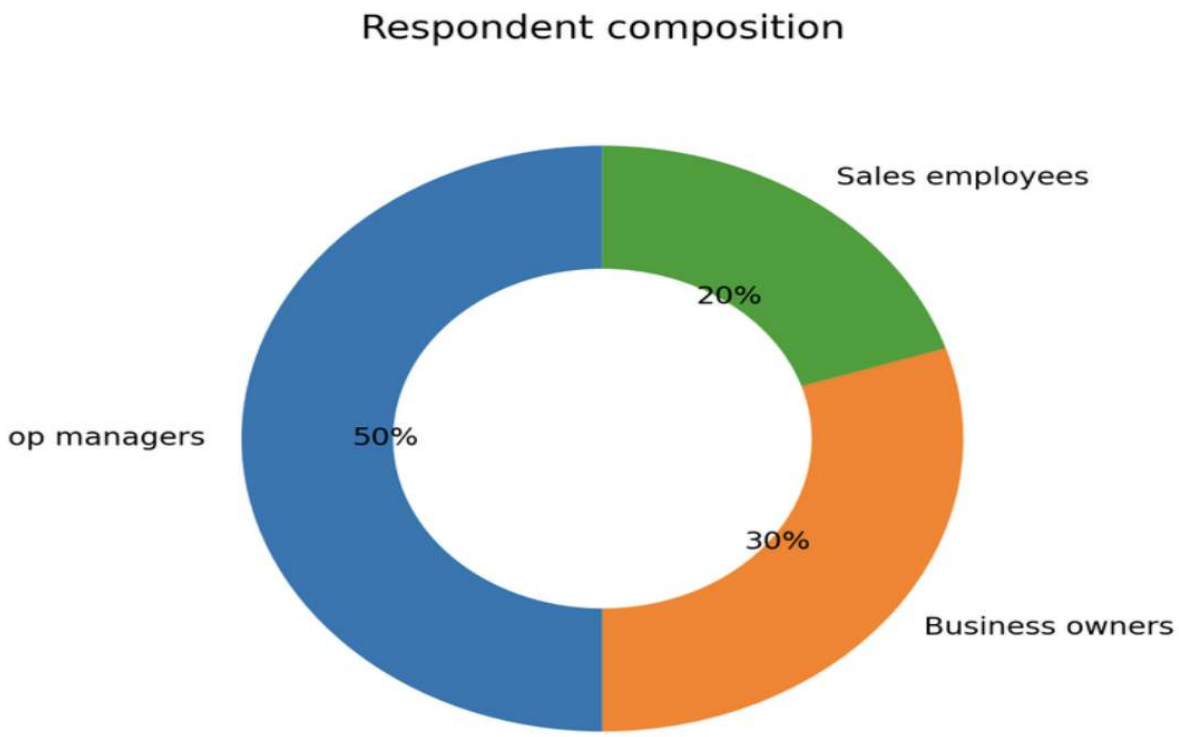


Figure 1. Approximate respondent composition. Percentages are based on the fieldwork description.

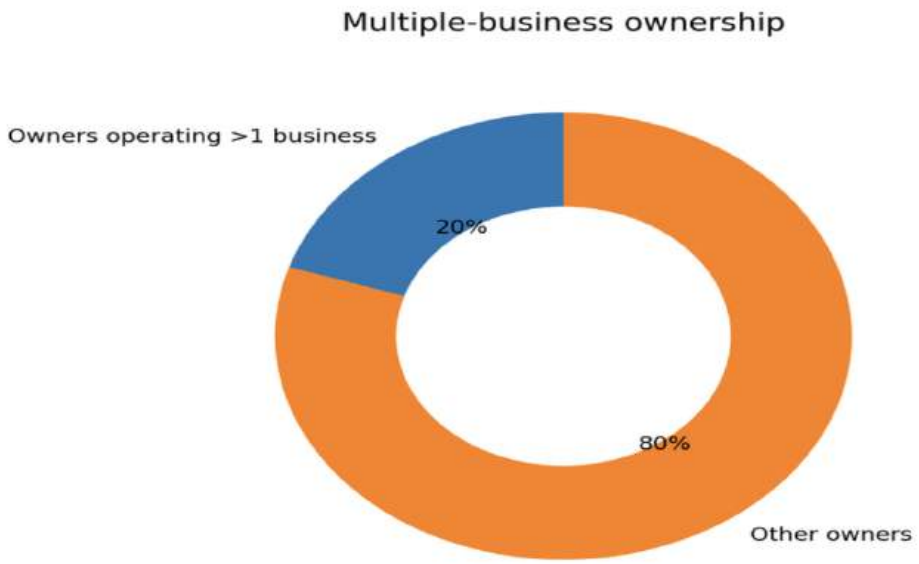


Figure 2. Approximate share of owners operating more than one business.

4.4 Entrepreneurial Background

Respondents represented different entrepreneurial pathways. Approximately 60% started their businesses independently from a low-resource starting position, 30% continued or developed an existing family business, and 10% received substantial family resources or had significant advantages at the beginning. The exact percentages should be treated as approximate because the original field data was not recorded as a formal statistical dataset.

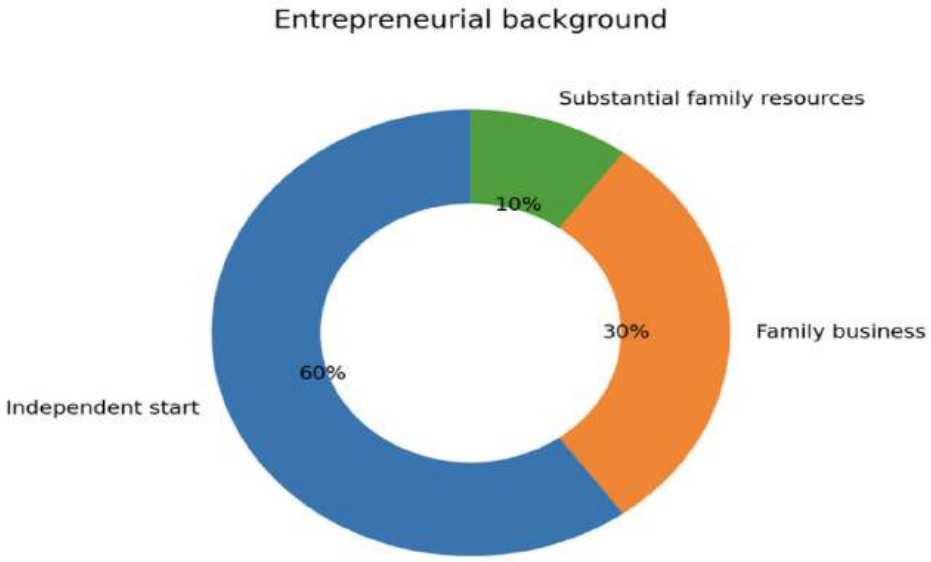


Figure 3. Approximate entrepreneurial pathways.

4.5 Business Age

Many businesses began between approximately 2015 and 2020. Some businesses began earlier, around 2010. The research also included newer entrepreneurs who established businesses around 2022 to 2025. Several of these newer entrepreneurs had previously worked in Sardoba for approximately 3 to 5 years.

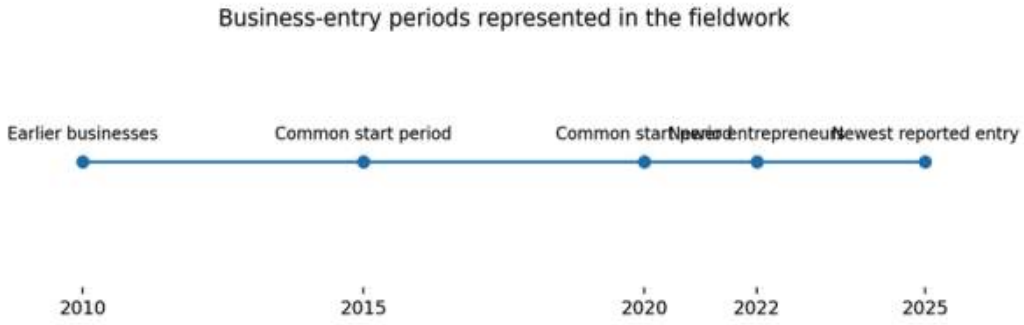


Figure 4. Business-entry periods represented in the fieldwork.

4.6 Financial Information

The study does not use reported profit as its primary financial measure because most respondents were unwilling to disclose their monthly profits. Respondents were more willing to discuss rent, employee wages, starting capital, loans, inventory, and business size. Monthly rent in Sardoba was reported at approximately 3 million to 15 million UZS, depending on location and shop size. Employee wages were reported at approximately 3 million to 5 million UZS per month, with experience and customer-handling ability influencing compensation.

5. Findings

5.1StartingWith Limited Resources

A substantial proportion of entrepreneurs reported starting with limited resources. Approximately 60% of respondents were classified as having started from zero or with limited initial support. Some entrepreneurs started with very small amounts of capital. Others borrowed money from relatives and friends. Approximately 5% to 8% of the businesses were reported to have used relatively large loans during the startup period. Respondents described loans as a major source of pressure when the business could not generate sufficient revenue to cover rent and repayment obligations.

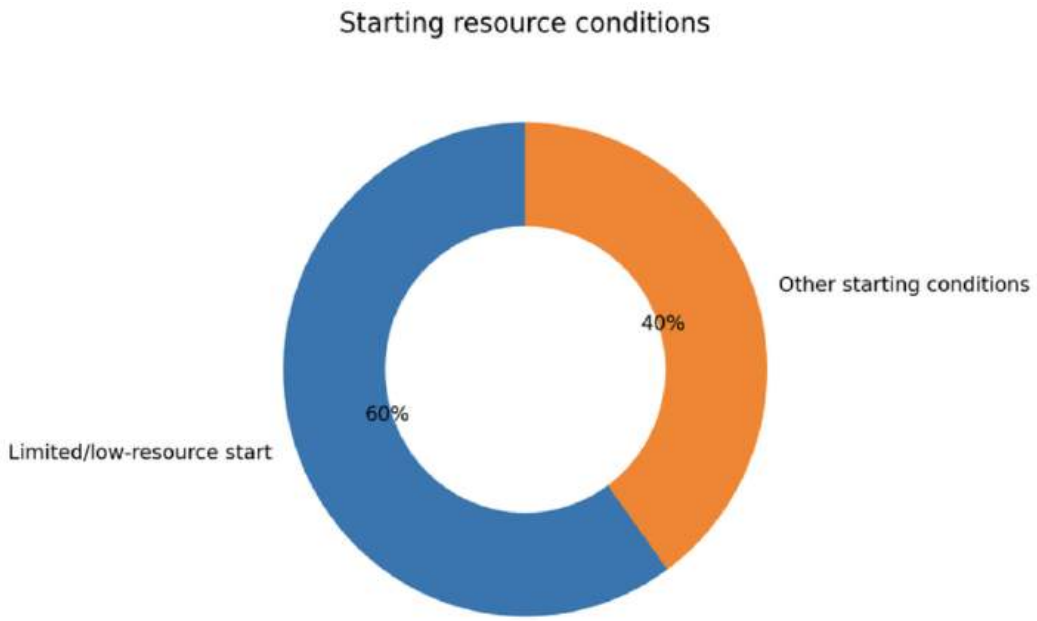


Figure 5. Approximate share classified as starting from zero or with limited initial support.

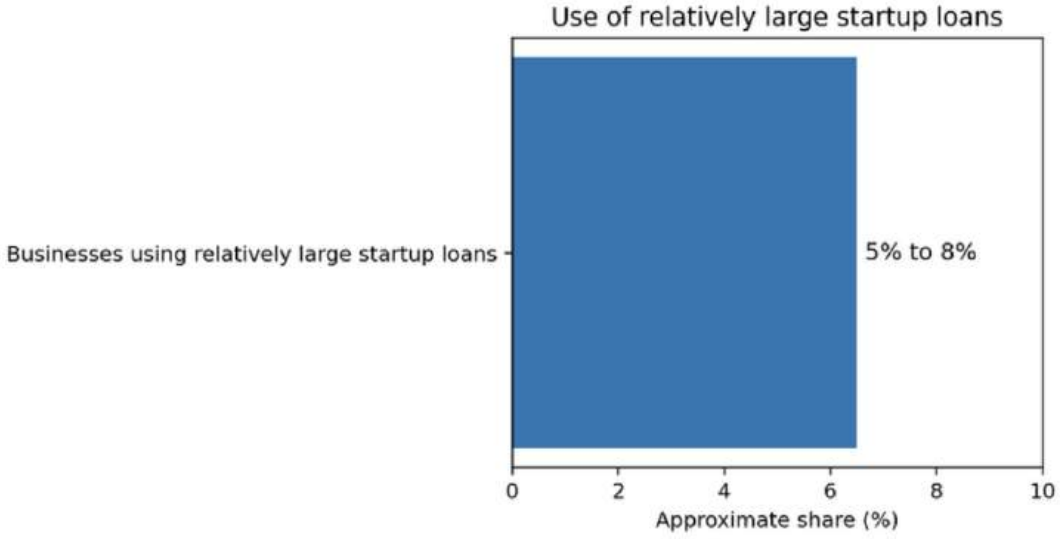


Figure 6. Approximate range of businesses reported to have used relatively large startup loans. The chart uses the midpoint of the reported 5% to 8% range for visualization.

However, some successful entrepreneurs used borrowed funds strategically, particularly to purchase products they expected to sell quickly. This distinction suggests that access to capital alone does not determine business performance. How entrepreneurs use capital may also matter.

5.2 The Importance of Experience

Experience appeared repeatedly throughout the fieldwork. Entrepreneurs who began

without previous retail experience reported difficulties with finding suitable locations, understanding customers, building a customer base, choosing products, managing inventory, understanding competition, and negotiating with suppliers. Respondents suggested that developing these capabilities often required two to five years.

In contrast, younger entrepreneurs who opened businesses between 2022 and 2025 often had several years of previous experience working in Sardoba. They described learning customer communication, product selection, supplier identification, price negotiation, inventory management, employee management, location selection, margin calculation, competitor analysis, handling difficult customers, and expansion decisions. This supports H1 at the exploratory level.

5.3 Retail Employment as Entrepreneurial Training

Several younger entrepreneurs had worked as employees in Sardoba for approximately three to five years before starting their own businesses. They generally did not receive high compensation immediately. Respondents described an initial period of approximately one to two years during which workers primarily developed skills. After gaining sufficient experience, employees could receive paid positions. Some eventually used their experience to establish their own businesses.

Retail employment → skill development → customer relationships → business knowledge → entrepreneurial entry

However, experience did not automatically lead to entrepreneurship. Some experienced employees deliberately remained employed because they feared insufficient startup capital, losing their savings, competition, or failure to generate profit. This finding is important because it shows that experience and entrepreneurial entry are different outcomes.

5.4 Communication and Customer Relationships

Communication was one of the most frequently emphasized skills in the fieldwork. Respondents described skilled sellers as people who greet customers, ask what the customer needs, listen carefully, identify the customer's actual problem, recommend appropriate products, explain differences between products, provide honest information, avoid unnecessary arguments, treat customers personally, and encourage customers to return.

Respondents reported that customers often returned to the same seller because they trusted the seller's recommendations. Many businesses reported approximately 10 to 15 loyal customers. Respondents often defined a loyal customer as someone who returned within approximately one or two days.

Communication → trust → repeat purchases → customer loyalty

However, because the current research does not have standardized customer-level

measurements, this relationship should be treated as an observed pattern rather than a statistically proven causal relationship.

5.5 Customer-Oriented Product Selection

When customers requested products that a business did not stock, sellers sometimes recorded the request and obtained the product later. Some sellers tested new products through small initial purchases. For example, if observed demand suggested that approximately 20 units were needed, a seller might initially purchase approximately 25 units. This approach allowed sellers to test demand, reduce inventory risk, respond to customer preferences, identify new trends, and avoid excessive unsold inventory. This represents a form of demand-driven inventory management.

5.6 Competition

Competition was particularly strong in Sardoba because many businesses operated close to each other and sold similar products. Respondents reported that businesses generally competed through communication, product quality, product variety, trending products, recommendations, and faster service. Price competition appeared more limited. Respondents stated that similar products often had similar prices, while differences existed in quality and service. This suggests that non-price competition may be particularly important in this retail environment.

5.7 Location and Relocation

Location emerged as another major factor. Businesses reported difficulties related to high rent, low customer traffic, distance from residential areas, competition, insufficient space, customer purchasing power, and landlord decisions. Several businesses relocated. Many reported improved customer numbers after moving to better locations.

However, relocation did not always produce positive results. Some businesses moved from smaller shops with established customer bases to larger locations farther away. After moving, they lost customers because existing customers were unwilling to travel farther. In some cases, another entrepreneur reportedly opened a business in the former location and captured the previous customer base. This suggests that location value is not determined by physical size alone. An established location can contain valuable customer relationships that are difficult to transfer.

5.8 Rent and Business Survival

Rent was repeatedly identified as a major business pressure. Sardoba rents were reported at approximately 3 million to 15 million UZS per month. Respondents explained that high rent could force sellers to raise prices or operate with very small margins. If customer traffic was insufficient, a large proportion of business resources could be consumed by rent.

Some businesses therefore moved to smaller or cheaper locations. However, cheaper rent did not automatically mean better performance. Some businesses accepted higher rent because the location generated more customers. This suggests that the relevant factor may be rent relative to customer-generating capacity, rather than rent alone.

5.9 Family Business and Family Support

Approximately 30% of the sample was described as continuing a family business. These entrepreneurs often began by working in the family shop. They gained retail experience, customer relationships, product knowledge, selling skills, and responsibility. Parents gradually gave them greater responsibility for managing the business.

This pathway differs from simply giving a young person startup capital. The family business provided an existing environment in which entrepreneurial skills could develop.

Family support also appeared in another form. Some families gave entrepreneurs time to become successful rather than immediately demanding income from them. However, family pressure could also contribute to closure decisions. Some families told struggling entrepreneurs to close their businesses because of rent, debt, or lack of profit. Interestingly, some entrepreneurs interpreted this pressure as motivation and continued operating.

5.10 Business Survival and Mentality

Approximately 40% of respondents reported having seriously considered closing their business. The reasons included high rent, competition, low customer demand, financial pressure, and family pressure.

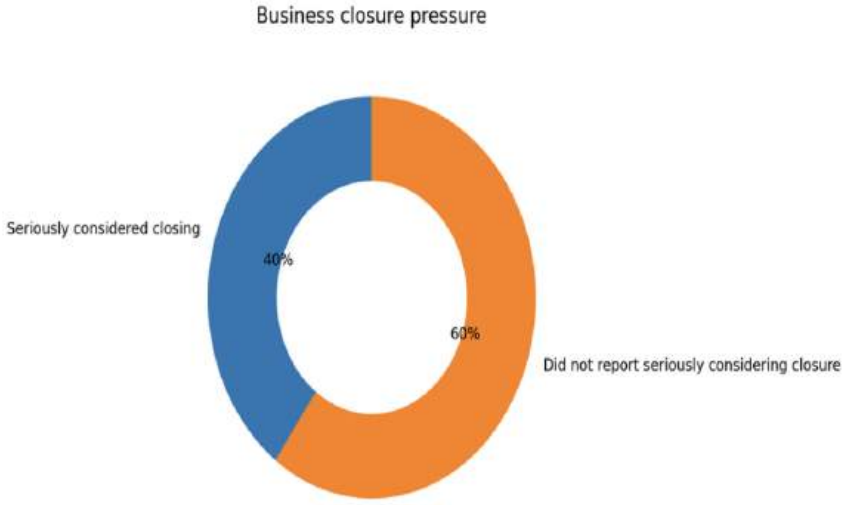


Figure 7. Approximate share of respondents who reported seriously considering closure.

However, some continued because they had already spent years developing the business, believed it could eventually become profitable, or saw opportunities created by competitors leaving. Some surviving businesses responded to competitor closures by increasing inventory,

adding products previously sold by competitors, maintaining similar prices, and continuing to operate while competitors exited.

Your observation that mentality matters is therefore relevant. For academic precision, this paper defines the construct as entrepreneurial resilience and adaptive mindset. This refers to the tendency to continue operating, adapt products and strategies, and respond to difficulties rather than immediately abandoning the business.

6. Conceptual Framework

The fieldwork supports the following exploratory model:

Starting conditions	Entrepreneurial capabiliti	Aedsaptive behavior	Business outcomes
Capital Family background Previous experience Location	Communication Customer understanding Supplier knowledge Inventory management Pricing Competitor analysis	Product experimentation Customer-driven selection Relocation Inventory adjustment Response to competitor exit	Customer loyalty Customer traffic Business stability Survival Expansion

Entrepreneurial resilience and adaptive mindset influences how the owner responds when facing rent, competition, low demand, and financial pressure.

This framework is stronger than simply saying successful businesses have a good mentality. It identifies specific mechanisms through which starting conditions may be translated into business outcomes.

7. Discussion

The preliminary evidence suggests that small retail business survival in Namangan cannot be explained by capital or location alone. Businesses operate within a combination of financial, geographic, social, and behavioral conditions.

Capital affects whether entrepreneurs can enter the market and maintain inventory. However, respondents described experience as important for converting available resources into effective decisions.

Customer relationships also appear important. Sellers who understand customer needs can respond to demand, recommend products, and develop repeat customers.

Location presents both opportunities and risks. A high-traffic location can generate customers but also impose high rent. Conversely, a cheaper or larger location can perform poorly if it separates the business from its established customer base.

The evidence also suggests that entrepreneurs respond differently to similar pressures. Some close when competition and rent become difficult. Others reduce personal consumption, change their products, continue operating, and take advantage of competitors leaving. This supports the potential importance of entrepreneurial resilience and adaptability.

8. Limitations

8.1Approximatesample size

The sample was approximately 90 to 100 businesses rather than a precisely documented random sample.

8.2 Survivorship bias

The study primarily surveyed businesses that were still operating. Businesses that had already closed were difficult to include. Therefore, the research cannot accurately estimate the overall business failure rate.

8.3 Self-reported information

Much of the research relies on respondents' descriptions of their experiences.

8.4 Limited financial information

Most respondents did not disclose profit. Therefore, the study should not claim that a particular factor directly increases profit unless additional financial evidence supports it.

8.5 Researcher judgment

Some measures, such as business success, were partly evaluated through observations of shop size, customer traffic, product variety, and communication. A future study should use standardized measurements.

8.6 Local scope

The study focuses on Namangan and cannot automatically represent all businesses in Uzbekistan.

9. Preliminary Conclusion

The field research indicates that the survival and growth of small retail businesses in Namangan are influenced by a combination of financial resources, experience, location, customer relationships, communication skills, product selection, competition, family support, and entrepreneurial resilience. The findings particularly highlight the importance of experience and customer-oriented behavior. Experienced sellers reported developing knowledge about customers, suppliers, inventory, pricing, competitors, and locations before becoming entrepreneurs.

The research also identifies an important distinction between having business resources and knowing how to use them. Entrepreneurs with limited resources could survive by adapting their products, building customer relationships, reducing personal consumption, and continuing through difficult periods.

The preliminary evidence therefore suggests that entrepreneurial resilience and adaptive behavior may play an important role in business survival. However, further quantitative analysis is necessary to determine which factors have the strongest statistical association with business outcomes.

10. Data and Figure Transparency

All charts in this report use only percentages or ranges explicitly stated in the supplied field-research narrative. No missing percentages were invented. Where the text gives an approximate range, the figure is labeled as approximate. Percentages should therefore be interpreted as descriptive fieldwork estimates rather than population estimates.

Measure	Reported value	Interpretation
Sample	Approximately 90 to 100	Exploratory field sample
Shop managers	50%	Approximate respondent share
Business owners	30%	Approximate respondent share
Sales employees	20%	Approximate respondent share
Independent starts	60%	Approximate entrepreneurial pathway
Family business	30%	Approximate entrepreneurial pathway
Substantial family resources	10%	Approximate entrepreneurial pathway
Limited/zero starting support	60%	Approximate classification
Multiple businesses	20%	Approximate owner share
Seriously considered closure	40%	Approximate respondent share
Relatively large startup loans	5% to 8%	Approximate reported share
Sardoba monthly rent	3 to 15 million UZS	Reported range
Employee wages	3 to 5 million UZS/month	Reported range
Loyal customers	10 to 15	Reported typical count per business
Experience before own business	3 to 5 years	Reported for several newer entrepreneurs
Skill-development period	1 to 2 years	Reported initial development period
Experience to develop capabilities	2 to 5 years	Respondent estimate

Suggested next-stage statistical validation

For journal submission, the next stage should convert the field notes into a respondent-level dataset. Each business should receive one row and standardized variables should measure experience, rent burden, location quality, customer loyalty, adaptive product behavior, capital access, resilience, survival status, and growth indicators. This would allow descriptive statistics, cross-tabulations, correlation analysis, and regression models to test H1 to H6 rather than presenting the hypotheses as established effects.

Prepared from the supplied field-research material. Percentages and ranges are retained as approximations where the underlying field dataset was not provided.

Paying with Less Friction: UPI and The Shift in Indian Consumer Spending
By Saksham Bhartiya

Abstract

The rapid growth of India's Unified Payments Interface (UPI) has fundamentally reshaped the country's financial ecosystem, yet existing research offers limited insight into how the accessibility of digital payment systems directly influences everyday consumer spending behaviour. This study investigates the research question: how has the easy accessibility of UPI and digital payment systems influenced consumer spending patterns among Indians? It examines three hypotheses related to the substitution of traditional payment methods, changes in transaction frequency, and shifts in transactional behaviour across consumption categories. Using a mixed-methods approach combining a survey of 46 respondents with secondary literature, the study finds that UPI primarily shapes consumer behaviour through friction reduction rather than new demand creation, encouraging higher transaction frequency, partial displacement of cash, and deeper integration of digital payments into low-value, high-frequency expenditure categories such as transport, food, and groceries. These findings carry practical implications for policymakers seeking to expand financial inclusion, businesses operating in high-UPI-penetration sectors, and researchers studying the behavioural dimensions of digital payment adoption in emerging economies.

Keywords: UPI adoption, Consumer behavior, Indian consumer, Digital Payments, Fintech

Introduction

Digital payments, specifically the Unified Payments Interface (UPI), have significantly changed consumer purchasing trends and the financial ecosystem within India. UPI enables rapid, affordable, and multi-platform transactions, utilizing an interbank payment system to become the face of India’s digital economy. In FY24 (fiscal year 2023-24), UPI recorded ₹200 lakh crore in transactions, which was a 138% increase since FY18 showing its growth. The system is used by over 20 billion transactions each month and is used for 84% of digital retail payments in India (Sinha et al. 29-30) Its expansion results in the majority of the population having access to digital payments, raising the question of how this may impact the spending habits of everyday customers.

The scale of UPI's growth is best observed through NPCI (National Payments Corporation of India) transaction data. UPI transaction values rose from ₹2,16,242.97 crore in early 2020 to ₹23,48,037.12 crore by January 2025, representing nearly an eightfold increase over five years (NPCI). By March 2026, NPCI recorded a single-month transaction volume of 22,641.11 million transactions valued at ₹29,52,542.05 crore, with 705 banks now live on UPI, reflecting both the deepening institutional integration of the platform and its continued growth as the dominant infrastructure for everyday payments in India (NPCI).

So far, research on digital payments and financial technology (fintech) in India has been bucketed into three main areas: financial inclusion, transaction efficiency, and user friendliness. Research shows that by getting rid of middlemen, decreasing wait times, and facilitating smooth payments, digital payment platforms lower transaction costs and therefore promote an increase in consumer activity (Mohn). Furthermore, research of behavioural economists has shown that online payments get rid of the "pain of paying" since they are less perceptible than cash payments, resulting in more frequent and impulsive purchases (Kenton). Research further shows that UPI's low-cost and interoperability encourages adoption across vendors of all types.

Although current research indicates a substantial correlation between financial inclusion and the adoption of digital payments, its findings about consumer spending patterns are not sufficiently thorough. According to prior research, rising income levels and macroeconomic expansion, rather than the availability of digital payments alone, are the factors to higher consumption (Kenton). Others emphasise the significance of network externalities, according to which UPI's value rises as more people tend to use it, driving up transaction volumes and overall usage (Srivastav).

By investigating how consumer purchasing habits are directly influenced by the accessibility of UPI and digital payment systems, this study aims to provide key insights into behavioural changes caused by UPI adoption. While earlier research acknowledges the importance of accessibility in boosting adoption, it does not sufficiently examine how features like low transaction fees and rapid payments impact daily spending frequency and small value transactions. Additionally, behavioural changes driven by reduced transactional barriers have received little empirical attention, particularly when it comes to consumers' everyday usage patterns. This study attempts to offer a more thorough understanding of how accessibility translates into behavioural change by concentrating on these factors.

This study's main purpose is to examine how Indian consumers' purchasing patterns have been impacted by the ease of use of UPI. It primarily aims to investigate whether enhanced convenience and decreasing transaction friction have resulted in changes in spending frequency, payment preferences, and consumption behaviour. The research question that directs this study investigates is: How has the easy accessibility of UPI and digital payment systems influenced consumer spending patterns among Indians?

Methods

This study employs a multi-method approach, combining primary survey data with secondary sources to examine whether UPI accessibility has influenced consumer spending behaviour. Primary data was collected through a standardised questionnaire distributed to 46 respondents across varying age groups, income levels, and demographics. The questionnaire was designed to capture individual payment preferences before and after UPI adoption, self-reported changes in transaction frequency, and spending behaviour across everyday consumption categories. Secondary data was drawn from government publications, Reserve Bank of India

(RBI) reports, institutional studies, and peer-reviewed research to contextualise the survey findings within broader national trends.

The sample was structured to capture variation across key demographic variables including age, income, and gender, allowing for segmented analysis beyond aggregate trends. The survey was distributed across WhatsApp groups and promoted via Instagram, and data collection was conducted over a short window of two days before the form was closed, which accounts for the relatively modest number of 46 responses. Respondents were questioned to reflect on their payment behaviour both prior and post adoption of UPI, making the survey retrospective in nature. This design enables a before and after comparison within a single cross-sectional snapshot.

Data Analysis

Hypothesis 1: The widespread availability of UPI has accelerated the substitution of traditional payment methods with digital transactions

Prior to UPI, respondents relied on cash more regularly than digital payment methods such as debit cards and netbanking, which were used less frequently daily. Cash does have inherent transactional frictions, including the need to physically handle currency, visit ATMs, and carry exact change. Additionally, it is the most psychologically transparent method of payment, resulting in the highest level of "pain of paying," an unpleasant response to spending that makes customers hesitant to initiate transactions (Loewenstein and Prelec, 4). Even before UPI, digital payments were available, but most of the respondents stated that they used them at best once a week, indicating they were not yet regular tools for daily spending.

As illustrated in Figure 1, the line chart depicts the percentage change in usage frequency across all payment methods following UPI adoption, revealing a clear and uneven redistribution of transactional behaviour. The most major trend is the 25% drop in cash usage for the "multiple times a day" category, the most dramatic drop among any method or frequency band, confirming that UPI's most direct competition was with cash at the highest frequency of use. This is supported by a corresponding 12% rise in weekly cash usage, suggesting that cash has not been eliminated as an option but has been pushed down the frequency spectrum, now reserved for less routine or more deliberate transactions.

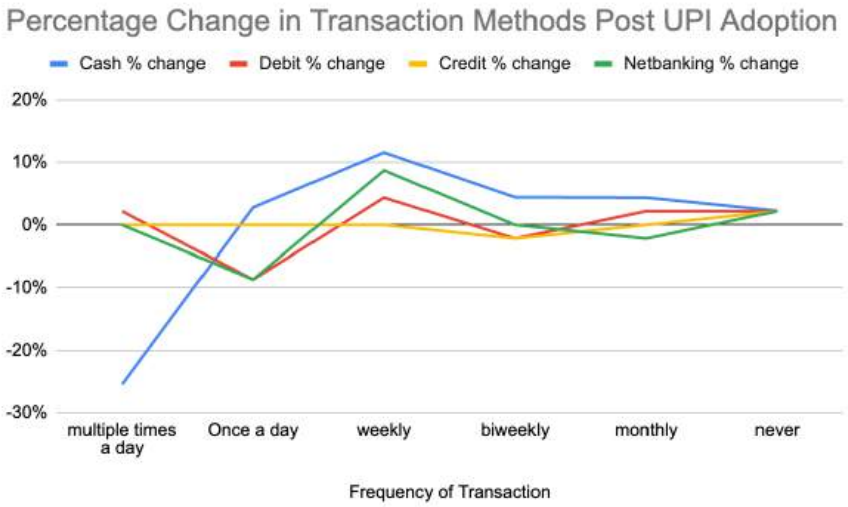


Figure 1: Percentage change in transaction methods post UPI adoption

On the other hand, debit, credit, and netbanking exhibit relatively small variations across all frequency bands, rarely surpassing 9% in any direction. The functionality linked with these approaches, which typically entail larger, more planned purchases rather than routine small-value transactions, have not been greatly impacted by UPI, according to this study. Interestingly, both debit and netbanking exhibit a 9% decrease in once-daily usage, indicating some partial substitution in that frequency band, but the impact is negligible in comparison to the displacement posited in cash. When considered together, Figure 1 lends credence to the idea that UPI's substitution effect has been strongest at the high-frequency end of the spectrum, exactly where cash historically predominated, while leaving other digital payment methods mostly unaffected.

The frequency landscape alters significantly after UPI implementation. When compared to cash at its pre-UPI peak, UPI has the highest concentration of daily users. Crucially, the use of cash has dispersed over lower frequencies rather than entirely vanishing, indicating partial displacement rather than complete substitution. Rather than just switching one-for-one, consumers appear to have added UPI transactions to their residual cash habit, which suggests that the overall daily transaction volume has increased rather than just being reallocated. This is corroborated by research on UPI adoption, which shows that users completed 40–50% more transactions than they did prior to adoption, mostly due to the removal of friction in small-value payments (Gupta, 1268).

Taken together, the evidence supports Hypothesis 1 in a nuanced way. UPI has not fully replaced cash but has clearly become the dominant payment mode for high-frequency use, displacing cash from its position as the default transactional tool. The shift is most visible at the top end of the frequency spectrum, while cash has retreated toward weekly and monthly usage patterns.

Hypothesis 2: Easy access to UPI has increased the number of daily financial transactions conducted by consumers

The respondents' self-reported evaluation of how their transaction frequency altered after employing UPI offers the clearest evidence for this theory. With 47.8% reporting a significant rise and 34.8% reporting a modest increase, over 82% of respondents claim an overall increase in frequency. Only 15.2% claim that there has been no change, and only a marginal fraction say it has decreased. This is crucial since it captures both personal attribution and observed conduct. According to the respondents, there is a perceived causative relationship rather than a coincidental correlation between the increase in transaction frequency and UPI access. This finding is in line with a poll of 276 participants, which revealed that 75% of respondents claimed that UPI had increased their expenditure (Gupta and Dev 1).

Change in transaction frequency based on age

As illustrated in Figure 2, the age-segmented data reveals two notable patterns. First, respondents under 35 report comparatively smaller increases in transaction frequency than those above 35. Across the under 18, 18-24, and 25-35 age groups, approximately 33% reported their frequency "increased a lot", with the remaining majority reporting only a slight increase or no change at all. In contrast, respondents aged 36-49 showed a notably higher proportion, with around 52% reporting "increased a lot", and this figure rises further among those aged 50 and above, where approximately 75% reported a significant increase in transaction frequency. This pattern is likely since younger consumers were already digitally active before UPI, having used netbanking, debit cards, or early mobile wallets, meaning UPI represented an incremental improvement rather than a transformative shift in how they transacted. In contrast, for respondents above 35, many of whom were predominantly cash users, experienced a more dramatic behavioural shift upon the adoption of UPI, as it offered a fundamentally different and frictionless alternative to their established habits. Second, despite this variation, a majority across all age groups still report at least a slight increase in frequency, with even the lowest-reporting groups showing combined increase rates of around 75%, confirming that UPI's impact is broad-based and not limited to any single demographic.

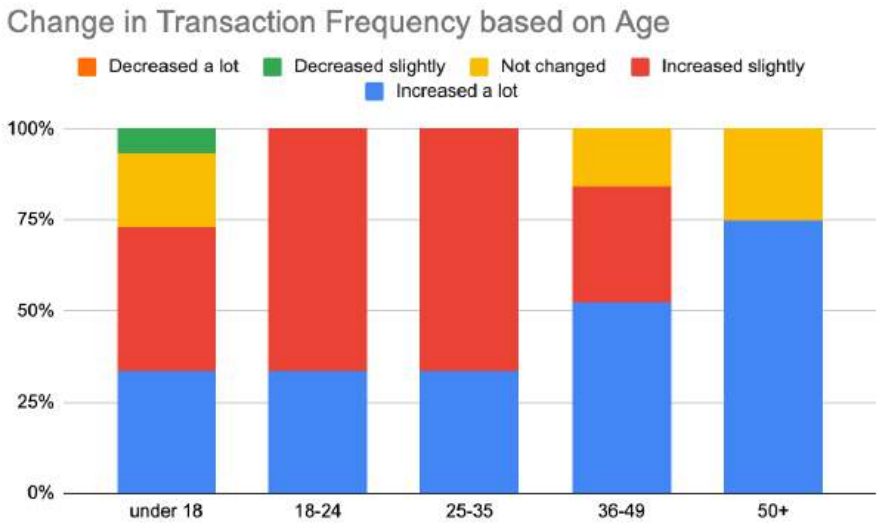


Figure 2: Change in transaction frequency based on age

Change in transaction frequency based on income

As illustrated in Figure 3, the income segmentation produces an equally interesting and somewhat counterintuitive pattern. Respondents with no income show a relatively modest increase, with only around 25% reporting "increased a lot" and a notable share reporting no change or even a decrease, likely reflecting limited spending capacity regardless of payment method availability. In stark contrast, respondents earning up to ₹50,000 annually show 100% reporting an increase in frequency, with the entire bar composed of "increased a lot", suggesting that UPI has had its most transformative access led impact at the lowest income level, enabling small-value, high-frequency transactions that were previously impractical with digital tools, such as paying an autorickshaw fare or buying street food.

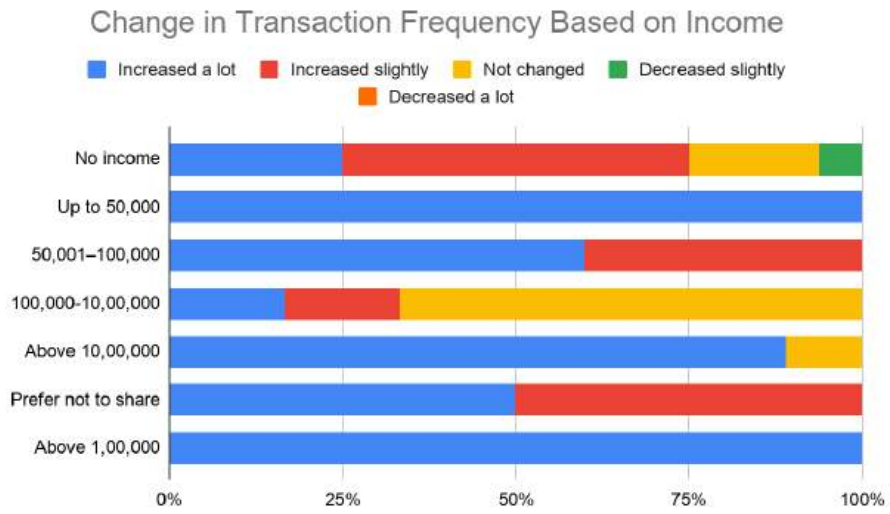


Figure 3: Change in transaction frequency based on income

The ₹50,001-₹1,00,000 bracket also shows a strong response, with approximately 60% reporting "increased a lot" and the remainder reporting a slight increase, indicating broad behavioural change across low-income groups. The most counterintuitive finding appears in the ₹1,00,000-₹10,00,000 bracket, where only around 17% report "increased a lot" and a substantial portion report no change at all, suggesting that this middle-income group, having already had access to debit cards and netbanking, experienced the least marginal shift in transactional behaviour following UPI adoption. High-income individuals earning above ₹10,00,000, however, show approximately 88% reporting an increase, with the majority in the "increased a lot" category, reflecting that UPI's speed and convenience aligns with their already active consumption lifestyle. Respondents earning above ₹1,00,000 per month show 100% reporting "increased a lot", that is the strongest response across all income groups. This suggests that UPI's impact operates through different mechanisms across income groups: access led for lower incomes and convenience led for higher incomes.

Hypothesis 3: UPI usage has altered transactional behaviour patterns across everyday consumption categories

Beyond the increase in transaction frequency, the data further reveals that UPI has become integrated into consumers' financial practices. As demonstrated in Figure 4, while UPI usage is widespread, majority of respondents continue transacting a relatively modest share of their monthly expenditure through UPI, with 30.4% reporting that only 1%-15% of their monthly spending is conducted through UPI and another 30.4% reporting 16%-30%, meaning over 60% of respondents are channelling less than a third of their monthly expenditure using UPI. This suggests that while UPI has become a habitual payment tool, it is being used predominantly for smaller and more frequent transactions rather than capturing the majority of consumers' monthly spending.

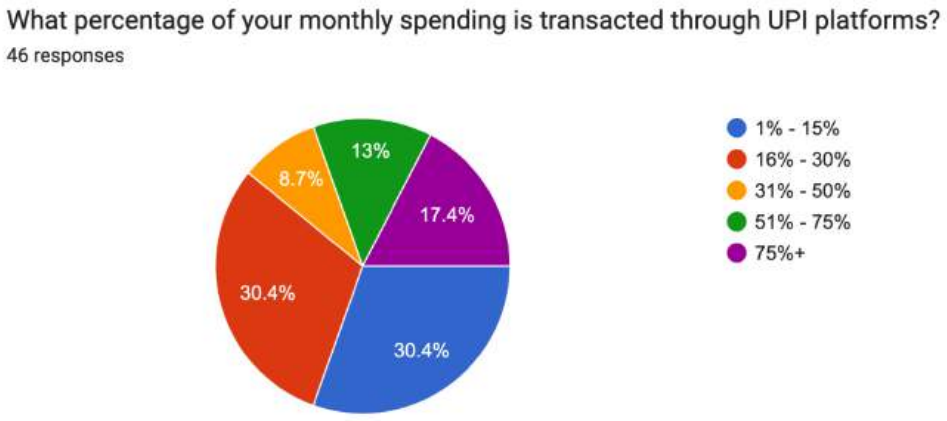


Figure 4: Percentage of monthly spending transacted through UPI

This is supported by Figure 5, which provides insight that UPI's deepest penetration is most evident in inherently low-cost, high-frequency spending categories. Transport emerges as the highest penetration category at 93.5%, followed closely by food delivery and restaurants at 91.3%, groceries at 82.6%, and peer-to-peer (P2P) transfers at 58.7%. These are the categories typically characterised by small transaction values and daily recurrence, reinforcing the idea that UPI is being used most heavily where individual transaction amounts are low, however, frequency is high. Shopping (65.2%) and entertainment (58.7%) occupy the mid-range of the radar chart, while categories involving larger or more structured payments such as bills (30.4%), health and medical (23.9%), and education (21.7%) sit at the innermost rings, indicating that UPI has not yet become the dominant method for higher-value expenses. Studies confirm that UPI has increased transaction frequency, stimulated low-value high-frequency purchases, and reduced payment friction at checkout, thus influencing consumer purchasing behaviour across both online and offline retail (Logambal et al. 898).

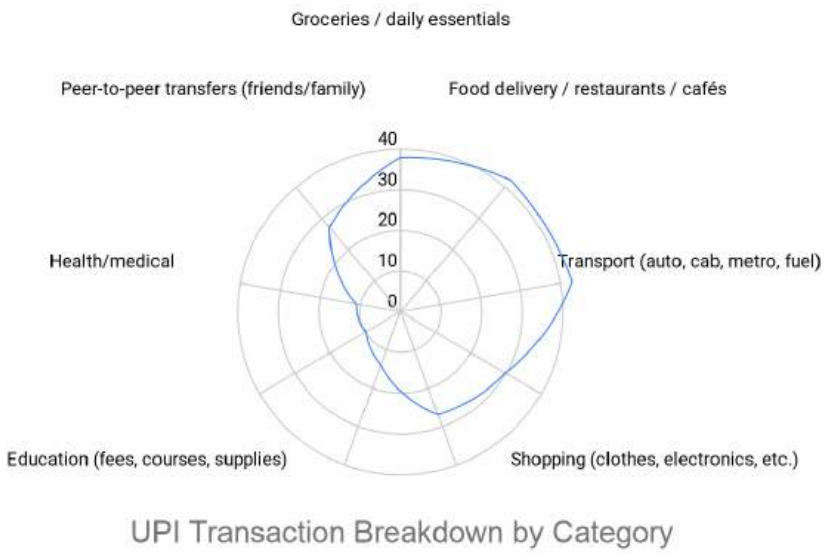


Figure 5: UPI Transaction Breakdown by Category

A significant aspect of this pattern is that the highest penetration categories are necessity based rather than discretionary in nature. This finding suggests that UPI's influence on spending behaviour is not primarily driven by impulse or luxury consumption, but by the routinisation of digital payments within everyday necessity spending. Once UPI becomes embedded in daily essentials, it naturally accumulates a high transaction count almost by default, independent of any change in underlying consumption desire. However, its growing presence in shopping and entertainment categories also hints at a secondary effect that once consumers are comfortable with UPI in necessity categories, they naturally extend it to more discretionary spending as well, gradually broadening its footprint across the full range of consumer expenditure.

Discussion

Summary of Findings

The first major takeaway is the age-related pattern in UPI adoption. The data shows that respondents aged above 35, particularly those 50 and above, report the most dramatic increases in transaction frequency following UPI adoption, with approximately 75% of those aged 50 and above reporting that their frequency "increased a lot." In comparison to younger respondents across the under 18, 18-24 and 25-35 brackets, where only around a third say the same. This could be because younger consumers were already digitally active before UPI, having used other digital payment methods, so UPI was an incremental shift rather than a transformational one for them. While, among older consumers, who depended almost exclusively on cash, the behavioural disruption upon adoption was much greater. This is consistent with the diffusion of innovation theory that states that later adopters undergo more radical behavioural shifts than early adopters who have already made incremental adjustments (Kim).

The second notable takeaway is the counterintuitive pattern across income groups. The most defining behavioural change is not in the middle income bracket, one might expect, but in the two ends of the income spectrum. All the lower-income respondents who earn up to Rs. 50,000 per annum have reported an increase in the frequency with UPI leading the way since it enables small value transactions that earlier digital tools failed to facilitate. Higher income respondents earning more than Rs 10,00,000 also have around 88% reporting an increase driven by convenience and seamless integration of UPI into an already active consumption lifestyle. The increase is the lowest with just about 17% reporting “increased a lot” for middle income respondents in the bracket of ₹1,00,000-₹10,00,000. This could be due to this group already having partial digital access via debit cards and netbanking before UPI, reducing the marginal impact of its adoption. This finding suggests that UPI's impact is not uniform but functions through different mechanisms depending on income level: access-led for lower incomes and convenience-led for higher incomes.

The third major takeaway concerns the pattern of spending categories. As illustrated in Figure 5, UPI's deepest penetration was in transport (93.5%), food delivery and restaurants (91.3%), and groceries (82.6%), all of which are low-cost, high-frequency, need based categories. This is consistent with Figure 4, which highlights that over 60% of respondents are channeling lower than 30% of their total monthly expenditure through UPI, suggesting that while UPI is used at high frequencies, it is used predominantly for smaller transactions. This combination of high frequency and low individual transaction value explains how UPI drives up transaction counts without necessarily capturing a large share of total monthly spending.

Comparison with Existing Literature

The findings of this study broadly support the existing literature on digital payment adoption and consumer behaviour in India, while adding important nuance in several areas. The literature highlights that digital payment platforms lower transaction costs by eliminating intermediaries and reducing wait times, thereby promoting increased economic activity (Mohn). This study supports this view, as the highest UPI penetration is seen in everyday transactional

categories where friction reduction has the greatest impact, namely transport, food, and groceries.

The behavioural economics literature similarly predicts that digital payments reduce the "pain of paying" in comparison to cash, leading to more frequent and potentially more impulsive purchases (Kenton). The finding that over 82% of respondents report increased transaction frequency following UPI adoption is consistent with this prediction, and the sharpest increase seen among older, previously cash-dependent respondents further validates the friction reduction argument. Research found that approximately 75% of participants in a survey of 276 respondents reported increased spending due to UPI, a figure closely aligned with the 82% found in this study (Gupta and Dev 1).

However, the finding that middle-income respondents show the least increase in transaction frequency offers a partial challenge to the broader narrative that UPI drives universal behavioural change. This suggests that the availability of prior digital alternatives may moderate UPI's impact, a nuance not sufficiently captured in studies that treat UPI adoption as uniformly transformative across all demographic groups. The literature on network externalities argues that UPI's value grows as more users and merchants adopt it, and this is reflected in the category penetration data, where UPI has become most deeply embedded in sectors with the widest merchant acceptance, such as transport and food delivery.

This study contributes to the understanding of how payment infrastructure shapes consumer behaviour, with practical implications for policymakers, fintech companies, and businesses operating in India. The finding that UPI has its strongest access-led impact among low-income groups highlights its potential as a tool for financial inclusion, enabling populations that were previously excluded from digital finance to participate in everyday digital transactions. For businesses, the category data suggests that sectors such as transport, food delivery, and groceries are where UPI integration delivers the greatest transactional volume, and investment in seamless UPI checkout experiences in these categories is likely to yield the highest returns.

Conclusion

This study examined how increased access to UPI and digital payment systems has changed consumer spending patterns among Indians. It tested three hypotheses relating to payment method substitution, transaction frequency, and category-level spending behaviour. The findings broadly support all three hypotheses.

Overall, the results suggest that UPI has altered the ease and frequency of spending more substantially than the composition of consumption. People are not purchasing different goods and services because of UPI. Rather, they are making similar purchases more often and with less friction. Cash has stepped back from its role as the default method for routine payments. After UPI adoption, 82% of respondents report higher transaction frequency, and the bulk of UPI usage sits in low-value, recurring expenditure like transport, food delivery, and groceries.

The demographic patterns reveal an interesting finding that the effects of UPI adoption are uneven across population groups. Older consumers and low-income groups, both heavily

reliant on cash before UPI, show the biggest behavioural shifts. For low-income users especially, UPI did not just make payments easier, it made certain transactions possible that may previously have been difficult. By contrast, middle-income consumers, who already had debit cards and netbanking, exhibited relatively limited behavioral change. These findings are consistent with transaction cost theory, the pain of paying literature, and network externality research would all predict, but this research adds something those studies miss: the magnitude of the shift depends entirely on consumers' prior payment practices (Mohn). For policymakers, the takeaway is clear, rural and semi-urban areas with high cash dependency are where UPI infrastructure investment will have the most impact. For businesses, the data points to discretionary categories like shopping and entertainment as the next major areas of UPI-driven growth.

The study is subject to several limitations. The sample size of 46 respondents, collected over a two-day window through WhatsApp and Instagram, limits the statistical generalisability of the findings. The sample likely skews toward urban, educated, and digitally literate respondents, which would underrepresent the experience of rural or semi-urban consumers where UPI adoption dynamics may differ considerably. The self-reported and retrospective nature of the data also introduces the possibility of recall bias and social desirability bias, as discussed in the Research Design section.

Future research could address these limitations by employing larger and more demographically diverse samples, incorporating rural and semi-urban respondents to capture a fuller picture of UPI's reach. Longitudinal studies tracking the same respondents over time, rather than relying on retrospective self-reporting, would provide stronger causal evidence of behavioural change. Additionally, combining survey data with actual transaction records from banking institutions or payment platforms would allow researchers to move beyond self-reported behaviour and verify the patterns identified here with objective data. Examining the relationship between UPI adoption and overall household expenditure levels, rather than just transaction frequency, would also offer a more complete picture of UPI's macroeconomic impact on consumer spending in India.

Works Cited

Gupta, G. G. (2025, December). IMPACT OF DIGITAL PAYMENTS & UPI ADOPTION ON FINANCIAL BEHAVIOUR. IJPREMS. Retrieved July 24, 2026, from https://www.ijprems.com/ijprems-paper/impact-of-digital-payments--upi-adoption-on-financial-behaviour

Gupta, R., & Dev, H. (2024, May). From Cash to Cashless: UPI's Impact on Spending Behavior among Indian Users May 2024. ResearchGate. Retrieved June 24, 2026, from https://dl.acm.org/doi/pdf/10.1145/3613905.3651050

Kenton, W. (2026, April 25). Consumption Function: Formula, Assumptions, and Implications. Investopedia. Retrieved July 18, 2026, from https://www.investopedia.com/terms/c/consumptionfunction.asp

Kim, J. Y. (n.d.). World Bank Group [Digital Financial Inclusion]. Retrieved March 23, 2026, from https://www.worldbank.org/en/topic/financialinclusion/publication/digital-financial-inclusion

Loewenstein, G., & Prelec, D. (1998, February 1). The Red and the Black: Mental Accounting of Savings and Debt. InformsPubsOnline. Retrieved July 24, 2026, from https://pubsonline.informs.org/doi/10.1287/mksc.17.1.4

Logambal, R., Parthiban, S., Ananthi, P., & K, G. (n.d.). A Study On The Role Of India's Digital Payment Systems In Shaping Consumer Purchase Decisions. European Economic Letters. Retrieved July 24, 2026, from https://www.eelet.org.uk/index.php/journal/article/view/4212

Mohn, E. (2023). Transaction cost theory. EBSCO. Retrieved July 18, 2026, from https://www.ebsco.com/research-starters/business-and-management/transaction-cost-theory

Sinha, J., Mehrotra, P., Ginoria, S., Vohra, J., Kanwar, A., & Mandhata, V. (2025, October 8). UPI – The Global Benchmark for Digital Payments. Boston Consulting Group. Retrieved July 18, 2026, from https://www.bcg.com/publications/2025/india-upi-the-global-benchmark-for-digital-payments

Srivastav, A. K. (2024, December 23). Network Externalities. WallStreet Mojo. Retrieved July 18, 2026, from https://www.wallstreetmojo.com/network-externalities/

Transaction Costs. (2019, November 17). CFI. Retrieved July 18, 2026, from https://corporatefinanceinstitute.com/resources/economics/transaction-costs/

Financing the Future: Implementation of Value Capture in the NYC Subway System
By Ray Zhang

Abstract

This paper aims to highlight a realistic approach for how the subway system of New York City (NYC) can improve its service to satisfy riders. It argues that the majority of current issues of the Metropolitan Transportation Authority's (MTA) stem from the systemic outdatedness of the subway system's infrastructure, through analysis of the 2025-2029 capital plan and its outlined goals for the MTA. Further analysis of the MTA's financial documents demonstrates that in order for the MTA to sustainably fund its infrastructure improvements, it needs a restructuring of its financial system to reduce its over-reliance on government subsidies. The solution is introduced by examining the success of the Hong Kong metro, specifically the success of the value capture system. Analysis of this system through the Cervero paper shows that not only does value capture allow for financial independence, it also benefits the transit agency, the city, and the people in general. Therefore, a practical implementation of value capture mechanisms in NYC is suggested, along with specific examples of when and how they should be used. This is accompanied by a discussion of previous attempts of public financing for the MTA, and the lessons NYC can learn from these. Finally, the proposed solution is evaluated with the potential drawbacks of value capture through previous literature and suggests how to correct them, thus ensuring a successful implementation.

Keywords: value capture, transportation infrastructure, public transportation, metro systems, public financing, infrastructure finance, urban planning in NYC, transit-oriented development, urban economics, New York City subway, Metropolitan Transportation Authority

Author Biography

Ray Zhang is a rising senior at the Dwight School in New York City. He is very interested in the intersection of urban planning and economics, particularly through a public policy lens. His research focuses on improving public transportation in New York City.

Introduction

As urbanization increases globally (United Nations), it becomes imperative for cities to promote sustainable urban development. This means that cities should be designed with an emphasis on efficiency and reducing waste, as well as prioritizing long-term well-being. One of the core aspects of sustainable development is the advancement of public transit systems. Public transportation not only reduces adverse environmental effects but can also reduce traffic and save money. In particular, metro systems are crucial for improving cities, especially in highly dense urban settings (Cervero et al. 109).

One of the most prominent subway systems in the world is the New York City Subway. Operated by the Metropolitan Transportation Authority (MTA), the subway provides public

transportation services to millions of people every day (Governor's Press Office). However, the subway system faces a number of issues, mainly stemming from outdated infrastructure. In addition, the MTA has consistently faced financial struggles, leaving it incapable of funding the infrastructure upgrades it desperately needs without over-reliance on the government.

In contrast, the metro system of Hong Kong is remarkably modern and efficient. However, the key aspect that distinguishes Hong Kong's system is its financial robustness. In contrast to the MTA, which relies on government subsidies to stay profitable, the MTR (Hong Kong's metro agency) uses the concept of value capture to earn revenue. Value capture refers to the practice of transportation services capturing the property value they create through the construction or upgrades of transit systems. In Hong Kong, this is done through a model known as the Rail and Property model, or R+P for short.

Through first establishing the necessity of the MTA to improve its infrastructure and therefore, its financial system, this paper then evaluates the success of the MTR through value capture in order to argue that value capture can and should be implemented in New York City. Due to the special circumstances of NYC, a mixed approach to value capture is presented, with different mechanisms for different scenarios. With reference to previous examples of projects that used approaches similar to value capture, as well as future projects, the paper outlines how the MTA could learn from the success of the MTR.

The Current Situation in New York City

Home to a population of around 8.5 million, New York City (NYC) is the largest city in the United States. Known worldwide for being a hub of finance, culture, academics, and entertainment, it is considered one of, if not the most, important cities in the world (Lankevich). However, NYC is also known for having some of the worst traffic in the world, with it being ranked #4 among the most congested cities in the world in the 2025 Inrix Traffic Scorecard (INRIX). In order to avoid the traffic, many New Yorkers take public transportation, which is provided by the MTA.

The MTA operates a vast transportation network around NYC, including the New York City Transit Authority, which operates the subway and bus system. In particular, the NYC subway system is known worldwide for its long history as well as its extensive coverage. However, the service of the subway has faced many complaints, mainly due to its outdated infrastructure. In fact, the MTA faces a systemic need to replace and repair aging infrastructure, which is the root cause of most of its service-related issues.

The 2025-2029 Capital Plan of the MTA provides insight into the extent of outdated infrastructure that is currently in use in the subway system, as well as its negative impact on various aspects of its service. The Capital Plan serves as an indicator of how the MTA plans to spend its financial resources, thus reflecting its priorities for the coming few years. It can be understood as having two main goals: to replace outdated infrastructure and to modernize the system, both of which fall under the general goal of upgrading infrastructure.

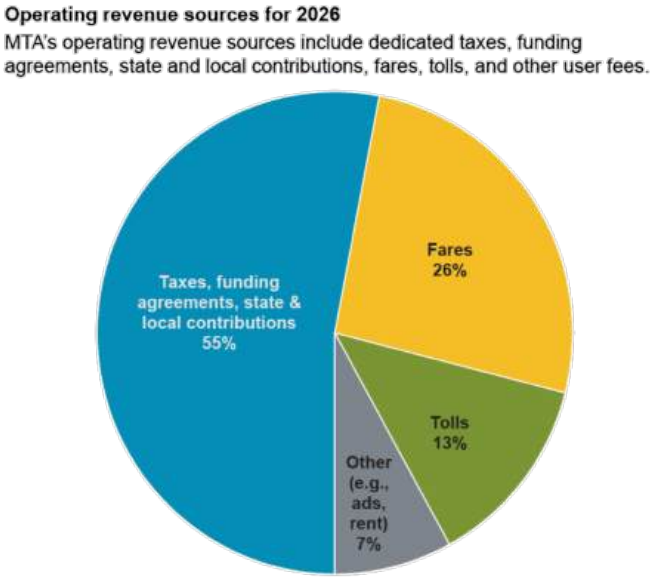
In the Capital Plan, the MTA outlines the systems that have significant antiquated infrastructure and the problems that this creates. For example, one of the most notable complaints regarding the NYC subway are delays, as around 1 in 5 trains did not arrive on time in 2024 (Office of the New York State Comptroller [OSC], *Trends in New York City Subway Delays*); this is in stark contrast to cities like Hong Kong, which has nearly 100% of trains arriving on time (Leong). One of the largest causes of delays is the signal system, which causes around 10% of all delays and 27% of major subway incidents, which is where 50 or more trains are delayed (*Trends in NYC Subway Delays*). These signal issues are mainly due to the system's age, as the 2025-2029 Capital Plan notes that some sections of the subway system still use signals that are up to 80 years old, and roughly 69% of signals still use the outdated fixed block signaling system. The modern signaling system, known as Communications-Based Train Control (CBTC), is limited to just a few select lines. Another priority of the MTA is upgrading subway cars, as around 1 in 4 subway cars are past their 40-year lifespan. The MTA notes that newer cars are more comfortable, less likely to break down (6x less), and require less frequent and costly repairs, as well as being necessary for CBTC implementation. Other highlighted infrastructure needs include upgrades to maintenance facilities, power systems, bridges, and tunnels through which the subway passes. In addition, large percentages of tracks, line equipment, and line structures also need to be replaced (MTA, *2025-2029 Capital Plan*).

The other main goal of the MTA is to modernize the infrastructure and service of the subway. In particular, the Capital Plan includes a number of ideas to improve stations and the overall accessibility of the subway, such as through the current Second Ave Line Phase 2 Expansion and the creation of the Interborough Express (IBX). These improvements to stations include adding accessibility features, such as elevators and ramps, as well as safety features. Safety has been a source of concern for many riders, as there has been an increase in transit assaults and discourse regarding the safety of the subway, particularly after the COVID-19 pandemic. However, most experts agree that while the number of violent crimes has risen since the pandemic, the total number of incidents remains very low, especially compared to the rate of violent crime outside of the subway (Mehranbod et al.). Even though the statistics do not show a significant increase in crime, public perception of the subway as dangerous has increased, potentially impacting the number of riders on the subway and damaging the MTA's image (Lane). In the Capital Plan, the MTA seeks to change this perception by adding platform fencing, upgrading and increasing security cameras, and installing modern fare gates. In particular, these gates have been a priority for the MTA, as fare evaders not only cost the MTA nearly $1 billion in revenue each year (Citizens Budget Commission) but are also linked to higher crime rates (Moses). Thus, the analysis of the Capital Plan reveals that for the MTA to achieve its goals of improving its service, it must upgrade and modernize its infrastructure.

The Financial Woes of the MTA

With such an extensive list of upgrades, the MTA's biggest constraint to achieving its goals is money. The Capital Plan is a $68.4 billion investment, which is funded by federal, state,

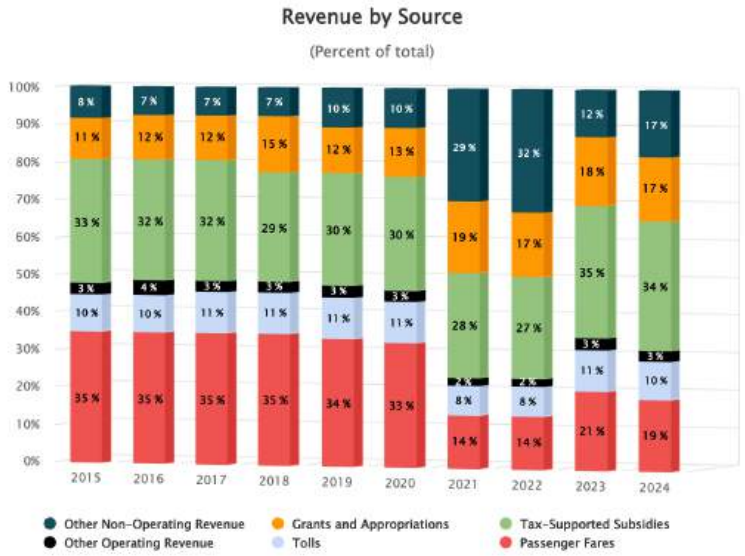
and local funding, in addition to funding from the capital lockbox (*2025-2029 Capital Plan*). The MTA Capital Lockbox is a fund that is dedicated to capital plan budgets and comprises revenue from certain taxes as well as tolls from the recently established congestion relief zone. All spending related to the Capital Plan is considered part of the capital budget, which is separate from the operating budget: the budget that finances the day-to-day operations of the MTA. Thus, the MTA must manage two separate budgets: one to maintain the current level of service and one to improve it ("Understanding the MTA's Operating and Capital Budgets"). However, an analysis of the MTA's finances reveals that it is overly reliant on subsidies and needs structural reform of its financial system.



This figure shows the breakdown of the MTA operating budget for 2026. Note that only a small percentage of the operating budget comes from fares, and none of the capital budget ("Understanding the MTA's Operating and Capital Budgets").

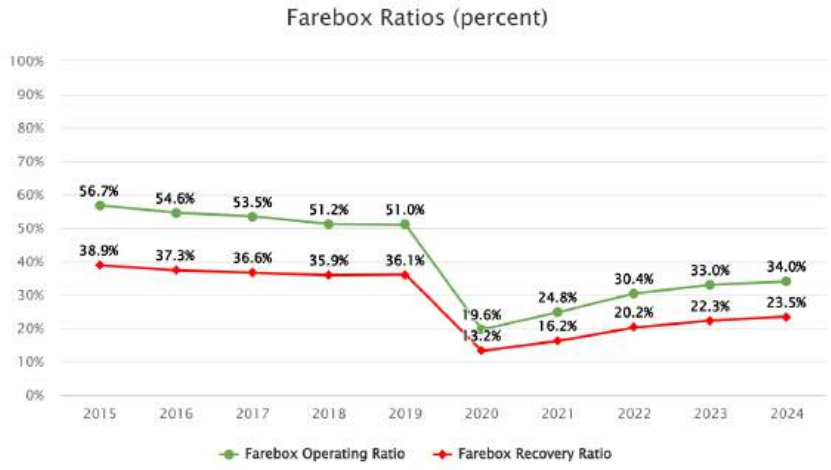
The main problem with the MTA's finances is that the operating budget is largely reliant on government funding, as well as the capital budget. This means that the MTA is extremely vulnerable to any potential delays or pauses in funding from the government. This has already been seen in the current dispute with the Trump administration over the funding of the Second Ave Subway project (Negesse). In order to understand the profitability of the MTA, one must look at its sources of income and expenses. An analysis of the most recent financial performance report of the MTA shows that it has two primary sources of income for its operating budget: revenue and subsidies. The MTA earns operating revenue from collecting fares from the subway, buses, and commuter rails, as well as tolls from the bridges and tunnels it operates. The other source of income is subsidies. These come in a variety of forms, both as direct subsidies from the state and local governments and as components of various taxes. This money is distributed through entities like the Metropolitan Mass Transportation Operating Assistance (MMTOA), which consists of various tax revenues, and other direct sources, such as the Payroll Mobility Tax

and revenue from casinos. The MTA's expenses consist of operating expenses and debt (MTA Finance Committee).



This figure shows a breakdown of the MTA's revenue by source over the past decade. Note that passenger fares have substantially declined as a source of revenue (Office of the Controller 162).

Through an analysis of the 2024 Annual Comprehensive Financial Report, it can be seen that over the past decade, the MTA has become more reliant on government subsidies and non-operating revenue streams to support its budget. This is mainly due to a decrease in the revenue it earns from fares, an ill omen for the MTA's ambitious spending goals. A transit authority that earns the majority of its revenue from fares is more financially independent, but the MTA is far from this goal.



This figure shows the trend of farebox ratios in the last decade. Note how the farebox ratios have not recovered to pre-pandemic levels (Office of the Controller 163).

The decline in revenue from fares can be understood through the farebox ratios, as shown in the graph above. The Farebox Operating Ratio is the ratio of operating expenses that fares cover, while the Recovery Ratio is the ratio of total expenses (including debt) that fares cover. The Farebox Recovery Ratio (FRR) is extremely important as it takes into account long-term costs for the subway system. Even before the pandemic, the FRR has always been less than 50%, meaning that fares do not cover even half of the MTA's long-term expenses. The main reason that the FRR has remained low after the pandemic is a sustained decrease in ridership. In 2015, the MTA estimated a total of 1.7 billion subway riders total over the course of the year ("Subway and Bus Ridership For 2019"), while in 2025, the MTA only reported around 1.3 billion subway riders (Governor's Press Office). Since the FRR serves as an indicator of how self-reliant the MTA is for its funding, its decline means that the MTA must establish a more financially robust model. In order to achieve its long-term goals of improving its service, it must look to a new source of revenue that is not dependent on government subsidies.

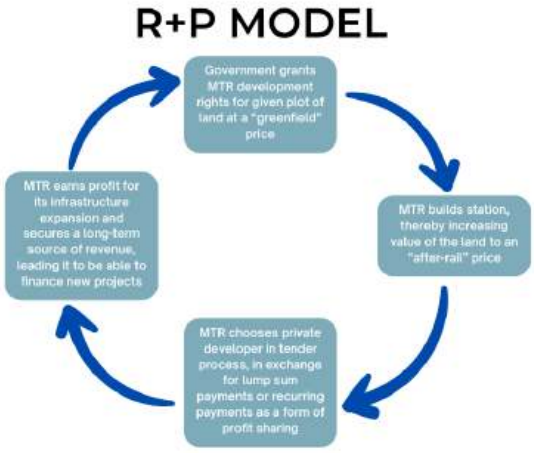
Value Capture Model

In order to find a solution to the MTA's financial problems, examples of successful subways from around the globe should be considered. In stark contrast to the MTA and its financial burdens is the Mass Transit Railway (MTR), the rapid transit system of Hong Kong. Known worldwide as one of the most efficient metro systems in the world, the MTR boasts a 99.9% on-time rate for its trains (Auyeung). It carries nearly 5 million passengers each day, around 60% of the Hong Kong population (Transport Department). However, Hong Kong's metro system is mainly distinguished by its financial robustness. In 2024, the MTR recorded a staggering HK$16 billion profit, which is ~$2 billion USD. Unlike the MTA, which only increased its net position through government subsidies, the MTR is able to earn revenue from more than transit operations (Mass Transit Railway). This means that even if the government were to cut funding, the MTR would be relatively unaffected and financially stable.

The secret to this financial independence and the MTR's success lies in the Rail plus Property model. The R+P model is a value capture model where some of the value generated from metro stations is captured by the MTR. This works since property values increase in areas where new metro stations are constructed. Instead of all this generated value being captured by private developers, the MTR retains a significant portion of it through a public-private partnership. This is particularly useful in Hong Kong, as it is one of the most densely populated cities in the world. Since land supply is limited, an increase in accessibility through the construction of a new station would lead to a sharp increase in commercial activity and property prices. In addition, the large population of Hong Kong naturally demands more rail stations and provides a high level of transit usage (Cervero & Murakami).

The way that this mechanism works is through collaboration between the government and the MTR. The government grants the MTR exclusive development rights to the land at and around where the MTR will build its station. Crucially, the MTR purchases these rights at a "greenfield" price, or at a value before development. After the MTR builds the station, thus

increasing the value of the land, the MTR then sells the development rights to private developers at a higher price (Cervero & Murakami). The MTR chooses a developer through a tender process, which allows it to control how the property will look and what kind of development will take place (Leong). Due to the population density in Hong Kong, the change in value from before rail to after rail is quite high, and the MTR manages to capture much of it. In addition, the MTR usually also has an agreement to share profits with the private developers. This could come in many forms, but it is usually either a fixed percentage or a lump sum. This means that the MTR receives value not only upfront (from selling development rights) but also on a recurring basis (from future shared profits). This model ensures the long-term sustainability of the MTR, as past expansions fund new projects (Cervero & Murakami).



Above: Diagram of how the R+P model functions, original visual

The R+P model should also be understood as a form of transit-oriented development. Transit-oriented development (TOD) is a form of urban planning in which development is centered around public transit. Specifically, the goal of TOD is to create mixed-use areas that discourage automobile use and support neighborhoods with easy accessibility to public transit stations. Transit-oriented development is increasing in popularity, as it is seen as a primary way to build sustainable cities. However, there are many obstacles to implementing TOD in the real world. The primary obstacle is conflicting interests. While in most places, the government, private developers, and transit operators have different interests, the R+P model manages to align these different actors through a public-private partnership. In places like the U.S., when a transit agency builds a new station, it often captures relatively little of the value it generates. Instead, private developers benefit without having to work with the government or rail agencies. This disincentivizes transit agencies from undertaking new projects and leads to structural losses for agencies like the MTA. In contrast, the institutional support of the R+P model allows for the successful implementation of TOD. Thus, the R+P model is not only a mechanism that

financially benefits the MTR; rather, it serves as a system that benefits society as a whole (Cervero & Murakami).

The application of value capture for transit agencies has not been limited to Hong Kong. A city facing infrastructure challenges similar to those of NYC is London, where the transit agency has used value capture to help fund major infrastructure projects (London Assembly). Notably, the London Crossrail, a major rail project that connected passengers from West to East London, was funded through a form of value capture known as a Business Rate Supplement (BRS). The BRS, a property tax on commercial buildings, was used to generate £4.1 billion for funding, although cost overruns have significantly increased this figure (Buck). A notable difference between London's use of value capture and that of the MTR is that London's approach targeted all commercial properties across London, based on the reasoning that all businesses would benefit from the increased productivity resulting from the Crossrail project. Therefore, value capture has been successful in public financing even for cities with legacy infrastructure.

Application in New York

In NYC, the ideas of value capture and transit-oriented development can and should be utilized. By helping the MTA financially, not only does the public transit system improve, but also society as a whole. Without a radical change to the MTA's financial model, the proper improvements that need to occur will never materialize. In addition, the MTA's budget may decrease in the coming years due to the overall financial situation of the New York State government. Federal budget cuts and increasing debt mean that the state of New York may have to lower subsidies for the MTA ("DiNapoli: State Faces $34.3 Billion Cumulative Budget Gap"). Thus, it becomes imperative for the MTA to become more financially independent and less reliant on government subsidies.

However, there are still many obstacles to implementing value capture in New York. In particular, the largest issue and the hardest to fix is the lack of unified institutional control. In NYC, any form of value capture would be greatly complicated by the high levels of fragmentation present. While the MTA is run by the state, any projects that are undertaken have to be approved by the city. Real estate is mostly controlled by private developers, and taxation is usually administered by both the city and state. In addition, the government in Hong Kong owned much of the undeveloped land as well as the development rights. This is not the case in NYC, where the government owns relatively little land. Finally, NYC has extremely complex zoning laws, which are regulated by the city. However, while these issues are obstacles, they do not prevent value capture entirely. Given the complexity of the situation in New York City and the diversity of potential actors involved, a hybrid approach is the most appropriate course of action. This means that unlike the R+P model, there will be no "one-size-fits-all" approach. Instead, this paper suggests several possible mechanisms of value capture that the MTA could utilize for different scenarios, depending on the target area.

In this paper, value capture has been defined as any type of public financing tool where a share of the value that is generated through transit improvements is recovered. Prior scholars

have separated value capture into different types, without an academic consensus. This paper will divide value capture mechanisms into two categories: direct and tax-based. Direct value capture includes any form of value capture where transit agencies directly collect payment from the private sector, typically in exchange for an incentive. This category encompasses joint development, density bonuses, and other mechanisms that feature a negotiable transaction between the public and private sectors. Tax-based mechanisms include all value capture models where the value returned to the transit agency is collected through the imposition or increase of a tax. Both categories of value capture are applicable in NYC and should be used in conjunction.

One of the ways value capture could be implemented in NYC is through a direct model that is similar to the R+P model in Hong Kong. Instead of providing development rights (which the NYC government does not have), the government could instead sell air rights. While the land may not be owned by the government, the city has the power to grant additional air rights. In New York, air rights refer to unused space above a plot of land that can be used for development. These air rights are measured using the Floor Area Ratio (FAR), defined as the ratio of the total area that is allowed to be developed divided by the lot area (Decode Real Estate). In NYC, FAR is particularly valuable as it allows for the construction of skyscrapers. Thus, increasing FAR is an enormous incentive for private developers.

Through a mechanism similar to R+P, the city government could grant the MTA FAR for free or at a 'greenfield' price. After the MTA constructs a station, it could then choose one or a few private developers to which it could sell the FAR, but at an increased price due to the value the station created. In practice, this would most likely be a one-time sale. While it would be more financially robust if the MTA was able to jointly develop properties or collect a percentage of future profits, it would likely require the establishment of a real estate division within the MTA. This is something that is rather unlikely in the near future but perhaps could be a long-term goal for the MTA. If the MTA managed to secure long-term returns on its property developments, it would become much more sustainable and better protected from risk.

This method, characterized by the MTA directly deciding which developers to partner with, will be known as the R+P method. This method is best suited for areas that are currently known as 'transit deserts', or areas without extensive public transportation infrastructure. In New York, these areas are often found in the outer boroughs. Thus, joint development and the MTA offering development bonuses are best suited for areas of NYC which are currently lacking urban development. These areas will have more opportunities for private developers and the public sector to create new real estate. These areas would also give the MTA more power to set prices, as the area is reliant on transit improvements to boost property value.

Thus, the R+P method could potentially be used in the planned IBX. This project is a proposed rapid transit line that will connect Brooklyn and Queens. The IBX serves a currently underserved area and will greatly reduce travel time for commuters ("Interborough Express"). The IBX is a perfect opportunity for the MTA to utilize FAR value capture because the area is currently underdeveloped. This means that the MTA expansion would lead to a large increase in property value, potentially leading to an influx of private developers. Outside of Manhattan, FAR

is generally lower due to the lack of skyscrapers (Department of City Planning). Therefore, developers could have an incentive to buy more FAR in order to construct mid-rise or high-rise buildings. In practice, the MTA would first acquire the right to increase FAR of a certain region around the new station, then find developers with which it could partner. This method is the one that gives the MTA the most control, thus it is the one that is (potentially) the most profitable. However, this is also the most difficult method and is limited in application.

Another related method of direct value capture that could be utilized is upzoning. Upzoning is the practice of changing zoning laws to increase density (Planopedia). The main form of upzoning is increasing the FAR of a given lot, but other ways include increasing maximum height or reducing constraints such as the amount of area that must be dedicated to parking. In value capture, upzoning can be offered as an incentive in areas near new metro stations. In practice, this would mean that the government or the MTA would upzone certain areas in exchange for a lump sum or recurring payment to the MTA. This differs from the previous mechanism because the MTA would not be involved in the development of the property, but only as the recipient of the benefit private developers give in the upzoning deal.

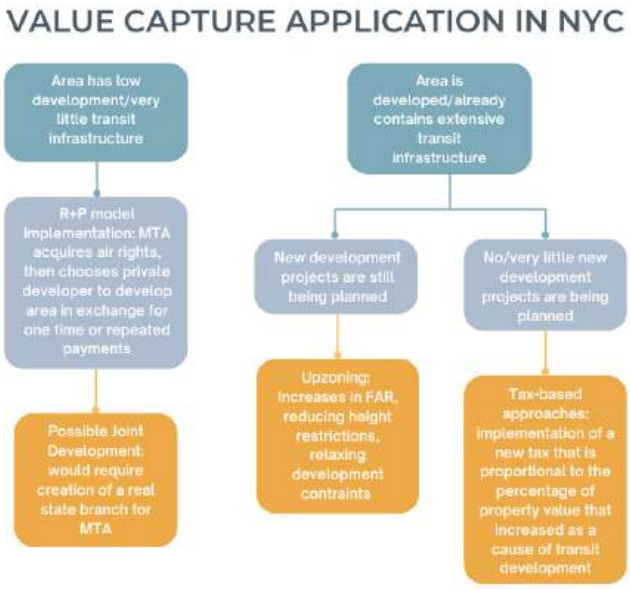
Upzoning is not a new concept, but rather a method that has already been used in NYC. It featured notably in the construction of One Vanderbilt, a skyscraper located in Midtown that was finished in 2020. The recent project faced initial hurdles regarding FAR, as it required an extremely large FAR to be profitable. This was more than the zoning requirements at the time allowed, so the city agreed to grant One Vanderbilt additional FAR in exchange for the developers to finance transit improvements at nearby Grand Central (Pham). In total, the developers spent $220 million to upgrade the facilities of the Grand Central subway station ("MTA announces Grand Opening of New Entrances"). This deal was successful, as One Vanderbilt has proved to be profitable, with a $4.7 billion valuation (SL Green Realty Corp). However, the MTA could still benefit more from these sorts of transactions. Instead of simple station renovation, the MTA should ask for funds that can be reinvested into general infrastructure improvements. This would allow the MTA more flexibility over what they want to spend their budget on, as improving stations may not be their top priority.

The mechanism of upzoning should be used for scenarios where there are new projects in already developed areas with MTA coverage. This would prevent joint development, as the existing transit infrastructure would mean that developers are not reliant on the MTA for their development to succeed. In addition, the R+P model would be used when there are multiple potential developers, while upzoning would be used when the developer has already been decided. Upzoning is all about negotiation, as too large of a request could lead to the developer deciding that it does not need the extra FAR. The government must work to balance the needs of the MTA to secure funds along with the developer's desire to earn profit.

Finally, the MTA could use a wide range of tax-related value capture methods. Predominantly, this is done through Tax Increment Financing (TIF). TIF is a mechanism in which infrastructure investments are financed through debt (bonds) and paid for with tax revenue after development. This type of method has been used in NYC, most notably at Hudson Yards,

where a crucial part of development was the extension of the 7 train to improve accessibility. In order to finance this project, the NYC government created the Hudson Yards Infrastructure Corporation (HYIC) which issued bonds (Ten Year Capital Strategy). After various funds and financial institutions bought the bonds, the new subway station was completed. Then, private developers began to construct high-rises. In exchange for the value created by the metro station, the developers had to pay the city. This money was then used to pay off the bonds, so the private developers indirectly financed the MTA expansion. In a proper TIF system, the money that the developers paid would be in the form of taxes. In the case of Hudson Yards, developers paid Payments in Lieu of Taxes (PILOTs) instead. PILOTs are payments that go directly to repaying bonds and are considered more predictable and stable than traditional property taxes (Municipality of Princeton). Nevertheless, the Hudson Yards project did not generate enough revenue during its early years due to the 2008 Global Financial Crisis, which meant that the city faced a major risk. Eventually, Hudson Yards became successful, but this experience pointed out the risk of TIF-style value capture in comparison to R+P.

Tax-based methods should be used for areas that have existing transit infrastructure and do not have significant new development projects. The case of Hudson Yards demonstrated the pitfalls of TIF-style mechanisms, particularly the level of risk for the public sector. With the current financial precarity of the MTA, an inability to pay back bonds would be catastrophic for the agency. Instead, tax-based value capture should simply be an increase in taxes for a defined region near a subway station. The amount of this tax could be calculated through a comparison of the property values of the region before and after the construction of the station. The tax would then be proportional to the percentage increase in value due to transit infrastructure for a given property. Although there is always hesitancy about raising taxes, New York is key for businesses; corporations are unlikely to leave unless the value capture tax is unreasonably expensive. As long as an appropriate combination of the suggested value capture methods is used, the city could generate a robust revenue source that can fund upgrades to the MTA's infrastructure and improve the quality of its service.



Above: Visual of how and when each value capture mechanism should be used, original visual

Evaluation & Literature Review

While this paper has focused on value capture, there are other methods and steps that the MTA should take to improve its financial position. One of the most prominent issues is the extremely high cost of construction and maintenance in NYC, for which there are various solutions. The Manhattan Institute has suggested nightly shutdowns to cut down maintenance costs (Harris). While a full shutdown might prove problematic due to the constant flow of people, partial shutdowns could and should be implemented as a way to cut costs. NYU's Marron Institute has also suggested that by giving the MTA more authority in planning projects, costs can be consistently reduced, especially for large-scale infrastructure projects (Goldwyn et al.). Another proposed action for the MTA is asset recycling (Breydo). While asset recycling can generate significant revenue for the MTA, it still does not solve the fundamental issue: the MTA is not incentivized to improve transit because it loses money on new projects. Thus, the most important solution for the MTA must be to implement a policy of value capture that creates a long-term source of funding for the agency.

Previous research on value capture has shown that its success depends on strong legislation (Sepe). Without proper legal support, value capture will not be able to achieve its goal of institutional change. An analysis of six transit agencies in Europe and North America, including the 7 line extension in NYC, found that four of the six cases of value capture relied on institutional change, through the creation of new authorities that manage the planning of a region (Salon et al.). However, NYC was not one of these cases; in fact, institutional challenges were highlighted as the largest issue to the implementation of value capture. This is mainly due to the

complications between the state and the city government, both of which need to approve any potential value capture mechanisms. Additionally, government officials have expressed hesitation in interviews regarding further instances of value capture (van Zoest and Daamen).

It is also important to note the criticism of value capture regarding its potential to induce gentrification and displacement in affected neighborhoods, particularly in transit-desert regions. While studies regarding the nature of the link between transit spending and gentrification are not conclusive, evidence suggests that value capture policies often lead to the displacement of lower-income residents. In particular, Zuk et al. found that across studies, new residents were consistently wealthier and more educated than the residents they replaced, even in studies that found no statistically significant increase in displacement. Thus, it is necessary for any application of value capture in NYC to mitigate this effect as much as possible, such as through the allocation of affordable housing and the protection of key historic neighborhoods. In addition, as upper-income residents move into new upzoned housing, they vacate their previous units, which can become more affordable housing over time. This process, known as filtering, is most applicable when large-scale upzoning occurs, as the increase of the housing market supply may be significant enough to lower existing prices (Büchler & Lutz).

However, not all developments are negative. A recent bill has been introduced in the New York State Senate to make value capture for the MTA permanent and easier to implement (Senate Bill S4203). There is also increasing academic research on what factors allow for the success of value capture. In addition to institutional change, Van Zoest & Daamen have suggested that the support of the beneficiary (private sector) is key to the expansion of value capture. Overall, as value capture becomes increasingly relevant, institutional change and public opinion will follow.

Conclusion

While there are still significant barriers to value capture, a hybrid approach with these ideas has the potential to be successful in NYC. Through effective collaboration between the public and private sectors, the city can adapt lessons from the success of Hong Kong and create a financially self-sustaining metro system. By learning from previous examples of value capture in New York, as well as case studies from other cities around the world, the MTA can effectively choose a value capture mechanism for new projects. Over time, successful instances of value capture will allow for systemic change. Each new usage will face fewer hurdles until eventually, it becomes ingrained in the financial model of the MTA. During this period of change, it is key to reduce bureaucratic elements as much as possible. Interagency communication is always difficult, and efficiency will be vital to establish a proper reinvestment vehicle. In addition, steps should be taken to prevent corruption and other instances of misconduct that may occur with the addition of a new source of revenue. By keeping in mind and preventing these issues, the MTA will be able to revolutionize its financial model, improving it for the future.

This paper has shown that the MTA is in a situation where it must improve its infrastructure so that it can achieve its goal of providing proper transportation to NYC, and that it must first reform its funding system to afford these upgrades. By analyzing the successful

implementation of value capture in Hong Kong, London, and previous examples in New York, it is clear that value capture can be implemented on a large scale in NYC. Through a mixed approach of different mechanisms, value capture can generate a dependable source of revenue for the MTA. While there are definitely concerns with value capture, as well as other approaches to the MTA's financial problem, these are not reasons for it to be ruled out. Instead, with proper implementation, as well as clear guidelines and communication, value capture can become the most effective tool for the MTA's future.

Works Cited

Auyeung, Rex. Chairman's Letter. MTR, 7 Mar. 2024, https://www.mtr.com.hk/archive/corporate/en/investor/annual2023/E10.pdf.

Breydo, Lev. "Infrastructure Finance for the Public Good: How Asset Recycling Can Untangle the New York MTA's $50 Billion Debt Load." Journal of Law and Mobility, vol. 2024, 2024, University of Michigan Law School Scholarship Repository, https://repository.law.umich.edu/jlm/vol2024/iss1/2.

Buck, Martin. "Crossrail Project: Finance, Funding and Value Capture for London's Elizabeth Line." Proceedings of the Institution of Civil Engineers - Civil Engineering, vol. 170, no. 6, Nov. 2017, pp. 15–22, https://doi.org/10.1680/jcien.17.00005.

Büchler, Simon, and Elena Lutz. "Making Housing Affordable? The Local Effects of Relaxing Land-Use Regulation." Journal of Urban Economics, vol. 143, Sept. 2024, article 103689, https://doi.org/10.1016/j.jue.2024.103689.

Cervero, Robert, et al. Beyond Mobility: Planning Cities for People and Places. Island Press, 2017, pp. 1–11, 58, 109-131.

Cervero, Robert, and Jin Murakami. "Rail + Property Development: A Model of Sustainable Transit Finance and Urbanism." UC Berkeley Center for Future Urban Transport: A Volvo Center of Excellence, May 2008, eScholarship, https://escholarship.org/uc/item/6jx3k35x.

Citizens Budget Commission. No Fare: The Costs of MTA Fare and Toll Evasion. CBCNY, 11 Sept. 2025, https://cbcny.org/research/no-fare.

Comrie, Leroy. Senate Bill S4203. New York State Senate,
3 Feb. 2025, https://www.nysenate.gov/legislation/bills/2025/S4203.

Decode NYC. "Air Rights New York City - a Detailed Guide." Decode Real Estate, 2022, https://decodenyc.com/blog/air-rights-new-york-city-a-detailed-guide.

Department of City Planning. "Residence Districts." NYC.gov, 2025, https://www.nyc.gov/content/planning/pages/zoning/zoning-districts-guide/residence-districts.

"DiNapoli: State Faces $34.3 Billion Cumulative Budget Gap Through State Fiscal Year 2029." Office of the New York State Comptroller, 8 Aug. 2025, https://www.osc.ny.gov/press/releases/2025/08/dinapoli-state-faces-343-billion-cumulative-budget-gap-through-state-fiscal-year-2029.

Goldwyn, Eric, et al. "The New York Case." Transit Costs Project, NYU Marron Institute of Urban Management, 11 Feb. 2023, https://transitcosts.com/city/newyork-case-study/.

Governor's Press Office. "Governor Hochul Highlights Record-Breaking Year of Performance and Ridership for the MTA in 2025." New York State Governor, 2025, https://www.governor.ny.gov/news/governor-hochul-highlights-record-breaking-year-performance-and-ridership-mta-2025.

Harris, Connor. Five Cheap Ways to Improve NYC Subway Operations. Manhattan Institute, July 2020, https://media4.manhattan-institute.org/sites/default/files/five-ways-to-improve-nyc-subway-operations-CH.pdf.

INRIX. "Global Traffic Scorecard." INRIX, 2025, https://inrix.com/scorecard/.

"Interborough Express." MTA, 2026, https://www.mta.info/project/interborough-express.

Lane, Charles. “New Yorkers Feel Somewhat Safer in Public, but Subway Fears Persist, Survey Finds.” Gothamist, 5 June 2025, https://gothamist.com/news/new-yorkers-feel-somewhat-safer-in-public-but-subway-fears-persist-survey-finds.

Lankevich, George. “New York City.” Encyclopædia Britannica, 28 July 1998, https://www.britannica.com/place/New-York-City.

Leong, Lincoln. “The ‘Rail Plus Property’ Model: Hong Kong’s Successful Self-Financing Formula.” McKinsey & Company, 2 June 2016, https://www.mckinsey.com/capabilities/operations/our-insights/the-rail-plus-property-model.

London Assembly, Budget and Performance Committee. Euston We Have a Problem: Mind the Funding Gap. Nov. 2025, https://www.london.gov.uk/sites/default/files/2025-12/Capital%20funding%20and%20delivery%20report%20-%20Final.pdf.

Mass Transit Railway. Consolidated Statement of Profit or Loss. MTR Corporation Limited, 2025, https://www.mtr.com.hk/archive/corporate/en/investor/annual2024/E30.pdf.

Mehranbod, Christina, et al. “New York City Sees Uptick in Transit Assaults Since Pandemic Start.” Columbia University Mailman School of Public Health, 29 Apr. 2024, https://www.publichealth.columbia.edu/news/uptick-nyc-transit-assault-rate-during-covid-pandemic-has-not-returned-pre-pandemic-levels-despite-subway-safety-plan.

Metropolitan Transportation Authority. 2025-2029 Capital Plan: The Future Rides With Us. 17 June 2025, https://www.mta.info/document/174186.

Moses, Dean. “Fare Evaders Driving Increase in More Serious NYPD Transit Crime, Officials Say.” amNewYork, 11 Feb. 2024, https://www.amny.com/news/fare-evaders-nypd-transit-crime-increase/.

“MTA Announces Grand Opening of New Entrances to Grand Central From One Vanderbilt Avenue.” MTA, 9 Dec. 2020, https://www.mta.info/press-release/%20mta-announces-grand-opening-of-new-entrances-to-grand-central-from-one-vanderbilt-avenue.

MTA Finance Committee. Financial Performance Report. MTA, 23 Mar. 2026, https://www.mta.info/document/201741.

Municipality of Princeton. “Frequently Asked Questions.” Municipality of Princeton, 2025, https://www.princetonnj.gov/m/faq?cat=47.

Negesse, Gelila. “MTA Sues Trump Administration for Second Avenue Subway Funding.” Politico, 17 Mar. 2026, https://www.politico.com/news/2026/03/17/mta-sues-trump-administration-00832615.

Office of the Controller. Annual Comprehensive Financial Report: For the Years Ended December 31, 2024 and 2023. 27 June 2025, https://www.mta.info/document/178446.

Office of the New York State Comptroller. Trends in New York City Subway Delays. Sept. 2025, https://www.osc.ny.gov/files/reports/pdf/report-10-2026.pdf.

Pham, Diane. “One Vanderbilt Tower Receives Unanimous Approval From City Council.” 6Sqft, 27 May 2015, https://www.6sqft.com/one-vanderbilt-tower-receives-unanimous-approval-from-city-council/.

Planopedia. “What Is Upzoning?” Planitzen, https://www.planetizen.com/definition/upzoning.
Salon, Deborah, et al. “Can Location Value Capture Pay for Transit? Organizational Challenges of Transforming Theory into Practice.” Urban Affairs Review, vol. 55, no. 3, June 2017, pp. 743–71, https://doi.org/10.1177/1078087417715523.
Sepe, Alexander. City Planning With Land Value Capture: A Comparative Analysis of Capture Policies in New York, London and Toronto. 2019, https://yorkspace.library.yorku.ca/server/api/core/bitstreams/2760dfc3-f1e1-4aa3-8be4-77eb97352a6e/content. Major Paper.
SL Green Realty Corp. “SL Green and Mori Building Co., Ltd. Complete Investment in One Vanderbilt Avenue.” SL Green, 21 Nov. 2024, https://slgreen.com/sl-green-and-mori-building-co-ltd-complete-investment-in-one-vanderbilt-avenue/.
“Subway and Bus Ridership For 2019.” MTA, 2020, https://www.mta.info/agency/new-york-city-transit/subway-bus-ridership-2019.
Ten Year Capital Strategy. “Hudson Yards Infrastructure Corporation.” Ten Year Capital Strategy, City of New York, 2023, https://tycs.planning.nyc.gov/hudson-yards-infrastructure-corporation.
Transport Department. Railways. Hong Kong Transport Department, 2025, https://www.td.gov.hk/en/transport_in_hong_kong/public_transport/railways/index.html.
“Understanding the MTA’s Operating and Capital Budgets.” MTA, 19 Feb. 2026, https://www.mta.info/article/understanding-mtas-operating-and-capital-budgets.
United Nations. “Urbanization.” United Nations, 2018, https://www.un.org/development/desa/pd/content/urbanization-0.
van Zoest, S., and T. A. Daamen. “Explaining Value Capture Implementation in New York, London, and Copenhagen: Negotiating Distributional Effects.” Urban Affairs Review, vol. 60, no. 5, 2024, pp. 1349–81, https://journals.sagepub.com/doi/epub/10.1177/10780874231222174.
Zuk, Miriam, et al. “Gentrification, Displacement, and the Role of Public Investment.” Journal of Planning Literature, vol. 33, no. 1, July 2018, pp. 31–44, https://www.urbandisplacement.org/wp-content/uploads/2021/08/gentrification.pdf.

Works Consulted

Reddy, A. V., et al. “Subway Productivity, Profitability, and Performance: A Tale of Five Cities.” Transportation Research Record: Journal of the Transportation Research Board, vol. 2143, no. 1, 2010, pp. 48–58, https://doi.org/10.3141/2143-07.
Sharma, Rohit, and Peter Newman. “Urban Rail and Sustainable Development Key Lessons From Hong Kong, New York, London and India for Emerging Cities.” Transportation Research Procedia, vol. 26, 2017, pp. 92–105, https://doi.org/10.1016/j.trpro.2017.07.011.
Xu, Yidan. Under What Circumstances Will Land Value Capture Work to Finance Public Transit? Based on Case Studies of Hong Kong, Tokyo and New York City. 2015. Master’s Thesis. https://doi.org/10.7916/d8df6qc4.

Cashless Economies and Consumer Spending: A Comparative Study of Kazakhstan and the United States By Islam Gumar

Abstract

The rapid growth of digital payment technologies has changed how consumers purchase goods, transfer money, and manage everyday finances. This paper examines the relationship between cashless payment adoption and consumer spending behavior through a comparison of Kazakhstan and the United States. Kazakhstan has developed a highly integrated digital financial environment, particularly through platforms such as Kaspi.kz, which combine payments, banking, shopping, transfers, and other financial services. The United States has a more fragmented payment system in which consumers use credit and debit cards alongside services such as Apple Pay, Google Pay, PayPal, Venmo, and Zelle. Existing research suggests that cashless payments can make transactions faster and more convenient, but they can also reduce the psychological awareness associated with spending physical money. Studies of digital payments have connected this lower payment salience with increased consumption and, in some circumstances, overspending. At the same time, digital payments can support financial inclusion, electronic commerce, and economic efficiency. By comparing Kazakhstan and the United States, this paper argues that the effects of cashless payments depend not only on payment technology itself, but also on financial knowledge, consumer behavior, access to credit, and the structure of each country's digital payment ecosystem.

1. Introduction

Money has changed significantly during the past several decades. Although cash remains an important form of payment, consumers increasingly use debit and credit cards, mobile wallets, banking applications, QR codes, and instant electronic transfers to complete everyday transactions. A person can now purchase groceries, order food, pay utility bills, send money to another person, or shop online without physically handling money. This transition has made payments faster and more convenient, but it has also raised questions about whether the method people use to pay can influence how much they spend. Research in behavioral economics suggests that payment methods can affect consumer decision-making because cash and digital payments create different psychological experiences. When consumers pay with cash, they physically see money leaving their possession. Digital payments can make this exchange less visible. Schomburgk et al. (2024) describe this relationship as the "cashless effect," or the tendency for consumers to spend more when using cashless payment methods rather than cash. Their meta-analysis brought together evidence from 71 studies, providing broader support for the existence of the effect while also showing that its strength varies depending on the situation and has weakened over time. This evidence suggests that the transition toward cashless payments is not only a technological change. It may also affect consumer behavior and spending decisions. Kazakhstan and the United States provide an interesting comparison because both countries have experienced significant growth in digital payments, but their payment environments have

developed differently. Kazakhstan has undergone particularly rapid financial digitalization and has become strongly associated with integrated financial platforms such as Kaspi.kz. Within one digital ecosystem, consumers can make payments, transfer money, shop online, use banking services, and access other financial products. Recent research on Kazakhstan's fintech development identifies Kaspi Pay as an especially important part of the country's digital financial environment (Almabekova & Jaxybekova, 2025). The United States, in comparison, has a more fragmented payment system. American consumers can choose among traditional credit and debit cards, bank transfers, Apple Pay, Google Pay, PayPal, Venmo, Zelle, and other services. Digital payments in the United States have therefore developed alongside an already mature banking and credit card system rather than primarily around a single integrated platform. These differences make the two countries useful cases for examining how different forms of financial digitalization can shape consumer experiences.

The purpose of this paper is to compare the development of cashless payments in Kazakhstan and the United States and review existing evidence on how digital payment technologies may influence consumer spending behavior. The comparison focuses on three main areas: the structure and development of digital payment ecosystems, the behavioral effects of cashless transactions on consumers, and the broader economic benefits and risks associated with payment digitalization. The paper does not assume that using Kaspi, Apple Pay, PayPal, or another digital service automatically causes consumers to overspend. Instead, it considers how payment convenience, financial knowledge, mental accounting, access to credit, and the reduced visibility of money may influence spending decisions. This approach is important because existing research shows that the relationship between digital payments and consumption is complex. Cashless technologies can increase economic efficiency and improve access to financial services, while at the same time creating conditions in which consumers may pay less attention to individual expenditures (Ahn & Nam, 2022; Hou et al., 2021; Schomburgk et al., 2024).

2. Digital Payments and Consumer Behavior

To understand why digital payments might affect consumer spending, it is useful to consider the psychological experience of making a payment. Paying with physical cash creates a direct and visible exchange. A consumer gives up bills or coins at the same moment that a product or service is received. Because the loss of money is immediately noticeable, cash can make the cost of a purchase psychologically clear. Digital payments change this experience because the exchange becomes less physical. Tapping a smartphone, scanning a QR code, or selecting a saved payment method can complete a transaction within seconds without the consumer physically seeing money disappear. Behavioral economists have studied this relationship through concepts such as mental accounting and the "pain of paying." Prelec and Loewenstein (1998) explain that consumers psychologically connect payments with consumption and that changing the timing or visibility of payment can alter this relationship. When the

payment becomes less noticeable or more separated from consumption, the psychological discomfort associated with spending can become weaker.

More recent research provides evidence that these behavioral mechanisms remain relevant in the digital economy. Hou et al. (2021) studied digital payments and household consumption through the framework of mental accounting and found that digital payment adoption was associated with increased household consumption. Their interpretation suggests that digital payment methods can change how consumers mentally organize and evaluate expenditures. Evidence from the United States also supports a relationship between mobile payment use and spending behavior. Ahn and Nam (2022) analyzed data from 21,457 consumers in the U.S. National Financial Capability Study and found that mobile payment users were more likely to report behaviors associated with overspending. Financial knowledge moderated this relationship, meaning that consumers with stronger financial knowledge were better positioned to control some of the risks associated with convenient payment technology. However, financial knowledge did not completely eliminate the relationship between mobile payments and overspending. These findings are especially useful for the present comparison because they suggest that technology and individual financial characteristics can influence spending behavior at the same time.

The broader literature also suggests that the relationship between cashless payments and spending is not limited to one particular country or payment technology. Schomburgk et al. (2024) synthesized a large body of previous research and found an overall cashless effect, although the size of the effect differed across studies and contexts. More recent research has also examined the psychological processes behind digital spending. Faraz and Anjum (2025), for example, investigated digital payment use and impulse buying and emphasized the importance of the reduced visibility of money during digital transactions. Together, these findings provide a reasonable theoretical explanation for why a consumer might behave differently when paying digitally. A purchase of the same monetary value can feel different depending on whether a consumer hands over physical cash or completes the transaction with a quick digital action.
At the same time, the relationship between digital payments and consumer behavior should not be interpreted as evidence that cashless technology is inherently harmful. Digital payments provide substantial benefits to consumers and businesses. They can reduce the time required to complete transactions, facilitate online shopping, provide electronic records of purchases, allow people to transfer funds across long distances, and connect consumers to additional financial services. Digital payments can also contribute to financial inclusion when they make formal financial tools available to people who previously faced barriers to traditional banking. Research on fintech increasingly shows that digital payments can function as an entry point to broader financial participation (Hong et al., 2025). The important issue, therefore, is the tradeoff created by convenience. Reducing transaction friction makes legitimate economic activity easier, but the same reduction in friction may remove some of the psychological barriers that encourage consumers to think carefully before spending. Understanding this tradeoff is essential when

comparing Kazakhstan and the United States because both countries have embraced digital payments while developing very different payment ecosystems.

3. Kazakhstan's Transition Toward a Cashless Economy

Kazakhstan provides an important example of rapid financial digitalization in a relatively young market economy. Over a comparatively short period, everyday financial activity has moved increasingly toward mobile banking, electronic transfers, QR payments, and digital commerce. A central feature of this transformation is the growth of integrated fintech platforms, particularly Kaspi.kz. Kaspi differs from a simple mobile wallet because its ecosystem combines several activities that consumers might otherwise perform through separate companies or applications. Users can make payments, transfer money, access banking products, shop through an online marketplace, and interact with merchants through the same broader platform. Research examining fintech development in Kazakhstan identifies Kaspi Pay as an important part of the country's financial technology environment and connects the expansion of digital financial services with growing consumer adoption and trust (Almabekova & Jaxybekova, 2025). As a result, the transition away from cash in Kazakhstan is closely connected to the development of an integrated digital financial ecosystem rather than simply to the replacement of cash with payment cards.

This integration has several potential economic advantages. When consumers can make purchases, send transfers, pay bills, and access financial services through a familiar digital environment, the time and effort required to complete transactions can decrease. Businesses can also benefit when digital systems make it easier to receive payments and participate in electronic commerce. In addition, electronic transactions create digital records, which can make financial activity more visible within the formal economy. Research on digital financial inclusion in Kazakhstan has emphasized the importance of consumer trust and platform-based financial intermediation in the expansion of electronic payments (Imanbayeva et al., 2026). Digital platforms can therefore contribute to economic modernization not only by replacing cash but also by connecting consumers, merchants, banking services, and online commerce. This is especially significant for Kazakhstan because digital finance has become part of a broader transformation in how consumers interact with financial institutions and businesses.

However, the same integration that makes Kazakhstan's payment system convenient also creates questions about consumer behavior and market concentration. When shopping, banking, payments, and other financial activities are available through one platform, the number of steps separating a consumer from a purchase can become very small. Behavioral research from other payment contexts suggests that lower transaction friction and reduced payment salience may affect spending decisions (Hou et al., 2021; Schomburgk et al., 2024). It would be inappropriate to assume that these findings automatically prove that Kaspi causes Kazakh consumers to spend more, since direct evidence establishing that causal relationship remains limited. Still, the psychological mechanisms identified in the broader literature provide a reason to examine the issue. Recent research has also raised questions about Kaspi's increasingly important position

within Kazakhstan's financial landscape and what platform concentration could mean for the country's financial system (Bibi & Yerzhan, 2026). Kazakhstan therefore demonstrates both sides of cashless development: integrated platforms can make economic activity faster and more accessible, but their growing importance also raises questions about consumer decision-making, competition, and dependence on digital financial infrastructure.

4. Digital Payments in the United States

The United States has also experienced a major shift toward digital payments, but its payment environment differs considerably from Kazakhstan's. Instead of relying on one dominant integrated platform, American consumers generally choose among many separate payment methods and financial providers. Credit and debit cards remain central to everyday transactions, while mobile wallets such as Apple Pay and Google Pay allow consumers to connect existing cards to smartphones and other devices. Services such as PayPal and Venmo provide additional options for online purchases and person-to-person transfers, while Zelle connects transfers more directly to participating banks. This creates a relatively fragmented digital payment ecosystem in which banks, card networks, technology companies, and fintech firms compete and cooperate to process consumer transactions. The result is a system that offers considerable choice and convenience, but does not have the same level of platform integration found in Kazakhstan's Kaspi-centered ecosystem.

The behavioral effects of mobile payments are particularly relevant in the United States because digital payment adoption exists alongside an established consumer credit system. Ahn and Nam (2022) examined data from 21,457 participants in the U.S. National Financial Capability Study and found that mobile payment users were more likely to exhibit several forms of overspending behavior. Their findings are important because they suggest that the convenience of mobile payments may have consequences beyond consumers simply choosing one payment technology over another. Financial knowledge moderated the relationship, but it did not completely remove it. This indicates that consumer characteristics remain important even when payment technology makes transactions easier. The result is also consistent with the broader cashless effect identified by Schomburgk et al. (2024), whose meta-analysis found that consumers generally tend to spend more with cashless payment methods than with physical cash. These findings do not establish that every American who adopts mobile payments will increase spending, but they provide evidence that payment technology can influence financial behavior. Another important characteristic of the U.S. market is the relationship between digital payments and consumer credit. Credit cards have long allowed American consumers to separate the moment of consumption from the moment when the full financial cost must be paid. Newer fintech products can extend this separation further. Bian et al. (2023, 2024) examine the expansion of e-wallets and Buy Now Pay Later services and show how digital payment platforms can interact with credit expansion, consumer spending, and indebtedness. This relationship matters because payment convenience and credit availability are conceptually different but can reinforce each other. A digital wallet can reduce the effort required to complete a transaction,

while credit can reduce the immediate financial constraint associated with that transaction. As a result, consumer spending within a cashless environment may be influenced by both the psychological experience of payment and the financial structure behind the payment method. The American case therefore illustrates that the development of a cashless economy involves more than technological innovation. Digital payments can make transactions faster, support online commerce, and give consumers multiple ways to manage and transfer money. At the same time, their interaction with credit and consumer financial behavior creates potential risks. This distinction is important when comparing the United States with Kazakhstan. Kazakhstan's digitalization has developed strongly around an integrated platform ecosystem, while the United States combines digital wallets and payment applications with a mature network of banks, cards, and credit products. Both systems reduce the need for physical cash, but they do so through different institutional structures. These differences may influence how consumers experience spending and how the benefits and risks of cashless transactions appear in each economy.

5. Kazakhstan and the United States: Comparative Analysis

Kazakhstan and the United States share an important similarity: in both countries, payment technology has reduced the amount of effort required to complete many everyday transactions. Consumers can make purchases, pay bills, send money, and access financial services without handling physical currency. This convenience can reduce transaction costs for both consumers and businesses and can support the expansion of digital commerce. From a behavioral perspective, however, greater convenience can also change how consumers experience the act of spending. The broader literature on cashless payments suggests that reducing the physical visibility of money can weaken the psychological connection between purchasing a product and experiencing its financial cost. Schomburgk et al. (2024) found an overall cashless effect across 71 studies and 392 effect sizes, although the magnitude varied across different settings. Hou et al. (2021) similarly connected digital payment adoption with household consumption through mental accounting. These findings provide a common theoretical framework that can be applied when considering both Kazakhstan and the United States.

The most important difference between the two cases is the structure of their digital payment ecosystems. Kazakhstan's system is characterized by a relatively high degree of integration around platforms such as Kaspi.kz. Payments, transfers, banking services, shopping, and other financial activities can occur within the same broader ecosystem. The U.S. system is more fragmented. A consumer might hold a checking account with one bank, use a credit card issued by another institution, make a purchase through Apple Pay, send money through Venmo or Zelle, and shop through an unrelated online marketplace. The difference is economically meaningful because platform structure can influence transaction costs and consumer behavior. An integrated ecosystem can make movement between financial activities particularly seamless, while a fragmented ecosystem gives consumers more choices but requires interaction among several financial institutions and technology providers. The available research does not establish

that one structure necessarily produces higher consumer spending than the other. Instead, the comparison demonstrates that countries can move toward cashlessness through different institutional paths.

There is also an important difference in how payment technology interacts with consumer credit. The United States has a mature credit card market, and newer digital payment products operate alongside widespread access to consumer credit. This can make it difficult to separate the behavioral effect of a digital payment interface from the effect of borrowing. Bian et al. (2023, 2024) show why this distinction matters by examining how e-wallets and Buy Now Pay Later services can connect payment innovation with credit expansion and consumer indebtedness. Kazakhstan's integrated fintech environment also contains financial products beyond simple payments, but its institutional development and platform structure differ from those of the United States. A careful comparison should therefore avoid treating Apple Pay and Kaspi as equivalent technologies. The relevant economic question is how each ecosystem combines payment convenience with banking, commerce, and access to financial products.

The comparison also demonstrates why claims about national spending should be made carefully. It would be tempting to compare the rapid adoption of digital payments with changes in consumption and conclude that cashlessness causes consumers to spend more. However, national consumption is influenced by income, inflation, employment, interest rates, access to credit, demographic characteristics, and many other factors. Existing behavioral research provides stronger support for the narrower argument that payment methods can influence how consumers perceive and make individual spending decisions (Ahn & Nam, 2022; Prelec & Loewenstein, 1998; Schomburgk et al., 2024). Kazakhstan and the United States can therefore be compared most effectively at the level of payment systems and behavioral mechanisms rather than through an unsupported claim that consumers in one country spend more because their economy is more cashless.

6. Economic Benefits and Risks of Cashless Economies

One of the clearest economic benefits of digital payments is greater transactional efficiency. Cash transactions require physical currency to be produced, transported, stored, counted, and exchanged. Digital systems can complete many of these transactions almost instantly and can support purchases that would be difficult to conduct with physical cash, particularly online transactions and long-distance transfers. For consumers, this means greater convenience and easier access to goods and services. For businesses, digital payments can simplify transactions and expand access to customers who increasingly expect electronic payment options. Digital financial platforms may also provide a gateway to additional financial services. Hong et al. (2025), for example, examine how fintech participation can develop from digital payment use toward broader platform-based financial participation. Cashless infrastructure can therefore influence economic activity beyond the payment itself by connecting consumers with a wider financial ecosystem.

Digital payments may also contribute to financial inclusion, although this benefit depends on access to technology and financial infrastructure. Mobile financial services can make formal financial transactions more accessible to consumers who may have difficulty using conventional banking services. Kazakhstan provides a particularly relevant example because its fintech expansion has combined mobile payments with broader financial and commercial services. Research examining digital financial inclusion in Kazakhstan emphasizes the importance of trust and platform-based financial intermediation in shaping participation in electronic payment ecosystems (Imanbayeva et al., 2026). In the United States, the institutional structure is different, but digital payment services similarly allow consumers to transfer money and participate in digital commerce through multiple channels. The potential benefit is therefore similar across the two cases, even though the systems through which consumers receive it differ.

The main consumer-level risk is that greater convenience may reduce attention to spending. Behavioral economics does not suggest that consumers completely stop understanding the value of money when they use a smartphone or payment card. Rather, the argument is that the form and timing of payment can change how strongly consumers experience the cost of a purchase. Prelec and Loewenstein (1998) established an important theoretical foundation for understanding this relationship between payment and consumption. More recent research applies similar ideas to digital environments. Hou et al. (2021) connect digital payments with consumption through mental accounting, while Ahn and Nam (2022) find an association between mobile payment use and overspending. Faraz and Anjum (2025) similarly examine how the reduced visibility of money in digital transactions can contribute to impulse purchasing. The consistency of these findings across different research designs makes the behavioral consequences of cashlessness economically relevant, even though the size of the effect varies across consumers and contexts.

There are also broader risks that differ between payment ecosystems. In Kazakhstan, the importance of an integrated platform can create concerns about market concentration and reliance on a relatively small number of digital intermediaries. Bibi and Yerzhan (2026) discuss Kaspi's prominent position within Kazakhstan's financial landscape in the context of platform capitalism and monetary sovereignty. In the United States, the challenge is somewhat different because payment services are distributed across numerous banks, technology companies, card networks, and fintech providers. Consumer credit adds another dimension, particularly when frictionless digital payments are combined with products that allow purchases to be financed or repayment to be delayed (Bian et al., 2023, 2024). The economic risks of cashlessness therefore cannot be reduced to overspending alone. Competition, financial literacy, consumer protection, access to credit, privacy, and platform structure all influence the consequences of digital payment adoption.

7. Discussion

The comparison between Kazakhstan and the United States demonstrates that a "cashless economy" is not a single economic model. Both countries have experienced substantial payment

digitalization, but the institutional structures behind that transformation differ. Kazakhstan illustrates how a highly integrated fintech ecosystem can combine payments, commerce, transfers, and financial services within a relatively seamless consumer experience. The United States illustrates a more fragmented model in which digital wallets and transfer applications operate alongside established banks, credit cards, card networks, and other financial institutions. These differences are important because payment technology does not influence consumers in isolation. Its effects depend on the broader system in which transactions occur. Platform design, access to credit, financial knowledge, competition, and consumer habits can all influence the relationship between cashless payments and spending.

The evidence reviewed in this paper nevertheless suggests that there is a meaningful behavioral dimension to payment digitalization. Prelec and Loewenstein (1998) provide a theoretical explanation for why the connection between payment and consumption can influence decision-making. Hou et al. (2021) connect digital payment adoption with consumption through mental accounting, while Ahn and Nam (2022) find that mobile payment users in the United States are more likely to exhibit overspending behaviors. The larger meta-analysis by Schomburgk et al. (2024) strengthens the argument by finding an overall cashless effect across a broad body of prior research. At the same time, the meta-analysis indicates that the effect varies across circumstances. This is important because it prevents the evidence from being interpreted too broadly. The literature supports the conclusion that payment methods can influence spending behavior, but it does not support the claim that every consumer will spend more whenever cash is replaced by a digital alternative.

Causality is particularly important for the Kazakhstan and United States comparison. Consumers who choose digital payment technologies may differ from consumers who prefer cash in age, income, technological familiarity, financial knowledge, shopping habits, and access to credit. At the national level, consumer spending is also influenced by broader macroeconomic conditions that have little to do with payment technology. For these reasons, this comparative review cannot establish that the expansion of Kaspi in Kazakhstan or mobile wallets in the United States directly caused an increase in national consumer expenditure. Instead, the strongest conclusion is that cashless technologies alter the environment in which spending decisions are made. They reduce certain transaction costs and can make payments less physically salient, creating conditions that may encourage consumption for some consumers.

This limitation also reveals an important gap in the existing literature. There is a substantial body of research on cashless payment behavior generally, but considerably less peer-reviewed research directly comparing consumer spending effects in Kazakhstan and the United States using equivalent measures. Much of the Kazakhstan-specific literature focuses on fintech development, digital financial inclusion, or the structure of the country's platform economy, while stronger consumer-level evidence on overspending is available from the United States and other markets. Future research could address this gap by comparing similar groups of consumers in both countries or by examining spending behavior before and after adoption of particular payment technologies. For the purposes of the present review, however, the lack of

identical data should not be hidden. Instead, it reinforces the paper's central argument that similar payment technologies can operate within very different economic and institutional environments.

8. Conclusion

The transition from cash to digital payments represents more than a change in how consumers complete transactions. It changes the infrastructure connecting consumers, businesses, financial institutions, and digital platforms, while potentially changing the psychological experience of spending. Kazakhstan and the United States illustrate two different paths toward a more cashless economy. Kazakhstan's development has been strongly influenced by integrated fintech platforms such as Kaspi.kz, while the United States has developed a more fragmented ecosystem involving banks, card networks, digital wallets, peer-to-peer payment services, and consumer credit. Despite these differences, both systems share a common objective of making financial transactions faster and more convenient.

Existing research suggests that this convenience creates both economic benefits and behavioral risks. Digital payments can lower transaction friction, facilitate electronic commerce, improve access to financial services, and connect consumers to broader financial ecosystems. At the same time, research on mental accounting, payment salience, mobile payments, and the cashless effect suggests that less visible forms of payment can make spending feel less immediate and may contribute to higher consumption or overspending in some circumstances (Ahn & Nam, 2022; Hou et al., 2021; Prelec & Loewenstein, 1998; Schomburgk et al., 2024). The comparison therefore does not suggest that cashless economies are inherently beneficial or harmful. Instead, the economic consequences of digital payments depend on how payment systems are designed, how consumers use them, and how factors such as financial knowledge, credit availability, competition, and regulation interact with technological convenience. Kazakhstan and the United States provide useful contrasting cases for understanding how the continuing transition toward cashless economies may shape both consumer behavior and the broader economic environment.

Works Cited

Ahn, S. Y., & Nam, Y. (2022). Does mobile payment use lead to overspending? The moderating role of financial knowledge.

Almabekova, Z., & Jaxybekova, G. (2025). Comparative analysis of the development of financial technologies in Kazakhstan, China, the United Kingdom, and Singapore. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5253586

Bian, W., Cong, L. W., & Ji, Y. (2023). The rise of e-wallets and buy-now-pay-later: Payment competition, credit expansion, and consumer behavior (NBER Working Paper No. 31202). National Bureau of Economic Research. https://www.nber.org/papers/w31202

Bian, W., Cong, L. W., & Ji, Y. (2024). The rise of e-wallet super-apps and buy-now-pay-later (NBER Working Paper No. 33178). National Bureau of Economic Research. https://www.nber.org/papers/w33178

Bibi, S., & Yerzhan, I. (2026). Monetary sovereignty and dual extraction: Platform capitalism, CBDC, and the new financial extractivism in Kazakhstan. Journal of Post Keynesian Economics. https://doi.org/10.1080/01603477.2026.2709396

Faraz, N., & Anjum, N. (2025). Spendception: The psychological impact of digital payments on consumer purchase behavior and impulse buying. Behavioral Sciences, 15(3), 387. https://doi.org/10.3390/bs15030387

Hong, C. Y., Lu, X., & Pan, J. (2025). Financial inclusion via FinTech: From digital payments to platform investments. Management Science. https://doi.org/10.1287/mnsc.2024.04423

Hou, T., Hou, K., Wang, X., & Luo, X. (2021). Why do I give money to unknown people? An investigation of online donation and forwarding intention. Electronic Commerce Research and Applications, 47, 101055. https://doi.org/10.1016/j.elerap.2021.101055

Prelec, D., & Loewenstein, G. (1998). The red and the black: Mental accounting of savings and debt. Marketing Science, 17(1), 4–28. https://doi.org/10.1287/mksc.17.1.4

Schomburgk, L., Belli, A., & Hoffmann, A. O. I. (2024). Less cash, more splash? A meta-analysis on the cashless effect. Journal of Retailing, 100(3), 382–400.

Who Really Wins from The Name, Image, and Likeness (NIL)? Athletes' Gains and University Finances By Aristote Gideon Ophir

Abstract

On July 1st, 2021, the NCAA allowed college athletes to be paid for their name, image and likeness (NIL). This paper asks whether that change created new economic value or simply moved value that already belonged to universities into the hands of the athletes.

I compare Notre Dame and The University of Texas at Austin (UT Austin) from 2016 to 2025, using endowment size as a proxy for institutional financial health, total enrollment as a proxy for student demand, and average seasonal score and win rate for on-field performance. Additionally, publicly available social media accounts from key football athletes from Ohio State, UT Austin, and Notre Dame track social media engagement and brand deals on either side of the policy change. I leverage ordinary least squares regressions to estimate the NIL effect on endowment and enrollment separately for each school.

Athletes gained enormously: engagement and endorsement income rose sharply even for players whose on-field production was unchanged. On the university side, endowments were unchanged, while win rates rose at both schools. This gives suggestive evidence that there was a positive-sum change due to NIL rather than merely a transfer away from universities.

1.0 Introduction

Historically, it has been assumed that the financial health of a major American university and the success of its football program have been treated as deeply interdependent. A winning season was never just a sporting result; it spilled into the institution as a whole, driving up student applications, boosting alumni donations, and increasing merchandise sales, ticket revenue, and profitable TV deals. This interdependence is so well known that it even has a specific name, the "Flutie effect" (Garthwaite et al.). After a spike in applications, Boston College reportedly saw a single famous play in 1984. Using this information, they linked it to an athlete's value effectively belonging to the school. The performance generated revenue, prestige and international visibility for the school. But is this still the case in today's world? July 1st, 2021 changed all these arrangements.

On July 1st 2021, the National Collegiate Athletic Association (NCAA) passed a law allowing college athletes to be paid by brands through sponsorships. For the first time ever, a college athlete could wake up and legally get paid for who they are.

College athletes are now allowed to be paid for their Name, Image, and Likeness (NIL); this meant athletes could receive endorsements, social media deals, and appearances. An athlete can now make millions of dollars as young as 17 years old just for posting online or for wearing something as simple as a specific T-shirt. Within one year, the market reached just under $1 billion in value. For so long, the common idea that college athletes should not be paid was the norm; why should they get paid before making it professionally? It effectively takes away the

incentive to work harder to make it to the “Big Leagues.” This change in ideology dramatically altered the amateur Division I college-level football system.

Before the NIL, athletes like Justin Fields, for example, would have amazing games with incredible statistics. Take the Sugar Bowl in January 2021: 385 yards and 6 touchdowns, in one game. His social media has only around 4 thousand likes, no brands, normal game attire. Nothing flashy. Playing amazingly because the incentive is to make it to the NFL. One year later, a similar position star player. CJ Stroud, in the Rose Bowl in January of 2022, 573 yards, 6 touchdowns, the Ohio State all-time record. At this moment in his life, he has a $2.5 million NIL valuation, a Mercedes deal where he can have any car he wants, and millions of likes on his social media. Where is the incentive to make it professionally if, in this position, he already has a more successful career than some NFL athletes economically?

The NIL shifted financial power toward athletes, but universities are still central; football programs are massive revenue generators: tickets, merchandise, TV deals, and donations. The question begs the idea: does having NIL-famous athletes help or hurt the university's bottom line? Does a Heisman-contender with millions of Instagram followers bring more donors to a powerhouse like Notre Dame, or does it distract from team cohesion and results? These questions matter because college football isn't just a game to a university - it's a revenue engine that feeds admissions, donations, and the school’s national brand. So the answer to “who benefits from NIL” isn't rhetorical; it decides whether the entire money model behind college still works the same way every economic party assumes it does.

The NIL sounds like a win for every party in this situation on the surface. But if a university used to benefit from an athlete's image for free and now that athlete monetizes themselves - who really wins? In this article, I will show that NIL primarily created new economic value for individual players rather than draining it from universities. Making it closer to a value-creating situation than a trade-off.

This paper hypothesizes that the introduction of NIL primarily created new economic value for individual athletes rather than redistributing existing value away from universities. It predicts that NIL produced large and immediate financial gains for athletes, while leaving universities’ financial health and on-field performance largely unchanged - making NIL closer to a positive-sum transfer than a zero- or negative-sum transfer.

1.1 NIL and the University: Two Sides Of the Same Coin.

Notre Dame, UT Austin, and Ohio State are all consistently elite football programs; they all have large endowments, and all produced athletes with clear pre- and post-NIL social media trails that are trackable. They represent different regions and different conference structures. The time period examined will be from 2016-2025, 2016-2020 is pre-NIL, 2021-2025 is post-NIL. July 2021 is the dividing window. All these universities have been affected by the NIL.

The NIL creates winners and losers, and the outcome depends on the athletes' marketability, their on-field performance, and how financially equipped their university already was. For athletes, the NIL has created a clear financial and media win, especially for high

performers. For universities, the picture is far more complicated- endowment trends don't follow a clean post-NIL pattern, and team performance doesn't dramatically shift either.

1.2 Literature Review

This paper contributes to a growing economic literature on the financial structure of college athletics and the compensation of college athletes. One prominent branch reviews the economic rents generated under amateurism–the long-standing rule barring athletes from sharing in the revenue they produce. Garthwaite et al. (2020) show that the bulk of athletic revenue is generated by football and basketball in universities, and at the same time, those athletes don't make that much of the money; the surplus is often redirected to coaching salaries, facilities, and non-revenue sports. They further find that athletic revenue is largely branched out from the university's broader finances (Garthwaite et al.). This arrangement was challenged in the Supreme Court in the National Collegiate Athletic Association V Alston (2021), in which the Supreme Court unanimously struck down NCAA limits on education-related benefits, and where one of the justices went further, openly doubting whether the NCAA's whole system of not paying athletes was even legal.

A second branch asks whether athletic success translates into measurable gains for the institutions themselves. Chung (2013) models the "advertising" value of winning football, showing that on field success raises the quantity and quality of students' applications - the effect popularly known as the "Flutie effect" . Anderson (2017) specifically shows this using a propensity score design built on bookmaker spreads to show that football success raises athletic donations, applications, academic, reputation, and incoming student quality. A related branch debates whether athletes within this system are harmed by it: Heckman and Loughlin (2021) use longitudinal panel data to argue that student athletes graduate at rates equal to or higher than comparable non-athletes and earn similar or better early career wages, concluding that participation promotes social movement rather than complete exploitation (Heckman and Loughlin).

Across these branches, the existing literature is very often concentrated on the Pre-NIL era, frames athlete compensation in largely zero-sum terms, and tends to examine single outcomes in isolation; observational work on NIL continues to be scarce, since the policy took effect in July 2021. This paper complements that literature by testing the positive sum hypothesis across Notre Dame, the University of Texas at Austin, and Ohio State over the time period from 2016 to 2025, asking whether NIL created new economic value for athletes without draining university endowments or on-field performance–connecting the rent sharing, university–outcomes, and athlete outcomes that work that has been done earlier has dealt with separately.

2.0 Data and Methodology

This section aims to explain where the main 4 data sources came from and the evidence they provide. Social media, endowments, NIL deals, on-field performance.

2.1 Notre Dame- Endowment & Football performance DATA

A university's endowment is the total value of a university's long-term investment fund–donations invested over time to generate returns.

A university's endowment is the best available proxy for university financial health over time. It reflects donor confidence, athletic success in this context, and institutional prestige.

Endowment figures are taken from each university's investment office: Notre Dame's from the Notre Dame Investment Office (investment.nd.edu) and UT Austin's from UTIMCO, the University of Texas/Texas A&M Investment Management Company (utimco.org/reports/general-endowment-fund).

It is measured in Billions of USD. Endowments are affected by stock market conditions, big one-time donations, and investment strategy–not just athletic success. This could skew analysis based on these results, which presents a limitation.

Year	Endowment	Average seasonal score	
2016	10.41	28.54	
2017	11.79	32.62	
2018	13.1	31.15	
2019	13.78	34.23	
2020	13.31	34.25	
2021	20.3	32.54	
2022	16.73	28.31	
2023	20.25	36.08	
2024	22.1	36.85	
2025	20.1	38.77	

Figure 2.1.1.0 (Notre Dame Endowment VS Average Seasonal Score)

Figure 2.1.1.0 (Notre Dame Endowment VS Average Seasonal Score) Explanation

Figure 2.1.1.0 reports Notre Dame's endowment alongside its average seasonal score for every year from 2016 to 2025. The two columns are placed side by side so that the endowment trajectory can be read against on-field output for the same season, with the 2021 row marking the boundary between the pre- and post-NIL periods.

Pre- NIL (2016-2020): endowment grows steadily from $10.41 billion to $13.31 billion for Notre Dame, and scores stay consistent in the 28-34 point range- Notre Dame is a stable, well-funded program before NIL even exists.

2021: Endowment jumps dramatically to $20.3 billion; however, average scores actually slightly dip to 32.54, the financial boom did not translate to better performance. In 2022, endowment drops to $16.73 billion, and scores drop to 28.31- the worst season in the entire dataset, happening within one year into NIL.

From 2023-2025, both endowment and scores recover and grow together, average scores peaking at 38.77 in 2025, which is Notre Dame's best average. Notre Dame shows no clean positive relationship between NIL and either metric in the short term, but in the longer term both

metrics improve, suggesting NIL's benefit for universities may be delayed rather than immediate. The reasoning for no positive effect in the short term may be due to adjustment problems and possibly Covid-19.

Overall, these trends are positive, or at least not negative, for the universities specifically. Notre Dame's endowment grew across the full period despite the 2022 dip, so at no point does NIL appear to weaken the school's finances; this supports the prediction that NIL did not hurt university finances–their endowments kept growing after NIL, just like they did before. The data does not say much about the players themselves, since endowment and scores measure the institution rather than the athletes; the question is answered by the social media data in section 2.3. As for the future: because both metrics recovered and rose together after 2022, the early downturn looks like a contemporary adjustment and did not last for a long time. This suggests universities like Notre Dame stay financially healthy through the NIL era. This prediction is tentative because endowments depend mostly on markets and donations rather than on athletics.

2.2 UT Austin - Endowment & Football performance DATA

Year	endowment size	average seasonal score	
2016	8.79	31.92	
2017	9.65	29.17	
2018	10.68	31.67	
2019	13	35.00	
2020	15.7	41.33	
2021	22.2	35.25	
2022	21.9	35.67	
2023	23.36	35.08	
2024	25.99	34.83	
2025	28.4	29.58	

Figure 2.2.1.1 (UT Austin Endowment VS Average Seasonal Score)

Figure 2.2.1.1 presents the same two series for UT Austin over the identical 2016–2025 window, allowing a direct comparison with Notre Dame in Figure 2.1.1.0.

Pre-NIL (2016-2020) Endowment grows from $8.79 billion to $15.7 billion, and scores range from 19-41; importantly, UT's highest-scoring season in the entire dataset is in 2020, which is Pre-NIL, this complicates the narrative that NIL improves performance.

Post-NIL endowment keeps growing every single year, reaching $28.4 billion by 2025 - this is the strongest and most consistent endowment growth of both universities.

Endowment grows strongly and consistently post-NIL while football performance actually gets worse. These two lines move in completely opposite directions after 2021, which directly challenges the assumption that NIL boosts both university finances and on-field performance simultaneously.

The fact that UT Austin's endowment grew while their scores declined does not necessarily mean NIL hurt their performance. Endowments are driven by investments and donor

money, not by wins and losses on the field–these are two separate things that don't always move together. Additionally, UT Austin's move to the SEC in 2024 meant they were suddenly playing against much tougher opponents, which naturally brought their scoring averages down. So the money kept growing while the scores dropped simply because the two things are measuring completely different aspects of the university's success.

Overall, this is financially good for the university: UT's endowment grew the strongest of any school with little to no NIL era dip, even as scores fell–which only reinforces that money and performance are unrelated. The player side is covered in section 2.3. Looking ahead, the endowment is meant to keep climbing, while the scoring drop is extremely likely to represent the shift towards harder opponents in the move towards the SEC rather than NIL.

2.3 Instagram and Social Media Analytics

Instagram is the best source for overall athlete behavior and statistics pre- and post-NIL. All Posts are public and provide the most information.

NIL deals were found using news articles announcing players' new record-breaking deals and Instagram pages of the players announcing their achievements by signing with companies.

Table 2.3.1

Athlete	University	Period	Game Performance	Likes	NIL Deals
Justin Fields	Ohio State	Pre-NIL	385 yds, 6 TDs	~4k	None
CJ Stroud	Ohio State	Post-NIL	573 yds, 6 TDs	Millions	Mercedes, Express, etc.
Bijan Robinson	UT Austin	Pre-NIL	183 yds, 4 TDs	~15k	None
Bijan Robinson	UT Austin	Post-NIL	179 yds, 2 TDs	~80k	Raising Cane's, Lamborghini
Ian Book	Notre Dame	Pre-NIL	224 yds, 1 TD	~20k	None
Michael Mayer	Notre Dame	Post-NIL	11 catches, 2 TDs	N/A	Levi's, Rhoback

Figure 2.3.1.2 summarises six athlete-seasons drawn from the three schools, pairing each player's headline game performance with their Instagram engagement and their brand partnerships. The table is deliberately structured around matched pairs — the same athlete either side of July 2021, or two comparable athletes at the same school in different periods — so that performance can be held roughly constant while the NIL variable changes.

Bijan Robinson

One year apart, Bijan has nearly identical on-field performances in both games - yet his Instagram grew from 15 k all the way to 80 k on average, an over 5x increase. He went from zero

brand deals to Lamborghini, Raising Cane's, and his own mustard company. Same player, same production, completely different financial reality. This proves NIL created opportunity independent of performance.

Justin Fields VS CJ Stroud

Both players had record-breaking performances in Ohio State bowl games; their statistics are unimaginable. But Justin Fields, 4 thousand likes on average, lower than the Ohio State fanbase, CJ Stroud, millions of likes, $2.5 million NIL valuation, and a Mercedes deal. The only meaningful difference between them was NIL existing: same university, same position, same level of performance, yet an entirely different financial outcome, demonstrating that NIL alone was sufficient to transform a college athlete's economic reality.

Ian Book vs Michael Mayer

Both athletes represented Notre Dame in high-stakes games, yet their relationship with personal branding could not have been more different. Ian Book, posting exclusively Notre Dame gear and journey content, accumulated around 20k likes with zero commercial identity beyond the university itself. Michael Mayer, by contrast, carried Levi's and Rhoback activewear partnerships alongside his record-breaking 11-catch, 2-touchdown performance. It is worth noting that Book and Mayer play different positions, making a direct statistical comparison unfair — however, the contrast in their public identities tells its own story. Pre-NIL athletes built their brand around the university. Post-NIL athletes built their own.

Taken together, these cases convey the idea that NIL has imported the NFL's star-driven model into college American football. In professional sports, individual stars are effectively the product that is sold by the administration. Fans buy tickets and jerseys to support their favorite player and increase their fame. The NIL creates this within college football; a player like CJ Stroud, with millions of followers and a Mercedes deal, creates a fanbase just for himself. Bringing in fans and money that don't necessarily support the team he's in but just the player. Adding this "stardom" to the NIL.

In my view, though, this is not a healthy development for the players themselves. College athletes are still very young, often only seventeen or eighteen. This occurs at a developmental stage where the main focus should be football, not money or the distractions outside of the sport. It is too early in their life, and too low a level of the sport, to be pointing the athletes' focus on something else. The incentive that benefits them financially may pull their attention away from the very development a successful jump to the NFL requires.

2.4 Enrollment Data

Total enrollment is the second institutional outcome in this paper. Where endowment measures accumulated wealth, enrollment measures stranding with prospective students. This matters because the demand for students is the main road that athletic success is supposed to be beneficial for an institution. - Chung (2013) treats winning football as advertising that raises

applications, and Anderson (2017) finds the same for applications and incoming student quality. If NIL moved attention from the school's brand to the athletes, that mechanism should weaken, and enrollment is where it would show.

Enrollment comes from the National Center for Education Statistics (NCES) through the college navigators database - Notre Dame and UT both report total enrollment for the fall of the stated year, matching each figure to that football season.

Notre Dame moved from 12,393 students in 2016 to 13,042 in 2024; UT Austin from 51,331 to 53,864. Both series are complete through 2024; the 2025 figures had not been released at the same time of writing, which is why the regressions run 2016-2024 rather than the full 2016-2025 window used in the earlier figures.

2.5 Data limitations

Enrollment reflects a supply decision, not demand. The most serious limitation of the enrollment variable is that it does not measure how many students want to attend a university — only how many that university decided to admit. Notre Dame keeps enrollment low by choice rather than necessity. Selectivity is part of what the school sells, so a small class is a deliberate policy, and its enrollment accordingly moves within a band of about six percent across the entire period. UT Austin faces the same constraint through state funding and physical capacity.

This has a direct consequence for how the suggestive evidence should be read. If NIL did raise student interest in these schools, the extra demand would be absorbed into selectivity — more applications competing for the same fixed number of places — and would never appear in the enrollment figures at all. The variable is capped from above. That makes it uninformative in both directions: a flat enrollment series cannot be treated as evidence that NIL failed to attract students, and a rising one would more likely reflect a decision to expand than a response to demand.

Endowment is dominated by markets. Endowment is accumulated invested wealth, not annual performance. Its yearly movement comes mostly from asset prices and occasional large gifts. Athletic revenue is small relative to the whole institution and reaches the endowment only indirectly.

Tuition is not comparable across the two schools. Notre Dame reports one private price that rises every year. UT Austin's figure is in-state tuition, set by the Board of Regents and effectively frozen after 2022. These are different objects — one a market price, one administered — and neither is a net price after aid.

Average seasonal score is crude. Using points scored ignores opponent quality and does not separate a win from a loss. UT's 2024 move to the SEC changed opposition strength mid-sample, and the 2020 seasons include postponed and cancelled games, so those averages rest on fewer fixtures.

There is no control group. The rule changed for every Division I school on the same day, so no untreated comparison exists. The post-2021 shift necessarily absorbs everything else in those years — pandemic recovery, conference realignment, market cycles.

*The athlete cases are selected on visibility***.** The NIL deals in Section 2.3 were found through news coverage and players' own announcements, both of which favour the biggest deals. They are not a representative sample, and like counts also rose platform-wide over the period.

3.0 Data analysis

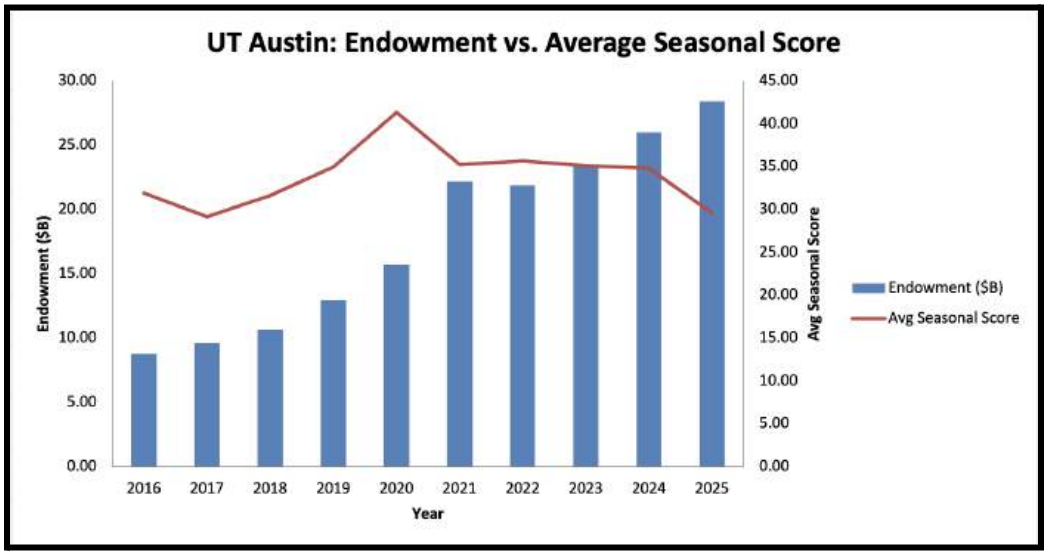


Figure 3.1 Graph Presenting Endowment VS Average Seasonal Score for UT Austin

Figure 3.1 plots UT Austin's endowment against its average seasonal score across the 2016-2025 seasons. The contrast between the two series is immediate: the endowment bars rise in every single year, showing an extremely steady climb. While the scoring line peaks in 2020 - before NIL - and drifts downwards from there. After the July 2021 dividing line, the two measures move in visibly opposite directions. Endowment accelerates just as on-field scoring falls off. This divergence is the clearest single piece of evidence that on-field performance and institutional wealth are largely unrelated: if the two were linked, the lines would track each other, and here they do the reverse. The more likely explanation is that endowment is driven by investment returns and donor giving rather than wins and losses, so the two simply measure different things. The reason this pattern matters for the hypothesis is that it is the opposite of what a zero-sum account would predict. If NIL had transferred value away from the institution — athletes capturing endorsement income that would previously have flowed to the school through merchandise, ticketing, and donor enthusiasm — the visible signature would be a break in the endowment trajectory at or shortly after July 2021: slower growth, a flattening, or a reversal. UT Austin shows none of these. The endowment climbs in every post-NIL year and does so faster than it did before. Whatever athletes gained, it did not come out of this particular

pocket. The scoring decline should not be read as NIL damaging performance: UT Austin's 2024 move to the SEC placed the team against much harder opponents, which pulls scoring averages down regardless of any NIL effect. The endowment grew more strongly here than anywhere else in the dataset, with no visible NIL era dip, which supports the prediction that universities' financial health continued on its pre-NIL trajectory with no decline attributable to NIL. Looking ahead, the endowment trend suggests continued growth, while the scoring drop most plausibly reflects the SEC transition rather than anything to do with NIL, though this forecast stays tentative given that endowment track markets and donations far more than athletics.

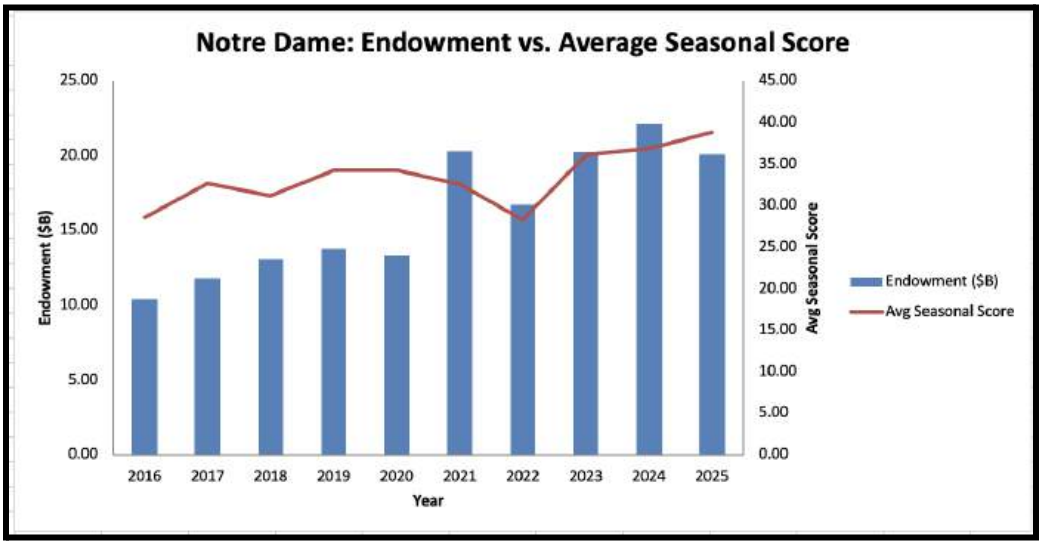


Figure 3.2 Graph presenting Endowment VS Average Seasonal Score for Notre Dame

Figure 3.2 plots Notre Dame's endowment against its average seasonal score across 2016-2025. Across the pre-NIL year, the two series sit in a stable band - the endowment bars step up modestly from 2016 to 2020 while the scoring line holds in a consistent range, the profile of a well-funded, established program. The chart then shows the two measures briefly coming apart around the policy implementation. The endowment bar jumps sharply in 2021 while the scoring line actually dips; in 2022, both fall together to the lowest scoring in the dataset. From 2023 onward, though, the bars and the line climb in step, with scoring reaching its dataset high at the right edge of the chart. Read as a whole, this graph shows no clean short-term link between NIL and another measure, but a longer-term pattern in which both recover and rise together - suggesting any NIL benefit is delayed rather than immediate, with the 2021-2022 dip more plausibly explained by the pandemic disruption than NIL itself. Crucially, at no point do the bars trend downward, so nothing suggests NIL weakened the financial health of schools. Supporting the prediction that schools remained financially healthy post-NIL. The player side question is left to the social media section, since endowment and scoring measure the institution rather than the athletes. Looking ahead, the recovery after 2022 of both suggests Notre Dame stays financially

healthy during the NIL era, though the forecast stays tentative given that endowments depend far more on markets and donations than on athletics.

3.3 Suggestive evidence of NIL's influence on top athletic schools

Points scored is a limited measure of performance because it depends on the teams that you end up playing against. The win rate is more direct. Notre Dame won 74.6% of its games pre-NIL and 81.7% of its games post-NIL. UT Austin won 57.9% before and 73.3% after - an improvement of over fifteen percentage points from before to after. It improved this much even as the scoring average fell, which confirms the scoring drop came from the harder SEC games it was playing again after its move.

Draft output tells the same story. Notre Dame sent 25 players to the NFL in the pre-NIL window and 27 after, essentially flat. UT Austin went from 11 to 33. Both schools remained strong pipelines to the pro level, and Austin became a far better one. This matters because the worry raised in the introduction of the NIL was that athletes would just stop trying at a college level because they are already making almost as much as the pros. Two cautions apply. Draft classes lag the seasons that produced them, so genuinely post-NIL recruits only begin to appear around 2024, and Steve Sarkisian took over Texas in 2021, the same year the turtle changed, making his recruiting overhaul at least as plausible an explanation as the NIL. Since Notre Dame doesn't have as big a jump, Texas' increase may be explained by the difference in administration behind it.

All of this should be read carefully, because these are three of the most advantaged programs in the country. NIL value comes from attention, and attention was already very concentrated on these teams. A policy letting players monetize their appearance will obviously work more in universities, showing off the players' appearance for free, and this costs the schools nothing. The positive sum relationship will always appear here. Smaller programs are a different story. A school without an audience cannot generate real endorsement value for its athletes, while its best players now have more reason to transfer somewhere that can pay them. NIL may create value at the top and drain talent from the bottom. This paper's conclusion applies to elite programs and should not be stretched to college football as a whole. Figures like 3.1 and 3.2 cannot be reproduced for those schools anyway, since their endowments are far smaller and their athletic departments usually lose money — athletic revenue would be the right variable there, which is left to future work.

4.0 Empirical Approach

The figures in Section 3 provide evidence on how university finances and on-field performance moved around the July 2021 policy change. What they show is not a clean correlation but a divergence: endowments continued their upward trajectory at both schools while average seasonal scores followed no consistent post-NIL pattern. These graphical patterns are only descriptive, and the visible year-to-year movements may reflect market conditions, conference realignment, or the pandemic rather than NIL itself. To assess the relationship more

systematically, this section estimates a regression that isolates the NIL-era shift while holding constant the other factors that plausibly move the outcome over time. Because the trend lines differ across the two schools, the model is estimated separately for each, allowing for heterogeneous effects rather than imposing a single common response.

To test whether the introduction of NIL affected the financial health of different universities, this paper estimates an ordinary least squares (OLS) regression separately for each school, using annual data from 2016 to 2024. The model takes endowment and enrollment as the outcome variables and isolates the NIL effect while holding constant the other factors that plausibly move an endowment over time. The specification is:

$$\text{Endowment}_t = \beta_0 + \beta_1 \cdot \text{NIL}_t + \beta_2 \cdot \text{Tuition}_t + \beta_3 \cdot \text{Year}_t + \varepsilon_t$$

$$\text{Enrollment}_t = \beta_0 + \beta_1 \cdot \text{NIL}_t + \beta_2 \cdot \text{Tuition}_t + \beta_3 \cdot \text{Year}_t + \varepsilon_t$$

- *Endowment*$_t$ — the outcome (dependent) variable: the university's endowment in billions of USD in year *t*. This is the proxy for institutional financial health.

- *Enrollment*$_t$ — the outcome (dependent) variable: total enrollment in year *t* to account for institutional size. This is the proxy for salience of the institution to prospective students.

- β_0 — the intercept: the baseline endowment level when the other variables are held at zero. It anchors the regression rather than carrying a direct interpretation.

- *NIL*$_t$ — the key explanatory variable: a dummy equal to 0 in the pre-NIL years (2016–2020) and 1 in the post-NIL years (2021–2024), switching on at the July 2021 policy change.

- β_1 — the coefficient of interest: the shift in endowment associated with the NIL era, *holding tuition, enrollment, and the year trend constant*. A coefficient that is zero or positive supports the hypothesis that NIL did not reduce university financial health; a large negative coefficient would indicate the opposite.

- *Tuition*$_t$ / β_2 — tuition and fees in year *t*, included as a control. β_2 captures how much endowment moves with tuition, absorbing the effect of the school's pricing and revenue base.

- *Year*$_t$ / β_3 — a linear time trend. β_3 captures the average yearly change in endowment that occurs independently of NIL — critically, this separates NIL's effect from the general

upward drift of endowments over time, so that β_1 does not simply pick up the fact that later years happen to have larger endowments and/or enrollment.

- ε□ — the error term: everything else affecting endowment and enrollment that the model does not capture, such as stock-market conditions, one-time major donations, and investment strategy.

5.0 Results

Table 5.1 - Endowment Regressions

	Notre Dame	**UT Austin**
NIL	7.90**	4.51**
	(2.62)	(1.10)
Tuition	0.0041	0.0018
	(0.0021)	(0.0012)
Year	−7.62	1.22*
	(4.20)	(0.33)
Constant	15160.27	−2461.23
	(8359.05)	(646.75)
R^2	0.951	0.991
Adjusted R^2	0.922	0.985
Observations	9	9

NIL is positive and statistically significant at the 5% level for both schools: 7.90 at Notre Dame (p=0.030) and 4.51 at UT Austin (p=0.009), with R^2 of 0.951 and 0.991.

But these numbers cannot be taken at face value because of a lack of statistical power. I regressed the effect of NIL on endowment size, but I only have 9 years of Data points for each school. Furthermore, given that the R^2 for both schools is close to a value of 1, this means that my regression line almost perfectly fits through those 9 data points. Realistically, when accounting for other variables that can impact endowment size that I cannot control for, the R^2 is likely to decrease. Thus, these regression results should be interpreted with caution because they suffer from omitted variable bias. What these first pass-through coefficients do suggest is that the

impact of the NIL goes counter to my hypothesis. The NIL positively impacts the financial health of a university, rather than reducing it.

Table 5.2 - Enrollment Regressions

	Notre Dame	**UT Austin**
NIL	116.06	1279.15*
	(235.94)	(327.77)
Tuition	−0.124	−2.506**
	(0.185)	(0.369)
Year	315.26	613.76**
	(377.65)	(97.27)
Constant	−616950.42	−1160448.04
	(752225.85)	(193295.78)
R^2	0.922	0.963
Adjusted R^2	0.876	0.941
Observations	9	9

Figure 5.2 reports both enrollment regressions. The two schools give different answers. Notre Dame's NIL figure — about 116 students — is smaller than its own margin of error, so it is no different from zero. UT Austin's figure of roughly 1,279 students is significant, but that significance is misleading.

Notre Dame's flat result is what section 2.5 predicted: the school sets its own class size, so enrollment cannot rise just because demand does. UT Austin's problem is different. In the same model, adding one dollar to tuition appears to cost the school more than two students, which cannot be true. The year comes out just as strong. UT Austin's enrollment climbs steadily every year, and the model hands that climb to whichever variable happens to fit it best.

6.0 Discussion

Endowment Outcomes

The endowment results support the hypothesis, but only in its weaker form. The prediction was never that endowments would stay flat — both schools' endowments grew substantially over the period. It was that NIL would not break the trajectory they were already on, and the data is consistent with that: no NIL coefficient is negative at either school, and neither series shows a slowdown or reversal after July 2021. What the results cannot show is that NIL caused endowments to grow, since the coefficients are far too large to be believable and the models cannot separate NIL from the passage of time.

Endowment is driven mainly by markets, so it is a blunt instrument — well suited to detecting large financial shocks and poorly suited to detecting small ones. This only works in the paper's favour if a link between athletics and endowment exists, and it does. Humphreys and Mondello (2007) find endowment donations rise with a program's winning percentage, and Meer and Rosen (2009) find a conference championship raises alumni general-purpose giving by about seven percent. If NIL had in fact taken away value from universities, the break would have been clearly seen. Here it is not.

Enrollment Outcomes

The enrollment models are the weakest aspect of the paper, and the reason is structural rather than statistical. Neither NIL nor student demand is the main driver here: class size is set by internal policy, decided by university administrations on the basis of capacity, funding, and strategic priorities. Notre Dame's null and UT Austin's positive coefficient are therefore equally uninformative — one because the number is held fixed by policy, the other because the model assigned a steady administrative growth trend to the NIL dummy.

This means the student demand channel identified by Chung (2013) and Anderson (2017) has not really been tested. Whether NIL weakened the link between athletic success and the appeal for the universities remains very open. The only way to answer this could be through application counts and percentages of college tour applications. Universities have a fixed amount of people to accept, no matter how many applications, so enrollment will not change. This means it cannot demonstrate the Flutie effect. If application rates are looked at, they could demonstrate a percentage increase in the attraction a university has for new students.

Overall (Larger Picture)

Taken together, the three strands point the same way. Athletes gained enormously, as section 2.3 shows. Universities did not lose financially. On-field performance and NFL output improved rather than declined. For these schools, NIL looked like a positive-sum change: value that was previously going uncollected is now collected by student-athletes, instead of value being taken from other institutions being taken and handed over.

This reframes the Flutie effect instead of overturning it. Athletic success still generates institutional benefits; NIL simply added a second layer on top, in which the athlete captures the commercial value of their own image while the school keeps everything it had before. The money was being left on the table, not traded. The simple definition is that pre-NIL there was

money going to waste every year being left on the table unanswered and unused, now that money is traded towards athletes for their benefit, highlighting the positive sum trade of the implication of NIL. There are several limitations that have to be stated directly. First, this holds for elite programs and not for the rest of college football, for the reasons set out in section 3.3. Second, the NIL variable only marks when the policy came in, not how much money actually moved. It is a simple on/off switch. So Notre Dame and UT Austin are treated equally even though the NIL market in different schools is nowhere near the same size; this can skew results and also should be kept in mind when evaluating and scrutinizing data.

A key limitation is that this research paper only relied on data for only 2 elite universities; therefore, it is unclear if the external validity will hold when this research paper is applied to other schools. To my knowledge, I only leveraged less than 1% of the data on the set of football schools that were impacted by the NIL. Therefore, the credibility of my estimates should be challenged. Future work should widen the sample to more schools, use athletic department revenue, and track applications for specific motivations instead of enrollment.

Lastly, a limitation when testing causality of the impact of NIL on university-wide macroeconomic variables is that I relied on a core OLS regression model. This means that my coefficients of interest are NOT capturing causality but rather mere correlations. In the regression model, I only controlled for year and tuition. There are clearly other variables that can impact the endowment and the tuition channel. Thus my model may suffer from severe omitted variable bias. In the future, with more econometric knowledge, I will employ more rigorous causal empirical strategies to test the impact of NIL on key outcome variables of interest.

7.0 Conclusion

This paper questioned whether NIL created new economic value or simply moved that value that already existed away from universities into the hands of these fresh-out-of-puberty athletes. Across the three strands of evidence provided, the answer points towards the first. The athlete side is the clearest. Bijan Robinson put up nearly identical numbers in the game a year apart, but a year later he had deals with Lamborghini and Raising Cane's. Justin Fields and CJ Stroud, both the star players of a CFB powerhouse, but one had a $2.5 million valuation because he was an athlete post NIL. The only difference between the two athletes is that the NIL exists.

The university side shows no matching loss. Endowments at Notre Dame and UT Austin kept climbing after July 2021, with no break, slowdown, or reversal at the policy change, and no NIL coefficient came out negative in either regression. Win rates went up at both schools, and UT Austin's draft output tripled. Whatever athletes gained, it did not come out of the schools

This is what makes a positive-sum rather than a zero-sum. The commercial value of a player's image was always there; the schools were just getting it for free. NIL did not take anything off the table; it let the athletes pick up what was already sitting on it. The Flutie effect still holds; NIL just added a second layer on top of it.

The conclusion has real limits. It covers two elite programs out of hundreds, the model shows correlation and not causation, and the enrollment channel was never properly tested

because schools set their own class sizes. Smaller programs may see the opposite result entirely. And a positive sum in dollars is not the same as welfare - seventeen and eighteen-year-olds now managing brands alongside their development is a question no regression can answer.

In April 2026, Donald Trump signed an executive order called "Urgent National Action to Save College Sports." It tells the NCAA to bring in transfer limits, eligibility caps, and NIL guardrails by August 1st 2026, and backs that up by threatening federal grants and contracts for schools that do not comply. It only applies to schools with over 20 million dollars in athletics revenue, so it targets exactly the kind of programs I looked at. The reasoning behind it is that NIL has created a financial arms race that is draining universities.

My data does not support that reasoning, at least not at the top. Notre Dame and UT Austin both got richer after 2021, not poorer. If I had been advising him, I would have said the money problem is not at the top of the sport; it is at the bottom, where schools have no audience to sell and are losing their best players to the schools that do. An order aimed at the biggest programs does not fix that, and capping what athletes can earn takes value away from the group my paper shows actually gained.

It is also worth noting he had already signed a first order in July 2025 and then went much further in 2026, which suggests the first one did not do much without Congress behind it. A lot of the "college sports is dying" language comes from the NCAA, coaches and athletic directors, and all of them lost power when NIL arrived. That does not automatically make them wrong, but the loudest version of the story is coming from the people with the most to lose, and it did not match what I found in the numbers. That is a reason to check the data before accepting the framing.

To actually influence policy, I would need better evidence than two schools. Three things I would research next: athletic department revenue at smaller programs, which is where the damage should show up if it exists; coaching salaries before and after NIL, since coaches were the ones collecting the money athletes now get; and whether the same pattern holds in basketball, where rosters are far smaller so each player is worth more.

Works Cited

Anderson, Michael L. "The Benefits of College Athletic Success: An Application of the Propensity Score Design." Review of Economics and Statistics, vol. 99, no. 1, 2017, pp. 119-34.

Chung, Doug J. "The Dynamic Advertising Effect of Collegiate Athletics." Marketing Science, vol. 32, no. 5, 2013, pp. 679-98.

Garthwaite, Craig, et al. "Who Profits from Amateurism? Rent-Sharing in Modern College Sports." National Bureau of Economic Research, Working Paper 27734, 2020.

Heckman, James J., and Colleen P. Loughlin. "Are Student-Athletes Exploited?" National Bureau of Economic Research, Working Paper 29072, 2021.

Humphreys, Brad R., and Michael Mondello. "Intercollegiate Athletic Success and Donations at NCAA Division I Institutions." Journal of Sport Management, vol. 21, no. 2, 2007, pp. 265-80.

Meer, Jonathan, and Harvey S. Rosen. "The Impact of Athletic Performance on Alumni Giving: An Analysis of Microdata." Economics of Education Review, vol. 28, no. 3, 2009, pp. 287-94.

National Collegiate Athletic Association v. Alston. 594 U.S. 69. Supreme Court of the United States, 2021.

Weak-Form Efficiency in Developed Versus Emerging Markets: A Fama-French Five-Factor Analysis By Kevin Yuan

Abstract

The weak form of the efficient market hypothesis theorizes that market pricing reflects past data. The confirmation of this data would indicate predictable trends in stock market pricing that can be exploited to achieve abnormal returns. The hypothesis has been tested extensively in developed and emerging markets in isolation; results generally show a trend towards efficiency over time. However, there is little research directly comparing developed and emerging markets in aggregate using similar datasets. This paper seeks to address this problem by using the Fama-French five-factor model to analyze developed and emerging market data across a 35-year period in an attempt to identify market excess returns, an indicator of market inefficiency. This paper hypothesized that developed markets would exhibit weak form efficiency to a greater extent than emerging markets. The hypothesis was tested using a variety of statistical tests: an ADF test to ensure stationarity, as well as autocorrelation, vector autoregression, and impulse response tests to examine the relationship between factors. The data showed that developed markets generally exhibited weak form efficiency, while emerging markets did not. As such, there is an inability to reject the null hypothesis.

Introduction

The efficient market hypothesis, theorized by economist Eugene F. Fama in 1970, introduced three forms of market efficiency: weak, semi-strong, and strong. A market exhibiting weak-form efficiency would have asset prices that reflect historical data, markets with semi-strong form efficiency have asset prices that reflect publicly available information, and markets with strong form efficiency have asset prices that reflect all information (Fama, 1970).

This paper focuses on weak-form markets by examining the predictive power of past market data on present returns. Specifically, an autoregression was run on the Fama-French five-factor model to examine the predictive power of past excess returns on present excess returns. The establishment of a statistically significant link between these factors can provide a basis for the rejection of weak-form market efficiency. A rejection of the weak-form efficiency indicates the presence of arbitrage opportunities, signalling to investors the presence of methods to "beat the market" by examining past data to consistently achieve above-average returns.

The hypothesis for the data is that developed markets generally exhibit market efficiency more so than emerging markets. This is due to a variety of factors, including having greater rates of high-frequency trading—30-70% in developed markets as compared to 5-10% in emerging markets (Ekinci and Ersan, 2022). As high-frequency trading improves market efficiency by increasing the speed at which stock pricing readjusts following major events (Brogaard, 2014), the increased proportion of high-frequency trading in efficient markets helps contribute to their efficiency.

Another driver of higher efficiency in developed markets is increased liquidity. Of the ten largest stock exchanges by market capitalization, seven are based in developed countries as

classified by the dataset used (Royal, 2025; Fama and French, n.d.). Larger market capitalizations imply higher trading volumes (Lo and Wang, 2000), therefore indicating greater liquidity, a factor that increases the efficiency of markets (Chung and Hrazdil, 2010).

Literature Review

The literature on market efficiency has traditionally been quite mixed, but has recently trended towards confirming the efficiency of markets. The first paper written on this topic found compelling evidence of market efficiency on Dow Jones Industrial Average stocks (Fama, 1970). However, later papers examining stock price predictability in the small-cap NYSE-AMEX exchange, the New York Stock Exchange, Nasdaq, and the London Stock Exchange found evidence of stock price predictability (Lo & MacKinlay, 1998; Lo et al., 2002; Mills, 1991). Notably, Lo & MacKinlay (1998) found that larger stocks exhibited less autocorrelation (and therefore greater efficiency), which they stated was largely due to the behavior of larger stocks, such as more frequent trading.

More recently, a 2007 paper examining 8 stock markets in Asia found that 5 were predictable, 2 were unpredictable and 1 was inconclusive (Hoque et al, 2007) - symbolizing a global trend towards market efficiency. This trend has continued into the 2010s, with two papers published in that decade finding that stock markets were generally unpredictable and therefore efficient, at least in developed markets. A paper examining the French stock market found that returns were generally unpredictable and efficient, except for brief inefficient periods triggered by major world events (Boya, 2019). Albano et al. (2013) further captures this sentiment, finding evidence that the Italian stock market is efficient.

This paper revisits the efficient market hypothesis by examining more recent data across developed and emerging markets. Moreover, this paper includes data from more recent activity, which may continue to show the impact of stock market trends towards market efficiency. This paper also addresses a gap in the existing literature by examining the market dynamics of both developed markets and emerging markets in aggregate and as direct comparisons to each other using the Fama-French five-factor model.

Methodology

Two representative samples of stock market returns across 23 developed and 23 emerging economies were accessed via Kenneth R. French's Fama/French 5 Factors for Developed Markets and Kenneth R. French's Fama/French 5 Factors for Emerging Markets datasets respectively. These two samples included market data for their respective economies from July 1990 to April 2025. Each dataset was then used to create 18 portfolios. The stocks were first sorted three times by book-to-market ratio, operating profitability, and investment. The stocks in the 70th percentile and above in each category were categorized as value, robust, and aggressive respectively. The stocks in the 30th percentile and below were categorized as growth, weak, and conservative respectively. The stocks in the top 90% by market cap were categorized as big, and

the ones in the bottom 10% were categorized as small (Fama & French, n.d.). Then, using the portfolios, the factors were constructed as follows:

$$SMB_{(B/M)} = \frac{1}{3}(Small\ Value + Small\ Neutral + Small\ Growth) - \frac{1}{3}(Big\ Value + Big\ Neutral + Big\ Growth)$$

$$SMB_{(OP)} = \frac{1}{3}(Small\ Robust + Small\ Neutral + Small\ Weak) - \frac{1}{3}(Big\ Robust + Big\ Neutral + Big\ Weak)$$

$$SMB_{(INV)} = \frac{1}{3}(Small\ Conservative + Small\ Neutral + Small\ Aggressive) - \frac{1}{3}(Small\ Conservative + Small\ Neutral + Small\ Aggressive)$$

$$SMB = \frac{1}{3}(SMB_{(B/M)} + SMB_{(OP)} + SMB_{(INV)})$$

$$HMB = \frac{1}{2}(Small\ Value + Big\ Value) - \frac{1}{2}(Small\ Growth - Big\ Growth)$$

$$RMW = \frac{1}{2}(Small\ Robust + Big\ Robust) - \frac{1}{2}(Small\ Weak - Big\ Weak)$$

$$CMA = \frac{1}{2}(Small\ Conservative + Big\ Conservative) - \frac{1}{2}(Small\ Aggressive - Big\ Aggressive)$$

RF represents the rate of return on risk-free assets. In this dataset, this was measured using one-month US treasury bills. Mkt-RF represents excess returns over risk-free assets, compensating investors for taking on higher risk in the stock market. SMB, HML, RMW, and CMA each represent the increased returns of companies with smaller market capitalizations, companies with high book to market ratios (undervalued), companies with robust operating profitability, and conservative internal investments over companies that exhibit the opposite properties respectively. The companies were sorted into each of these groups at the end of June every year based on performance from the past year (Fama & French, n.d.).

To measure this, the persistence of excess market returns was examined. Specifically, if there were a change in excess market returns, investors in an efficient market would adjust their investment strategy accordingly, leading to a change in stock pricing that matches the new rate of returns. Therefore, the existence of persistent excess returns, measured through the predictive ability of past returns on current returns, provides evidence to reject the efficient market hypothesis.

The data was analyzed using vector autoregressive (VAR) analysis. In a vector autoregressive model, past data is used to determine correlations with present data using the following equation for a model with k variables:

$y_t = c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \ldots + \Phi_p y_{t-p} + \epsilon_t$, where:

- y_t is a k x 1 matrix showing the values of the variables at time t
- c is a k x 1 matrix of constant terms
- Φ is a k x k matrix of coefficients for each past lag
- ϵ is a k x 1 matrix showing the error for the model which is assumed to be white noise

Unlike traditional autoregressive models, VAR models represent each term as a vector, allowing the coefficient vectors ϕ to show the impact of past values of variables on the present values of other variables.

To qualify this data for autoregressive analysis, stationarity is required. The augmented Dickey-Fuller test was therefore first used to validate first order difference and confirm stationarity (Cheung and Lai, 1995). Then, summary statistics were used to identify larger trends in stock market data. Next, Lag-5 autoregression was used to identify the predictive power of each factor on excess market returns. Afterwards, Impulse-response tests were used to identify the persistence of excess market returns following large shocks. Finally, autocorrelation and partial autocorrelation tests were used to identify statistically significant past predictors of current excess market returns.

Results and Discussion

Firstly, stationarity exists in the datasets for both the developed and emerging markets. An ADF test examining excess returns for both sets of markets gave a p-value of 0 and ADF statistics of -19.49 (figure 1) and -16.50 (figure 2) for developed and emerging markets respectively. This confirms the stationarity of the datasets and therefore provides grounds to reject the null hypothesis for both sets of markets.

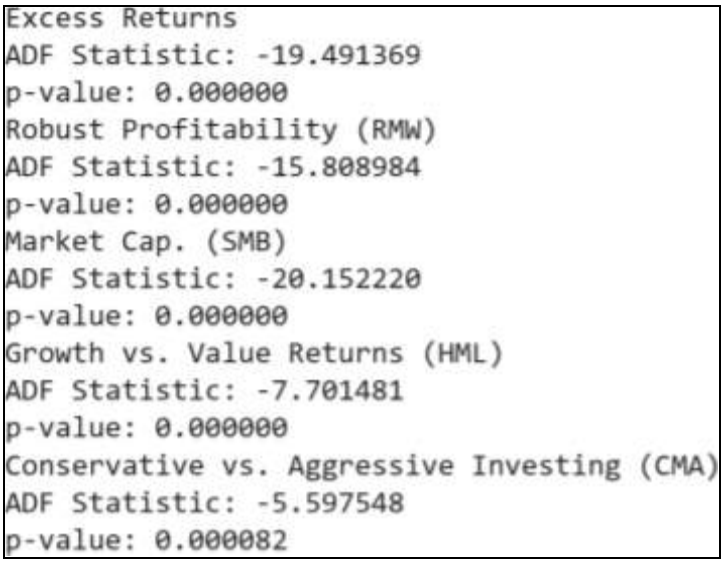


Figure 1: ADF test results for developed markets have highly negative ADF statistics, which are all supported with low p-values indicating statistical significance. Together, they indicate stationarity, which provides grounds to reject the null hypothesis for developed markets.

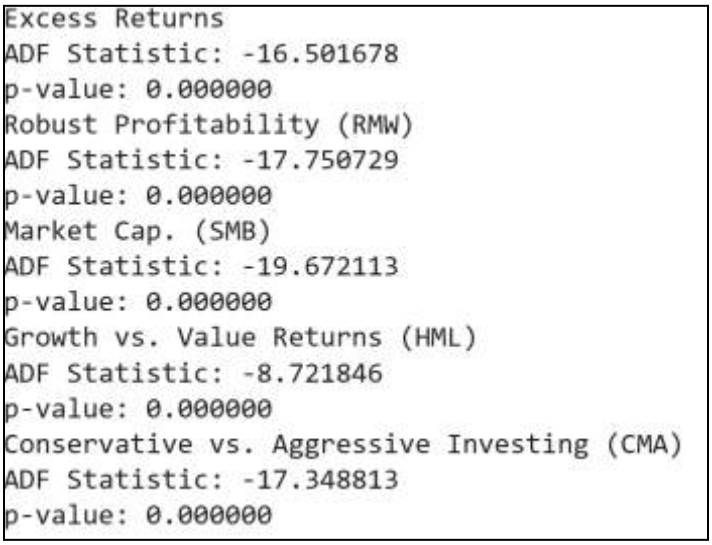


Figure 2: ADF test results for emerging markets have highly negative ADF statistics, which are all supported with p-values of 0 indicating statistical significance. Together, they indicate stationarity, which provides grounds to reject the null hypothesis for emerging markets.

Autocorrelation tests indicate excess market return persistence in emerging markets but not in developed markets. In developed markets, autocorrelation and partial autocorrelation tests performed using the excess market returns showed no statistically significant correlations at the 95% level between current and past excess returns in any capacity (figure 3). On the other hand, emerging markets have statistically significant results in autocorrelation and partial autocorrelation tests (figure 4).

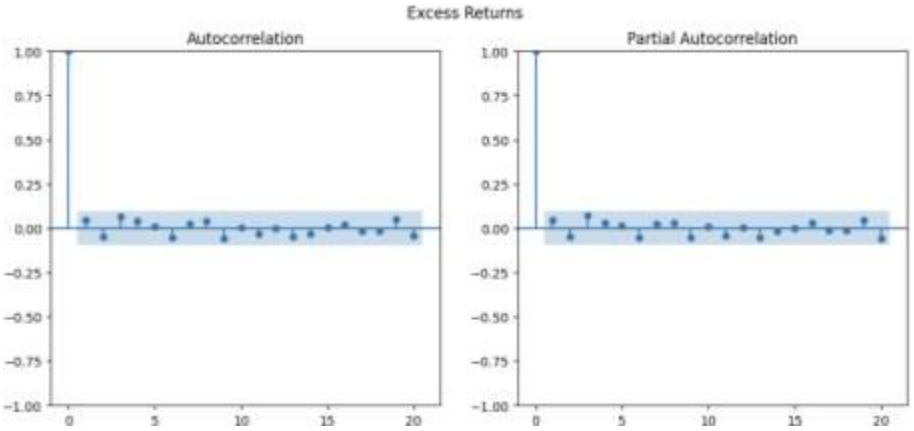


Figure 3: Autocorrelation and partial autocorrelation tests for developed markets show no statistically significant results at the 95% level.

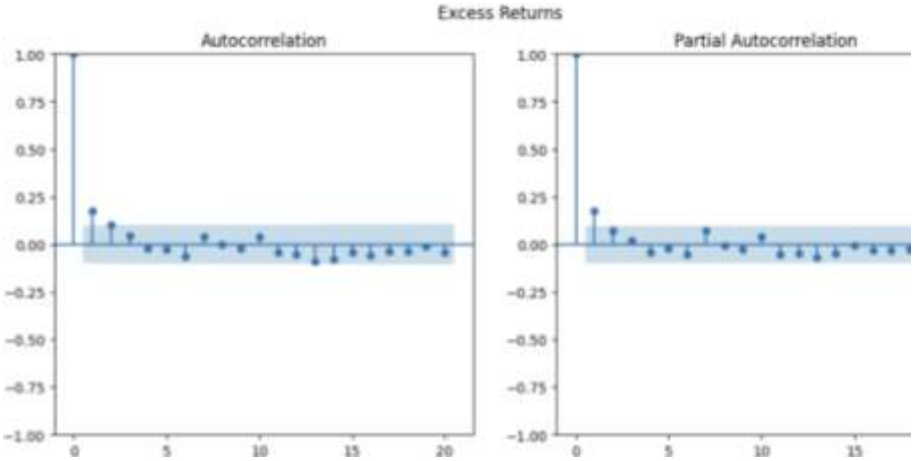


Figure 4: Autocorrelation and partial autocorrelation tests for emerging markets show that excess returns the month prior are correlated with present excess returns in a statistically significant manner at the 95% level.

A VAR analysis using lag-5 variables found statistically significant predictors of excess market returns in both developed and emerging markets at the 95% level, although emerging markets were more statistically significant. In developed markets, lag-2 market excess returns were somewhat statistically significant (p=0.045) and had a coefficient of -0.12 (figure 5). This suggests that over time, developed markets tend to overcorrect for changes in excess market returns, creating small hype cycles that die down and lead to decreased returns. This predictability does not support the efficient market hypothesis for developed markets. However, it is still important to note the limitations of this finding, notably the relatively high p-value of 0.045 and t-stat of -2.0, indicating that these findings may not hold in all scenarios. In emerging markets, lag-1 excess market returns acted as statistically significant (p=0.013) predictors of excess market returns, with a coefficient of 0.15 (figure 6). This suggests that excess market returns persist in emerging markets as they take more time to adjust asset prices to match the new excess returns, allowing changes in excess market returns to be perceivable in the next month. This supports the findings of the autocorrelation and partial correlation tests, as both suggest a

relationship between excess market returns in present months and in the month prior. Therefore, this does not support the efficient market hypothesis for emerging markets.

```
  Summary of Regression Results
==================================
Model:                         VAR
Method:                        OLS
Date:           Sat, 11, Oct, 2025
Time:                      07:45:59
----------------------------------------------------------------------
No. of Equations:         5.00000    BIC:                    8.26290
Nobs:                     413.000    HQIC:                   7.49734
Log likelihood:          -4244.87    FPE:                    1093.64
AIC:                      6.99644    Det(Omega_mle):         805.941
----------------------------------------------------------------------
Results for equation Mkt-RF
==============================================================================
                coefficient       std. error           t-stat            prob
------------------------------------------------------------------------------
const             0.921330         0.283538            3.249           0.001
L1.Mkt-RF        -0.006447         0.060526           -0.107           0.915
L1.SMB            0.139742         0.124875            1.119           0.263
L1.HML            0.012611         0.132146            0.095           0.924
L1.RMW           -0.061480         0.168936           -0.364           0.716
L1.CMA           -0.174881         0.189141           -0.925           0.355
L2.Mkt-RF        -0.123174         0.061575           -2.000           0.045
L2.SMB           -0.028276         0.124309           -0.227           0.820
L2.HML            0.167510         0.132458            1.265           0.206
L2.RMW           -0.145671         0.169437           -0.860           0.390
L2.CMA           -0.249133         0.189366           -1.316           0.188
L3.Mkt-RF         0.005342         0.061794            0.086           0.931
L3.SMB           -0.000445         0.124051           -0.004           0.997
L3.HML           -0.006428         0.132784           -0.048           0.961
L3.RMW            0.002139         0.169789            0.013           0.990
L3.CMA           -0.138803         0.189094           -0.734           0.463
L4.Mkt-RF        -0.012472         0.060587           -0.206           0.837
L4.SMB           -0.050969         0.124551           -0.409           0.682
L4.HML           -0.011672         0.133428           -0.087           0.930
L4.RMW           -0.041726         0.168107           -0.248           0.804
L4.CMA           -0.223091         0.191918           -1.162           0.245
L5.Mkt-RF        -0.028738         0.060699           -0.473           0.636
L5.SMB           -0.039922         0.121769           -0.328           0.743
L5.HML            0.033606         0.132441            0.254           0.800
L5.RMW           -0.191537         0.167178           -1.146           0.252
L5.CMA           -0.084353         0.189675           -0.445           0.657
==============================================================================
```

Figure 5: Lag-5 tests demonstrate a limited negative correlation between present and lag-2 market excess returns that is somewhat statistically significant (p=0.045) at the 95% level. This implies that the market often overcorrects for changes in excess market returns, leading to excess market returns having a negative predictive power on excess returns 2 months in the future.

Summary of Regression Results

Model:	VAR		
Method:	OLS		
Date:	Sat, 11, Oct, 2025		
Time:	07:32:44		
No. of Equations:	5.00000	BIC:	9.26631
Nobs:	389.000	HQIC:	8.46684
Log likelihood:	-4174.50	FPE:	2814.99
AIC:	7.94172	Det(Omega_mle):	2036.97

Results for equation Mkt-RF

	coefficient	std. error	t-stat	prob
const	0.766965	0.405141	1.893	0.058
L1.Mkt-RF	0.153565	0.061650	2.491	0.013
L1.SMB	-0.010331	0.172657	-0.060	0.952
L1.HML	0.011753	0.171970	0.068	0.946
L1.RMW	-0.147471	0.248615	-0.593	0.553
L1.CMA	0.063451	0.199407	0.318	0.750
L2.Mkt-RF	0.063123	0.063332	0.997	0.319
L2.SMB	-0.029544	0.172910	-0.171	0.864
L2.HML	-0.128465	0.176303	-0.729	0.466
L2.RMW	-0.151665	0.249145	-0.609	0.543
L2.CMA	0.049141	0.196775	0.250	0.803
L3.Mkt-RF	-0.002779	0.064364	-0.043	0.966
L3.SMB	-0.083899	0.172986	-0.485	0.628
L3.HML	0.030287	0.173089	0.175	0.861
L3.RMW	-0.326760	0.251730	-1.298	0.194
L3.CMA	-0.023084	0.194081	-0.119	0.905
L4.Mkt-RF	-0.042380	0.065094	-0.651	0.515
L4.SMB	-0.037166	0.169212	-0.220	0.826
L4.HML	-0.175916	0.177018	-0.994	0.320
L4.RMW	-0.159306	0.253300	-0.629	0.529
L4.CMA	0.099123	0.197264	0.502	0.615
L5.Mkt-RF	0.007675	0.064270	0.119	0.905
L5.SMB	0.305121	0.159029	1.919	0.055
L5.HML	-0.046108	0.173685	-0.265	0.791
L5.RMW	0.107531	0.249810	0.430	0.667
L5.CMA	-0.107906	0.198152	-0.545	0.586

Figure 6: Lag-5 tests for emerging markets show a positive correlation between present and lag-1 excess market returns that is statistically significant (p=0.013) at the 95% level. This implies that excess returns for emerging markets are driven by excess returns the month prior, indicating persistence.

Correlation matrices for both markets show generally low correlations between the residuals of excess market returns (Mkt-RF) and many other factors. However, the residuals between excess Mkt-RF and SMB in emerging markets, as well as Mkt-RF and both RMW and CMA in both markets are unusually correlated. This is largely countered by strong data showing the independence of Mkt-RF, but may nonetheless be worth examining in future literature.

Developed markets

Correlation matrix of residuals

	Mkt-RF	SMB	HML	RMW	CMA
Mkt-RF	1.000000	0.000411	-0.125781	-0.365204	-0.377979
SMB	0.000411	1.000000	0.069362	-0.291354	-0.006314
HML	-0.125781	0.069362	1.000000	-0.019937	0.721324
RMW	-0.365204	-0.291354	-0.019937	1.000000	0.076361
CMA	-0.377979	-0.006314	0.721324	0.076361	1.000000

Emerging markets

Correlation matrix of residuals

	Mkt-RF	SMB	HML	RMW	CMA
Mkt-RF	1.000000	-0.306378	0.076301	-0.230007	-0.312535
SMB	-0.306378	1.000000	0.031608	-0.134301	0.157205
HML	0.076301	0.031608	1.000000	-0.496345	0.418234
RMW	-0.230007	-0.134301	-0.496345	1.000000	-0.233316
CMA	-0.312535	0.157205	0.418234	-0.233316	1.000000

Figure 7: Correlation matrices for developed and emerging markets.

Impulse response tests with 5-period horizons were also used to show that shocks to excess market returns take longer to die down in emerging markets than in developed markets. In developed markets, shocks to excess market returns dissipate quickly, returning to the zero level in 1 month (figure 8). This suggests that developed markets quickly adapt to and address shocks to excess returns. This supports the efficient market hypothesis for developed markets. However, in emerging markets, shocks to excess market returns take far longer to dissipate, returning to zero in around 4 months (figure 9). In both graphs, other factors do not have a statistically significant correlation with excess market returns. This suggests that emerging markets do not quickly adapt to shocks to excess market returns, allowing them to persist in a statistically significant manner for long periods of time. As such, this does not support the efficient market hypothesis for emerging markets and reflects the presence of market frictions which may have contributed to inefficiencies.

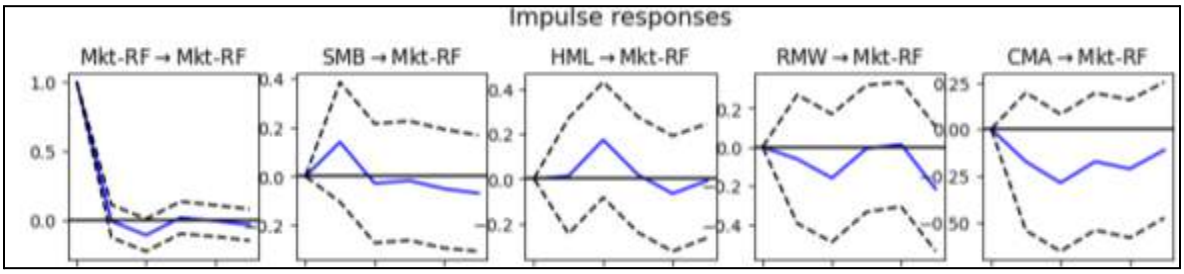


Figure 8: Impulse-response tests for developed markets show a rapid return to normalcy following shocks to excess market returns. This implies developed markets are efficient and can quickly adapt to excess market returns. The other factors do not have a statistically significant impact on excess returns. The full impulse response graphs are in the appendix. Dotted lines indicate confidence intervals.

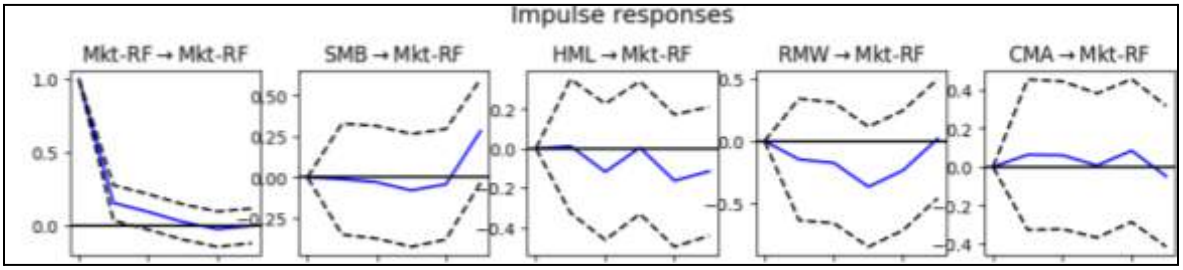


Figure 9: Impulse-response tests for emerging markets show a persistence of excess market returns before returning to normal levels. This supports the idea of excess market returns being persistent in emerging markets. The other factors do not have a statistically significant impact on excess returns. The full impulse response graphs are in the appendix. Dotted lines indicate confidence intervals.

The evidence analyzed in this paper suggests that developed markets largely exhibit weak-form efficiency. Markets quickly adapt to new information and excess returns die down rapidly, as demonstrated with the autocorrelation, partial autocorrelation, and impulse-response tests. The VAR model largely confirmed this, but added a specific circumstance to show that developed markets do not entirely demonstrate efficiency. However, the same cannot be said for emerging markets, where shocks to excess market returns persist before dying down in a

statistically significant manner, as shown in the autocorrelation, partial autocorrelation, VAR, and impulse-response tests.

The findings of this support the hypothesis that developed markets exhibit weak-form efficiency more so than emerging markets. The increased efficiency in developed markets may point to the fact that developed markets have higher trading volumes and are therefore more liquid, leading to greater efficiency and a better ability to both adapt to past trends and respond to new data. This also supports the findings of Boya (2019) and Albano et al. (2013), who argued that stock markets were generally efficient in the developed economies they examined.

Conclusion

In this paper, the analysis of the Fama-French five-factor model of developed and emerging economies using autocorrelation, partial autocorrelation, VAR, and impulse response tests show that developed markets exhibit weak-from efficiency more so than emerging markets.

However, there are certain limitations to this research. Firstly, the returns analyzed include data dating as far back as the 1980s, with all years analyzed having equal weighting. If the observation of a trend towards market efficiency holds, then both datasets would have been impacted by past inefficiencies that may not exist now. As such, findings of inefficiencies may not accurately reflect the current conditions of more efficient markets.

Secondly, there is the possibility of an incomplete analysis caused by omitting variables. While small, the correlation matrix of residuals displayed nonzero coefficients, especially between Mkt-RF and RMW/CMA in developed markets. Thus, residual uncorrelation may not be fully established and there may be factors that have been unaccounted for, leaving room for future analysis using additional variables.

Finally, the logical premise of this paper is based on the assumption that stock price predictability is a good proxy for market efficiency. Although predictability has been found to have a negative association with stock pricing (Eom et al, 2008), potential changes in the relation between the two factors or unpredictable data may hinder the construct validity of using stock price predictability as an indicator of efficiency.

Acknowledgements

I would like to thank Qi Xuan Khoo for help and advice during the research process.

Works Cited

Albano, G., La Rocca, M., & Perna, C. (2013). Testing the Weak Form Market Efficiency: Empirical Evidence from the Italian Stock Exchange. In Neural Nets and Surroundings: 22nd Italian Workshop on Neural Nets, WIRN 2012, May 17-19, Vietri sul Mare, Salerno, Italy (pp. 227-236). Berlin, Heidelberg: Springer Berlin Heidelberg.

Boya, C. M. (2019). From efficient markets to adaptive markets: Evidence from the French stock exchange. Research in International Business and Finance, 49, 156-165.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan. "High-frequency trading and price discovery." The Review of Financial Studies 27.8 (2014): 2267-2306.

Cheung, Y. W., & Lai, K. S. (1995). Lag order and critical values of the augmented Dickey–Fuller test. Journal of Business & Economic Statistics, 13(3), 277-280.

Chung, D., & Hrazdil, K. (2010). Liquidity and market efficiency: A large sample study. Journal of Banking & Finance, 34(10), 2346-2357.

Ekinci, C., & Ersan, O. (2022). High-frequency trading and market quality: The case of a "slightly exposed" market. International Review of Financial Analysis, 79, 102004.

Eom, C., Oh, G., & Jung, W. (2008). Relationship between efficiency and predictability in stock price change. Physica a Statistical Mechanics and Its Applications, 387(22), 5511–5517. https://doi.org/10.1016/j.physa.2008.05.059

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. The journal of Finance, 25(2), 383-417.

Fama, E. F., & French, K. R. (n.d.). Fama/French 5 factors (Developed) [Data set]. Kenneth R. French Data Library, Tuck School of Business at Dartmouth. Retrieved November 21, 2025, from https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_5developed.html

Hoque, H. A., Kim, J. H., & Pyun, C. S. (2007). A comparison of variance ratio tests of random walk: A case of Asian emerging stock markets. International Review of Economics & Finance, 16(4), 488-502.

Lo, A. W., & MacKinlay, A. C. (1988). Stock Market Prices do not Follow Random Walks: Evidence from a Simple Specification Test. The Review of Financial Studies, 1(1), 41–66. http://www.jstor.org/stable/2962126

Lo, A. W., Mamaysky, H., & Wang, J. (2000). Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. The journal of finance, 55(4), 1705-1765.

Mills, T. C. (1991). Assessing the predictability of UK stock market returns using statistics based on multiperiod returns. Applied Financial Economics, 1(4), 241–245. https://doi.org/10.1080/758522008

Royal, J. (2025). The world's largest stock exchanges: 10 biggest by market capitalization. Yahoo Finance. https://finance.yahoo.com/news/world-largest-stock-exchanges-10-181701563.html

Scroll, Signal, Buy: An Integrated Analysis Of Identity, Psychology, Ethics, And Branding In Fashion Consumption By Ami Decelles

Abstract

Where fashion once changed with the seasons, apps like Shein now refresh their inventory by the hour, training consumers to treat their clothing less like a purchase and more like a scroll. Consumerism in retail and fashion has expanded rapidly, driven by e-commerce, social media, and fast fashion's relentless pace of production, though few studies examine how identity, psychology, ethics, and marketing interact as a unified system. Using a literature review of peer-reviewed sources organized into four thematic sections, this paper asks how these forces collectively shape consumer purchasing decisions. Findings reveal a persistent attitude-behavior gap in ethical consumption, driven by price sensitivity, habit, and skepticism toward corporate sustainability claims. Psychologically, emotions routinely override rational decision-making, as consumers shop to regulate mood, seek belonging, and experience immediate gratification, needs the fast fashion industry is designed to exploit. Marketing reinforces this dynamic by operating beneath conscious awareness, using store layout, visual merchandising, manufactured scarcity, and aesthetic manipulation to construct desire rather than convey product information. Identity emerges as a cross-cutting driver, with individuals using fashion to construct and signal selfhood across social contexts, making consumption an inherently communicative act. Taken together, these dimensions are deeply interconnected: psychological vulnerability amplifies susceptibility to marketing, identity-based motivations override ethical intentions, and manufactured desire feeds impulsive behavior. Consumer behavior, this review concludes, is not a product of isolated forces but a complex, socially embedded human experience requiring integrated analysis.

Introduction

Consumerism has expanded rapidly in recent decades, driven by the growth of the retail and fast-fashion industries (Rosely and Ali). The rise of global supply chains, e-commerce platforms, and social media marketing has made products more accessible than ever before, allowing consumers to purchase items quickly, cheaply, and frequently (Dehosse). Fast-fashion brands, in particular, have transformed the pace at which trends are produced and consumed, introducing new styles nearly weekly (Rosely and Ali). The environmental toll of this accelerated cycle is immense, with discarded garments piling into landfills at a rate that the industry's recycling efforts have yet to meaningfully offset (Rosely and Ali). This constant turnover encourages consumers to view clothing and other goods as disposable rather than durable, often without awareness of the environmental and social destruction that cycle leaves behind. (Turkki). Consumer identity refers to the way individuals use products and brands to construct and express who they are (Turkki). When newness becomes the measure of relevance, purchasing shifts from a practical decision to an identity performance, one that must be continuously updated to keep pace with the cycle. This pressure creates fertile ground for impulse buying, as consumers reach for the latest trends not out of need, but out of fear of falling behind. As a result, purchasing is often driven less by necessity and more by a desire to keep up

with rapidly changing trends (Turkki).

The increased accessibility of inexpensive, trend-driven products has contributed to widespread overconsumption, making the study of consumer behavior increasingly important (Turkki). Understanding why individuals make certain purchasing decisions is essential not only for businesses aiming to meet demand but also for researchers and policymakers seeking to address the broader consequences of these consumption patterns (Rosely and Ali). These consumption patterns carry significant economic, social, and environmental consequences, ranging from individual spending habits to broader issues such as labor practices and sustainability (Rosely and Ali). On an economic level, increased consumer spending can stimulate growth and drive innovation (Dehosse). Still, it can also encourage debt accumulation and financial instability among individuals who feel pressured to keep up with rapidly changing trends (Turkki).

The normalization of frequent purchasing, especially within industries like fast fashion, often leads consumers to prioritize quantity over quality, resulting in shorter product lifespans and repeated spending cycles (Rosely and Ali). Socially, these patterns can shape identity and status, as individuals may use consumption to signal belonging, success, or personal values (Turkki). However, this can also reinforce inequality, as access to trend-driven goods is not evenly distributed (Cerchia and Piccolo). In addition, global supply chains that support mass consumption frequently rely on low-wage labor, raising concerns about worker exploitation, unsafe working conditions, and a lack of regulation in manufacturing regions (Cerchia and Piccolo). Environmentally, the consequences of overconsumption are particularly severe (Cerchia and Piccolo). The constant demand for new products contributes to resource depletion, increased carbon emissions, and large amounts of waste (Cerchia and Piccolo). For example, the fast fashion industry is known for its heavy use of water, synthetic materials, and chemical dyes, all of which contribute to pollution and environmental degradation (Cerchia and Piccolo).

The culture of disposability, in which items are quickly discarded after minimal use, further exacerbates landfill overflows and undermines efforts toward sustainability (Cerchia and Piccolo). As a result, consumer behavior is no longer just a matter of personal choice (Turkki). Still, it has become closely tied to global environmental challenges and the urgent need for more sustainable practices (Dehosse). Within this context, several key concepts have emerged as central to understanding consumer decision-making, including consumer identity, psychological drivers, ethical considerations, and the influence of marketing (Turkki). Purchases are often influenced not just by practical needs but by how well a product aligns with a person's self-image or desired identity (Turkki). For instance, someone may choose environmentally friendly products to reflect a commitment to sustainability, while others may prioritize luxury brands to convey status or success (Turkki). Despite the range of research in this area, many studies focus on only one of these dimensions at a time, resulting in a fragmented understanding of consumer behavior (Dehosse). This lack of an integrated perspective has contributed to a notable gap in the literature, particularly in explaining the disconnect between consumers' ethical beliefs and their actual purchasing decisions (Rosely and Ali).

Furthermore, there is limited synthesis across the intersecting roles of identity, psychology, and marketing. To address these gaps, this paper asks: *How do identity, psychological motivation, ethical concern, and marketing strategy interact to shape fashion consumption?* This question is explored through a literature review organized into four thematic sections, using a comparative and synthesis-based approach to analyze how these factors interact.

Methods

This study employs a literature review methodology to examine existing research on consumer behavior in retail and fashion contexts. Sources were gathered from academic databases, including PubMed and Google Scholar, and searches for relevant sources were conducted between March and April 2026. The search process was guided by specific keywords and phrases, such as "consumer identity theory," "impulse buying behavior," "attitude-behavior gap," "brand manipulation," and "fast fashion consumption." Inclusion criteria required sources to be peer-reviewed academic articles relevant to consumer decision-making, particularly those addressing psychology, ethics, branding, or identity. Non-academic sources, duplicated studies, and articles published before 2010, were excluded, as sources older than 15 years were deemed unlikely to reflect the current pace of digital retail, social media marketing, and fast fashion's accelerated production styles. The selected sources were then organized into four thematic sections: consumer ethics, psychological drivers, branding influence, and identity expression, each comprising four studies. Key data extracted from each study included the research question, methodology, findings, and limitations. All sources were managed and organized using Zotero to ensure consistency and efficiency throughout the research process.

Literature Review

Consumer Choice, Ethics, and Resistance

Ethical consumption in the fashion industry refers to purchasing behaviors guided not only by price and style, but also by moral considerations such as environmental sustainability, labor conditions, and corporate responsibility (Dehosse). Increasing awareness of these issues should lead consumers to favor more sustainable and ethical brands (Rosely and Ali). However, research shows a significant disconnect between consumers' stated values and their actual purchasing behavior (Turkki). This disconnect is commonly referred to as the "attitude-behavior gap," which is a central challenge in understanding ethical consumption (Turkki).

One key explanation for this gap is the structure of the fast fashion industry itself. Rosely and Syed Ali examined how fast-fashion business models influence consumer decision-making. Their research focused on how fast fashion encouraged impulsive purchasing through low prices, rapid product turnover, and trend-driven marketing. These conditions reduced the likelihood that consumers will engage in deliberate, ethical decision-making before purchasing. Instead, consumption became immediate and emotionally driven, often prioritizing convenience and affordability over sustainability concerns. Although the study highlights important behavioral

patterns, its reliance on secondary research limits its ability to capture detailed, real-world purchasing behavior across different consumer groups (Rosely and Ali).

The disconnect between ethical awareness and actual behavior is further explored in Turkki's study of the attitude-behavior gap in fashion consumption. Turkki investigated why consumers who express concern for ethical and sustainable fashion do not consistently translate these attitudes into purchasing decisions. The study identified a persistent gap between intention and action, suggesting that ethical attitudes are often overridden by competing influences such as price, habit, convenience, and social norms. Turkki emphasized that ethical awareness alone is not sufficient to change consumer behavior, as purchasing decisions are shaped by a complex combination of psychological and situational factors. However, because the study is largely conceptual, it does not fully measure behavioral outcomes across large or diverse populations (Turkki).

Building on the idea of inconsistent ethical behavior, Dehosse examined how consumers understand and engage with ethical purchasing. Dehosse found that while consumers often express support for ethical fashion and can articulate what ethical consumption means, their actual behavior is inconsistent and situational. Ethical considerations tend to compete with more immediate concerns such as price, convenience, and personal preference. As a result, ethical consumption is often selective rather than consistent across all purchasing decisions. This reinforces the idea that ethical consumption is not a fixed behavioral pattern but rather a flexible, context-dependent practice. The study's broad, literature-based approach, however, limits its ability to measure specific behavioral trends in real-world settings (Dehosse).

In addition to individual decision-making factors, trust and transparency within the fashion industry also play a significant role in shaping consumer behavior. Cerchia and Piccolo, in their analysis of ethical consumers and codes of ethics in the fashion industry, emphasized the importance of credibility in corporate sustainability claims. Their research highlighted that many consumers are skeptical of brands' ethical and environmental messaging due to concerns about "greenwashing" and a lack of transparency. This skepticism reduces the likelihood that consumers will base purchasing decisions on ethical considerations, even when they value sustainability in principle. When consumers cannot verify whether ethical claims are genuine, they are less likely to prioritize them in their consumption choices. This further weakens the relationship between ethical awareness and actual behavior (Cerchia and Piccolo).

Across all four studies, several consistent themes emerge. One of the most significant barriers to ethical consumption is price sensitivity. Sustainable and ethically produced fashion items are often more expensive than fast fashion alternatives, making them less accessible to a wide range of consumers (Cerchia and Piccolo). Even individuals who support ethical values may be unable or unwilling to pay higher prices consistently (Cerchia and Piccolo). Accessibility also plays a major role, as ethical fashion options are often less visible and less readily available than mainstream alternatives (Dehosse). This structural imbalance reinforces reliance on fast fashion, even among ethically aware consumers (Turkki).

Another recurring issue is skepticism toward corporate sustainability claims. As

awareness of "greenwashing" increases, consumers become more cautious about trusting brands' ethical messages (Cerchia and Piccolo). This lack of trust undermines the effectiveness of ethical labeling and corporate codes of conduct, further complicating decision-making (Turkki). As a result, even well-intentioned consumers may disengage from ethical consumption altogether due to uncertainty about what constitutes a truly ethical choice (Dehosse).

Taken together, these studies demonstrate that ethical considerations alone are relatively weak predictors of consumer behavior in the fashion industry (Turkki). While many consumers express concern about sustainability and labor ethics, these attitudes are frequently outweighed by external constraints such as cost, convenience, and trust (Turkki). The persistence of impulsive purchasing behavior, combined with structural limitations in the fashion market, reinforces the gap between ethical intention and actual consumption (Turkki).

Ethical consumption in fashion is highly inconsistent and context-dependent (Cerchia and Piccolo). Rather than reflecting a stable set of values translated into action, consumer behavior is shaped by a dynamic interaction between ethical awareness and practical constraints (Cerchia and Piccolo). As a result, the attitude-behavior gap remains a defining feature of consumer choice in the fashion industry, highlighting the limits of individual ethical intention in driving sustainable consumption patterns (Turkki).

Emotional and Psychological Drivers of Consumption

Emotional and psychological consumption refers to purchasing behavior driven not by practical need but by internal states, mood, emotion, identity, social belonging, and psychological impulse (Samadi). Rather than following rational cost-benefit reasoning, consumers in this mode respond to feelings, desires, and unconscious motivations (Preethi and Gupta). This pattern is especially visible in fast fashion, where low prices, constant innovation, and aggressive marketing create environments that repeatedly trigger spontaneous, unplanned purchases (Hsu et al.). The accessibility and speed of fast fashion make it a near-perfect vehicle for emotionally driven consumption, and understanding the psychology behind it is essential to understanding the industry itself (Mair).

Mair's 2014 paper provides a theoretical framework for understanding non-rational fashion consumption. Rather than offering new empirical data, Mair critically examines the cognitive and social psychological foundations of fashion demand, arguing that the dominant economic assumption, that consumers are rational agents, is simply not realistic. Consumers do not have access to all available information, nor do they have the cognitive capacity or motivation to process it perfectly. Instead, social relationships, values, identity, and emotion consistently distort decision-making. Mair finds that many individuals buy far more fashion items than they need due to impulse and emotional buying, with many purchases remaining unworn. She frames fashion as satisfying two simultaneous and paradoxical psychological needs: the desire for originality and individuality on one hand, and, on the other, the desire for conformity and social belonging. Mair also raises an important methodological concern, that the self-report surveys used in most consumer research are fundamentally limited. She raises this

concern because consumers often cannot accurately recall the true, emotionally driven reasons behind their purchases due to cognitive biases and the unreliability of memory. This limitation applies broadly across the field (Mair).

Preethi and Gupta take a more granular look at the emotional triggers that drive impulsive apparel purchases specifically, conducting a literature review across both online and offline retail contexts, Their review finds that emotions play a pivotal role in impulsive buying: positive emotions such as excitement or happiness can lead to spontaneous purchases, while negative emotions like stress or sadness can equally drive impulsive apparel buying more frequently, and situational factors (urgency, scarcity, influencer marketing, and store environment) act as powerful amplifiers of emotional impulse. The paper also identifies a significant gap in the literature: the specific psychological processes triggered by social media and online advertising on impulsive buying remain underexplored. Because this paper is a literature review rather than an original empirical study, it does not produce primary data, and its conclusions depend on the quality and consistency of the studies it analyzes (Preethi and Gupta).

Samadi's 2019 study investigates how individuals' mood and emotional states influence their clothing purchase behavior, and how clothing choices in turn reflect personality and emotion. Using a mixed-methods approach combining both quantitative and qualitative data, Samadi found strong relationships between mood and clothing preference, and significant links between an individual's emotional state and what they choose to wear. Mood emerged as an important predictor of clothing preference, even when emotional feelings were more restrained, suggesting that how someone feels on a given day shapes not just what they wear, but also what they are drawn to buy. While the study offers valuable insight into the mood-clothing connection, its relatively limited sample size and mixed-methods design constrain the applicability of its findings. It also does not isolate specific emotions deeply enough to determine which moods most strongly predict purchase behavior (Samadi).

When Samadi focuses on the individual emotional experience, Hsu and colleagues' 2025 meta-analysis takes a much broader view, synthesizing empirical evidence from studies published between 2004 and 2024 to examine the psychological and behavioral determinants of fast fashion consumption. Integrating the Theory of Planned Behavior with brand-related constructs (perceived scarcity, perceived quality, and self-congruity) the authors used a random-effects meta-analysis and found that brand attitude is the strongest predictor of purchase intention, while self-congruity with fashion brands significantly enhances all components of planned behavior. Perceived quality, meanwhile, influences both cognitive and social evaluations. Crucially, the study confirms that the psychological mechanisms underlying fast fashion consumption, including emotional impulses, brand identification, and peer influence, are far more complex than traditional rational decision-making models suggest. As a meta-analysis, however, the study is constrained by the quality and variety of other studies. It mainly focuses on brands, which may overlook other feelings, such as mood or shopping for comfort (Hsu et al.).

Beyond the individual findings of each paper, several cross-cutting themes emerge. First, shopping functions as emotional regulation: multiple studies confirm that consumers use

unplanned purchases to manage or escape negative emotional states, a behavior commonly described as retail therapy (Preethi and Gupta). Second, instant gratification is a central motivator (Samadi). Mair traces this back to early theories of conspicuous consumption. She notes that fast fashion, with its low prices and relentless innovation, has made gratification-driven buying accessible across all socioeconomic groups, while online shopping has further reduced the friction between impulse and purchase (Mair). Third, habit formation plays a meaningful role. Samadi's findings suggest that repeated emotional associations between mood and clothing preference solidify into habitual buying patterns over time (Samadi). Preethi and Gupta note that past emotionally satisfying purchases reinforce the behavior loop, making impulsive fashion buying increasingly automatic (Preethi and Gupta).

Taken together, these four studies converge on a single central insight: emotions routinely override rational decision-making in fashion consumption (Preethi and Gupta). Mair dismantles the economic fiction of the rational consumer. Samadi shows that even daily mood meaningfully shapes what people buy. Preethi and Gupta map the specific triggers that ignite impulse purchases. Finally, Hsu and colleagues confirm at scale that emotional and identity-based constructs are stronger predictors of purchase than deliberate reasoning. Consumption in fashion, then, is not primarily a practical act (Hsu et al.). People buy clothes to manage their emotions, signal their identities, seek belonging, and experience immediate pleasure (Samadi). As a result, the fast fashion industry is structurally designed to meet and exploit those needs (Preethi and Gupta).

Branding and Psychological Power

Branding influence in the fashion industry refers to the ways in which visual, aesthetic, and psychological marketing strategies shape consumer perception, desire, and purchasing behavior, often independently of a product's actual qualities (Bruderman). Marketing psychology examines how retailers and brands leverage cognitive and emotional tendencies to guide consumer decisions, from the physical layout of a retail store to the construction of luxury brand identity (Sher). Together, these forces represent some of the most powerful and least visible mechanisms driving consumption in the fashion industry, operating beneath conscious awareness and raising significant ethical concerns.

One foundational explanation for marketing's psychological power lies in the strategic use of aesthetics. Bruderman, in his philosophical analysis of aesthetic manipulation in consumer culture, examined how the appearance of products and marketing materials functions as a form of manipulation. Drawing on neo-Marxist theory and cultural criticism, Bruderman argued that the use of aesthetics as a manipulative tool is both complex and compelling, serving as a veil that camouflages certain aspects of consumerist activities. His framework categorizes aesthetic manipulation into benign, strategic, and restraining forms, with the study concluding that the visual presentation of consumer products has become so normalized that it obscures difficult ethical questions about the social and environmental consequences of production (Bruderman). Because the study is entirely conceptual and philosophical, it does not include empirical data or

measure actual consumer responses, which limits its ability to confirm how aesthetic manipulation operates across specific real-world purchasing contexts.

Building on the idea that the retail environment itself functions as a psychological tool, Florea and colleagues investigated how visual merchandising and store layout influence customer shopping decisions in fashion retail. Their study collected survey data from 488 respondents in the South Muntenia region of Romania and analyzed responses using structural equation modeling. Results confirmed that store layout, visual merchandising, and service quality have a significant positive effect on consumer satisfaction. More specifically, exterior lighting and brand signage emerged as particularly strong predictors of purchase decisions, and mannequins influenced the buying behavior of 73% of respondents. The study concluded that retailers who optimize product arrangement, window displays, and lighting create psychologically engaging environments that increase foot traffic, encourage impulse purchases, and build customer loyalty (Florea et al.). A notable limitation, however, is that the sample was geographically restricted to one region of Romania, thereby limiting the generalizability of the findings to other cultural and economic contexts.

While the studies above focus on how marketing manipulates consumers in practice, Sher directly addressed the ethical dimensions of these tactics. In his framework for assessing immorally manipulative marketing tactics, Sher reviewed existing academic and philosophical debates on persuasion and manipulation to develop operational criteria for evaluating when a marketing tactic crosses an ethical line. The study concluded that a marketing tactic is manipulative if it is intended to motivate by undermining what the marketer believes is the audience's normal decision-making process, either by deception or by playing on a vulnerability. Importantly, Sher argued that manipulation is not inherently immoral, but becomes so when it bypasses rational consumer agency without consent, offering marketers a practical tool for distinguishing between legitimate persuasion and exploitation. Because the framework is theoretical rather than empirical, however, its real-world application remains difficult, as it requires accurate knowledge of marketers' intentions, which are rarely transparent or verifiable (Sher).

Extending the discussion of psychological influence specifically to the luxury sector, Woodside and Ko examined the core tenets of fashion marketing theory from the perspective of economic psychology. Using multiple regression analysis and drawing on available empirical literature, the authors proposed that impactful fashion marketing moderates the relationships among price, consumer demand for a brand by psychological customer segments, and, subsequently, firm profitability. Their research found that consumers high in chronic desire for conspicuous consumption and rarity respond more intensively to luxury branding cues regardless of price, and that the impact of manipulated rarity (high versus low) increases dramatically among consumers highest versus lowest in chronic desire for rarity. These findings demonstrate how branding strategies can be precisely calibrated to exploit psychological predispositions, particularly in luxury markets. As a largely literature-based theoretical framework, however, the study does not draw from a single original dataset, limiting the direct empirical weight of its

conclusions (Woodside and Ko).

Across all four studies, several consistent themes emerge. First, branding and marketing shape consumer perception of quality far more than the actual attributes of products do. Environmental cues such as lighting, aesthetic design, and visual merchandising create associations with value and prestige that frequently override objective evaluation (Florea et al.; Bruderman). Second, manufactured scarcity and exclusivity function as powerful psychological levers, particularly among customers predisposed to seek status through consumption, allowing brands to command premium pricing well beyond any functional product value (Woodside and Ko). Third, consistent aesthetic experiences, from storefront signage to luxury brand identity, cultivate long-term brand loyalty by building emotional associations that persist across purchasing decisions (Florea et al.; Woodside and Ko).

Taken together, these studies demonstrate that marketing exerts substantial psychological control over consumers in the fashion industry. Whether through the philosophical lens of aesthetic manipulation (Sher), the practical mechanisms of store design (Florea et al.), the ethical analysis of decision-making interference (Bruderman), or the strategic use of luxury identity and scarcity (Woodside and Ko), branding consistently constructs desire and perceived value rather than simply conveying product information. The persistence of these strategies across both luxury and everyday retail contexts underscores the depth of marketing's psychological reach and raises important ethical questions about the boundaries of consumer autonomy (Sher; Bruderman). Rather than reflecting a neutral exchange of information, fashion marketing functions as a carefully engineered system of psychological influence, shaping what consumers want, what they believe products are worth, and ultimately, what they choose to buy.

Identity, Self-Expression, and Social Signaling

Identity-based consumption refers to the practice of purchasing goods not simply to satisfy functional needs, but to express, construct, and communicate a sense of self to others (Esaine). In the context of fashion, clothing and personal appearance serve as one of the most visible and socially legible forms of self-expression available to individuals. What people wear communicates who they are, which groups they belong to, and what values and aspirations they hold (Hokkanen et al.). Because fashion operates simultaneously as a personal and public practice, it sits at the intersection of individual identity and social signaling, making it a particularly rich area for understanding how consumption functions as a form of communication (Kuo).

Hokkanen extended this inquiry into the postmodern dimensions of brand and identity by examining how fashion brands function as tools for constructing self and social identities in contemporary consumer culture. Drawing on Consumer Culture Theory (CCT), a study on how people use brands and products to shape their identity and social connections, and a qualitative literature review, the study explored how postmodern consumers form deeply personal relationships with brands, using them to express both who they are and who they wish to become. Brands have become so interwoven with consumption that today's postmodern

consumers often have deeply personal relationships with brands, as meanings are constructed in consumers' minds in different settings, situations, and psychological states. Hokkanen argued that the consumer is an active coproducer of brand meaning and that fashion consumption serves not merely as a reflection of identity but as an ongoing process of identity construction and negotiation. The study further noted that fashion consumption from a social identity perspective helps in understanding how consumers form and maintain brand relationships and perceive the brand. A significant limitation, however, is that the study is literature-based and qualitative in scope, conducted specifically in the Finnish context, which the author acknowledges limits its generalizability to other cultural settings (Hokkanen et al.).

One of the most grounded examinations of this dynamic comes from Ortiz Esaine, who investigated how fashion consumption shapes identity formation among young people in an emerging social class context. Using a qualitative, ethnographic methodology conducted over 24 months in Lima Norte, Peru, the study included in-depth interviews with 20 participants, two focus groups, and field observation across shopping malls, universities, and public spaces. The findings revealed that identity in postmodern consumer culture is fluid and contextual rather than fixed. Young people in Lima Norte used fashion as a strategic tool to navigate different social theaters, presenting themselves differently at work, at university, and in their neighborhoods. Commercial brands, beyond the real need for consumer objects, acted as mythologies that give meaning to the theater of social roles, creating a symbolic reference system that operates by creating empathies or distinctions between young people, but reinforcing the identities present. The study found that fashion consumption functioned as a language allowing access to social groups, marking the limits between what is accepted and what is exclusive. A key limitation of this research is that its qualitative, single-city methodology limits the generalizability of findings beyond the specific cultural and socioeconomic context of Lima Norte (Esaine).

Kuo approached the relationship between identity and consumption from a quantitative perspective, developing and testing a framework in which retail brand personality influences consumer behavioral outcomes through identity and social identity theories. The study applied identity and social identity theories to develop and test a framework in which retail brand personality influences consumer outcomes, specifically positive word-of-mouth and patronage intention toward the retailer, through public and private self-congruity, strengthened by shopping conspicuousness, and retail brand identification. Using a 2x2 between-subjects questionnaire design administered to a sample of US consumers and analyzed through structural equation modeling, the study found that the “genuine” dimension of retail brand personality was the most influential predictor of consumer behavior, operating both directly and indirectly through self-congruity and brand identification. Notably, the high shopping conspicuousness situation strengthened the relationship between public self-congruity and the overall retail brand identification, suggesting that the social visibility of shopping amplifies identity-driven consumption. A limitation of this study is that its sample was drawn exclusively from US consumers, which may reduce its applicability to markets with different cultural orientations toward brand identity and social conformity (Kuo).

Zuliarni and colleagues examined identity-based consumption from a different angle, investigating how global fashion brands compete with local alternatives through the lens of social and cultural identity. The study explored the antecedents of consumers' purchasing intentions for local fashions, mainly considering buying intentions for global alternatives, price, and product design, based on a non-probability survey of 260 young adults and hierarchical multiple regression analysis. The findings revealed that stronger buying intentions toward global fashion brands significantly reduced consumers' willingness to purchase local substitutes, suggesting that global brands carry aspirational identity values that local products struggle to compete with. However, the study also found that when local products offer a distinctive design that aligns with consumers' cultural and social identities, purchase intention increases. The better the product design, the more unique it is, and the more it represents consumers' cultural and social identities, the more likely they are to want it, The study's generalizability is limited by its focus on young Indonesian consumers in an e-commerce context, and the authors note that future research should expand to include a wider variety of samples, product categories, and national contexts (Zuliarni et al.).

Across all four studies, a consistent and important theme emerges: consumption is not simply a transaction but a form of communication. Whether expressed through the chameleon-like wardrobe strategies of emerging-class youth in Lima (Esaine), the symbolic brand relationships of postmodern consumers (Hokkanen et al.), the identity-driven loyalty patterns of retail shoppers (Kuo), or the competition between global prestige and local cultural pride (Zuliarni et al.), fashion choices consistently serve as signals to others about who one is and where one belongs. Social context plays a central role in this process: the visibility of shopping situations, the aspirational pull of global brands, and the pressure of social group norms all shape how identity is communicated through consumption (Kuo). Together, these studies also highlight how identity-based consumption can reinforce social hierarchies, as the desire for inclusion and distinction drives consumers toward brands and products that carry the greatest social legibility.

In summary, identity is a central and persistent driver of consumer choice in the fashion industry. Individuals use clothing and brand affiliation to construct, maintain, and signal their sense of self across different social contexts, making consumption an inherently social and communicative act (Hokkanen et al.). The research reviewed here demonstrated that this process is shaped by both deeply personal motivations and powerful external social forces, and that understanding the relationship between identity and consumption is essential to understanding why people buy what they buy.

Discussion

This literature review set out to answer a deceptively simple question: how do identity, psychology, ethics, and marketing drive our shopping experiences? The research reveals that there is no single, clear answer. Consumer behavior is not the product of one isolated force but rather the outcome of several deeply intertwined factors, each shaping and being shaped by the

others. Taken together, the four thematic sections of this review paint a picture of a modern consumer who is emotionally complex, socially motivated, psychologically vulnerable, and often inconsistent.

One of the most striking findings across the literature is just how persistent the gap is between what people say they value and what they actually do. Across the studies reviewed in the ethics section, consumers repeatedly expressed genuine concern for sustainability, fair labor practices, and environmental responsibility (Rosely and Ali). And yet, when standing in front of a rack of affordable, trendy clothing, those values routinely become subordinate to price, convenience, and habit (Turkki; Rosely and Ali). This is not simply a matter of hypocrisy or indifference. Rather, it reflects the reality that ethical consumption does not exist in a vacuum. It competes with financial constraints, with deeply ingrained shopping habits, and with a fast fashion industry that has been deliberately structured to make impulsive, frequent purchasing feel natural and even necessary (Rosely and Ali). When consumers also cannot trust whether a brand's sustainability claims are genuine, even their best intentions are undermined (Cerchia and Piccolo). The attitude-behavior gap, then, is not a personal failure. It is a structural one.

This structural dimension becomes even clearer when viewed alongside the psychological findings. The research in this area consistently challenges the assumption that consumers are rational agents who weigh costs and benefits before making decisions. Mair was direct in dismantling this idea, arguing that cognitive limitations, emotional states, and social pressures mean that truly rational consumer decision-making is far more of an ideal than a reality. Samadi reinforced this by showing that something as ordinary as a person's mood on a given day can meaningfully shape what they feel drawn to buy. Preethi and Gupta mapped the specific emotional triggers: excitement, stress, boredom, and the fear of missing out. These triggers convert a browsing session into an unplanned purchase. Preethi and Gupta found that scarcity cues and urgency messaging were among the most effective in prompting immediate, impulsive purchases. And at the broadest level, Hsu and colleagues confirmed through meta-analysis that emotional and identity-driven constructs outperform deliberate reasoning as predictors of purchase behavior.

What this body of research makes clear is that emotions are central to consumer behavior, and not a peripheral factor. People shop to feel better, to reward themselves, to cope with stress, to experience a fleeting sense of control or pleasure (Hokkanen et al.). The fast fashion industry, with its low prices and constant influx of new styles, has positioned itself perfectly to serve these needs (Mair). Every new drop, every flash sale, every notification on a shopping app is designed to intercept a consumer at a moment of emotional susceptibility (Preethi and Gupta). Understanding this dynamic matters not just academically but practically. If we want to understand why people struggle to shop more ethically or more intentionally, we have to start by acknowledging the emotional architecture that shopping is built upon.

The branding section deepened this understanding considerably. Marketing, as the studies by Bruderman, Florea and colleagues, Sher, and Woodside and Ko all demonstrated, is not merely informational. It is psychological. Retail environments are carefully engineered through

lighting, layout, visual merchandising, and signage to create atmospheres that encourage lingering and impulse buying (Bruderman). Luxury brands go further still, using manufactured scarcity and carefully curated aesthetics to position their products as symbols of status and exclusivity (Florea et al.). The result is that the perceived value of a product is often constructed almost entirely by marketing rather than by the product's actual qualities (Woodside and Ko). Consumers are not simply discovering what they want. In many cases, they are being told what to want, and the messaging is sophisticated enough that the instruction feels like personal preference (Sher).

This raises important ethical questions that the research, particularly Sher and Bruderman, begins to address but does not fully resolve. At what point does persuasion become manipulation? When a store layout is designed to maximize time spent browsing, or when artificial scarcity is manufactured to trigger urgency, is the consumer making a free choice? These questions do not have easy answers, but they matter enormously for how we think about consumer autonomy. The fact that marketing operates largely beneath conscious awareness makes it all the more powerful and all the more worthy of scrutiny (Sher; Bruderman).

The identity section brought these threads together in a particularly compelling way. Across all four studies reviewed (Hokkanen, Ortiz Esaine, Kuo, and Zuliarni and colleagues), consumption emerged not as a passive act of satisfying needs but as an active, ongoing process of self-construction and social communication. People use fashion to tell the world who they are, which groups they belong to, and what they aspire to become (Kuo). This is true across vastly different cultural and socioeconomic contexts, from postmodern brand culture in Finland to the wardrobe strategies of upwardly mobile young adults in Lima (Esaine). What changes across these contexts is not the impulse itself but the specific brands, aesthetics, and styles that carry social legibility in a given setting (Hokkanen et al.).

This identity dimension also helps explain the persistence of the attitude-behavior gap in new ways. If fashion is a language through which people communicate belonging and aspiration, then choosing a less trendy or more sustainable option carries social costs that go beyond mere aesthetics (Woodside and Ko). It risks misalignment with one's social group or with the identity one is actively trying to construct (Florea et al.). This is why ethical consumption cannot be addressed through awareness campaigns alone; it requires grappling with the deeply human need to be seen, accepted, and understood by others.

Importantly, these four dimensions (ethics, psychology, branding, and identity) are not parallel but interconnected. Psychological vulnerability to emotional triggers makes consumers more susceptible to branding manipulation. Identity-based motivations intersect with social norms to override ethical intentions. Manufactured desire through marketing feeds into impulsive purchasing patterns rooted in emotional regulation. Each force amplifies the others, and together they create a consumer landscape that is remarkably resistant to simple interventions. Telling people to shop more ethically, for example, is unlikely to be effective if it does not also account for the emotional, social, and identity-based needs that shopping fulfills.

These are notable limitations in the existing literature that future research should address.

Many of the studies reviewed relied on retrospective, self-report data, which Mair rightly flagged as methodologically problematic, given that consumers often cannot accurately articulate the true reasons behind their purchasing decisions. Several studies were also geographically or demographically narrow, limiting their generalizability across different cultural contexts. Additionally, relatively few studies attempted to integrate multiple dimensions of consumer behavior simultaneously, leaving the interaction effects between identity, psychology, ethics, and marketing underexplored. Future research would benefit from mixed-methods designs that track real purchasing behavior over time, rather than relying solely on stated intentions or retrospective self-reports.

There are also important limitations within this review that should be acknowledged. As a narrative review rather than a formal scoping or systematic review, this paper does not attempt to capture every existing study related to consumer behavior, fast fashion, identity, or ethical consumption. The selection of sources was shaped by thematic relevance and the need to build a cohesive discussion across multiple dimensions of shopping behavior, which means that some perspectives and findings were inevitably left out. In addition, because the review draws from studies across different disciplines, including psychology, sociology, marketing, and consumer research, there are differences in methodology, terminology, and theoretical frameworks that can make direct comparison difficult. The review also places significant emphasis on contemporary fast fashion and digitally mediated shopping environments, which may overlook how these dynamics operate differently in other retail sectors or in non-Western contexts where consumer culture is shaped by different social and economic realities. Finally, interpretation is an unavoidable part of narrative analysis. While every effort was made to represent the research accurately and fairly, the synthesis of findings is still influenced by subjective judgment regarding which themes were most significant and how they connect to one another. These limitations do not undermine the value of the review, but they do highlight the importance of viewing its conclusions as a part of a broader and ongoing conversation about the complexity of modern consumer behavior.

Ultimately, what this review reveals is that shopping is one of the most psychologically and socially loaded activities in modern life. Meaningful change will require interventions that address all four dimensions simultaneously: regulating the desire manufactured by marketing, dismantling the status hierarchies that drive overconsumption, making sustainable options financially accessible, and holding brands accountable to transparent labor and environmental practices. It is a space where personal identity meets social expectation, where emotional impulse meets ethical intentions, and where individual choice meets the extraordinary power of engineered marketing environments. Understanding consumer behavior fully requires holding all of these dimensions in view at once. Not as competing explanations, but as parts of a single, complicated human experience. If researchers, policymakers, and even consumers themselves can begin to see shopping in this more integrated way, there is real potential to move toward patterns of consumption that are not only more sustainable and ethical, but more conscious and self-aware.

Works Cited

Bruderman, Eli. “The Nature of Aesthetic Manipulation in Consumer Culture.” South African Journal of Philosophy, vol. 35, no. 2, June 2016, pp. 210–23. DOI.org (Crossref), https://doi.org/10.1080/02580136.2016.1172750.

Cerchia, Rossella Esther, and Katherine Piccolo. “The Ethical Consumer and Codes of Ethics in the Fashion Industry.” Laws, vol. 8, no. 4, Sept. 2019, p. 23. DOI.org (Crossref), https://doi.org/10.3390/laws8040023.

Dehosse, Solène. A Study of Ethical Fashion Consumption. 2020. Stellenbosch University, Master’s Thesis. Stellenbosch, https://scholar.sun.ac.za/items/8ef19630-a2dd-4ca4-af7a-b529c1c2cb98.

Esaine, Nicolás Ortiz. “Anthropology of Consumption, Fashion and Identity in Emerging Class.” Revista Gestão Inovação e Tecnologias, vol. 11, no. 4, July 2021, pp. 3120–33. DOI.org (Crossref), https://doi.org/10.47059/revistageintec.v11i4.2355.

Florea, Nicoleta-Valentina, et al. “The Influence of Fashion Retailers on Customer Psychology Using Visual Merchandising and Store Layout to Improve Shopping Decision.” Journal of Theoretical and Applied Electronic Commerce Research, vol. 20, no. 1, Feb. 2025, p. 40. DOI.org (Crossref), https://doi.org/10.3390/jtaer20010040.

Hokkanen, Silja, et al. Fashion Brands and Consumption in Postmodern Consumer Culture.

Hsu, Shu-Chuan, et al. Fashion Consumption: A Meta-Analytic. 2026.

Kuo, Ya-Hui. “The Retail Brand Personality—Behavioral Outcomes Framework: Applications to Identity and Social Identity Theories.” Frontiers in Psychology, vol. 13, Aug. 2022, p. 903170. DOI.org (Crossref), https://doi.org/10.3389/fpsyg.2022.903170.

Mair, Carolyn. “A Cognitive and Social Psychological Perspective on the Demand for Fashion.” [University of Quebec, Montreal], 2014, 18th International Conference of the Association for Cultural Economics International (ACEI).

Preethi, D. Maria Nirmal, and Chithambar Gupta. THE POWER OF EMOTION: DRIVING IMPULSE BUYS IN FASHION RETAIL. vol. 22, no. 1, 2024.

Rosely, Nurhidayah, and Sharifah Faridah Syed Ali. “A Challenge Towards Sustainable Fashion Consumption: Fast Fashion and Impulsive Purchase Behaviour.” International Journal of Accounting, Finance and Business, vol. 8, no. 47, 2023, pp. 63–76, https://doi.org/10.55573/IJAFB.084707.

Samadi, Sepideh. “The Effects of Mood and Emotion on Consumer Purchase Behavior of Buying Clothes.” International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 12S, Dec. 2019, pp. 770–74. DOI.org (Crossref), https://doi.org/10.35940/ijitee.L1178.10812S19.

Sher, Shlomo. “A Framework for Assessing Immorally Manipulative Marketing Tactics.” Journal of Business Ethics, vol. 102, 2011, pp. 97–118, https://doi.org/10.1007/s10551-011-0802-4.

Turkki, Reetta. The Attitude-Behavior Gap in Fashion Consumption. 2023. Tampere University, Bachelor’s Thesis. Tampere University,

https://trepo.tuni.fi/bitstream/handle/10024/149655/TurkkiReetta.pdf;jsessionid=E8E89A6079B6D82F17EB50C77AAF58E6?sequence=2.

Woodside, Arch G., and Eunju Ko. “Luxury Fashion Theory, Culture, and Brand Marketing Strategy.” Advances in Culture, Tourism and Hospitality Research, edited by Eunju Ko and Arch G. Woodside, Emerald Group Publishing Limited, 2013, pp. 1–14. DOI.org (Crossref), https://doi.org/10.1108/S1871-3173(2013)0000007004.

Zuliarni, Sri, et al. “The Impact of Buying Intention of Global Fashion on Local Substitute: The Role of Product Design and Price.” Heliyon, vol. 9, no. 11, Nov. 2023, p. e22160. DOI.org (Crossref), https://doi.org/10.1016/j.heliyon.2023.e22160.

Comparative Analysis of the Financial Sustainability of Three Global Pension Systems: Lessons from Sweden and Singapore for U.S. Social Security Reform By Anisha Kalaria

Abstract

This paper presents a comparative policy analysis of the financial sustainability of the retirement pension systems of the United States, Sweden, and Singapore. Sweden's notional defined contribution (NDC) framework and Singapore's Central Provident Fund (CPF) are examined for design features that could inform reform of the U.S. Social Security system, which faces a projected trust fund depletion in the 2030s. Drawing on the strengths of each system, the paper proposes a partially privatized reform model for the United States that combines individual retirement accounts with a notional defined contribution PAYG component, and evaluates the feasibility, trade-offs, and political context of implementing such a reform.

1. Introduction

About 68 million Americans, roughly one in five U.S. residents, receive Social Security benefits (Social Security Administration, "Fast Facts and Figures"). As a cornerstone of the U.S. retirement system, Social Security provides financial stability for millions of retirees, survivors, and disabled individuals. But with the population aging, a declining worker-to-be beneficiary ratio, and a projected trust fund depletion by the 2030s, the system's long-term sustainability is increasingly in question. America's pay-as-you-go (PAYG) model may not survive without reform, risking reduced benefits for future retirees and higher taxes for current workers.
This paper asks: how do Sweden, Singapore, and U.S. pension policies compare in long-term financial sustainability? It examines several global pension systems to identify strengths, weaknesses, and areas for improvement in the United States, with the goal of informing the future of Social Security.

2. Comparative Analysis

Because approaches to social welfare vary significantly worldwide, a comparative approach is well suited to this question. Just as Germany's welfare model has informed U.S. policy in other areas, the strengths of global pension systems can inform the United States' own approach to public benefits. Sweden's semi-privatized system and Singapore's mandatory, state-administered defined-contribution system offer contrasting models to learn from. Sweden and Singapore were chosen because both are similar to the United States in economic development — high GDP per capita, deep integration in global trade, high human development, and high levels of innovation — which lets this comparison act as a "constant," reducing the number of variables that could confound the outcome. Their aging populations and socioeconomic parallels with the U.S. further increase the likelihood that lessons from their frameworks are transferable.

2.1. Point of Comparison

This paper focuses on the Old-Age Insurance portion of the Social Security program because it is the largest component of the system and has the greatest impact on financial security. This part of the program provides retirement income to workers who have paid into the system throughout their working lives.

Edwin E. Witte, known as the "father of Social Security," explained the principle of old-age security based on social insurance in The Development of the Social Security Act: "Social insurance rests fundamentally on the principle of protection for all who contribute against the common economic hazards of life, particularly those arising from old age. It is a collective measure that provides not for dependency, but security." An effective pension system, in other words, acts as a social safety net that keeps the elderly economically stable.

This paper compares the pension programs of the U.S., Sweden, and Singapore on financial sustainability, allowing a large-scale, multi-perspective analysis of how well the U.S. system functions relative to global standards and providing a framework for identifying global best practices.

3. Background on the United States, Sweden, Singapore

This background section first summarizes the origins of social welfare and Social Security and the challenges facing wealthy economies today, then outlines the pension programs of the U.S., Sweden, and Singapore — their frameworks, goals, political context, and the demographic and economic factors shaping them. This foundation supports the comparative analysis that follows.

3.1. History

The idea of "social insurance" paved the way for public pension benefits. The concept began in 19th-century Europe as an extension of the continent's social welfare tradition, first adopted by German Chancellor Otto von Bismarck in 1889 at the urging of Emperor William I, who wrote that "those who are disabled from work by age and invalidity have a well-grounded claim to care from the state." The German system mandated participation and contributions, providing retirement and disability benefits and later sickness insurance and workers' compensation, aiming to secure a prosperous workforce.

The idea reached the United States through the Social Security Act of 1935. The Great Depression had left the country in economic distress, prompting President Franklin D. Roosevelt's New Deal, a series of recovery measures. The Social Security Act, one of its most significant components, provided financial support to the elderly, disabled, unemployed, and dependent, offering immediate relief to families and helping restart the economy.
Economic security had been a concern well before the Depression. Earlier state-level proposals, such as Louisiana Senator Huey Long's "Share Our Wealth" plan and Francis Townsend's "Old Age Revolving Pension Plan," based pensions on financial need, but restrictive eligibility made them difficult to qualify for and insufficient once received (Social Security Administration, "Historical Background").

Taking office in 1932, Roosevelt proposed economic security through social insurance rather than welfare assistance. The goal was to help the elderly provide for their own retirement amid the instability of the Depression, funded through mandatory payroll contributions. Workers would contribute until reaching the full retirement age of 67, when they became eligible for benefits.

The proposal drew favorable reception from conservatives and liberals alike: "It was also a very moderate alternative to the radical calls to action that were so common in the America of the 1930s" (Social Security Administration, "Historical Background"). Signed into law in 1935, the Social Security Act has remained crucial to Americans' well-being for nearly a century, today providing benefits to roughly 67 million U.S. residents (Center on Budget and Policy Priorities).

3.2. The Problem

America's new social insurance system changed the relationship between citizens and the federal government. Where 1930s government pensions had gone only to those who served in government, Social Security let most of the public receive payments regardless of need. Mandatory payroll contributions throughout one's working life were meant to make benefits feel personally earned. In reality, not enough was set aside in the decades after inception, so America's PAYG system funds current retirees with current workers' contributions rather than their own past contributions. Unlike a pyramid scheme, participation is legally mandated rather than dependent on recruiting new voluntary contributors, and benefits are backed by ongoing payroll tax revenue and a dedicated trust fund rather than solely by new entrants — but this PAYG structure has still made the program politically difficult to reform or scale back.

When payroll contributions began in the 1930s, revenue built up in a reserve fund with no benefits yet going out; the first scheduled payout, planned for 1942, was moved up to 1940 as Congress increased benefits. Trust Fund surpluses were invested, as required by law, in special-issue U.S. Treasury securities, which the government must redeem with interest to help pay future benefits. For decades Social Security collected more in payroll taxes than it paid out, but by 2010, with fewer workers per retiree, it began paying out more than it collected, drawing down trust fund reserves. Repaying those securities indirectly adds to the national debt and federal borrowing.

These financial problems have been worsened by rising life expectancy and slowing birth rates, leaving more retirees per worker. The payroll tax cap of $168,600 further limits revenue growth (Social Security Administration, "Maximum Taxable Earnings"), since rising incomes mean a shrinking share of total earnings is taxed. Meanwhile, benefit costs keep climbing with inflation while wages remain largely stagnant.

As a result, the Social Security Trust Fund has been contributing to the national debt: "As of December 2022, Social Security was responsible for $2.7 trillion of the US government's $6.18 trillion intragovernmental debt" (Social Security Administration). This threatens future solvency, as "reserves would be projected to become depleted in 2035 and 83 percent of

scheduled Social Security benefits would be payable at that time, declining to 73 percent by 2098" (Social Security Administration, "A Summary of the 2024 Annual Reports").

Social Security also faces other issues, including a structure that disadvantages single heads of household, fueling ongoing controversy. Spending roughly 21 percent of the national budget, or $1.4 trillion, it plays a critical role in the U.S. economy and in millions of Americans' lives, making its solvency, efficiency, and effectiveness vital to the nation's financial and social well-being. This paper explores how Sweden, Singapore, and the U.S. differ in financial sustainability, coverage, and benefits.

3.3. The United States

Social Security remains one of the most widely debated programs in the United States, providing financial security for millions of retirees, disabled individuals, and survivors, yet its future is politically polarizing. Most debate concerns long-term sustainability amid rising costs from an aging population, with proposals ranging from raising the retirement age and reducing benefits to raising payroll taxes or partially privatizing the system. The left views Social Security as crucial to reducing poverty among older Americans and favors expanding benefits while raising payroll taxes on high earners. The right views it as a financial burden requiring reform and greater personal responsibility to supplement retirement savings.

Social Security operates as a PAYG system: benefits paid to retirees, disabled individuals, and survivors are funded directly by payroll taxes on today's workforce. Workers and employers each contribute 6.2% of wages under the Federal Insurance Contributions Act (FICA), for a combined 12.4%, up to an income cap of $168,600 as of 2024 (Social Security Administration, "Maximum Taxable Earnings").

Each year of employment earns a worker up to four credits, and 40 credits (10 years of work) are required for eligibility. "One credit is granted up to $6,920 for every $1,730 earned in 2024" (Connett). Benefits can begin at age 62, but early claiming reduces monthly payments — a worker starting at 62 in 2024 would receive about 30% less than at full retirement age (66–67) — while delaying past full retirement age increases payments 8% per year up to age 70. Benefits are based on a worker's 35 highest-earning years (AIME); the average monthly benefit was $1,869.77 as of June 2024 (Connett).

Social Security is split into the Old-Age and Survivors Insurance (OASI) Trust Fund for retirees and the Disability Insurance (DI) Trust Fund for disability beneficiaries, overseen by a six-member board including the secretaries of Treasury, Labor, and Health and Human Services, the Commissioner of Social Security, and two public representatives appointed by the President (Connett). Social Security also funds Medicare, the federal health program for Americans 65 and older and eligible disabled individuals, through a third trust fund managed by the Centers for Medicare & Medicaid Services (CMS).

Other retirement programs, covered by the Employee Retirement Income Security Act (ERISA), include defined benefit plans, which promise a monthly benefit based on service and salary and are insured through the Pension Benefit Guaranty Corporation (PBGC), and defined

contribution plans such as 401(k)s and 403(b)s. "In 2022, 48% of private industry workers participated in a defined contribution plan, resulting in a take-up rate of 73%" (United States, Bureau of Labor Statistics). In these plans, "the employee or the employer (or both) contribute to the employee's individual account under the plan, sometimes at a set rate, such as 5 percent of earnings annually" (U.S. Department of Labor), with contributions invested and fluctuating in value.

3.4. Sweden

In Sweden, political debate over welfare payments centers on benefits for immigrants, who make up about 20% of the population under the country's open-border policy. Debate centers on how quickly immigrants should be expected to contribute relative to when they can receive benefits, with some arguing they should demonstrate economic contribution first; others argue immigrants strain the system since newcomers from low-income backgrounds may need healthcare, unemployment assistance, and housing subsidies while not yet integrated into the labor market. The Swedish left favors an inclusive system for all residents, while the right favors reduced benefits for immigrants.

Sweden once ran a fully pay-as-you-go system similar to the United States, but sustainability issues and changing demographics prompted reform in the 1980s and 1990s. Its current semi-privatized system balances benefits with contributions.

Sweden's Pension System is now the cornerstone of its welfare state. "It provided pension benefits for 1,036,311 men and 1,167,073 women, or roughly 22% of the total population in 2017, [and] 2,917,171 men and 2,816,981 women of working age (5,734,152 contributors in total) paid in pension contributions to the public pension system in 2017, or approximately 57% of Sweden's roughly 10.1 million population that year... In total, approximately 79% of the entire population either paid contributions or received benefits from the national pension system" (Lundberg). The framework has three main components, visualized in Figure 1 (Notas de Pensiones).

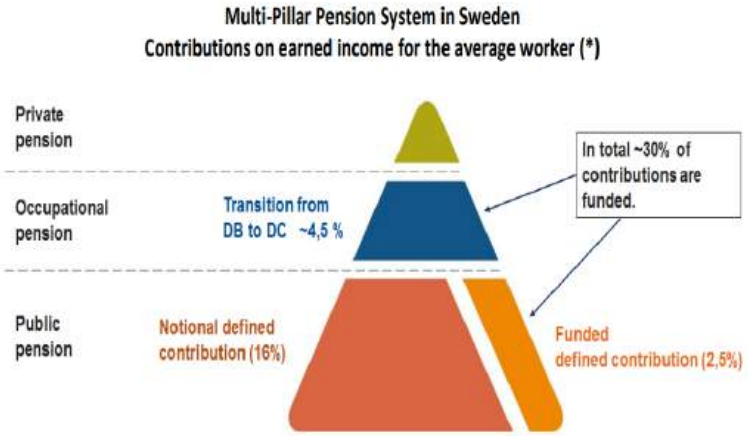


Figure 1. Sweden's Pension System Structure. Source: "Notas de Pensiones." Fiapinternacional, Mar. 2022, www.fiapinternacional.org/wp-content/uploads/2022/03/Nota-de-pensiones-61-Mar-2022-ESP.pdf.

This Public Pension System has three parts: Income Pension, Premium Pension, and Guarantee Pension. The Income Pension, a PAYG benefit based on lifetime earnings, uses current workers' contributions to fund current retirees. The Premium Pension is invested into individual accounts, with workers choosing among five fund types — Equity, Fixed Income, Mixed, Generation, and Ethical and Sustainable Funds — based on risk tolerance and goals, with returns tied to performance. Figure 2 shows 2024 returns and assets of Sweden's AP buffer funds (AP1, AP2, AP3, AP4, AP6), separate reserves that help finance the PAYG Income Pension rather than the individually invested Premium Pension (Fixsen). The Guarantee Pension provides a minimum income floor for those with little or no savings, available to anyone who has lived or worked in Sweden for at least three years.

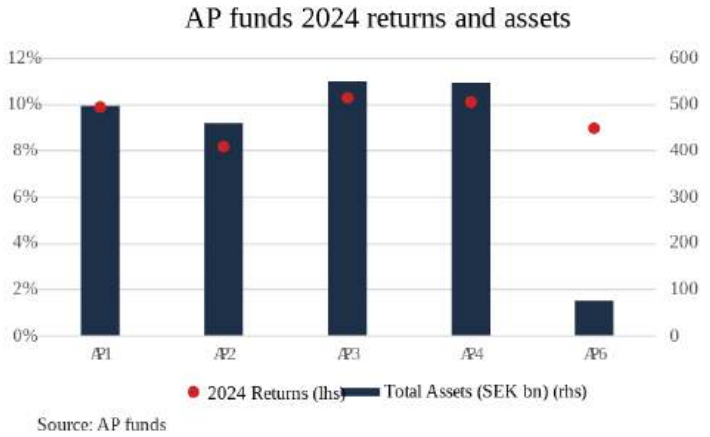


Figure 2. 2024 Returns and Assets of Sweden's AP Funds. Source: Fixsen, Rachel. "AP3 Tops Swedish Buffer 2024 Returns; AP6 Still Restrained by Forex Hedging." IPE, IPE, 24 Feb. 2025, www.ipe.com/news/ap3-tops-swedish-buffer-2024-returns-ap6-still-restrained-by-forex-hedging/10129024.article.

Unlike most countries, Sweden has no payroll tax; instead employers pay Social Security contributions equal to 31.42% of salary (reduced to 10.21% for employees 66 or older). Of an individual's pensionable income, 18.5% funds retirement: 16% to the Income Pension and 2.5% to the Premium Pension, balancing investment risk with a guaranteed safety net.
Workers can begin drawing a reduced pension as early as 63, with full benefits from 65 and increased benefits for delaying further; reduced benefits are available from 60, and workers aged 60–64 who cut their hours qualify for partial pensions. Pensionable income for the Income Pension is capped — for 2023, the income base amount was SEK 74,300, and annual income above SEK 599,250 earned no public pension.

Sweden's Social Security also provides tax-funded universal healthcare, unemployment benefits, sickness benefits, parental benefits, housing allowances, disability benefits, and social assistance.

3.5. Singapore

In Singapore, political debate over the pension system centers on whether the Central Provident Fund (CPF) is sufficient for an aging population, with discussion focused on the

adequacy of mandatory contributions, withdrawal flexibility, and coverage for low-income and disadvantaged workers. Overall, the CPF enjoys broad public support.

Many experts consider the CPF one of Asia's most developed Social Security systems for its structured approach to retirement, healthcare, and housing security. Established in 1955, it requires mandatory contributions from both employee and employer into individual accounts, at rates of 37% of monthly wages above S$50 for employees 55 or younger, 31% for ages 55–60, 22% for ages 60–65, 16.5% for ages 65–70, and 12.5% for 70 and older (Central Provident Fund Board, "How Much CPF").

The CPF includes three accounts — Ordinary (OA), Special (SA), and MediSave (MA) — shown in Figure 3 (Koh et al.). The OA funds major purchases like housing, education, and investments; the SA supports post-retirement financial needs; and the MA covers health insurance premiums, including MediShield Life and ElderShield/CareShield Life. Allocation shifts with age: "During your younger years," more contributions go to the OA to support a home purchase; "as you grow older, more CPF contributions are allocated to your Special Account (SA) and MediSave Account (MA) to meet growing retirement and healthcare needs" (Central Provident Fund Board, "How Much CPF").

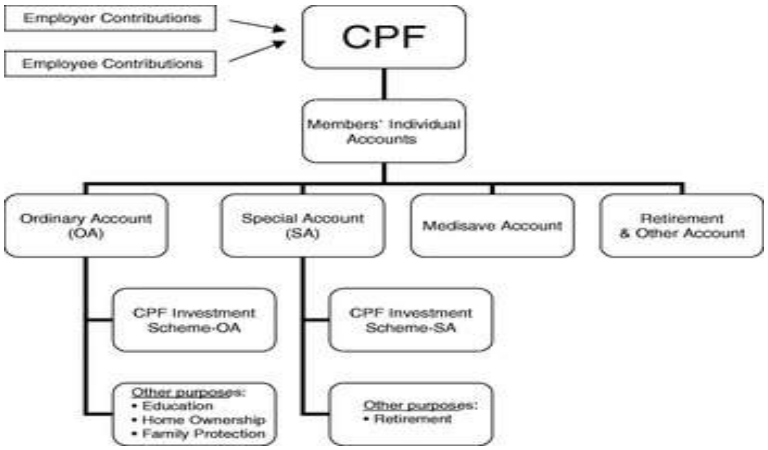


Figure 3. Singapore's CPF Structure. Source: KOH, BENEDICT S. K. et al. "Investment Patterns in Singapore's Central Provident Fund System." Journal of Pension Economics and Finance 7.1 (2008): 37--65. Web.

These accounts are managed by the CPF Board, overseen by the Ministry of Manpower. CPF savings are transferred to the government, which issues Special Singapore Government Securities to the CPF Board — bonds paying guaranteed interest above typical bank rates while carrying less risk than market investing. The funds raised are managed as part of Singapore's foreign reserves by the Government of Singapore Investment Corporation (GIC), one of the world's largest sovereign wealth funds, which invests globally across equities, real estate, bonds, and other assets. CPF members are shielded from this investment risk and guaranteed a predetermined return: "The interest rate for Special, MediSave and Retirement Account is 4.08% per annum from 1 July 2024 to 30 September 2024" (Central Provident Fund Board, "How Much CPF"), the same for all members, with extra interest on the first $60,000 of combined balances

(capped at $20,000 from the OA) credited to the Retirement Account. This structure secures members' savings while letting the government invest efficiently for long-term growth.

At 55, the OA and SA combine into a Retirement Account (RA) used for CPF LIFE, a lifelong annuity providing monthly retirement payouts. Members can start benefits between 65 and 70 with five years of vesting service, or as early as 55 with at least 10 years of vesting service (at a 3% annual reduction before age 65). Special Retirement Benefits are available at 62 with 25 years of service (Central Provident Fund Board, "What Is the CPF Retirement Sum").

4. Comparative Methodology

To ensure a rigorous comparison, this paper evaluates each pension system's financial sustainability using five criteria: (1) solvency — whether projected revenues cover projected benefit obligations over a defined horizon; (2) demographic resilience — capacity to absorb changes in fertility, mortality, and the worker-to-beneficiary ratio; (3) benefit adequacy — whether payouts sustain a basic standard of living in retirement; (4) administrative costs — the share of spending consumed by operating the system rather than paying benefits; and (5) distributional effects — how costs and benefits are shared across income groups and generations. Applying these criteria consistently across the U.S., Sweden, and Singapore allows a structured rather than impressionistic comparison.

Sweden and Singapore differ from the United States in population size, labor-market structure, tax systems, and welfare-state design — Sweden's roughly 10 million and Singapore's roughly 6 million are far smaller than the U.S.'s 330 million, and both maintain more centralized policymaking than the U.S. federal system. Still, policy transfer remains appropriate because this paper draws on specific design mechanisms — automatic balancing formulas, notional accounts, mandatory individual savings — rather than transplanting entire welfare-state models. These modular mechanisms have already been adapted across varied institutional contexts, suggesting they can inform incremental U.S. reform despite differences in surrounding architecture.

5. Main Findings

A comparative analysis of the Social Security systems in the United States, Sweden, and Singapore reveals significant differences in structure and financial sustainability. Sweden's pension system combines a notional defined contribution (NDC) Income Pension — a PAYG scheme crediting contributions to notional individual accounts — with a smaller, fully funded Premium Pension (2.5% of pensionable income invested in funds of the individual's choosing), together supporting long-term fiscal stability. Singapore's CPF relies on mandatory savings and government-managed investments. The U.S. system, by contrast, operates on a PAYG model facing solvency challenges as benefit obligations grow amid an aging population. These findings point to reforms combining successful strategies from Sweden and Singapore: a semi-privatized system allocating part of payroll taxes to individual retirement accounts and the rest to NDC accounts operating as PAYG, improving U.S. sustainability while preserving the system's core competencies.

This paper first surveys the history of social welfare and the pension systems of the U.S., Sweden, and Singapore, including their political context; then compares the three systems' approaches to long-term financial sustainability; and finally proposes policy reforms for the United States that account for the current political landscape.

5.1. Financial Sustainability

Most U.S. policy debate on Social Security concerns its financial sustainability, with most policymakers viewing it as unsustainable long term — a concern shared by Sweden and Singapore amid similar demographic and economic pressures. This section compares financial sustainability across the three countries using the five criteria from the Comparative Methodology: solvency, demographic resilience, benefit adequacy, administrative costs, and distributional effects.

It is worth defining exactly what "financial sustainability" means here. According to Stephen C. Goss, there are two ways to analyze it:

"The first considers only the simple question of whether currently scheduled dedicated tax revenue is sufficient to adequately finance currently scheduled benefits in the law, without any modification to the law. The second considers whether the current structure of the program, with a defined benefit reflecting career-average earnings levels and a flat payroll tax up to a specified earnings level, is viable for the future" (Goss).

This section addresses the second question, evaluating each system's long-term structural viability amid an aging population, lower birth rates, and longer life expectancies.

5.1.1. United States

The U.S. The Social Security System is the primary source of retirement income for millions of Americans, acting as a safety net for retirees, disabled individuals, and the spouses of deceased workers. Yet it remains a topic of deep political polarization. Republicans tend to favor a more market-based or privatized approach giving individuals greater personal responsibility over their retirement savings, while Democrats favor a government-reliant system. Progressives generally propose lifting the payroll tax cap so higher earners pay more into the system, while conservatives favor gradually raising the retirement age and privatizing Social Security through personal accounts similar to 401(k)s.

Demographic Resilience

Despite this divide, both parties largely agree the system in its current form is not financially sustainable. Its PAYG model means current workers fund current retirees' benefits, a problem as the population ages: in 1960 there were 5.1 workers per beneficiary; by 2024 that had dropped to about 2.8 (Peter G. Peterson Foundation). This imbalance will worsen as the Baby Boomer generation retires — "about 12,000 people will turn 65 every day in the next year... by 2030, all baby boomers will be 65 or older" (Veradigm) — while rising life expectancy, from

about 68 years in the 1950s to 79 in 2024 (Harvard Health), means benefits must be paid out longer, further straining the fund.

Solvency

Social Security is funded primarily through payroll taxes: a 6.2% FICA tax on employee wages matched by employers (12.4% total), or 12.4% for the self-employed. As of 2024, income above $168,600 is exempt from this tax (Social Security Administration, "Maximum Taxable Earnings"), limiting revenue growth.

Past FICA surpluses were pooled into a trust fund now used when tax revenue falls short. Per the 2024 Trustees' report, "The Old-Age and Survivors Insurance (OASI) Trust Fund will be able to pay 100 percent of total scheduled benefits until 2033... At that time, the fund's reserves will become depleted and continuing program income will be sufficient to pay 79 percent of scheduled benefits" (Social Security Administration, "A Summary of the 2024 Annual Reports"). Economic pressures such as wage stagnation and declining labor force participation compound the problem. Per the Economic Policy Institute, "Since the late 1970s, wages for the bottom 70 percent of earners have been essentially stagnant, and between 2009 and 2013, real wages fell for the entire bottom 90 percent of the wage distribution" (Mishel), driven by financial deregulation, declining union density, labor policy changes, and high unemployment — leaving these workers contributing less in payroll taxes than they otherwise could.

Benefit Adequacy

Still, the system has real strengths. Cost-of-Living Adjustments (COLAs) tie benefits to the Consumer Price Index for Urban Wage Earners and Clerical Workers, raising benefits when the index rises year-over-year without ever reducing them if inflation is flat or negative, helping preserve purchasing power.

Administrative Costs

The trust funds also earn interest on their U.S. Treasury holdings — a modest but real supplement to FICA revenue — and the SSA operates efficiently: administrative costs are only about 0.5% of annual benefits, well below privatized or semi-privatized systems elsewhere (Center on Budget and Policy Priorities), maximizing funds available for benefits.

Distributional Effects

The system is also progressive, replacing a higher share of prior earnings for low-wage workers than high earners: "Benefits for a low earner (with 45 percent of the average wage) retiring at age 65 in 2024 provide $15,477 a year, replacing about half of their prior earnings. But benefits for a high earner (with 160 percent of the average wage) provide $33,769, replacing about one third of prior earnings, though they are larger in dollar terms than those for the low-wage worker" (Center on Budget and Policy Priorities). Low earners below a $18,765

income threshold in 2024 also receive a Special Minimum Benefit (Social Security Administration, "Social Security Benefit Tables"), helping ensure a baseline of support.

Coverage is broad, extending to about 96 percent of U.S. workers (United States, Government Accountability Office), including retirees, disabled workers, dependents of deceased or disabled workers, legally working non-citizens, government employees, military personnel, and the self-employed. This breadth is generally a strength, though it strains the system somewhat, since roughly one-fourth of public employees are historically exempt from Social Security taxes (United States, Government Accountability Office), offsetting some contributions against benefits paid. On balance, while the system is not sustainable long-term without reform given demographic shifts, a depleting trust fund, and economic pressures, its COLA system, trust fund interest, low administrative costs, and broad coverage remain real strengths.

5.1.2. Sweden

Solvency

Sweden's public finances are among the strongest in the Eurozone: its debt-to-GDP ratio has stayed below the Eurozone average since 1995 and stood at 31.20% in 2023 (Trading Economics), far below the Eurozone average of 88.6% (European Commission) — due in part to its robust pension program. "The returns from the normal income pension is around 2% per year, but from the private accounts the average Sweden has made an impressive average return of roughly 10% a year since its inception in 1995, despite the dot-com crash, the financial crisis and the pandemic" (Norberg). This resilience through economic instability stems from the system's flexible design, built around a notional defined contribution (NDC) structure.

This semi-privatized NDC system links an individual's Premium Pension benefits to lifetime contributions. The Income Pension follows a PAYG model similar to the U.S., where this year's contributions pay this year's retirees — leaving its long-term sustainability dependent on demographics and life expectancy. The Premium Pension offsets this: 2.5% of pensionable income goes annually into an individually chosen investment fund, with returns paid out as benefits after retirement.

Benefits are also indexed to wage growth so the government doesn't over-promise when growth slows; the Premium Pension account's return "is determined by the change in the value of the premium pension funds chosen by the insured" (Pensionsmyndigheten). A life expectancy divisor further adjusts payouts so lifetime benefits stay proportional to contributions as life expectancy rises, and an Automatic Balancing Mechanism, introduced in 2001, cuts future pensions if liabilities exceed assets. Together these "automatic stabilizers" have helped push pension fund assets to 98.7% of Sweden's GDP (Statista Research Department), while social expenditure stays around 25–30% of GDP.

Demographic Resilience

Despite this, Sweden faces real demographic pressure: "Of Sweden's 10.5 million inhabitants, 20 per cent have passed the standard retirement age of 65. This number is projected to rise to 23 per cent by 2040" (Nordea), straining the Income Pension as fewer workers support more retirees. Life expectancy compounds this — "among the highest in the world: 81.58 years for men and close to 84.9 years for women" (Nordea) — meaning longer payout periods. In response, Sweden raised the earliest pension age: "The first of January 2020, the earliest age to receive an old age pension was increased by one year for both women and men, from 61 years to 62 years" (Regeringskansliet), easing financial strain.

Benefit Adequacy

Other challenges include volatility in Premium Pension investments, since market exposure can lower returns and benefits when choices underperform; Sweden has responded by narrowing the fund menu and improving oversight to steer savers toward low-cost, reliable options. The Automatic Balancing Mechanism (ABM) has its own tradeoff — cutting benefits when liabilities exceed assets can hit lower-income retirees hardest, lowering living standards and public confidence — though Sweden continues refining it to smooth reductions and limit hardship.

Buffer fund (AP Fund) sustainability is a further concern; these funds cushion economic downturns by smoothing fluctuations in contributions and payouts, but a prolonged downturn or period of low returns could still strain the system. Overall, Sweden's financial sustainability is strong and projected to remain stable, supported by automatic adjustment mechanisms, ongoing balancing of retirement age against life expectancy, and the Premium Pension program, despite its reliance on economic performance and market volatility.

Administrative Costs

Comparable administrative-cost data for Sweden's pension system was not available in the sources consulted for this paper, but the system's multi-tiered structure — combining a centrally administered PAYG Income Pension with individually managed Premium Pension accounts — introduces more operational complexity than Singapore's single-agency CPF or the United States' single PAYG program, suggesting administrative costs are plausibly higher, though further data would be needed to confirm this.

Distributional Effects

Sweden's Guarantee Pension provides a redistributive floor for those with little or no lifetime earnings, funded through general taxation rather than payroll contributions, offsetting some of the income-linked design of the Income and Premium Pensions. As discussed earlier, though, extending this guarantee to recently arrived immigrants who haven't yet contributed substantially remains a politically contentious distributional question.

5.1.3. Singapore

Solvency

Singapore's pension system is widely considered financially sustainable due to its fully funded, defined-contribution design, in which individuals save for their own retirement rather than relying on others' contributions, through the Central Provident Fund (CPF). The CPF "serves to meet our retirement, housing and healthcare needs" (Ministry of Manpower), and "as of 2023, CPF managed around US$571 billion — 143% of GDP" (Bhagwat). It comprises four accounts — Ordinary (OA), Special (SA), MediSave (MA), and Retirement (RA) — funded through mandatory contributions: for employees 55 and under, employers contribute 17% and employees 20% of wages, totaling 37% of monthly payroll (Ministry of Manpower). Funds in the OA and SA can be invested through the CPF Investment Scheme, in which the government channels CPF contributions into non-tradable, government-guaranteed Special Singapore Government Securities (SSGS), which are then invested globally in bonds, infrastructure, real estate, and other assets by the Government of Singapore Investment Corporation (GIC).

Demographic Resilience

The CPF's strength comes from crediting individual and employer contributions to individual accounts — the opposite of the West's PAYG model — which minimizes reliance on future workers to fund current retirees and largely insulates the system from demographic pressures like aging populations or shrinking workforces. Because funds cannot be transferred between accounts, individuals bear responsibility for their own savings, reducing the government's exposure to financing debt, future tax increases, or benefit cuts to meet pension obligations.

Benefit Adequacy

GIC's investments are generally low-risk given their highly diversified portfolio. Its stated risk management objectives are to ensure "the investment strategies are consistent with the risk tolerance set by our client, the Singapore Government"; that "policies, guidelines, and control processes are in place to reduce the likelihood of significant losses"; that "any reputational impact due to our actions is carefully managed"; and that "the risks associated with each investment are well-understood, and that we are adequately compensated for taking those risks" (International Forum of Sovereign Wealth Funds).

GIC's risk management objectives are to ensure:

1. The investment strategies are consistent with the risk tolerance set by our client, the Singapore Government;
2. Policies, guidelines, and control processes are in place to reduce the likelihood of significant losses;
3. Any reputational impact due to our actions is carefully managed;

4. The risks associated with each investment are well-understood, and that we are adequately compensated for taking those risks" (International Forum of Sovereign Wealth Funds).

These objectives prioritize long-term returns over short-term gains, reflecting GIC's mandate to manage CPF assets cautiously. This conservative strategy lets the fund guarantee returns even during downturns, and on average makes GIC's investments less risky than the fund options available to individuals in Sweden's semi-privatized system.

Members also receive guaranteed interest on each account — 2.5% for the OA and 4% for the SA, MA, and RA (Central Provident Fund Board, "What Is the CPF Retirement Sum") — ensuring adequate retirement benefits even when investments underperform, and protecting savings from inflation. The CPF's multiple pillars, covering retirement, healthcare, housing, and insurance, make it a holistic safety net, with separate accounts letting individuals meet different needs without drawing down their Retirement Account.

Distributional Effects

Singapore also supports low-income earners through supplements and top-ups that boost CPF balances for low-wage workers and older citizens with inadequate RA savings: "The threshold for low-income families ranges from S$500 to S$3,000 in terms of household income per capita" (Chan). Schemes like the Workfare Income Supplement (WIS) and Silver Support Scheme, funded by GIC's investment surplus, extend coverage to everyone regardless of income while preserving sustainability.

Administrative Costs

Even so, the CPF faces challenges: whether Retirement Account funds are adequate given inflation, system complexity, and investment restrictions. Many members fall short of the Minimum Sum needed for basic needs — "In 2011, only 45 percent of those 33,644 active members who turned 55 in 2011 had sufficient cash balances to meet the Minimum Sum requirement of S$112,000 (in 2003 prices)" (Asher) — a gap that rising living costs and stagnant wages can widen. Guaranteed interest rates can also be outpaced by inflation, which peaked at 5.5% in August 2024 (Kok), potentially requiring future rate adjustments. The system's complexity — multiple accounts with different rules for withdrawals, contributions, and investment — can make it hard for members to fully use what CPF offers, and GIC's conservative investment approach limits returns for members who might accept more risk for more reward, since individual investment options remain restricted.

The Singaporean government continues adjusting the system — changing contribution rates for older workers, raising the retirement age, and revising the Minimum Sum — to keep it financially strong. Overall, the CPF is robust thanks to its fully funded design, low-risk GIC investments, guaranteed interest, multi-account structure, and low-income support schemes, though adequacy, inflation, complexity, and investment restrictions remain ongoing concerns.

6. Policy Reform

Having examined the pension programs of the United States, Sweden, and Singapore, it is clear each system has its own advantages and drawbacks. A comparative approach makes it possible to identify what works well in each and use it to build a reform plan that improves the financial sustainability of America's Social Security program — integrating strengths from Sweden's Notional Defined Contribution system and Singapore's mandatory savings and investment framework into a comprehensive reform addressing U.S. sustainability challenges.

6.1. Reform Plan

First, the United States can adopt both a Notional Defined Contribution (NDC) system like Sweden's and a Mandatory Funded Component like Singapore's, shown in Figure 4. An NDC system combines traditional pay-as-you-go (PAYG) principles with a defined-contribution structure, recording worker contributions in individual notional accounts. This would require transitioning from a complete PAYG model to a partially funded one, where each worker's 6.2% payroll tax contribution is tracked in a notional account that earns a notional interest rate tied to economic factors like wage growth. Benefits would then be calculated from total lifetime contributions and notional interest, with the system continually adjusting for life expectancy. While still PAYG at its core, an NDC model offers greater transparency, since payouts are directly linked to lifetime contributions, minimal political risk from limited government intervention, and a self-balancing structure that automatically adjusts for wage growth, rising life expectancy, inflation, and labor force participation.

Second, the employer's 6.2% Social Security contribution would fund individual retirement accounts, similar to Singapore's CPF, invested by a government-appointed board within the Treasury Department in a diversified mix of low-risk stocks and index funds. This lets individuals build personal retirement savings while preserving the PAYG safety net, increasing their ownership over their retirement funds and potentially generating higher returns than the current system with low volatility, while reducing reliance on future taxpayers to fund current retirees.

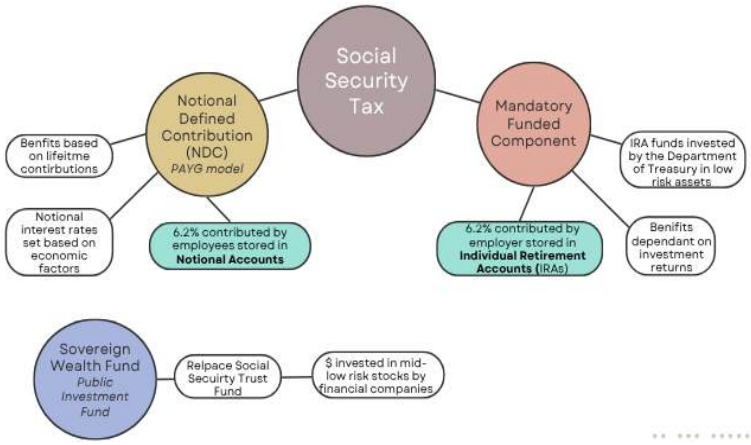


Figure 4. Structure of Proposed Social Security Reform (Hybrid Component). Source: Diagram created by the author

Contribution rates should also rise in two ways: incremental 0.5% increases to the payroll tax rate until the system is adequately funded, and removing the taxable earnings cap ($168,600 in 2024) for the PAYG portion so higher earners contribute proportionally, raising revenue without disproportionately burdening lower earners. This mirrors Sweden, which has no cap on contributions to its income-based pension (though its Premium Pension does have one) — a model the U.S. could adapt by keeping a contribution ceiling for individual retirement accounts while removing it for the PAYG portion.

As shown in Figure 5, retirement age should also be automatically indexed to life expectancy. Life expectancy is currently 79.25 years (Macrotrends), while the U.S. retirement age has remained 66 since it was last raised in 1983 (Social Security Administration, "Historical Background") — a growing gap that strains the system as retirees draw benefits longer. Automatic indexation would keep the system balanced, while early retirement would remain available at reduced benefits, similar to Sweden's gradual, demographically-tied adjustments. Workers in physically demanding occupations should be exempt, retaining full benefits at 66. As the retirement age rises, measures should encourage greater workforce participation among older workers, such as tax credits for employers who hire or retain them and added benefits for older workers who keep working. This would delay retirement, increasing NDC/PAYG contributions while reducing near-term benefit payouts and boosting individual retirement account contributions.

Benefit structure reform should also reduce the replacement rate for high earners. "In 2024, the Social Security replacement rate for high earners in the US is 35.5%" (Social Security Administration, "Social Security Benefit Tables"), compared to 39% for the average earner retiring at 65; lowering the high-earner rate to around 30% would help fund the system without disadvantaging lower earners. Cost-of-living adjustments (COLA) should be preserved, and to protect middle-income earners, the replacement-rate change would apply only to the top bracket, those earning $626,350 or more (Durante).

Finally, a Sovereign Wealth Fund (SWF) should be established from a portion of the existing Social Security Trust Fund, operating separately from individual retirement accounts under professional management. Investing more assertively than the individual accounts — across equity, infrastructure, and other assets — the SWF would aim to outperform current low-yield treasury holdings, gradually reducing reliance on payroll taxes and adding security to the PAYG portion.

This approach is proven feasible by Singapore's own sovereign wealth fund, the Government of Singapore Investment Corporation (GIC), a government-owned entity that reported a 20-year annualized real return of 3.9% as of March 2024 (GIC), demonstrating its contribution to national financial sustainability — a model the U.S. could adapt to support the PAYG portion of Social Security.

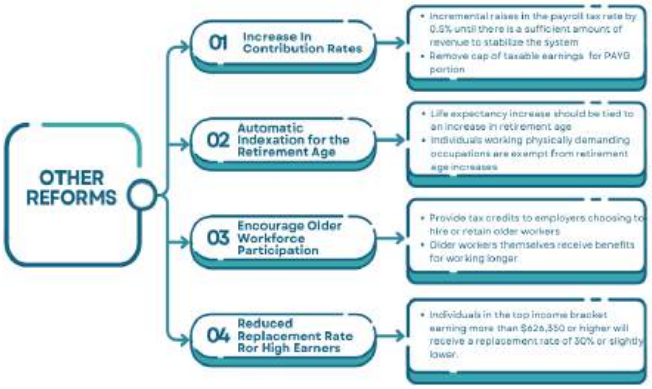


Figure 5. Structure of Proposed Social Security Reform (Other Component). Source: Diagram created by the author

6.2. Past Privatization Attempts and Lessons for Reform

Past attempts to privatize Social Security under Presidents George H. W. Bush and Bill Clinton offer relevant lessons. Bush believed Americans should be able to invest their FICA contributions themselves, declaring, "We should trust Americans by giving them the option of investing part of their Social Security contributions in private accounts" (Galston). His plan, detailed alongside the February 2005 State of the Union address, would have let workers divert a third of their payroll tax (up to 4% of earnings, capped at $1,000) into private accounts invested in five index funds, shifting automatically into lifecycle funds as workers aged and converting mostly to lifetime annuities at retirement (Galston). It failed for three main reasons: public skepticism, since most people saw Social Security's long-term solvency as a problem but not an urgent one, leaving Bush — already at roughly 50% approval — without enough support after launching the plan with little congressional consultation and winning over few Democrats, mainly libertarians and economic conservatives rather than social conservatives; unresolved transition costs, since even a partial shift from PAYG to advance-funded personal accounts was estimated to cost a trillion dollars (Galston); and a lack of bipartisan support.

This reform plan learns from that failure by adopting a hybrid model that keeps a PAYG structure while introducing a government-managed Notional Defined Contribution (NDC) system and Individual Retirement Accounts (IRAs), a less drastic shift that should improve public acceptance. Because the government still manages the guaranteed NDC benefits, it addresses public concerns about security while increasing transparency and limiting retirees' market risk exposure.

6.3. The Transition

Although this plan differs substantially from President Bush's proposal, some of the same risks remain, including market exposure. Even though individual retirement account and Sovereign Wealth Fund investments are relatively low-risk, downturns could still affect retirement savings — particularly for individual retirement accounts, which directly determine benefits and could leave retirees with less if markets are down at retirement, though markets have historically recovered over time as consistent contributions continue. The transition would

also raise administrative costs, since managing notional accounts, individual retirement accounts, and the SWF requires additional capital.

Most significantly, transitioning from a purely PAYG system to a hybrid model may strain finances in the short term, since current retirees still need to be paid while a portion of the tax that once funded them — 6.2% from the employer — shifts to individual retirement accounts. The remaining 6.2% funding the NDC PAYG portion would not be sufficient alone, creating temporary strain and requiring transitional measures such as phased benefit adjustments for older workers who can't fully benefit from the new system.

A phased, five-year implementation plan can ease this strain. Phase one focuses on passing legislation to create the NDC system, IRAs, and SWF, developing an investment strategy and regulatory framework for managing IRA and SWF funds, and beginning the gradual, demographically-indexed increase in retirement age. Phase two tracks worker contributions to notional accounts while continuing PAYG payouts, gradually shifting the employer's 6.2% contribution into IRAs.

The PAYG taxable earnings cap would then phase out gradually, rising $25,000 per year until fully removed, followed by tax credits for employers who hire or retain older workers and establishment of the SWF, seeded with a portion of the Trust Fund invested across various assets. Phase three completes the transition to the hybrid NDC model, with new retirees receiving benefits under the new structure, IRAs and the SWF fully capitalized and generating adequate returns, and the high-earner replacement rate gradually reduced while COLA adjustments continue. Phase four evaluates the reforms' impact, monitoring IRA and SWF returns to confirm the system is self-sustaining. This phased approach, illustrated in Figure 6, balances stability with reform.

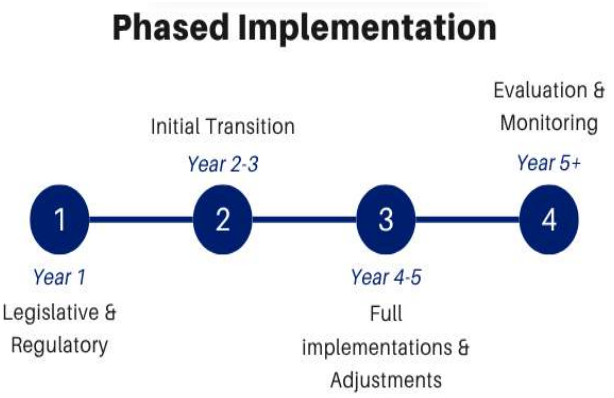


Figure 6. Phased Implementation Plan for Social Security Reform. Source: Diagram created by the author

6.4. Limitations & Advantages

This reform plan offers a comprehensive path to improving financial sustainability, transparency, and individual ownership, but it carries real limitations. These include the short-term financial strain from removing the taxable earnings cap and redirecting 6.2% of payroll taxes to IRAs; continued exposure of low-risk stock and index fund investments to downturns, which could reduce IRA benefits during those periods; higher implementation and

administrative costs as the Treasury and Social Security Administration expand fund management roles for the NDC accounts, IRAs, and SWF; and limited benefit for older workers nearing retirement, who would face the retirement age increase without enough time to build meaningful IRA savings.

Still, the advantages outweigh these limitations. The NDC system is self-balancing, adjusting benefits to economic conditions and life expectancy, while tying retirement age to life expectancy keeps the system solvent as longevity rises — without abolishing physically demanding occupations, which remain exempt. IRA investments in low-risk stocks and index funds should generate higher returns than the current purely PAYG system, reducing long-term reliance on payroll taxes and letting workers build personal savings alongside Social Security benefits, with diversification and dollar-cost averaging limiting risk even during downturns. The SWF, investing more aggressively than the IRAs or NDC system, adds a further funding source beyond the traditional Trust Fund, while tax credits for employers who hire and retain older workers should increase older workforce participation and ease financial strain.

6.5. Public Perception

Certain parts of this reform plan are likely to be more popular than others with the public and Congress — worth considering given what it would take to get this reform approved. Individual retirement accounts should be popular with the public, since most Americans value personal ownership and control over retirement savings, and the NDC system's transparency, linking lifetime contributions directly to benefits, should reduce uncertainty about solvency and build confidence. Tax credits for employers and older workers should also be popular for incentivizing continued workforce participation.

In Congress, the hybrid approach — a partial PAYG system alongside a privatized portion — offers a reasonable, bipartisan compromise between Democratic and Republican priorities. The Sovereign Wealth Fund should appeal to those concerned about the deficit and long-term solvency, and tying retirement age to life expectancy should be popular for reducing long-term benefit costs.

Eliminating the taxable income cap and reducing PAYG replacement rates, however, would likely be unpopular with high earners, whose tax burden stays the same while benefits fall, and with those who see Social Security as an earned benefit rather than welfare. Eliminating the PAYG earnings cap may also face broader opposition from Republicans and moderate Democrats wary of raising taxes on high earners, while government-managed investment could concern conservatives worried about mismanagement or political interference, making the Sovereign Wealth Fund a harder sell.

In short, individual retirement accounts, the NDC system, COLA adjustments, and hiring tax credits are the portions most likely to pass, while the Sovereign Wealth Fund, retirement age indexation, eliminating the PAYG tax cap, and reducing high-earner replacement rates are more controversial.

6.6. Political Climate

In today's polarized political and economic climate, public skepticism of privatization runs high. With Social Security supporting 71.6 million Americans and remaining one of the most popular government programs, a plan resembling "privatization" would likely face strong Democratic opposition (Social Security Administration). Republicans, accordingly, have shifted from Bush-era privatization proposals toward alternatives like raising the retirement age or cutting costs to improve solvency.

To address these concerns and distinguish this plan from past privatization attempts, its security features can be emphasized: individual retirement accounts can be framed around "government-guaranteed" minimum returns protecting retirees from market volatility, and the Sovereign Wealth Fund as managed by a public firm mandated to prioritize long-term stability over short-term gains.

Low confidence in government institutions and financial markets also fuels resistance to privatization. The 2008 financial crisis showed the public how damaging a downturn can be to retirement accounts, and recent market volatility has reinforced skepticism about exposing retirement savings to market risk — reflected in the Conference Board Consumer Confidence Index, which "declined by 5.4 points in January to 104.1 (1985=100)" (The Conference Board). Still, with 79% of U.S. adults wanting action on Social Security's solvency as of 2024, there is likely real openness to reform ideas (National Institute on Retirement Security). Because this plan uses government-controlled investment rather than shifting all risk to individuals, and combines a hybrid model with a PAYG safety net rather than full privatization, it should avoid the criticisms that doomed Bush's proposal while preserving the current system's familiarity. Together with a more risk-tolerant new generation of Americans (Millennials and Gen Z), these bipartisan features give this plan a real chance at a different outcome than past privatization attempts.

7. Conclusion

The comparative analysis of Social Security systems in the United States, Sweden, and Singapore reveals different paths to financial sustainability. Sweden's notional defined contribution system and Singapore's Central Provident Fund both balance state oversight with individual responsibility to secure long-term solvency. Drawing on these models, this paper proposes a reform plan for the U.S. system using a partially privatized model with individual retirement accounts alongside a notional defined contribution PAYG portion as a safety net. Though the transition would take time, it offers a possible answer to the long-standing concern of a depleting Social Security Trust Fund.

President Trump has pledged to protect Social Security benefits while opposing cuts, underscoring how central these benefits remain to millions of American retirees. Current political discourse continues to wrestle with reforming the system to ensure sustainability without sacrificing benefits amid rising costs, falling revenues, and inflation — reflecting the broader challenge of balancing fiscal responsibility with social welfare.

Works Cited

"ASEAN Briefing. Social Insurance in Singapore." ASEAN Briefing, www.aseanbriefing.com/doing-business-guide/singapore/human-resources-and-payroll/social-insurance. Accessed 23 Feb. 2025.

"Governance of Migrant Integration in Sweden." European Website on Integration, June 2024, migrant-integration.ec.europa.eu/country-governance/governance/sweden_en.

"Latest Monthly CPF Salary Ceiling, Allocation and Contribution Rates: 2025." Endowus, 14 Feb. 2025, endowus.com/insights/cpf-allocation-rates-how-it-changes-as-you-grow-older.

"More than Three in Four U.S. Adults Believe the Social Security System Needs to Change." Nationwide Mutual Insurance, Nationwide Mutual Insurance, 31 July 2024, news.nationwide.com/adults-believe-social-security-system-needs-to-change/.

"Notas de Pensiones." Fiapinternacional, Mar. 2022, www.fiapinternacional.org/wp-content/uploads/2022/03/Nota-de-pensiones-61-Mar-2022-ESP.pdf.

"Payroll Taxes in Sweden [2025 Guide]." Horizons, 22 June 2024, joinhorizons.com/countries/sweden/hiring-employees/employee-taxes/.

"Policy Basics: Top Ten Facts about Social Security." Center on Budget and Policy Priorities, 31 May 2024, www.cbpp.org/research/social-security/top-ten-facts-about-social-security.

"Policy Basics: Understanding the Social Security Trust Funds." Center on Budget and Policy Priorities, 24 July 2024, www.cbpp.org/research/social-security/understanding-the-social-security-trust-funds-0.

"Report Highlights." GIC, 2024, report.gic.com.sg/.

"Speechwriting, White House Office of: Speech Drafts Box." Reagan Library, www.reaganlibrary.gov/public/digitallibrary/smof/speechwriting-speechdrafts/box-002/40-534-5709750-002-001-2018.pdf. Accessed 23 Feb. 2025.

"The Swedish Social Security System." Business Sweden, 2021, www.business-sweden.com/contentassets/41249a64a04e4309b7637b00c526eef8/the-swedish-social-security-system-2021-publ.pdf.

"Types of Retirement Plans." DOL, www.dol.gov/general/topic/retirement/typesofplans. Accessed 23 Feb. 2025.

Asher, Mukul G. "Chapter 12. Singapore's Pension System: Challenges and Reform Options." Equitable and Sustainable Pensions: Challenges and Experience, International Monetary Fund, 8 Mar. 2014, www.elibrary.imf.org/display/book/9781616359508/ch012.xml.

Bhagwat. [Author's original source not located -- verify before submission.]

Central Provident Fund Board. "How Much CPF Contributions to Pay." CPFB, 2 Jan. 2025, www.cpf.gov.sg/employer/employer-obligations/how-much-cpf-contributions-to-pay.

Central Provident Fund Board. "What Is the CPF Retirement Sum?" CPFB, 21 Feb. 2024, www.cpf.gov.sg/member/infohub/educational-resources/what-is-the-cpf-retirement-sum.

Chan. [Author's original source not located -- verify before submission.]

Connett, Wendy. "Social Security Explained: How It Works and Types of Benefits." Investopedia, Investopedia, 15 Aug. 2024, www.investopedia.com/terms/s/socialsecurity.asp#toc-the-future-of-social-security.

Durante, Alex. "2025 Tax Brackets." Tax Foundation, 2 Oct. 2022, taxfoundation.org/data/all/federal/2025-tax-brackets/.

European Commission. [Eurozone government debt-to-GDP average] [Author's original source not located -- verify before submission.]

European Commission. [Sweden pensionable income allocation rates] [Author's original source not located -- verify before submission.]

Fixsen, Rachel. "AP3 Tops Swedish Buffer 2024 Returns; AP6 Still Restrained by Forex Hedging." IPE, IPE, 24 Feb. 2025, www.ipe.com/news/ap3-tops-swedish-buffer-2024-returns-ap6-still-restrained-by-forex-hedging/10129024.article.

Galston, William A. "Why the 2005 Social Security Initiative Failed, and What It Means for the Future." Brookings, 21 Sept. 2007, www.brookings.edu/articles/why-the-2005-social-security-initiative-failed-and-what-it-means-for-the-future/.

Goss. [Author's original source not located -- verify before submission.]

Harvard Health. [U.S. life expectancy figure] [Author's original source not located -- verify before submission.]

International Forum of Sovereign Wealth Funds. [Author's original source not located -- verify before submission.]

Koh, Benedict S. K., et al. "Investment Patterns in Singapore's Central Provident Fund System." Journal of Pension Economics & Finance, Cambridge University Press, 16 Nov. 2007, www.cambridge.org/core/journals/journal-of-pension-economics-and-finance/article/abs/investment-patterns-in-singapores-central-provident-fund-system/B36C7333BDA01525575E4A872B092FF9.

Kok. [Author's original source not located -- verify before submission.]

Lundberg, Kristoffer. "Retirement: Lessons from the Swedish Reforms." Fondapol, 4 Jan. 2020, www.fondapol.org/en/study/retirement-lessons-from-the-swedish-reforms/.

Macrotrends. "U.S. Life Expectancy 1950-2025." Macrotrends, www.macrotrends.net/global-metrics/countries/usa/united-states/life-expectancy. Accessed 2025.

Ministry of Manpower. [Singapore] [Author's original source not located -- verify before submission.]

Mishel, Lawrence. "Causes of Wage Stagnation." Economic Policy Institute, 6 Jan. 2015, www.epi.org/publication/causes-of-wage-stagnation/.

National Institute on Retirement Security. [79% figure -- possibly the same source as Kenneally, K. above, which cites 87%; reconcile which figure is correct before submission.]

Norberg, Johan. "How Sweden Saved Social Security." Cato.org, 22 Feb. 2023, www.cato.org/commentary/how-sweden-saved-social-security.
Nordea. [Sweden retirement-age population and life expectancy figures] [Author's original source not located -- verify before submission.]
Pensionsmyndigheten. "The Swedish Pension System." Pensionsmyndighetens Startsida, www.pensionsmyndigheten.se/other-languages/english-engelska/english-engelska/pension-system-in-sweden. Accessed 23 Feb. 2025.
Peter G. Peterson Foundation. [Worker-to-beneficiary ratio figures] [Author's original source not located -- verify before submission.]
Quinnell, Kenneth. "Five Causes of Wage Stagnation in the United States." AFL-CIO, 15 Jan. 2015, aflcio.org/2015/1/15/five-causes-wage-stagnation-united-states.
Regeringskansliet. [Sweden minimum pension age policy] [Author's original source not located -- verify before submission.]
Shutt, Jennifer. "Social Security and Medicare: Where Do Harris and Trump Stand?" Missouri Independent, 18 Oct. 2024, missouriindependent.com/2024/10/18/social-security-and-medicare-where-do-harris-and-trump-stand/.
Skatteverket. [Sweden employer pension contribution rate] [Author's original source not located -- verify before submission.]
Social Security Administration. "A Summary of the 2024 Annual Reports." Trustees Report Summary, 2024, www.ssa.gov/oact/trsum/.
Social Security Administration. "Fast Facts and Figures about Social Security, 2023." SSA, 2023, www.ssa.gov/policy/docs/chartbooks/fast_facts/2023/fast_facts23.pdf.
Social Security Administration. "Historical Background and Development of Social Security." Social Security History, www.ssa.gov/history/briefhistory3.html. Accessed 23 Feb. 2025.
Social Security Administration. "Maximum Taxable Earnings." SSA, 2025, www.ssa.gov/benefits/retirement/planner/maxtax.html.
Social Security Administration. "Social Security Benefit Tables." Benefit Tables, www.ssa.gov/oact/progdata/tableForm.html. Accessed 23 Feb. 2025.
Springs, Abby. "Voters Overwhelmingly Oppose GOP Efforts to Privatize Social Security." Data For Progress, Data For Progress, 27 July 2023, www.dataforprogress.org/blog/2022/12/22/voters-overwhelmingly-oppose-gop-efforts-to-privatize-social-security.
Statista Research Department. [Author's original source not located -- verify before submission.]
Sutton, K. "Social Security." Benefit Tables, www.ssa.gov/oact/progdata/tableForm.html. Accessed 23 Feb. 2025.
The Conference Board. "US Consumer Confidence." The Conference Board, 1 Feb. 2025, www.conference-board.org/topics/consumer-confidence.
Trading Economics. [Sweden debt-to-GDP figure] [Author's original source not located -- verify before submission.]

United States, Bureau of Labor Statistics. "Employer Costs for Employee Compensation, Retirement Benefits: Access, Participation, and Take-up Rates, Private Industry Workers, March 2022." U.S. Department of Labor, 2023, www.bls.gov/opub/mlr/2024/highcharts/data/erisa_chart-4.stm.

United States, Government Accountability Office. "Social Security: Coverage of Public Employees and Implications for Reform." GAO, 9 June 2005, www.gao.gov/products/gao-05-786t.

Veradigm. [Baby Boomer retirement rate figures] [Author's original source not located -- verify before submission.]

Weaver, R. Kent. "Social Security Administration." Social Security Administration Research, Statistics, and Policy Analysis, 1 May 2005, www.ssa.gov/policy/docs/ssb/v65n4/v65n4p38.html.

Firm Scale and Advertising Efficiency in the Apparel Sector
A Mixed-Methods Comparative Case Study of Paid Social Media and Out-of-Home Advertising By Ethan Moran

Abstract

Businesses do not select advertising channels in isolation. An online-only apparel start-up and a department store have different budgets, data availability, distribution systems, and measurement toolkits. This research investigates how firm scale alters tradeoffs between paid social media and out-of-home (OOH) advertising in the retail apparel category. With publicly available campaign records too sparse to conduct reliable large-sample regression analysis, this study employs a structured mixed-methods comparison of seven reported campaigns. The reported lift includes standard metrics of return on advertising spend (ROAS), customer acquisition cost (CAC), sales lift, visits, lift in awareness, and purchase intent. This is combined with a qualitative assessment of campaign goals, geographic fit to target audience, privacy limits, source quality, and attribution techniques. Results indicate that paid social remains financially accessible for independent brands, but its role as a low-risk testing channel has become less certain, as Apple's App Tracking Transparency (ATT) has weakened conversion-based optimization for many small e-commerce firms. OOH is also not solely the province of large firms. Smaller companies can deploy hyper-local, guerrilla, event-based, or programmatic digital OOH, while larger brands can tie national- or regional-level OOH efforts to owned channels like stores, e-commerce, and brand lift tracking studies. Firm scale influences feasible tactics and quality of measurement more than it dictates a single best channel. Paid social media and OOH are more effectively used as complements than as direct substitutes.

Keywords: firm scale, advertising efficiency, paid social media advertising, out-of-home advertising, apparel sector, incrementality testing, attribution bias, marketing mix modeling

1. Introduction

Advertising is a risky investment. A firm incurs costs today in hopes that advertising will lift present sales, future demand, or brand equity. The key question is therefore not whether an advertisement grabs attention, but whether it generates incremental value that exceeds its cost, margin, and the returns that could be achieved through alternative uses of the same resources. The comparison between paid social media advertising and out-of-home (OOH) advertising is easier when both channels report similar types of outcomes.

Firms can buy paid social media advertisements in increments as small as a single day's advertising budget. They can link clicks to website activity and purchases fairly directly. By contrast, OOH media—billboards, transit wraps, street furniture, event marketing, or digital displays near sidewalks—has proven harder to tie to transactions. As a result, marketers report OOH impacts on awareness, search, visits, or sales changes within the area it reaches.

Advertising responded to digital capabilities by telling a straightforward story: paid social media advertising must be cheaper because targeting is better and reporting is easier. The reality

is more complex. Digital dashboards illuminate some outcomes, but what is easy to observe does not equal causal impact. Platforms can attribute a sale to an advertisement even when those purchases would have occurred regardless. Meanwhile, OOH may drive earlier stages of the funnel. Measured against this pathway, a last-click report undervalues OOH and overstates the effect of digital exposures near purchase.

Brand scale exacerbates the problem. Many independent apparel companies operate with limited cash reserves, narrow assortments, and significant pressure to recoup ad expenditures quickly. Paid social allows a small starting budget and easy creative iteration. Recent privacy changes have degraded this learning process, particularly for smaller firms. This still leaves entry costs low, in the literal sense that a firm can always start with a tiny daily budget. But learning is not cheap if the advertiser needs substantial data, first-party information, or cleaner tests to understand whether advertisements actually drove profitable purchases.

The same can be said for OOH. Large retailers may buy programmatic or city-wide OOH across hundreds of locations. Small brands could instead use local, guerrilla, event-based, or programmatic digital placements when their target audience is near. This does not imply that every small brand should invest in OOH. For a growing company, OOH should not be equated with national-scale billboard buys.

The clothing sector is informative because apparel and accessories are purchased by both retail shoppers and e-commerce audiences. Social media can place products alongside creators and customer content while also providing direct links to purchase. OOH presents brands in public spaces and creates local awareness near stores or events. These can benefit firms at any stage. Apparel also includes luxury-label firms, department stores, design-focused independents, and wholesale giants attached to mass retailers. Limiting the scope to clothing excludes industries with different average prices, regulations, margins, and purchasing frequencies.

This paper studies how brand scale influences marketers' choice between paid social media and OOH advertising in the apparel sector. It argues that paid social provides independents with low barriers to entry and easy-to-access response data, but advertising's post-ATT performance may be less predictable than prior accounts of low-cost experimentation suggest. OOH presents options for small or big businesses, but scale changes the tactics marketers can use, evolving from local, guerrilla, one-time, or event-based placements to programmatic buys tied to broader inventory or brand-building efforts. Lastly, social media and OOH too often act as complements instead of substitutes. Managers should focus on how each channel supports a campaign's objective, not insist that every observed outcome can be distilled into clicks or conversions.

2. Literature Review and Economic Framework

2.1 Cost, Targeting, and Causal Measurement

The optimal level of advertising occurs when the marginal expected contribution of another advertising dollar is higher than its marginal cost and the return that can be earned elsewhere. Paid socializing reduces the minimum scale necessary to operate at that level. An

advertiser can budget daily, test creatives, segment audiences, and move budgets with ease. This flexibility is valuable if a firm is young with little customer history and cannot stomach a botched national rollout. However, a low cost of entry does not imply that a test will provide trustworthy answers.

Digital targeting can increase efficiency by limiting impressions to users that the platform predicts are more likely to purchase. Targeting is, however, an imperfect signal. Privacy regulations, idiosyncratic blips from device switching or algorithm updates, and other noise can diminish the reliability of both delivery and reporting. Tian et al. even demonstrate that cookie-based measurement misses any lift that occurs outside of the tracked environment. Absent perfect information, a dashboard could display behavior very accurately yet still measure only a fraction of the advertisement's overall economic impact.

That is why attributed lift and incremental lift must be differentiated. Attributed Return on Advertising Spend (ROAS) measures total revenue assigned to an advertisement divided by campaign spend. Incremental ROAS measures how much additional revenue the advertisement drove versus what would have sold otherwise. Lewis and Rao empirically derive this conclusion using 25 large field experiments. They find that advertising responses are hard to detect because ordinary variation in sales can dwarf campaign effects. Gordon et al. came to this conclusion when observational advertising estimates were compared to randomly assigned Facebook banner experiments. Advertisers should take away a simple lesson from these papers: seeing a 3x ROAS on a dashboard provides useful information but should not be regarded as evidence that the advertisement caused all of that sales lift.

Randomized incremental-lift tests provide the cleanest causal comparison. Firms running ads on paid social can withhold the advertisement from a holdout group. Out-of-home advertisers can run comparisons across cities, neighborhoods, stores, or time periods. This study design is optimal, but impractical in some cases. It requires sufficient traffic to accumulate enough observations, a persistent treatment platform, or media-owner cooperation, and access to both online and offline sales data. Smaller firms may lack enough web or storefront traffic to detect a precise lift. Also, companies might be reluctant to deny impressions to paying customers.

2.2 Privacy Regulation, Measurement Tools, and Unequal Digital Returns

Firm scale impacts digital advertising via many channels: more data, larger budgets, distribution realities, and organizational scale. Bigger firms tend to have more customers, more transactions, and more data to work with when building audiences or assessing tests. Smaller firms often enjoy lower entry prices paid for social presence. But small brands may disproportionately rely on conversion-optimization systems outside of their control.

Consider one piece of direct evidence, presented by Aridor et al. Examining e-commerce firms during Apple's ATT policy rollout, they find reduced click-through rates on conversion-optimized Meta advertisements. More importantly, they show revenue losses were focused on smaller firms that were more dependent on Meta pre-ATT. Paid social remains viable, but its performance has become less predictable. The ease, cheapness, and precision of

acquisition have simply moved from reliable to uncertain for small brands. Tighter budgets may force more first-party data, more repeated experimentation, and less rigid timelines for learning.

Measurement need not be a pipe dream either. Runge et al. recently introduced Robyn, an open-source market mix modeling (MMM) package designed to make aggregate measurement more usable for digital-first firms. Software like Robyn reduce barriers to entry and can help firms better leverage privacy-friendly data. They will not fix issues such as needing good time-series data, ensuring variation in ad spend, making careful model specifications, and applying human intuition. Open-source MMM can democratize access to measurement, but it will not solve all internal validity problems for small advertisers with limited historical data.

2.3 OOH, Programmatic Buying, and Brand Capital

Taylor et al. explain billboard effectiveness through selective attention and retail-gravity ideas: an ad is more useful when it reaches relevant people and appears near a practical destination. That older framework remains helpful, but OOH has changed. Digital screens, programmatic buying, mobile-location panels, day-parting, contextual triggers, and dynamic creative have made many OOH campaigns more flexible. Vangelov argues that AI and digital OOH can improve campaign relevance by using contextual and city-level data, especially in smart-city environments. This does not prove that programmatic OOH (pOOH) is always inexpensive, but it does show that OOH is no longer only a fixed, long-term billboard purchase.

For small apparel brands, this development matters because there are now more ways to enter OOH. A brand might buy a short programmatic placement near a concert, a shopping district, a campus, or a pop-up event. It might also use a mural, storefront installation, projection, or local transit placement. Farooqui provides supporting, though limited, evidence that some guerrilla marketing tactics can affect clothing purchase behavior. The better claim is not that small firms can use OOH just like national brands. The better claim is that smaller firms can use narrower OOH tactics when geography and audience fit are strong.

For large firms, OOH can do something different. A retailer with many stores and an established brand can convert public exposure through several paths. The customer might buy at a store, search online, visit the website, or respond to a later digital ad. OOH can also build brand capital, meaning familiarity, legitimacy, and preference that may reduce future acquisition costs. Those benefits are difficult to translate into immediate ROAS, but they are still economically relevant if they affect later demand.

2.4 Complex Channel Synergies and Triangulated Measurement

Recent studies are abandoning this binary framing altogether. Bell et al., who evaluate 1,083 campaigns, conclude that "Channel patterns, and not individual channels per se, are what power higher-performing media plans." Similarly, Dinner et al., who examine a high-touch apparel retailer, find substantive cross-channel effects between traditional advertising, online display, and paid search. Rather than identifying one channel as superior, these studies reveal a more nuanced relationship between the channels: OOH builds awareness or local salience. Paid social performs well in search, retargeting, or purchases. Any broader claims about channel

superiority should therefore be treated cautiously unless they are supported by evidence on how the channels interact in practice.

Measurement suffers from the same fallacy. Hosahally et al. outline a "post-cookie" measurement framework that leverages marketing mix modeling, multi-touch attribution, and incrementality testing. The authors are clear that each methodology has its uses. Attribution is convenient for day-to-day monitoring but inherently overweights trackable touchpoints. MMM captures aggregate planning and privacy-friendly measurement but is noisy at the user or campaign level. Incrementality tests establish the strongest causal links between ad and outcome but can be expensive or underpowered. By triangulating these three methods into a cohesive measurement plan, marketers can avoid relying on a single methodology to explain all advertising effects. Together, these approaches show why advertising measurement is strongest when attribution, aggregate modeling, and experimental evidence are used to check one another rather than when any single method is treated as definitive.

2.5 Research Gap

Previous comparisons have suggested that digital advertising provides targeted reach that can be measured while traditional advertising provides broad reach that cannot. That is true but does not go far enough. It does not capture how privacy regulation, open-source measurement standards, pOOH, local guerrilla techniques, firm scale, and channel interaction alter the value of those attributes. It also risks implying that the easiest outcomes to measure are the most effective, when they are not. This study addresses that gap by comparing similar campaigns from within one product category, separating direct-response effects from branding and behavioral effects, and evaluating each public source.

3. Methodology

3.1 Research Design and Case Selection

This study employs a structured mixed-methods comparative case design with secondary evidence. Quantitative data are records of reported spending and outcomes, when available. Qualitative data are binary coding of firm scale, campaign objective, distribution, geographic scope, reversibility, privacy exposure, cross-channel design, attribution, and source quality. A case-study design is used rather than a regression due to the fact that public campaign reports rarely disclose the same variables. Forcing incompatible measures into a regression model would imply false precision.

Cases were included if they met four criteria: (1) describe an apparel or accessories business, (2) identify a paid-social or OOH campaign, (3) report at least one measurable outcome, and (4) provide enough context to understand the result. This resulted in the final sample of seven cases: Freja, Heaven Mayhem, Poppy Lissiman, H&M, Absolute Fashion, UGG, and Intimissimi. These cases were selected to represent independent or growth-stage brands and established firms, as well as direct-response and brand or behavioral objectives. The

sample is purposive rather than representative and thus does not attempt to estimate average industry returns.

3.2 Outcomes, Evidence Grades, and Triangulation

Results were separated into two categories. Direct-response outcomes include ROAS, customer-acquisition cost, sales lift, and incremental revenue. Brand and behavioral outcomes include awareness, purchase intent, website traffic, and store visits. The outcomes are reported in their original form rather than converted into one index. A store visit is not revenue, and purchase intent is not a purchase. Converting those measures into profit would require assumptions about margins, later conversion, and customer lifetime value that the public sources do not provide.

The A-D evidence tiers are a classification system and were created to make the quality of evidence visible. They are not meant to follow GRADE, a formal evidence-rating system often used in healthcare research, but instead serve as a study-specific way to compare the reliability of public advertising evidence. The framework is closer in spirit to evidence-based management: it asks whether the source is independent, whether a counterfactual is clear, whether the method is transparent, and whether the outcome is economically meaningful. Kozielski and Wilczewska discuss the importance of using indicators in an evidence-based management approach, and Andersen et al. emphasize validation concerns in case-based research. Those ideas informed the grading system used here.

Grade A is reserved for independently reported evidence with a disclosed causal design, a credible control group, and a clearly defined economic outcome. Grade B identifies company or founder reporting documented by credible independent journalism or an official source, with a clear quantitative outcome but no independent audit. Grade C covers a platform, agency, or measurement-provider case with a described method and a quantitative result but a direct commercial incentive. Grade D includes an incomplete promotional case that is useful only as an illustration.

Table 1 applies this evidence-tier system to the seven campaign cases used in the study and summarizes the main outcome, grade, and bias concern for each case.

Because this study relies on public campaign reports rather than confidential transaction-level data, it uses an adapted form of triangulation. Reported metrics are evaluated alongside case context, source incentives, attribution method, and evidence quality, rather than accepted as standalone proof of campaign effectiveness.

Three forms of triangulation were used to reduce bias. Source triangulation compared founder or company claims with independent journalism, academic research, or official campaign documentation when available. Data triangulation compared different outcome families without treating ROAS, awareness, and visits as interchangeable. Methodological triangulation combined numerical campaign results with qualitative evidence about objectives, distribution, privacy conditions, geography, and attribution design. A conclusion was kept only when it still made sense after Grade D evidence was removed and Grade C evidence was treated

as commercially reported rather than independently verified. Investigator triangulation was not possible because the study was conducted by one researcher, and that remains a limitation.

Table 1. Selected Campaign Evidence

Firm and scale	Channel	Reported outcome	Grade and main bias concern
Freja, independent/growth	Paid social	About $150 per day initially; about 2x ROAS in year two and later a consistent 3x ROAS.	B. Founder-reported metrics documented by Vogue Business; not audited or incremental.
Heaven Mayhem, independent/growth	Paid social	$15 daily starting budget; about $2.50 per reported new customer.	B. Founder-reported acquisition metric; attribution method not disclosed.
Poppy Lissiman, independent/growth	Paid social	Up to $250,000 per month; returns above 6x before Apple privacy changes, followed by lower returns.	B. Founder report; useful for privacy context but not a controlled pre-post study.
H&M, established	Paid social	About 2x online and in-store incremental ROAS on top of business-as-usual campaigns.	C. Incremental language is useful, but the claim comes from Meta/award materials. Holdout design, confidence intervals, margin, and concurrent-media controls are not public.
Absolute Fashion, independent/growth	Local OOH	116% increase in website visits and 67% sales lift after a Chicago campaign.	D. Provider case without spend, date, or clear control. Strong risk of geographic self-selection and local demand bias.
UGG, established	Programmatic digital OOH	6% lift in purchase intent and awareness gains for a weather-relevant rain collection.	C. Provider-reported brand-lift study. No campaign spend, profit result, or full matching details. Geography was likely selected for product relevance.
Intimissimi, established	Programmatic digital OOH	53.55% store-traffic lift, about four days from exposure to visit, and $0.82 per reported incremental visit.	C. Vendor test/control case. Full matching, neighborhood selection, and visit-to-profit conversion are not public.

Note. Results are reported in their original form. They are not directly comparable because objectives, baselines, attribution windows, and source incentives differ.

3.3 Analytic Procedure and Treatment of Bias

The analysis was conducted in four steps. First, each result was taken at face value, including the type of outcome as reported (attributed, incremental, relative, or lift). Second, each case was compared only with other cases within the same outcome family. Third, each case was annotated for the same mechanisms as above: low entry cost, reversibility, targeting, privacy

exposure, physical distribution, pOOH flexibility, local relevance, guerrilla potential, brand-building, channel complementarity, and attribution quality. Fourth, each interpretation was evaluated against its grade of evidence and the relevant literature body.

Finally, two risks were accounted for separately, because failing to do so would make the results too positive overall. The first is the H&M example. It is described as incremental and reports online and in-store ROAS, making it stronger than a typical platform success story. But it is still Tier C because it is published through Meta and award materials and not through an independent academic or audited source. The paper therefore counts H&M as an example of how established firms can and have designed more granular omnichannel measurement, but not as proof of general 2x causal return from Meta ads.

The second risk is geographic self-selection in OOH. Brands do not place ads in random locations. They tend to choose rich neighborhoods, high-traffic shopping corridors, culturally active areas, weather-sensitive cities, or near their own stores. If their campaign there then reports higher visits in those areas, some of that lift will come from where the campaign was located rather than from the ad itself. For that reason, OOH results are interpreted as plausible behavioral or brand signals, not clean causal estimates, unless the same public source reports a strong matched-market or randomized design.

3.4 Preferred Measurement Roadmap

The case comparison above is descriptive. A suggested future design would use a triangulated measurement approach based on three forms of measurement. Attribution can help managers track clicks, purchases, and visits in the short run, as long as the result is labeled attributed, not causal. MMM can help estimate aggregate media effects with privacy-safe time-series data, particularly when tools like Robyn lower the software barrier for smaller firms. Incrementality tests can calibrate whether the advertising caused additional sales or visits. This triangulated roadmap is more realistic than saying every firm should run only randomized tests and more rigorous than relying only on platform dashboards.

4. Findings and Discussion

No campaign in the final sample received a Grade A evidence rating. This reflects the limited availability of independently reported apparel advertising campaigns with transparent causal designs and clearly defined economic outcomes.

4.1 Paid Social: Low Entry, but More Expensive Learning

The independent-brand cases show paid social's clearest advantage: it turns one large advertising decision into a sequence of smaller decisions. Heaven Mayhem reportedly began with a $15 daily Meta budget and acquired customers for about $2.50 each. Freja started at modest levels as well, spending less than about $150 per day during an early stage, reporting about 2x ROAS in its second year and a later consistent 3x ROAS (Schulz). These are founder reports, not audited experiments. Even so, they show a real operational feature. A young brand can enter the market without signing a long placement contract.

That flexibility should not be described as easy or low-risk without qualification. Poppy Lissiman reportedly spent as much as $250,000 per month and earned more than six times that amount before Apple's privacy changes, followed by lower returns afterward (Schulz). This is consistent with Aridor et al.'s broader finding that ATT weakened conversion-optimized advertising and hurt smaller Meta-dependent e-commerce firms more severely. A small brand can still start cheaply, but it may need more data and more patient testing to know whether paid social is creating profitable customers rather than just visible dashboard results.

The cases therefore support a more nuanced first proposition. Paid social remains accessible in entry cost and flexible in execution. It is not automatically efficient. A reported 3x ROAS may be inadequate if product cost, returns, fulfillment, discounts, creative production, and agency fees absorb most of the revenue. It may also include purchases that would have happened without the advertisement. Managers should judge campaigns by contribution margin and customer lifetime value, while using platform results as one input rather than the final answer.

4.2 The H&M Case and Commercial Incrementality Claims

H&M's case is informative because it shows how established firms can measure online and in-store behavior. The campaign reportedly produced about 2x online and in-store incremental ROAS on top of business-as-usual campaigns, along with improvements in consideration and product views (Meta; Shorty Awards). The word "incremental" is significant. It suggests a stronger design than a simple attribution dashboard.

Nevertheless, the claim is not treated as Tier A evidence. The public materials do not disclose the full holdout design, sample construction, confidence intervals, profit margin, treatment timing, or how other media activity was controlled. The campaign also involved Meta, H&M, and award-submission materials, all of which have incentives to highlight success. This does not make the result useless. It means the case should support a limited claim: large firms are better positioned to run omnichannel tests and report incremental metrics. It should not be interpreted as evidence that Meta advertising generally causes a 2x incremental return for apparel brands.

4.3 OOH for Smaller Firms: Local, Guerrilla, and Programmatic Options

The Absolute Fashion case challenges the simple idea that OOH only becomes viable after a firm reaches broad distribution. The smaller Chicago brand reportedly saw a 116% increase in website visits and a 67% sales lift after a local OOH campaign (Blue Line Media). The evidence grade is low because the case lacks spend, dates, and a transparent control. But it illustrates a reasonable mechanism: a small brand can concentrate exposure in a city where its audience, events, or retail activity are concentrated.

Guerrilla and hyper-local tactics fit this logic. A small clothing brand may use a mural, a projection, a pop-up, a storefront display, a college-area activation, or one billboard in a culturally relevant place. Heaven Mayhem's founder also described using a Sunset Strip billboard after beginning with very small Meta budgets (Schulz). These are not national-scale

strategies. They are focused attempts to create social proof, local attention, or a memorable brand cue.

Programmatic digital OOH adds another option. Rather than buying a large fixed campaign, a smaller advertiser may be able to buy digital screen inventory by time, location, or context. Vangelov supports the idea that AI and DOOH can use contextual data to improve campaign relevance. The available evidence does not show that pOOH is always inexpensive, so this study avoids making that claim. The more cautious point is that pOOH can reduce transaction friction and make OOH more flexible than older static billboard models.

4.4 OOH at Established Scale and the Self-Selection Problem

The established OOH cases show why distribution and measurement capacity still matter. UGG used location and weather triggers near relevant markets and retail contexts, then connected OOH exposure with later digital activity. The provider reported a 6% increase in purchase intent and awareness gains for the promoted rain collection (Vistar Media). The result is not a profit measure, but it matches a consideration objective and uses geography connected to product relevance.

Intimissimi provides the strongest behavioral OOH case in the sample. A U.S. digital OOH campaign ran from January 24 to February 9, 2025, and used mobile-location data to compare exposed and unexposed audiences. The provider reported a 53.55% lift in store traffic, an average four-day period from exposure to visit, and $0.82 per reported incremental visit (Blindspot). This provides stronger evidence than a simple awareness claim because it connects exposure to a store visit and reports a cost denominator.

Even here, geographic self-selection limits interpretation. A brand may choose screens near stores, shopping areas, wealthy neighborhoods, high foot-traffic streets, or places where its customers already spend time. A measured lift may partly reflect smart location choice rather than the ad's causal power. Smart location choice is valuable in practice, but it is not the same as random assignment. A stronger study would compare similar markets or stores, hold out some areas, and report whether the difference remained after baseline demand and neighborhood characteristics were considered.

4.5 Channel Complementarity and Measurement Triangulation

The clearest finding across cases is that paid social and OOH report different kinds of outcomes. Paid-social cases emphasize purchases, new customers, and ROAS. OOH cases emphasize visits, awareness, intent, and geographically observed sales lift. Calculating one average across those outcomes would be misleading. A store visit is not revenue, purchase intent is not a purchase, and attributed revenue is not necessarily causal revenue.

The channels may also work together. A consumer can see a public ad, search for the brand, receive a social ad, and then buy online or in a store. Dinner et al. show that cross-channel effects can matter for an apparel retailer, and Bell et al. show that high-performing media plans often rely on channel combinations. This makes the paper's question more subtle. The issue is

not whether OOH or paid social wins. The issue is how the firm can measure what each channel contributes at each stage of the customer journey.

A practical measurement system should be triangulated. Attribution can help show more short-term and trackable responses to a campaign. MMM can help estimate broader channel effects using aggregate data, especially when tools like Robyn make MMM more accessible. Incrementality tests should be used when decisions are large enough or uncertain enough to justify them. None of these methods is perfect alone. Together, they reduce the risk that a firm optimizes the channel that is easiest to observe rather than the channel that creates the most profit.

4.6 Revised Cross-Case Interpretation

Dimension	Independent or growth-stage firms	Established firms
Paid-social economics	Low entry cost and fast budget changes, but post-ATT optimization is less precise. Learning now depends more on first-party data, repeated tests, and contribution-margin thinking.	Larger datasets and omnichannel records support more credible incremental tests and portfolio optimization.
OOH tactic	Local, guerrilla, event-based, or pOOH placements tied to a narrow geographic hypothesis.	Broader, programmatic, or store-linked OOH connected to e-commerce reach, dynamic creative, and brand tracking.
Measurement approach	Attributed CPA/ROAS clearly labeled, first-party data, simple holdouts, and eventually open-source MMM when enough data exist.	Incremental online and offline profit, matched markets, brand lift, visitation, MMM, and multi-channel experiments.
Main risk	Platform dependence, limited statistical power, audience saturation, and reading attributed ROAS as profit.	Large fixed commitments, geographic self-selection, complex attribution, and over-crediting broad exposure.
Likely channel role	Paid social for response and learning; OOH or pOOH for local legitimacy and brand cues when geography is tight.	Paid social and OOH as complementary portfolio elements, often measured through a triangulated framework.

5. Practical Implications

Independent apparel firms should begin with a maximum acquisition cost based on contribution margin rather than an attractive but incomplete ROAS target. Paid social can still be used to test offers and creative materials, but a small daily budget should not be confused with a clean experiment. Brands should preserve first-party customer data, compare repeated periods instead of reacting to one short spike, and use simple holdouts when traffic is high enough. When an experiment is too small to be precise, the result should be treated as uncertain.

Smaller firms should also consider OOH more flexibly than has traditionally been assumed. They do not need national distribution to use the channel. They do need a strong local reason for the placement. A store opening, pop-up, campus event, fashion market, city with a concentrated customer base, or culturally relevant billboard can make more sense than a broad campaign. A measurable next step, such as branded search, a landing page, a store visit, or a matched-area comparison, should be planned before the campaign begins.

Established firms should separate performance and brand objectives before comparing channels. Paid social can be evaluated through incremental purchases across online and offline channels. OOH can be evaluated through visits, search, awareness, matched markets, and later sales. For both channels, the final question should be their incremental contribution to profit, not only attributed revenue. A mature advertiser should use attribution, MMM, and incrementality tests together instead of letting one dashboard decide the budget.

6. Limitations and Future Research

This study has several limitations. First, the sample is purposive and consists mainly of public success stories, so positive publication bias is likely. Failed campaigns are less likely to be published. Second, most cases are reported by founders, platforms, agencies, or measurement vendors with commercial incentives. Source and methodological triangulation reduce the risk of accepting a promotional claim at face value, but they cannot turn self-selected cases into representative evidence. Third, outcomes differ across campaigns and cannot be pooled into one average. Fourth, firm scale is classified according to independence, distribution, and organizational reach rather than a consistent revenue threshold because private financial data are often unavailable. Fifth, OOH cases face geographic self-selection because brands choose locations where they expect better performance. Sixth, investigator triangulation was unavailable.

Future research should adopt a stronger measurement design. Paid social audiences could be randomly divided into treatment and holdout groups. OOH could use randomized or carefully matched cities, store catchment areas, or launch periods. Researchers should prespecify outcomes, follow customers long enough to capture delayed purchases, and integrate online and offline transactions. The most useful economic outcome would be incremental contribution to profit: additional revenue multiplied by contribution margin, less advertising and campaign-specific operating costs. Brand lift, search, and visits should remain supporting mechanisms unless they can be linked to later profit.

A realistic future roadmap should not rely on one method. MMM can help with broad channel planning and privacy-safe aggregate measurement, especially as open-source tools reduce cost barriers. Multi-touch attribution can support short-term optimization if it is not mistaken for causality. Incrementality tests can calibrate whether the other methods are over- or under-crediting a channel. This triangulated approach is especially important in apparel, where a consumer’s path may move from public exposure to search to social media to website or store purchase.

7. Conclusion

This study examined how firm scale changes the economics of paid social and OOH advertising in the apparel industry. The findings do not support an across-the-board channel ranking. Paid social still has a low financial entry point and continues to provide small brands with a direct path to view responses. However, post-ATT, the channel is more volatile and less accurate than the classic low-stakes testing story. The ease of entry has expanded to become accessible in terms of initial cost, but not automatically accessible in terms of reliable measurement or profitability.

OOH is also not only a large-firm channel. Smaller apparel brands can access guerrilla, hyper-local, event-based, or programmatic digital OOH when geography and audience are tightly defined. As firms scale, OOH can expand to include broader, data-enabled, store-linked campaigns that can also contribute to awareness, traffic, and consideration. Scale changes the tactic and the quality of measurement more than it dictates whether the channel is appropriate at all.

The contribution of this study is methodological. Advertising channels should not be ranked by the outcomes they make easiest to observe. Paid social often reports purchases and ROAS, whereas OOH often shows visits, awareness, or purchase intent. Both are susceptible to bias. The more defensible approach is to consider what the campaign was designed to do, whether the evidence includes a counterfactual, and how the result translates to contribution profit. Until public campaign data become more complete, transparent, triangulated case comparisons provide more credible evidence than a regression built from incompatible measures or missing values.

Works Cited

Andersen, P. H., A. Dubois, and F. Lind. “Process Validation: Coping with Three Dilemmas in Process-Based Single-Case Research.” Journal of Business & Industrial Marketing, vol. 33, no. 4, 2018, pp. 539-549. https://doi.org/10.1108/JBIM-07-2016-0152.

Aridor, G., et al. “Evaluating the Impact of Privacy Regulation on E-Commerce Firms: Evidence from Apple’s App Tracking Transparency.” Management Science, 2025, advance online publication. https://doi.org/10.1287/mnsc.2024.06600.

Bell, J. J., F. Thomaz, and A. T. Stephen. “Beyond the Pair: Media Archetypes and Complex Channel Synergies in Advertising.” Journal of Marketing, vol. 89, no. 4, 2025, pp. 99-119. https://doi.org/10.1177/00222429241302808.

Blindspot. “Intimissimi & Blindspot: Driving a 53% Lift in Store Visits across the U.S.” 7 May 2026, https://seeblindspot.com/intimissimi-blindspot-driving-a-53-lift-in-store-visits-across-the-u-s/.

Blue Line Media. “Absolute Fashion.” Blue Line Media, n.d., https://www.bluelinemedia.com/case-studies.

Dinner, I. M., H. J. van Heerde, and S. A. Neslin. “Driving Online and Offline Sales: The Cross-Channel Effects of Traditional, Online Display, and Paid Search Advertising.” Journal of Marketing Research, vol. 51, no. 5, 2014, pp. 527-545. https://doi.org/10.1509/jmr.11.0466.

Farooqui, R. “The Role of Guerilla Marketing for Consumer Buying Behavior in the Clothing Industry of Pakistan Using Structural Equation Modeling (SEM).” South Asian Journal of Management Sciences, vol. 15, no. 1, 2021, pp. 52-68.

Gordon, B. R., et al. “A Comparison of Approaches to Advertising Measurement: Evidence from Big Field Experiments at Facebook.” Marketing Science, vol. 38, no. 2, 2019, pp. 193-225. https://doi.org/10.1287/mksc.2018.1135.

Hosahally, S., M. Bharadwaj, and A. Zaremba. “Measuring Digital Advertising in a Post-Cookie Era: A Study of Marketing-Mix Models, Attribution and Incrementality.” Journal of Digital & Social Media Marketing, vol. 12, no. 4, 2025, pp. 348-373.

Kozielski, R., and D. Wilczewska. “The Use of Marketing Indicators in Poland: Towards Evidence-Based Management.” Marketing of Scientific and Research Organizations, vol. 47, no. 1, 2023, pp. 1-16. https://doi.org/10.2478/minib-2023-0001.

Lewis, R. A., and J. M. Rao. “The Unfavorable Economics of Measuring the Returns to Advertising.” The Quarterly Journal of Economics, vol. 130, no. 4, 2015, pp. 1941-1973. https://doi.org/10.1093/qje/qjv023.

Meta. “H&M: Facebook Ads Case Study.” Meta for Business, n.d., https://www.facebook.com/business/success/hm.

Runge, J., et al. “Packaging Up Media Mix Modeling: An Introduction to Robyn’s Open-Source Approach.” arXiv, 2024, https://arxiv.org/abs/2403.14674.

Schulz, M. “Is the Era of Digital Brand-Building Over?” Vogue Business, 1 May 2025, https://www.vogue.com/article/is-the-era-of-digital-brand-building-over.

Shorty Awards. “H&M and Meta Partner on ‘Find Your Beach’ Summer Campaign.” Shorty Awards, n.d., https://shortyawards.com/17th/tk.

Taylor, C. R., G. R. Franke, and H.-K. Bang. “Use and Effectiveness of Billboards: Perspectives from Selective-Perception Theory and Retail-Gravity Models.” Journal of Advertising, vol. 35, no. 4, 2006, pp. 21-34. https://doi.org/10.2753/JOA0091-3367350402.

Tian, M., P. R. Hoban, and N. Arora. “What Cookie-Based Advertising Effectiveness Fails to Measure.” Marketing Science, vol. 43, no. 2, 2024, pp. 407-418. https://doi.org/10.1287/mksc.2023.1453.

Vangelov, N. “AI and Digital Out-of-Home Advertising in Smart Cities.” Smart Cities and Regional Development Journal, vol. 9, no. 2, 2025, pp. 55-61. https://doi.org/10.25019/hjab9w14.

Vistar Media. “UGG Boosts Purchase Intent with Weather Triggers.” Vistar Media, 22 Mar. 2023, https://www.vistarmedia.com/blog/ugg-case-study.

The Impact of Macroeconomic and Institutional Factors on Trade By Cathy Zhang

Abstract

This paper examines the main macroeconomic and institutional factors that influence international trade performance. Instead of focusing on single factors, this study uses cross-sectional data from 137 countries to analyze the effect of foreign direct investment (FDI), inflation, tariff rates, institutional quality, and labour market conditions on trade openness as a percentage of GDP. The analysis uses descriptive statistics, scatter plots, and regression models. The results show that lower tariff rates, higher levels of FDI, greater political stability, more effective governance, and higher wages are associated with higher levels of trade. In contrast, inflation and tariff rates exhibited a negative relationship with trade openness. The findings suggest that a country's trade performance depends not only on financial and economic conditions, but also on the quality of its institutions and labour markets. These results may be useful for policymakers, business managers, development agencies, international organizations, investors, and researchers to improve trade performance and support economic growth.

Keywords International trade, macroeconomic indicators, labour market, institutional quality

Introduction

Trade is a major driver of a country's economic growth and global influence. Historically, commerce has been used for mutual benefit, to indicate alliances, and as leverage for political negotiation. It has enabled economies to expand beyond domestic markets and is a key component of a country's growth. Thus, it is important for policymakers looking to deepen their country's integration into the global economy to understand the determinants of trade performance and how they are related to the macroeconomic conditions.

There have been many studies in the literature that investigates the various factors that influence trades. But, most of the existing literature is focused on individual variables. In practice, trade performance is affected by a multiplicity of factors at the same time, such as institutional qualities and labour market characteristics. These factors provide a broader picture of how open, competitive, and resilient an economy is to international markets.

The present study considers trade trade as a percentage of GDP, in relation to several key variables: foreign direct investment (FDI), inflation, tariff rate, political stability, government effectiveness, wage, interest rate, and exchange rate. Using cross sectional data from 137 countries as provided by the World Bank Development Indicator, the study analyzes macroeconomic, institutional, and labour market factors of trade.

The main objective is to enhance the understanding of the relationship between these indicators and international trade performance, thereby providing useful insight for policies aiming to promote the expansion of trade and economic development.

The empirical results show statistically significant positive relationships between trade and FDI, political stability, government effectiveness, and wage. In contrast, the evidence indicates that tariff rates and inflation are negatively correlated with trade.

This paper contributes to the existing literature on international trade and developing economies in several important ways. First, it integrates both macroeconomic indicators and institutional factors, rather than examining them separately. Second, it highlights the importance of governance and political stability in supporting trade performance. Third, this study also introduces wage as a labour market indicator, showing that labour market conditions may also influence trade outcomes. These findings suggest policies to promote trade should address not only macroeconomic conditions, but also the quality of institutions and domestic labour market conditions.

The rest of the paper is organized as follows: section 2 reviews the related literature, section 3 presents the empirical evidence and results, and section 4 concludes the paper.

Literature Review

A wide range of methods have been used to examine the relationship between trade and macroeconomic indicators. These methods include panel data analysis (Cong et al., 2024; Nam and Ryu, 2024), Granger causality test (Ridwan et al., 2025; Min, 2023), regression analysis (Lee, 2012; Badinger, 2009; Nzama et al., 2023; Kang and Dagli, 2018; Kennedy, 2007; Nam and Ryu, 2024), cointegration test (Abdi et al., 2024; Zhou, 1996), gravity model (Cheung et al., 2017; Bagozzi and Landis, 2012; Ismail, 2013; Kang and Dagli, 2018; Kumar et al., 2024), autoregressive distributive lag models (ARDL) (Kaukab and Anggara, 2024; Nwagu et al., 2024;), and theoretical models (Sampson, 2014; Marjit et al., 2008).

The variables used to explain trade performance also differed across studies. Frequently examined macroeconomic factors include GDP (Abdi et al., 2024; Badinger, 2009; Kaukab and Anggara, 2024; Le et al., 2023; Cong et al., 2024; Nwagu et al., 2024; Kennedy, 2007; Kumar et al., 2024), economic growth (Le et al., 2023), inflation (Kaukab and Anggara, 2024; Cong et al., 2024; Kennedy, 2007; Ridwan et al., 2025;), tariff rate (Cheung et al., 2017; Nam and Ryu, 2024), political stability (Le et al., 2023; Bagozzi and Landis, 2012;), openness (Abdi et al., 2024; Badinger, 2009; Kaukab and Anggara, 2024; Le et al., 2023; Bagozzi and Landis, 2012; Nzama et al., 2023; Ridwan et al., 2025; Nam and Ryu, 2024), exchange rate (Cong et al., 2024; Nwagu et al., 2024; Ridwan et al., 2025; Kumar et al., 2024), governance (Nzama et al., 2023), Foreign Direct Investment (Min, 2003; Abdi et al., 2024; Kaukab and Anggara, 2024; Nwagu et al., 2024; Nam and Ryu, 2024;), financial development (Nzama et al., 2023; Nam and Ryu, 2024), population (Badinger, 2009; Ridwan et al., 2025;), unemployment (Kennedy, 2007; Nam and Ryu, 2024), government debt (Cong et al., 2024;), wage (Sampson, 2014; Marjit et al., 2008;), central bank independence (Badinger, 2009;), labour force (Abdi et al., 2024;), Gross National Income (Lee, 2012;).

Various studies have also examined different countries and regions, including the OECD countries (Lee, 2012), Malaysia (Min, 2003), Somalia (Abdi et al., 2024), Indonesia (Kaukab and

Anggara, 2024), Bangladesh (Ridwan et al., 2025), India (Kumar et al., 2024), and Nigeria (Nwagu et al., 2024). These geographical varieties show that the influence factors of trade may differ between locations.

Several studies highlight the importance of monetary conditions and FDI: Nwagu et al. (2024) emphasized that interest rates changes significantly influence Nigeria's trade balance, showing how monetary stability supports competitiveness in exports. Similarly, Min (2003) found that FDI played a key role in Malaysia's export growth, which is further supported by Abi et al. (2024) who reported that FDI and domestic investment supported economic growth in Somalia. Le et al. (2023) extends this perspective by showing that FDI positively benefits from trade openness, suggesting that trade and FDI may influence one another. Cong et al. (2024) reinforces this finding by demonstrating that FDI is able to increase exchange rates through trade freedom.

However, the relationship among trade openness, FDI, and macroeconomic performance is not always positive. For example, Kaukab and Anggara (2024) found that trade openness increased inflation, while GDP and FDI reduced it in Indonesia. This suggests that the effect of trade and investments may depend on domestic economic conditions.

Institutional qualities are another important factor. Badinger (2009) found that greater trade and financial openness can reduce inflation and affect the relationship between output and inflation. To support this, Nzama et al. (2023) found that trade openness may improve government effectiveness, while financial openness can produce negative outcomes when appropriate regulations are absent. These findings suggest that countries require effective institutions and appropriate policies to gain the full benefits of trade and financial openness.

Beyond economic and institutional factors, it is suggested that socio-geographic factors, such as geography and history, can also dictate a country's trade performance. In particular, Kang and Dagli (2018) demonstrated that exchange rate movements, particularly currency depreciation, can boost exports, although it is important to also note a country's historical trade structure and export composition.

Meanwhile, Marjit et al. (2008) approach this examination using human variables: they computed the relationship between trade and various labour forces, including the impact of skilled and unskilled labour on a country's trade performance. The findings show that trade and fragmentation raise unskilled wages but may still increase or decrease the wage gap, suggesting the varied effects of trade liberalization on different income groups.

The extensive review of literature on determinant indicators of trade is presented in Table 1 below.

Table 1 Summarization of Literature

Authors	Year	Region	Key Variables	Key Findings	Method	Period
Jong-Eun Lee	2012	OECD	Gross National Income (GNI), oil price, Production	High demand and unemployment reduces trade.	Regression	1995 to 2008
Byung S. Min	2003	Malaysia	FDI, export, and import	FDI boosted exports.	Causality Analysis	1970-1990
Abdi et al.	2024	Somalia	GDP, FDI, Trade Openness, Capital, and Labour Force	FDI and domestic investment boost Somalia's economic growth.	Cointegration	1990-2020
Cheung et al.	2017	Global	Imports, Extensive Margin, Tariff, FTA, PTA, PSA, CU	PTAs increase trade by reducing both tariffs and non-tariff barriers.	Gravity Model, OLS	1998-2010
Harald Badinger	2009	Global	Openness, political instability, Capital, GDP,	Greater trade and financial openness lowers inflation and increases output–inflation tradeoff.	Regression (OLS, IV)	1985 - 2004
Kaukab & Anggara	2024	Indonesia	GDP, Inflation, Openness, RGDP (real gross domestic product), FDI	In Indonesia, trade openness increases inflation, while GDP and FDI reduce inflation.	ARDL model	1985 - 2022
Le et al.	2023	Asia-Pacific	Trade Openness, Political stability, CPI, corruption	Trade openness boosts FDI, while political stability has a negative correlation.	OLS, Panel Data Analysis	1990-2020
Bagozzi & Landis	2012	Global	Trade Openness, Political stability	International political ties reduce trade volatility more than trade agreements.	Gravity Model	1950-1999
Nzama et al.	2023	Global	Trade and Financial openness, government effectiveness	Trade openness improves government effectiveness, while financial openness has negative effects without strong policies.	Quantile Regression	2010-2020
Normaz wana Ismail	2013	Asia	R&D, Hi-tech Exports	Innovation driven by R&D and FDI boosts high-tech exports in Asian countries.	Gravity Model	2004-2009
Marjit et al.	2008	Global	Trade, Skilled and Unskilled Wage	Trade and fragmentation raise unskilled wages but may still increase or decrease the wage gap.	Theoretical Model	20th century

Kang & Dagli	2018	Global	Distance, Colonial Relationship, Common Language, Geographical Continuity	Currency depreciation boosts exports.	Gravity Model, Panel Regression	2001-2015
Cong et al.	2024	Global	Exchange Rate Index, Trade Freedom, GDP, Debt	This study finds a U-shaped relationship between trade freedom and exchange rate movement. Trade freedom boosts exchange rates through high-tech trade and FDI.	Panel Data, GMM Model	2015-2021
Nwagu et al.	2024	Nigeria	Trade Balance, monetary policy rate, GDP, FDI, Oil Price	Changes in policy rate significantly impact the trade balance.	ARDL Model	1981-2022
Su Zhou	1996	ASEAN, Japan, USA	Regional Trade, Interest Rate	ASEAN relies heavily on the US and Japan for trade and investment.	Cointegration Test	20th century
Ryan Kennedy	2007	Global	Accountability, Trade Policy, Level of Trade Liberalization, Inflation	Less government control and more political openness increases trade reform.	Regression Analysis	1991-1999
Ridwan et al.	2025	Bangladesh	Inflation, Exchange Rate, Population, Trade Openness, Economic Development	Inflation and population boost economic growth, while trade openness negatively impacts it.	Granger Causality Test	1990-2020
Nam & Ryu	2024	South East Asian Nations	Economic Growth, Trade Openness, Tariff, Import, Export, FDI, CPI, Employment	Trade volume boosts growth, while reducing trade barriers is not always beneficial.	Panel Data Regression	2000-2022
Kumar et al.	2024	India	GDP Per Capita, Income, Trade, Exchange Rate	Variables such as GDP, population, and currency value boost exports, while local production in importing countries lowers demand.	Gravity Model, Panel Cross-Section Dependence Test	1995-2021

Empirical Findings

Data

Appendix 1 presents data on trade and several economic indicators obtained from the World Bank World Development Indicators. Trade, measured as a percentage of GDP is compared against economic variables that include tariff rate (%), foreign direct investment (FDI) as a percentage of GDP, consumer price index as a proxy of inflation, political stability and absence of violence, government effectiveness, voice and accountability, medium and hi-tech manufacturing as a percentage of manufacturing value added, wage and salaried workers as a percentage of employment, real effective exchange rate index, and real interest rate (%).

Descriptive Statistics

Table 1. Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Trade	137	83.069	37.686	2.699	188.945
Tariff Rate	91	8.979	5.338	.4	26.31
FDI net	127	3.181	4.435	-12.323	28.175
Inflation	115	192.859	111.493	107.84	708.321
Political Stability	130	-.352	.923	-2.777	1.209
Government Effectiveness	130	-.367	.738	-2.228	1.619
Voice and Accountability	130	-.352	.819	-1.946	1.285
High-tech	100	19.444	13.504	.26	62.457
Exchange Rate	52	106.59	37.787	60.362	334.622
Interest Rate	67	.068	9.725	-36.833	39.184
Wage	129	53.428	25.399	5.454	99.605

Table 1 presents the descriptive statistics for the variables used in the study. As shown, trade has 137 observations, with a mean of 83.069 and a standard deviation of 37.686, ranging from a minimum of 2.699 to a maximum of 188.945. Tariff Rate has 91 observations with a mean of 8.979, and standard deviation is 5.338, ranging from a minimum of 0.4 to a maximum of 26.31. FDI net has 127 observations with a mean of 3.181, a standard deviation of 4.435, with a minimum of -12.323 to a maximum of 28.175. Inflation has 115 observations, with a mean of 192.859 and a standard deviation of 111.493. It ranges from a minimum of 107.84 to a maximum of 708.321. Political Stability has 130 observations recorded ,with a mean of -0.352 and a standard deviation of 0.923. This includes a minimum of -2.777 to a maximum of 1.209. Government Effectiveness has 130 observations with a mean of -0.367. Its standard deviation is 0.738, with a minimum of -2.228 to a maximum of 1.619. Voice and Accountability has 130 observations, with a mean of -0.352 and a standard deviation of 0.819, ranging from a minimum of -1.946 to a maximum of 1.285. High-tech has 100 observations, with a mean of 19.444 and a standard deviation of 13.504. This includes a minimum of 0.26 to a maximum of 52.457. Exchange Rate has 52 observations, with a mean of 106.59. Its standard deviation is 37.787 with a minimum of 60.362 to a maximum of 334.622. Interest Rate has 67 observations, with a mean

of 0.068. Its standard deviation is 9.725, ranging from a minimum of -36.833 to a maximum of 39.184. Finally, wage has 129 observations, with a mean of 53.428 and a standard deviation of 25.399, ranging from a minimum of 5.454 to a maximum of 99.605.

Model

To investigate the effect of macroeconomic, labour market, and institutional factors on trade, we estimate the following model by referencing existing studies (Lee, 2012; Badinger, 2009; Kennedy, 2007):

$$Y_i = B_0 + B_iX_i + u_i$$

where Y_i represents trade as a percentage of GDP for country (i). B_0 is a constant term, representing the expected value of the dependent variable when all independent variables are zero. B_i is the coefficient of each explanatory variable, which include tariff rate, FDI net, inflation, political stability, government effectiveness, voice and accountability, hi-tech, wage, exchange rate, and interest rate.

Findings

Figure 1. Scatterplots of Trade vs Macroeconomic Indicators

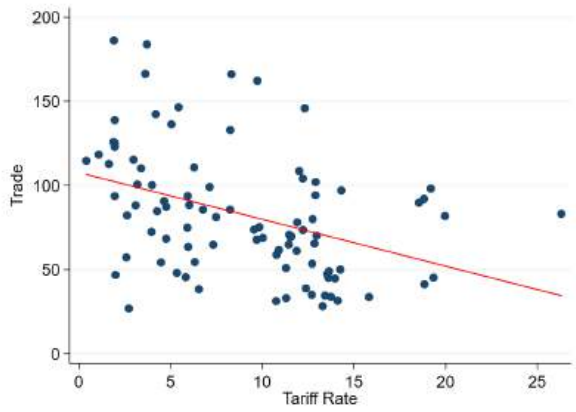
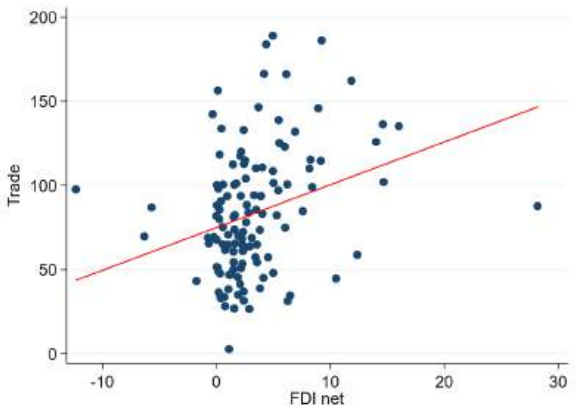
Panel a. Trade vs Tariff Rate	Panel b. Trade vs FDI net
	

Figure 1a shows a negative relationship between trade and tariff rate, suggesting that a higher level of tariff rate contributes to lower trade. Conversely, Figure 1b presents a positive correlation between FDI net and trade, suggesting that a higher level of FDI net contributes to higher trade.

Figure 2. Scatterplots of Trade vs Macroeconomic Indicators

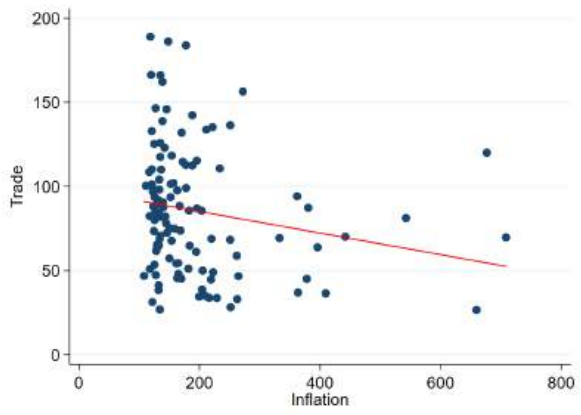
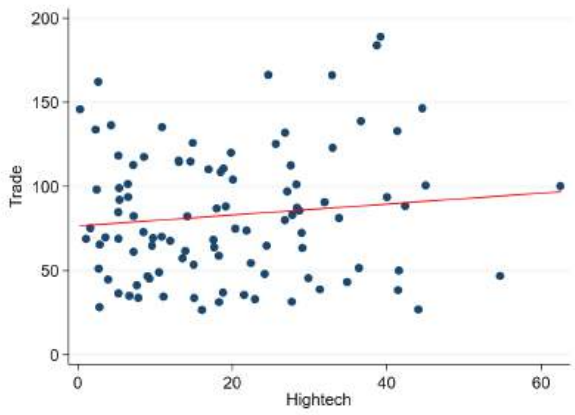
Panel a. Trade vs Inflation	Panel b. Trade vs Hightech
	

Figure 2a displays a negative relationship between trade and inflation, suggesting that higher levels of inflation contributes to lower trade. Meanwhile, high-tech and trade are shown to be slightly positively related in Figure 2b, indicating that higher levels of high-tech leads to higher levels of trade.

Figure 3. Scatterplots of Trade vs Macroeconomic Indicators

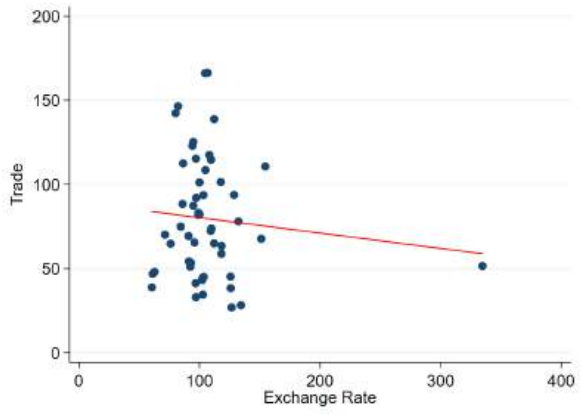
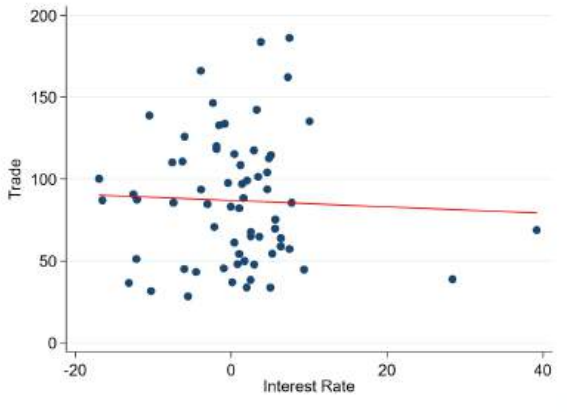
Panel a. Trade vs Exchange Rate	Panel b. Trade vs Interest Rate
	

Figure 3a displays a slightly negative relationship between trade and exchange rate, indicating that higher levels of exchange rate leads to lower levels of trade. Similarly, Figure 3b also displays a slightly negative relationship between trade and interest rate, suggesting that higher levels of interest rate leads to lower levels of trade.

Figure 4. Scatterplots of Trade vs Institutional Indicators

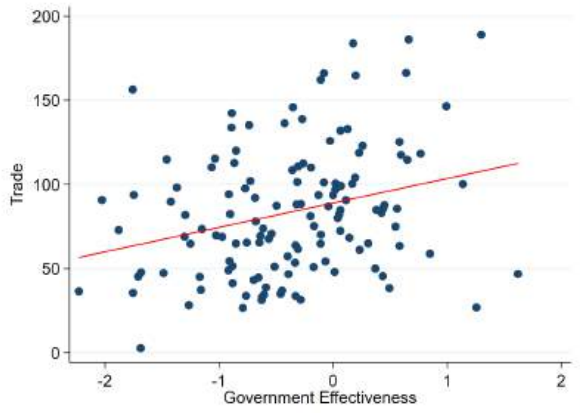
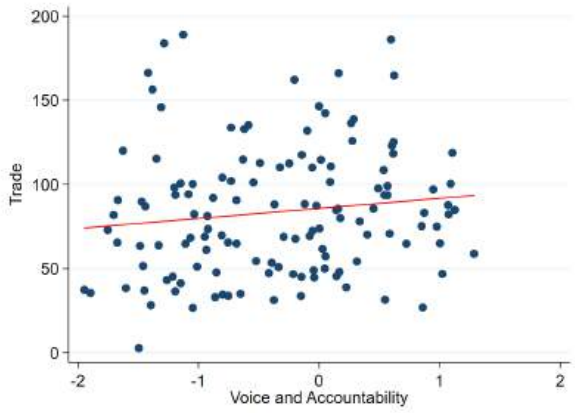
Panel a. Trade vs Government Effectiveness	Panel b. Trade vs Voice and Accountability
	

Figure 4a displays a positive link between trade and government effectiveness, suggesting that the higher level of government effectiveness, the higher the trade. Additionally, Figure 4b shows a similar positive connection between trade and voice and accountability, indicating that the higher level of voice and accountability, the higher the trade.

Figure 5. Scatterplots of Trade vs Institutional and Labour Market Indicators

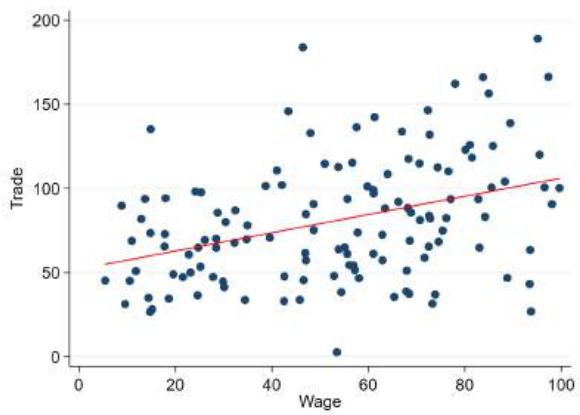
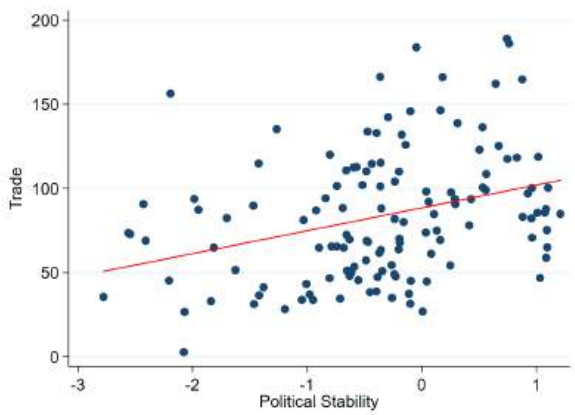
Panel a. Trade vs Wage	Panel b. Trade vs Political Stability
	

Figure 5a displays the positive relationship between trade and wage, indicating that countries with higher wages have higher levels of trade. Finally, Figure 5b demonstrates a positive connection between trade and political stability, supporting the idea that countries that are more politically stable possess higher levels of trade.

Table 1a. presents the result of five regression models. Trade is the dependent variable, tested against the independent variables of tariff rate, FDI net, inflation, political stability, and government effectiveness respectively. Statistically significant results are demonstrated as follows: at one percentage level, we indicated by (***), the 5 percentage level (**), and the 10 percentage level (*).

Table 1a. Regression Analysis for Trade

VARIABLES	(1) trade	(2) trade	(3) trade	(4) trade	(5) trade
Tariff Rate	-2.781*** (0.716)				
FDI net		2.541*** (0.884)			
Inflation			-0.002*** (0.000)		
Political Stability				13.578*** (3.453)	
Government Effectiveness					14.507*** (5.275)
Constant	107.687*** (7.799)	74.975*** (3.910)	86.033*** (3.496)	88.443*** (3.460)	88.990*** (4.076)
Observations	91	127	119	130	130
R-squared	0.159	0.089	0.044	0.107	0.078

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In Model 1 of Table 1a, the tariff rate has a negative and statistically significant relationship with trade, with a coefficient of -2.781. The findings suggest that the higher level of tariff rate is linked to lower levels of trade. In Model 2, FDI net has a positive and statistically significant coefficient of 2.54, indicating that the higher level of foreign direct investment in a country, the higher its trade. Model 3 shows a negative correlation between inflation and trade, with a coefficient of -0.002, implying that higher inflation rates lead to lower trade. In Model 4, political stability has a positive coefficient of 13.578, suggesting that the higher the political stability, the higher the trade. Similarly, Model 5 shows a positive relationship between government effectiveness and trade, with a coefficient of 14.507, highlighting the importance of institutional quality in commerce.

Table 1b. presents the results of five additional regression models, where the dependent variable is trade and the independent variables are voice and accountability, hi-tech, wage, and interest rate respectively.

Table 1b. Regression Analysis for Trade

VARIABLES	(1) trade	(2) trade	(3) trade	(4) trade	(5) trade
Voice and Accountability	5.983				
	(4.072)				
Hi-tech		0.325			
		(0.315)			
Wage			0.542***		
			(0.120)		
Exchange Rate				-0.091	
				(0.073)	
Interest Rate					-0.020
					(0.411)
Constant	85.765***	76.546***	51.972***	89.378***	86.474***
	(3.425)	(6.558)	(6.209)	(10.261)	(4.769)
Observations	130	100	129	52	67
R-squared	0.016	0.013	0.140	0.009	0.000

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Model 1 in Table 1b shows that voice and accountability has a positive but statistically insignificant relationship with trade, having a coefficient of 5.983. This indicates that voice and accountability may not directly influence trade. Model 2 also indicates a positive but insignificant relationship between hi-tech and trade, as supported by the coefficient of 0.325. Conversely, Model 3 shows that wage has a coefficient of 0.542, indicating a strongly positive and statistically significant relationship with trade. Model 4 shows a negative and insignificant relationship between exchange rate and trade, with a coefficient of -0.091. Finally, in Model 5, interest rate has a negative relationship with trade, having a coefficient of -0.020. However, similar to previous examples, this relationship is statistically insignificant.

Conclusion

The importance of trade and macroeconomic indicators have been presented by both previous research and the empirical evidence outlined in this paper. Trade is not only important for exchanging goods and services, but can also support economic growth and strengthen a country's position in the global economy. Hence, it is an important concept for policy makers to understand, aiding in the development of strategies that promote sustainable development, enhance international competitiveness, and effectively respond to global economic shifts.

This paper investigates how macroeconomic, institutional, and labour market factors are related to a country's trade. The main macroeconomic variables include tariff rates, FDI, inflation, exchange rate, hi-tech, and interest rate. The main institutional variables are political stability and government effectiveness, and the common labour market indicator is wage.

The empirical results highlight several important findings, which have been divided into two categories: first are the variables that exhibit strong positive correlations with trade, which are FDI net, political stability, government effectiveness, and wage. Second are the variables that are revealed to have strong negative correlations with trade, which are tariff rate and inflation. The results suggest that governments seeking to expand commercial capacity should not only focus on examining macroeconomic indicators, but also account for institutional and labour factors, which are also shown to be statistically significant variables.

This study can be expanded to explore how macroeconomic indicators and institutional and labour factors influence each other rather than trade. This paper also recommends that future studies employ more observational methods to format and analyze data, including panel data analysis. Additionally, future research may want to tighten the scope of data by specifically analyzing the trade dynamics of a specific country or region to produce more specialized findings and recommendations.

Works Cited

Abdi, A. H., Zaidi, M. A. S., Halane, D. R., & Warsame, A. A. (2024). Asymmetric effects of foreign direct investment and trade openness on economic growth in Somalia: Evidence from a non-linear ARDL approach. Cogent Economics & Finance, 12(1), 2305010.

Bagozzi, B. E., & Landis, S. T. (2015). The stabilizing effects of international politics on bilateral trade flows. Foreign Policy Analysis, 11(2), 151-171.

Cheong, J., & Tang, K. K. (2018). The trade effects of tariffs and non-tariff changes of preferential trade agreements. Economic Modelling, 70, 370-382.

Cong, S., Chin, L., & Kumarusamy, R. (2025). Does trade freedom affect exchange rate movement? A perspective of high-technology trade. The Journal of International Trade & Economic Development, 34(3), 552-572.

Kang, J. W., & Dagli, S. (2018). International trade and exchange rates. Journal of applied economics, 21(1), 84-105.

Kaukab, M. E., & Anggara, A. A. (2024). Does Trade Openness and Inflation Rate Have Dynamic Interconnected Patterns? An Autoregressive Distributed Lag (ARDL) Model Co-integration Approach. Journal of Economics, Finance and Management (JEFM), 3(1), 181-193.

Kennedy, R. (2007). Fragments of economic accountability and trade policy. Foreign Policy Analysis, 3(2), 145-169.

Le, A. N. N., Pham, H., Pham, D. T. N., & Duong, K. D. (2023). Political stability and foreign direct investment inflows in 25 Asia-Pacific countries: the moderating role of trade openness. Humanities and social sciences communications, 10(1), 1-9.

Lee, J. E. (2012). Macroeconomic determinants of the world trade disputes. Applied Economics, 44(33), 4301-4311.

Marjit, S., Beladi, H., & Chakrabarti, A. (2004). Trade and wage inequality in developing countries. Economic Inquiry, 42(2), 295-303.

Min, B. (2003). FDI and Trade. Journal of the Asia Pacific Economy, 8(2), 229-250.

Nam, H. J., & Ryu, D. (2024). Does trade openness promote economic growth in developing countries?. Journal of International Financial Markets, Institutions and Money, 93, 101985.

Nwagu, U. G., Uzonwanne, M., & Ezenekwe, U. (2024). MONETARY POLICY RATE ON TRADE BALANCE IN NIGERIA. Journal of Economic Development, 2(2), 145-161.

Nzama, L., Sithole, T., & Kahyaoglu, S. B. (2022). The impact of government effectiveness on trade and financial openness: the generalized quantile panel regression approach. Journal of risk and financial management, 16(1), 14.

Ravi Kumar, K. N., Reddy, K. G., Shafiwu, A. B., & Mohan Reddy, M. J. (2024). Trade determinants and opportunities for Indian rice: a dynamic panel gravity model perspective. Cogent Economics & Finance, 12(1), 2312367.

Ridwan, M., Tahsin, M. S., Al-Absy, M. S. M., Eleais, M., Ridzuan, A. R., & Mukthar, K. J. (2025). Economic alchemy: Unraveling the nexus between trade openness, inflation,

exchange rates, and economic growth in Bangladesh. International journal of economics and financial issues, 15(3), 244.
Sampson, T. (2014). Selection into trade and wage inequality. American Economic Journal: Microeconomics, 6(3), 157-202.
Wana Ismail, N. (2013). Innovation and high-tech trade in Asian countries. In Proceedings from International Conference on Recent Developments in Asian Trade Policy and Integration (pp. 1-19).
Zhou, S. (1996). Trade connections and interest rate linkages among ASEAN, Japan, and the USA: an empirical analysis. Applied Economics, 28(5), 617-630.

Appendix

Appendix 1. Source of the data

Indicator Name	**Variable**	**Definition**	**Source**
Trade	Trade (% of GDP)	Trade is the sum of exports and imports of goods and services measured as a share of gross domestic product.	WDI: World Bank
Tariff Rate	Tariff rate, applied, simple mean, all products (%)	Simple mean applied tariff is the unweighted average of effectively applied rates for all products subject to tariffs calculated for all traded goods.	WDI: World Bank
FDI Net	Foreign Direct Investment, Net in flow (% of GDP)	Foreign direct investment are the net inflows of investment to acquire a lasting management interest (10 percent or more of voting stock) in an enterprise operating in an economy other than that of the investor.	WDI: World Bank
Inflation	Consumer Price Index (2010 = 100)	Consumer price index reflects changes in the cost to the average consumer of acquiring a basket of goods and services that may be fixed or changed at specified intervals, such as yearly.	WDI: World Bank
Political Stability	Political Stability and Absence of Violence	Political Stability and Absence of Violence/Terrorism measures perceptions of the likelihood of political instability and/or politically-motivated violence, including terrorism, ranging from -2.5 to 2.5.	WDI: World Bank
Government Effectiveness	Government Effectiveness: Estimate	Government Effectiveness captures perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies, ranging from -2.5 to 2.5.	WDI: World Bank
Voice and	Voice and	Voice and Accountability captures perceptions of the	WDI: World Bank

Accountability	Accountability: Estimate	extent to which a country's citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association, and a free media, ranging from -2.5 to 2.5.	
Hi-tech	Medium and Hi-tech Manufacturing (% Manufacturing Value Added)	The proportion of medium and high-tech industry value added in total value added of manufacturing	WDI: World Bank
Wage	Wage and Salaried Workers, Total (% of Total Employment)	Wage and salaried workers (employees) are those workers who hold the type of jobs defined as "paid employment jobs," where the incumbents hold explicit (written or oral) or implicit employment contracts that give them a basic remuneration that is not directly dependent upon the revenue of the unit for which they work.	WDI: World Bank
Exchange Rate	Real Effective Exchange Rate Index	Real effective exchange rate is the nominal effective exchange rate (a measure of the value of a currency against a weighted average of several foreign currencies) divided by a price deflator or index of costs.	WDI: World Bank
Interest Rate	Real Interest Rate (%)	Real interest rate is the lending interest rate adjusted for inflation as measured by the GDP deflator. The terms and conditions attached to lending rates differ by country, however, limiting their comparability.	WDI: World Bank

How does Persuasive Messaging in Chinese EV Companies' Social Media Influence Viewer Engagement? By Yan Yu

1. Introduction

The EV industry in China is rapidly growing with increasing demand. However, many Chinese consumers still avoid purchasing Electric Vehicles (EVs) due to concerns about safety and performance, negative perceptions of EVs, or unwillingness to change their behavior. Yet EVs have many benefits, such as being more environmentally friendly in the long term and having innovative technical functions.

EVs are much more sustainable than their gasoline counterparts. Although they have higher carbon emissions during manufacturing due to battery production, their lifecycle emissions are 73% lower than those of gasoline cars (Negri and Bieker, 2025). Thus, an increase in EV sales will likely lead to more sustainable transportation. In addition to environmental benefits, the EV industry accounts for a substantial portion of the Chinese economy and has made a significant contribution to economic growth. China has the world's largest EV market, and expanding it will likely continue to improve China's economy (Quanshan and Selvaratnam, 2025).

Through an analysis of Xiaomi Auto's BiliBili advertisements and Lei Jun's Weibo page, this study demonstrates that Xiaomi's successful advertising campaign is strongly connected to the promotion of the car's performance capabilities and Lei Jun's authority and idolized social status.

2. Literature Review

2.1 Reviewing prior research

2.1.1 Consumer barriers against EVs

There are many factors that affect consumers' decisions to purchase EVs. Researchers have conducted a questionnaire survey in Shanghai on how various factors influence consumer attitudes towards EVs. Results show that the main buyers of EVs are between 31 and 40 years old, have high income and education, and live in the city center; consumers who avoided purchasing EVs were concerned about the performance of such vehicles (Zhao et al., 2022). An alternative study identified a number of barriers that negatively impact EV purchasing decisions. They identified functional barriers, including usage barriers, or when innovative products do not match existing consumer habits; value barriers, referring to the high prices and performance limitations that EVs possess; and risk barriers, pertaining to the perceived performance and safety risks that consumers hold against EVs. The study also identified tradition as a psychological barrier, meaning the consumer's concern that adopting EVs would disturb their existing lifestyles (Xue et al., 2024). Additionally, frugal consumers sensitive to the quality of a product are more likely to avoid EV purchasing, for the reason that they tend to emphasize the potential drawbacks of EVs, such as battery life and driving range. On the other hand,

ecologically conscious individuals notice more benefits in purchasing EVs thanks to their environmentally friendly characteristics (Wu et al., 2021). Another noticeable hindrance towards EV adoption is consumers' inclinations towards Internal Combustion Engine (ICE) vehicles, which have lower costs, more supporting infrastructure, and longer range; many users also believe that ICEs are more desirable because of their aesthetics, social image, sound, and performance (Krisha, 2021). These studies help us understand the various barriers to EV adoption. There are also noticeable patterns between the sources. Most sources agree to an extent that consumers' concerns towards the performance and safety of EVs negatively influence their decisions to purchase EVs. Some sources also corroborate psychological factors within consumers, such as the risk of change in their lifestyles, their level of environmental consciousness, and their subjective attitudes towards or against EVs. Understanding these key consumer barriers is a vital step in our research, as it helps us analyze how advertisements address these consumer barriers.

2.1.2 Prior intervention attempts

Many studies have been conducted examining messaging techniques to engender consumer purchase intentions for EVs. In a field experiment conducted by Porsche, a German sports vehicle company, researchers tested the effectiveness of different message frames in emails encouraging workers to adopt EVs. The results showed that illustrating financial and cost-saving benefits of EVs had the most powerful impact on consumer decisions (Decrinis et al., 2023). A similar online survey based in Australia found that concrete messages evoking pride in EV ownership have a significant effect on consumers, as they lower perceived functional barriers and risks (Mulcahy et al. 2025). The two studies both corroborate the conclusion that messages lowering perceived risks – whether related to finances, function, or safety – contribute to a positive effect on purchase intentions. In other words, intangible barriers, such as perceived risks and benefits, also pose a significant hurdle to EV adoption among consumers. The literature suggests that consumers face a dual barrier when deciding to purchase EVs: they must overcome their tangible performance and safety concerns, while their ingrained psychological biases amplify or downplay their concerns. Thus, analysing the effectiveness of different nudges to change consumers' negative attitudes towards EVs and designing an effective nudge will significantly improve consumer decisions towards buying EVs.

2.1.3 Marketing techniques in EV companies

Understanding the marketing strategies in EV manufacturers provides critical information to their brand and target audience. Xiaomi, for example, mainly targets relatively young users looking for innovative technology. For this reason, Xiaomi's cars emphasize function and performance, including high-density batteries with longer lifespans, autonomous driving systems, and large horsepower. Furthermore, Xiaomi also targets its existing fanbase, which it built from selling cheap and reliable appliances such as air purifiers and smartphones. This fanbase has a strong loyalty to Xiaomi, making it one of the brand's main audiences (Zheng,

2025). Although not directly connected to persuasive messaging, knowing how Xiaomi persuades its audience to buy its products is important because it provides context as to how its advertisements can synthesize with the brand's principles, key messages, and design.

2.2 Research Objectives

Although there are many studies aiming to address these barriers by examining holistic marketing plans, the importance of marketing interventions and nudges remains overlooked in this field. The main gap lies in analyzing the effectiveness of different EVs' advertising methods on consumer purchase intentions, which could serve as a crucial deciding factor for EV adoption. Our research will build upon the existing studies by providing a concise and accurate summary of the key barriers against EV adoption and addressing the gap in the effectiveness of advertising techniques. By comparing the engagement data from various advertisements and posts on Xiaomi's social media, we will be able to categorize the advertisements and provide statistically accurate conclusions on the most effective method. This way, we will be able to design an advertisement that nudges consumers to have a positive attitude towards EVs, which will decrease their perceived risks and ultimately increase their chances of purchasing EVs. If EV companies use effective nudges in their marketing, our research will have significant benefits to China's economy and sustainability. This is because effective marketing and persuasive messaging on a large scale will increase EV sales in Chinese EV companies domestically and internationally, leading to economic growth in China and less carbon emissions in the long term.

3. Methods

This study examines the use of persuasive techniques in advertisements targeting consumers' decisions to purchase electric vehicles (EVs). Specifically, we analyze how various marketing strategies are employed to influence consumer attitudes and behaviors. The research uses a content analysis approach, systematically coding advertisements to identify recurring persuasive elements.

3.1 Advertisement Sample & Data Collection

A sample of 30 advertisements was collected from Xiaomi Auto's BiliBili page between 2023 and 2025. These advertisements were selected based on the following inclusion criteria:

- The ad explicitly promoted emotional appeal, aesthetics, performance, function, authority, or environment.
- The ad was published within the last two years to ensure relevance.
- Engagement data (e.g., likes, shares, views) was accessible.

During our examination of Xiaomi's BiliBili content, we realized that videos featuring the presence of Lei Jun (CEO of Xiaomi) amassed far more engagement data compared to others.

Thus, we collected 20 posts from Lei Jun's Weibo from February to July 2025 to analyze how he uses social media to gain influence.

The final sample consisted of 50 ads/posts spanning 2 different platforms and representing advertisements from Xiaomi's auto industry. The full list of advertisements analyzed, including URLs and coding metrics, can be found in Appendices A and B.

3.2 Coding Framework & Variables

To analyze the persuasive techniques in the advertisements from Xiaomi's BiliBili, a coding framework was developed based on prior literature on advertising and consumer persuasion. The coding scheme included the following categories:

- Emotional Appeal
- Aesthetics
- Performance
- Function
- Authority Figures
- Environment

We also developed a separate coding framework for Lei Jun's Weibo, as his page focuses on his holistic social image rather than Xiaomi EVs. This coding scheme included the following groups:

- Achievement
- Call to Action
- Nationalism
- Personal Life
- Kindness
- Gratitude

3.3 Coding Approach: Binary vs. Likert Scale

To measure the presence and intensity of persuasive techniques, two different coding approaches were used:

- **Binary Coding (Yes/No):** Some categories were coded dichotomously, with a value of **1** indicating the presence of the persuasive technique and a value of **0** indicating its absence.
- **Likert Scale (1-5):** Other categories were coded using a **Likert scale** to assess the strength of the technique within each advertisement:
    - 1 = Not present at all
    - 2 = Weak presence
    - 3 = Moderate presence
    - 4 = Strong presence
    - 5 = Dominant feature of the ad

The choice of coding scheme was determined based on the media platform and the complexity of each persuasive technique. For example, emotional appeal was measured using a Likert scale as different advertisements evoke varying levels of intensity. On the other hand, the environment metric was measured on a binary scale, as the amount of environmental presence within an ad is subjective. All posts on Lei Jun's Weibo are measured using a binary scale because Weibo posts include much less information, and therefore, it is harder to assess the intensity of each metric.

3.4 Engagement Metrics

For advertisements where engagement data was available, the following metrics were recorded:

- Views: Total number of times the ad was watched.
- Likes: Number of user interactions expressing approval.
- Shares: Frequency with which users shared the ad.
- Saves/Favorites: Number of users who saved the ad to their media collection.
- Comments: User-generated discussions related to the advertisement.
- Reposts (Weibo only): Frequency with which users reposted a post onto their own page

Engagement data was extracted and analyzed to identify trends in user response to different persuasive techniques.

Refer to Table 1 and Table 2 below for a comprehensive coding scheme for Xiaomi Auto and Lei Jun, respectively, including definitions and examples**.** See the appendix for official coding.

Table 1: Coding Scheme (BiliBili - Xiaomi Auto)

The table below provides a detailed overview of the coding scheme used for ads from Xiaomi Auto, including definitions, examples, and coding format (Binary: 0/1 or Likert Scale: 1-5).

Category	Definition	Example	Coding Format
Emotional Appeal	The ad evokes emotions such as suspense, happiness, excitement.	Quick cuts of a car zooming past the camera to evoke excitement.	Likert (1-5)
Aesthetics	The ad focuses on the exterior or interior look of the car.	An advertisement with close-up shots on the exterior of a car.	Likert (1-5)
Performance	The ad emphasizes the capabilities of the car, including speed and durability.	A video showing a car's top speed.	Likert (1-5)
Function	The ad highlights the UX/Comfort features of the car.	An advertisement showcasing a car's reclining seat.	Likert (1-5)
Authority	The ad uses a driver, scientist, or celebrity to add credibility.	"9 out of 10 drivers recommend this car."	Likert (1-5)

Environment	The ad presents the product's sustainability or connection to the environment.	An electric car driving in a clear environmental landscape.	Binary (0-1)

Table 2: Coding Scheme (Weibo – Lei Jun)

The table below provides a detailed overview of the coding scheme used for posts from Lei Jun, including definitions, examples, and coding format.

Category	Definition	Example	Coding Format
Achievement	The post showcases a milestone, award, or statistic illustrating Xiaomi's success.	"Xiaomi SU7 sales hit 50000 units."	Binary (0-1)
Call to Action	The post reminds or urges users to purchase Xiaomi EVs.	"Pre-order sales close by 8PM. Order now!"	Binary (0-1)
Nationalism	The post appeals to users' identification and support for China.	A post mentioning Xiaomi's contribution to China's EV market.	Binary (0-1)
Personal Life	The post focuses on moments in Lei Jun's life.	A post of Lei Jun going to the gym.	Binary (0-1)
Kindness	The post publicizes Lei Jun's help to the public or people in need.	A post of a donation certificate to a charitable organization.	Binary (0-1)
Gratitude	The post expresses thankfulness and appreciation to clients, partners, or organizations.	"Thank you to everyone for helping the Xiaomi SU7 reach 50000 units.	Binary (0-1)

4. Results

4.1 Data Analysis

To examine the use of persuasive strategies across advertisements, a multi-step analytical approach was used. Julius AI was employed to assist with all data exploration, statistical testing, and pattern detection.

Descriptive statistics were first calculated to summarize the overall prevalence and intensity of persuasive techniques across the ad sample. Comparative analyses were then conducted to determine whether the use of these techniques varied by key categorical factors such as platform or ad format. Finally, predictive analyses were used to identify which persuasive strategies were most associated with audience interaction.

- Descriptive Analysis: Mean scores (or frequency counts) were calculated for each persuasive technique to identify the most and least prevalent strategies used across the sample.

- Comparative Analysis: Differences in persuasive technique usage were examined across variables such as platform or target behavior using statistical tests, including chi-square or ANOVA, depending on how the data were coded.
- Predictive Analysis: Regression or correlation analyses were conducted to assess which persuasive techniques were most strongly associated with higher audience engagement.
- Qualitative Insights: In some cases, thematic patterns in ad visuals or messaging were noted to add context to the statistical results, offering a deeper understanding of how persuasive techniques were applied.

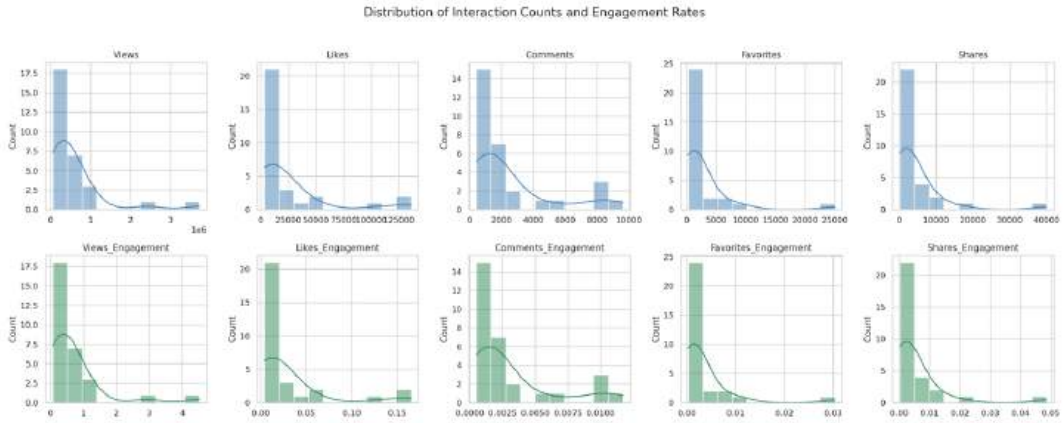
4.2 Descriptive Overview – Xiaomi BiliBili

A total of 30 Xiaomi advertisements were analyzed, drawn from BiliBili. Thirty-seven percent are for the SU7 model, 47% for the SU7 Ultra, and 20% for YU7. The rest are not focused on any model. Xiaomi has 823,000 followers. The descriptive statistics for all advertisement engagement metrics can be found in Table 3. On average, an advertisement has about 567,600 views, 24,880 likes, 2,484 comments, 2,726 favorites, and 4,479 shares. Normalizing for follower amount, an ad earns about .69 views per follower, .03 likes per follower, .003 comments per follower, .0003 favorites per follower, and .005 shares per follower. The distributions for each of these variables can be found in Figure 1. As can be seen, all five metrics are sharply right-skewed and attract modest numbers, while there are a few outliers.

Table 3. Descriptive Statistics of Xiaomi's Bilibili Advertisements

	count	mean	std	min	max
Views	30	567600	748495.2	57000	3715000
Likes	30	24880.9	36223.71	3450	136000
Comments	30	2484.033	2675.143	408	9641
Favorites	30	2726.2	4766.088	256	25000
Shares	30	4479.5	7726.601	194	40000
Views_Engagement	30	0.69	0.909	0.069	4.514
Likes_Engagement	30	0.03	0.044	0.004	0.165
Comments_Engagement	30	0.003	0.003	0	0.012
Favorites_Engagement	30	0.003	0.006	0	0.03
Shares_Engagement	30	0.005	0.009	0	0.049

Figure 1. Distribution of Engagement Data in Xiaomi's Bilibili Advertisements



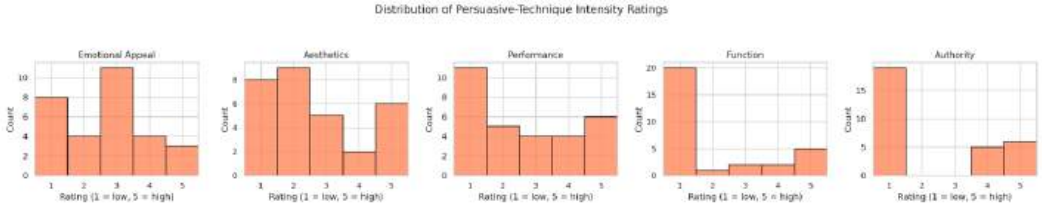
4.2.1 Persuasive Techniques

To begin the analysis, I examined the frequency of the use of environmental benefits. Most ads did not have an explicit environmental benefit message (87%). Next, I examined the average intensity of each persuasive technique (coded on a continuous scale) across all advertisements. See Table 4 for the results. The results reveal a balanced, but moderate use of the tactics, such that emotional appeals average a 2.7, and aesthetics and performance are both around 2.6. Function is the least utilized technique at a mean of 2, while authority averages a 2.3. Figure 2 provides the distributions for each of these techniques. For emotional appeal, scores tended to skew toward the low-to-moderate ratings. Aesthetics was similar but also had a number of ads given 4's or 5's. Performance was evenly distributed among 2-5's, but had a large number of ads given a 1. Function and authority were mostly dominated by scores of 1.

Table 4. Presence of Persuasive Techniques in Xiaomi's Bilibili Advertisements

	count	mean	std	min	max
Emotional Appeal	30	2.67	1.3	1	5
Aesthetics	30	2.63	1.47	1	5
Performance	30	2.63	1.59	1	5
Function	30	2.03	1.61	1	5
Authority	30	2.3	1.76	1	5

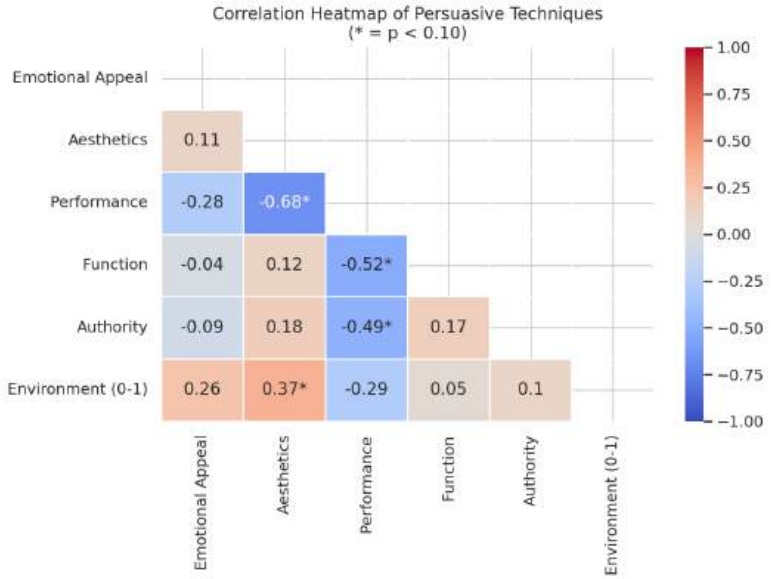
Figure 2. Distribution of Persuasive Techniques in Xiaomi's Bilibili Advertisements



Next, I correlated all of the persuasive techniques to see how they occur together in advertisements. See Figure 3 below or a representation of these results. In this figure, a heatmap provides the Pearson correlation coefficient for each relationship. If the relationship is significant at a p<.10 value, an asterisk appears next to the correlation coefficient.

As can be seen in the figure, performance is negatively correlated with aesthetics, function, authority, and environment, such that as the use of performance appeals increases, the use of the other appeals decreases. Additionally, there is a statistically significant relationship between environment and aesthetics such that aesthetic appeals are more likely to occur when an environmental appeal is present. Taken together, there are two apparent clusters; the first is a visually rich, eco-focused ad versus a fact and performance-driven message.

Figure 3. Correlation Heatmap of Persuasive Techniques in Xiaomi's Bilibili Advertisements



4.2.2 Predictive Insights

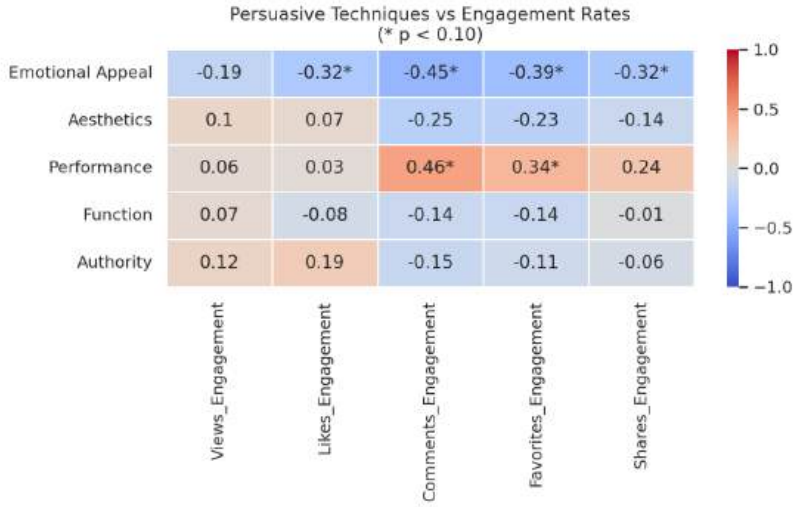
To examine whether specific persuasive strategies influenced audience engagement, we conducted predictive analyses. First, we ran several t-tests comparing the mean scores on all engagement variables when environmental benefits were included versus not included. See Table 5 for the results. This means the environmental present groups appear much larger across all engagement metrics; however, the huge variability and the small sample size for the environment not-present group affected the significance level. With larger sample sizes, we may see this trend become significant. When outliers were removed from this analysis, the results remained the same.

Table 5. T-Tests of Engagement Variables in Xiaomi's Bilibili Advertisements

Variable	n(Env=0)	n(Env=1)	Mean0	Mean1	t	p
Views	26	4	484384.6	1108500	-0.71	0.528
Likes	26	4	22798.46	38416.75	-0.471	0.668
Comments	26	4	2422.231	2885.75	-0.256	0.812
Favorites	26	4	2795.192	2277.75	0.275	0.794
Shares	26	4	4350	5321.25	-0.23	0.83
Views_Engagement	26	4	0.589	1.347	-0.71	0.528
Likes_Engagement	26	4	0.028	0.047	-0.471	0.668
Comments_Engagement	26	4	0.003	0.004	-0.256	0.812
Favorites_Engagement	26	4	0.003	0.003	0.275	0.794
Shares_Engagement	26	4	0.005	0.006	-0.23	0.83

Next, we correlated the continuously coded persuasive techniques with the engagement metrics. See results in Figure 4. The results demonstrate an interesting story where emotional appeals hurt engagement all around, with negative correlations with likes, comments, favorites, and shares per follower. Performance appeals however, help engagement with a positive association with comments and favorites per follower.

Figure 4. Correlation between Persuasive Techniques and Viewer Engagement in Xiaomi's Bilibili Advertisements



Finally, multiple regressions were run with all six persuasive messaging techniques as predictors of the engagement metrics. In these analyses, only the model for comments per follower was significant (p=.05). When examining this model, emotional engagement was the only predictor, having a negative impact on comments per follower. Every one-point increase in

the emotional appeal rating (say, moving from 2 → 3 on the 0-4 scale) is associated with a 0.001 drop in the comments-per-follower rate.

4.2.3 Qualitative Insights

In addition to the coded variables for BiliBili content, we observed several qualitative patterns:

- **Lei Jun's Presence (CEO of Xiaomi):** Ads featuring Xiaomi CEO Lei Jun generated significantly more raw engagement data compared to other persuasive techniques.
- **Simple Spotlight:** Xiaomi uses several short advertisements, each emphasizing one important function, such as the voice control feature and the artificial engine sound, or one performance metric, including an acceleration and a steering test.
- **Tone and Style:** Many ads used emotional music and dramatic shots to create short documentary-style videos, often highlighting a specific aspect of the car or its overall appeal.
- **Environmental Appeal:** Ads with the "environment" metric present tended to focus on the presence of the environment in the video rather than mentioning the car's sustainability efforts.
- **Connection to Audience:** Certain ads were designed to appeal to specific audiences by resonating with their demographics or nationality. This included an ad telling the story of a female engineer at Xiaomi, and a video of the YU7 driving through iconic Chinese roads.

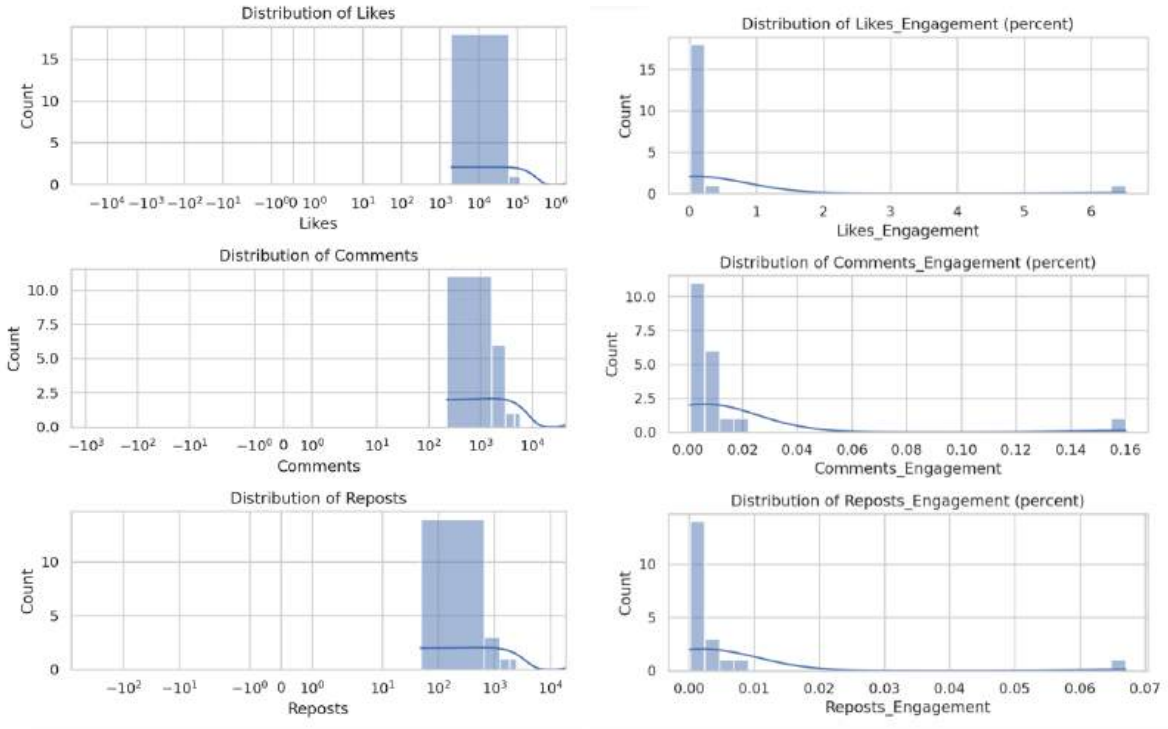
4.3 Descriptive Overview – Lei Jun's Weibo

A total of 20 Weibo posts from Lei Jun were collected. Lei Jun has 26,846,000 followers. The descriptive statistics for all engagement data can be found below in Table 6. The average post has around 107,033 likes, 3,727 comments, and 1,460 reposts, which converts to around 3.987E-03 likes per follower, 1.388E-04 comments per follower, and 5.438E-05 reposts per follower. The distributions of this data are shown below in Figure 5. All metrics are skewed to the right, some including a few outliers.

Table 6. Descriptive statistics in Lei Jun's Weibo

Variable	count	mean	std	min	max
Likes	20	107033.3	387199.242	1945.0	1750000.0
Comments	20	3727.1	9312.7730	225.0	43000.0
Reposts	20	1459.95	3934.436	49.0	18000.0
Likes_Engagement	20	3.987E-03	1.442E-02	7.245E-05	6.519E-02
Comments_Engagement	20	1.388E-04	3.469E-04	8.381E-06	1.602E-03
Reposts_Engagement	20	5.438E-05	1.466E-04	1.825E-06	6.705E-04

Figure 5. Distribution of engagement data in Lei Jun's Weibo Posts



4.3.1 Persuasive Techniques

We analyzed the frequency of our six binary variables in Lei Jun's posts, as shown in Table 7. The frequency of most variables was relatively similar, being around 15–25%. The achievement metric, however, was present 9 times, accounting for 45% of the time. Personal life was the metric with the lowest frequency, with a count of 3, equaling 15% of the data.

Table 7. Frequency of Persuasive Techniques in Lei Jun's Weibo Posts

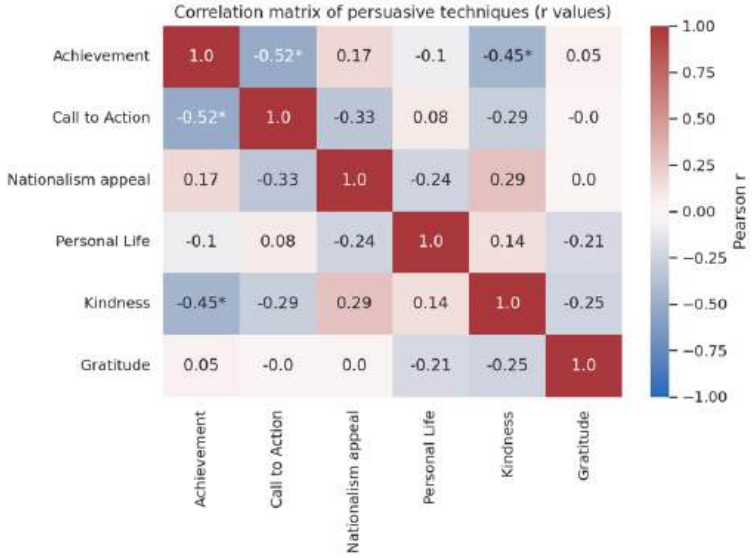
Technique	Count	Percentage
Achievement	9	45%
Call to Action	5	25%
Nationalism Appeal	5	25%
Personal life	3	15%
Kindness	4	20%
Gratitude	4	20%

Again, we correlated the techniques to visualize their occurrence together in Lei Jun's posts. The results are represented as a heatmap in Figure 6 below, providing the Pearson correlation coefficient for each relationship. Relationships with a $p < 0.10$ value are marked with asterisks.

The heat map shows that achievement has statistically significant negative correlations with call to action and kindness, such that as achievement is present, the other two techniques are not. This suggests that achievement is more often present by itself than in combination with

others. However, other relationships are statistically insignificant and do not provide detailed insights.

Figure 6. Correlation Heatmap of Persuasive Techniques in Lei Jun's Weibo Posts



4.3.2 Predictive Insights

Then, we correlated the presence of each persuasive technique with the engagement metrics, as shown in Figure 7. The results show a positive, statistically significant correlation between kindness and the engagement variables, and little to no correlation between the other persuasive techniques. However, when we repeated the calculations excluding outliers (see Figure 8), the outcome was different. The only significant relationship was a negative correlation between call to action and likes, and achievement and gratitude had small correlations with likes.

Figure 7. Correlations between Techniques and Viewer Engagement in Lei Jun's Weibo Posts

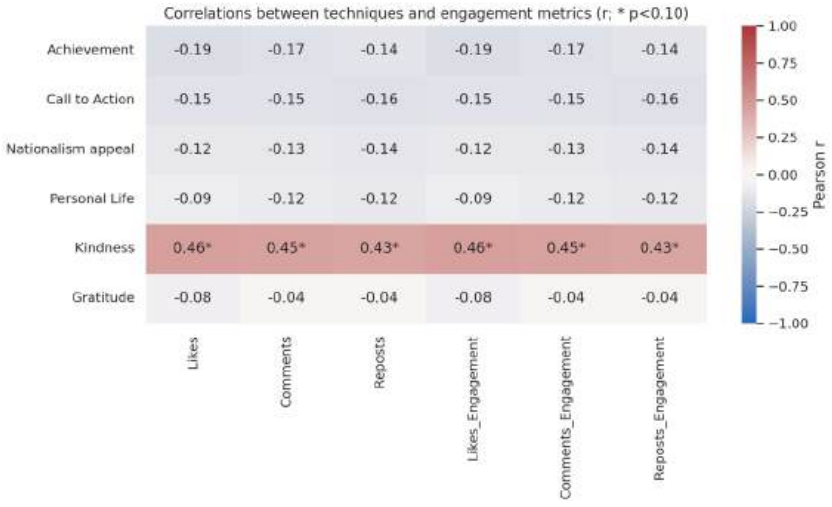
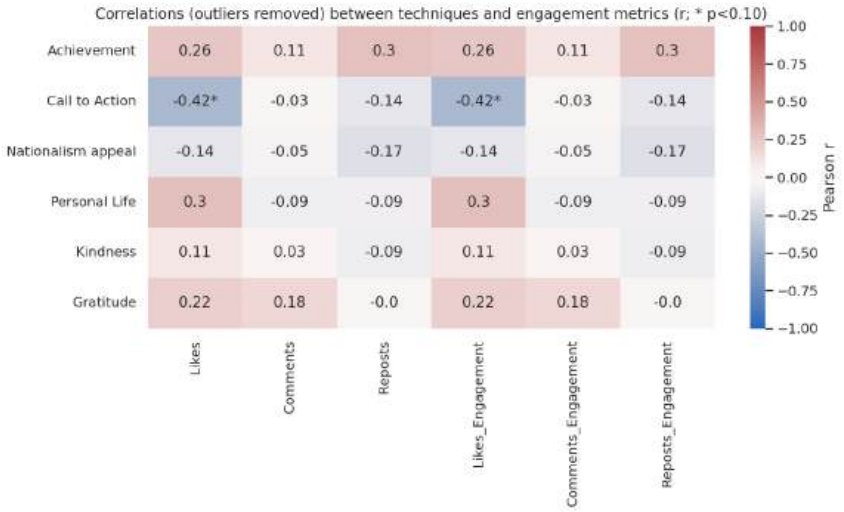


Figure 8. Correlations between Techniques and Viewer Engagement in Lei Jun's Weibo Posts With Outliers Removed



4.3.3 Qualitative Insights

We also noticed qualitative patterns throughout Lei Jun's Weibo posts:

- **Personal Involvement:** Many posts, including posts without mentions to personal life, have Lei Jun's personal opinions, involvement, or experiences with Xiaomi products.
- **Nationalist Appeal:** Lei Jun frequently includes short phrases in his posts that praise or encourage China, as well as describe Xiaomi's efforts to support the Chinese economy.
- **Emphasizing Statistics:** Several posts spotlight a ranking, such as the SU7's grade in a safety test, and statistics, including sales milestones or car performance data.

4. Discussion

4.1 Summary of Results

Understanding the driving incentives and barriers towards EV adoption has been a topic of great interest in recent years. This paper contributes to the literature by analyzing which persuasive messaging frames in social media engender viewer engagement. In Xiaomi's BiliBili advertisements, most techniques are moderately used, with emotional appeals being the most common, followed by aesthetics and performance. Our data shows, however, that emotional appeals hurt engagement, and most other techniques have little to no correlation with engagement data; in contrast, performance appeals showed a strong positive relationship with engagement and generated promising and significant numbers – namely, more comments, shares, and favorites.

4.2 Connection to Previous Literature

Our finding that performance appeals corroborate more engagement data aligns with previous literature, as multiple researchers have identified concerns towards EV performance and safety as a significant barrier against consumer decisions. Thus, advertisements downplaying these concerns and highlighting technical capabilities generate higher viewer engagement as they mitigate a key consumer concern. However, the negative relationship between emotional appeals and engagement data differs from previous research, as various sources identify emotional appeal to have a positive effect in advertisements. This unexpected result could be due to several factors: for one, BiliBili's relatively young user base may be less emotionally moved by these appeals, and could therefore be less engaged with the video; another possible reason is that emotional appeal aspects in advertisements overshadow the technical specs or performance of the car, a key incentive for purchasing Xiaomi cars; finally, many other EV companies, such as Li Auto and BYD, employ similar emotional appeal approaches, so ads that focus on this technique may lack distinctiveness.

4.3 Limitations

Although our efforts have led to valuable insights on EV advertising, there are still various limitations within our work. The most obvious shortcoming is the lack of sample size. With a total of twelve persuasive techniques to analyse, a sample of 50 advertisements may not be enough to provide compelling evidence and reveal trends. Thus, many of our identified correlations are not statistically significant. To add, our team consisted of only one coder, so only specific categories were coded for; there are many other categories that were left out, such as personal gain/loss, social proof, and comparison with competitors. Aside from that, our sample only focuses on Xiaomi, so we are uncertain if our findings can be applied to other companies. We encourage future researchers to build on our work by addressing these limitations. With a larger team of researchers, a similar study could be conducted on a larger scale with more persuasive techniques, advertisements, and a selection of companies. This would provide more definitive results and offer insights into how the effectiveness of techniques varies across different EV companies.

5. Designing a High-Engagement Advertisement

Using our findings from the study, we created an advertisement by collecting clips from Xiaomi Bilibili advertisements and editing them into a montage. Namely, the strong positive association between performance and viewer engagement, the negative correlation with generic emotional appeals, and Lei Jun's strong influence as an authority figure. This 30-second video, designed to be released on Xiaomi's Bilibili and advertised on television and streaming platforms, targets Xiaomi's intended audience of young consumers interested in innovative technology and people concerned with EV performance and safety. The full video can be found in this Youtube Link: https://youtu.be/_llEy7QLYUA. See Table 8 for an explanation of the elements included in the advertisement.

Table 8. Application of Research Findings to Advertisement Design

Research Finding	Application	Intended Outcome	Frame
Generic emotional appeals showed a negative correlation with engagement data.	Instead of using emotional appeal as a dominant technique, we used it as a supplementary feature to emphasize the car's performance capabilities. The ad opens with a montage of a driver flooring the throttle pedal, accelerating the car	To evoke feelings of exhilaration in the viewer, strengthening their perception of the car's performance and speed.	 
Performance appeals showed a strong positive correlation with viewer engagement	The ad includes a shot displaying the top speed of the Xiaomi SU7 and a safety test of the car treading through water unscathed.	To address consumer concerns of EV performance and safety by demonstrating the shocking technical capabilities of the Xiaomi SU7.	 
Lei Jun's presence in advertisements generated significantly more engagement data.	The ad ends with a shot of Lei Jun drifting the Xiaomi SU7 drifting. As he steps out of the car, takes off his helmet, revealing his face.	Using is powerful authority presence can generate a significant amount of engagement data.	 

6. Conclusion

Based on the results of our findings, we suggest that Chinese EV companies include the following elements in their advertising:

1. Showcase the vehicle's technical capabilities and/or safety considerations. Many consumers avoid EV purchase due to concerns about EV performance and safety risks.
2. Use authority or celebrity figures in advertisements to boost the vehicle's credibility and reputation. Influential figures help engender demand for products.

Ultimately, this analysis of Xiaomi's advertisements provides valuable insights into the effectiveness of persuasive techniques in Chinese EV markets. Effectively advertising EVs could be a crucial factor driving future EV adoption, improving China's sustainability and economy.

Works Cited

Negri, M., & Bieker, G. (2025). Life-cycle greenhouse gas emissions from passenger cars in the European Union: A 2025 update and key factors to consider. (n.d.). International Council on Clean Transportation. Retrieved August 19, 2025, from https://theicct.org/publication/electric-cars-life-cycle-analysis-emissions-europe-jul25/

Quanshan, G., & Selvaratnam, D. P. (2025). The impact of new energy vehicles on china's economic development. International Journal of Research and Innovation in Social Science, IX(II), 2427–2449. https://doi.org/10.47772/IJRISS.2025.9020191

Zhao, X., Ma, Y., Shao, S., & Ma, T. (2022). What determines consumers' acceptance of electric vehicles: A survey in Shanghai, China. Energy Economics, 108, 105805. https://doi.org/10.1016/j.eneco.2021.105805

Xue, Y., Zhang, X., Zhang, Y., & Luo, E. (2024). Understanding the barriers to consumer purchasing of electric vehicles: The innovation resistance theory. Sustainability, 16(6), 2420. https://doi.org/10.3390/su16062420

Wu,D.; Yu,L.; Zhang,Q.; Jiao, Y.; Wu, Y (2021). Materialism, Ecological Consciousness and Purchasing Intention of Electric Vehicles: An Empirical Analysis among Chinese Consumers. https://doi.org/10.3390/su13052964

Krishna, G. (2021). Understanding and identifying barriers to electric vehicle adoption through thematic analysis. Transportation Research Interdisciplinary Perspectives, 10, 100364. https://doi.org/10.1016/j.trip.2021.100364

Decrinis, L., Freibichler, W., Kaiser, M., Sunstein, C. R., & Reisch, L. A. (2023). Sustainable behaviour at work: How message framing encourages employees to choose electric vehicles. Business Strategy and the Environment, 32(8), 5650–5668. https://doi.org/10.1002/bse.3441

Mulcahy, R., Fleischman, D., De Villiers Scheepers, M. J., & Reichenbach, N. (2025). Proud to go electric: Overcoming perceived functional barriers to EV adoption through congruent messaging frames. Transportation Research Part F: Traffic Psychology and Behaviour, 114, 1–15. https://doi.org/10.1016/j.trf.2025.05.019

Zheng, X. (2025). Analysis of new energy vehicle marketing strategy—Taking xiaomi as an example. SHS Web of Conferences, 218, 01008. https://doi.org/10.1051/shsconf/202521801008

Appendices

Appendix A: Table of Collected Bilibili Advertisements

Ad Link	Source	Car Model	Followers	Views	Likes	Comments	Favorites	Shares	Emotional Appeal	Aesthetics	Performance	Function	Authority	Environment (0-1)
Hello, Xiaomi SU7!	Bilibili	Xiaomi SU7	823,000	508,000	10,000	1,646	1,296	3,073	3	5	1	1	1	1
Meet Xiaomi SU7 Ultra	Bilibili	Xiaomi SU7 Ultra	823,000	139,000	11,000	968	844	737	3	5	2	1	1	0
Xiaomi SU7 Drift Mode	Bilibili	Xiaomi SU7	823,000	140,000	3,718	408	439	300	3	2	2	4	1	0
Xiaomi YU7 | Debut	Bilibili	Xiaomi YU7	823,000	165,000	3,840	958	352	1,214	2	5	3	1	1	0

Xiaomi SU7 Ultra prototype	Official certified video from the Nürburgring		823,000	2,398,000	134,000	9,641	25,000	40,000	1	1	5	1	1	0
Xiaomi SU7 Ultra prototype	Official uncut Nürburgring footage	Bilibili	Xiaomi SU7 Ultra	823,000	847,000	45,000	8,651	9,806	9,932	1	1	5	1	1	0
Xiaomi YU7 Unboxing with Perry		Bilibili	Xiaomi YU7	823,000	131,000	5,997	1,279	1,570	2,361	1	3	1	5	4	0
Xiaomi SU7 Ultra "Space Silver"		Bilibili	Xiaomi SU7 Ultra	823,000	765,000	46,000	2,099	3,157	6,181	3	5	1	1	1	0
The Sound of Xiaomi SU7 Ultra		Bilibili	Xiaomi SU7 Ultra	823,000	433,000	4,522	425	721	391	4	3	2	5	1	0
Xiaomi SU7 Design Philosophy		Bilibili	Xiaomi SU7	823,000	154,000	4,217	1,235	484	958	3	4	1	3	4	1
Xiaomi SU7 Ultra	Official uncut Nürburgring footage	Bilibili	Xiaomi SU7 Ultra	823,000	685,000	37,000	8,110	6,977	7,975	1	1	5	1	1	0
Kazunori Yamauchi #XiaomiSU7Ultra test-drive		Bilibili	Xiaomi SU7 Ultra	823,000	360,000	13,000	1,826	1,725	1,718	2	2	3	1	5	0
Xiaomi YU7 24-Hour Endurance Challenge		Bilibili	Xiaomi YU7	823,000	202,000	5,999	1,659	639	503	3	2	4	1	1	0
Xiaomi YU7 Design Philosophy		Bilibili	Xiaomi YU7	823,000	146,000	6,071	1,051	617	623	3	5	1	1	4	0
Xiaomi SU7 Ultra Lap Record Zhejiang		Bilibili	Xiaomi SU7 Ultra	823,000	301,000	12,000	2,443	1,041	1,445	1	1	5	1	1	0
Xiaomi SU7 Ultra Lap Record Shanghai		Bilibili	Xiaomi SU7 Ultra	823,000	704,000	23,000	5,182	2,601	5,513	1	1	5	1	1	0
Xiaomi YU7 Drives on Three Major National Roads		Bilibili	Xiaomi YU7	823,000	57,000	3,450	598	275	254	5	2	3	1	1	1
Documentary on Xiaomi Female Engineers		Bilibili	N/A	823,000	262,000	29,000	1,169	2,284	1,402	5	1	1	1	4	0
Xiaomi SU7 Ultra	Extreme Battery Test	Bilibili	Xiaomi SU7 Ultra	823,000	366,000	14,000	1,274	1,520	1,070	3	1	5	1	1	0
Xiaomi SU7 Ultra	Media Review	Bilibili	Xiaomi SU7 Ultra	823,000	144,000	4,623	499	256	194	2	3	2	2	5	0
Lei Jun Drifts the Xiaomi SU7		Bilibili	Xiaomi SU7	823,000	137,000	7,720	550	888	2,774	5	2	2	1	5	0
30 Seconds of Xiaomi's Powerful Chassis		Bilibili	Xiaomi SU7	823,000	307,000	8,086	1,222	977	2,201	3	2	1	5	1	0
Entering the "Two-Second Club"		Bilibili	Xiaomi SU7	823,000	422,000	11,000	1,567	1,273	3,335	4	2	4	1	1	0

Xiaomi SU7 Ultra | Top speed test	Bilibili	Xiaomi SU7 Ultra	823,000	152,000	3,895	719	423	275	4	2	4	1	1	0
Xiaomi EV Steering Test	Bilibili	Xiaomi SU7	823,000	1,104,000	16,000	4,391	1,185	4,913	4	2	4	1	1	0
“Hey, Siri" Voice control	Bilibili	Xiaomi SU7	823,000	505,000	15,000	2,923	1,922	11,000	2	3	1	5	4	0
Xiaomi SU7's Nine Colors - Which is your favorite?	Bilibili	Xiaomi SU7	823,000	941,000	106,000	1,400	3,849	2,059	1	4	1	1	5	0
Lei Jun Showcases the Xiaomi SU7	Bilibili	Xiaomi SU7	823,000	3,715,000	136,000	8,064	7,056	17,000	3	5	1	4	5	1
Lei Jun | Beijing to Shanghai Driving Test	Bilibili	Xiaomi YU7	823,000	478,000	17,000	1,025	1,133	1,851	1	1	3	3	5	0
Xiaomi's Smart Cabin Awaits You.	Bilibili	Xiaomi SU7	823,000	360,000	9,289	1,539	1,476	3,133	3	3	1	5	1	0

Appendix B: Table of Collected Weibo Posts

Ad Link	Content Type	Followers	Likes	Comments	Reposts	Achievement	Call to Action	Nationalism appeal	Personal Life	Kindness	Gratitude
https://weibo.com/1749127163/PDAAh7wIE	Photos	26,846,000	1,945	225	49	0	0	0	0	0	0
https://weibo.com/1749127163/PDw3AazFP	Video	26,846,000	2,906	321	102	0	0	1	0	1	0
https://weibo.com/1749127163/PCJY9CTaY	Photos	26,846,000	4,973	1,334	188	0	1	0	0	0	0
https://weibo.com/1749127163/Pli9GfwXY	Text	26,846,000	1,750,000	43,000	18,000	0	0	0	0	1	0
https://weibo.com/1749127163/PkAMDqZU2	Photos	26,846,000	32,000	1,882	567	0	0	0	0	0	1
https://weibo.com/1749127163/PiuvH3mpi	Photos	26,846,000	6,327	553	119	1	0	0	0	0	0
https://weibo.com/1749127163/PnF0CtUkA	Video	26,846,000	24,000	2,794	496	0	0	1	0	1	0
https://weibo.com/1749127163/PkckLxHnU	Photos	26,846,000	13,000	1,049	228	1	0	1	0	0	0
https://weibo.com/1749127163/Pve3CzsnD	Photos	26,846,000	5,560	754	131	0	1	0	0	0	0
https://weibo.com/1749127163/PyUsW6xIt	Photos	26,846,000	11,000	1,602	312	0	1	0	0	0	1
https://weibo.com/1749127163/Pyvq4CIaH	Photos	26,846,000	86,000	3,078	1,311	1	0	1	0	0	1
https://weibo.com/1749127163/Pyr5FEEa6	Photos	26,846,000	27,000	2,391	453	1	0	0	0	0	0
https://weibo.com/1749127163/PxPoD0Mko	Photos	26,846,000	16,000	2,270	1,090	0	1	0	0	0	0
https://weibo.com/1749127163/PwB7Maq1J	Photos	26,846,000	26,000	1,137	516	0	0	0	1	1	0
https://weibo.com/1749127163/Pgk5cic80	Photos	26,846,000	34,000	1,809	505	1	0	0	1	0	0
https://weibo.com/1749127163/PtMfsn7gZ	Photos	26,846,000	14,000	1,359	894	1	0	0	0	0	0
https://weibo.com/1749127163/Pj8iNCC9b	Photos	26,846,000	24,000	2,072	1,212	1	0	0	0	0	0
https://weibo.com/1749127163/Pzbokth51	Photos	26,846,000	5,955	751	113	0	1	0	1	0	0
https://weibo.com/1749127163/PyncrvcmU	Photos	26,846,000	45,000	5,110	2,403	1	0	0	0	0	1
https://weibo.com/1749127163/PvyLC3lec	Photos	26,846,000	11,000	1,051	510	1	0	1	0	0	0

When Spending Stops Feeling Like Spending: Digital Payments, Buy Now Pay Later, and Generation Z's Financial Behavior By Gadzhi Bilalov

Abstract

Generation Z has entered adulthood in a financial environment where purchases can be completed almost instantly through mobile wallets, contactless payments, and Buy Now Pay Later (BNPL) services. Although these technologies provide convenience and greater payment flexibility, research suggests that they may also change how consumers experience the cost of spending. This narrative review examines the extent to which digital payment and BNPL use is associated with a reduced perception of spending cost among Gen Z and considers the cognitive mechanisms that may explain the financial behaviors that follow. Evidence from surveys, qualitative research, experiments, meta-analyses, and financial data suggests that cashless payments can reduce the psychological discomfort and visibility associated with paying and may contribute to greater or more impulsive spending. Three mechanisms are considered in particular: mental accounting, psychological ownership, and financial overconfidence. BNPL may intensify these effects by separating consumption from the full financial cost of a purchase. However, financial literacy, self-control, socioeconomic circumstances, and platform design may influence these relationships. Overall, the evidence suggests that understanding Gen Z's financial behavior requires examining not only how much young consumers spend, but also how digital payment technologies change the psychological experience of spending.

1. Introduction

The way consumers pay for goods and services has changed significantly as cash has increasingly been supplemented by cards, contactless payments, mobile wallets, and other digital financial technologies. BNPL services extend this development by allowing consumers to receive a product immediately while dividing its price into smaller future payments. These technologies provide real economic benefits, including convenience, faster transactions, easier online shopping, and greater flexibility in managing short-term expenses. At the same time, they may change the psychological experience of purchasing. Cash creates a visible exchange in which consumers physically give up money. Digital payments can reduce this physical connection, while BNPL can separate the purchase from much of its financial cost in time. This raises a particularly important question for Gen Z, a generation that has developed many of its financial habits in an environment where digital transactions are normal rather than new.

A growing body of research suggests that the method of payment can influence how costly spending feels. Broekhoff and van der Cruijsen (2024) found that electronic payments generally produced less "pain of paying" than cash, meaning that consumers experienced less psychological discomfort when paying electronically. Schomburgk et al. (2024) reached a related conclusion in a meta-analysis of 71 studies, identifying a small but statistically significant cashless effect in which consumers generally spent more when using cashless methods. Importantly, the size of this effect was not identical across contexts and appeared to weaken over

time. Experimental evidence further suggests that digital payment may do more than simply reduce an unpleasant feeling. Ma et al. (2024) found that reduced pain of paying helped explain mobile payment effects under some purchasing conditions, while under others a "pleasure of payment" emerged from the smooth digital transaction experience. Together, these findings suggest that payment technology can alter both how consumers perceive financial sacrifice and how they experience the transaction itself.

The behavioral relationship may become more complicated when BNPL is introduced. Unlike an ordinary mobile payment, BNPL does not only make payment less physically visible. It also separates consumption from the full payment in time. A consumer may receive a product immediately while initially seeing only one installment rather than experiencing the full price as an immediate financial loss. Several cognitive mechanisms may help explain why this matters. Mental accounting describes how consumers psychologically organize and evaluate money, while psychological ownership concerns the degree to which people experience money or possessions as belonging to them. Khan and Belk (2024) argue that the transition from tangible toward digital money has important implications for psychological ownership because physical cash can be touched, counted, and visibly surrendered in ways that digital money cannot. Financial overconfidence introduces a different potential vulnerability. Consumers who believe they are better at managing money than their objective financial knowledge suggests may be less likely to question whether numerous small digital transactions or future BNPL installments are accumulating into a larger financial obligation. Financial literacy can provide some protection, but evidence indicates that knowledge does not necessarily remove the behavioral relationship between mobile payments and overspending (Ahn & Nam, 2022).

This paper therefore asks: To what extent is Gen Z's use of digital payment systems and Buy Now Pay Later services associated with a reduced perception of spending cost, and what cognitive mechanisms might explain the financial habits that result? Rather than assuming that digital payments directly cause irresponsible behavior, this review examines evidence across different research methods and financial environments. It focuses primarily on mental accounting, psychological ownership, and financial overconfidence while considering financial literacy, self-control, financial circumstances, and platform design as factors that may strengthen or weaken these relationships. Existing studies have documented cashless spending effects and separately developed psychological explanations for how consumers experience digital money. This review brings those bodies of evidence together and considers psychological ownership in particular as a possible integrative explanation for spending patterns already observed across the digital payment literature.

2. Literature Review

Mental accounting and the pain of paying provide the most established starting point for explaining why payment methods may influence spending. Mental accounting refers to the ways consumers mentally categorize, track, and evaluate financial resources rather than treating every dollar as completely interchangeable. The broader relationship between consumption and

payment was formalized by Prelec and Loewenstein (1998), who described how the timing and psychological connection between paying and consuming can influence financial decisions. Digital payments can weaken this connection because money is no longer physically exchanged and transactions can be completed with minimal effort. Recent evidence supports this general mechanism while adding important qualifications. Broekhoff and van der Cruijsen (2024) found lower pain of paying for electronic payments compared with cash, while Ma et al. (2024) showed experimentally that reduced pain does not explain every mobile payment effect. Their findings suggest that mobile payment can also create a positive transaction experience, which they describe as the "pleasure of payment." Faraz and Anjum (2025) approach the problem through their concept of "Spendception," arguing that the reduced visibility of money in digital transactions can contribute to impulse purchasing. At a broader level, Schomburgk et al. (2024) provide meta-analytic evidence that cashless payment is associated with somewhat greater spending across a large body of research. Taken together, these studies suggest that digital payment can weaken some of the psychological friction traditionally associated with spending, although the magnitude and precise mechanism vary across situations.

Psychological ownership provides a second explanation for why the disappearance of physical money may matter. Psychological ownership refers to the feeling that something is "mine," even when the strength of that feeling is not determined entirely by legal ownership. Physical cash provides unusually strong ownership cues because consumers can hold it, count it, organize it, and watch it leave their possession. Khan and Belk (2024) use this perspective to examine how digital currencies change the relationship consumers have with money. Their qualitative evidence suggests that tangible cash can create stronger feelings of possession, while digital payment shifts the consumer's interaction toward an application or payment interface. This mechanism may complement rather than replace the pain-of-paying explanation. If consumers experience weaker ownership over digital money, giving it up may create a smaller perceived loss. At the same time, digital environments can generate psychological ownership over the things being purchased. Research on digital goods suggests that consumers can develop meaningful feelings of ownership even toward intangible products, and stronger ownership can increase willingness to purchase and reduce price sensitivity (Kučinskas, 2024). The combination creates an interesting possibility: digital consumers may experience weaker psychological attachment to the money being surrendered while simultaneously developing strong attachment to the product or digital experience being acquired.

Financial overconfidence provides a third mechanism, although it should be distinguished carefully from financial literacy. Financial literacy concerns actual knowledge and skills, while overconfidence concerns a mismatch between what a consumer knows and what the consumer believes they know. This distinction matters because digital financial tools often appear easy to control. A consumer may believe that several small mobile transactions or BNPL installments are manageable even when their combined financial impact is larger than expected. Research using large U.S. financial datasets suggests that mobile payment behavior can interact with individual financial characteristics. Ahn and Nam (2022), using data from 21,457 participants in

the National Financial Capability Study, found that mobile payment users were more likely to demonstrate overspending behaviors. Greater financial knowledge moderated this association but did not eliminate it. The student's source collection also identifies recent research connecting financial overconfidence and mobile payment use with high-cost borrowing, suggesting that subjective confidence may become particularly important when convenient payment technologies are combined with risky financial choices. This evidence does not establish that digital payments create overconfidence. Instead, it suggests that preexisting overconfidence may make some consumers more vulnerable to the reduced friction of digital finance.

BNPL brings these mechanisms together because it changes both how a payment is presented and when the financial sacrifice occurs. A traditional cash purchase connects consumption and payment almost simultaneously. A credit card already weakens this connection, but BNPL can make the separation particularly visible by dividing a total price into smaller installments displayed at checkout. Research reviewed in the student's materials links BNPL with impulsive purchasing, normalization of debt, financial strain, and reduced attention to total obligations among younger consumers. Qualitative work by Threadgold et al. (2025) is especially relevant because it examines how BNPL technologies become incorporated into the financial lives of young people and how platform features can make debt appear more routine and less threatening. Other recent research has linked BNPL use with financial fragility and adverse financial outcomes, although such findings should be interpreted as associations unless the underlying research design supports causal conclusions. This distinction is important because people who adopt BNPL may already face greater financial constraints than non-users. BNPL can provide useful flexibility for such consumers while simultaneously making it easier to accumulate several payment obligations whose combined cost is less obvious at the moment of purchase.

Across these areas, the literature points toward a common pattern without establishing one universal causal pathway. Digital payment research documents lower pain of paying and a modest cashless spending effect; psychological ownership research provides a potential explanation for why losing digital money may feel different from surrendering physical cash; financial research shows that knowledge and confidence can influence how consumers respond to payment technologies; and BNPL research demonstrates how delayed payment can further separate consumption from financial sacrifice. These mechanisms also operate within broader circumstances such as income, financial socialization, self-control, debt, and economic insecurity. The main gap is therefore not a complete absence of evidence about digital spending. Rather, the literature remains fragmented across different explanations and research methods. This review proposes that reduced spending-cost perception is best understood as the product of interacting mechanisms rather than one isolated psychological effect. Psychological ownership may be particularly useful for connecting the physical disappearance of money with the behavioral spending patterns documented elsewhere, but the available evidence remains insufficient to claim that ownership has been empirically established as the direct cause of Gen Z's digital spending behavior.

3. Methods

This study uses a narrative review approach to examine the relationship between digital payment technologies, BNPL services, perceived spending cost, and financial behavior among Gen Z consumers. A narrative review was selected because the purpose of the paper is not to estimate a single numerical effect across studies, but to compare findings from different research traditions and determine whether they support a common behavioral explanation. The literature considered in this review includes research on mobile payments, cashless spending, BNPL, mental accounting, pain of paying, psychological ownership, financial literacy, financial overconfidence, impulsive buying, and financial fragility. Particular attention is given to studies involving younger consumers or university students, although broader consumer research is included when it provides evidence about mechanisms relevant to the research question.

The evidence reviewed comes from several different methodological approaches. These include large consumer surveys, qualitative interviews and focus groups, experimental studies, economic models, meta-analyses, and studies using behavioral or financial data. For example, Ahn and Nam (2022) analyzed more than 21,000 participants from the U.S. National Financial Capability Study, while Cobla and Osei-Assibey (2018) examined mobile money use among 506 university students in Ghana. Schomburgk et al. (2024) provide broader evidence through a meta-analysis of 71 previous studies, while Ma et al. (2024) used experimental and EEG evidence to investigate the psychological processes involved in mobile payment. Qualitative research is also important because concepts such as psychological ownership and perceived spending cost are partly subjective experiences that cannot always be captured through transaction data alone. Using different types of evidence makes it possible to compare whether similar patterns appear when the same general problem is studied through different methods.

Rather than organizing the evidence study by study, the review groups findings around three proposed cognitive mechanisms: mental accounting, psychological ownership, and financial overconfidence. Evidence concerning BNPL is then considered as a specific payment environment in which these mechanisms may become especially relevant because payment is divided and delayed. Financial literacy, self-control, socioeconomic conditions, and financial socialization are treated primarily as contextual or moderating factors rather than as consequences of digital payment adoption. This distinction is necessary because the available evidence does not demonstrate that using mobile payments or BNPL causes consumers to become less financially literate or more overconfident. Instead, these individual characteristics may help explain why similar payment technologies produce different outcomes for different consumers.

There are important limitations to this approach. This paper is a narrative review rather than a systematic review or meta-analysis, so it cannot claim to identify every relevant study or calculate a pooled causal effect. The studies also differ considerably in their populations, countries, definitions of digital payment, and measures of financial behavior. Evidence from university students in Ghana, young BNPL users in Australia, or general consumers in the United States cannot automatically be generalized to all members of Gen Z. Furthermore, much of the

available evidence is observational, meaning that an association between digital payment use and overspending does not necessarily demonstrate that the payment method caused the behavior. These limitations are considered throughout the analysis, particularly when connecting psychological theories with observed financial outcomes.

4. Results and Synthesis

Across the literature reviewed, the most consistent finding is that reducing the physical and psychological friction of payment is associated with changes in spending behavior. Schomburgk et al. (2024) found a small but statistically significant overall cashless effect across a large body of prior research, while Broekhoff and van der Cruijsen (2024) found that electronic payment methods were generally associated with lower pain of paying than cash. Evidence involving younger consumers points in a similar direction. Cobla and Osei-Assibey (2018) found higher expenditure among university students who actively used mobile money, even after accounting for several socioeconomic factors. Ahn and Nam (2022) also found that U.S. mobile payment users were more likely to demonstrate overspending behaviors. These studies differ substantially in population and methodology, so their results should not be treated as measurements of one identical effect. However, their general direction is consistent with the argument that payment methods can influence spending behavior rather than functioning as completely neutral channels through which predetermined purchases are completed.

The evidence also suggests that reduced spending-cost perception cannot be explained solely by one mechanism. Mental accounting and pain of paying provide the clearest initial explanation. A digital transaction reduces many of the physical cues that make the loss of money visible, while a frictionless interface may further reduce the attention given to the payment itself. Ma et al. (2024) add an important qualification by showing that the mobile payment effect is not always explained simply by reduced pain. Their experiments indicate that under some conditions a positive "pleasure of payment" can also contribute to purchasing. Psychological ownership provides a complementary explanation. Khan and Belk (2024) suggest that physical cash creates forms of psychological ownership that become altered as money moves into digital environments. Considered together, these findings suggest a possible process in which consumers experience less psychological attachment to the money being surrendered, less discomfort during the transaction, and a smoother or even more rewarding purchasing experience. The literature does not yet establish this sequence as a tested causal pathway, but findings across different methodologies are consistent with parts of it.

BNPL appears to add another layer by changing the temporal structure of payment. Digital wallets can make a payment less tangible, but BNPL additionally divides the financial obligation across future periods. This can direct attention toward a smaller installment rather than the full purchase price and create greater separation between receiving a product and experiencing its complete financial cost. The young-consumer literature reviewed in the student's research materials repeatedly connects BNPL with impulsive purchasing, normalization of debt, and difficulty keeping track of multiple repayment obligations. Threadgold et al. (2025), for

example, examine how BNPL technologies are incorporated into young people's financial lives and emphasize the role of platform design and the normalization of digital debt. Other recent work has associated BNPL use with financial fragility and adverse outcomes. These findings do not mean that BNPL is harmful for every consumer. Installment payments can provide useful short-term flexibility, particularly when they substitute for more expensive credit. The risk appears more significant when payment fragmentation combines with limited resources, impulsive purchasing, or difficulty monitoring several obligations simultaneously.

Finally, financial literacy and overconfidence appear to influence these relationships without fully explaining them. Ahn and Nam (2022) found that greater financial knowledge moderated the association between mobile payment use and overspending but did not eliminate it. This finding supports an important distinction for the present review. Consumers can understand financial concepts while still responding behaviorally to the design of a payment system. Overconfidence may create an additional problem when consumers believe their ability to manage spending is stronger than their actual financial knowledge or behavior suggests. At the same time, financial outcomes cannot be understood only through individual cognition. The student's source materials also identify financial socialization, mental health, income constraints, and broader financial fragility as relevant factors affecting younger consumers. The overall synthesis therefore suggests an interaction rather than a single cause: digital payment design can reduce spending friction, cognitive mechanisms can affect how that reduced friction is perceived, and personal or economic circumstances can determine whether the result is harmless convenience or problematic financial behavior.

5. Discussion and Analysis

The evidence reviewed provides qualified support for the central research question. Gen Z's use of digital payment systems and BNPL is associated in the literature with conditions that can reduce the perceived cost of spending, but the evidence is stronger for some parts of this argument than others. The cashless effect and reduced pain of paying are supported by research using multiple methodologies, including meta-analysis, consumer surveys, and experiments (Broekhoff & van der Cruijsen, 2024; Ma et al., 2024; Schomburgk et al., 2024). There is also evidence connecting mobile payment use with overspending among both student and broader consumer populations (Ahn & Nam, 2022; Cobla & Osei-Assibey, 2018). What is less firmly established is whether these relationships operate differently for Gen Z specifically and whether psychological ownership is the mechanism directly responsible for observed spending increases. The evidence therefore supports an association between increasingly frictionless payment and reduced spending salience more strongly than it supports a single Gen Z-specific causal explanation.

The most useful contribution of psychological ownership is that it may explain something that pain of paying alone does not fully capture. Pain of paying describes the discomfort consumers experience when surrendering money, while psychological ownership addresses why surrendering one form of money may feel more significant than surrendering another. Physical

cash can be held, counted, divided, and visibly depleted. Digital balances remain numbers on a screen, and many digital transactions require only a tap or confirmation. Khan and Belk's (2024) framework therefore provides a useful lens for interpreting behavioral findings from other studies. If physical interaction strengthens a sense that money is personally possessed, removing that interaction may reduce the subjective sense of loss when money is spent. This interpretation remains theoretical because the studies documenting cashless spending increases generally do not directly test psychological ownership as the mediating variable. Still, it offers a plausible bridge between research on the subjective experience of money and research measuring actual spending behavior.

Mental accounting and overconfidence may strengthen this process in different ways. When payments are divided across apps, cards, digital wallets, and future installments, tracking the total cost of consumption may become more cognitively demanding. BNPL makes this especially clear because one purchase creates several future obligations rather than one immediate financial sacrifice. Overconfidence can then become relevant if consumers believe they are successfully managing these obligations despite incomplete tracking. Financial literacy helps complicate the picture. If knowledge completely eliminated the cashless spending effect, the problem could largely be treated as one of financial education. Instead, Ahn and Nam (2022) found that financial knowledge reduced but did not remove the association between mobile payments and overspending. This suggests that educational interventions may be useful but insufficient when the design of the transaction itself changes consumer behavior. Better knowledge may need to be combined with tools that make total expenditure and future obligations more visible at the moment consumers make decisions.

This interpretation has implications for BNPL platform design and consumer protection. Many current digital interfaces emphasize convenience, speed, and the affordability of individual installments. From a behavioral perspective, however, showing a small installment more prominently than the total obligation could increase the separation between the perceived and actual cost of a purchase. A possible policy response would therefore focus not simply on providing more information, but on how that information is presented. Displaying the total outstanding BNPL balance across purchases, clearly presenting the full purchase price alongside each installment, or providing reminders of simultaneous repayment obligations could increase payment salience without removing access to installment credit. Such interventions would need empirical testing before their effectiveness could be assumed. Nevertheless, they follow logically from evidence suggesting that visibility, timing, and transaction experience influence spending decisions.

Several limitations prevent stronger conclusions. First, the evidence comes from different countries and populations, and Gen Z should not be treated as a financially uniform global group. A university student in Ghana, a BNPL user in Australia, and a young American mobile-wallet users face different credit systems, income conditions, cultural expectations, and regulatory environments. Second, many studies identify associations rather than causation, and consumers who choose BNPL or mobile payments may already differ from non-users. Third, the cashless

effect itself appears to vary across contexts and has weakened over time (Schomburgk et al., 2024). Finally, this narrative review cannot empirically test whether psychological ownership mediates the relationship between payment method and Gen Z spending. Future research could address this directly by measuring psychological ownership, perceived payment cost, and actual spending behavior within the same study and comparing cash, immediate digital payment, and BNPL conditions. Such research would provide a stronger test of whether the theoretical connection proposed in this review explains the behavioral patterns already documented across the literature.

6. Conclusion

This review examined the extent to which Gen Z's use of digital payment systems and Buy Now Pay Later services is associated with a reduced perception of spending cost and explored the cognitive mechanisms that may help explain the financial behaviors that result. Overall, the evidence suggests that digital payments can make spending feel less immediate and less psychologically costly than physical cash. BNPL may strengthen this effect because it not only removes many of the physical cues associated with payment but also separates the moment of purchase from the full financial cost. However, these relationships are not universal. Financial knowledge, self-control, economic circumstances, previous financial habits, and the design of payment platforms can all influence how consumers respond to digital payment technologies. The literature also suggests that mental accounting, psychological ownership, and financial overconfidence can work together rather than functioning as completely separate explanations. Digital payments may weaken the visibility of individual expenses, while the loss of physical interaction with money can reduce some of the sense of ownership associated with cash. At the same time, consumers who are confident in their ability to manage money may underestimate how quickly small transactions and BNPL obligations accumulate. The main contribution of this review is therefore to bring these mechanisms together and use psychological ownership as a possible bridge between the changing experience of money and the spending patterns documented in previous research. Future studies should test this relationship directly by examining psychological ownership, perceived spending cost, financial confidence, and actual spending behavior within the same group of Gen Z consumers. This would help determine whether increasingly frictionless payment systems are simply changing how young people pay or are also changing how they perceive the cost of spending.

Works Cited

Ahn, S. Y., & Nam, Y. (2022). Does mobile payment use lead to overspending? The moderating role of financial knowledge. Computers in Human Behavior, 134, 107319. https://doi.org/10.1016/j.chb.2022.107319

Broekhoff, M. C., & van der Cruijsen, C. (2024). Paying in a blink of an eye: It hurts less, but you spend more. Journal of Economic Behavior & Organization, 221, 110–133. https://doi.org/10.1016/j.jebo.2024.03.017

Cobla, G. M., & Osei-Assibey, E. (2018). Mobile money adoption and spending behaviour: The case of students in Ghana. International Journal of Social Economics, 45(1), 29–42. https://doi.org/10.1108/IJSE-11-2016-0302

Faraz, N., & Anjum, N. (2025). Spendception: The psychological impact of digital payments on consumer purchase behavior and impulse buying. Behavioral Sciences, 15(3), 387. https://doi.org/10.3390/bs15030387

Khan, J., & Belk, R. (2024). Money you could touch: Cash and psychological ownership. Qualitative Market Research: An International Journal, 27(5), 820–841. https://www.emerald.com/qmr/article-abstract/27/5/820/1220083

Kučinskas, G. (2024). Consumer responses to diverse digital goods: The role of psychological ownership in life planning apps, music streaming services, and game skins. Journal of Industrial Engineering & Management Research. https://jiemar.org/index.php/jiemar/article/download/529/408

Ma, Q., Tan, Y., He, Y., Cheng, L., & Wang, M. (2024). Why does mobile payment promote purchases? Revisiting the pain of paying, and understanding the implicit pleasure via selective attention. PsyCh Journal. https://doi.org/10.1002/pchj.765

Prelec, D., & Loewenstein, G. (1998). The red and the black: Mental accounting of savings and debt. Marketing Science, 17(1), 4–28. https://doi.org/10.1287/mksc.17.1.4

Schomburgk, L., Belli, A., & Hoffmann, A. O. I. (2024). Less cash, more splash? A meta-analysis on the cashless effect. Journal of Retailing, 100(3), 382–400. https://www.sciencedirect.com/science/article/pii/S0022435924000216

Threadgold, S., Shannon, B., Haro, A., Cook, J., & Coffey, J. (2025). Buy Now, Pay Later technologies and the gamification of debt in the financial lives of young people. Journal of Cultural Economy. https://doi.org/10.1080/17530350.2024.2346210

The Economics of Semiconductor Subsidies: A Comparative Review of the United States, European Union, and China By Matvei Ivashchenko

Abstract

Semiconductors have become a central target of industrial policy as governments seek to strengthen economic resilience, technological competitiveness, and access to strategically important technologies. The United States, European Union, and China have each committed substantial public resources to developing domestic semiconductor capabilities, but their policies reflect different economic priorities and positions within the global semiconductor value chain. This comparative review examines the economic rationale, policy design, potential benefits, and risks of semiconductor subsidies across the three regions. It argues that semiconductor subsidies can be understood partly as a response to market failures associated with high capital requirements, research spillovers, supply-chain concentration, and the economic costs of major disruptions. At the same time, subsidies create risks of inefficient investment, international subsidy competition, excess capacity, and government support for firms or projects that may not remain competitive without continued assistance. The comparison demonstrates that the United States emphasizes domestic manufacturing and technological leadership, the European Union combines capacity expansion with strategic autonomy, and China employs a broader state-directed strategy aimed at technological self-sufficiency. The paper concludes that the economic effectiveness of semiconductor subsidies depends less on the scale of government spending alone than on targeting, coordination, and whether public investment creates durable productive and technological capabilities.

Keywords: semiconductor subsidies, industrial policy, United States, European Union, China, supply chain resilience, strategic autonomy, technological competition.

1. Introduction

Semiconductors occupy an unusual position in the modern economy because they are both intermediate goods and foundational technologies. They are not generally purchased by consumers for their own sake, yet they are embedded in products ranging from smartphones and automobiles to industrial equipment, telecommunications infrastructure, medical devices, and advanced computing systems. Their economic importance has increased further with the expansion of artificial intelligence, cloud computing, autonomous technologies, and other data-intensive industries. At the same time, semiconductor production is organized through a highly specialized global value chain in which design, fabrication, equipment production, materials, assembly, and testing are distributed across different firms and countries. This specialization has generated substantial efficiency and innovation, but it has also produced geographic concentration at several critical stages of production. As the student's preliminary research correctly identifies, this creates exposure to disruptions arising from geopolitical

conflict, natural disasters, trade restrictions, and other shocks (García-Herrero & Poitiers, 2022; Sutter et al., 2023).

The economic consequences of this concentration became especially visible during the semiconductor shortages associated with the COVID-19 pandemic. Disruptions in production occurred at the same time that demand for electronics and digital technologies changed rapidly, creating shortages that spread into downstream industries. Automobile manufacturers were particularly exposed because modern vehicles depend on numerous semiconductor components, but the effects extended across consumer electronics and industrial production. The experience demonstrated an important characteristic of semiconductor supply chains: disruption in a relatively small number of upstream facilities can create economic consequences across industries representing a much larger share of output. Governments consequently began to view semiconductor capacity not simply as a private investment decision made by individual firms, but as an issue involving national economic resilience and strategic capacity. This shift contributed to a broader return of industrial policy in advanced economies, where governments increasingly use subsidies, tax incentives, public research funding, and other interventions to influence the location and composition of private investment (Millot & Rawdanowicz, 2024).
The United States, European Union, and China provide particularly useful cases for studying this development because all three have adopted major semiconductor policies while approaching the industry from different economic positions. The United States remains exceptionally strong in semiconductor design, intellectual property, and parts of the semiconductor equipment industry, but its share of global manufacturing has declined substantially over several decades. The CHIPS and Science Act represented an effort to attract fabrication capacity, expand research and development, and reduce dependence on geographically concentrated foreign production. The European Union faces a different problem. Europe possesses important capabilities in semiconductor equipment, research, automotive chips, and specialized manufacturing, yet remains dependent on non-European suppliers for many advanced semiconductors. The European Chips Act therefore seeks to expand European capacity while strengthening what policymakers describe as technological sovereignty or strategic autonomy. García-Herrero and Poitiers (2022), however, caution that European subsidies must be targeted carefully rather than simply attempting to reproduce every stage of semiconductor production domestically.
China represents a third model. Its semiconductor policy forms part of a longer state-led effort to develop technologically strategic industries and reduce dependence on foreign suppliers. Chinese policymakers have supported semiconductor development through investment funds, tax incentives, government procurement, financing, research support, and other forms of industrial policy. The objective is broader than attracting individual fabrication plants. It involves developing domestic capabilities across multiple stages of the semiconductor ecosystem. This strategy has become increasingly important as the United States and some of its partners have imposed restrictions affecting China's access to advanced chips and semiconductor manufacturing technologies. Consequently, semiconductor policy now operates at the intersection of conventional industrial economics and geoeconomic competition. Decisions about

subsidies influence investment and production, while export controls and technology restrictions influence which firms and countries can participate in particular parts of the value chain (Sutter et al., 2023).

These developments raise a broader economic question about the appropriate role of government in strategically important industries. Semiconductor manufacturing involves enormous fixed costs, uncertain technological returns, substantial research and development expenditure, knowledge spillovers, and potential social benefits from supply-chain resilience that an individual firm may not fully capture when making an investment decision. These characteristics provide economic arguments for government intervention. Yet the existence of a rationale for intervention does not establish that every subsidy produces a net economic benefit. Subsidies can shift costs from firms to taxpayers, distort investment decisions, encourage governments to compete against one another for the same multinational firms, and support production that would otherwise be commercially uncompetitive. This paper therefore compares semiconductor subsidy policies in the United States, European Union, and China to evaluate their economic rationales, differences in policy design, and principal benefits and risks. Rather than asking which region spends the most, the analysis focuses on a more important economic issue: under what conditions can semiconductor subsidies strengthen long-term productive capacity and resilience, and under what conditions might they instead produce inefficient allocation and costly subsidy competition?

2. Economic Rationale for Semiconductor Subsidies

The most straightforward economic argument for semiconductor subsidies begins with the unusually high cost of entering advanced semiconductor manufacturing. Modern fabrication facilities require investments measured in billions of dollars, while production equipment becomes technologically obsolete relatively quickly as manufacturing processes advance. Firms must therefore commit enormous amounts of capital before knowing with certainty what future demand, prices, or technological standards will look like. These fixed costs create significant barriers to entry and favor firms that already possess scale, technical expertise, and established relationships with suppliers and customers. From a conventional economic perspective, high fixed costs alone do not necessarily justify subsidies. However, the semiconductor industry also generates knowledge spillovers through research, engineering expertise, workforce development, and supplier networks. Because individual companies may be unable to capture the entire social return generated by these investments, private investment can theoretically remain below the level that would maximize broader economic benefits. This is one of the traditional market-failure arguments for industrial policy (Millot & Rawdanowicz, 2024).

A second rationale concerns supply-chain resilience. Firms generally optimize supply chains according to private considerations such as cost, reliability, inventory requirements, and expected profitability. Governments, however, may assign additional economic value to maintaining access to critical inputs during severe disruptions. Semiconductor shortages can generate costs far beyond the semiconductor industry's own lost output because chips are

required by numerous downstream sectors. This creates a potential divergence between private and social incentives. A semiconductor producer deciding where to locate a fabrication facility does not necessarily receive compensation for the resilience that additional geographic diversification provides to automobile manufacturers, defense suppliers, or communications companies. Government support can therefore be interpreted as an attempt to subsidize a positive externality, namely the broader economic security created by diversified productive capacity. Recent OECD work similarly emphasizes that resilience policies must account for the costs of disruptions while avoiding the assumption that complete domestic self-sufficiency is necessarily the most efficient solution (Thakur-Weigold & Miroudot, 2024).

The third rationale is technological competition. Semiconductor manufacturing is connected to innovation in artificial intelligence, telecommunications, advanced computing, defense systems, and other technologies with potentially large productivity effects. Semiconductor research can therefore generate dynamic benefits that extend beyond the immediate value of chip production. Governments may use subsidies to create industrial clusters in which fabrication facilities attract suppliers, engineers, research institutions, and complementary investment. Once established, such clusters can potentially generate agglomeration effects that make future investment more productive. This logic helps explain why semiconductor policy in all three regions extends beyond direct payments for factories. Research funding, workforce programs, tax incentives, infrastructure investment, and support for related technologies are frequently combined with manufacturing subsidies. The policy objective is not simply to produce a larger number of chips in the short run, but to influence the location of technological capabilities over a much longer period.

National security adds another dimension, although it should be distinguished from purely economic arguments. Advanced semiconductors are important inputs into military systems and critical infrastructure, while leadership in chip design and manufacturing can influence capabilities in artificial intelligence and high-performance computing. Sutter et al. (2023) show that U.S. semiconductor policy increasingly reflects the connection between commercial supply chains, technological leadership, and national-security concerns. Similar considerations appear in European discussions of strategic autonomy and China's efforts to reduce technological dependence. From an economic perspective, national-security externalities make conventional cost comparisons incomplete. A domestically produced semiconductor may cost more than an imported alternative but still provide value if policymakers believe domestic capacity reduces the expected cost of a severe geopolitical disruption. The difficulty is that this insurance value is hard to measure. Governments can consequently justify substantial spending using benefits that are real in principle but uncertain in magnitude.

These arguments explain why semiconductor subsidies can have an economic rationale, but they also establish the standard against which such policies should be evaluated. The relevant question is not whether semiconductors are important. Their importance is well established. The question is whether a particular intervention addresses a genuine market failure or resilience problem more effectively than alternative policies. Poorly targeted subsidies may simply transfer

public resources to profitable corporations or move investment from one subsidizing country to another without meaningfully increasing global resilience. Subsidies may also produce excess capacity if governments simultaneously encourage similar investments based on optimistic projections of future demand. The economic case for semiconductor industrial policy is therefore strongest when public support creates capabilities with broader social benefits that private markets would otherwise underprovide. It becomes weaker when subsidies primarily protect domestic producers from competition or encourage investments whose costs persist after their strategic value has declined.

3. The United States: Reshoring, Innovation, and the CHIPS Act

The contemporary U.S. semiconductor strategy reflects a significant change in the federal government's approach to industrial policy. Although American firms remain highly competitive in semiconductor design, intellectual property, electronic design automation, and manufacturing equipment, semiconductor fabrication has become increasingly concentrated outside the United States. The Congressional Research Service notes that the United States is particularly dependent on Taiwan for advanced logic chips and South Korea for advanced memory production, while Taiwan accounts for a very large share of global leading-edge fabrication capacity (Sutter et al., 2023). The student's preliminary research identifies this dependence as one of the main motivations behind current U.S. policy. The economic concern is not simply that semiconductor manufacturing takes place abroad. International specialization can lower production costs and improve efficiency. Rather, policymakers increasingly view extreme geographic concentration as a source of systemic risk because disruption in a small number of foreign production centers could affect a much wider range of American industries.

The CHIPS and Science Act of 2022 represents the principal U.S. policy response to this concern. The legislation authorized approximately $52.7 billion for semiconductor manufacturing incentives, research, workforce development, and related programs, including $39 billion in financial incentives for manufacturing. It also established a 25 percent advanced manufacturing investment tax credit for qualifying semiconductor investments (Sutter et al., 2023; Taylor, 2023). Importantly, the policy does not seek to replace private semiconductor investment with government production. Instead, public funding is intended to alter the economics of locating capital-intensive semiconductor projects in the United States. Building and operating fabrication facilities in the United States can be more expensive than doing so in established Asian manufacturing clusters, meaning that even firms headquartered in the United States may have financial incentives to locate production elsewhere. Grants and tax credits reduce part of this cost difference and attempt to attract private investment that might otherwise occur abroad. In this sense, the CHIPS Act uses public resources to influence the geographic distribution of private capital rather than creating a state-owned semiconductor industry.

The U.S. strategy also extends beyond fabrication. Semiconductor competitiveness depends on continuous research, specialized workers, manufacturing equipment, advanced packaging, materials, and the ability to move new technologies from laboratories into commercial

production. For this reason, CHIPS Act programs combine manufacturing incentives with substantial support for research and development. This broader structure matters economically because a fabrication plant by itself does not necessarily create a sustainable semiconductor ecosystem. If equipment, engineering talent, suppliers, and future generations of technology remain concentrated elsewhere, subsidizing individual factories may produce only temporary gains in manufacturing capacity. Zhou et al. (2022) therefore argue that the success of the CHIPS and Science Act depends partly on whether public investment strengthens innovation capabilities and coordination between government, industry, and research institutions. More recent analysis similarly treats the Act as a combination of manufacturing policy and technological strategy rather than a conventional short-term subsidy program (Hyatt et al., 2025).

Alongside this, the U.S. strategy contains a second policy instrument that distinguishes it from a simple reshoring program: restrictions on the transfer of certain advanced semiconductor technologies to China. Manufacturing incentives encourage investment within the United States, while export controls attempt to preserve technological advantages by limiting Chinese access to advanced computing chips and selected semiconductor manufacturing equipment. These policies are economically connected. If the United States subsidized domestic technological development while advanced capabilities could be transferred without restriction to strategic competitors, policymakers argue that some of the security benefits of the subsidies would be weakened. However, export restrictions also create costs. Semiconductor companies operate in global markets and can lose revenue when access to major customers is restricted. Reduced international sales can, in turn, affect the resources firms have available for research and development. U.S. policy therefore faces a tension between maintaining an open global market that supports scale and innovation and restricting technological diffusion for national-security purposes (Patrignani et al., 2025).

The economic evaluation of the CHIPS Act consequently depends on what standard is used to define success. If success means eliminating U.S. dependence on foreign semiconductor production, the objective would be extremely costly and potentially unrealistic because the semiconductor value chain is too internationally specialized for complete self-sufficiency to be efficient. If the objective is instead to reduce concentrated vulnerabilities, preserve technological capabilities, and establish additional domestic production at strategically important stages, the policy has a more defensible economic rationale. Early research has found evidence of increased semiconductor investment and employment following the shift in U.S. industrial policy, although it remains too early to determine the full long-term return on the subsidies (Erten et al., 2025). The central challenge will be determining whether the resulting facilities remain internationally competitive after the initial incentives have been absorbed. A subsidy can attract a factory, but long-term industrial competitiveness requires productivity, innovation, skilled labor, supplier networks, and sufficient demand to sustain that factory after government support becomes less important.

4. The European Union: Strategic Autonomy and the European Chips Act

The European Union entered the current semiconductor subsidy competition from a different position. Europe is not absent from the global semiconductor industry. It possesses important capabilities in semiconductor research, manufacturing equipment, automotive and industrial chips, power semiconductors, and specialized production. ASML, headquartered in the Netherlands, occupies an especially important position in advanced lithography equipment, while European firms such as Infineon, STMicroelectronics, and NXP maintain significant positions in particular semiconductor markets. Nevertheless, the EU remains dependent on external suppliers for many advanced chips and has a relatively limited position in leading-edge logic fabrication. The pandemic-era shortages therefore raised concerns not only about the quantity of semiconductor production located in Europe but also about the types of semiconductor capabilities available within the region. Johnston and Huggins (2023) describe this challenge in terms of both industrial policy and the development of regional semiconductor clusters, emphasizing that European competitiveness depends on strengthening existing areas of specialization rather than treating semiconductor manufacturing as a single homogeneous industry.

The European Chips Act, which entered into force in 2023, represents the EU's main policy response. The initiative seeks to mobilize more than €43 billion in public and private investment and has been associated with the ambition of increasing the EU's share of global semiconductor production to 20 percent by 2030. The policy combines support for manufacturing with research, pilot lines, design capabilities, skills development, coordination mechanisms, and measures intended to monitor and respond to supply disruptions. This makes the European strategy similar to the U.S. approach in its recognition that semiconductor policy cannot consist solely of factory subsidies. At the same time, the institutional structure is more complicated because industrial policy in Europe operates through both EU-level institutions and individual member states. Large semiconductor projects may therefore involve European programs, national subsidies, state-aid rules, and private investment simultaneously. This multilevel structure can provide substantial resources but can also make policy coordination more difficult (Johnston & Huggins, 2023).

The economic concept underlying much of the European strategy is "open strategic autonomy." The term reflects an attempt to balance two objectives that can pull policy in different directions. Europe wants to reduce excessive dependence on external suppliers in strategically important technologies, but it does not seek complete separation from global markets. This distinction is especially important for semiconductors because no major economy currently possesses a fully autonomous supply chain. Semiconductor production depends on equipment, materials, intellectual property, design expertise, manufacturing facilities, and customers distributed across multiple countries. Monsees (2025) describes this as a fundamental paradox in European semiconductor governance: efforts to increase technological sovereignty take place within an industry whose economic structure is based on international interdependence. For the EU, the practical objective is therefore better understood as reducing

critical dependencies and strengthening bargaining power rather than achieving complete semiconductor self-sufficiency.

This is also where García-Herrero and Poitiers' analysis, which the student selected in the preliminary assignment, becomes particularly useful. They argue that European semiconductor subsidies need to be better targeted and caution against interpreting a larger share of global production as an economic objective in itself (García-Herrero & Poitiers, 2022). From an economic perspective, Europe does not necessarily benefit from subsidizing every stage of semiconductor production simply because other governments are doing so. Public support is more likely to generate lasting benefits when it reinforces areas in which Europe possesses technological capabilities, creates positive spillovers, or addresses clearly identified supply vulnerabilities. Attempting to reproduce the entire global semiconductor value chain domestically would require enormous resources and could divert public funding toward activities in which European production has no durable comparative advantage. The relevant question is therefore not how many chips Europe can manufacture, but which semiconductor capabilities generate the greatest economic and strategic value for the region.

The European approach consequently illustrates one of the central difficulties of modern industrial policy. Subsidies may be justified as a response to strategic dependence, yet international subsidy competition can itself weaken economic efficiency. If the United States, EU member states, China, Japan, South Korea, and other economies all offer increasingly generous incentives for semiconductor investment, multinational firms may be able to negotiate larger public support packages by comparing potential locations. Governments can then end up transferring a greater share of investment costs to taxpayers without producing a proportional increase in global semiconductor capacity or resilience. There is also an internal European distributional issue. Wealthier member states have greater fiscal capacity to subsidize large industrial projects, potentially concentrating new investment in countries that can offer the largest support packages. The long-term success of the European Chips Act should therefore be assessed not merely by whether new fabrication facilities are announced, but by whether the policy strengthens innovation, reduces economically significant vulnerabilities, complements existing European specialization, and produces capabilities that remain viable without permanent public support (Barbieri Góes & Viesti, 2025; Johnston & Huggins, 2023).

5. China: State-Led Industrial Policy and Semiconductor Self-Sufficiency

China's semiconductor strategy differs from those of the United States and European Union because large-scale government involvement in the industry predates the recent global expansion of semiconductor subsidies. While the United States and EU significantly expanded semiconductor industrial policy following the supply disruptions and geopolitical tensions of the early 2020s, China had already identified semiconductors as a strategic industry within a broader effort to move its economy toward higher-value technological production. Semiconductor development became especially prominent under policies associated with Made in China 2025 and other national technology strategies. The economic motivation is closely connected to the

structure of China's economy. China is one of the world's largest markets for semiconductors because of its enormous electronics manufacturing sector, yet for many years domestic production has been unable to satisfy demand for many advanced chips and semiconductor technologies. This imbalance created both an economic dependence on imported technology and a strategic vulnerability that Chinese industrial policy has attempted to reduce (Ding, 2025; Sutter et al., 2023).

A central feature of the Chinese model is the variety of instruments through which the state supports semiconductor development. Rather than relying on a single legislative package comparable to the U.S. CHIPS Act, support has been distributed through national and local investment funds, tax incentives, preferential financing, research support, government procurement, land and infrastructure policies, and other forms of assistance. One of the most visible mechanisms has been the China Integrated Circuit Industry Investment Fund, commonly known as the Big Fund. Its successive financing rounds have directed capital toward semiconductor manufacturing and other parts of the domestic supply chain. At the same time, provincial and municipal governments have developed their own semiconductor programs, creating a multilayered industrial policy system. This allows China to mobilize substantial financial resources and support multiple parts of the semiconductor ecosystem simultaneously. Recent research describes this approach as a form of state-directed technological upgrading in which government institutions attempt to coordinate capital, firms, research capacity, and industrial development around strategic technological objectives (To & Lee, 2026).

The economic record of this strategy, however, is mixed. Chinese support has contributed to an expansion of domestic semiconductor capacity and has helped firms develop capabilities in chip design, mature-node manufacturing, packaging, and parts of the semiconductor equipment and materials industries. Yet large public expenditures have not automatically produced leadership at every technological stage. Leading-edge semiconductor manufacturing is especially difficult to reproduce because technological capability depends on accumulated expertise and access to highly specialized equipment, software, intellectual property, materials, and production knowledge. Large subsidies can finance fabrication plants, but they cannot immediately create decades of tacit knowledge or replace every specialized foreign input. This distinction illustrates an important limitation of capital-centered industrial policy. Financial resources can accelerate investment, but technological catch-up depends on whether firms convert those resources into productivity, innovation, and commercially competitive capabilities. Recent systematic research on semiconductor industrial policy likewise finds that national subsidies alone have limited effectiveness when they are not accompanied by strong regional innovation ecosystems and technological linkages (Cai et al., 2026).

China's incentives have changed further as semiconductor trade has become more closely connected to U.S.-China technological competition. Beginning with earlier restrictions and expanding substantially from 2022 onward, the United States imposed controls intended to limit China's access to certain advanced computing chips and semiconductor manufacturing technologies. Cooperation with countries that control other critical parts of the semiconductor

equipment supply chain has increased the significance of these restrictions. For China, this changes the economic calculation behind semiconductor subsidies. Domestic production is no longer encouraged only because policymakers want a larger domestic technology industry. It is also viewed as insurance against the possibility that foreign technologies may become unavailable. Ding (2025) finds that Chinese semiconductor policy under U.S. restrictions has increasingly emphasized domestic substitution and adaptation. In economic terms, external restrictions raise the perceived value of technological self-sufficiency even when producing a technology domestically is initially more expensive than importing it.

At the same time, stronger pressure for self-sufficiency can create significant efficiency costs. Semiconductor production benefits from global specialization because different regions have developed comparative advantages in different stages of the value chain. Attempting to reproduce each of these capabilities domestically can require large investments in industries where local firms initially operate at higher costs or lower productivity. Government support can also create soft budget constraints, allowing unsuccessful firms or projects to survive longer than market conditions would otherwise permit. Local governments competing to establish semiconductor clusters may duplicate investments, and access to subsidized capital can encourage entry by firms that lack the technological capability required for long-term success. China therefore demonstrates both the potential strength and the central risk of aggressive semiconductor industrial policy. Sustained public support can accelerate technological catch-up and create domestic capabilities, but the amount of capital invested is not itself evidence of economic efficiency. The more useful measure is whether subsidies eventually produce firms and technologies capable of operating competitively without continuous protection or extraordinary government support (Koopman & Huang, 2025; Patrignani et al., 2025).

6. Comparative Analysis: United States, European Union, and China

The three cases reveal a significant convergence in economic policy. Despite major differences in political institutions and economic systems, the United States, European Union, and China have all concluded that market outcomes alone are insufficient to determine the geographic distribution of semiconductor capacity. Each now uses public resources to influence investment, research, technological development, and supply-chain organization. This convergence is important because industrial policy was previously more strongly associated with state-led development models, while the United States and European Union generally placed greater emphasis on market allocation and international specialization. The semiconductor industry illustrates how concerns about resilience, technological spillovers, and national security have changed this calculation. Governments increasingly accept some loss of short-term allocative efficiency in exchange for what they expect to be greater long-term resilience and technological capacity (Koopman & Huang, 2025; Patrignani et al., 2025).

However, the economic problem each region is attempting to solve is different. The United States is seeking to rebuild manufacturing capacity while preserving existing strengths in semiconductor design, intellectual property, and manufacturing equipment. Its policy can

therefore be characterized primarily as a combination of reshoring and technological leadership. The European Union is attempting to reduce selected dependencies while building upon existing strengths in equipment, automotive and industrial semiconductors, research, and specialized manufacturing. Its strategy is more explicitly framed around strategic autonomy within continued international interdependence. China, in contrast, faces stronger technological dependence across several advanced parts of the semiconductor value chain and has increasingly emphasized localization and self-sufficiency. These differences mean that superficially similar subsidies may have different economic objectives. A subsidy for a new U.S. fabrication facility may be intended to diversify leading-edge production geographically, while a Chinese subsidy for domestic semiconductor equipment may be designed to replace access to a foreign technology that could be restricted.

The regions also differ in how industrial policy is administered. U.S. semiconductor incentives are organized substantially through federal legislation, tax policy, and competitive awards, although state and local governments provide additional incentives. European policy is divided between EU institutions and member states, producing a more decentralized structure constrained by European competition and state-aid rules. China operates through a combination of central government strategy, state-linked investment funds, provincial and local governments, state-owned financial institutions, and private or partially state-supported firms. These institutional differences affect both coordination and accountability. Centralized strategic direction can mobilize investment rapidly but can also amplify errors if resources are directed toward poorly selected projects. More decentralized systems may provide greater competition among projects but can produce duplication and subsidy competition between jurisdictions. There is consequently no obvious institutional model that eliminates the information problems inherent in industrial policy.

A particularly important difference concerns the meaning of resilience. For the United States and European Union, resilience generally means reducing excessive concentration while maintaining substantial international supply-chain relationships with allies and trading partners. Neither economy appears likely to reproduce the entire semiconductor supply chain domestically. China's policy has moved closer to technological self-reliance because export controls have demonstrated that access to some foreign technologies cannot be assumed. Nevertheless, even China remains integrated into global semiconductor markets. This highlights the limitations of the term "self-sufficiency." Modern semiconductor production involves such a large number of specialized inputs that complete national independence would be extremely expensive and technologically difficult. A more realistic policy objective is therefore selective redundancy: maintaining alternative sources or domestic capabilities in areas where disruption would impose especially large economic or security costs (García-Herrero & Poitiers, 2022; Sutter et al., 2023).

The comparison also reveals a collective-action problem. A semiconductor subsidy can appear rational from the perspective of an individual government while producing less desirable outcomes when many governments adopt the same strategy simultaneously. If one country

subsidizes production to correct a genuine vulnerability, additional capacity may increase resilience. If many countries subsidize similar facilities, however, the result can be excess capacity and escalating competition for the same private investment. Firms can benefit by locating projects where governments provide the most favorable packages, while taxpayers absorb an increasing portion of investment costs. Patrignani et al. (2025) describe the semiconductor ecosystem as being reshaped simultaneously by subsidies and restrictions, making investment decisions increasingly dependent on public policy rather than production economics alone. This does not mean semiconductor subsidies are necessarily inefficient. It means their effectiveness must be evaluated relative to the policies of other countries, because semiconductor industrial policy creates cross-border effects that no individual government fully controls.

7. Economic Benefits and Risks of Semiconductor Subsidies

One of the strongest potential benefits of semiconductor subsidies is increased supply-chain resilience. The economic disruption created by a semiconductor shortage can greatly exceed the market value of the missing chips because downstream producers may be unable to complete much more valuable products without them. In this setting, additional production capacity or geographic diversification can provide an insurance benefit to the wider economy. Firms making private location decisions may not capture this benefit, creating a possible rationale for public intervention. Semiconductor subsidies can also encourage companies to maintain capacity that appears relatively expensive during normal conditions but becomes highly valuable during a disruption. From this perspective, comparing subsidized domestic production only with the normal-market price of imported chips understates the potential benefit. The relevant comparison should also incorporate the expected economic cost of severe shortages and supply interruptions (Sutter et al., 2023).

A second potential benefit is the creation of technological and regional spillovers. Semiconductor facilities require specialized engineers, technicians, suppliers, research relationships, infrastructure, and complementary services. A major investment can therefore generate economic activity beyond the company receiving the original subsidy. Semiconductor clusters may improve labor-market matching, encourage supplier entry, and facilitate knowledge transfer between firms and research institutions. Johnston and Huggins (2023) emphasize the role of regional clusters in determining whether European semiconductor industrial policy can create sustainable economic benefits. A similar logic underlies U.S. investments in research centers and workforce development. If subsidies contribute to an ecosystem that continues attracting private capital and generating innovation, their social return may exceed the direct financial return generated by the subsidized facility itself.

However, subsidies also create an opportunity cost because government resources used to support semiconductor firms cannot simultaneously be used elsewhere. The relevant economic question is therefore not whether a semiconductor factory creates jobs or output. Almost any large investment will do so. The question is whether public funds generate greater social benefits

through semiconductor subsidies than through alternative investments or through leaving resources in the private economy. This becomes particularly important when subsidies are offered to large, profitable multinational companies that might have invested even without government assistance. If a project would have occurred in substantially the same form without public support, the subsidy produces a large transfer to the firm but relatively little additional investment. Evaluating additionality, meaning the extent to which government support actually changes firm behavior, is consequently essential for assessing industrial policy.

A further risk is that political incentives can distort project selection. Governments may favor highly visible fabrication plants because factory announcements generate immediate political benefits, even when less visible investments in research, workforce development, equipment, materials, or infrastructure might produce higher long-term returns. Industrial policy can also create lobbying incentives as firms compete to qualify for government support. Once a subsidized facility has been established, policymakers may face pressure to continue supporting it even if its economic performance is weak because closure would create concentrated local job losses. These dynamics can transform temporary support for market failures into persistent protection of uncompetitive production. China's experience with local semiconductor investment illustrates some of these risks, but they are not unique to China. Any government attempting to select strategically important projects faces imperfect information about future technologies, demand, and commercial viability.

Finally, semiconductor subsidies can generate international externalities through trade and investment diversion. A new factory attracted to one country may represent genuinely additional global capacity, but it may also represent investment moved from another location because the receiving government offered a larger incentive. When multiple governments respond by increasing their own subsidies, industrial policy can develop characteristics of a bidding contest. The resulting equilibrium may leave the geographic distribution of production only moderately changed while significantly increasing the share of project costs borne by governments. The economic case for semiconductor subsidies is therefore strongest when they address identifiable market failures, create durable technological spillovers, or reduce clearly defined vulnerabilities. It is weakest when subsidies become an objective in themselves or when success is measured primarily by headline investment announcements rather than productivity, innovation, resilience, and long-term commercial viability.

8. Discussion

The comparison of the United States, European Union, and China suggests that the current expansion of semiconductor subsidies cannot be explained by a single economic objective. Governments are simultaneously attempting to increase supply-chain resilience, promote innovation, attract investment, strengthen domestic production, and respond to national-security concerns. These objectives overlap, but they are not identical. A policy that is effective at attracting fabrication capacity may not necessarily maximize innovation, while a policy designed to reduce dependence on foreign suppliers may increase production costs. The

central economic challenge is therefore one of policy design. Semiconductor subsidies are most defensible when they address identifiable market failures or vulnerabilities and generate benefits that private firms would not fully consider when making investment decisions. They are more difficult to justify when they primarily relocate production between countries or support facilities that remain commercially dependent on government assistance (Koopman & Huang, 2025; Patrignani et al., 2025).

One of the clearest findings of this comparative review is that semiconductor resilience should not be equated with complete self-sufficiency. The semiconductor value chain developed through decades of international specialization, with different countries and firms establishing strong positions in design, fabrication, equipment, materials, assembly, and testing. Attempting to reproduce all of these capabilities within a single economy would sacrifice many of the efficiency gains generated by specialization. The more economically realistic objective is diversification at strategically important points of the supply chain. For the United States, this means rebuilding selected manufacturing capabilities while maintaining its existing strengths in design and semiconductor equipment. For the European Union, it means combining additional capacity with established strengths in equipment, research, automotive chips, and specialized semiconductor production. China faces stronger incentives for localization because restrictions on access to foreign technologies have increased the economic value of domestic alternatives. Even in the Chinese case, however, complete semiconductor autonomy remains difficult because of the industry's technological complexity and international interdependence (Ding, 2025; García-Herrero & Poitiers, 2022).

The analysis also suggests that the scale of a subsidy program is an incomplete measure of industrial-policy effectiveness. Large public commitments can generate investment announcements relatively quickly, but the more important outcomes emerge over a much longer period. Semiconductor fabrication facilities must remain productive after construction, workers must acquire specialized skills, domestic suppliers must become competitive, and research investments must contribute to future technological generations. This distinction is particularly important because semiconductor technology changes rapidly. A facility that appears strategically important when a subsidy is awarded can become less valuable if demand changes or manufacturing technology advances. Policy evaluation should therefore emphasize productivity, technological capabilities, private follow-on investment, knowledge spillovers, and the durability of semiconductor clusters rather than simply the nominal value of announced projects. Recent scholarship on semiconductor ecosystems similarly emphasizes that technological sovereignty depends on networks of firms, institutions, skills, and research capabilities rather than manufacturing capacity alone (Cai et al., 2026).

Another important implication concerns the interaction between subsidies and international economic competition. The United States, European Union, and China are not designing semiconductor policies independently. Each government's actions affect the incentives facing the others. U.S. subsidies encourage competing economies to offer their own incentives, while U.S. export controls strengthen China's incentives to invest in domestic substitutes.

European policymakers, meanwhile, face pressure to prevent investment from moving toward more heavily subsidized locations in the United States or Asia. This creates the possibility of a semiconductor subsidy race in which government intervention becomes partly defensive. From the perspective of an individual jurisdiction, matching foreign subsidies can appear rational because failing to do so could result in lost investment. Collectively, however, escalating incentives may transfer more economic rents to semiconductor firms without generating equivalent social benefits. The semiconductor case therefore demonstrates how industrial policy can produce strategic interactions similar to those found in international tax competition and trade policy (Patrignani et al., 2025). Simultaneously, the evidence does not support the opposite conclusion that governments should simply return to the pre-pandemic structure of semiconductor globalization. The concentration of critical production stages represents a genuine economic vulnerability, and recent geopolitical restrictions demonstrate that access to technology can be influenced by political relationships as well as market prices. The appropriate policy question is therefore not whether governments should intervene at all, but how intervention can be disciplined. Subsidies should be connected to measurable objectives, coordinated with research and workforce policies, and periodically evaluated against changing technological and market conditions. International coordination among economically aligned countries may also reduce unnecessary duplication by allowing governments to specialize while diversifying the most critical vulnerabilities. Under this approach, semiconductor policy would aim for resilient interdependence rather than either unrestricted dependence or complete national self-sufficiency.

9. Conclusion

The United States, European Union, and China have entered a new period of semiconductor industrial policy in which governments are willing to commit substantial public resources to influence where and how strategically important technologies are produced. Although all three regions use subsidies to strengthen their semiconductor industries, their underlying economic positions and objectives differ. The United States seeks to restore selected manufacturing capacity while maintaining technological leadership, the European Union seeks greater strategic autonomy while building on specialized areas of existing strength, and China has pursued a broader strategy of technological self-sufficiency that has become increasingly important as access to some foreign technologies has been restricted. This comparative review finds that semiconductor subsidies can have a legitimate economic rationale when they address knowledge spillovers, high investment barriers, concentrated supply-chain vulnerabilities, and strategic risks that private firms may not fully incorporate into their decisions. However, the existence of these benefits does not mean that all semiconductor subsidies are economically efficient. Poorly targeted support can encourage duplication, protect uncompetitive production, create dependence on continued government assistance, and contribute to costly international subsidy competition. The long-term success of semiconductor industrial policy will therefore depend less on which government spends the most and more on whether public support creates productive capacity, technological knowledge, skilled labor, competitive firms, and resilient

supply networks that remain valuable after the initial subsidies have been provided. The semiconductor industry consequently offers a broader lesson for modern industrial policy: economic resilience does not require eliminating global interdependence, but it may require governments to manage its most significant vulnerabilities more deliberately.

Works Cited

European Commission. (2022). European Chips Act: Communication, regulation, joint undertaking and recommendation. European Commission. https://digital-strategy.ec.europa.eu/en/policies/european-chips-act

European Parliament & Council of the European Union. (2023). Regulation (EU) 2023/1781 of the European Parliament and of the Council of 13 September 2023 establishing a framework of measures for strengthening Europe's semiconductor ecosystem. Official Journal of the European Union, L 229, 1–53. https://eur-lex.europa.eu/eli/reg/2023/1781/oj

García-Herrero, A., & Poitiers, N. (2022). The EU Chips Act: How to make Europe fit for the digital age. Bruegel. https://www.bruegel.org/policy-brief/eu-chips-act-how-make-europe-fit-digital-age

Huggins, R., Johnston, A., Munday, M., & Xu, C. (2023). Competition, open innovation and growth challenges in the semiconductor industry: The case of Europe's clusters. Science and Public Policy, 50(3), 531–547. https://doi.org/10.1093/scipol/scad005

Hyatt, K., Ryle, P. M., & McKnight, M. A. (2025). Semiconductor production, geopolitics and the CHIPS Act of 2022: A theoretical analysis. Digital Policy, Regulation and Governance, 27(1), 1–18. https://www.emerald.com/dprg/article/27/1/1/1240520

Johnston, A., & Huggins, R. (2023). Europe's semiconductor industry at a crossroads: Industrial policy and regional clusters. European Urban and Regional Studies, 30(3), 255–262. https://doi.org/10.1177/09697764231165199

Khan, S. M., Mann, A., & Peterson, D. (2021). The semiconductor supply chain: Assessing national competitiveness. Center for Security and Emerging Technology, Georgetown University. https://cset.georgetown.edu/publication/the-semiconductor-supply-chain/

Kleinhans, J.-P., & Baisakova, N. (2020). The global semiconductor value chain: A technology primer for policy makers. Stiftung Neue Verantwortung. https://www.interface-eu.org/publications/the-global-semiconductor-value-chain

Organisation for Economic Co-operation and Development. (2023). Risks and opportunities of reshaping global value chains. OECD Publishing. https://www.oecd.org/

Patrignani, L., Savini Zangrandi, M., & Schivardi, F. (2025). Between sanctions and subsidies: Reshaping the semiconductor ecosystem. Bank of Italy Occasional Papers. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5252872

Sutter, K. M., Sargent, J. F., Jr., & Singh, M. (2023). Semiconductors and the CHIPS Act: The global context (CRS Report No. R47558). Congressional Research Service. https://crsreports.congress.gov/product/pdf/R/R47558

The White House. (2022). CHIPS and Science Act will lower costs, create jobs, strengthen supply chains, and counter China. https://www.whitehouse.gov/

U.S. Congress. (2022). CHIPS and Science Act of 2022, Pub. L. No. 117-167, 136 Stat. 1366. https://www.congress.gov/117/plaws/publ167/PLAW-117publ167.pdf

U.S. Department of Commerce. (2023). CHIPS for America: A strategy for the CHIPS for America Fund. National Institute of Standards and Technology. https://www.nist.gov/chips

Varas, A., Varadarajan, R., Goodrich, J., & Yinug, F. (2021). Strengthening the global semiconductor supply chain in an uncertain era. Boston Consulting Group & Semiconductor Industry Association. https://www.semiconductors.org/strengthening-the-global-semiconductor-supply-chain-in-an-uncertain-era/

Xiong, W., Wu, D. D., & Yeung, J. H. Y. (2025). Semiconductor supply chain resilience and disruption: Insights, mitigation, and future directions. International Journal of Production Research. https://doi.org/10.1080/00207543.2024.2387074

Yeung, H. W.-c. (2024). From regional to global and back again? A future agenda for regional evolution and (de)globalised production networks in regional studies. Regional Studies. https://doi.org/10.1080/00343404.2024.2316856

Gender Inequality in Business Leadership: Hiring and Pay Gaps By Nakshatra Sakhamuri

Abstract

This project examines the prejudices against women in senior leadership roles within the business field. Currently, women continue to be underrepresented in senior management positions, indicating the level of wage gaps and evaluation standards that have been formed through societal and cultural barriers that hinder further advancement. This marginalization stems from social and systemic prejudice, and from unrepresentative skill levels, education, and experience. This paper explores what factors contribute to the underrepresentation of women in senior leadership positions, whether women are compensated equally and evaluated under similar standards as men, and how this underrepresentation affects company performance and economic outcomes. The project also examines the ratio of men versus women in senior positions, and how women of color are hired even less than white men or white women. By excluding workers based on gender or race, companies deprive themselves of a range of ideas and perspectives, leading to less innovation, teamwork, and problem-solving ability, slowing innovation and hindering economic development. Systemic biases, cultural prejudices, and unequal evaluation standards drive the underrepresentation of women in leadership. These factors limit a company's innovation and corporate profitability.

Keywords: Wage Disparity, Gender Diversity, Workplace Bias, Gender Inequality, Executive Representation, Structural Barriers

Introduction

Women remain underrepresented in senior leadership positions within business management, despite increasing participation in the workforce over the past several decades (McKinsey & Company and LeanIn.org). While women have made significant progress in education and employment, they continue to hold a disproportionately small share of executive and top management roles across industries. In addition to representation gaps, research has consistently documented compensation disparities between men and women, as well as differences in promotion rates and performance evaluations.

These disparities are influenced by structural and cultural factors. Societal expectations, historical norms regarding gender roles, and workplace biases can shape hiring, promotion, and leadership opportunities. Women, particularly of color, often face compounded barriers related to both gender and race. As a result, leadership pipelines remain uneven, and access to executive decision-making roles is not equally distributed.

Understanding the scope of gender disparities in business leadership is critical not only as a matter of equity but also because leadership composition can influence organizational performance, innovation, and economic outcomes. This context provides the foundation for examining why women remain underrepresented in leadership and how these disparities affect business management and the broader economy. Around 2.7 billion women are restricted from

having the same job options and decisions as men (UN Women). This occurs due to traditional gender norms, concerns about the safety of women in the workplace, and bans on women imposed by nations.

Currently, women continue to be underrepresented in senior management positions, indicating the enduring level of wage gaps and identity standards that are formed through societal and cultural barriers that hinder further advancements and liberties. This marginalization of females stems from social and systemic bias and prejudice, unrepresentative of skill levels, education, and experience.

This phenomenon is shown through McKinsey's senior partners' analysis titled "*The Broken Rung*." Kweilin Ellingrud, Lareina Yee, and María del Mar Martínez emphasize the disadvantages faced by women, even though they are subject to having averagely higher GPA. However, when they try to apply for a role in the workplace, men are usually favored more. Among every 100 men who are promoted, only around 80 women are promoted, while 60 Black women and 74 women of color are promoted. (McKinsey & Company). This "broken rung" aberration causes the trailing inequalities in career occupation alongside men.

Men are gaining more value for the education that women are pursuing at exceptionally higher rates, leading to more men in the workforce at a remarkable 53% (74.86 million), while women are lagging at 47% (58.81 million) for full-time workers. Men are consistently favored in the labor market disproportionately, and this deviation makes it onerous to catch up to these advantaged males (McKinsey & Company).

As a result, men take greater advantage of and benefit from the favored system and the value the altered imbalance has. In essence, this management deficit corroborates that the corporate environment is not lacking female competence, but instead equality.

My paper identifies the systemic deficits and the trajectory towards impartiality. This research paper explores the following questions:

What factors contribute to the underrepresentation of women in senior leadership positions in business management? How does the deficiency of women in leadership positions affect the overall company performance and success outcomes, and are women provided with adequate compensation for their work compared to the standards that are set for men?

Systematic biases, cultural prejudices, and unequal evaluation standards drive the underrepresentation of women in leadership; these methods limit company innovation, suppress corporate profitability, and hinder economic development.

Literature Review

Representation of Women in Leadership

Representation of women in leadership is categorized in three different sections: the corporate sphere, education, and the political sphere. In the corporate world, there continues to be imbalances in gender composition, especially at senior levels in the workplace. While women consequently showcase higher levels of academic enrichment and commitment, they persistently

face barriers to their career advancement. The American Institute for Boys and Men states that 8.9 million women are currently college undergraduates, while only 6.5 million men are (American Institute for Boys and Men). The leadership gap is especially shown through both the public and private sectors.

According to the Pew Research Center, women make up 28% of the U.S. congressional members and about a third of state legislators. Additionally, only eight Democratic and four Republican women serve as governors nationwide. While former President Joe Biden's Cabinet consisted of 12 women out of the 25 in Cabinet or Cabinet-level positions (48%), 11% of Fortune 500 company CEOs are women, and 30% of Fortune 500 board members are females (Pew Research).

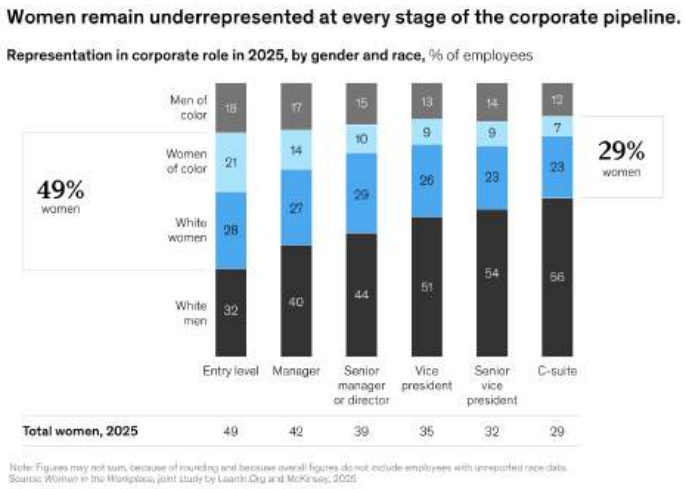


Figure 1. Representation of Women in corporate positions by level. Source: McKinsey & Company and LeanIn.Org, *Women in the Workplace 2025*.

This corporate constraint starts to emerge as employees try to rise through the ranks. Data from McKinsey & Company (Figure 1) showcases a decline in female representation through a company's power structure. Women make up 49% of the labor force in entry-level positions, and as they attempt to climb this ladder, their numbers decline immensely to 42% for a managerial position, 39% for senior manager/director, 35% for vice president, and 32% for senior vice president. As these female employees try to reach the highest executive role, known as the C-suite position, female representation drops to 29% (Figure 1). Although 29% is the total representation of women, including women of color, the real number of women of color drops to almost 7% out of 29%, showcasing the biases among companies (McKinsey & Company). These statistics prove that at each rank of a company's hierarchy, the talent that women provide is overlooked at every step of the way.

Compensation and Pay Gaps

Representation of women in leadership is categorized in three different sections: the corporate sphere, education, and the political sphere. In the corporate world, there continues to be imbalances in gender composition, especially at senior levels in the workplace. While women consequently showcase higher levels of academic enrichment and commitment, they persistently face barriers to their career advancement. The American Institute for Boys and Men states that 8.9 million women are currently college undergraduates, while only 6.5 million men are (American Institute for Boys and Men). The leadership gap is especially shown through both the public and private sectors.

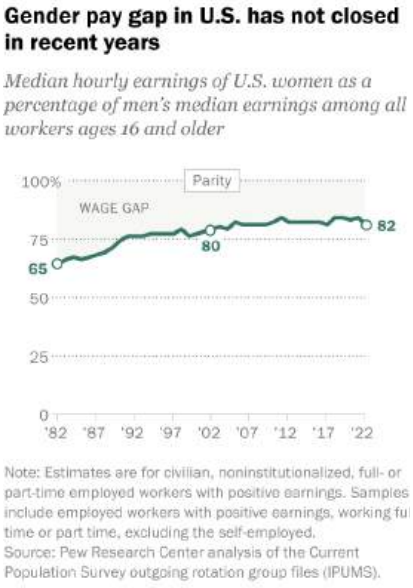


Figure 2. Average hourly wages of US Women from 1982-2022. Source: Pew Research Center, The *Enduring Grip of the Gender Pay Gap*, 2023.

The absence of women in senior executive roles is tied to wage inequality, fewer opportunities, and less female advocacy by women. The wage gap is caused by the underrepresentation of women in senior positions, as men hold the majority of high-paying executive roles, keeping the average female salary lower than that of men. The equality of both men and women in the workplace leads to high-level and high-impact roles, including higher income and technological advancements. However, the lack of women drives the pay gap because fewer women are in these higher-paying roles. The Pew Research Center (Figure 2) shows that the gender wage gap has remained stagnant for decades. For every dollar earned by men, women on average earned only 82 cents in 2022. In 2002, women only earned 80 cents to that dollar ratio, which is only a minimal increase, even 20 years later (Figure 2). When surveyed about this discrepancy, half of US adults attributed this to the treatment imposed on women by employers (Pew Research). These statistics showcase a constant cycle of the lack of women in executive roles, which keeps the national average of income/wages earned by women at a lower rate.

Career Trajectories and Gender Diversity

The salary gap between men and women is heavily interconnected due to a lack of support and opportunities. Research from McKinsey & Company and LeanIn.Org shows that entry-level women are half as likely to gain sponsorship (when a senior employee actively advocates for a junior employee's career advancement) as men, which immensely helps advocate for career advancement and promotional opportunities among companies (McKinsey & Company and LeanIn.Org). This advocacy helps employees significantly advance their careers, and women seem to be at a disadvantage since sponsors help double the chance of advancement compared with what would occur without a sponsor. Furthermore, women are also less likely to gain the ability to get important training opportunities than men. Data (Figure 3) show that 34% of senior-level men were offered the chance to participate in career and leadership training, while only 20% of women were offered the chance, altering the career progression that they can largely benefit from (Figure 3). These persisting inequalities alter career trajectories, preventing women from achieving their full potential.

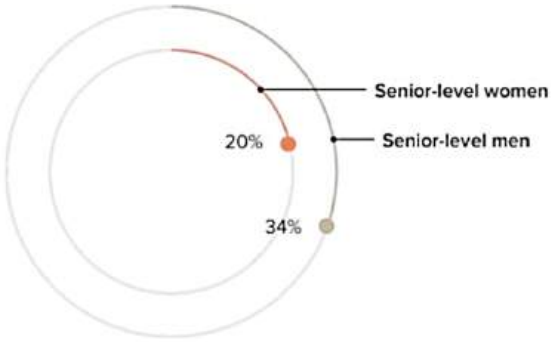


Figure 3. Percentage of senior-level employees offered career and leadership training, based on gender. Source: McKinsey & Company and LeanIn.org, *Women in the Workplace 2025*.

Structural and Social Barriers

The disparities women face in reaching these higher roles stem from cultural and structural obstacles. According to the National Institute of Health, gender biases affect women's leadership especially in higher roles. A double standard is established among women: when women display assertive traits, they are often called “bossy” and “aggressive”; however, when men possess these same traits, they are often commended. In other areas of the world, these same biases and hypocrisy are shown. Research from Pakistan showcases similar discriminatory barriers, including nepotism and other factors, with 65% of women emphasizing gender‑specific obstacles in their career advancement. (National Institute of Health).

These structural barriers are also constantly compounded by racial discrimination, where specific ethnicities are left with less support than women regularly receive. McKinsey & Company and LeanIn.Org showcase that mid-level Asian women are consistently receiving less

assistance to build their careers among their senior colleagues. Compared to Black, White, and Latina women, who average career advancement tips, sponsorships, and more at 34%, Asian women receive those same career advancement tips and benefits 24% of the time, missing out on career support that could boost their careers. This deficit occurs from racial and gender biases. For example, Asian women are frequently held back through prejudices that they don't speak out, but when they do speak out, they are judged and held back, making career advancement extremely hard.

Although overt discrimination has been driven out of most up-to-date and modern organizations, subtle and concealed gender discrimination still exists, shown through various work practices and cultural norms according to the National Institute of Health (National Institute of Health). Discrimination against women is so ingrained in companies that it is difficult to detect. These deconstructions are going to be essential to changing bias that can remove the barriers and preconceptions which keep women from moving into senior leadership.

Methods

Research Type

This study was based on a literature-based economic analysis that spans both qualitative and quantitative research to investigate the discriminatory barriers due to gender throughout business management. The research gathered is not experimental but is based on secondary sources and data collection. The methodology is based on reputable, low-bias sources, including non-governmental websites. Using multiple sources allows for a larger scope of analysis among different organizations, time periods, and various viewpoints.

The data/research collected in this study were based mainly on global management consulting firms, authoritative sources, and institutional sources. Most of the sources span multiple sectors, such as economic, labor, and global. The main references used throughout this study were McKinsey & Company, which focuses on delivering primary industry- and economic-based research (McKinsey & Company). This source uses mainly quantitative data and helps organizations and companies to achieve lasting improvements and navigate through complex market trends and advancements. Pew Research Center, a nonpartisan, secondary source that focuses more on factual statements, uses public polls, demographic research, and other studies to inform its audience (Pew Research). To add on, UN Women is an authoritative source on gender statistics and women's rights. This source intends to close the gender gap and advance equality and support for women globally (UN Women). The National Institute of Health, which is an authoritative primary source, is evidence-based and advocates for equality among all individuals and eliminating systemic discrimination (National Institute of Health). Hult International Business School, which is an institutional and authoritative source, focuses on gender equality in the workplace and conducts first-hand experiments (Hult International Business School).

Data Analysis

The data was grouped into multiple sub-segments to find trends and commonalities among the sources. These sub-segments organized the research into wage gaps, the percentage of senior-level women, social and cultural barriers, and company performance with diversity. Research from all sources was used to compare and contrast trends in the market. The comparisons were used to verify the validity of workplace gaps. Using quantitative data, the study analyzes a company's financial return before and after diversity was implemented in the corporate economy by improving innovation, communication, and the overall company environment.

Limitations

This research paper utilizes a literature-based technique rather than an empirical approach. Since this study depends on secondary sources, I was unable to collect authentic and original data or interview any officials directly. In some instances, audiences can interpret certain definitions of senior leadership in different ways, leading to misconceptions and variances. The paper was also limited since most companies do not disclose their gender wages or promotion statistics, which means that this paper is restricted to reports and outside sources. However, this study largely displays these inconsistencies by using large-scale sources that are known to conduct their own research and experiments. This study consists of visual aids and a comprehensive analysis to contextualize concepts and definitions.

Results

Data from many research centers (such as Pew Research Center), government agencies (like the National Institute of Health), and independent organizations (including McKinsey & Company) show the distinct corporate hierarchy and barriers to female representation. This research reveals that while women possess higher educational achievements, the disparities are shown by cultural prejudices, unequal access to opportunities, and more.

Leadership Representation Trends

The relationship between a company’s structure and its female involvement can be revealed by the “broken rung.” This definition, previously stated in the introduction, correlates with the number of women compared to men who get hired and promoted. According to McKinsey & Company, although women make up half of the workforce overall, their representation drops with each promotion, with nearly a third working in the C-suite and even less for people of color. In top positions, women hold only one-ninth of CEO positions, one-third of seats on Fortune 500 boards, and one-fourth of seats in Congress, according to the Pew Research Center.

Compensation Findings

The wage gap correlates with the deficit of gender representation in executive leadership roles. Over the span of two decades, the average female wage has gone up only 2.5%, according to Pew Research. (Pew Research). In addition, UN Women states that globally, the wage gap based on gender is estimated to be around 20%. To put it in simpler words, women earn only 80% of what men earn, but these numbers don't fully exemplify the extent of gender gaps and the profundity of the systemic inequalities (UN Women). The lack of equal training opportunities is showcased through the unequal rate of senior-level men receiving training and more opportunities, while one-fifth of these senior women receive them.

Performance Outcomes

Increasing gender diversity improves the economy and a company's revenue. By expanding across every race, gender, and orientation, the US could bring in $7 trillion into the economy and GDP. Companies with gender diversity earn 15% more than the median. For every 10% increase in diversity at senior levels, EBIT (Earnings Before Interest and Taxes) continues increasing by 3.5%. (McKinsey & Company). Additionally, companies with three or more women in leadership roles are more likely to achieve strong operational and financial outcomes, and diversity is an indicator of high-performing companies (Hult International Business School). Businesses with more diverse leadership are more statistically likely to increase profits and operate at a higher level.

Barriers Identified Across Studies

These limitations are established through historically ingrained societal barriers rather than any talent deficiencies. To add on, racial prejudices are especially apparent among Asian women, with less than a quarter of them receiving career-building connections from senior colleagues (McKinsey & Company). A shift toward reducing these systemic and cultural biases is critical to increase corporate profit and the global economy. While systemic and racial barriers prevent businesses from profiting, increasing corporate diversity is essential to propel economic growth.

Discussion

The presented research findings indicate that the underrepresentation of women in corporate hierarchies is predominantly a result of structural inefficiencies instead of a lack of experience and skills. Through large syntheses of these patterns, the causes of a company's leadership are shown. These restrictions on diversity and imbalances can also lead to a company's economic system shattering.

Economic Consequences of Underrepresentation

The long-term penalties imposed by corporations for excluding women are especially harmful and disruptive to these executive teams. This gender shortage and removal of talent occurs through the lack of female recruitment during the early stages of promotion. This barrier

prevents highly skilled and competent employees from reaching decision-making roles. Companies with undiversified leadership constantly experience intellectual lag, which limits creativity, innovation, and stalls other beneficial skills. The homogeneous use of corporate leadership systems creates a deficit in critical cognitive and motor skills, including circumstantial thinking and problem-solving (Harvard Division of Continuing Education). Companies that lag in gender and racial diversity are around 66% less likely to reach their full potential. This deficit of unique perspectives restricts and limits innovation and can carry financial burdens (Forbes).

To add on, women make up the majority of consumer purchases. According to Forbes, women make up 85% of purchases in recent years (Forbes). Male-dominated executive boards often face many limitations in innovation and consumer outreach. However, integrating more female perspectives in executive roles can help overcome this creative gap. This gender diversity can largely help businesses align their strategies and outcomes with current trends and develop a profitable market strategy. Gender-balanced and racially diverse teams promote a collective and collaborative workplace environment, which benefits a company's assets, communication, and strategies.

Broader Economic Impacts

The restriction of female advancement through training and other assistance enforces serious employment bottlenecks, which reduce the possibility of increasing global Gross Domestic Product (GDP). The current labor market continually works through a workplace imbalance. While most countries spend fortunes educating women, these inequalities limit the return of money put into women, without being able to receive a return on investment. Furthermore, when educated women are pushed into lower positions, it leads to fewer economic opportunities and returns when high-level education and skills are left out of executive decision-making. Instead of working on this disparity, males who are often less qualified are constantly rewarded (McKinsey & Company and LeanIn.Org), especially opportunity-wise. This lowers the quality of corporate leadership, while leaving highly valued skills underappreciated and underutilized.

As gender wage gaps, lack of social protections, and regulatory barriers combine, they eliminate the authority of females, consumer-wise, and their purchasing capacity. Although the main consumers and buyers are women, these restrictive hurdles suppress the earnings of women. This suppression makes an economy suffer extensively from less consumer spending and stagnant or decreased GDP growth.

Evaluation Standards and Organizational Culture

Workplace evaluations are constantly influenced by a double standard, especially prominent through behaviour. When women try to influence or exhibit decisive and assertive behaviours like their male colleagues, they face severe backlash and are constantly penalized opportunity-wise. This standard creates a gender barrier that forces women to act compliantly, which can limit promotional advancement.

These biased standards limit the utilization and access to executive pathways needed to level up the hierarchy. Historically and in the present, women are consequently undermined for their competency, leading to many entry-level women being overlooked for corporate advancement opportunities, including sponsorship among executives and training.

Moreover, women of color suffer from large cultural prejudices that can harm their imposition of rules, their societal expectations, and more. These cultural barriers prove the systems of privilege and prejudice rather than meritocracy.

Conclusion

Summary of Key Findings

This paper showcases that the deficits of females in executive roles are driven by structural constraints instead of a lack of skills and education. This inequality limits female advancement and promotional opportunities during the initial levels of the corporate hierarchy, creating an imbalanced access to opportunities. A constant gender wage gap and unequal access to resources such as sponsorship and training are linked to representation disparities.

Implications for Businesses

Retaining a rigid executive structure often has negative impacts on modern companies through their operations and finances. The absence of women often isolates businesses from their main customer bases and hinders product/market innovation. Teams that are gender-diverse, especially in higher roles of management, largely maximize assets, boost a business's profitability, and improve collaboration and communication in the workplace.

Policy and Organizational Recommendations

Businesses must stray away from gender-specific norms to achieve a fully diverse, equitable company. Companies should make policies that ensure equal allocation of funds for both males and females. These policies would reinstate equal investments for executive sponsorships and opportunities for training and leadership programs for women. In addition, companies should add a legislative framework designed to work to change as well as inform employees of the managerial constraints.

Final Reflection on Economic Growth and Equity

To conclude, abolishing gender inequalities is not an issue of social equality, but a need for change to institute economic stability globally. By minimizing the gender labor gaps, the usage of idle female talent and experience can benefit the national GDP extensively by adding trillions to the economy. The change to structural equity can vastly improve a business's performance and profitability, transform gender roles in corporations, and increase economic outcomes.

Works Cited

Carter, Sandy. "Who Runs the World? Women Control 85% of Purchases, 29% of STEM Roles." Forbes, 7 Mar. 2024, https://www.forbes.com/sites/digital-assets/2024/03/07/who-runs-the-world-women-control-85-of-purchases-29-of-stem-roles/.

Carucci, Ron. "One More Time: Why Diversity Leads to Better Team Performance." Forbes, 24 Jan. 2024, https://www.forbes.com/sites/roncarucci/2024/01/24/one-more-time-why-diversity-leads-to-better-team-performance/.

Dame Vivian Hunt, et al. "Why Diversity Matters." McKinsey & Company, McKinsey & Company, 2015, https://www.mckinsey.com/featured-insights/diversity-and-inclusion/why-diversity-matters.

Duchene, Kate. "The Transformative Power of Women in Leadership Roles." Forbes, 22 Jan. 2025, https://www.forbes.com/sites/kateduchene/2025/01/22/the-transformative-power-of-women-in-leadership-roles/.

Flipsnack. "Women in the Workplace 2025." Flipsnack, 2025, https://www.flipsnack.com/78C9ACFF8D6/women-in-the-workplace-2025.

Horowitz, Juliana Menasce. "How Americans See the State of Gender and Leadership in Business." Pew Research Center, 27 Sept. 2023, https://www.pewresearch.org/short-reads/2023/09/27/how-americans-see-the-state-of-gender-and-leadership-in-business/.

Johns, Merida L. "Breaking the Glass Ceiling: Structural, Cultural, and Organizational Barriers Preventing Women From Achieving Senior and Executive Positions." Perspectives in Health Information Management / AHIMA, American Health Information Management Association, vol. 10, no. Winter, 2013, p. 1e, https://pmc.ncbi.nlm.nih.gov/articles/PMC3544145/.

Kochhar, Rakesh. "The Enduring Grip of the Gender Pay Gap." Pew Research Center, Pew Research Center, 1 Mar. 2023, https://www.pewresearch.org/social-trends/2023/03/01/the-enduring-grip-of-the-gender-pay-gap/.

Kweilin Ellingrud, et al. The Broken Rung. Harvard Business Press, 2025.

Mary Sharp Emerson. "Why Gender Equity in the Workplace Is Good for Business - Professional & Executive Development | Harvard DCE." Professional & Executive Development | Harvard DCE, 27 Mar. 2020, https://professional.dce.harvard.edu/blog/why-gender-equity-in-the-workplace-is-good-for-business/#Why-Gender-Equity-Matters.

Reeves, Richard, and Will Secker. “Degrees of Difference: Male College Enrollment and Completion.” American Institute for Boys and Men, 29 Mar. 2024, https://aibm.org/research/male-college-enrollment-and-completion/.
Reynolds, Katie. “Women in Business: Advantages, Challenges, and Opportunities Hult Blog.” Hult Blog, 14 Sept. 2017, https://www.hult.edu/blog/women-in-business-advantages-challenges-and-opportunities/.
Schaeffer, Katherine. “For Women’s History Month, a Look at Gender Gains – and Gaps – in the U.S..” Pew Research Center, Pew Research Center, 27 Feb. 2024, https://www.pewresearch.org/short-reads/2024/02/27/for-womens-history-month-a-look-at-gender-gains-and-gaps-in-the-us/.
UN Women. “Facts and Figures: Economic Empowerment.” UN Women, Feb. 2024, https://www.unwomen.org/en/what-we-do/economic-empowerment/facts-and-figures.

The Effect of Emotional Marketing Language on Consumer Brand Attachment
By Vaishnavi Singh

Abstract

Modern branding increasingly relies on marketing language, rather than product features, to build emotional bonds with consumers. Brand attachment has emerged as a strong predictor of loyalty and repeat purchase behavior. Although prior research has examined elements such as narrative structure and pronoun choice, few studies have directly compared specific types of sentiment-based language to determine which are most effective in building brand attachment. This study examines the extent to which sentiment-based marketing language—specifically experience focused messages, symbolic messages, inclusive language and emotional slogans—affects consumers' brand attachments, using Coca-Cola as a case study.

A total of 60 participants, recruited through convenience sampling, completed an 18-question online survey to measure their reactions to Coca-Cola marketing examples using sentiment-based language.

Results showed that sentiment-based language had a generally positive but uneven effect on brand attachment, with occasional consumers responding more strongly than daily consumers. Contrary to expectations, participants rated factual, objective information as more persuasive than sentiment-based language overall. Sentiment-based language also proved more effective for consumers over 25 than among participants under 25, while no clear pattern emerged by gender.

These findings suggest that emotional branding strategies do not affect all consumers equally, and that their effectiveness depends on a consumer's existing relationship with the brand and their age. Future research with larger, more geographically diverse samples and multiple brands could help determine how generalizable these patterns are.

Keywords: sentiment-based language, Coca-Cola, brand attachment, marketing strategistics, emotional branding

Introduction

Today's branding relies on a mix of product features and the way the product is marketed, with companies starting to build stronger relationships through the usage of language rather than just describing features of the product itself. Brand attachment is the emotional bond that ties a brand to a consumer's sense of self. It has become one of the strongest predictors of brand equity: the overall value a brand derives from consumers' awareness, loyalty, and positive perceptions. In fact, brand attachment outperforms brand attitudes in explaining loyalties and repeat purchase behavior (Park et al. 2). Companies that manage to connect with customers emotionally often see those same customers become measurably more valuable, with emotionally engaged customers found to be 52% more valuable than merely customers that were satisfied in terms of revenue and market shares (Magids et al.). As this emotional side of

branding takes on more strategic importance, marketers lean harder into sentiment based, narrative, and identity driven language as the primary tool to build that connection.

The existing research into this field backs the claim up consistently, showing that how a brand speaks to consumers shapes the relationship just as much as the product it's selling. Narrative processing theory argues that consumers interpret brand messages by mapping them into stories they already hold in their memories, meaning that a narratively structured ad can create or deepen consumers' brand connection almost by default (Escalas 169). This is further supported by more targeted linguistics research. In a series of experiments using real brand posts on social media like Facebook, including second person pronouns increased both consumer involvement and brand attitude, an effect driven by consumers self referencing more when referred to as "you" (Cruz et al. 111-112). Collectively, these findings indicate that language itself and not just physical products drive brand attachments in consumers.

However, a caveat of most previous studies is that they looked mostly at broader categories of brand language (narrative structure, pronoun choices), mostly in isolation. There's comparatively little research directly comparing specific types of sentiment based language against each other (experience focused messaging, inclusive phrases etc.) to see which types affect attachment to brands. So this gap isn't just whether emotional language works, it's what kinds work in what capacity and for whom they work for.

This study tries to close the gap by looking at the extent to which sentiment based marketing language affects consumers' brand attachment, utilizing Coca-Cola as a case study. Specifically it asks to what degree experience-focused messages, symbolic messages, inclusive language and emotional slogans strengthen brand identity and consumer loyalty. By surveying consumers on their reactions to real examples of Coca-Cola's own marketing language, this research aims to take the broader theories of narrative and emotional branding and turn them into something more specific, in order to gain a clearer sense of what kind of language builds lasting attachment to brands, and how it achieves that.

That said, this is only a preliminary study, and it comes with many limitations. The sample size. It is fairly small and pulled from limited demographics and a smaller geographic pool, which makes it harder to apply the findings in a more generalized sense. It's also worth noting that the focus on a single brand means the results may say more about Coca Cola's relation with sentiment based language than with businesses as a whole. As a result, the findings here should be taken as more of an exploratory step into the topic rather than a finalized discussion on how emotional language works across different industries.

Research Design

This study relies on a survey as its primary method to test whether sentiment based marketing language affects consumer brand attachment. A survey format is better suited to this research question than approaches such as interviews, as it allows for the collection of a wider, more standardized range of data from multiple types of consumers as opposed to individual perspectives that aren't as standardized in nature. A survey also allows for each participant to be

exposed to the same Coca-Cola marketing examples and uses consistent scales, which makes it possible to compare results across different demographics in a way that qualitative analysis cannot easily replicate.

The participants of this survey were recruited through convenience sampling, primarily through the researcher's school community and the researcher's parents' workplaces in Rocklin, California. In total, 61 participants completed the survey over a two-month period. To participate, respondents needed some baseline familiarity with the brand, as somebody who has never been exposed to the Coca-Cola products or its ads would have had little basis on which to evaluate them. Other than these requirements, no additional criteria excluded particular individuals from participating in the survey. This sample is appropriate for the research question because the goal isn't to represent all consumers nationally, but to test whether sentiment-based marketing language produces a measurable effect on brand attachment at all. A convenience sample recruited through personal networks is a reasonable starting point for this type of exploratory research, although it places clear limitations on the extent to which the findings can be generalized.

The questionnaire for this survey consists of 18 questions, structured to move from general to specific. The first section asks about demographic characteristics, including age, gender, and general soda consumption habits, to establish the composition of the sample. The second section focuses on participants' existing relationship with Coca-Cola, specifically, asking how often they drink it, whether it's their default soda choice, and what factors actually influence that preference. These questions help establish a baseline measure of brand attachment before any language-specific questions come into play. The third section is where participants were shown authentic examples of Coca-Cola's own marketing language across the four categories central to this study:experience-focused messaging, symbolic messaging, inclusive language, and emotional slogans. The section also includes two additional questions using purely informational messaging as an objective point of comparison. Each question in this section is tied directly to one of the study's four hypotheses, allowing responses to be grouped and compared by category to determine which type of language produces the strongest attachment response. All questions in the survey either use multiple-choice or five-point Likert-scale formats, which keeps the data standardized and easier to compare, while also keeping the survey short enough that fatigue isn't a real concern for participants.

The survey was built and distributed through Google Forms, mainly because the platform made it easy to reach a wide range of participants. The survey remained open for responses over a two-month period and took each participant roughly 5 to 10 minutes to complete. At the beginning of the form, participants were informed that the survey was intended for academic research only and that all of their responses would be recorded anonymously. No personally identifiable information was collected beyond the demographic categories. Participants provided their consent by voluntarily completing the form after being informed of the survey's purpose and format upfront.

Results

Hypothesis 1

Hypothesis 1 predicted the use of sentiment based marketing language would have a substantial impact on consumers' brand attachment. To test this, participants were shown experience focused messages modelled after Coca-Cola's own advertising and asked to rate how strongly it increased their likelihood of purchasing the product. Overall, the results point to a generally positive effect, with most participants choosing ratings of 3, 4 or 5 on the five point scale (Figure 1). As shown in Figure 1, about 29.5% of participants reported a moderate impact whilst roughly 23% reported a relatively high impact, suggesting that sentiment based language does resonate with a meaningful portion of consumers rather than just a small minority.

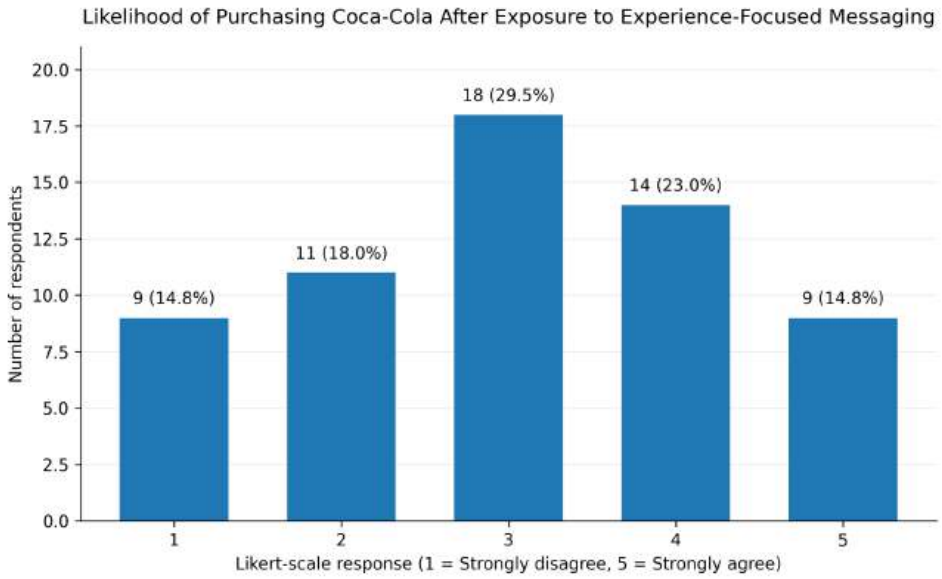


Figure 1. Likelihood of purchasing Coca-Cola in response to experience-focused messages

However, this effect was not consistent across every group. When responses were broken down by how frequently they drank soda, occasional consumers showed the strongest response to sentiment based messaging while daily consumers showed the weakest response.

As illustrated in Table 1, occasional consumers reported the strongest response to sentiment-based messaging, with a mean score of 3.68. This was higher than the mean for rare consumers, which was 3.01, and substantially higher than the scores for frequent and daily consumers, at 2.40 and 1.75, respectively. The pattern suggests that sentiment-based messages may be most persuasive among consumers who drink soda occasionally, while their effectiveness declines as consumption becomes more frequent.

Table 1. Response to Sentiment-based messaging by consumption frequency

Response to Sentiment-based messaging by consumption frequency	**Mean ± SD**
Rarely (1–2 times per month)	3.01 ± 1.19
Occasionally (1–2 times per week)	3.68 ± 0.63
Frequently (3–6 times per week)	2.40 ± 1.52
Daily	1.75 ± 0.35

Taken together, these findings suggest that Hypothesis 1 is partially supported: sentiment based marketing language does appear to have a positive impact on brand attachment, but the strength of that impact depends on how frequently the consumers already engage with the brand.

This pattern may reflect the fact that daily consumers already have an established relationship with the brand, so emotionally framed messages have less room to shift their perception, whereas occasional consumers may be more inclined towards that change as their preferences are less fixed. This finding confirms Cambell and Keller's study, where it found that consumers rely more heavily on existing knowledge when evaluating familiar brands. Consequently, attitudes generated by a particular advertisement have less influence on evaluations of familiar brands than on unfamiliar ones (2003).

Hypothesis 2

Hypothesis 2 proposed that sentiment-based marketing language would have a stronger effect on consumer loyalty than purely objective, fact based information. Participants were asked to rate their likelihood of purchasing Coca-Cola based on factual information, such as "Coca-Cola has refreshed the world for over 130 years" and "Coca-Cola offers multiple zero-sugar options for health-conscious consumers." In the meantime, participants were also presented with sentiment-based messages, such as "Together, we share more than a drink. We share moments" and "the spark that lights up moments." Contrary to the hypothesis' prediction, participants generally reported higher mean scores for the objective information questions than the sentiment based question with means of objective based information ranging from 2.9 for frequent consumers to 4.03 for occasional consumers with means of sentiment based information ranging between 2.54 for frequent consumers and 3.56 for occasional consumers (Figure 2). The findings indicate that factual information may carry more persuasive weight than emotionally framed language, at least when the two are compared directly.

The one exception to this pattern occurred amongst daily consumers, who rated the sentiment based messages slightly higher than the objective messages, with the mean of objective information being 1.75 and the mean of sentiment based messaging being around 1.85. This exception should be interpreted cautiously though, as both scores among daily consumers were low overall, suggesting that neither type of messaging strongly moves this group, rather

than indicating that sentiment based language was especially effective for daily consumers. Based on these results, Hypothesis 2 was not supported.

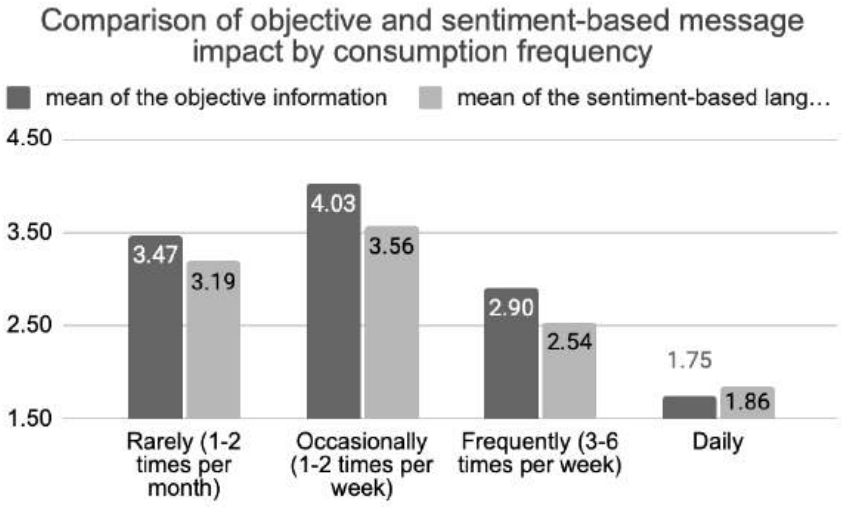


Figure. 2 Comparison of objective and sentiment-based message impact by consumption frequency

Hypothesis 3

Hypothesis 3 predicted that sentiment based marketing language would be applied more effectively towards certain groups of consumers than others. When broken down by age, the results showed a fairly clear pattern: younger participants (ages 14-18) had the lowest mean scores, reflecting participants' average responses on the five point likert scale, while the 25-34 age group had the highest mean at approximately 3.89, and the 35-50 age group also scored relatively high, as demonstrated in Table 2. The 25-34 result is worth examining more closely, however, since this age range is unusually broad in terms of life stage.

Table 2. Average effectiveness of sentiment based information by age

Age Group (years)	Mean ± SD
14–18	3.01 ± 1.14
18–24	3.14 ± —
25–34	3.89 ± 0.16
35–50	3.66 ± 1.15

Broadly speaking, older consumers above 25 years old appear more responsive to messaging built around brand values, emotional meaning, reliability, and quality, while younger

consumers between 14-18 years old may weigh more concrete factors like price and taste more heavily in their purchasing decisions.

When the data was broken down by gender instead, however, no clear or consistent pattern emerged between male and female participants, indicating that gender may not be as meaningful of a factor in how sentiment-based language is received, at least within this sample. As a result, Hypothesis 3 was supported with respect to age differences, but not supported with respect to gender differences.

Discussion and Conclusion

Overall, sentiment-based marketing language did strengthen brand attachment, but not evenly across all consumers. The clearest pattern was the gap between occasional and daily Coca-Cola drinkers: occasional consumers responded most strongly to sentiment-based messaging, while daily consumers showed the weakest response of any group. This pattern may reflect the fact that daily consumers already have an established relationship with the brand, so emotionally framed messages have less room to shift their perception, whereas occasional consumers may be more inclined toward that change as their preferences are less fixed. This finding confirms Campbell and Keller's study, which found that consumers rely more heavily on existing knowledge when evaluating familiar brands; consequently, attitudes generated by a particular advertisement have less influence on evaluations of familiar brands than on unfamiliar ones (301).

That same reliance on existing brand knowledge may help explain why, contrary to expectations, participants generally rated factual, objective information as more persuasive than sentiment-based language. One possible explanation for these results as a whole is that emotional influence operates more subtly than factual influence, and may be harder for participants to consciously identify and rate in a survey, even when it continues to shape their attitudes in less measurable ways. In other words, familiarity may make consumers less swayed by emotional appeals and more inclined to lean on concrete, verifiable claims when asked to consciously evaluate a message.

Age, however, produced a different pattern from consumption frequency. Sentiment-based language proved most effective among consumers aged 25–34 and 35–50, while the youngest participants (14-18) showed the weakest response, suggesting that older consumers respond more to messaging built around brand values, emotional meaning, reliability, and quality, while younger consumers may weigh more concrete factors like price and taste more heavily. Unlike the other age brackets, which tend to represent a more uniform stage of life, the 25–34 group likely includes a mix of people still early in their independence alongside those who are more established, such as parents or working professionals with more purchase experience. It's possible that this diversity in life experience is part of what drove the group's higher mean, since more established consumers within that range may respond more to sentiment-based cues tied to values, nostalgia, or identity, while less experienced consumers in the same bracket may lean more toward practical factors like price. If sentiment-based responses and objective-leaning

responses were both pulled upward by different subsets of this group, it could help explain why 25–34 outperformed even the 35–50 bracket. No similarly meaningful pattern emerged by gender, suggesting that age may be a more influential demographic factor than gender in shaping responses to sentiment-based marketing language.

These findings should be interpreted in light of several limitations. The sample size was relatively small and drawn largely from the researcher's local school community and family networks in Rocklin, California, which limits the geographic and demographic diversity of the results. The study's exclusive focus on Coca-Cola also means the findings may reflect something specific to that brand's marketing rather than sentiment-based language in general. Despite these constraints, this study offers useful exploratory evidence that sentiment-based marketing language can meaningfully shape brand attachment, and that its effectiveness depends on factors such as how often a consumer already engages with a brand and their age group. Future research could build on these findings by surveying larger and more geographically diverse samples, examining multiple brands rather than a single case study, incorporating interviews or longer surveys to capture more nuanced reactions, conducting research to track attachment over time, or testing consumers' reactions to entirely new advertisements rather than real, already-familiar ones.

Work Cited

Campbell, Margaret C., and Kevin Lane Keller. "Brand Familiarity and Advertising Repetition Effects." Journal of Consumer Research, vol. 30, no. 2, Sept. 2003, pp. 292–304. ResearchGate, https://www.researchgate.net/publication/24099213_Brand_Familiarity_and_Advertising_Repetition_Effects.

Cruz, Ryan E., James M. Leonhardt, and Todd Pezzuti. "Second Person Pronouns Enhance Consumer Involvement and Brand Attitude." Journal of Interactive Marketing, vol. 39, 2017, pp. 104–116. https://repositorio.uchile.cl/bitstream/handle/2250/149711/Second-Person-Pronouns.pdf?sequence=1&isAllowed=y.

Escalas, Jennifer Edson. "Narrative Processing: Building Consumer Connections to Brands." Journal of Consumer Psychology, vol. 14, no. 1&2, 2004, pp. 168–180. https://sharingiscaringsouls.wordpress.com/wp-content/uploads/2018/04/09-05.pdf.

Magids, Scott, Alan Zorfas, and Daniel Leemon. "The New Science of Customer Emotions." Harvard Business Review, vol. 93, no. 11, Nov. 2015, pp. 66–74, 76. https://inspirefire.com/wp-content/uploads/2019/08/HBR_The-New-Science-of-Customer-Emotions.pdf.

Park, C. Whan, et al. "Brand Attachment and Brand Attitude Strength: Conceptual and Empirical Differentiation of Two Critical Brand Equity Drivers." Journal of Marketing, vol. 74, no. 6, Nov. 2010, pp. 1–17. https://www.jstor.org/stable/25764280?googleloggedin=true&seq=1.

Priced for Perfection: Retail Investor Psychology and the Valuation of Unprofitable AI Companies By Ayaan Jindal

Abstract

OpenAI and Anthropic have confidentially filed for potential initial public offerings, and both could enter the public market at valuations approaching $1 trillion. Though neither OpenAI nor Anthropic consistently makes a profit, and OpenAI is expected to book a massive $14 billion loss in 2026, many people are still extremely eager to buy their stocks. This paper will use behavioral economics (the study of financial behavior that is not always fully rational) to examine four factors that help explain the excessive hype surrounding the two companies: loss aversion, herd behavior, FOMO (fear of missing out), and narrative economics, which explains why people sometimes invest based on stories rather than hard facts. I will also explore the role of apps like Robinhood, which have made it extremely easy for retail investors to enter the market. My main claim is that the excessive hype surrounding these two AI companies stems largely from psychological and social factors, rather than from their financial numbers. Yet this does not automatically mean that their stocks are a bad investment. Rather, it means that many people may be buying for reasons that have little to do with the fundamental value of the two companies.

Introduction

OpenAI and Anthropic filed paperwork to go public in the summer of 2026. In May 2026, Anthropic reached a private valuation of $965 billion, while OpenAI's valuation was reported at around $852 billion, with some reports suggesting it could seek a valuation close to $1 trillion (Forbes; SmartAsset). But while both companies are valued at hundreds of billions of dollars, neither one consistently makes a profit. OpenAI, in particular, is projected to lose around $14 billion in 2026, and it may not become profitable until 2029 or 2030 (Barchart). For its part, Anthropic has reported one profitable quarter, but it warned that the result may not be sustainable once temporary discounts on computing power expire (SmartAsset).

The basic question this paper investigates is why investors are so willing to buy shares in loss-making companies such as OpenAI and Anthropic, sometimes before the stocks are even publicly available (Barchart). This is not to say that the companies have no value. In fact, both are valued above many established public companies. However, the value of a private company depends not only on its current financial performance, but also on expected future growth and other factors that investors believe may affect its value. The gap between the financial performance of these companies and the prices investors may be willing to pay is not simply a matter of math. It can also be explained through behavioral economics, which studies how people actually behave when making financial decisions. In addition to examining this gap, this paper also investigates the role that phone-based investment apps have recently played in making it easier for retail investors to act on their excitement.

The Psychology Behind Regular People Investing

For decades, the old-school way of doing finance has operated under the assumption that people are rational, that they weigh the risks and rewards of any given action, and then make a smart and informed decision. However, according to behavioral economics, real people do not always operate in this way. Rather, we are often driven by emotions and guided by mental shortcuts that help us make decisions quickly. For those who invest in the stock market, these tendencies can lead to predictable mistakes. And for those looking to invest in AI stocks, a few of these mistakes are particularly important.

Loss aversion means that losing money tends to hurt more than gaining the same amount feels good. This can cause people to behave differently. On one hand, the fear of losing can prompt individuals to avoid reasonable risks. On the other hand, it can push them into riskier investments because they fear missing out and regretting it later. A 2022 study, "Herding and Loss Aversion in Stock Markets," found that retail investors' decisions were strongly affected by loss aversion and herd behavior. The study also found that FOMO intensified the effects of both biases ("Herding and Loss Aversion in Stock Markets"). Herd behavior, or following the crowd, describes the tendency of people to copy what others are doing, even in financial markets. Humans have long depended on groups for safety, so people may be naturally inclined to follow the crowd even when making investment decisions ("The Sheep in the Stock Market").

FOMO further intensifies loss aversion, herd behavior, and other behavioral biases, especially in the era of social media, where investors are constantly exposed to the gains of others in real time. A 2024 review of the effects of FOMO on investment behavior found that social media amplifies FOMO by displaying other people's wins, which can encourage overtrading, gambling on stocks, and the creation of bubbles, especially in meme stocks, stocks that surge mainly due to online hype and coordinated retail buying rather than company performance ("The Effects of FOMO on Investment Behavior in the Stock Market"). Researchers at UC Riverside also cited Bed Bath & Beyond's 2021 stock surge as an example of a stock rising because of hype and peer pressure rather than because the company was performing well ("Expert Insights on Herd Instinct").

Finally, there is the issue of social comparison, which can increase risk-taking. In one experiment involving 807 experienced investors, those who were told that their friends were earning higher returns began taking greater risks and trading more often. They also felt worse about their own performance, no matter how well they were actually doing ("The Influence of Upward Social Comparison on Retail Trading Behaviour"). This matters for AI stocks because financial media outlets regularly publicize such outcomes. Business Insider, for example, interviewed early Nvidia retail investors whose gains funded cars, vacations, and a down payment on a second home, with one investor's total $62,000 stake growing to roughly $3 million (Altchek and Sor). Coverage like this can make ordinary investors feel as though they are falling behind, even when their own results are reasonable.

Stories Sell Stocks: Narrative Economics

There is also a larger point made by economist Robert Shiller. The economy is shaped in part by stories that spread from person to person. These "viral" stories can powerfully change people's perceptions of money and business opportunities, often without being supported by facts. Shiller calls this "narrative economics." This can be seen when investors continue to pour money into companies that are currently losing money because the story surrounding the company seems strong enough to justify the risk (Shiller).

These narratives do not appear on their own. They are produced through a combination of company messaging, media coverage, and investor commentary. Executives shape the story directly through product launches, public appearances, and investor presentations, while journalists and analysts amplify and reinforce it. In the case of AI, company leaders have repeatedly framed their technology as a fundamental shift in how the economy will operate, and financial media have echoed that framing. This means the AI narrative is partly organic, driven by genuine public interest in tools people already use, and partly manufactured by the companies and institutions that benefit from investor enthusiasm.

The same patterns appeared during the dot-com bubble in the late 1990s. Netscape went public in 1995 even though the company was not yet profitable, helping begin a trend in which companies were valued more for growth and hype than for current earnings ("AI Hype vs. Dot-Com Bubble"). In 1999, 62 percent of US venture capital money went to unprofitable companies. By 2000, the number of IPOs on the NASDAQ had more than doubled before the market collapsed, wiping out huge amounts of value ("AI Hype vs. Dot-Com Bubble"). A similar pattern appeared in 2020 and 2021, when 70 to 80 percent of companies that went public were not profitable, compared with only about 20 percent in 2009 (Nasdaq). People often respond to these warning signs by saying that "this time is different" (Nasdaq).

The story of AI presents it as a once-in-a-lifetime shift in the economy, similar to the rise of the internet in the 1990s. That story is so strong that investors may be willing to accept prices that do not closely match a company's current earnings or cash flow. One economist said in 2026 that valuations were high, retail investors were piling in, and the market felt increasingly frothy, although he did not believe that all the conditions for a full bubble had yet appeared (Lichtenberg).

Where the AI Narrative Comes From

Not every AI narrative is created equal. The stories told about the use of AI can vary depending on who tells them. Researchers Ozturkcan and Bozdağ describe "AI washing" as the "deliberate or negligent exaggeration of a system's artificial intelligence capabilities." When a company touts the use of AI in marketing their product, sometimes those claims can be exaggerated and even mislead consumers. In the past, companies such as Coca-Cola have been accused of using vague AI-related claims to market their products, such as their Y3000 drink (Ozturkcan and Bozdağ). Even in the realm of investing, there have been instances of "AI washing," in which companies make exaggerated claims about the use of AI in their investment

strategies, as the SEC recently charged two firms in the space (Ozturkcan and Bozdağ). OpenAI and Anthropic, however, are a different case. Not only do the companies have working products used by hundreds of millions of people around the world, but they also generate real revenue, even if it is not yet sufficient to meet their high valuations. This means OpenAI and Anthropic are not making claims about the current capabilities of their products; instead, the focus is on future capabilities, and those claims are not always presented with the necessary uncertainty.

There is already a lot of hype around superintelligence, and AI in general, and Sam Altman's own words show why. At the India AI Impact Summit in February 2026, Altman said that on OpenAI's current trajectory, early superintelligence could arrive within a couple of years, and that by 2028, more of the world's thinking capacity could reside inside data centers than in human brains ("Altman Sees Superintelligence Arriving in a Couple of Years"). He did add that AI development has already produced plenty of surprises and will likely produce more, but that qualifier came after the headline prediction, not before it, and it was far less specific than the timeline itself.

Financial coverage of these claims tends to amplify them rather than test them, echoing the FOMO effect described earlier in this paper: a CEO's superintelligence timeline is reported the same way an earnings beat would be, as fact rather than forecast. A 2026 study of U.S. firms tracked which companies' AI disclosures were backed by actual patents and technology spending and which were largely promises, then measured how each group performed afterward. The study found that the "promising" disclosures initially attracted investor interest, but this led to negative returns that remained weak as the gap between the AI-related promises and actual results continued to widen (Song et al.). This matches Shiller's narrative economics: financial markets reward a good AI story in the short run but eventually punish claims that outpace results. OpenAI's and Anthropic's predictions of near-term superintelligence and their trillion-dollar valuations remain, for the most part, unproven.

Dan Kotliar, a sociologist who studies AI hype, suggests three questions for figuring out whether an AI-related narrative is about developments already taking place or about something expected to happen down the road: who is behind the narrative, what do they stand to gain from it, and is the claim really about the technology's current abilities or its imagined future ones? Applying these questions to OpenAI and Anthropic yields mixed answers. Both companies do have real technology and real revenue, so the excitement is not built on nothing. At the same time, their valuations still seem to rest on more than just current performance, since both companies would need growth far beyond anything they have shown so far to justify the prices investors are paying. The boldest predictions about that future growth, in turn, come from the companies' own executives, the people who stand to gain the most if investors believe them. Whether these executives genuinely believe their own predictions or are knowingly overselling them is impossible to know for certain, and it matters either way ethically. But for a retail investor deciding whether to buy shares, that distinction matters less than it might seem. Shiller's narrative economics does not require anyone to lie for a story to move prices; it only needs to be compelling and repeated often enough. In that sense, the AI narrative is justified to the extent

that the underlying technology is real, but results do not yet justify it. The gap between the two is exactly what a "priced for perfection" valuation is betting will close.

The 2026 AI IPO Situation

Here is a sense of the magnitude of what is being attempted: OpenAI, Anthropic, and SpaceX are aiming for a combined valuation of around $3.8 trillion, which is about 10 percent of the entire US stock market (Intellectia). To justify those valuations at more typical market levels, the three companies would need to earn about $150 billion in annual profit. Instead, they are currently losing about $20 billion per year combined (Intellectia).

Figure 1 compares the private valuations of OpenAI and Anthropic with the amount of revenue each company generates per year. The gap between their current revenue and projected valuations is extremely large. This helps explain why these companies are often described as being “priced for perfection,” since they would need to achieve very strong future growth for their valuations to make sense to investors (Klover.ai).

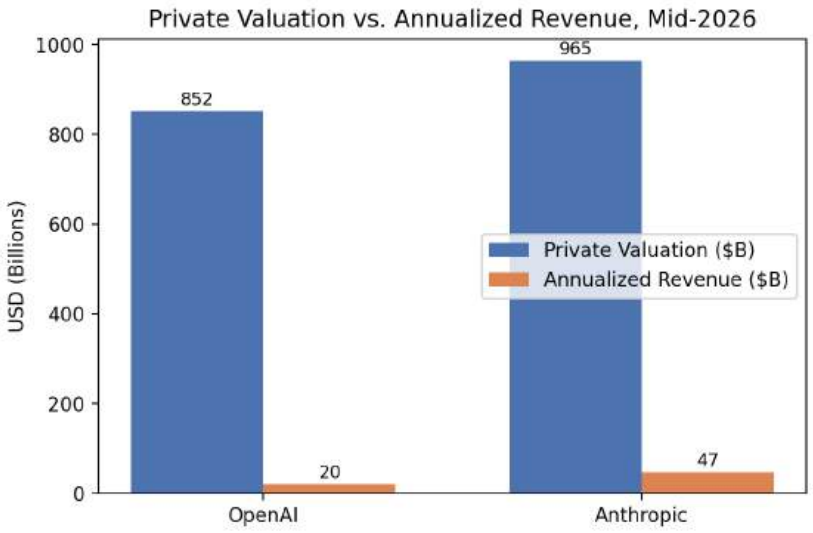


Fig. 1: Private value versus yearly revenue for OpenAI and Anthropic, mid-2026. Numbers from Forbes and SmartAsset.

For our purposes, it is more helpful to look at the first of the three companies to go public: SpaceX. When the space exploration company began trading on June 12, 2026, it priced its shares at $135 each, giving it an initial valuation of about $1.77 trillion. On the first day of trading, the stock rose by about 19 percent and closed at $160.95, giving the company a market value above $2 trillion (SmartAsset; IndMoney). The share price later climbed to a high of $225.64 before falling by roughly 36 percent. By July 10, 2026, it had dropped to about $145.30 (Investing.com). This raises the question of whether OpenAI and Anthropic, both priced for perfection on similarly unproven growth stories, could experience a similar decline once their shares begin trading (Barchart).

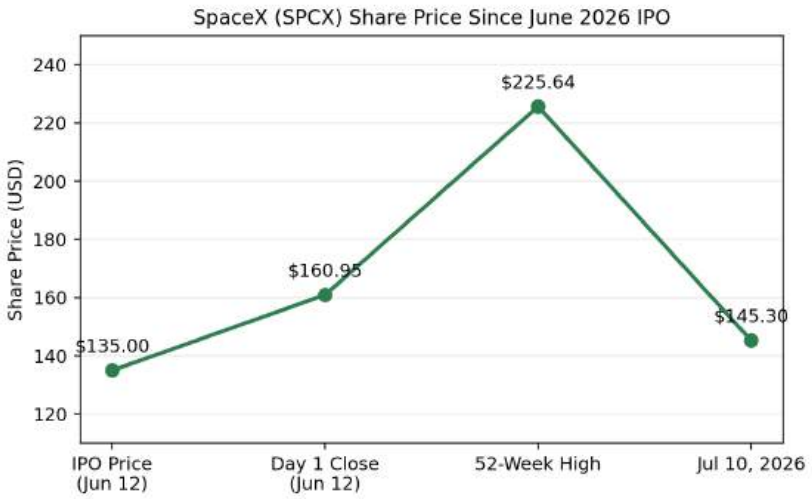


Fig. 2: SpaceX (SPCX) share price at key points following its June 2026 IPO, through July 10, 2026. Data from Investing.com.

How Investment Apps Lowered the Barrier to Entry

It used to be that IPO shares at the initial price were mostly reserved for large institutions and wealthy clients. Regular investors could usually buy only after the stock had already begun trading, often at a higher price (NerdWallet). That changed with SpaceX's June 2026 IPO. Retail investors could request shares through platforms such as Robinhood, Fidelity, Charles Schwab, SoFi, and E*TRADE, and Robinhood did not require a minimum account balance (CNBC; Robinhood; Motley Fool).

Robinhood's CEO said that retail investors now have a real seat at the table in IPOs. In some recent deals, retail investors received 20 to 30 percent of the available shares, which is very different from earlier IPOs, when they received almost nothing (Yahoo Finance). SpaceX therefore gave ordinary investors access to IPO shares at the same time, and often at the same price, as larger investors. OpenAI and Anthropic had not yet confirmed whether they would use the same approach, but SpaceX established a model they could follow (NerdWallet).

FOMO and herd behavior can affect stock prices only when people can act on those emotions quickly. In the past, buying into a highly anticipated IPO required significant money or connections, so the reactions of ordinary investors had less influence on the market. Now, investors can request shares through an app in a very short amount of time, sometimes without any minimum balance. This gives emotions such as FOMO and herd behavior a much more direct path to affecting demand and price.

Counterargument: Is This Behavior Actually Irrational?

It is worth acknowledging another perspective: AI prices may be high without necessarily forming another bubble. One economist uses a bubble checklist called the "Four Horsemen": overpricing, widespread belief in the hype, large amounts of new stock being issued, and heavy speculative money flows. He argues that three of the four are clearly present in 2026, but one is

missing because companies are buying back their own shares instead of flooding the market with new ones. He sees this as a sign that the people running these companies may not believe their shares are overpriced (Lichtenberg).

Some analysts have made a similar argument about Anthropic. They describe its valuation as aggressive but note that its revenue-to-value ratio is somewhat similar to that of other fast-growing software companies. At the same time, the valuation leaves very little room for mistakes (Klover.ai). In other words, the price may be high, but it could still reflect a risky and genuine bet on technology that investors believe will change everything. If that belief proves correct, the investment could perform well even if it looks extreme right now.

I do not have a definite answer about whether AI company valuations are reasonable. Until these companies trade publicly for several years, it will be impossible to know for sure. What the research does show, however, is that even if the bet proves correct, many retail investors may still be buying for the same reasons that led people into past bubbles: FOMO, herd behavior, social comparison, and belief in a powerful story. It is therefore possible for a company to be a strong investment while many individual investors are still buying it for weak or irrational reasons. Both things can be true.

Applying These Patterns to OpenAI and Anthropic

You have probably heard of FOMO, or the fear of missing out, before. In the case of OpenAI and Anthropic, it may be even stronger than it was in some earlier bubbles because many people already know and use their products. ChatGPT alone has more than 900 million weekly users (Barchart). That makes the fear of missing out more powerful because investors are not betting on a completely unfamiliar idea. They are betting on a product they already use and understand.

Herd behavior also appears in the order in which these companies go public. Analysts have suggested that whichever company lists first may capture a large share of investor attention simply because it is first, regardless of whether it has the strongest fundamentals (Klover.ai). In that situation, investors may follow the crowd toward the first major AI IPO instead of carefully comparing the companies.

Loss aversion helps explain another part of the demand. Many people feel that they missed out on the early gains of companies such as Nvidia or Amazon and do not want to miss the next major opportunity. In the past, ordinary investors often could not access IPO shares at the initial price. Now that retail investors may be able to buy through investment apps as soon as shares are offered, the fear of missing out on another major success can turn into a purchase almost immediately (Klover.ai).

This also fits Shiller's idea that narratives can drive stock prices. News coverage often presents these IPOs as part of a major historical shift rather than simply as shares in two individual companies. As a result, investors may judge them based more on future possibilities than on their current financial performance (Bloomberg).

Conclusion

People investing in unprofitable AI companies such as OpenAI and Anthropic are following a pattern that has appeared with other highly anticipated technologies for decades. What is different now is how easy it has become to act on that excitement through phone-based investment apps. Loss aversion, herd behavior, FOMO, and social comparison all help explain why investors may chase these companies even as they lose money. Narrative economics also helps explain why the larger story surrounding AI can be powerful enough to make people overlook traditional warning signs.

This does not mean that OpenAI or Anthropic are necessarily bad investments. Both companies are growing quickly, and their future earnings may eventually justify their current valuations (Klover.ai). However, much of the current demand appears to be driven by the AI narrative and by the psychological and social pressures discussed throughout this paper. For that reason, the excitement surrounding these companies should be treated separately from their actual financial value. Hype alone should not be taken as proof that a company is worth its price.

SpaceX offers an early example of this pattern. Its stock rose sharply after going public, then fell by roughly 36 percent from its peak. It remains to be seen whether OpenAI and Anthropic will experience a similar rise and decline, or whether demand for their shares will be supported by stronger financial results. We will not know for sure until they go public and trade for some time. Still, because it is now easier than ever for retail investors to participate, the behavior of these stocks will be especially important to watch.

Works Cited

"AI Hype vs. Dot-Com Bubble: A 25-Year Market Lens." CKGSB Knowledge, Cheung Kong Graduate School of Business, 2026, english.ckgsb.edu.cn/knowledge/professor_analysis/dot-com-to-deepseek-25-year-tech-bubble-comparison-for-ai-er/.

Altchek, Ana, and Jennifer Sor. "They Bought Nvidia Stock Before the Hype. It's Paid for Cars, Vacations, and a Deposit on a Dream Home." Business Insider, 24 Apr. 2024, markets.businessinsider.com/news/stocks/nvidia-stock-retail-investors-gains-generational-wealth-ai-chips-gpus-2024-4.

"Altman Sees Superintelligence Arriving in a Couple of Years." MIT Sloan Management Review Middle East, 19 Feb. 2026, https://www.mitsloanme.com/article/altman-sees-superintelligence-arriving-in-a-couple-of-years/

"Best Brokers for IPO Access." NerdWallet, 2026, www.nerdwallet.com/investing/best/brokers-for-ipo-access.

"The Effects of FOMO on Investment Behavior in the Stock Market." ResearchGate, Oct. 2024, www.researchgate.net/publication/385392144_The_Effects_of_FOMO_on_Investment_Behavior_in_the_Stock_Market.

"Expert Insights on Herd Instinct." UCR School of Business, University of California, Riverside, 1 Aug. 2024, business.ucr.edu/news/2024/08/01/expert-insights-herd-instinct.

"Herding and Loss Aversion in Stock Markets: Mediating Role of Fear of Missing Out (FOMO) in Retail Investors." International Journal of Emerging Markets, Emerald Publishing, 2022, www.emerald.com/ijoem/article-abstract/17/7/1720/146285.

"How Can Retail Investors Take Part in the SpaceX IPO?" The Motley Fool, 28 May 2026, www.fool.com/investing/2026/05/28/how-can-retail-investors-invest-in-spacex-ipo/.

"How the IPO Bubble Burst, and Where Do We Go from Here?" Nasdaq, 17 Sept. 2022, www.nasdaq.com/articles/how-the-ipo-bubble-burst-and-where-do-we-go-from-here.

"The Influence of Upward Social Comparison on Retail Trading Behaviour." PMC, National Center for Biotechnology Information, www.ncbi.nlm.nih.gov/pmc/articles/PMC10733348/.

"IPO AI Market: Anthropic vs. OpenAI [Analysis] [2026]." Klover.ai, 3 June 2026, www.klover.ai/ipo_ai_market_anthropic_vs_openai_analysis_2026/.

Kotliar, Dan M. "Can't Stop the Hype: Scrutinizing AI's Realities." Information, Communication & Society, vol. 29, no. 3, 2025, pp. 828-849, doi.org/10.1080/1369118X.2025.2531165.

Lichtenberg, Nick. "'We're Not in a Bubble Yet' Because Only 3 Out of 4 Conditions Are Met, Top Economist Says. Cue the OpenAI IPO." Fortune, 1 Feb. 2026, fortune.com/2026/02/01/is-ai-a-bubble-top-economist-says-no-because-ipos-fraud/.

"OpenAI & Anthropic IPO: What's Confirmed, What's Still Speculation." Barchart, 2026, www.barchart.com/story/news/3116486/openai-anthropic-ipo-what-s-confirmed-what-s-still-speculation.

"OpenAI Stock IPO: Valuation, Timeline and Investment Options." SmartAsset, 2026, smartasset.com/investing/openai-stock-ipo.
"OpenAI Vs Anthropic IPO: How They Compare And What We Know." Forbes, 2026, www.forbes.com/sites/investor-hub/article/openai-vs-anthropic-ipo-comparison/.
Ozturkcan, Selcen, and Ayşe Aslı Bozdağ. "Responsible AI in Marketing: AI Booing and AI Washing Cycle of AI Mistrust." International Journal of Market Research, vol. 67, no. 6, 2025, journals.sagepub.com/doi/10.1177/14707853251379285.
"Retail Investors Get Direct Access to SpaceX IPO Through Major Brokerage Platforms." CNBC, 21 May 2026, www.cnbc.com/2026/05/21/retail-investors-get-direct-access-to-spacex-ipo-through-major-brokerage-platforms.html.
Robinhood. "Robinhood Is Democratizing IPOs." Robinhood, 20 May 2021, www.robinhood.com/us/en/newsroom/robinhood-is-democratizing-ipos/.
"Robinhood Is Opening the SpaceX IPO to Everyday Investors. What That Means for the Stock." Yahoo Finance, 5 June 2026, finance.yahoo.com/markets/stocks/articles/robinhood-opening-spacex-ipo-everyday-041100617.html.
"The Sheep in the Stock Market: How Herd Behavior Shapes Investment Trends." The Decision Lab, thedecisionlab.com/insights/finance/the-sheep-in-the-stock-market.
Shiller, Robert J. Narrative Economics: How Stories Go Viral and Drive Major Economic Events. Princeton University Press, 2019.
Song, Xiapeng, et al. "AI Washing: Strategic Disclosure and Backlash." Finance Research Letters, vol. 95, 2026, www.sciencedirect.com/science/article/pii/S1544612326002151.
"SpaceX Stock: IPO Date, Share Price & News." Investing.com, 10 July 2026, www.investing.com/equities/spacex.
"SpaceX, OpenAI & Anthropic IPOs 2026 | Dates, Valuations, Risks & Market Impact." IndMoney, 2026, www.indmoney.com/blog/us-stocks/spacex-openai-anthropic-ipo-explained.
"The $3.8 Trillion AI Mega IPO Wave of 2026: SpaceX, OpenAI & Anthropic Investment Guide." Intellectia, 2026, intellectia.ai/blog/ai-mega-ipo-wave-2026-spacex-openai-anthropic.
"Why OpenAI and Anthropic IPOs May Be Dangerous for Retail Buyers." Bloomberg Businessweek, 19 May 2026, www.bloomberg.com/news/articles/2026-05-19/why-openai-and-anthropic-ipos-may-be-dangerous-for-retail-buyers.

Domestic and Global Drivers of the Brazilian Stock Market: The Role of Interest Rates, Political Stability, and International Market Conditions By Pietro Ribas

Abstract

This paper examines how interest rates, politics, and the global markets influence the performance of the Brazilian stock market, which is represented by the Bovespa index. Because Brazil is classified as an emerging market, it is dependent on foreign capital investment and has an economy concentrated in the commodity sector, it is very reactive to shifts in those three forces. This paper uses a quantitative analysis of market and economic data and a qualitative evaluation of political and historical events to examine the relationship between the three factors. Furthermore, it analyzes how the demand for commodities and the US Federal Funds Rate impact capital flows in and out of Brazil and how political events like the 2015/2016 economic crisis and the 2018 presidential election were both events that influenced investor confidence in the market and increased volatility. The paper concludes that interest rates, politics, and the global markets rarely act alone and often are interconnected and reinforce one another, making it crucial to look at all factors when understanding Ibovespa.

1. Introduction

The Brazilian stock market is represented by the Bovespa Index, or Ibovespa, on the B3 exchange and was the largest stock exchange in Latin America and the twentieth-largest in the world as of December 2024, with a market value of over $783 billion (Baker McKenzie, 2026). As Brazil is classified as an emerging market, the Brazilian financial system is heavily affected by both domestic policies and conditions brought by foreign countries; these factors distinguish it from more developed markets. In the market, investors often shift their assets based on their confidence and sentiment about political decisions, changes in the interest rate, and the global market, since those forces have a significant impact on capital flows, companies' earnings, and overall market performance. This raises the question that this paper examines: How do interest rates, politics, and global markets influence the Brazilian stock market? This paper argues that the Brazilian stock market is significantly influenced by domestic interest rates, political decisions, and global market conditions, as these factors shape investor confidence, capital flows, and overall economic stability.

2. Objectives of the Study

- Analyze key factors affecting the Brazilian stock market.
- Examine how interest rates influence investor behavior and stock prices.
- Evaluate how politics and government decisions affect market confidence.
- Assess the impact of global markets, capital flows, and commodity prices.
- Explain how these factors work together to shape Ibovespa performance.
- Show why Brazil's emerging market status makes it more sensitive to volatility.

3. Research Methodology

The research includes a mixture of quantitative and qualitative analysis to study the relationship between macroeconomic conditions and the Brazilian stock market. Qualitative data were used to study political events, investor confidence, and sentiment, which were used to interpret and analyze the quantitative numbers. Quantitative evidence includes interest rate histories and index performance to measure economic indicators over time. By using both approaches, this research could relate the relationship between market patterns and the political and economic outcomes, resulting in a more complete conclusion than either approach could do alone.

4. Overview of the Brazilian Stock Market

Located in São Paulo, B3 is Brazil's official stock exchange. It is the market in which financial assets such as stocks, real estate funds, and exchange-traded funds (ETFs) are bought and sold. It was created in 2017 through the merger of BM&FBOVESPA and CETIP, resulting in the company B3, which stands for Brasil, Bolsa, Balcão. Furthermore, B3 is responsible for regulating and managing the financial market, ensuring that the exchange and sale of various assets are safe, transparent, and controlled. Since its creation in 1968, Brazil's stock market has been tracked by the Bovespa Index, or Ibovespa, the main stock performance indicator (B3, n.d.). Like the S&P 500 in the USA, the Ibovespa is the standard indicator and reference for Brazilian equities for investors or analysts.

The Ibovespa is a total-return index that is market-capitalization-weighted and free-float-adjusted, meaning that companies with larger market capitalizations move the index more than those with smaller market capitalizations (B3, n.d.). Only the shares of the companies that are available for trade are included, and this measures the total gains, the stock price appreciation, plus the dividends, and not only the movement of the stock prices. These three components are important to understanding the index overall. Larger companies with more shares available are more relevant to how the index moves; for example, a 3% drop in Petrobras is more significant to the index than a 3% drop in a small-cap company. For a stock to be eligible for the index, several components are considered, such as price (no penny stocks) and Average Daily Trading Volume (ADTV). It has to be available for sale on 95% of the days B3 operates, excluding stocks that are rarely traded and go without buyers or sellers for some time. The trading volume must be worth at least 0.1% of the total money being exchanged in the market in a given time. This is another filter aspect, meaning that it is not enough for the assets to be sold regularly, but it must also matter to the market in terms of how much money is traded (B3, n.d.). The Ibovespa index is reviewed every four months so that all stocks qualify for the criteria and that it has the most traded companies in the stock market. The stocks that are part of the index represent about 80% of B3's financial trading volume (B3, n.d.) and are heavily concentrated in the financial, energy, and commodities sectors. Consequently, the index is mainly focused on three dominant sectors of the index, so if they gain or lose money, it determines the market movement. Because the index is concentrated in these three sectors, it is more affected by

interest rates, global trade, and political changes than a more balanced index, making the Bovespa index sensitive to shifts in these forces.

Foreign capital dependence, institutions that are still developing, a higher equity risk premium, and high market volatility are connected characteristics that reinforce each other, maintaining Brazil's status as an emerging market and leaving it more structurally fragile. Those vulnerabilities are what make the Brazilian market index more sensitive to political shifts, interest rates, and the global market than other more developed market indexes. In 2024, Brazil's GDP had a 3.4% increase, and in the 2022/2023 period, the growth exceeded the IMF prediction by over 2 percentage points, showing investors that Brazil is a country with investment potential. However, with a real GDP per capita of $10,616, it was still a lot below more developed economies. It showed how Brazil was still structurally dependent on foreign capital, is in a development stage, and does not have a fully mature market like countries such as the US and Germany, which characterizes it as an emerging market. This combination of having the potential to grow but still being structurally immature is what makes Brazil an emerging market. This explains how foreign investment, political decisions, and global markets have a larger impact on the Bovespa index than on indexes of more developed economies.

The consequences of the events of Dilma's impeachment and March 2020 illustrate how the Bovespa Index is significantly affected by political decisions and global market conditions, as these factors influence investor sentiment and confidence in the government and global capital invested. In 2016, during Dilma Rousseff's impeachment, the Ibovespa was very volatile and sensitive. Most investors did not have confidence in her economic decisions, but her impeachment made them hope and expect a government with choices that were more pro-market. Consequently, they started reinvesting more money into the market, causing the Ibovespa to shift up even though the overall economy of Brazil was in recession (Branco, 2016). The opposite result occurred in March 2020, when COVID-19 spread globally, and the World Health Organization (WHO) declared a pandemic. As a result, both international and Brazilian investors moved their money into safer assets. When investors took their money out of global capital markets, the index dropped by about 30%, taking out R$1.1 trillion from Brazilian companies in just one month (*G1*, 2020). Both examples indicate that the short-term volatility of the index reflects political decisions and global market stability, as they impact investor confidence and capital investment.

5. Impact of Interest Rates on the Brazilian Stock Market

In emerging market economies like Brazil, changes in domestic interest rates directly influence not only the country's overall financial state but also stock market stability. The Selic rate is a base interest rate that is set by the Monetary Policy Committee (Copom) of the Central Bank of Brazil. It is used to hold down inflation when there is a rapid increase in prices and to stimulate the economy when it is not growing. When the Selic rate is high, investors tend to reduce their equity holdings because they can get a similar or higher return with lower risk, like the Tesouro Selic. Also, when the rate is up, people choose to move their money to safer assets

that are considered to be risk-free, instead of putting their money in stocks, where there is a risk premium. When the Selic rate is high, it is more expensive for companies to borrow money and capital, which forces Brazilian businesses to hold back on their investments and ultimately limits the growth of the stock market. A 2024 ARDL study of Brazilian stock market factors from 2010 to 2024 discovered that the Selic rate has a significant negative long-term influence on the Ibovespa (Vieira & Ferrando, 2024). As shown in Figure 1, this opposite relationship is visible on most shifts for almost two decades, especially during 2015 and 2016, when the Selic rate increased to 14.25% while the Ibovespa was falling, and in 2020 and 2021, when the Selic rate decreased to 2% while the market rose substantially. This pattern visually evidences the ARDL findings, as it represents these economic cycles and indicates that this relationship between Ibovespa and the Selic rate did not happen only once but is recurring in the Brazilian market. As a result, these shifts in the Selic rate are responsible for dictating the country's financial state and also for market volatility, significantly affecting investor sentiment. Conversely, when the rate is lowered, firms and investors borrow more and invest more money in the stock market, contributing to a rise in the Ibovespa. Furthermore, this effect does not impact the stock market equally because market volatility is different in every sector. Sectors that depend on borrowing money, such as construction and real estate, will suffer more when there is an increase in interest rates because the demand will decrease. During this same 2014–2016 period, the construction sector was hit disproportionately hard. The construction company PDG Realty, which was one of the largest firms in the industry with a market capitalization in 2010 of R$14.4 billion, filed for judicial recovery in 2017 with a debt of almost R$7.9 billion and a valuation of R$158 million. Similar events happened in the following years, with other large construction companies like Viver and Rossi Residencial. Ultimately, these changes in the Selic rate have significant outcomes in the country's economy and in the volatility of the market, also significantly impacting investor confidence.

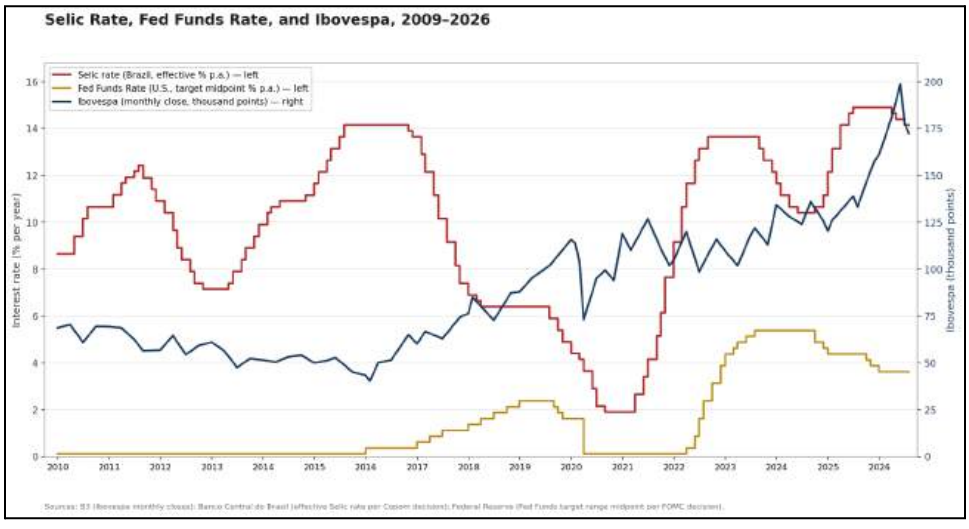


Figure 1

As shown in Figure 1, this opposite relationship is visible in most shifts over almost two decades, especially during 2015 and 2016, when the Selic rate increased to 14.25% while the Ibovespa was falling, and in 2020 and 2021, when the Selic rate decreased to 2% while the market rose substantially. This pattern visually evidences the ARDL findings, as it represents these economic cycles and indicates that this relationship between Ibovespa and the Selic rate did not happen only once but is recurring in the Brazilian market. As a result, these shifts in the Selic rate are responsible for dictating the country's financial state and for market volatility, significantly affecting investor sentiment. Conversely, when the rate is lowered, firms and investors borrow more and invest more money in the stock market, contributing to a rise in the Ibovespa. Furthermore, this effect does not impact the stock market equally because market volatility is different in every sector. Sectors that depend on borrowing money, such as construction and real estate, will suffer more when there is an increase in interest rates because the demand will decrease. During this same 2014–2016 period, the construction sector was hit disproportionately hard. The construction company PDG Realty, which was one of the largest firms in the industry with a market capitalization in 2010 of R$14.4 billion, filed for judicial recovery in 2017 with a debt of almost R$7.9 billion and a valuation of R$158 million. Similar events happened in the following years, with other large construction companies like Viver and Rossi Residencial. Ultimately, these changes in the Selic rate have significant outcomes in the country's economy and in the volatility of the market, also significantly impacting investor confidence.

6. Influence of Politics on the Brazilian Stock Market

Political stability is essential to maintaining investor confidence in emerging markets like Brazil, whereas government instability leads to market volatility and higher interest rates. When a government has poor fiscal control, it increases uncertainty for investors, forcing them to demand higher returns or withdraw their money from the country. For example, when government spending exceeds what the government collects, an increase in the country's debt creates uncertainty for investors about tax changes, inflation, or whether the government can repay the debt; all of these risks discourage investors from investing in the Brazilian stock market. Agencies like Fitch Ratings have often pointed out that one factor that limits investor confidence in Brazil is the Brazilian government's GDP-debt ratio, which creates more financial risk and keeps B3 stock prices low (Fitch Ratings, 2026). A clear example occurred between 2015 and 2016, when credit rating companies like S&P, Fitch, and later Moody's downgraded Brazil's credit debt to “junk” status while it was under Rousseff's crisis rule, with the government having a budget deficit that exceeded 10% of GDP and debt that was expected to go over 80% of GDP. Because investors saw the risk in the market, they anticipated the move by pulling assets out of Brazil, which resulted in a heavy drop in Ibovespa for the rest of the year. In summary, weak financial management and political instability created significant problems for investors in the Brazilian stock market, as shown in the downgrades in 2015 and 2016 that resulted from the market's risk expectations from the market.

Political elections are a major catalyst for market volatility as investors shift their assets based on their confidence in the policies and administration of a candidate. After the first round of presidential elections in October 2018 in Brazil, with the result that candidate Jair Bolsonaro received the largest share of votes, the Bovespa index rose 4.6% in only one day, being the largest increase in Ibovespa in more than one year (CNBC, 2018). Furthermore, in the following weeks of October, the market grew 8% in the Ibovespa and 18% in the iShares MSCI Brazil ETF (EWZ); at the same time, global markets fell about 9% in the iShares MSCI ACWI ETF (ACWI). This shift in the market happened because investors were excited about Bolsonaro's promises of privatizing state companies, more independence of the central bank, and the appointment of the University of Chicago economist Paulo Guedes as minister of the economy, signals of a more market-oriented economic direction following a change in the party in power. As a result, political elections can be the reason for market shifts and volatility as investors move their money based on their policy and political expectations.

7. Role of Global Markets in Influencing Brazil's Economy

Brazil's capital flows, domestic economic conditions, and the Ibovespa's movement are substantially shaped by global monetary policies such as the U.S. Federal Reserve decisions. Historically, when the US interest rates increase, foreign investors tend to move their money out of emerging markets like Brazil into US assets to get safer and risk-free returns, which causes the Real to lose value against the dollar (Matschke et al., 2023). When the Real value is depreciated, the financial pressure on companies with dollar-denominated debt increases. This example was clear in June 2015, when Petrobras had US$111 billion in foreign-currency debt, 83% of its total debt (Bruegel, 2015), which resulted in investors seeing more risk in the company because the cost of paying the debt in dollars and reais increased, taking the price of Petrobras stocks down and finally pulling the Ibovespa overall down, as it is one of the companies with the biggest weight in the index. During the US Federal Reserve interest rate hike between March 2022 and July 2023, Brazilian assets held up surprisingly well, as the Brazilian Selic rate had already been increased since February 2021 (Federal Reserve Bank of Dallas, 2023), increasing the difference with the US rate and serving as a protective measure. As shown in Figure 1, the Selic rate began to increase in February 2021, more than a year before the Fed funds rate hike started in March 2022, considerably increasing the gap between the US and Brazil interest rates during this period. Figure 1 illustrates how this difference is still ongoing, with the Fed rate significantly lower than the Selic even after both rates became stable, indicating Brazil's high interest rates prevented foreign investors from pulling their assets out of the country even though capital flight usually happens when the U.S. raises interest rates. This protective gap had narrowed substantially by June 2026, however, with the possible US rate hikes, the Ibovespa fell from a 1% gain to a loss of 0.70% in one day because the foreign investors pulled their assets (News Directory 3, 2026), while in 2025, when there was a cut in the interest rate, companies like Vale and Bradesco had a rise in their stocks (Trading Economics, 2025). Overall, this shows how the

Brazilian stock market is largely affected by international interest rate decisions, which directly shape whether foreign capital enters or leaves the market.

Commodity companies have a large weight in Ibovespa, causing the index to be significantly shaped by global demand and prices of products like oil, iron ore, and agricultural goods (FOREX.com, n.d.). The Brazilian economy is largely made up of the exportation of commodities. A substantial part of the revenue generated by Petrobras goes back to support fiscal accounts and reserves, while Vale is one of the world's top iron ore exporters and sells to countries with large iron ore demand, like China (FOREX.com, n.d.). When global demand falls, the Brazilian market drops. In 2024, Vale lost about $17 billion in value because of the decrease in iron ore prices and the Chinese economy slowing down (Kolodny, 2025). Conversely, when there are problems with the global market, Brazil can still benefit. In the Ukraine-Russia war, at the same time as most world stocks fell, Ibovespa rose 8% at the beginning of 2022 due to Brazil being a major commodity supplier. These examples indicate that global market problems affect Brazil as a major global commodity supplier (Mais Retorno, 2022). In the same way that drops in demand can take the market down, the market can also be pushed up by shortages of products in other countries.

8. Interconnection of Factors

Interest rates, political influence, and the global market are often analyzed separately in terms of their effects on the Bovespa index. These factors are not independent; they reinforce and intensify one another, resulting in outcomes larger than what any one factor alone could produce. Brazil's increase in interest rates and political crisis did not happen as independent events. Factors between 2015 and 2016 both contributed to one another, worsening the Brazilian situation. A major factor that significantly contributed to the issue is the political crisis, which can be highlighted by several weak political decisions that led large credit agencies to downgrade Brazil to junk status, a budget deficit that exceeded 10% of the GDP, and debt that was expected to go over 80% of the GDP. At the same time, the Selic rate increased from 7.25% to 14.25% in a few months, which was due to decisions aimed at controlling inflation. It is not a coincidence that both events happened during the same period. As the Brazilian government lost confidence in its ability to manage its finances, investors took their assets out of Brazil, which weakened the country even more and increased prices even further, continuing the inflation that drove the central bank to increase the interest rate even further. This relationship between political influence and increasing interest rates is what made this crisis so significant for businesses that depended on borrowing money, such as PDG Realty, which was one of the largest Brazilian construction companies. The company's failure is not the result of high interest rates alone, nor is it the result of political instability; it happened because both factors were present simultaneously within the company. This shows that, to understand the Ibovespa, these factors must be considered together instead of being evaluated independently.

9. Conclusion

The Brazilian stock market is shaped by three major factors: Brazilian interest rates, political stability, and the global market. Interest rates largely affect how attractive stocks are compared to low-risk investments and how much companies borrow money. The political influence and stability impact the confidence of investors in keeping their assets in the country instead of pulling them out. The global market influences the stock market through foreign interest rates and global commodity prices and demand, which determine the capital flows in Brazil. These factors are interconnected and mostly don't happen by themselves, for example, a political issue can result in an increase in the interest rate, or a global market event can be the reason for a change in Brazil's own economic decisions, meaning that significant changes in the stock market often result from multiple issues combined instead of one. Because Brazil is an emerging market, it is more vulnerable to these forces together than a more developed economy, making it important for investors who are evaluating the Brazilian stock exchange to consider interest rates, politics, and the global market not as independent factors but as connected forces, since their combined movement has historically resulted in significant shifts in Ibovespa.

Works Cited

Agência Petrobras. (2026, March 6). Lucro líquido da Petrobras alcança R$110 bilhões em 2025. Petrobras. Retrieved July 27, 2026, from https://agencia.petrobras.com.br/w/negocio/lucro-l%C3%ADquido-da-petrobras-alcan%C3%A7a-r-110-bilh%C3%B5es-em-2025

B3 S/A - o que é, significado e definição | Bora Investir. (2025, July 23). Bora Investir B3. Retrieved June 18, 2026, from https://borainvestir.b3.com.br/glossario/b3-s-a/

Bovespa fecha em queda nesta terça e tem pior mês em mais de 20 anos | G1. (2020, March 31). G1. Retrieved June 30, 2026, fromhttps://g1.globo.com/economia/noticia/2020/03/31/bovespa.ghtml

Bovespa Index (Ibovespa). (n.d.). B3. Retrieved June 17, 2026, from https://www.b3.com.br/en_us/market-data-and-indices/indices/broad-indices/ibovespa.htm

Branco, M. (2016, August 31). Bolsa e dólar caem em dia de votação final do impeachment. Agência Brasil. Retrieved June 30, 2026, from https://agenciabrasil.ebc.com.br/politica/noticia/2016-08/bolsa-e-dolar-caem-em-dia-de-votacao-final-do-impeachment#:~:text=Mariana%20Branco%20%2D%20Rep%C3%B3rter%20da%20Ag%C3%AAncia,da%20ex%2Dpresidenta%20Dilma%20Rousseff.

Brazil Credit Rating Cut to Junk by S%P Amid Budget Strain. (n.d.). Bloomber. Retrieved July 23, 2026, from https://www.bloomberg.com/news/articles/2015-09-09/brazil-credit-rating-cut-to-junk-by-s-p-amid-budget-strain

Brazil has debt rating cut to 'junk' status as problems mount. (2015, September 10). The Guardian. Retrieved July 22, 2026, from https://www.theguardian.com/business/2015/sep/10/brazil-has-debt-rating-cut-to-junk-status-as-problems-mount#:~:text=as%20problems%20mount-,This%20article%20is%20more%20than%2010%20years%20old,10%20Sep%202015%2002.26%20EDT

Brazil Overview: Development news, research, data. (2025, October 15). World Bank. Retrieved June 17, 2026, from https://www.worldbank.org/en/Country/brazil/overview

Brazil: playing with fire. (2015, November 6). Bruegel. Retrieved July 27, 2026, from https://www.bruegel.org/blog-post/brazil-playing-fire

Carranza, Cano, & Nishimura. (2025, June 20). Emerging Markets in Latin America Are Not Just More of the Same. MSCI. Retrieved June 17, 2026, from https://www.msci.com/research-and-insights/blog-post/emerging-markets-in-latin-america-are-not-just-more-of-the-same

Com crise, mais incorporadoras podem pedir recuperação judicial. (2016, September 09). andré torres advogados associados. Retrieved July 23, 2026, from https://www.andretorres.adv.br/com-crise-mais-incorporadoras-podem-pedir-recuperacao-judicial

Com dívidas de R$7,7 bi, construtora PDG pede recuperação judicial. (2017, February 23). Correio Brasiliense. Retrieved July 23, 2026, from https://www.correiobraziliense.com.br/app/noticia/economia/2017/02/23/internas_economia,576059/problemas-da-pdg-nao-se-restringem-a-obras-paradas.shtml

Emerging-market countries insulate themselves from Fed rate hikes. (2023, August 8). The Federal Reserve Bank of Dallas. Retrieved July 27, 2026, from https://www.dallasfed.org/research/economics/2023/0808

Fitch Rating. (2026, February 10). Brazil's Rising Spending Offsets Tax Efforts, Limiting Fiscal Gains. Retrieved July 22, 2026, from https://www.fitchratings.com/research/sovereigns/brazils-rising-spending-offsets-tax-efforts-limiting-fiscal-gains-10-02-2026#:~:text=We%20forecast%20Brazil%20will%20post,category%20forecast%20median%20of%2053.6%25.

Ibovespa Falls and Dollar Rises After Fed Signals Interest Rate Stance. (2026, June 17). newsdirectory3.com. Retrieved July 27, 2026, from https://www.newsdirectory3.com/ibovespa-falls-and-dollar-rises-after-fed-signals-interest-rate-stance/

Impacts of the Russian-Ukraine Conflict on Global Agriculture Commodity Prices, Trade, and Cropland Reallocation. (n.d.). Choices Magazine. Retrieved July 27, 2026, from https://www.choicesmagazine.org/choices-magazine/theme-articles/turmoil-in-global-food-agricultural-and--input-markets-implications-of-russias-invasion-of-ukraine/impacts-of-the-russian-ukraine-conflict-on-global-agriculture-commodity-prices-trade-and-crop

Matschke, J. (2023, October 19). Capital Flows and Monetary Policy in Emerging Markets around Fed Tightening Cycles. Federal Reserve Bank of Kansas City. Retrieved July 27, 2026, from https://www.kansascityfed.org/research/economic-review/capital-flows-and-monetary-policy-in-emerging-markets-around-fed-tightening-cycles/

Moody's surprises with a two-notch Brazil rating cut. (2016, February 25). Business Standard. Retrieved July 23, 2026, from https://www.business-standard.com/amp/article/reuters/moody-s-surprises-with-two-notch-brazil-rating-cut-116022500082_1.html

Overview of exchange. (2026, February 1). Baker McKenzie. Retrieved June 17, 2026, from https://resourcehub.bakermckenzie.com/en/resources/cross-border-listings-guide/latin-america/sao-paulob3-formerly-bmfbovespa/topics/overview-of-exchange

PDGR3: de maior incorporadora à recuperação judicial; veja como a empresa está hoje. (2021, September 21). Eu quero investir. Retrieved July 23, 2026, from https://euqueroinvestir.com/educacao-financeira/pdg-realty-pdgr3-incorporadora-recuperacao-judicial

The relationship between political instability and stock market performance: An analysis of the MSCI index in the case of Pakistan. (2023, October 19). PMC. Retrieved July 22, 2026, from https://pmc.ncbi.nlm.nih.gov/articles/PMC10586669/

Vale Shares Shed $17 Billion as Iron Ore Slump Foils Turnaround. (n.d.). Bloomberg. Retrieved July 27, 2026, from https://www.bloomberg.com/news/articles/2025-01-27/vale-shares-shed-17-billion-as-iron-ore-slump-foils-turnaround

Vale volta a ser maior produtora global de minério Este trecho é parte de conteúdo que pode ser compartilhado utilizando o link https://valor.globo.com/empresas/noticia/2026/01/28/vale-volta-a-ser-maior-produtora-global-de-minerio.ghtml ou as ferramentas o. (2026, February 1). Valor Economico. Retrieved July 27, 2026, from https://valor.globo.com/empresas/noticia/2026/01/28/vale-volta-a-ser-maior-produtora-global-de-minerio.ghtml

WFH Case Study By Ajinkya Patil

Abstract

Work from home (WFH) policies are a strategic variable considered by both firms and investors yet empirical links between these policies and firm performance remain thin. This case study addresses that gap by comparing two similar large market cap tech firms, Amazon and Meta, with differing WFH policies from 2018 to 2024. I collected firm level data on stock performance, market cap, net income, and profit margins and compared them between the respective companies. Results showed that the flexible firm, Meta, demonstrated a sharper post pandemic recovery and higher stock returns relative to Amazon, consistent with the literature citing lower quits and stable productivity for digital work under hybrid remote models. While limited to two tech firms, findings imply flexibility can correlate with a competitive advantage where work is able to be digitised. Future research should expand across various different sectors and incorporate individual level data on productivity and tenure to test the effects of WFH under these conditions.

Introduction

Over the past several years, the nature of work in large organizations has undergone an extreme change, which was intensified by the COVID-19 pandemic. Driven by improvements in digital collaboration and increased flexibility for employees, remote and hybrid work models have shifted from being uncommon practices to long-term choices. This shift has raised ongoing debate among corporate leaders and investors about whether remote work should remain a permanent strategy or give way to a full return to the office.

The question usually revolves around three main issues: performance, retention, and shareholder value. Mentioned in the Bloom et al., 2024 study, advocates for in office work often believe that working face to face in a shared location helps build collaboration, connection, and fosters a strong workplace culture. Alternatively, supporters of flexibility point to studies suggesting that people can be just as productive or even more so, when working away from the office rather they thrive when in a comfortable and personal setting. While some claim that physical presence is crucial for fostering connections, flexible positions are a powerful tool for attracting and retaining talent in competitive job markets. Ultimately, For shareholders, the challenge is weighing the cost saving and wider access to talent that flexibility offers against potential risks to team cohesion and long term performance.

This paper provides insight into whether or not workplace flexibility affects shareholder returns and the drivers behind the changes. As companies continue to question how work models impact performance, the financial impact of WFH policies are becoming increasingly strategic. Understanding this relationship helps executives optimize their company policies and gives investors a new variable to determine if a company is worth investing in. Building on this, due to an increase in flexibility and the opportunity for employees to work in a setting that may be better suited for their needs, researchers have conjectured that companies who have flexible

WFH policies yield higher stock returns annually and maximize their employees efficiency than companies who don't.

To determine the relationship between workplace flexibility and firmstock performance, this study draws on a variety of both quantitative and qualitative data sources spanning 2018 to 2024. The goal is to compare companies with flexible WFH policies such as Meta to those enforcing in-office mandates like Amazon, using stock returns and revenues as the financial markers. These two companies provide a useful comparison as Meta has been one of the most vocal champions of flexible and remote work, even reducing office footprints, while Amazon has publicly mandated office returns and tied performance to physical presence. Both are large market cap tech firms with global scale, strong investor attention, and an abundance of available data, making them ideal for a controlled case study.

Literature Review

Measuring the relationship between workplace flexibility and firm performance is complicated by potential biases in both employee surveys and corporate disclosures. For instance, workers who value flexibility may be more likely to report positive experiences with WFH, causing job satisfaction to be mixed up with results surrounding self reported productivity, or focusing on their individual performance without fully accounting for the holistic impact. Large-scale corporate performance datasets also rarely capture the many nuances of workplace policies.

This segment looks at what current, already existing research and data exists about the connection between workplace flexibility and company performance, with a clear focus on stock returns and investor outcomes. Additionally, this section points out clear gaps in the existing research leading the way to how this paper's analysis will add to the wider discussion. Since remote and hybrid policies affect several areas including employee retention, productivity, and market value, the review draws from a variety of sources including experiments, datasets, government statistics, and general market analyses. The different studies all provide crucial evidence and context necessary in determining if workplace flexibility creates real, measurable benefits for companies and their investors. First we look at sources that cover how workplace flexibility influences employee output and company performance, testing if remote and hybrid work systems improve or diminish productivity. Studies in this section provide direct insight into the results of flexible policies, showing how they lead to measurable benefits for certain sectors.

Empirical evidence does suggest productivity improvements from hybrid arrangements. Bloom et al., 2024 conducted a mass randomized controlled experiment at Trip.com involving 1,612 employees across engineer- ing, marketing, and support divisions. The participants were randomly assigned either to a hybrid schedule, which involved working two days working from home per week, or they were instructed to stay fully in of- fice over a six month period. The study found that hybrid work produced no statistically significant change in performance metrics such as their manager-assigned performance grades, or their promotion rates over a two year period. However, it resulted in a 33% reduction in employee quit rates, in addition to 54%

decrease among employees with long commutes. The data also showed a slow improvement in manager approval for hybrid positions following the trial, indicating a slow shift in perceptions among company supervisors. The authors relate the rise in retention to better work life balance and lower overall stress levels, achieved without any noted drop in employee productivity. A clear limitation here, is that this trial was done at a single Chinese travel services company, which may limit applicability to both other industries, regions, and work cultures.

U.S. Bureau of Labor Statistics, 2023 provides broad evidence along with empirics supporting that the impact of remote work is extremely industry dependent. The authors found that the shift to remote work during and after the COVID 19 pandemic produced productivity gains in information, finance, and professional services, but caused declines in sectors such as manufacturing, transportation, and warehousing. These results show the nuances in remote work and indicate that the benefits of workplace flexibility are uneven and highly dependent on the type of the work, the degree of required physical presence, and the reliance on collaboration. BLS offers a unique viewpoint from other sources covering the more holistic impacts of WFH as it covers the sectors that have been negatively impacted in addition to those who have been affected oppositely. The sector- specific lens offered by BLS is useful for this study because it highlights where remote or hybrid work is most likely to improve efficiency, which in turn could contribute to stronger stock returns in certain sectors. The limitation is that the BLS data does not report for individual companies making it necessary to link these productivity patterns with firm-level market performance data from the other sources.

In contrast, this following section covers the role of workplace flexibility in employee satisfaction and turnover, showing how these factors connect to the previously mentioned company outcomes. For this study, the most relevant information comes from the source WFHResearch and its data on employee quit rates and satisfac- tion levels These measures help link flexibility directly to retention outcomes so that we can later draw its correlation with performance outcomes.

WFHResearch, n.d., is an ongoing source that collects mass survey data on workplace situations, employee preferences, and policy shifts within companies. Its reliability mainly comes from both its scale and methods of research. Their surveys reach thousands of people across various industries, occupations, and regions. Furthermore, the project is led by well known researchers in economics and its findings are published and cited by various different reports. Both its monthly and quarterly reports draw from thousands of different respondents across several sectors and geographies recording policy types and employee satisfaction ratings. WFHResearch.com's analysis by industry, occupation, and policy type are useful in identifying which sectors maintain high flexibility and for connecting those flexibility patterns with stock performance data from market sources like Yahoo Finance. The primary limitation here is that while the dataset offers extensive insight about policies, it does not include any company specific financial information. This gap means that, on its own, the source cannot directly link flexibility to shareholder returns.

While evidence on productivity and retention shows that flexible work can benefit both employees and firms, these studies have lacked empirics and missed connecting those outcomes directly to stock markets. To address this gap in information, these sources offer a direct link between workplace models and market outcomes covering how flexible work policies are associated with stronger performance and higher stock returns.

Huang et al., 2023 conducted a different large-scale randomized study examining how both remote and hybrid work arrangements influence firm performance in the post COVID period. Using a dataset of publicly listed companies across several different industries, they proxied workplace flexibility by searching for remote job postings, then compared this to their company level financial data, including stock returns, profitability, and market cap growth between 2020 and 2022. Their experiments, held constant for industry, size, and previous performance, indicated that the technology companies with more remote or hybrid listings had significantly higher annual stock returns than companies with more office based policies i.e less remote job listings. This pattern was especially clear in those organizations where competition is greater and where jobs and work can be carried out remotely with no compromises in productivity. The authors argue that the ability to attract and retain skilled workers from a wider geographic area, along with reduced facility costs, are the likely mechanisms driving this improved performance. However, the study's reliance on job postings as a proxy for actual workplace policies introduce more limitations, postings may reflect hiring intentions rather than typical company norms.

Beyond the individual firm studies completed by Huang, Li, and Xiao (2023) , broader datasets are needed to see if the relationship between flexibility and performance holds across the market.

FlexIndex, 2024 provides that perspective through collecting and analyzing workplace flexibility data for S&P 500 companies, assigning each firm a score for flexibility based on both the frequency and structure of the different work options. The dataset covers several different types of work and measures different variables like average in office days per week and general industry wide flexibility trends. The 2024 report finds that companies who ranked in the top quartile for flexibility had median three year total stock returns around 18 percent higher than companies in the lower quartile. This correlation applied to most industries but was especially clear in tech and media firms. For example, high flexibility tech firms not only outperformed their less flexible competitors in share price appreciation but also reported higher employee satisfaction and lower voluntary turnover rates in the survey data. However, the analysis is presented primarily as high-level visualizations, it does not provide any real datasets or clear tenure statistics which limits the ability to conduct different testing on the relationship between flexibility and performance. Taken together, the previous studies point towards a consistent link being that flexibility supports retention, improves performance, and in several cases increases revenue growth. Since much of this advantage comes through productivity gains, it is important to support that connection in broader financial literature.

Ang et al., 2017 creates a direct link between productivity improvements and stock market performance, providing the main basis for connecting workplace flexibility to shareholder returns. Through using a global dataset of publicly traded firms across multiple industries, they analyze changes in total productivity and connect these with annual stock returns over yearly periods. Their findings show a statistically significant positive relationship between productivity gains and returns and their test controls for size, industry, and various market conditions. Moreover, firms in the top quartile of productivity growth experienced, on average, roughly 7% higher annual stock returns compared to those in the lowest quartile. The authors claim that productivity in- creases efficiency and offers a competitive advantage, both of which raise company value in investor eyes. The limitation is that the study does not single out why exactly productivity improved, factors like tech innovation or general policies are mixed into the analysis, leaving the need for more specific research and studies.

Finally, this section analyzes how investors react to changes in company workplace policies, focusing on short term market reactions and their potential long term effects. This perspective is essential because it helps clarify if changes in the market reflect investor judgments about flexibility itself, or if policy announcements are overshadowed by other firm specific factors.

Flynn and Nair, 2024 show how the stock market reacts when companies change their workplace flexibility policies. They look at the 3000 biggest companies ranked by market cap that announced return to office mandates, increases in remote work, or shifts in hybrid positions between 2021 and 2023. Using company press releases and news reports, they confirm the timing of each announcement and then measure abnormal stock return data in the three day window around it. Overall, they find no consistent market reaction and, on average, investors neither rewarded nor punished companies for changing their policies in the short term. Still, there are some patterns by industry, tech and other similar firms saw small negative results following stricter return to office rules, while firms with large physical establishments saw neutral or slightly positive stock results. The authors also find that companies who had downward trends in recent stock performance and more physical office costs were more likely to enforce return to office mandates. Although the paper focuses on short-term effects, it helps clarify that the long-term impact on returns, retention, and productivity is something they don’t provide insight on.

Outside of academic research, there is also a growing belief that remote and hybrid work can improve productivity and profitability, and including this kind of qualitative information is useful since it shows how investors tend to interpret the trade offs of flexibility.

Forbes Council, 2021 offers a more opinion based perspective on remote work, highlighting its potential to increase profitability through cost reduction. The article explains that remote and hybrid work models allow companies to enter a broader and more diverse talent pool, reduce general expenses related to office space and tools, and increase workforce satisfaction by providing greater flexibility. This increased efficiency, Rojas believes, can enhance overall competitiveness and long term growth potential. While the article does contain any relevant

empirics or analysis, its value comes from offering a unique and clear possible perspective of why organizations would find remote working to be an attractive model. Ultimately, it highlights cost savings and talent attraction as reasons that could influence investor perception and due to that, increase stock performance. The key limitation is its lack of quantitative evidence and its lack of information related to hybrid work models, making it important to strengthen its claims with the empirics mentioned in our case study.

Materials & Methods

This study is quantitative since it focuses on analyzing data and statistics to compare stock returns across two different work models. A qualitative method wouldn't provide the clear, measurable results needed for this research. I considered using mixed methods, but since the goal is to find patterns in financial performance between matched companies, a data-driven approach is the best fit.

Starting around 2023, companies who continue to allow their employees to work from home will begin to gain sharper improvements in stock performance than companies who enforce their RTO mandates. If workplace flexibility truly enhances retention and productivity, then those advantages should eventually be reflected in stronger financial outcomes. To answer this hypothesis, I collected data for the years 2018 through 2024 across six key categories, revenue, profit margins, market capitalization (in trillions of USD), net income (in millions of USD), annual stock return percentages, and tenure or job satisfaction indicators for the year of 2024 exclusively. Together, these categories allow for both a company and market level analysis on how different workplace models translate into performance. The financial data was drawn from Amazon's and Meta's annual reports, SEC filings, Macrotrends, and Yahoo Finance, while workforce insights came from sources such as Visual Capitalist and WFHResearch. This combination of quantitative and qualitative sources provides the baseline needed to compare these two firms with contrasting workplace policies allowing us to evaluate if flexibility correlates with improved outcomes.

I compiled data for the years 2018 through 2024 across 6 key categories:

- Revenue
- Profit Margins
- Market Capitalization (in USD trillions)
- Net Income (in millions USD)
- Annual Stock Return (%)
- Tenure rates for 2024

I collected metrics for both Amazon which instituted their return-to-office policy in 2023 and Meta which allows flexible WFH in order to compare the financial outcomes and employee-related trends of two companies with differing work models.

Primary data sources included

- Amazon Annual Reports (2018–2024)

From Amazon's annual reports, I collected revenue, operating income, and net income results. I also noted the share count and headcount of employees where available. Overall, these reports gave me a good baseline for Amazon's long term trends across the different categories we measured.

- Yahoo Finance, 2025

Yahoo Finance is a market data website that collects the publicly available finance information for listed companies, including stock prices, shareholder returns, market cap, and earnings data. The platform's wide amount of data allows for specific comparisons across firms and sectors, and its functionality is crucial for deciding whether firms with flexible workplace policies have outperformed their peers in the markets over the observed period. However, while Yahoo Finance is a good tool for market performance tracking, it provides no direct insight into specific corporate policies or employee retention rates. This creates a gap between observed market outcomes and the causes behind them.

- Meta Platforms Annual Reports and Security and Exchange committee Filings (2018–2024)

From Meta's filings, I recorded their revenue, operating income, total net income, and share count from 2018 to 2024. These filings were essential for understanding why Meta's earnings trajectory did not always align with its stock market performance in certain years.

- Macrotrends.net

Macrotrends.net was mainly used for general analysis surrounding profit margins. While the site collects publicly reported data, I verified all key statistics against official SEC (security and exchange committee) filings before including them in the dataset. The primary strength of Macrotrends is its ability to provide multi year datasets for both Amazon and Meta which offered a quick way to compare the twos profits. However, I treated it as a secondary source, using it mainly to validate my calculations instead of a main data provider.

- CompaniesMarketCap.com

All data used in this study was publicly available and sourced from reputable financial databases or corporate disclosures. No private or identifiable employee data was collected. The company classifications surrounding workplace flexibility were based entirely on the collected

official statements from press releases, SEC fillings, and publicly released HR policies to avoid misinterpretation and eliminate the need for proxies to record tenure rates similar to how Huang Li and Xiao performed their test.

- Visual Capitalist

Visual Capitalist was used for workforce related context in 2024, more specifically, employee tenure and retention statistics from LinkedIn and resume datasets. While only available for one year, this dataset was useful in framing employee side impacts of flexibility policies. I used the information descriptively rather than quantitatively, to show how pressures and turnover rates may have influenced company policy decisions during the period studied.

- FinancialCharts.com

This source presented Stock Price information in a cleanly formatted way that allowed me to find annual prices of both Amazon and Meta for the years we compared. The stock return formula shown below measures how much a company's stock price changes from one year to the next. It does this by subtracting the previous year's stock price from the current year's stock price, then dividing by the current year's price to give the return as a percentage. The annual growth in revenues formula captures how much a company's revenue increased compared to the year before, subtracting the prior year's revenue from the current year's revenue and dividing by the stock price for that year to standardize the result.

Stock return formula – For each year t in the sample:

$$Stock\ Returns_t = (Stock\ price_t - Stock\ price_{t-1})/Stock\ price_t$$

Annual Growth in Revenues:

$$Annual\ Growth\ in\ Revenues_t = (Revenue_t - Revenue_{t-1})/Stock\ price_t$$

All financial data was cleaned and organized in a spreadsheet to allow for a cross-company comparison. Here is a subset of the data collected.

Table 1: Revenue Comparison Between Amazon And Meta

	Revenue (billions of dollars)		Annual Growth Rate in Revenue	
Revenue (billions of dollars)	Amazon	Meta	Amazon	Meta
2018	232.89	55.84		
2019	280.52	70.7	20.45%	26.61%
2020	386.06	85.97	37.62%	21.60%
2021	469.82	117.93	21.70%	37.18%
2022	513.98	116.61	9.40%	-1.12%
2023	574.79	134.9	11.83%	15.68%
2024	637.96	164.5	10.99%	21.94%

Table 2: Profit Margin Comparison Between Amazon and Meta

Profit Margins	Amazon	Meta
2018	4.10%	26.30%
2019	4.10%	26.10%
2020	5.50%	33.90%
2021	6.10%	33.00%
2022	6.40%	23.20%
2023	5.00%	22.20%
2024	8.00%	35%

Table 3: Market Cap Comparison Between Amazon and Meta (Trillions USD)

Market Cap in Trillion USD	Amazon	Meta	Annual Growth Rates in Market Cap (%)	Amazon	Meta
2018	0.7375	0.374	2018		
2019	0.9202	0.585	2019	24.77%	56.42%
2020	1.634	0.778	2020	77.57%	32.99%
2021	1.691	0.922	2021	3.49%	18.51%
2022	0.8569	0.32	2022	-49.33%	-65.29%
2023	1.57	0.91	2023	83.22%	184.38%
2024	2.081	1.49	2024	32.55%	63.74%

Table 4: Net Income Comparison Between Amazon and Meta (Millions USD)

Net Income Millions USD	Amazon	Meta
2018	10,073	22,110
2019	11,588	18,490
2020	21,331	29,150
2021	33,364	39,370
2022	-2,722	23,200
2023	30,425	39,100
2024	59,248	62,360

Table 5: Annual Stock Return Comparison Between Amazon and Meta (%)

Annual stock return %	Amazon	Meta
2018	26.3	-26
2019	20.1	57
2020	71.6	33
2021	4.6	35
2022	-50.7	-64
2023	77	110
2024	46.3	55

Results

In February 2023, Amazon enforced RTO mandates in place for its corporate employees. If our hypothesis is correct, being that workplace flexibility plays an important role in stock performance, then Meta, which continued to allow remote work, should show stronger gains. Since stock performance is a key way to measure the effects of WFH policies, we compared annual stock returns for Amazon and Meta through 2024, paying close attention to any differences before and after Amazon's RTO policy began. As shown in Chart 1 Annual Stock Returns (%), Meta's stock rose more sharply than Amazon's, supporting our hypothesis.

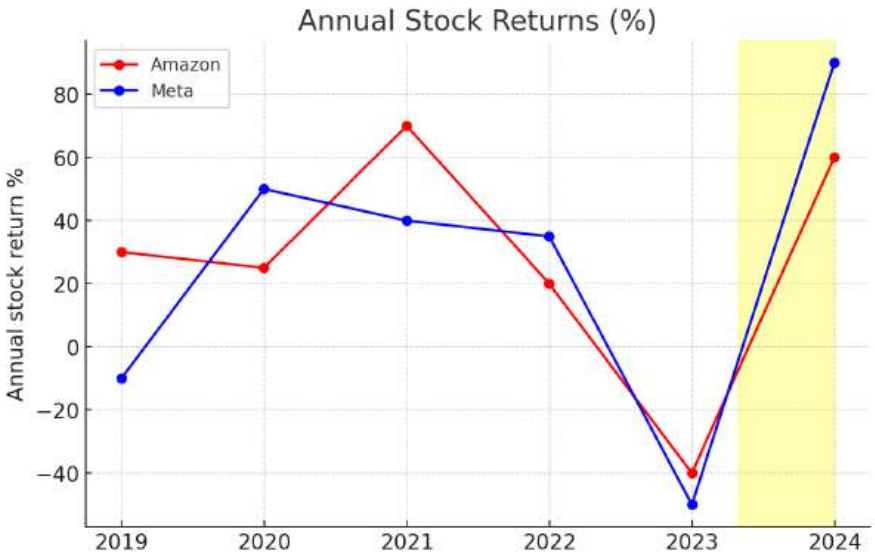


Figure 1: Annual Stock Returns (%)

Here, the stock returns mirror the major investment strategies. Amazon and Meta followed similar paths but with different magnitudes. Both saw strong gains in 2020–2021, large declines in 2022–2023, and a sharp increase into 2024. Meta's rebound was especially strong, delivering nearly 90% returns compared to Amazon's 60%. This suggests that, while both stocks were hit hard during the downturn, Meta offered higher rewards to shareholders during the recovery. Notably, an article by the Wall Street Journal mentions that upon Amazon's announcement of their RTO mandate, the stock dropped 8.27% supporting our narrative that the mandate was met with investor displeasure.

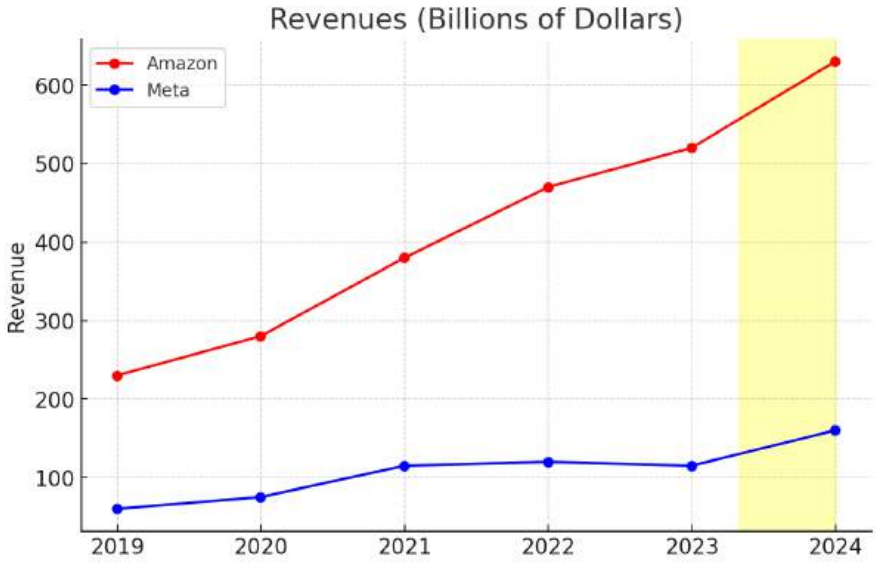


Figure 2: Annual Stock Returns (%)

Here we see Amazon's revenue advantage over Meta. Amazon's revenues rose steadily each year, climbing from just over $200 billion in 2019 to more than $600 billion in 2024. Meta's revenues, by comparison, are much smaller, only reaching about $150 billion by 2024. While Meta grew rapidly in some years, its scale is still far below Amazon's, underscoring that the two companies operate at very different magnitudes, even if their growth rates sometimes align. Considering that Amazon has always had a greater revenue than Meta in this graph, in contrast, a clear conclusion is not available from this graph as the grand difference between revenues has always existed. Since no conclusion could be made, looking at the Growth Rates in Revenue (%) combats this,closing the different magnitudes these companies operate on between the two companies and showing that these two companies follow comparable trends despite this grand difference in revenue.

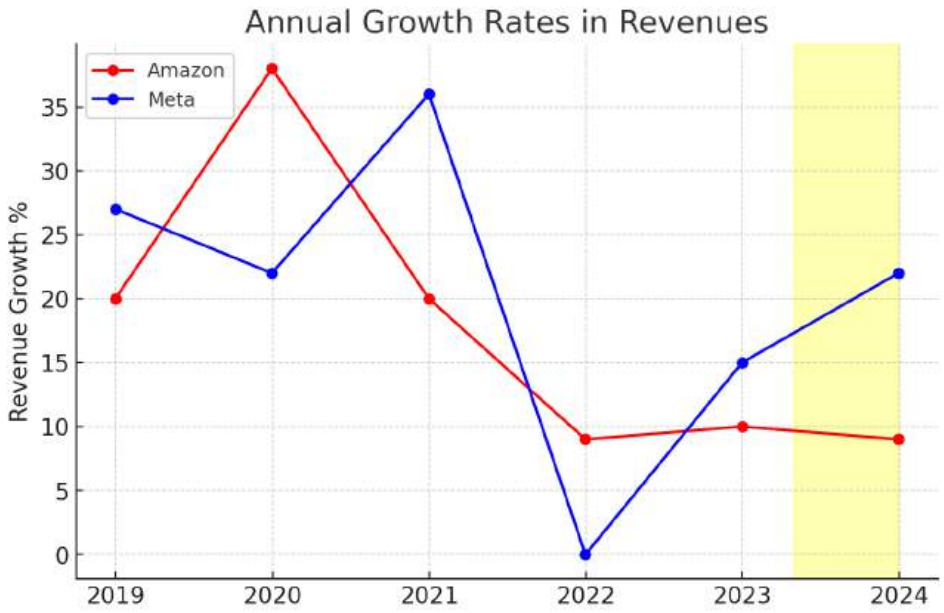


Figure 3: Annual Stock Returns (%)

This chart shows extreme volatility as Amazon saw rapid growth in 2019–2020 and Meta from 2020-2021, but sharp declines in 2021–2022, especially for Meta (nearly – 60%). Meta's market cap rebounded strongly in 2023, with growth above 15%, while Amazon's growth was more muted at around 10%. This difference signals that Meta's flexibility was a competitive advantage during recovery years, while Amazon's rigidity may have dampened investor enthusiasm. Since the highlighted portion signified when Amazon's RTO mandates went into effect, this indicates that there was a slight negative impact associated with them as shown in the decrease from 2023-2024. Here, the comparison clearly shows how workplace policy influences confidence in a company's ability to retain talent and sustain growth.

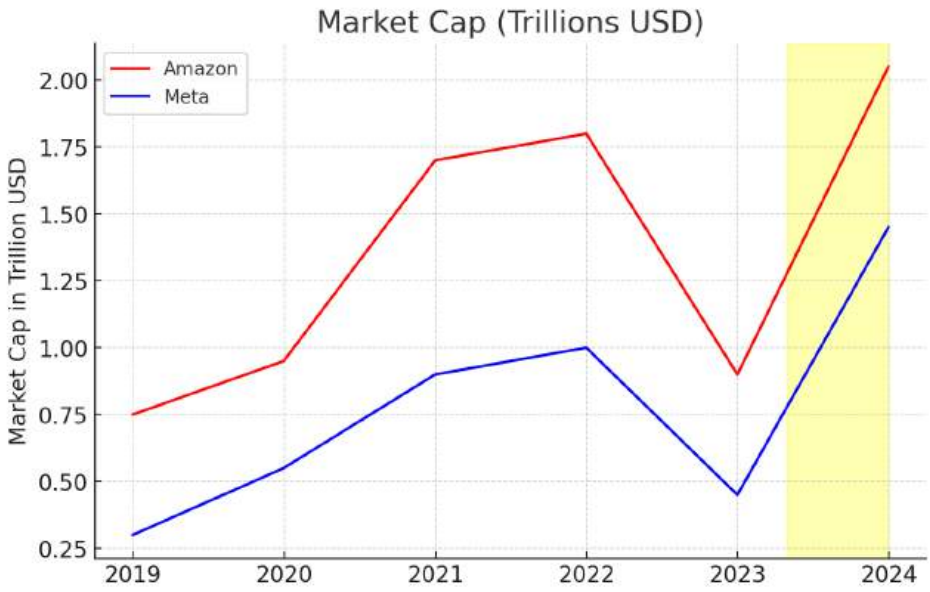


Figure 4: Annual Stock Returns (%)

For Market Cap, the dips from 2022 -2023 reflect a broad slowdown, but the almost identical recovery between Meta and Amazon by 2024 don't suggest much of a difference in market cap caused by the RTO mandates enforced by Amazon. Market cap reflects the overall company value which is driven by many other factors like certain trends and investor sentiment, so we wouldn't necessarily expect to see the effects of WFH policies there. Instead, those effects become clearer when we shift from market cap to growth rates, where differences between Amazon and Meta begin to show."

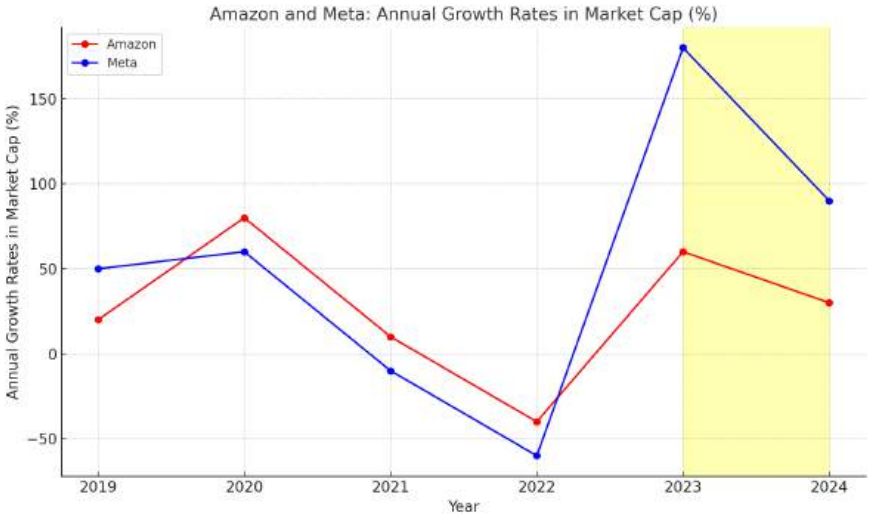


Figure 5: Annual Stock Returns (%)

Offering a clear competitive advantage during recovery years Meta's continued flexibility supports our hypothesis. Conversely, Amazon's rigidity lowered the expectations of investors. Both companies saw rapid growth in 2019–2020, but sharp declines in 2021–2022, especially for Meta (nearly –50%). Meta's market cap rebounded strongly in 2023, with growth above 180%, while Amazon's growth was lower at around 60%. This difference signals that Meta's recovery impressed investors more, while Amazon's stock was steadier and less explosive. Proving that hybrid flexibility may have contributed to stronger investor confidence in Meta's long-term potential, this graph is also critical in showing how WFH policies can influence not only productivity but also market value.

Discussion

Both Chart 1 and Chart 2 highlight how flexibility in the workplace shaped outcomes for Amazon and Meta in the years following COVID 19. Chart 1 illustrates how the two companies moved nearly identically through the pandemic, but began to diverge in 2023 when Amazon enforced stricter RTO policies. Upon the policies being enforced, Meta's stock returns pulled ahead, suggesting that investors viewed its continued flexibility as a strength, while Amazon's mandate was seen as a relative disadvantage. Chart 2 reinforces this trend on the revenue side, Amazon's growth rate slowed after 2023, while Meta's began to climb upward again. Putting these together, the patterns support my thesis that maintaining WFH policies is linked to stronger overall company performance, especially during periods of market recovery.

These results connect directly to the broader literature. Bloom et al., 2024 found in their randomized controlled trial at Trip.com that hybrid policies produced no measurable drop in productivity but drove a 33% reduction in quits overall and a 54% reduction for long commuting employees. That empiric helps explain the divergence we see in Chart 1, Meta's continued flexibility likely improved retention during a tight labor market, stabilizing teams and reducing the costs associated with turnover. Additionally, we likely don't see the initial transition to WFH as beneficial due to the other global events occurring at the time such as Covid. This analysis also matches the survey data from WFHResearch, n.d., which shows that remote and hybrid options consistently rank among the strongest predictors of employee satisfaction across industries. By contrast, Amazon's stricter mandates may have introduced friction, both among employees and financially, leading to slower post 2023 recovery as shown in Charts 1 and 2. Here, the implications are twofold. For companies, the diverging results between Amazon and Meta indicates that workplace flexibility is more than an HR preference; it can shape financial outcomes in critical ways. Firms that maintain hybrid or remote options may be better positioned to retain employees, sustain revenue growth, and deliver stronger returns during recovery periods, mainly in industries where digital work translates directly into output. The second implication is how for investors, flexibility provides an additional lens for evaluating long term stability, as companies that balance retention and productivity through the WFH models may outperform stricter peers. At a broader level, these results strengthen the connection drawn in the literature between employee dynamics such as turnover rates and satisfaction and firm level

outcomes like productivity and stock performance, suggesting that flexibility should be treated not only as a cultural factor but as a driver of this competitive advantage.

Taking into account the nature of this analysis being based on a case study between two companies, these findings should be interpreted with caution. The analysis focuses on only two firms, Amazon and Meta, which limits the ability to generalize across all industries or business models. Their divergence after 2023 may reflect more than just workplace policy. It should definitely be noted that there are several variables that play roles for the two companies' stock prices and revenue in ways that are not publicly available. In addition, the data used here captures financial outcomes such as revenues, market cap, and stock returns, but it cannot fully isolate the drive causes like changes in employee productivity, collaboration, or morale that change those results. That being said, the link between flexibility and performance, while strongly suggestive, cannot be treated as 100% cause and effect.

Conclusion

This case study compared Amazon and Meta from 2018 to 2024 to analyze how workplace flexibility, specifically remote and hybrid work policies, affect shareholder returns. Using financial data, retention statistics, and the existing literature surrounding the topic, I analyzed whether companies that allow employees to work from home perform better in the stock market than those that enforce in-office mandates.

The evidence supports my thesis that workplace flexibility is associated with stronger stock performance, improved recovery trends, and greater investor confidence. Meta’s continued support of hybrid work correlated with sharper post-pandemic growth, while Amazon’s stricter RTO approach was met with slower gains and a weaker rebound. These patterns align with broader research linking flexibility to higher retention, stable productivity, and enhanced company value.

To strengthen this conclusion, future research can expand the sample beyond two companies and explore other industries where the impact of flexibility may be more nuanced like healthcare or education. Additional analysis could also analyze specific productivity metrics and work quality, offering a wider understanding of how flexible work influences both company culture and financial performance.

Ethical Considerations

All data used in this study was publicly available and sourced from reputable financial databases or corporate websites. No private or identifiable employee data was collected. Company classifications on workplace flexibility were based entirely on official statements and publicly documented HR policies to guarantee accurate categorization

Works Cited

Ang, T. C., Lam, F. Y. E., & Wei, K. C. J. (2017). Understanding the firm-productivity effect in stock returns. EFMA Annual Meetings. https://www.efmaefm.org/0EFMAMEETINGS/EFMA%20ANNUAL%20MEETINGS/2017-Athens/papers/EFMA2017_0242_fullpaper.pdf [Accessed: 2025-09-22].

Bloom, N., Han, R., & Liang, J. (2024). Hybrid work is a win-win-win for companies, workers, and society. [Accessed: 2025-09-22].

FlexIndex. (2024). Workplace flexibility data for S&P 500 (2024 report). [Accessed: 2025-09-22].

Flynn, G., & Nair. (2024). Do markets care about return-to-office? Evidence from the Russell 3000. Journal of Financial Economics. https://www.sciencedirect.com/science/article/abs/pii/S0304405X19302259 [Accessed: 2025-09-22].

Forbes Council. (2021, October). How remote work can increase business profits. [Accessed: 2025-09-22].

Huang, W., Li, N., & Xiao, Y. (2023). Remote work and firm financial performance during and after COVID-19. Journal of Corporate Finance, 80. https://www.sciencedirect.com/science/article/pii/S0927537123001379 [Accessed: 2025-09-22].

U.S. Bureau of Labor Statistics. (2023). The rise in remote work and its impact on productivity. Beyond the Numbers, 13. https://www.bls.gov/opub/btn/volume-13/remote-work-productivity.htm [Accessed: 2025-09-22].

WFHResearch. (n.d.). WFHResearch.com. [Accessed: 2025-09-22].

Yahoo Finance. (2025). Yahoo Finance. [Accessed: 2025-09-22].

Segregated Work, Unequal Earnings: Occupation, Payment Systems, and the Gender Earnings Gap in Bombay City Cotton Mills, 1921-1937 By Anaisha Poddar

Abstract

This paper examines the gender earnings gap in Bombay City cotton mills between 1921 and 1937, focusing on how the gap changes when women and men are compared within the same occupations and payment systems. Using the TextileLab India textile-wages database and contemporary official inquiries, the analysis finds that women's average daily earnings were only around 54 to 59 per cent of men's when all detailed observations are combined. However, when large groups of women and men with the same original occupation title and payment system are compared, the female-male earnings ratio rises to around 87 to 95 per cent. This suggests that occupational segregation was a major part of the overall gender earnings gap. The pattern was not uniform across jobs. In several time-rated spinning occupations, women's and men's earnings were close, while sizable gaps remained in piece-rated winding. Piece work therefore mattered mainly because women were heavily concentrated in departments where earnings depended on output and the availability of work, not because piece rates always paid women less. The evidence supports a cautious conclusion: gender inequality in Bombay mill earnings was produced largely through the organisation of work, access to occupations, and uneven exposure to insecure forms of employment.

Introduction

Bombay was one of the largest industrial cities in colonial India. Its cotton mills turned raw cotton into yarn and cloth through many separate stages of work, and the wage question cannot be understood without this division. Women and men were rarely spread evenly across the mill. Women were especially common in winding and reeling, while men dominated weaving, warping, supervisory work, and many spinning jobs. Pay was organised differently too. Time-rated workers were mainly paid according to time worked, while piece-rated workers' earnings depended on output. In the 1934 General Wage Census, 80.52 per cent of female process workers in Bombay City were piece-rated, compared with 49.76 per cent of male process workers (Labour Office 76-78). An average wage for all women and an average wage for all men thus mixes together different occupations, payment systems, hours, output, and access to work.

This paper uses the TextileLab India textile-wages database to separate these differences as far as the historical sources allow. The main quantitative result is large. In the detailed Bombay City observations with usable worker counts, women's recorded average daily earnings were about 54 per cent of men's in 1921, 54 per cent in 1923, 59 per cent in 1934, and 55 per cent in 1937. When the comparison is restricted to large groups with the same original occupation title and the same payment system, the ratios rise to about 87, 95, 87, and 93 per cent. That does not mean occupational segregation 'explained' a precise percentage of the gap, as such. The matched comparison changes which workers are being compared. It shows that the

industry-wide gap was closely connected to where women and men stood in the mill's occupational structure, but it cannot identify one exact causal share.

The same-job evidence is not identical across occupations. Time-rated ring-spinning jobs such as doffing and siding show female and male averages that were often close. Winding is different. In several large piece-rated winding groups, women earned noticeably less than men with the same reported title. Even here, the official inquiries themselves warn against treating the number as simple proof of unequal pay for identical work. Winding and reeling could differ by yarn, machine, hours, output, and the amount of work available. The Textile Labour Inquiry Committee found that these departments often contained more workers than the available work required, so a piece worker's observed daily earnings could be low because there was not enough work for a full day (Government of Bombay 127-28). The strongest conclusion, then, is about the organisation of work rather than one universal wage rule.

Bombay Cotton Mills and the Gendered Labour Market

Bombay's cotton mills were predominantly male, and the female share was falling by the middle of the 1930s. Morris's employment series places women at 20.49 per cent of the Bombay cotton-mill workforce in 1921, 22.62 per cent in 1926, and 22.86 per cent in 1930. The share then fell to 18.94 per cent in 1934 and 15.92 per cent in 1937 (Morris 66, 217-18). At the same time, the women who remained in the mills were concentrated in a small number of departments. Kumar shows that winding and reeling were heavily female, while women rarely worked as spinners and were almost absent from weaving (81-83). In this paper, occupational segregation means the separation of women and men into different jobs. It can create a large overall earnings gap even if pay in some jobs shared by both sexes is much closer, because high-paid and low-paid occupations are not divided equally.

The wider economy, too, made this job structure important. Bombay mills sold different kinds of yarn and cloth in markets where demand changed quickly. Chandavarkar shows that millowners often adjusted production to short-term market conditions and relied heavily on labour rather than continuously replacing workers with the newest machinery. Mills differed in machinery, products, finances, and local labour relations, so they did not face exactly the same conditions. Labour costs were still one major area where management could respond when business became difficult (Chandavarkar 278-79, 313-16, 343). This pressure became especially visible in the early 1930s. Many mills cut wages in 1933-34, and the 1934 wage census in Bombay had to be taken in October rather than May because a major strike disrupted the original schedule. By then the Bombay Millowners' Association had announced minimum consolidated time rates for several jobs, but pay still varied between mills (Labour Office 17, 82-83; Morris 172-74).

Workers did not simply accept these changes. Bombay had major strikes in 1928 and 1929, communist and other trade-union organising, and repeated disputes over wages and the organisation of work. Women also participated. Kumar describes women winders at the Jacob Sassoon Mill striking in 1935 over wages, management harassment, and changes in hours

(95-96). The useful issue, then, is not whether women were naturally passive, but how much bargaining power different workers had. A worker in a skilled or difficult-to-replace job faced management from a different position than a worker in a crowded department where many people depended on the same limited amount of work.

The supply of labour to Bombay was also gendered. Many mill workers migrated from rural western India. Dixit and van Nederveen Meerkerk argue that rural production often encouraged men to migrate for factory work while women's labour remained important in agriculture and household production (247-49). Some poorer women did enter the mills, but many were channelled into low-paid winding and reeling. This did not make their earnings unimportant to the household. Kumar notes that many female mill workers were widows or major earners, and some supported unemployed husbands and children (82-83, 87-90). Men and women thus did not enter the urban labour market with the same migration patterns, family responsibilities, or outside options, and these differences helped keep the job structure divided by gender.

Data and Method

The main quantitative source is the TextileLab India textile-wages database. It compiles wage information from Bombay Labour Office inquiries and later official investigations. The database is not a list of individual workers. Each row is a reported observation for an occupation or group, usually an average wage or average earning. LAB_N gives the number of workers behind that average when the historical source reports it. A row with an average daily earning and LAB_N = 500, for instance, represents one group average based on 500 workers, not 500 separate wage records. The database manual also warns about the term 'piece'. In many rows, 'piece' refers to the average earnings of workers paid according to output; it does not necessarily record the literal price paid for one unit of yarn or cloth (van Nederveen Meerkerk and Dixit, Manual 3-6).

The surveys are not identical across years. The 1921 and 1923 enquiries covered firms employing a large majority of workers in the main textile centres. The 1926 enquiry covered a much smaller sample. The 1933 departmental enquiry was narrower again and included only selected occupations. The 1934 General Wage Census was much broader and used factory muster rolls to separate men, women, children, time workers, and piece workers. The 1937 Textile Labour Inquiry used another method: some time-worker information came from wage notices displayed in mills, while piece-worker earnings were still drawn from muster records, and the published summary tables mainly describe day-shift workers in July 1937 (Labour Office i-ii, 17; Government of Bombay 9-10). So a difference between two years can partly reflect survey coverage, and not only a real change in work or earnings.

To make the comparison as consistent as possible, the main sample keeps mechanised cotton-textile observations for Bombay City only. It excludes the wider suburban area and Kurla because those locations were not always surveyed in the same way as the city. It keeps adult men and women, daily earnings, and observations clearly marked as either time-paid or piece-paid.

Broad 'all operatives' rows are removed so that detailed occupational rows are not counted a second time. Exact logical duplicates are also removed. The 1926 data need another precaution: many LAB_N entries are written as 'xxx'. The manual explains that these are combined summary figures rather than normal worker counts, so the rows remain in the analysis file but are excluded from worker-weighted calculations (van Nederveen Meerkerk and Dixit, Manual 6).

Matching occupations is the most delicate step. TextileLab provides a historical occupation title, OCCSOURCE, and a modern standardised category, OCCSTAN. The standardised category is useful for broad grouping, but it can combine jobs that were not actually the same. For example, a broad category can place creelers and reelers together even though the tasks differ. The officials who produced the 1934 census faced the same problem: earlier enquiries had found more than a thousand local and vernacular job names for roughly 150 occupations, so the Labour Office had to construct standard lists (Labour Office i-ii). For the narrow comparison, I match women and men only when they have the same original OCCSOURCE title, the same payment system, and the same technology when the source explicitly records a technological difference.

A matched cell is called 'large' only if it includes at least 50 men and 50 women. Within each year, the matched ratio also uses the same occupational mix for both sexes. Each included cell is weighted by its combined male and female worker count when constructing the male and female matched averages. This avoids allowing the matched result to be driven simply by women and men being distributed differently across the jobs that survive matching. Even this is not proof of identical work. Two workers with the same title could still use different machines, yarn counts, raw materials, products, or local work arrangements, and mills could have different wage scales (Morris 155-59; Chandavarkar 313-16). The matched results are, as such, careful descriptive comparisons and not a controlled experiment.

The wage measure is nominal average daily earnings. It is not adjusted for prices or the cost of living, and it is not identical to a basic daily wage rate. The 1934 report calculated average daily earnings by dividing the wages received by a group by the days actually worked. A worker's daily earnings could change, then, because the wage rate changed, because hours changed, because output changed under piece work, because not enough work was available, or because bonuses and attendance rules changed (Labour Office 81-82). For this reason, the paper uses the term earnings gap and does not assume that every difference is a difference in the rate paid for the same quantity of labour.

Results: Broad Averages and Matched Jobs

Table 1 gives the central result. The first ratio is based on the detailed observations in the cleaned Bombay City sample and weights group averages by the number of workers when LAB_N is usable. The second keeps only large matched occupation-payment cells and gives the same cell mix to women and men. A ratio of 1.00 would mean equal average daily earnings in the groups being compared. A ratio of 0.54 means that women's average was 54 per cent of men's.

Table 1

Women's reported daily earnings as a share of men's in Bombay City cotton mills (1.00 = equal average earnings)

Year	All detailed observations: F/M ratio	Large matched job groups: F/M ratio
1921	0.543	0.865
1923	0.535	0.954
1934	0.585	0.870
1937	0.552	0.929

Source: Author's calculations from the TextileLab India textile-wages database. Large matched cells require at least 50 men and 50 women with the same original occupation title and payment system; technology is also matched when explicitly recorded. The 1926 and 1933 results are not shown because worker-count coverage and occupational coverage make them unsuitable for the headline comparison.

The contrast in Table 1 is too large to ignore, but it should not be turned into a causal percentage. The two columns do not, as such, contain exactly the same workers. Matching removes occupations that are strongly male or strongly female, and the matched groups are not intended to reproduce the entire workforce. The defensible statement is narrower: women and men looked much more unequal when the comparison crossed the full hierarchy of mill jobs than when it stayed inside the smaller group of occupations that both sexes shared in substantial numbers.

The occupational structure helps explain why. In the 1934 census, Bombay City had 12,900 female winders and 5,455 male winders, and 6,314 female reelers compared with 1,565 male reelers. Across the Bombay Presidency, about 82 per cent of winders were women in 1934, and by July 1937 92.6 per cent of day-shift winders were women. Reeling was even more strongly female. Warping, by contrast, was entirely male in the 1937 inquiry, and weaving in Bombay was effectively a male occupation (Labour Office 14-15, 106-8; Government of Bombay 29-32, 38-41). The average female mill worker and the average male mill worker were, more often than not, workers in different jobs. This is why an industry-wide female wage cannot automatically be read as the wage paid to a woman for the same work as a man.

The clearest matched evidence comes from time-rated ring-spinning jobs. Doffers removed full yarn packages and replaced them so production could continue. Siders tended ring-spinning frames; the sources distinguish single-side and double-side siding. In 1934 the female-male earnings ratio was 0.944 for doffers, 0.969 for single-side siders, and 1.005 for double-side siders. In 1937 the corresponding ratios were 0.976, 0.970, and 1.037. These figures do not prove that every woman and man had the same machine, workload, or experience. They do show that in several large shared time-rated jobs, the average gender gap was small compared

with the overall ratio of roughly 0.55. The official 1934 report points in the same direction for wider jobs across the Presidency (Labour Office 102-3).

Table 2

Selected large matched female-male daily earnings ratios (1.00 = equal average earnings)

Year	Occupation	Payment system	F/M ratio
1934	Doffers	Time	0.944
1934	Single-side siders	Time	0.969
1934	Double-side siders	Time	1.005
1934	Grey winders	Piece	0.799
1934	Colour winders	Piece	0.813
1934	Pirn winders	Piece	0.803
1937	Doffers	Time	0.976
1937	Single-side siders	Time	0.970
1937	Double-side siders	Time	1.037
1937	Pirn winders	Piece	0.633

Source: Author's calculations from the cleaned TextileLab Bombay City sample. Every row shown contains at least 50 workers of each sex and matches the original occupation title and payment system.

Winding is the main exception to the near-equality seen in the ring-spinning examples. In the large Bombay City piece-rated groups in 1934, the female-male ratio was about 0.799 for grey winders, 0.813 for colour winders, and 0.803 for pirn winders. The official report also gives the money averages. Piece-rated grey winders averaged about 12 annas 6 pies per day for men and 9 annas 11 pies for women; colour winders averaged 1 rupee 0 annas 6 pies for men and 13 annas 5 pies for women; pirn winders averaged 14 annas 5 pies for men and 11 annas 7 pies for women (Labour Office 106-8). Before decimalisation, 16 annas made one rupee and 12 pies made one anna. By 1937, male comparator groups in several winding categories had become very small because the occupation was even more female. Pirn winding still had a large male comparator group, and there the female-male ratio was only about 0.63.

These winding differences are important, but the wage table by itself does not tell us why they existed. Grey, colour, and pirn winding were listed separately because the material and work were not identical. The 1937 committee also warned that the amount of work available in winding was rarely constant and that average earnings could move for this reason (Government of Bombay 30-32). Matching by title is, hence, better than comparing all women with all men, but it still does not prove that a male and female winner produced the same output under the same conditions. The within-title gap is evidence that needs explanation, not automatic proof of direct wage discrimination.

Piece Rates and the Availability of Work

Piece payment was strongly linked to gender. In Bombay City in 1934, 80.52 per cent of female process workers were piece-rated, compared with 49.76 per cent of male process workers (Labour Office 76-78). This mainly reflected occupational concentration. Winding and reeling, where women were numerous, commonly used piece payment, while many jobs in spinning, supervision, and engineering were time-rated. Payment system was thus part of the gendered job structure: women were more likely to be in occupations where the day's earnings moved with output and the amount of work available.

That does not mean piece work always pays less. Some highly paid male jobs, especially weaving, were piece-rated. The 1937 inquiry also found occupations in which piece workers earned more than time workers; for one-head drawing tenters, it reported lower average earnings for time-rated workers in centres using both systems (Government of Bombay 17). The distinction matters. The evidence supports the statement that women were more exposed to piece work. It does not support the stronger statement that piece payment, by itself, always reduced earnings.

The effect of piece payment depended on how work was organised. More output could raise earnings, but the worker also carried more risk when production slowed or less work was assigned. Literal piece prices were difficult to compare between mills because machines, materials, yarn counts, and products differed. The 1934 census did not attempt one general table of piece prices for the entire industry for this reason. Morris also shows large differences between mills for the same named occupation. In the 1926 survey, the gap between the lowest- and highest-paying mills was about 87 per cent for grey winders and 73 per cent for reelers (156-59). A job title, in other words, did not come with one Bombay-wide piece rate.

Reeling and winding show most clearly how payment method and work availability could interact. The 1934 census said Bombay mills employed women in these departments considerably beyond the amount of work required. Many women in the same departments were allowed shorter days, usually about eight to eight and a half hours rather than the ten-hour day common in much of the industry. The report linked this arrangement to spreading employment among more women (Labour Office 13-15). A woman's lower daily earnings could, thus, reflect a lower price, fewer hours, less work, a different machine or material, lower output, or several of these at the same time.

The Textile Labour Inquiry Committee made the problem of insufficient work even clearer. It argued that actual earnings in reeling and winding did not give a reliable measure of what a worker could earn in a full day because the number of workers often exceeded the work available, and the amount of work changed from month to month. Workers were reported to prefer sharing the available work rather than seeing some of their group quickly retrenched. The committee recommended a guaranteed minimum time payment when piece earnings fell too low: 11 annas per day for reelers and grey winders, and 12 annas for high-speed, pirn, and colour winders (Government of Bombay 127-28). This recommendation matters, because it shows that

officials themselves saw low daily earnings as partly connected to underemployment within the factory.

Piece rates were, as such, part of the mechanism, but not a universal cause. Many women were concentrated in departments where employers could keep a flexible labour force and where available work could directly affect the day's earnings. Under piece payment, this instability appeared immediately in the wage packet: if a worker produced less or was given less to produce, earnings fell. The gendered outcome came from the combination of occupational segregation, excess labour in female departments, irregular workloads, and payment by output.

Work Organisation, Gender, and Bargaining Power

A wider explanation begins with the relationship between millowners and workers. Owners controlled the factories and tried to keep production profitable, while workers depended on wages and tried to protect jobs, earnings, and working conditions. This did not create the same conflict in every mill. Machinery, products, raw materials, finances, and labour relations differed from factory to factory. Recruitment and supervision were often handled through jobbers, who could recruit workers, allocate work, train workers, and enforce discipline. Mills also relied on badlis, substitute workers who filled in for absentees and formed a flexible pool of labour. These arrangements helped management adjust employment, but they also produced disputes over hiring, favouritism, discipline, and job security (Chandavarkar 278-98).

Rationalisation added another source of pressure. In this period it could mean new machinery, but it could also mean reorganising work so fewer workers handled the same production, making one worker tend more machines, changing work groups, or standardising wages. These changes could raise output per worker while also removing jobs or increasing strain. The technical details mattered. The 1937 committee noted that a single-side sider did not necessarily tend one side because the worker was less skilled; on some coarse yarn, the machinery made double-side work impractical (Government of Bombay 124-25). It is risky, therefore, to turn a job label or a machine count into a simple measure of individual effort.

Women faced this pressure from a weaker position in the occupational hierarchy. Their share of spinning work fell while winding became more strongly female. Van Nederveen Meerkerk and Dixit argue that employers used work organisation and gendered ideas about skill as part of broader cost-control strategies ("Human Beings" 186-91). The wage data fit part of this pattern: female and male earnings were close in several larger sider jobs, while winding showed a much larger female disadvantage. But the wage records cannot tell us what every employer intended. They do not prove that managers deliberately moved women into low-paid work for the purpose of discrimination. They do show that women were concentrated in lower-paid departments where piece payment and irregular work were common.

The route to better jobs also mattered. The 1937 committee heard complaints that doffers sometimes stayed in low-ranked jobs for years while outsiders were brought into better positions (Government of Bombay 359-60). Movement up the job ladder was, thus, contested even among male workers. If women had still fewer routes into weaving, warping, higher spinning jobs, or

supervision, the mill's wage structure could keep producing an overall gender gap even where women and men in one shared job received almost the same daily earnings.

Gender cannot be reduced to employer policy alone. Family migration, rural employment, household needs, and law affected who could take particular jobs and shifts. Morris argues that one reason women's share of Bombay mill employment fell after 1931 was the growth of night-shift work. Women were legally barred from mill night work, so expanding night production created work that men could take but women could not (65-67). This also affects the evidence: the 1937 wage inquiry mainly reports day-shift wages even though 99 mills with about 70,000 workers operated night shifts (Government of Bombay 9-10). The same fact is, then, part of both the historical explanation and the survey limitation.

Women's weaker position should not be confused with passivity. Women struck, challenged managers, and resisted retrenchment. In reeling and winding, workers sometimes preferred to share a limited amount of work among more people rather than allow a smaller group to keep full work while others lost their jobs. That choice could protect employment for more workers and still lower each worker's daily earnings. It is a clear example of how worker resistance could exist together with weak bargaining power. The outcome was shaped by constrained choices rather than by women simply accepting whatever management decided.

Limits of the Evidence

Several limits prevent a stronger causal claim. First, daily earnings are not hourly wage rates. Many female reelers and winders worked shorter days in 1934, so part of a daily earnings difference may reflect fewer hours. Second, the age categories were not identical for boys and girls: older teenage boys could be placed in a separate 'big lads' category, while girls above fourteen could be counted as women. Third, piece earnings depend on more than the piece price. Yarn quality, machine speed, product type, work assigned, experience, and individual output could all matter. The TextileLab database also does not systematically identify permanent workers and badlis, job tenure, or individual productivity. These missing details matter when two group averages differ.

The surveys create further limits. The 1933 inquiry is too narrow to represent the entire Bombay workforce. Many 1926 worker-count entries are combined summary codes rather than usable numbers. The 1937 inquiry did not tabulate every occupation, used different collection methods for time and piece workers, and mainly reports day-shift wages. The money values are also nominal. They show the rupees, annas, and pies recorded at the time, but not what those earnings could buy. A wage ratio can, then, measure inequality in recorded earnings without directly measuring living standards.

These limitations do not, however, remove the value of the evidence. They define the conclusion that is justified. The TextileLab data show that the large overall gender earnings gap was closely connected to the mill's occupational structure. They show that women were more exposed to piece payment and that some winding jobs had large female disadvantages even within the same reported title. The official enquiries suggest a mechanism in which piece

payment interacted with irregular work and excess labour in female departments. The evidence cannot fully separate hours from wage rates, measure individual productivity, or prove employer intent. It is strongest when used to describe the structure of inequality and then read together with historical evidence about how work was organised.

Conclusion

The gender earnings gap in Bombay's cotton mills between 1921 and 1937 was not, in fact, one single gap with one single cause. In the detailed TextileLab observations, women as a group earned only a little over half as much per day as men in the four headline years. But women and men were usually working in different occupations. When the comparison is limited to large groups with the same original job title and payment system, the gap becomes much smaller. Time-rated doffers and ring siders are the clearest examples: in 1934 and 1937, female and male averages were often within only a few percentage points of each other. Occupational segregation is, as such, central to understanding the broad wage gap. A large part of what appears as a male-female earnings difference at the industry level is also a difference between the jobs in which men and women worked.

Piece payment mattered, but not because piece work always paid less. Women were far more likely to be piece-rated because they were concentrated in winding and reeling. In those departments, daily earnings depended on piece prices, but also on hours, machinery, materials, output, and whether enough work was available. The official committee's warning that excess staffing made actual reeling and winding earnings a poor measure of a full day's earning capacity is especially important. It gives a concrete reason why low female daily earnings could arise inside a piece-rate system without one rule paying every woman a fixed percentage less than a man doing the same named job.

The best explanation, then, combines gender with the organisation of production and worker bargaining power. Millowners faced uncertain markets and strong incentives to control costs. They used flexible employment, different wage systems, changing workloads, and rationalisation. Workers organised and resisted, so management did not have unlimited power, but workers did not bargain from equal positions. Gender affected which jobs people could enter, whether they could work nights, how easily they could move into better occupations, and whether they were concentrated in departments with irregular work. Winding remains an important case where a substantial within-title gap survived, and it should not be dismissed. At the same time, the official evidence requires caution because job titles did not guarantee identical hours, output, machinery, or access to work. The strongest defensible conclusion is that occupational segregation was a major structure through which gender inequality appeared in Bombay mill earnings, while piece rates increased insecurity in some heavily female departments rather than acting as a universal cause of low women's wages.

Works Cited

Chandavarkar, Rajnarayan. The Origins of Industrial Capitalism in India: Business Strategies and the Working Classes in Bombay, 1900-1940. Cambridge UP, 1994.

Dixit, Aditi, and Elise van Nederveen Meerkerk. "Supply of Labour during Early Industrialisation: Agricultural Systems, Textile Factory Work and Gender in Japan and India, ca. 1880-1940." Indian Economic and Social History Review, vol. 59, no. 2, 2022, pp. 223-255. https://doi.org/10.1177/00194646221083806.

Government of Bombay. Report of the Textile Labour Inquiry Committee, 1937. Vol. II: Final Report. Government Central Press, 1940.

Kumar, Radha. "Family and Factory: Women in the Bombay Cotton Textile Industry, 1919-1939." Indian Economic and Social History Review, vol. 20, no. 1, 1983, pp. 81-96.

Labour Office, Government of Bombay. General Wage Census, Part I—Perennial Factories, Third Report: Report on Wages, Hours of Work and Conditions of Employment in the Textile Industries (Cotton, Silk, Wool and Hosiery) in the Bombay Presidency (excluding Sind), May 1934. Government Central Press, 1937.

Morris, Morris David. The Emergence of an Industrial Labor Force in India: A Study of the Bombay Cotton Mills, 1854-1947. U of California P, 1965.

van Nederveen Meerkerk, Elise, and Aditi Dixit. "Human Beings Are Too Cheap in India: Wages and Work Organization as Business Strategies in Bombay's Late Colonial Textile Industry." International Review of Social History, vol. 70, suppl. S33, 2025, pp. 161-192. https://doi.org/10.1017/S0020859024000579.

Van Nederveen Meerkerk, Elise, and Aditi Dixit. Manual Database Textile Wages_India. TextileLab, Feb. 2024.

Van Nederveen Meerkerk, Elise, and Aditi Dixit. TextileLab India Textile-Wages Database. TextileLab, 2024.

The Impact of Artificial Intelligence on Labour Markets By Arya Khanna

Abstract

With its extensive use across nearly every sector of the global economy, artificial intelligence (AI) is revolutionizing the labour market. Companies are racing to adopt LLMs into the workforce, drastically altering existing labour due to job creation and displacement. By assessing how recent technological developments affect employment trends, wage dynamics and qualification requirements, this paper researches the economic impact of AI on job markets. The data shows that while AI raises productivity and creates employment in high-skill occupations, it poses a serious risk to workers performing repetitive and routine tasks, particularly in low-skill industries. The paper concludes that the impact of AI's and will depend less on the technology itself than on how quickly workers, firms and governments adapt through skills investment and organizational redesign.

1. Methodology

This paper uses secondary-research methodology. The research process was divided into four stages. First, a literature review was conducted to identify credible sources on AI and labour markets, including sources such as; the International Monetary Fund (IMF), the Organisation for Economic Co-operation and Development (OECD), the International Labour Organization (ILO), (Goldman Sachs and McKinsey & Company, the Penn Wharton Budget Model, the Budget Lab at Yale, and Microsoft's New Future of Work Initiative. Each source was assessed for reliability by considering its institutional affiliation, methodology, sample size where reported, and date of publication. Secondly, employment trends, skill requirements and displacement patterns were compared across sectors and economies using estimates provided by each source. Third, the findings were evaluated together to identify points of commonality and disagreement between sources, since predictions of AI's labour-market impact vary considerably. This design has limitations. AI capability is changing rapidly, so quantitative projections may become outdated. Also, international labour statistics are not always directly comparable across countries and task-level exposure to AI does not automatically translate into job loss, since firms and workers can adapt. Finally, a miniature AI model was constructed with the purpose of identifying jobs with exposure. The model would base the exposure score by identifying the creative level of each task.

2. Introduction

In November 2022, OpenAI released ChatGPT, marking the point at which AI shifted from specialist technology to a vital business tool. OpenAI demonstrated that generative AI could complete not only simple searches but also writing, coding, analysis and other subsequent tasks. As a result, firms accelerated adoption across customer service, finance, software development, marketing and administration. One of the main impacts of Artificial Intelligence is that it primarily affects the tasks inside jobs rather than entire occupations at once (Goldman

Sachs Research, 2023). As firms adopt AI to reduce costs and raise productivity, labour demand appears to be shifting away from routine, administrative and repetitive tasks and towards roles that require digital skills, problem
solving, creativity and human judgement.

3. Main Findings

3.1 Task-Level Exposure and Automation Risk

One key finding across multiple sources is that AI's effects are concentrated in specific types of tasks, rather than spread evenly across the workforce. Goldman Sachs Research estimates that the equivalent of 300 million full-time jobs globally are exposed to some degree of AI automation, with AI reducing work hours by one fourth in the US alone according to Figure 1 (Goldman Sachs Research, 2023; Briggs). Some of the roles with the highest task exposure are office and administrative support, legal services, architecture and engineering, while occupations requiring substantial physical or manual input such as construction, maintenance and cleaning are far less exposed. According to the Penn Wharton Budget Model, in 40 per cent of the occupations it analysed, at least half of current tasks could become replaceable, with office and administrative support (75 per cent), business and financial operations (68 per cent) and computing and mathematics (63 per cent) being the most exposed occupational groups (Arnon, 2025; Knowledge at What ton, 2025). Similarly, the ILO estimates that, across G7 countries, 6.5 per cent of jobs (around 25 million) are highly exposed to generative AI, while a further 28 per cent of employment (109 million jobs) is likely to be transformed (OECD/ILO, 2026).

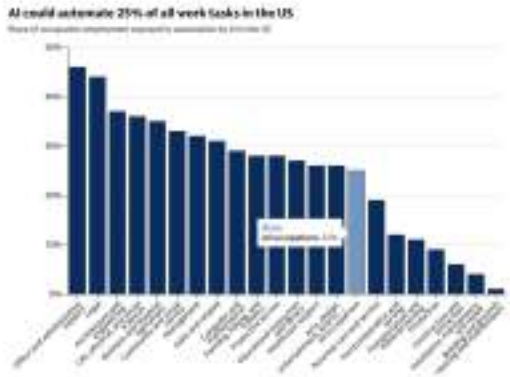


Figure 1: *Graph showing the percentage of automation in each (Goldman Sachs, 2026).*

This distribution of exposure is shown in Figure 2, which breaks down the entire US workforce by the share of job tasks that AI could perform. Only 1.1 per cent of employment sits in occupations where AI could complete 100 per cent of tasks, while a further 25.5 per cent falls into the 90 to 99 per cent range and 15.4 per cent into the 50 to 90 per cent range. This means that only 40 per cent of US employment is exposed to generative AI, while the remaining 58 per cent faces limited exposure (Arnon, 2025).

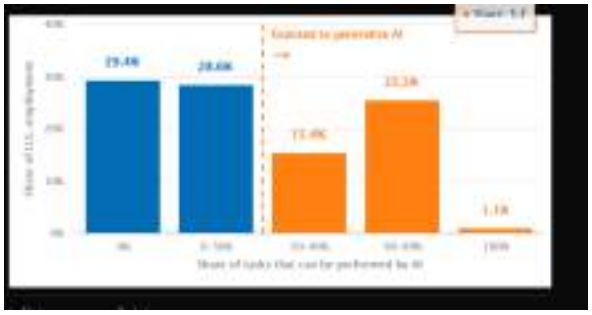


Figure 2: *Distribution of US employment by AI automation potential (Arnon, 2025).*

A common misconception is that "exposure" automatically results in job losses. However, in reality it refers to where task content is most likely to change (Tomlinson et al., 2025). High-exposure roles tend to involve analysis and structured decision-making while lower-exposure depends on physical dexterity that current AI systems cannot replicate.

Another common misconception is between automation and augmentation. Automation replaces a task outright, while augmentation helps a worker in completing it faster. While automation strategies remove cost from existing tasks, they cap the potential upside of AI adoption. On the other hand, strategies that involve augmentation and innovation can benefit substantially more (Brynjolf sson, 2022; Autor, 2022). This is important because it shows that the same technology can have very different outcomes depending on how firms choose to deploy it.

3.2 AI Capability Growth: The Accelerating Task Completion Horizon

One additional factor to consider is AI's rate of growth, as it influences the impact of exposure over the years gence systems are improving, reducing the time taken to complete longer, more complex tasks.

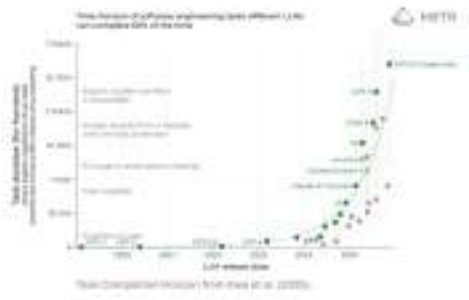


Figure 3: *METR*
task-completion horizon, reproduced in the Microsoft New Future of Work Report (2025, p. 6).

Secondly, while AI capabilities are developing quickly, companies typically take longer to redesign workflows, retrain employees and change organizational processes.
This helps to explain why the overall employment effects.

Researchers at METR have developed a benchmark that measures how long a human

would typically need to complete a software task. This is then measured against the length of time required by an AI model to complete the same task, with a 50% success rate. As Figure 3 shows, this “50 per cent task completion time horizon” has increased from just a few minutes for early models such as GPT 2 and GPT 3 to well over two hours for models released in 2025, with the rate of improvement is also increasing exponentially

Microsoft’s research estimates that the time AI takes to double its task completion ability is around seven months (Microsoft New Future of Work Report, 2025). This has two implications for the labour market analysis. First, it suggests that the task exposure estimates, most of which are based on AI capabilities from 2023 to 2024, may underestimate how many tasks could be automated in the next few years. This is because artificial intelli to date have been relatively small.

3.3 Productivity and Economic Growth

While AI may create risk of job displacement, it also can create economic benefit. The Penn Wharton Budget Model estimates that AI could increase productivity and GDP by around 1.5 per cent by 2035, rising to nearly 3 per cent by 2055 and 3.7 per cent by 2075. The largest increase in productivity growth is expected in the early 2030s before gradually becoming a smaller permanent effect (Arnon, 2025). Goldman Sachs Research estimates that AI could increase global GDP by around 7 per cent over ten years and raise US labour productivity by almost 1.5 percent per year during this period.

This is supported by Microsoft’s New Future of Work Report that found that AI can save individuals significant amounts of time. Small enterprises using ChatGPT reported saving around 40 to 60 minutes per day. However, these individual time savings do not necessarily lead to higher productivity across an entire organization, as AI may only be useful for certain tasks within a firm. This prediction is visible in Figure 4, as not all tasks benefit the same when using AI tools to assist.

McKinsey & Company estimates that this productivity boost from the use of generative AI could create an additional 4.4 trillion USD in revenue in the long run.

However, although 92% of companies plan to increase their AI investment, only 1% of companies have fully adopted AI (McKinsey & Company, 2025). This major gap is the primary reason why the effects on the labour market have been less than predictions.

This investment can already be seen in employment trends. Since 2022, jobs related to the construction of data centers have grown significantly. Utilities construction employment is up by 11.7 per cent, and electrical and HVAC contractors are both up by close to 8 per cent, compared with only 3.6 per cent growth for the overall economy (Goldman Sachs Research, cited in Yale Budget Lab data, 2026), as shown in Figure 5.

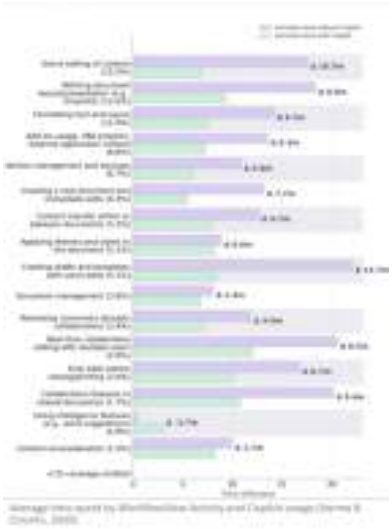


Figure 4: *Graph presenting the average time spent for tasks with and without Copilot usage (Verma & Counts, 2025, cited in Microsoft reports, 2025).*

3.4 Team Effects and the “Workslop” Problem

One interesting finding from Microsoft's 2025 New Future of Work Report is that AI is more effective in improving an individual's productivity, rather than team productivity (Schmutz et al., 2024, cited in Microsoft's New Future of Work Report, 2025). One laboratory study found that a single individual working with an AI assistant performed, on average, equally as well as a two person team without AI. This shows AI’s potential to replace some forms of collaboration, but also suggests that the effects of AI on team productivity are still uncertain (Dell’Acqua et al., 2025).

One explanation for this gap between individual and organisational productivity gains is due to a phenomenon researchers call "workshop". AI generated content that may seem useful actually contains incomplete, inaccurate or low quality information. This forces other employees to spend time in order to correct or redo the work. In a survey of 1150 US employees, 40% reported receiving “workslop” within the past month. “Workslop” is esti
mated to be 15% of all AI produced work (Microsoft New Future of Work Report, 2025, citing Niederhoffer et al., 2025). This is important for understanding the impact on

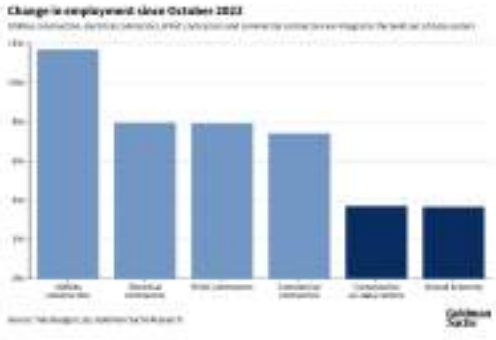


Figure 5: *Change in employment since October 2022 in construction roles responsible for building AI (Goldman Sachs Research, cited in Yale Budget Lab data, 2026).*

labor markets because artificial intelligence can generate the hidden forms of work. These hidden costs can cancel out productivity gains because employees are now required to do quality control verification tasks that did not exist before.

This evidence suggests that the productivity estimates discussed in Section 3 may fall short of the theoretical potential of AI. It also reinforces the distinction between automation and augmentation, as discussed in Section 1. Workslop creation of workslop may shift tasks to more review and quality control support based (Brynjolfsson, 2022, cited in Microsoft New Future of Work Report, 2025).

3.5 Global Divergence: Geography and Inequality

AI's impact on the labour market is not the same across countries. The Microsoft AI Economy Institute's Global AI Diffusion reports show that global AI adoption increased from around 16.3 per cent in the second half of 2025 to 17.8 per cent in the first quarter of 2026. In addition, 26 economies now have adoption rates above 30 per cent (Microsoft AI Economy Institute, 2026).

However, this growth has been uneven. AI adoption in the Global North has grown around twice as fast as in the Global South, increasing the gap between the two groups from around 9.8 to over 12 percent within a year (Microsoft AI Economy Institute, 2026), as shown in Figure 6. The IMF's research also highlights this difference, finding that one in ten job postings in advanced economies now requires at least one new skill, often related to AI, compared with one in twenty in emerging market economies (Georgieva, 2026).

This difference is important because advanced economies have more high skill and knowledge based jobs, making them more exposed to AI in the short term. However, they also tend to have stronger education systems, digital infrastructure and social support, which can help workers adapt. Lower income economies may face less immediate disruption because AI adoption is lower, but they may also have fewer resources to develop the skills needed to benefit from AI in the long term. This could increase economic inequality between countries over time.

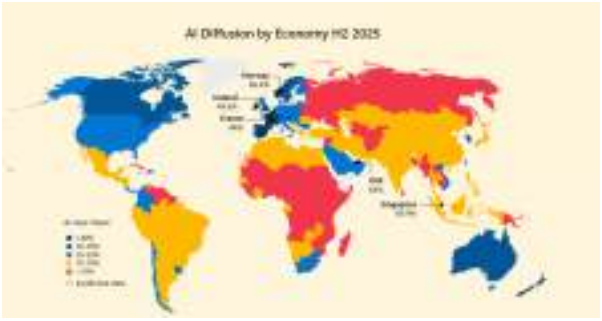


Figure 6: *World map showing global artificial intelligence diffusion and adoption rates by economy (Microsoft AI Economy Institute, 2026).*

Furthermore, language may also impact the effects of AI. Its adoption is growing fastest in countries where the main language is well supported by existing AI models and where digital infrastructure and education are stronger. In countries with less supported languages, users often switch to English when using AI, particularly in parts of Africa and Asia. This pattern is much less common in Europe and the Americas (Microsoft AI Economy Institute, 2025, cited in Microsoft New Future of Work Report, 2025). This language barrier is often overlooked in AI adoption statistics, but it may affect who can effectively use AI to improve their work.

Economist David Autor argues that AI could reduce skill gaps in knowledge intensive tasks, allowing more people to take on work that traditionally requires high levels of expertise, such as medical or legal decision support. This could lower the level of expertise needed to perform valuable tasks (Autor, 2024, cited in Microsoft New Future of Work Report, 2025).

However, this benefit is not guaranteed. Research suggests that AI can currently give greater advantages to already skilled users, particularly in open ended creative tasks. Addressing this gap will require improvements in AI technology as well as policies that ensure access to AI and its benefits are distributed more equally. Without these measures, differences in AI investment, access and benefits could contribute to greater inequality between workers and countries (Daepp et al., 2025; Microsoft AI Economy Institute, 2025, cited in Microsoft New Future of Work Report, 2025).

3.6 Policy Responses: Social Protection and Basic Income

Counteracting the costs of AI and ensuring productivity gains through AI adoption depends heavily on government policies. The IMF's Managing Director argues that the success of AI depends on investing in skills and governments should support workers through job changes and keeping markets competitive so that innovation benefits everyone (Georgieva, 2026).

This includes updating education systems to develop AI skills, creativity and judgement, which are harder to automate. Governments should also expand retraining programmes for workers whose jobs are affected by AI (OECD, 2025). One policy option

often discussed is universal basic income (UBI), which provides unconditional payment to all residents regardless of their employment status. Supporters argue that UBI could reduce poverty and provide income security for workers affected by AI driven unemployment (UNC-Chapel Hill, 2021). However, UBI would be difficult to implement on a large scale, and may result in voluntary unemployment (Britannica, 2026).

Overall, many sources supported a more targeted approach rather than introducing UBI. This includes investing in reskilling, strengthening unemployment support and continuing to monitor AI's actual effects on employment. This is particularly important for younger workers, who are more immediately exposed to AI. It is also consistent with research from the Budget Lab at Yale, which suggests that the scale of future disruption remains uncertain and that policies should be adjusted as more evidence becomes available.

3.7 A Mini Exposure Prediction Model: Testing AI's Own Assessment of AI Exposure

To verify the findings of the paper, a miniature AI model was to predict a job's exposure to AI automation. The model requires the user to answer questions about and returns a predicted exposure score from 0 to 100 per cent alongside a short explanation of the result. The aim was to test whether a simple set of inputs, of the kind an individual worker could report about themselves, could produce exposure predictions broadly consistent with the findings of Goldman Sachs, Wharton and the ILO discussed in Section 1.

The model was tested on ten illustrative job descriptions spanning a range of exposure levels reported in the literature. A field of questions about each job would be filled, as shown in Figure 7. Table 1 shows only those jobs the model predicted to have an exposure score of 70 per cent or above.



Figure 7: *Interface of the mini AI exposure prediction tool, showing the job characteristic inputs and an example prediction output (Author's own work, 2026).*

The output presented in Table 1 matches the pattern reported in the literature, since office and administrative support work is consistently identified as the occupational group with the highest task exposure, at around 75 per cent of tasks according to the Penn Wharton Budget Model (Arnon, 2025), and around 46 per cent according to Goldman Sachs Research (2023). None of the other job descriptions tested reached the 75 per cent threshold, largely because their descriptions referred to care, judgement, creativity or physical work, which reduced their scores.

Table 1: *Jobs and the risk of exposure.*

Job	Creativity	Technology	Effort	Predicted exposure	Risk
Data Entry Clerk	1	5	1	80.0%	High
Payroll Administrator	1	5	1	70.0%	High
Telemarketer	1	4	1	75.0%	High

4. Conclusion

AI is reshaping labour markets mainly by changing the tasks within jobs rather than removing entire occupations. Routine, text heavy, data heavy and predictable tasks face the greatest risk of automation, while roles requiring judgement, communication, creativity and human responsibility have been more resistant to displacement so far. Companies are slower to adopt AI than their investment plans suggest, and individual time savings do not always lead to higher productivity across organisations. This is partly because of quality control issues such as AI generated “workslop”. At the same time, AI systems are becoming capable of completing longer and more complex tasks, with independent research showing that this capability has been doubling roughly every seven months. This means that current estimates of AI’s impact on the labour market may underestimate its future reach.

Geography also affects these outcomes. Advanced economies are more exposed to AI in the short term, but they are also better positioned to benefit from it because of stronger education systems and digital infrastructure. Lower income countries and countries with less supported languages may face greater long term challenges if they cannot develop similar skills and infrastructure. Importantly, overall employment and unemployment data have not yet shown clear evidence of large scale AI driven disruption. This suggests that major changes to employment are likely to happen gradually rather than as a sudden shock.

The main conclusion is that AI will not affect all workers equally. The future of work will depend on whether workers, companies and governments treat AI as a tool that supports workers through investment in skills, worker involvement and social protection, rather than simply as a replacement for labour. Labour markets that are best prepared will be those that help people, especially those entering the workforce, learn to work alongside AI rather than compete against it.

Works Cited

[1] Arnon, A. (2025). The Projected Impact of Generative AI on Future Productivity Growth. Penn Wharton Budget Model, 8 September. https://budgetmodel.wharton.up enn.edu/issues/2025/9/8/projected-impact-of-generat ive-ai-on-future-productivity-growth

[2] Britannica (2026). Universal Basic Income (UBI): Pros, Cons, Debate, Arguments, & Income Equality. https: //www.britannica.com/procon/universal-basic-incom e-UBI-debate

[3] Budget Lab at Yale (2026). Tracking the Impact of AI on the Labor Market. https://budgetlab.yale.edu/researc h/tracking-impact-ai-labor-market

[4] Butler, J., Jaffe, S., Janßen, R., Baym, N., Hecht, B., Hofman, J., Rintel, S., Sarrafzadeh, B., Sellen, A., Vor voreanu, M. and Teevan, J. (eds.) (2025). Microsoft New Future of Work Report 2025. Microsoft Research Tech Report MSR-TR-2025-58. https://www.microsoft.com/ en-us/research/wp-content/uploads/2025/12/New-Fut ure-Of-Work-Report-2025.pdf

[5] Georgieva, K. (2026). New Skills and AI Are Reshaping the Future of Work. IMF Blog, 14 January. https://www. imf.org/en/blogs/articles/2026/01/14/new-skills-and-a i-are-reshaping-the-future-of-work

[6] Goldman Sachs Research (2023). How Will AI Affect the US Labor Market? https://www.goldmansachs.com/ins ights/articles/how-will-ai-affect-the-us-labor-market

[7] Knowledge at Wharton (2025). How AI Could Lift Productivity and GDP Growth. https://knowledge.wharton. upenn.edu/article/how-ai-could-lift-productivity-and-g dp-growth/

[8] Kwa, T., West, A., Snyder, J., Ivanov, N., Toit, C., Duffy, M., Rush, R., Blum, D., Yu, J., Wang, Y., Rein, D., Chan, A., Schmidt, L., Nunes, P., Ho, L., Vinayak, T., Hunter, K., McAleer, S. and Barnes, E. (2025). Measuring AI Ability to Complete Long Tasks. METR. As reproduced in Butler, J. et al. (eds.), Microsoft New Future of Work Report 2025, p. 6. https://www.microsoft.com/en-us/ research/wp-content/uploads/2025/12/New-Future-O f-Work-Report-2025.pdf

[9] Marcus by Goldman Sachs (2026). How Will AI Affect the US Labor Market? https://www.marcus.com/us/en /resources/heard-at-gs/how-will-ai-affect-the-us-labor -market

[10] McKinsey & Company (2025). Superagency in the Work place: Empowering People to Unlock AI's Full Potential at Work. https://www.mckinsey.com/capabilities/tech-a nd-ai/our-insights/superagency-in-the-workplace-emp owering-people-to-unlock-ais-full-potential-at-work

[11] Microsoft AI Economy Institute (2026). The State of Global AI Diffusion in 2026. https://blogs.microsoft.co m/on-the-issues/2026/05/07/the-state-of-global-ai-dif fusion-in-2026/

[12] Microsoft AI Economy Institute (2026). Global AI Adoption in 2025 – A Widening Digital

Divide. https://www. microsoft.com/en-us/corporate-responsibility/topics/ai-economy-institute/reports/global-ai-adoption-2025/
[13] OECD (2025). AI and the Future of Social Protection in OECD Countries. OECD Artificial Intelligence Papers, No. 42. Paris: OECD Publishing. https://www.oecd.org /content/dam/oecd/en/publications/reports/2025/06/ ai-and-the-future-of-social-protection-in-oecd-countri es_038f49ed/7b245f7e-en.pdf
Artificial Intelligence and Labour Markets Arya Khanna
[14] OECD (2025). Harnessing Artificial Intelligence in Social Security: Use Cases, Governance and Workforce Readiness. OECD Digital Government Studies. Paris: OECD Publishing. https://www.oecd.org/en/publications/harnessing-artificial-intelligence-in-social-security_b52405 c1-en.html
[15] OECD/ILO (2026). Compendium of Best Practices for a Human-Centered Development and Use of Artificial Intelligence in the World of Work. Geneva/Paris: ILO/OECD Publishing. https://www.ilo.org/sites/default/files/2026-01/Compendium%20of%20best%20practices%20for%20a%20human-centered%20development%20and%20use%20of%20AI%20in%20the%20world%20of%20work _1.pdf
[16] UNC-Chapel Hill (2021). The Pros and Cons of Universal Basic Income. https://www.unc.edu/discover/the-pro s-and-cons-of-universal-basic-income

Has PMJDY Improved Financial Inclusion? Evidence from Banks in India

By Ryan Bathwal

Abstract

Financial inclusion has now moved beyond only access towards actual use. This is important because having a bank account does not automatically mean that the person uses it regularly or meaningfully. India has done well in terms of access, as account ownership has increased strongly after 2014. But the outcome on usage is still not fully clear, as earlier studies and Global Findex data show that India has had a high share of inactive accounts. Therefore, this paper studies PMJDY not only as an account-opening scheme, but also as a process of account usage and account deepening.

The Pradhan Mantri Jan-Dhan Yojana, or PMJDY, is a flagship Government of India financial inclusion scheme launched in August 2014. Official data now show more than 58 crore PMJDY beneficiaries and deposits of more than ₹3 trillion. This paper uses PMJDY data and RBI DBIE bank branch data to study public sector banks, regional rural banks, and private banks. It first examines total beneficiaries, deposits, deposits per beneficiary, and deposits per bank. The branch data are taken from RBI's Database on Indian Economy, specifically the bank-wise and population group-wise number of functioning offices of commercial banks dataset (Reserve Bank of India). It then uses regression analysis to study whether PMJDY deposits are associated with beneficiaries, RuPay cards, and rural and urban branch presence. The results suggest that PMJDY deposits are positively associated with beneficiaries, RuPay cards, and rural branches, while urban branches show a negative association in the model. Overall, the paper studies PMJDY through account usage and account deepening, instead of only account ownership.

Introduction

Financial inclusion can be defined as the ownership and use of formal financial accounts, such as accounts in banks or other regulated financial institutions (Allen et al.). The concept has evolved over time. Earlier, financial inclusion was mainly understood as access to banks and basic account ownership. More recent research, however, emphasizes the actual use of accounts for savings, payments, borrowing, insurance, and digital transactions. This is important because financial inclusion helps people save and receive money in a safer and faster way, borrow from formal sources, and handle financial difficulties without depending on informal lenders. Globally, the World Bank's Global Findex shows major progress: adult account ownership increased from 51% in 2011 to 79% in 2024 (Klapper et al.). In India too, progress has been impressive. Account ownership increased from 35% of adults in 2011 to 89% in 2024, with much of the increase occurring after 2014 (Klapper et al.). One study similarly notes that adult account ownership went up from 35% in 2011 to 80% in 2017 (Prasuna et al.).

This increase was supported by the Pradhan Mantri Jan-Dhan Yojana (PMJDY), a National Mission for Financial Inclusion launched on 28 August 2014. Its main aim is to provide unbanked people with access to a basic bank account and other financial services at an affordable

cost. As of May 2026, official PMJDY data showed 58.18 crore beneficiaries, 40.48 crore RuPay debit cards, and deposits of more than ₹3 lakh crore (Pradhan Mantri Jan-Dhan Yojana). However, many studies also point out that just opening accounts is not sufficient, because accounts which are hardly used may not create much real impact in terms of financial inclusion (Singh et al.). This concern is also visible in Global Findex data, which reports that 16% of account owners in India did not have an active account in 2024 (Klapper et al.).

Keeping this in mind, this paper does not look at financial inclusion only through account opening. Instead, it studies PMJDY through indicators such as total beneficiaries, total deposits, deposits per beneficiary, deposits per bank, RuPay debit cards, and branch presence. The paper first describes how PMJDY has grown across public sector banks, regional rural banks, and major private banks. It then uses regression analysis to examine whether PMJDY deposits are associated with beneficiary expansion, RuPay card issuance, and rural and urban branch networks. The focus is therefore on understanding the relationship between PMJDY deposits and these access-related factors, rather than proving causation. This is important because earlier research suggests that account usage may increase gradually as people become more familiar with bank accounts, digital identity, RuPay cards, and payment systems over time (Agarwal et al.; Chopra et al.; Khandelwal et al.). While access has become widespread, Global Findex reports that 42.7% of adults in India made or received a digital payment in 2024, and rural account owners were still less likely than urban account owners to use digital payments (Klapper et al.). Therefore, studying deposits and related indicators can help show whether PMJDY has moved beyond account opening toward deeper and more meaningful financial inclusion.

Literature Review

Burgess and Pande studied whether the expansion of rural bank branches in India helped in reducing rural poverty. They used annual state-level data for 16 major Indian states from 1961 to 2000. Their method used the RBI's 1:4 branch licensing rule, where banks had to open branches in unbanked areas before opening more branches in already banked areas. The authors found that rural branch expansion reduced rural poverty, mainly through more rural credit and savings.

Demirgüç-Kunt and Klapper studied how financial inclusion can be measured across different countries. They used the first Global Findex Database, based on 2011 survey data. The authors found that having a bank account is only one part of financial inclusion, because actual use of accounts is very different across people and countries. Similarly, Allen et al. examined why people own and use formal bank accounts. The authors found that account ownership and usage were higher when banking was cheaper and financial service points were closer. They also found that low-fee accounts, easier documentation, and government payments through accounts could help poor and rural populations.

Prina studied whether poor households would open and actually use a simple savings account if access was made easier. The study used a field experiment in Nepal. The author found strong demand for basic accounts, with many households opening them, saving small amounts

regularly, and using their savings for health, food, and emergencies. Kochar examined whether India's branchless banking model, using Business Correspondents, improved financial inclusion in rural areas. The author found that Business Correspondent access significantly increased household savings and savings stored in bank accounts.

Agarwal et al. studied whether PMJDY only opened many accounts or also increased real account usage. They used administrative account-level data from the State Bank of India. The authors found that account opening was very high, but initial usage was low. However, account usage improved after six to twelve months. Dupas et al. studied whether simply giving unbanked people access to basic bank accounts led to real account use in Uganda, Malawi, and Chile. The authors found that while account opening increased, active usage still remained limited.

Singh et al. studied whether financial inclusion and PMJDY were linked with economic growth across Indian states. The authors found that financial inclusion was positively linked with economic growth, but PMJDY's effect was limited due to poor usage and dormant accounts. Khera et al. created new indexes for digital financial inclusion and found that digital financial services became an important driver of financial inclusion, especially in Africa and Asia-Pacific.

Demirgüç-Kunt et al. studied global access to formal financial services during the COVID-19 period. The report found that financial inclusion had grown strongly worldwide, but stressed that account ownership is only the first step. Klapper et al. studied financial inclusion in India using the Global Findex 2021 country brief, finding that India had achieved high account ownership, but account usage remained a major challenge. Ansar et al. also found that account ownership is not enough, linking high account inactivity in India partly to the large number of Jan Dhan accounts.

Agarwala et al. studied how efficiently Indian banks implemented PMJDY. The authors found that public sector banks performed better than private banks in promoting PMJDY-based financial inclusion. Singh and Prasad examined how PMJDY account holders adopted banking products after opening accounts, finding that social influence was highly important for fund transfers and ATM use. Prasuna et al. studied the level and causes of financial inclusion across Indian states, noting that financial inclusion improved over time, but big differences still remained across states.

Khandelwal et al. examined whether financial literacy was linked with financial inclusion in India after demonetization. They found a strong positive relationship between financial literacy and account ownership. Srivastava et al. found that PMJDY-led financial inclusion improved bank stability, especially through rural beneficiaries. Finally, Chopra et al. examined how India's digital public infrastructure, especially Aadhaar and UPI, affected financial inclusion, finding that these tools had a positive long-run effect.

Overall, the existing literature shows that financial inclusion cannot be understood only by counting how many bank accounts have been opened. Access is important, but actual usage, savings, digital payments, and account activity are also needed. This paper tries to fill a gap in the literature by studying PMJDY at the bank-wise level, focusing on account deepening through indicators such as deposits per beneficiary, total deposits, RuPay cards issued, branch presence,

and differences between bank types.

Data and Descriptive Statistics

This section uses annual PMJDY bank-type data from FY2014-15 to FY2025-26, based on the official PMJDY progress-report data (Pradhan Mantri Jan-Dhan Yojana). The data are divided into three bank categories: major private banks, public sector banks, and regional rural banks (RRBs). At the all-India level, PMJDY beneficiaries increased from about 145.4 million in FY2014-15 to about 583.5 million in FY2025-26, meaning the number of beneficiaries grew to nearly 4.0 times its FY2014-15 level. Total deposits increased from about ₹146.41 billion to about ₹3.07 trillion, or around 21.0 times the FY2014-15 level.

The four indicators used in this section are total beneficiaries, total deposits, deposits per beneficiary, and deposits per bank. These indicators give a basic descriptive picture of both expansion and possible account usage. The analysis in this section is only descriptive; regression analysis is presented in Section 4.

Beneficiary Expansion

Total beneficiaries by year and bank type (millions)

Financial year	Major Private Banks	Public Sector Banks	Regional Rural Banks
FY2014-15	6.04	114.01	25.32
FY2015-16	7.89	168.46	37.93
FY2016-17	9.15	226.15	46.38
FY2017-18	9.90	264.29	54.01
FY2018-19	12.41	281.25	59.92
FY2019-20	12.57	305.19	65.53
FY2020-21	12.53	333.90	75.57
FY2021-22	13.02	355.05	82.55
FY2022-23	14.06	382.16	90.48
FY2023-24	15.94	404.98	96.67
FY2024-25	17.95	428.45	103.52
FY2025-26	20.43	448.02	115.05

Public sector banks clearly lead in absolute numbers during the whole period. By FY2025-26, public sector banks had about 448.0 million beneficiaries, compared with about 115.0 million for RRBs and 20.4 million for major private banks. However, in terms of growth from the starting year, RRBs grew the fastest. Their beneficiary count grew to about 4.5 times its starting level, compared with about 3.9 times for public sector banks and 3.4 times for private banks.

Total Deposits

Deposits by year and bank type (₹ billion)

Financial year	Major Private Banks	Public Sector Banks	Regional Rural Banks

FY2014-15	8.68	114.54	23.19
FY2015-16	13.55	281.39	61.78
FY2016-17	20.98	492.66	116.09
FY2017-18	21.76	655.35	143.93
FY2018-19	28.20	769.31	165.94
FY2019-20	31.83	939.20	213.32
FY2020-21	42.83	1,131.46	281.22
FY2021-22	47.58	1,284.13	332.88
FY2022-23	55.81	1,544.61	388.32
FY2023-24	68.69	1,805.19	451.14
FY2024-25	74.09	2,016.26	513.52
FY2025-26	88.35	2,401.32	584.88

Public sector banks dominate in absolute terms. In FY2025-26, deposits with public sector banks were about ₹2.40 trillion, compared with about ₹584.88 billion for RRBs and ₹88.35 billion for major private banks. The growth rates show that RRB deposits increased the most proportionally over the full period, growing to about 25.2 times their starting level, compared with 21.0 times for public sector banks and 10.2 times for private banks.

Year-on-year growth rates by bank type (%)

Year	Beneficiaries : Private	Beneficiaries : Public	Beneficiaries : RRBs	Deposits: Private	Deposits: Public	Deposits: RRBs
FY2015-16	30.6	47.8	49.8	56.2	145.7	166.4
FY2016-17	15.9	34.2	22.3	54.9	75.1	87.9
FY2017-18	8.3	16.9	16.4	3.7	33.0	24.0
FY2018-19	25.3	6.4	10.9	29.6	17.4	15.3
FY2019-20	1.3	8.5	9.4	12.9	22.1	28.5
FY2020-21	-0.2	9.4	15.3	34.6	20.5	31.8
FY2021-22	3.8	6.3	9.2	11.1	13.5	18.4
FY2022-23	8.0	7.6	9.6	17.3	20.3	16.7
FY2023-24	13.3	6.0	6.8	23.1	16.9	16.2
FY2024-25	12.6	5.8	7.1	7.9	11.7	13.8
FY2025-26	13.8	4.6	11.1	19.3	19.1	13.9

Deposits per Beneficiary

Deposits per beneficiary by year and bank type (₹)

Financial year	Major Private Banks	Public Sector Banks	Regional Rural Banks
FY2014-15	1,436	1,005	916
FY2015-16	1,717	1,670	1,629
FY2016-17	2,294	2,178	2,503
FY2017-18	2,197	2,480	2,665
FY2018-19	2,273	2,735	2,770
FY2019-20	2,533	3,077	3,255
FY2020-21	3,417	3,389	3,721
FY2021-22	3,656	3,617	4,032
FY2022-23	3,969	4,042	4,292
FY2023-24	4,310	4,458	4,667
FY2024-25	4,127	4,706	4,961
FY2025-26	4,324	5,360	5,084

This indicator adjusts deposits for the number of account holders, giving a rough idea of account activity or depth. Deposits per beneficiary increased over time for all three bank categories. In FY2014-15, the values were about ₹1,436 for private banks, ₹1,005 for public sector banks, and ₹916 for RRBs. By FY2025-26, they had increased to about ₹4,324, ₹5,360, and ₹5,084, respectively. This suggests that deposits grew faster than the number of beneficiaries, which may indicate deeper use of PMJDY accounts over time.

Deposits per Bank

Deposits per bank by year and bank type (₹ billion)

Financial year	Major Private Banks	Public Sector Banks	Regional Rural Banks
FY2014-15	0.67	4.24	1.05
FY2015-16	1.04	10.42	2.81
FY2016-17	1.61	18.25	5.28
FY2017-18	1.67	24.27	6.54
FY2018-19	2.01	36.63	9.22
FY2019-20	2.27	52.18	14.22
FY2020-21	3.06	94.29	23.44
FY2021-22	3.40	107.01	27.74
FY2022-23	3.99	128.72	32.36
FY2023-24	5.28	150.43	37.59
FY2024-25	5.70	168.02	42.79
FY2025-26	6.80	200.11	48.74

This indicator divides total deposits by the number of banks in each category, showing how concentrated PMJDY deposits are. In FY2025-26, deposits per public sector bank were about ₹200.11 billion, compared with about ₹48.74 billion for RRBs and ₹6.80 billion for major

private banks. While public sector banks lead in level, RRBs show strong proportional improvement.

Regression Results and Analysis

This section uses regression analysis to study whether PMJDY deposits are associated with beneficiaries, RuPay cards, and branch presence. For the regression analysis, this section uses quarterly bank-level PMJDY data, while the previous section used annual bank-type data for descriptive analysis. It is important to note that this regression does not prove causation; it only shows association. The results should be understood as relationships in the data, not as proof that one variable directly causes another variable.

The dependent variable in this regression is PMJDY deposits, which serve as an aggregate indicator of account use rather than direct proof of account depth. Although deposits per beneficiary could be a useful metric for studying account depth, this study does not run a separate regression using it as the dependent variable. Because deposit data are only available at the aggregate level and not geographically disaggregated, regressing aggregate deposits per beneficiary against geographically disaggregated branch variables may not make complete economic or statistical sense. The independent variables are the factors which may be linked with PMJDY deposits: rural and semi-urban beneficiaries, urban and metro beneficiaries, RuPay debit cards issued, rural branches combined, and urban branches combined. Rural branches combined means rural and semi-urban branches taken together, and urban branches combined means urban and metropolitan branches taken together.

The regression is a log-log model, meaning logarithms are used to improve normality in the data, and the coefficients are interpreted as percentage associations. For example, a coefficient of 0.391 means that a 1% increase is associated with about a 0.39% increase in deposits, not a 39% increase. The method used is Ordinary Least Squares (OLS) regression, which tries to find the best-fitting line for the data by minimizing the overall error between actual deposits and the model's predicted values. The main hypothesis is that PMJDY deposits are associated with beneficiary expansion, RuPay card issuance, and branch presence, while the null hypothesis assumes no relationship. Beneficiaries, RuPay cards, and branches (especially rural/semi-urban) are expected to have a positive relationship with deposits due to their links with usage and financial inclusion.

The regression equation postulates that PMJDY deposits are explained by the independent variables:

$$\begin{aligned}\text{Log PMJDY Deposits} = {} & \beta_0 + \beta_1 \text{ Log Rural/Semi-urban Beneficiaries} \\ & + \beta_2 \text{ Log Urban/Metro Beneficiaries} \\ & + \beta_3 \text{ Log RuPay Cards} \\ & + \beta_4 \text{ Log Rural Branches} \\ & + \beta_5 \text{ Log Urban Branches}\end{aligned}$$

+ error term

The estimated equation is:

Log PMJDY Deposits =
-4.168
+ 0.391 Log Rural/Semi-urban Beneficiaries
+ 0.448 Log Urban/Metro Beneficiaries
+ 0.171 Log RuPay Cards
+ 0.837 Log Rural Branches
- 0.710 Log Urban Branches

Table 4.1: Regression Results

Variable	Coefficient	Significance
Intercept	-4.168	Significant
Log rural/semi-urban beneficiaries	0.391	Significant
Log urban/metro beneficiaries	0.448	Significant
Log RuPay debit cards issued	0.171	Significant
Log rural branches combined	0.837	Significant
Log urban branches combined	-0.710	Significant

Observations: 1,203; **R-squared:** 0.910; **Adjusted R-squared:** 0.909

Interpretation of Results

The R-squared value of 0.910 indicates that about 91% of the variation in PMJDY deposits is explained by the variables in the model, though other factors like income, government transfers, local economic conditions, financial literacy, and actual account activity are not included. All variables are statistically significant, suggesting that the relationships shown in the regression are unlikely to be only random. A positive coefficient signifies an association with higher deposits, while a negative one indicates lower deposits when controlling for other variables.

Specifically, a 1% increase in rural/semi-urban beneficiaries is associated with a 0.39% increase in PMJDY deposits, while a 1% increase in urban/metro beneficiaries is associated with a 0.45% increase. RuPay debit cards also show a positive relationship, with a 1% increase linked to a 0.17% rise in deposits; however, this is a smaller effect compared to beneficiary and branch

metrics, suggesting deposits are more strongly linked with those factors. Rural and semi-urban branch presence demonstrates the strongest positive association, where a 1% increase corresponds to a 0.84% rise in deposits, making sense given PMJDY's focus on financial inclusion in areas with lower banking access. Conversely, the coefficient for urban branches combined is negative (-0.710), a result that should be interpreted carefully. It does not imply that opening urban branches causes deposits to fall, but rather that after controlling for other factors, urban branch presence is associated with lower PMJDY deposits. A possible reason is that urban branches may serve a broader mix of non-PMJDY customers, such as salaried people, business customers, and regular savings account holders, meaning more urban branches may not strongly link with PMJDY deposits. Overall, the regression results suggest that PMJDY deposits are strongly associated with beneficiaries, RuPay cards, and branches.

Conclusion and Policy Implications

This study aimed to understand how PMJDY deposits, serving as an indicator of account usage, are associated with beneficiary expansion, RuPay card issuance, and bank branch networks. The results show clear relationships between these factors. Specifically, the numbers of rural and semi-urban beneficiaries, urban and metro beneficiaries, and RuPay debit cards, along with rural branch presence, are all positively associated with higher PMJDY deposits. This suggests that beneficiary expansion and card issuance are associated with stronger account use and may support account deepening. Conversely, urban branch presence was negatively associated with PMJDY deposits in this model, likely because urban branches serve a much broader, mixed base of non-PMJDY customers.

These findings offer important policy implications for banks and policymakers. The strong positive relationship between rural branches and deposits highlights the continued importance of physical rural branch networks. Even as banking technology advances, maintaining easy access to physical bank branches in rural and semi-urban areas remains crucial for encouraging newly banked populations to actively save and use their accounts. Furthermore, policymakers should recognize the changing role of debit cards. Although RuPay cards are positively linked to deposits, their relative impact is smaller than that of branch expansion. This likely reflects how digital payments have evolved far beyond only physical RuPay cards, with mobile banking and UPI becoming more important. Moving forward, promoting digital financial literacy should go hand-in-hand with traditional card issuance to ensure accounts remain active.

Finally, this research has a few limitations that should be noted. First, the study focuses on major commercial and rural banks but does not include small finance banks or microfinance institutions, which are increasingly important players in local financial inclusion. Second, the deposit data used are only available at the aggregate level and are not geographically disaggregated, meaning regional branch data cannot be perfectly matched with regional deposit amounts. Despite these constraints, the public availability of official PMJDY data is useful. This ongoing transparency provides a useful foundation for future research on financial inclusion policy.

Works Cited

Agarwal, S., et al. Banking the Unbanked: What Do 280 Million New Bank Accounts Reveal about Financial Access? Columbia Business School Research Paper no. 17-12, 2023, https://doi.org/10.2139/ssrn.2906523.

Agarwala, N., et al. "Efficiency of Indian Banks in Fostering Financial Inclusion: An Emerging Economy Perspective." Journal of Financial Services Marketing, vol. 29, no. 2, 2024, pp. 341–353.

Allen, F., et al. The Foundations of Financial Inclusion: Understanding Ownership and Use of Formal Accounts. Policy Research Working Paper no. 6290, World Bank, 2012.

Ansar, S., et al. The Importance of Financial Education for the Effective Use of Formal Financial Services. Policy Research Working Paper no. 10345, World Bank, 2023.

Burgess, R., and R. Pande. "Do Rural Banks Matter? Evidence from the Indian Social Banking Experiment." American Economic Review, vol. 95, no. 3, 2005, pp. 780–795.

Chopra, K., et al. "Digital Public Infrastructure as a Driver of Financial Inclusion in India." Discover Sustainability, vol. 7, 2026, article 155.

Demirgüç-Kunt, A., and L. Klapper. Measuring Financial Inclusion: The Global Findex Database. Policy Research Working Paper no. 6025, World Bank, 2012.

Demirgüç-Kunt, A., et al. The Global Findex Database 2021: Financial Inclusion, Digital Payments, and Resilience in the Age of COVID-19. World Bank, 2022.

Dupas, P., et al. "Banking the Unbanked? Evidence from Three Countries." American Economic Journal: Applied Economics, vol. 10, no. 2, 2018, pp. 257–297.

Khandelwal, A., et al. "Financial Literacy and Inclusion in India: Evidence from Household-Level Data after Demonetization." Journal of Emerging Market Finance, vol. 24, no. 3, 2025, pp. 331–359.

Khera, P., et al. Measuring Digital Financial Inclusion in Emerging Market and Developing Economies: A New Index. IMF Working Paper WP/21/90, International Monetary Fund, 2021.

Klapper, L., et al. The Global Findex Database 2021: India Country Brief. World Bank, 2022.

Klapper, L., et al. The Global Findex Database 2025: Connectivity and Financial Inclusion in the Digital Economy. World Bank, 2025, https://doi.org/10.1596/978-1-4648-2204-9.

Kochar, A. Branchless Banking: Evaluating the Doorstep Delivery of Financial Services in Rural India. Working Paper no. 566, Stanford Center for International Development, 2016.

Pradhan Mantri Jan-Dhan Yojana. "Progress Report." Department of Financial Services, Ministry of Finance, Government of India, 2026, https://www.pmjdy.gov.in/account. Accessed May 2026.

Prasuna, A., et al. "Unveiling the Factors Influencing Financial Inclusion in India: A Comprehensive Analysis." Cogent Economics & Finance, vol. 12, no. 1, 2024, article 2381682.

Prina, S. "Banking the Poor via Savings Accounts: Evidence from a Field Experiment." Journal of Development Economics, vol. 115, 2015, pp. 16–31.

Reserve Bank of India. “Bank-Wise and Population Group-Wise Number of Functioning Offices of Commercial Banks.” Database on Indian Economy, 2026, https://data.rbi.org.in/. Accessed July 2026.

Singh, B. P., et al. “Financial Inclusion, Pradhan Mantri Jan Dhan Yojna Scheme and Economic Growth: Evidence from Indian States.” Economic Notes, vol. 50, no. 3, 2021, e12186.

Singh, V. K., and R. Prasad. “Diffusion of Banking Products in Financial Inclusion Linked Savings Accounts: A Case Study Based on Pradhan Mantri Jan Dhan Yojana in India.” Global Business Review, vol. 25, no. 4, 2024, pp. 981–1001.

Srivastava, N., et al. “Financial Inclusion and Bank Stability: Evidence from the Indian Banking System.” Economics Letters, vol. 250, 2025, 112303.

The Impact of Perceived Authenticity in Athlete Endorsements on Consumer Trust and Purchase Intent By Rehaan Bhat

Abstract

Athlete endorsements are widely used by brands to influence consumer perceptions and their purchasing behaviour, but they actually depend on key factors like authenticity, trust, and emotional connection. This study takes Virat Kohli as the subject and uses 4 conditions to test four key metrics: authenticity, consumer trust, purchase intent, and emotional connection. A quantitative survey was conducted with 17 respondents across 4 conditions: mismatched, matched with no story, matched with a story, and an invested condition. Respondents rated authenticity, consumer trust, emotional connection, and purchase intent using a Likert Scale. Descriptive statistics, Pearson correlations, and linear regressions were used to further analyse the data. The invested condition yielded the best results, with authenticity, trust, and emotional connection all predicting purchase intent. Emotional connection was found to be the most consistent factor among all the conditions, and this was the main finding in the paper. An unexpected finding was the fact that the mismatched product, despite having the lowest mean scores, produced a very strong correlation of emotional connection and purchase intent. The matched with a story and matched without a story conditions yielded the same results, which was another unexpected finding. These findings show that the research question was centered around authenticity and its role in purchase intent, but rather, emotional connection may be a more important factor. For brands trying to build their relationships with consumers by using athletes, it may be more important for them to make genuine connections rather than product fit. Building deeper connections with the athlete will produce better results for driving sales.

Keywords: Perceived Authenticity, Athlete endorsements, Consumer trust, Purchase intent, Emotional connection, Brand-Athlete Fit

1. Introduction

Athlete endorsements have become an important part of a brand's marketing strategy. It allows a brand to use popular athletes and help associate the consumers with their product. An athlete can influence people's perception of a product by endorsing it. However, the effectiveness of an endorsement actually comes from whether the product matches the athletes' personality and gives an authentic view to the consumers. Research has consistently shown that celebrity endorsements can positively affect brand image, brand awareness, and purchase intent (Jatto, 2014). Moreira et al. (2023) further establish that sports celebrity endorsements on social media directly drive brand trust and purchase intention, confirming that athlete influencers represent one of the most effective promotional strategies available to brands today.

Perceived authenticity is becoming more and more important with the rise of social media. Consumers are more willing to buy products by watching an advertisement or a post on Instagram that is endorsed by their favourite athlete or celebrity. When an endorsement aligns

with an athlete's personality, career, interests, and identity, it comes off as more genuine. If the endorsement has nothing related to their career, it seems to be done for commercial gains. Chundawat & Makkalageri (2026) further established that consumer engagement and domain expertise, better product fit, is important for driving purchase intent. Farhat (2025) further established that authenticity actually boosts purchase intent. These different types of instances can influence consumer trust, emotional connection, and purchase intent. Understanding these different types of scenarios can help brands better understand what drives consumer behaviour.

Despite these findings, significant gaps remain in the existing literature. Jatto(2014) reported that the number of respondents for the two different social media platforms was unequal. Moreirea et al. (2023) reported that they did not dive into purchase intent and how brand attitude plays a role in it. Aydin (2024) reported that there is no differentiation in the products, all under one umbrella, so trying to get an idea of whether authenticity is relevant is inconsistent. These contradictions across the literature point to a clear gap: no existing study has systematically compared different types of athlete endorsements on a single platform using the same athlete to isolate the effect of endorsement type on perceived authenticity, consumer trust, emotional connection, and purchase intent.

This study uses Virat Kohli as the focal point across four different endorsements to examine the different types of responses to each endorsement. Virat Kohli was selected due to his strong presence in cricket and worldwide popularity. The four conditions were a mismatched endorsement, a matched endorsement without a story, a matched endorsement with a story, and lastly an invested endorsement, in which he owns a stake in the company.. By holding the athlete constant and only varying the type of endorsement, this study is able to isolate the effect of endorsement conditions on consumer responses.

The paper studies four key metrics: authenticity, consumer trust, emotional connection, and purchase intent. It determines consumers' results and which metric plays a key role in driving purchase intent. The research question is as follows: how do different types of athlete endorsements play a role in perceived authenticity, trust, emotional connection, and purchase intent?

Credibility and trust play an important role in how consumers perceive and respond to content shared by athletes on social media. Priyadarshini & Jonuja (2025) further established that adding a personal narrative or the element of storytelling for a product can improve the perceived authenticity and trust. Aydin (2024) further established that credibility and trust are built when the athlete tries to align personal narratives with the product. Pokhrel (2026) further established that in order for a brand and athlete to have more credibility and trust among the consumers, there has to be consistent engagement, stating that frequency and exposure are key elements. This study addresses these contradictions by examining one athlete across four distinct endorsement types to determine which factors most strongly predict purchase intent

The paper is structured into several sections. The methodology of the paper reports the aims, the research design and demographics of the respondents, how the data was analysed, and lastly an ethics portion to show that no wrongdoing was done during the paper. The results

section presents the findings made during the paper using different statistical tools like averages, correlations, and regression analysis. Then it is followed by the discussion, which allows for personal interpretation of the findings, and then ends with a conclusion and the references used in the paper.

2. Methodology

2.1 Aims

The aim of this study is to examine whether perceived authenticity, consumer trust, and emotional connection in athlete endorsements play a role in purchase intent. Using Virat Kohli as the subject, this study investigates how four different types of endorsements (i.e., mismatched, matched no story, matched with a story, and invested) affect consumer attitudes and purchasing behavior toward an endorsed product

The objectives of the research are as follows

- To evaluate the consumer perception of perceived authenticity, trust, emotional connection, and purchase intent across four different Virat Kohli endorsement conditions.
- To analyse whether perceived authenticity significantly predicts consumer trust and purchase intent for a product using correlation and regression analyses.
- To compare the different endorsements and their results across all four of the metrics (i.e., authenticity, trust, emotional connection, and purchase intent) to identify which endorsement condition produces the strongest consumer response.

2.2 Research Design

This study adopted a quantitative research design to examine the perceived authenticity in athlete endorsements and its effect on consumer trust and purchase intent. Virat Kohli was selected as the subject of the paper due to his strong media presence and prominent personality in cricket.

The data were collected through Google Forms, which were distributed through WhatsApp. The respondents were presented with four different endorsement posts of Virat Kohli: mismatched, matched with no story, matched with a story, and invested. After viewing each endorsement post, the respondents were asked to rate on a Likert scale their perceived authenticity, emotional connection, trust, and purchase intent.

The data were then compiled and analyzed according to participants' perceptions across the endorsements. Analytical techniques including descriptive statistics, correlation analysis, and regression analysis were used to identify trends and to determine how perceived authenticity influences consumer trust and purchase intent across the four conditions.

Table 1: Demographic characteristics of survey respondents

Variable	Sample (n)	Sample (%)
Gender		
Female	6	35.29
Male	11	64.71
Familiarity with Virat Kohli		
Familiar (yes)	17	100.00
Not Familiar (no)	0	0.00
Interest in Cricket		
Daily	3	17.65
Weekly	2	11.76
During big tournaments	7	41.18
Rarely	3	17.65
Never	2	11.76
Previously purchased celebrity endorsed product(s)		
Yes	12	70.59
No	5	29.41

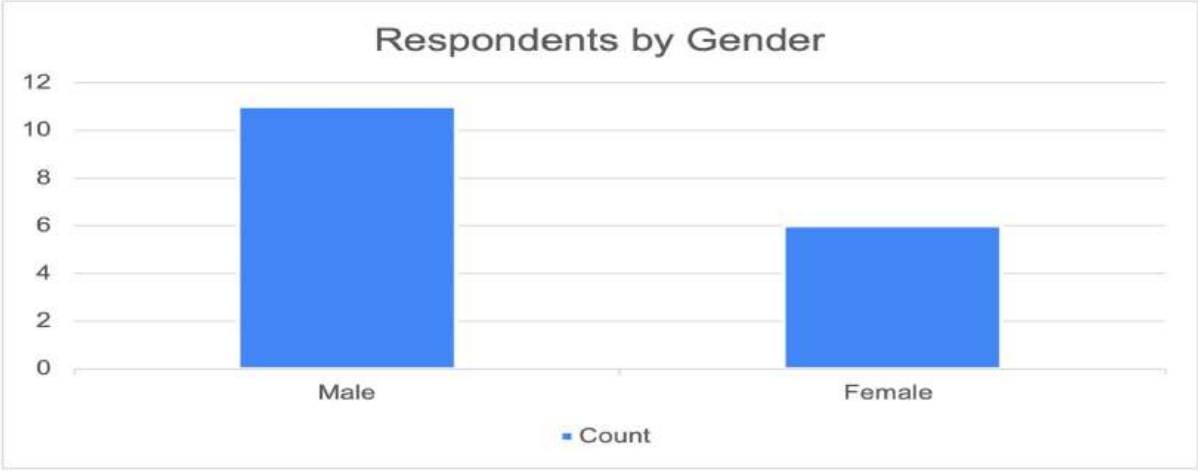


Figure 1: Gender Distribution of Respondents

The table shows the demographics of how many males and females there are. It also lists the familiarity with Virat Kohli, their interest in cricket, and whether the respondents have bought celebrity endorsements before. It represents each of these demographics with percentages as well.

2.3 Data Collection

2.3.1 Survey

To collect data from respondents, a Google Form was sent out via WhatsApp. This form first collected information about demographics (i.e., gender, age) and asked general questions about familiarity with Virat Kohli and prior celebrity endorsement purchases. It then presented four endorsements by Virat Kohli, with the same four questions asked for each condition.

a. Demographics

1. Age
2. Gender
3. Interest in Cricket
4. Do you know Virat Kohli
5. Have you bought any celebrity endorsement

b. Endorsements (Matched with a story, matched with no story, mismatched and invested)

1. How authentic does this endorsement feel to you?
2. How does this endorsement make you trust the athlete?
3. How emotionally connected to the product endorsed by the athlete are you?
4. How likely would you purchase the product

The same 4 questions were asked across all 4 of the endorsements to ensure consistency across the survey.

2.4 Data Analysis

2.4.1 Quantitative

In order to analyse the data it was taken from the Google Form, which was answered through a Likert scale. The data was then analysed by making pie charts, bar graphs, averages, correlations, and regression to further understand the data. Regression analysis was only conducted for variable pairs that showed a significant correlation.

2.4.2 Endorsement Conditions

Each respondent was presented with four endorsement posts featuring Virat Kohli, each representing a different level of perceived authenticity.

Matched Story Condition: It's an endorsement of Virat Kohli and Oakley, which uses the video to represent the Virat Kohli brand as someone who is driven and hardworking to achieve a goal.

Matched No Story Condition: It's an endorsement of Virat Kohli and TIDL, a relief spray. It aligns with his need for recovery, but the endorsement did not convey it properly/

Mismatched Condition: It's an endorsement of Virat Kohli and Digit Life Insurance; nothing related to the brand he has created through his career.

Invested Condition: It's an endorsement of Virat Kohli and One8, his apparel company, which he owns and runs.

2.4.3 Treatment of Missing Data

One participant did not complete the questions for two of the four endorsement conditions. These missing values were replaced by the mean score of the respective column, a method known as mean imputation (Tabachnick & Fidell, 2013).

2.5 Ethics

While writing the paper, all ethical responsibilities were maintained. The privacy of the respondents was protected at all times. The respondents were informed about the paper and its purpose before filling, making it completely voluntary.

No personal identification such as email address, name, or contact number was collected, ensuring all responses were completely anonymous. The data collected was solely used for research purposes.

All the posts of Virat Kohli were available and uploaded on the Google form as embedded YouTube videos. The findings of the paper are reported objectively without fabricating or manipulating the data.

3. Results

This section presents the findings of the survey analysis, which examines consumer responses across four Virat Kohli endorsement conditions. The findings are presented using descriptive statistics, correlation and regression analyses, and visual representations.

3.1 Descriptive Statistics

Table 1 presents the mean scores for perceived authenticity, trust, emotional connection, and purchase intent across the four endorsement conditions. Figures 1A through 1D display the response distributions for each metric individually.

Table 2: Average scores across four endorsement conditions (Perceived Authenticity, Trust, Purchase Intent, and Emotional Connection)

Condition	Authenticity	Trust	Purchase Intent	Emotional Connection
Mismatched	2.82	3.00	2.41	2.24
Matched No Story	3.71	3.82	3.18	2.76
Matched Story	3.41	3.59	2.82	2.59
Invested	3.76	3.29	2.71	3.12

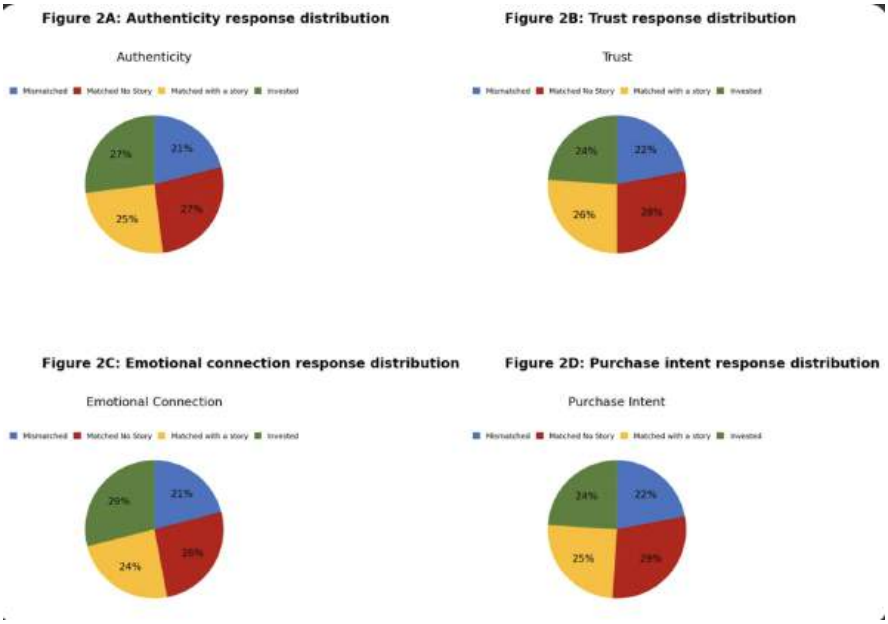


Figure 2: Response distribution of average scores across four Virat Kohli endorsement conditions

As shown in Table 1 and Figure 1, the invested condition produced the highest mean authenticity (M = 3.76) and emotional connection (M = 3.12) scores, while the matched no story condition produced the highest mean trust (M = 3.82) and purchase intent (M = 3.18) scores. The mismatched condition produced the lowest scores across all four metrics. Notably, the invested condition did not produce the highest purchase intent score despite having the highest authenticity score, suggesting that authenticity alone may not be sufficient to drive purchase intent. These patterns are explored further in the correlation and regression analyses below.

3.2 Correlation Analyses

Pearson correlation analyses were conducted separately for each of the four endorsement conditions to examine the relationships between perceived authenticity, trust, emotional connection, and purchase intent.

In the matched story condition, emotional connection was the only variable that showed a positive correlation with purchase intent ($r = 0.48$) and trust ($r = 0.38$). Perceived authenticity showed weak negative correlations with emotional connection ($r = -0.163$) and purchase intent ($r = -0.279$), suggesting that in this condition authenticity did not drive purchase intent. In the matched no story condition, perceived authenticity was strongly and positively correlated with trust ($r = 0.852$), while emotional connection was the only variable that correlated positively with purchase intent ($r = 0.460$). In the mismatched condition, emotional connection showed a strong positive correlation with purchase intent ($r = 0.850$), while trust showed positive correlations with both authenticity ($r = 0.613$) and purchase intent ($r = 0.571$). This was an unexpected finding as it was hypothesized that the mismatched condition would yield the weakest results.

Table 3: Pearson Correlation Matrix for the Invested Condition

Variables	Authenticity	Trust	Emotional Connection	Purchase Intent
Authenticity	1	-	-	-
Trust	0.58	1	-	-
Emotional Connection	0.38	0.62	1	-
Purchase Intent	0.73	0.76	0.70	1

As shown in Table 2, the invested condition produced the most consistent correlations across all four variables. Purchase intent was positively correlated with trust ($r = 0.759$), authenticity ($r = 0.731$), and emotional connection ($r = 0.705$), suggesting that when an athlete has a financial stake in a brand, consumers respond positively in all dimensions.

3.3 Regression Analyses

Regression analyses were conducted only for variable pairs that showed a meaningful correlation with purchase intent (i.e., $r > 0.5$). As no variable pairs in the matched no story condition met this threshold, no regression was conducted for that condition.

3.3.1 Matched Story Condition

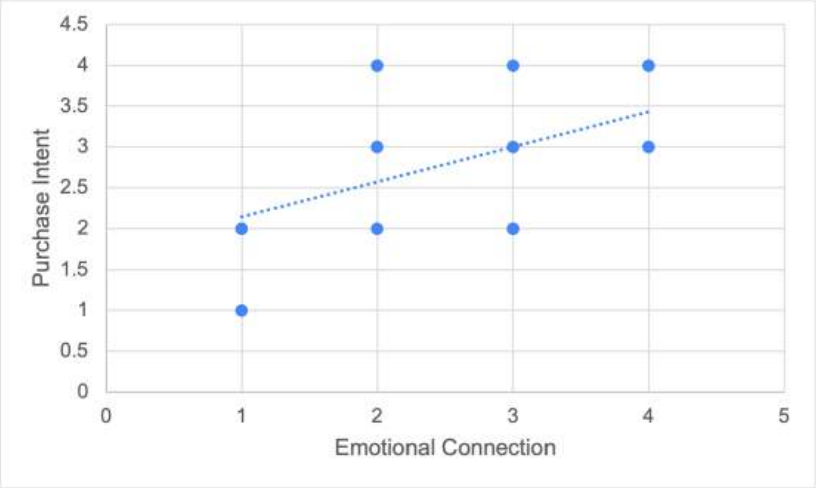


Figure 3: Linear Regression of Emotional Connection Predicting Purchase Intent in the Matched Story Condition

A simple linear regression was conducted to examine whether emotional connection predicted purchase intent in the matched story condition. As shown in Figure 2, emotional connection was a positive but marginally non-significant predictor of purchase intent ($B = 0.43$, $SE = 0.20$, $p = .051$), suggesting that while a trend exists, the relationship did not reach statistical

significance, possibly due to the small sample size.

3.3.2 Mismatched Condition

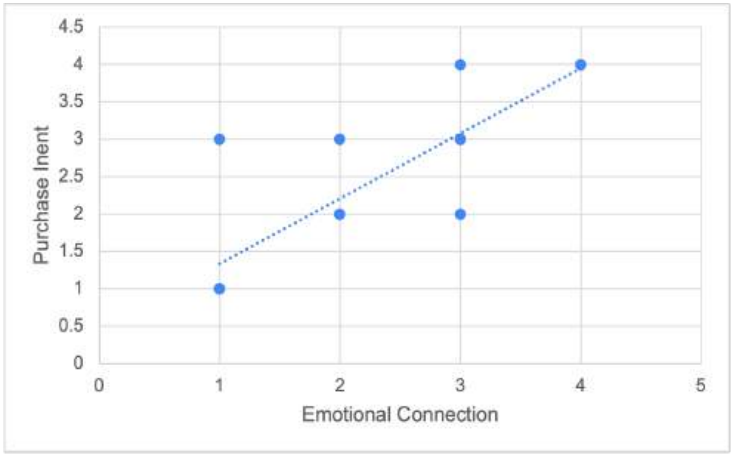


Figure 4. Linear Regression of Emotional Connection Predicting Purchase Intent in the Mismatched Condition

As shown in Figure 3, emotional connection was also a strong and significant positive predictor of purchase intent in the mismatched condition (B = 0.87, SE = 0.14, p < .001), indicating that for every one-point increase in emotional connection, purchase intent increased by 0.87 points. This was the strongest regression finding across all conditions.

3.3.3 Invested Condition

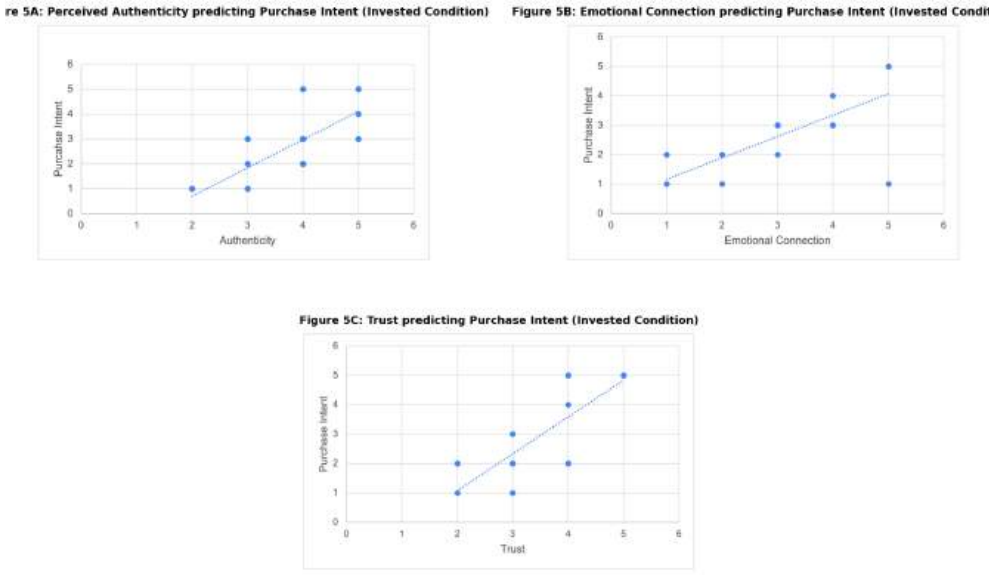


Figure 5. Predictors of purchase intent in the invested condition

Three regression analyses were conducted for the invested condition as authenticity, trust, and emotional connection all showed meaningful correlations with purchase intent. As shown in Figures 4, 5, and 6, emotional connection ($B = 0.73$, $SE = 0.19$, $p = .002$), trust ($B = 1.26$, $SE = 0.28$, $p < .001$), and perceived authenticity ($B = 1.14$, $SE = 0.27$, $p < .001$) were all significant positive predictors of purchase intent. These findings suggest that when an athlete has a financial investment in a brand, consumer trust, perceived authenticity, and emotional connection all independently contribute to driving purchase intent.

4. Discussion

4.1 Overview

The purpose of the study was to examine whether perceived authenticity, consumer trust, and emotional connection in athlete endorsements play a role in purchase intent. It was to study how four different types of conditions play a role in purchase intent. Emotional connection was a consistent predictor across all four conditions. The invested condition produced the strongest results compared to any other condition.

4.2 The Invested Condition Produced the Strongest Results

The strong correlations for authenticity, trust, emotional connection, and purchase intent were reflected by its significant regression values. It shows that all these variables are significant in predicting purchase intent. Farhat (2025) reports that a celebrity brand agreement is an essential necessity required for purchase intent, and even Moreira et al. (2023) report how a strong brand fit is essential so that consumers are not skeptical about the product. It can be argued that ownership signals a deeper and more genuine commitment, which consumers respond to in comparison to a regular endorsement.

4.3 The Mismatched Condition Produced Unexpected Results

Emotional connection was found to be the strongest single predictor of purchase intent across the whole study despite being a mismatched product. Aydin (2024) reported that athlete interactions are perceived as authentic regardless of product fit, which can be shown as a basis for unexpected results for the mismatched condition. Pokhrel (2026) reports that product fit is not a key element but rather the frequency with which an athlete interacts with consumers with a product is important for trust and brand fit. Since all respondents knew about the existence of Virat Kohli, it may have influenced them that even a mismatched product benefits from his emotional connection with fans.

4.4 The Story Condition Did Not Outperform Matched No Story

During the data analysis, it was found that the matched story and matched no story conditions produced almost identical emotional connection correlations with purchase intent. Priyadarshini and Jonjua (2025) reported in their study that matched stories are more effective. The reason the matched story did not produce better results might be because of a few reasons,

that are small sample size or the post not being compelling enough for the audience. Chundawat and Makkalageri (2026): the key finding for them was the credibility paradox, which states that trustworthiness alone does not drive purchase intent; it shows that the addition of a story does not help in increasing purchase intent.

4.5 Emotional Connection as the Consistent Driver

Across all four conditions, emotional connection was the variable most consistently correlated with purchase intent. Even the regression coefficients showed that these results were significant. This was unexpected since the research question was focused on perceived authenticity. Jatto (2014) reports that social media following helps build emotional connection, and since Virat Kohli is not only the most followed cricketer but also the third most followed athlete in the world shows a basis for why emotional connection is such a key variable. This finding is the most important finding of the paper due to the fact that it shows that authenticity and trust are not as significant for driving purchase intent in comparison to emotional connection.

5. Conclusion

Few limitations of this paper are the small sample size of 17 respondents, which limits generalizability. Since all the respondents were familiar with Virat Kohli may have inflated the emotional connection across all four conditions. Since the samples were collected from WhatsApp, it lacks a diverse sample of people. For future research, a larger and more diverse sample group may help in generalizability and understanding whether emotional connection plays a key role in predicting purchase intent. The same study can be tested with a less well-known athlete to see if emotional connection is still an important factor. Overall, this study highlights the importance of authenticity and emotional connection in shaping consumer responses to athlete endorsements.

Works Cited

Aydin, M. (2024). The Effects of Influencer Athletes on Sports Marketing in New Media. In Alıncak, F., & Özdal, M. (Eds.). Current Studies in Physical Education and Sports 2024. (pp. 121-127). ISRES Publishing.

Chundawat, R. S. & Makkalageri S. N. (2026). CELEBRITY ENDORSEMENT AND CONSUMER PURCHASE INTENTION IN THE URBAN SPORTSWEAR MARKET: Empirical Evidence from Puma Sports Shoes in Bangalore, India. International Journal of Science, Architecture, Technology, and Environment, 3(3). doi: https://doi.org/10.63680/ijsate0326015.012

Farhat, S. (2025). The Impact of Celebrity Endorsement, Celebrity Credibility,and Celebrity-Brand Congruity on Consumer Purchase Intention. LBS Journal of Marketing and Analytics, 1(1), 13-24.

Moreira, F. , Pereira, I. V. , Santos, J. D. & Pires, P. B. (2023). Sports celebrity endorsement impact on purchase intention. Galician Journal of Economics, 32(3), 1-21. doi: https://doi.org/10.15304/rge.32.3.8710

Jatto, O. I. (2014). Consumer Attitude towards Celebrity Endorsements on Social Media. Dublin Business School (MBA Dissertation), 1-92. https://esource.dbs.ie/server/api/core/bitstreams/da8b57cc-4b5b-4a82-8750-56ae8f7bc304/content

Pokhrel, P. (2026). Effectiveness of Athlete Endorsements in Enhancing Brand Awareness: Insights from Nepal. Nepalese Journal of Management Science and Research, 9(1), 199-221. Doi: https://doi.org/10.53056/njmsr-2026.9.1.010

Priyadarshini, U., & Jonjua, M. (2025). The Role of Media Storytelling and Narrative Building in Modern Sports. MediaSpace: DME Journal of Communication, 6(1), 21-29. Doi: https://doi.org/10.53361/dmejc.v6i01.04

On The Path to Financial Growth: Understanding Finance for Students
By Raima Sengupta

Abstract

This research paper studies financial literacy amongst young adults aged 17-22. Through researching financial topics that are important for students to understand as they are about to begin the journey of adulthood, including the following: credit, income, stocks, and student loan, this paper hopes to establish a platform of knowledge for students to reference and learn about financial literacy. The paper starts by identifying important terms necessary to understand and make financial decisions. Each term is defined and further explored to promote literacy amongst young adults. Then, a literature review was conducted to explore best practices for personal financial decisions. Finally, the research paper involved qualitative interviewing of two financial experts, both of whom are college finance professors, to gain evidence of best practice decisions young adults can make to improve their financial success. Between the qualitative interviewing and literature review, this paper hopes to empower students to better understand their own finances and make evidenced-based decisions for promoting wealth.

Introduction

Finances influence nearly every aspect of modern life, affecting both everyday decisions and long-term financial opportunities. In fact, money itself can affect every decision a person makes, especially as one gets older. Sure, an 8-year-old might not be thinking how much his ice cream costs on a warm sunny day, but on that same day 10 years later during the summer, that kid might be 18, filled with the impending doom of how to deal with their money as they enter the real world. When the term money comes to mind, most people naturally connect it to finances. However, this term can often be loosely thrown around. What is Finance?

When asked to define finance, many young adults struggle to provide a clear explanation beyond simply associating it with money. Authors of Financial Literacy among the Young report "fewer than one-third of young adults possess basic knowledge of interest rates, inflation, and risk diversification"(Lusardi, Mitchell, Curto, 2010). However, it is at this age where people should obtain knowledge of financial literacy, as"they assume full responsibility for managing and directing their financial affairs"(Breitback,Walstad, 2016) Surely there is some kind of space that deals with explaining finance to other students, so that we can make smart, educational decisions and not get lost in the sea of vocabulary and options presented to us. Thus, the intent of this paper is to articulate financial terms through the usage of research and simplistic wording for students aged 17-22, creating a platform that allows them to make smart financial decisions, gain independence, and understand what to do with the knowledge at hand.

First, there will be a deep dive of terms separated into five subcategories, highlighted as important for students to understand: credit, stocks, income, student loans, and student debt. In each section, additionally, a mini refectory section will be written about why this knowledge is pertinent to how to navigate the future for this category, specifically those planning to attend

college. To conclude, a reflection on the importance of financial literacy words will be presented and how this affects all people will be considered. Ultimately, this piece should become a resource to all students who are in the same boat, struggling to figure out financial literacy but willing to learn its value.

Methods

This paper aims to answer the research question of how financial literacy can be made more accessible to students, specifically highschoolers and undergraduates. through two primary methods. First, a literature review was conducted. This process was completed through research done on Google Scholar. Second, qualitative interviews were completed. These interviews were conducted on zoom and by email, where questions were emailed to the finance professors regarding student loans, debt, and student life.

Income

There is no clear definition for income, as discussed by John R. Brooks in his article titled, The Definition of Income. Personal income, in simple terms, all income collectively received by individuals/households in a country Personal income, also called gross income and/or individual income, is proportional to the current state of the economy (Szarowska). For example, during economic booms a person's personal income rapidly accelerates upward, and vice versa during recessions. In terms of calculation, all that is added is every different type of income a person has. The main component of personal income that separates from all other types of income is the fact that it is the total payment a person has before any taxes are paid, focusing on the amount of money an individual obtains. This includes payments such as salaries and wages(Federal Reserve Bank of San Francisco, 2000) Because of consumer spending's massive role in the economy, personal income is typically tracked by a quarter on an annual basis. Tracked by the Bureau of Economic Analysis, the agency does a tremendous job of comparing data from previous months as well as separating personal income to the different categories it is altered by to determine the trends arising and shifting in the current economy Overall, the main reason as for why personal income is so important when understanding/deriving financial literacy is that it plays a huge role in the economy, since it directly affects consumer spending. Understanding and analyzing the trends arising from the role of personal income allows financial analysts to understand the economy and components of the economy much better, especially individual's actions. It teaches students why having a job is so crucial for enhancing personal finance.(Szarowska)

Disposable Income

Not only is personal income important when understanding finance, but disposable income is just as integral. Disposable income is the money a person has to actually spend, after taxes have been subtracted. The main difference when comparing personal income and disposable income is the role of taxation(Federal Reserve Bank of San Francisco, 2000). One

useful way to visualize this relationship is as a ladder. The first step would be personal income, then the deduction from income by taxes, and then disposable income follows. Income does not follow such a linear path as it is altered and shifted through many factors, but this is a simplistic way of differentiating between the two types of income. Following this definition, disposable income is calculated by subtracting personal income tax,"a single worker taxation"(Szarowská, 2014). Disposable income reveals the money that a person has left to use however they choose, be it spending, investing or saving. As students join the workforce, it is important that they understand how their earnings are taxed and calculated, allowing for them to make smarter decisions based on their total amount of money remaining. However, it's important to note that disposable income does not directly tell a student or anybody for that matter how their earnings after being taxed should be spent.

Credit

The next term to define is credit, which is the ability to borrow money under the agreement that you will repay the debt later (Waugh, 2024). Essentially, it is a deal between the lender, the person who gives the money, services, goods, etc., and the borrower, the person who receives capital in exchange for a premium, defined by an interest rate . How a person's credit is defined, from 300-850, says a lot about the person and affects the rate at which a person can make certain financial decisions. For example if a person has good credit (defined as greater than 670), meaning they typically pay the debt on time, it can make it easier to apply for other loans, borrowing, insurance, housing, and even work. A bad credit score(lower than 579) would lead to the opposite effect, often making a person end up with no financial gains or autonomy.

Furthermore, there is interest involved in the transaction. Interest accrues when a person takes out a loan or makes some financial agreement, then they have to pay back the original money, goods, etc. an extra percentage on that amount established previously by the lender(Kern, 2017). Additionally, as previously mentioned, having good or bad credit can affect many aspects of a person's life. The determination of this is through things such as a person's history of borrowing and paying back debts. Additionally, credit history can be found in credit reports, given by different credit bureaus, including Experian, Equifax, and TransUnion.The information in credit reports given by these companies is used to assess a company's financial status(Chatterjee, Corbae, Dempsey, Rios-Rull, 2023). Obtained from different financial documents, such as income statements and credit scores, Segal asserts that Character, Capacity, Capital, Collateral, and Conditions are the 5 c's to help lenders determine who to give money to.

Character deals with a person's credit history, which has already been discussed. Just remember there is an increasingly proportional relationship between a person's credit history and credit overall. Another important component of how borrowers are addressed is through their credit score, which will be defined later. Next is capacity, determining a borrower's ability to pay back debt by looking at their debt-to-income ratio, DTI. To calculate DTI, add the monthly debt payments of the borrower and divide by the person's gross monthly income. Credit and DTI are inversely proportional, and most lenders typically prefer around less than 36% DTI.

Additionally, capital, essentially the money a person has, is highly viewed by lenders as it indicates investment levels. For example, those who invest more capital have higher investment levels, allowing the lenders to feel better regarding trust in giving credit. Additionally, collateral allows the lender to have the guarantee that they get something back if the borrower decides to "default" the loan. Collateral loans are supposedly not as risky as others since it's most of the time the reason money has been borrowed. Due to this, these specific loans typically have interest rates that are not as high. The best way to remedy collateral is by participating in a "loan agreement", in which the lender receives the "losses" after a default. Finally, the last C important for credit is conditions, which is when lenders look at all things surrounding the loan, including the current industry, a person's stability, their job, and more. This is extremely important as it helps the lender determine whether it is a good or bad decision to enter a deal with a borrower. These 5 c's determine the credentials of a borrower for credit. It is essential to understand these 5 key terms when learning about credit, as it allows others to better understand the process of lenders when making a decision to enter into a deal of credit(Segal).

Credit score is "an object that yields a ranking of individuals with regard to their probability of default on a particular contract"(Chatterjee, Corbae, Dempsey, Rios-Rull, 2023) Created by the Fair Isaac Corp (FICO), it reveals a person's history in terms of credit. Credit score ranges from 300-800. The higher the number the better chance you have of receiving a loan from a lender. To calculate this number, it involves a specific percentage of a person's payment history, amount owed, length of credit history, credit types, and if there is any new credit, typically done by the 3 largest credit bureaus located in America(Waugh, 2024).

Now there are many tips and tricks that can either help or harm a person's credit score. The following are how a person's score can drop: not making payments on time, using more than 30% of your credit, ending a credit card (increases ratio of credit utilization), and having unused credit. These factors can seriously harm students aged 17 to 22 when applying for a loan and can be a huge setback on their future. However, with the right steps, credit can increase as well: paying bills using specific timing, keeping the credit utilization ratio minimal, asking for help from lenders, join as a user on a friend's card, keeping check of your credit report, and/or having a secured credit card, where a cash deposit must be sent to the lender first, which the lender will not do anything to if payments are being made properly. Overall, it's important to understand how a credit score works, as it can hinder or help a person make important financial decisions, such as for education or housing. While this information may seem like a lot, there is even more to learn about credit. However, this is a beneficial and simple start to understanding credit that students need to know. Credit is one of the foundational concepts of financial literacy, and needs to be understood in order for students to do things such as applying for loans, finding housing, and even when making down payments(Garrett, Toumanoff, 2010).

Stocks

The next piece of financial literacy that will be discussed is stocks and the stock market. Stocks represent ownership in a corporation. They're created by a company in order for that

business to profit, because they are getting money from stockholders purchasing shares. Multiple stocks in a company are called shares(U.S. Securities and Exchange Commission). They allow the person who bought the unit, shareholders, to receive profits and assets of that corporation based on how many shares they own(Hayes).

Typically, stocks are purchased and sold through public exchanges. The company must become public first with initial public offering in order for stocks to be purchasable, and has to oblige with any laws created by the Securities and Exchange Commission. With a brokerage account, the stock's bid (purchasing price) and offer (selling price) is presented, which can fluctuate based on the current status of the market.

There are many advantages that come when owning stocks. For example, a shareholder has the ability to participate in shareholder meetings, can get dividends, part of the earnings of a company (this is not mandatory for corporations to pay), and have the ability to sell their shares. Dividends are the earnings of the company, represented with cash. The more shares an individual has means more power over a company's shares and the earnings, not the company. Additionally, people can earn profit off of stocks with capital appreciation or dividends. Dividends have already been explained, but capital appreciation is when the cost of a share rises. The profit is the amount of a share sold for deducted from the later price of it(Kirkpatrick, 2009).

When discussing stocks, there are two types that often arise: common and preferred. Common stocks are those that allow owners to participate in shareholder meetings and acquire dividends. People who own preferred stocks, however, cannot vote in meetings, but they have greater power over earnings and assets. They obtain dividends first and are considered first as well if something goes wrong with the company. However, no matter what kind of a stock a person owns, when companies decide to increase the numbers of shares in their business for money, then the stockholder loses power(U.S. Securities and Exchange Commission). While stocks have many advantages and disadvantages, the question of whether one should invest is up to them. Investing has risk, no matter what so there is no guarantee. However, stocks have proven to be one of the best to invest in due to their profit(Kirkpatrick, 2009). For example, the S&P 500 faced about an 18.4% gain from the start to end of 2020.

Stock Market

When most people think about stocks, they often associate the word stock market along with it. Peter Gratton thoroughly explains how the functions of the stock market, in which it, "consists of a primary market in which businesses first issue shares to investors and a secondary market in which investors trade shares"(Gratton). In greater detail, a primary market is one where people like investors buy shares straight from the company and have a private placement, where the shares are sold disregarding the registration process. The stock market is the platform where stocks, investment funds, real estate investment trusts, preferred stocks, derivatives, and American depositary receipts are sold and purchased, such as the New York Stock Exchange, which follows all regulations under the Securities and Exchange Commission in America, one of the federal bureaucracies(U.S. Securities and Exchange Commission). The stock market consists

of various exchanges of shares, rather than anything specific and has a large impact on the economy, as it alters the relationship of cash with businesses and those who invest in them. It regulates public companies, represents the performance of the economy, allows for investment of everyday people, and is a platform formed by the support and communication of investors, traders, and regulators.

In order to understand how the stock market functions, one must properly understand a stock exchange. In this process, shares are traded openly after the company becomes public. This is the type of exchange where trading is the most common, examples including NYSE and Nasdaq. There are specific rules and regulations that investors must oblige based on location.

Stocks do not only have to be exchanged this way, however. Another way would be through an over-the-counter market. Here, a person has the ability to exchange stocks straight to the investor with less regulation and publicity. It is important to note, however, that OTC markets are often for smaller companies with less liquidity and do not contain the correct criteria for the stock exchange because of this. This can be a setback for investors, because there tends to be less data for these smaller companies(U.S. Securities and Exchange Commission).

In simple terms, prices of stocks are determined through communication by investors and dealers. Usually though, investors tend to follow prices listed on brokerage platforms. However, there is an agreement that must be made when investing in stocks and this decision alters the demand of certain stocks. Prices of stocks are typically affected by fundamental and technical components, those that are based on the financial status of the company and those that are based on research of the market's performance and trends seen with stock pricing. When stocks have high prices, it typically means the company is successful but could also be due to dividends, share repurchases, and splits of stocks. When the price of stocks decreases, it represents the market value of that stock is not as high as it used to be. Additionally, companies have the ability to put the stock price higher, influencing investors to purchase shares(Beers, 2023).

When investing in the stock market, it's important to understand S&P 500, "a float-adujust market-capitalization-weighted index tracking the stocks of 500 leading US public companies"(Kenton, 2026). Indexes are benchmarks" for stocks as to how they are performing. According to S&P 500 Index: What It's for and Why It's Important in Investing, they represent 80% of the total equity in America's market value, and companies are chosen after much analysis of their capitalization and liquidity. This is then re-analyzed quarterly, where companies might shift based on the requirements to be part of the index. To determine the "weight" of each individual company in the index, take the company's market cap and divide it by the total market cap presented in the index. While the index might have originated by a private company, it has certainly expanded over time and now plays an integral role in determining the status of the market economy by many(S&P Global).

For Students

However, the audience of the paper is not the average American, it's students, specifically 18 years or older, who are trying to understand how to manage their finances and

wealth properly in order to be financially stable and learn key insights at such a young age. Well to simply state it, The stock market provides students with an opportunity to build long-term wealth, although successful investing requires patience, research, and an understanding of risk. If one starts off earlier, they can earn even more when buying shares and save up before getting college, helping to pay off student debt or college tuition. However, if one thing has been made clear, then it is the fact that the stock market is a risky business. In order to invest, it requires a proper understanding of the stock market. This, therefore, forces students who want to earn cash to research before any decision is made. As said by Lori Schock, former Director of the SEC's Office of Investor Education and Assistance, "planning for the future starts right now"(Schock). Early investing can also help students develop financial discipline and long-term decision-making skills.

Student Loans

The next section of this paper is arguably the most important to understand for 18 year olds and above: student loans and student loan debt. Here, the paper will dive into what student loans are, the rules and regulations for them, strategic ways to apply and utilise them, as well as advice from college professors on how to approach student loans and the debt that often follows. Student loans are a way for families to pay for college without having to pay as many financial difficulties that would have arisen if tuition had to be paid in 1 transaction. In order to be eligible for federal student loans, families must first complete the Free Application for Federal Student Aid (FAFSA) every year, pertaining to a family's financial status. However, private student loans simply need application submissions sent to any private lender, the person who receives money from the borrower. After completing this, the Student Aid Index, will be calculated determining the amount of financial aid necessary(Federal Student Aid). After this step, financial aid offices will be contacted at the college the student will be attending and determine the student's need after subtracting SAI from cost of attendance of a college(dorm, tuition, etc.) When a student is accepted, they will hear concurrently how much financial aid they qualified for. There are two types of loans to look out for in these letters of acceptance: direct subsidized and direct unsubsidized loans. Direct subsidized loans are given to students who need extremely high financial aid, its interest taken care of by the United States Department of Education as a student who is half-time and up to 6 months post graduation. Direct unsubsidized loans are open to all. Federal loans and private student loans each have advantages and disadvantages. For example, federal loans have low fixed interest rates and have greater leniency with payment. However, there is a cap on how much money can be borrowed. If parents choose the PLUS loan path, then there is more money involved, specifically from the parent or legal guardian. On the other hand, private student loans are always an option for students, needing the signature of a guardian and a relatively good credit score(here is why a good credit score is crucially important for students, as discussed earlier). It is important to note that these loans typically have higher, variable interest rates when compared to federal loans(Avery,Turner, 2012).

Qualitative Interviews

Student loan debt can often be viewed in differing lenses. Summarized by finance professor Jason Hall at The University of Michigan, "Student loan debt, like any other debt, allows people to access opportunities for investment (like buying shares and property) and consumption (like buying a car) that in the absence of debt would only be available to people who are already wealthy"(Hall)It can be seen as an attempt to bridge the gap between those who are fortunate and less fortunate, as well as enforces a highly supported opinion: education should be accessible to everyone, regardless of income. Professor Hall notes, additionally, that this debt does tend to pay off as students tend to receive a higher income after graduating than those who did not.

One question that tends to linger in the mind of students and their families is whether student loans are beneficial or harmful long term. To conduct further research, an interview was conducted with Professor Dario Laudati. Being an assistant professor of Finance at the Miami Herbert Business School, Professor Laudati has insightful information he generously offered for students. He notes that in past years, there used to be a great value of college premium: the students who went to college earned much more after graduating than those who had not. Yet, both professors also state that this is no longer the case, and the gap between the two groups has lessened due to the effects of time and policy changes. While there is no confirmation any more about the income, both professors recognize the value of a college education expands beyond the idea of how successful a person is. There is no definite answer to this question. Professor Hall takes an optimistic lens, as students, "can enter programs that would be beyond their reach in the absence of student loans, and therefore having to come up with upfront cash". However, both professors note the risk and danger that student loan debt can lead to. In the interview, Professor Laudati articulates that the mind of a student must be able to comprehend life from a macro and micro perspective: to look at life as a big picture, where one decision can affect anyone and everyone, while simultaneously understanding complex details about life. This skill, Professor Laudati claims, is only attained through college(Laudati).

The ages of 17 to 22 can often be the most confusing, unpredictable, years of a person's life. One suddenly goes from a teenager to adult, typically attends college, and all of a sudden had to gain a surge of independence. It feels as though life has spun on its axis, with no sense of guarantee regarding what their future will resemble. Taking Stock in Teen Trading by Lori Schock suggests investing is a great opportunity for students to get involved in the market or business world. Stocks, funds, and more are ways that people learn about harms and benefits from investing. Additionally, it teaches students the value of research before making crucial decisions(Schock).

At this point in life, specifically when entering college, Professor Laudati recommends that students should try to look at life from a macro and micro perspective. From a macro perspective, he suggests that students should think about how one decision can affect something globally. Opening the views and expanding the less allows people to become more aware and properly understand more about how the world functions, escaping this narrow bubble and this

perception that people are isolated. Next, students should also think about life from a micro, smaller lens. Looking at life from a macro lens reveals how every decision, from a small or large scale, has the ability to impact an individual. The combination of these two components makes students better prepared to tackle the real world, as they are more aware of how actions have indirect consequences and allows them to gain more clarity of how things in the world function.

When asked about college, both professors stated that the value of school should not be overlooked. Despite the chance of earning a lot after graduation, the importance of college is because of the lessons a student is taught. It is the place to make mistakes, grow as an individual, and learn more about surroundings. College teaches students skills no other experience can, therefore one should not view college as "transactional"

Professor Hall states that understanding one's preferences is necessary for determining what financial decisions an individual makes. Does the person care more about profit or satisfaction? Questions such as this are what students should ask themselves when understanding their finances. Understanding the opportunity cost, what a person must give up in exchange for choosing something else, is crucial to determining what someone's trajectory will look like. It is truly dependent upon a person's individual personality or characteristics to make the decisions that satisfy them the most.

Conclusion

Financial literacy is not about becoming the wealthiest instantly. It is about giving young adults the knowledge and confidence to make informed decisions before costly mistakes can occur. Whether students are building credit, investing for retirement, or financing their education, understanding these concepts today can shape the opportunities available to them for the time that is yet to come.

Works Cited

Avery, Christopher, and Sarah Turner. “Student loans: Do college students borrow too much—or not enough?” Journal of Economic Perspectives, vol. 26, no. 1, 1 Feb. 2012, pp. 165–192, https://doi.org/10.1257/jep.26.1.165.

Beers, Brian. “Understanding Stock Dividends: Payouts, Key Dates, and Payment Methods.” Investopedia, www.investopedia.com/ask/answers/102714/how-and-when-are-stock-dividends-paid-out.asp. Accessed 4 Aug. 2026.

Bird, Beverly. "Boost Your Credit Score: 5 Key Actions and Pitfalls to Avoid", www.investopedia.com/how-to-boost-your-credit-score-11729090. Accessed 3 Aug. 2026.

Bromberg, Michael. “Essential Investing Tips for Teens: Start Building Wealth Early.” Investopedia, www.investopedia.com/investing-for-teens-7111843. Accessed 1 Aug. 2026.

Breitbach, Elizabeth, and William B. Walstad. “Financial Literacy and Financial Behavior among Young Adults in the United States.” Economic Competence and Financial Literacy of Young Adults: Status and Challenges, edited by Eveline Wuttke et al., 1st ed., vol. 3, Verlag Barbara Budrich, 2016, pp. 81–98. JSTOR, https://doi.org/10.2307/j.ctvbkk29d.7. Accessed 21 Aug. 2026.

Brooks, John R. “The Definitions of Income.” HeinOnline, 8 Mar. 2021, access.heinonline.com/HOL

Chatterjee, Satyajit, et al. “A Quantitative Theory of the Credit Score .” Wiley, onlinelibrary.wiley.com/doi/full/10.3982/ECTA18771. Accessed 24 Aug. 2026.

Chen, James. “Common Stock: What It Is, Different Types, vs. Preferred Stock.” Investopedia, www.investopedia.com/terms/c/commonstock.asp. Accessed 4 Aug. 2026.

“Discuss the Definitions and Means of Calculating National Income, Personal Income, and Disposable Personal Income.” Federal Reserve Bank of San Francisco, 1 Mar. 2000, www.frbsf.org/research-and-insights/publications/doctor-econ/2000/03/national-personal-disposable-income-economic-indicators.

Dividends: What They Are, How They Work, and Important Dates, www.investopedia.com/terms/d/dividend.asp. Accessed 1 Aug. 2026.

Epstein, Lita. “A Beginner’s Guide to Applying for Student Loans.” Investopedia, www.investopedia.com/articles/personal-finance/091515/beginners-guide-student-loans.asp. Accessed 4 Aug. 2026.

GARRETT, DENNIS E., and PETER G. TOUMANOFF. “Are Consumers Disadvantaged or Vulnerable? An Examination of Consumer Complaints to the Better Business Bureau.” The Journal of Consumer Affairs, vol. 44, no. 1, 2010, pp. 3–23. JSTOR, http://www.jstor.org/stable/23859773. Accessed 21 Aug. 2026.

Gratton, Peter. “What Is the Stock Market and How Does It Work?” Investopedia, www.investopedia.com/terms/s/stockmarket.asp. Accessed 1 Aug. 2026.

Harbour, Sarita. “Credit Score Ranges: What Do They Mean?” Investopedia, www.investopedia.com/articles/personal-finance/081514/what-do-credit-score-ranges-mean.asp. Accessed 3 Aug. 2026.

Hayes, Adam. “Stocks: What They Are, Main Types, and How They Differ from Bonds.” Investopedia, www.investopedia.com/terms/s/stock.asp. Accessed 1 Aug. 2026.

Hershbein, Brad, and Kevin M. Hollenbeck. Student Loans and the Dynamics of Debt. W.E. Upjohn Institute, 2015. JSTOR, http://www.jstor.org/stable/j.ctvh4zh26. Accessed 24 Aug. 2026.

“Income & Saving.” Income & Saving | U.S. Bureau of Economic Analysis (BEA), www.bea.gov/resources/learning-center/what-to-know-income-saving. Accessed 1 Aug. 2026.

Kern, Abby M. “Credit Score Analysis.” OpenSIUC, 20 Apr. 2017, opensiuc.lib.siu.edu/cgi/viewcontent.cgi?article=2074&context=gs_rp&httpsredir=1.

Kirkpatrick, Charles D. Beat the Market: Invest by Knowing What Stocks to Buy and What Stocks to Sell. FT Press, 2009.

LUSARDI, ANNAMARIA, et al. “Financial Literacy among the Young.” The Journal of Consumer Affairs, vol. 44, no. 2, 2010, pp. 358–80. JSTOR, http://www.jstor.org/stable/23859796. Accessed 21 Aug. 2026.

Reiff, Nathan. “Advantages of Long-Term Stock Investment.” Investopedia, www.investopedia.com/articles/investing/052216/4-benefits-holding-stocks-long-term.asp. Accessed 4 Aug. 2026.

“S&P U.S. Indices Methodology.” S&P Dow Jones Indices, 13 July 2026, www.spglobal.com/spdji/en/methodology/article/sp-us-indices-methodology.

Segal, Troy. “5 Cs of Credit: What They Are, How They’re Used, and Which Is Most Important.” Investopedia, www.investopedia.com/terms/f/five-c-credit.asp. Accessed 3 Aug. 2026.

Sengupta, Raima, and Dario Laudati. “Interview With Professor Dario Laudati Regarding Student Debt.” 11 July 2026.

Sengupta, Raima, and Jason Hall. “Interview With Professor Jason Hall Regarding Student Debt.”

“Steps for Students Filling out the FAFSA Form.” Federal Student Aid, studentaid.gov/articles/fafsa-student-steps. Accessed 24 Aug. 2026.

“Stocks - FAQs.” Investor.Gov, www.investor.gov/introduction-investing/investing-basics/investment-products/stocks. Accessed 24 Aug. 2026.

Szarowska, Irena. “Personal Income Taxation in a Context of a Tax Structure .” Science Direct, www.sciencedirect.com/science/article/pii/S2212567114003918. Accessed 24 Aug. 2026.

“Taking Stock in Teen Trading.” Investor.Gov, www.investor.gov/additional-resources/spotlight/formerdirectorlorischock-directors-take/

taking-stock-teen-trading#:~:text=Learning%20about%20how%20the%20stock,buying%20a%20home%2C%20and%20retirement. Accessed 3 Aug. 2026.
Waugh, Evelyn. “What Is Credit?” Experian, 27 Jan. 2024, www.experian.com/blogs/ask-experian/credit-education/faqs/what-is-credit/.
“What Is a Credit Score?” Investopedia, www.investopedia.com/terms/c/credit_score.asp. Accessed 1 Aug. 2026.
“What Is the S&P 500 and Stock Market Average Return?” Fidelity, 6 Mar. 2026, www.fidelity.com/learning-center/trading-investing/sp-500-average-return?

How External Influences Affect Individual Investor Decision-Making, and How Individual Investors Can Better Control their Emotion By Cyrus Melnyk

Abstract

While individual investors are allocating their money in different ways, they all share the same goal of making a profit. This paper looks at how external market and internal psychological factors work together to shape the way individual investors make decisions. However, the reason the path for each dollar is different is that each person has a different psychological reaction to external influences. External factors like interest rates and market manipulation shape the information investors react to, but internal factors like emotions and biases can influence decisions just as much. When investors are hit with sudden price changes or unexpected news, the brain's stress response can take over, impairing rational decision-making and leading to patterns like loss aversion, overconfidence, and impulsive buying or selling. This paper dives deeper into how individual investors make decisions with their money, how external factors impact these decisions and sway their emotions, and how to better control emotions when investing. As more individual investors enter the market, understanding these influences becomes even more important. The information found in this paper was collected through reading books and other research documents on similar topics.

Introduction

The surface-level concept of investing seems very simple, with the common rule of buying low and selling high. However, if investing was so simple, everyone who invested would be rich. Many other external and internal factors such as overconfidence and confirmation bias play a role in the investing experience. An example of an external factor is interest rates, where the Federal Reserve raises interest rates, leading to stocks usually falling. Other external factors include inflation, where the value of the dollar decreases, market manipulation, where an authoritative person can influence people's decisions when investing to boost the value of a stock, and many other factors that impact investor decision-making. Internal factors, including emotions and cognitive behaviors, can impact the investment outcome as strongly as external factors. For example, impulsive investors may buy a stock too soon due to a sudden surge in price, just for it to drop again. Internal factors of investing are the most unknown and unpredictable variables in investing. Understanding the relationship between external influences and cognitive investor psychology is critical in modern investing. As more individual investors enter the market, the impact of these influences becomes even more significant.

When an unexpected change happens, the brain goes into an acute stress response. The amygdala, the brain's central controller of fear and emotional response, begins firing off more intensely in an attempt to regain control of the situation. This stress response can impair the prefrontal cortex, the brain region responsible for rational decision-making (Arnsten, 2009). In financial markets, rapid price changes, breaking news, or large market swings can trigger this response in investors. Instead of calmly analyzing information, the brain makes impulsive

decisions to try to mitigate the often temporary negative impact of these scenarios. These emotional reactions directly influence how individuals interpret market information and determine their investment decisions. Investors often become unpredictable and disregard their original strategy. These reactions create dangerous patterns of impulsively buying or selling out of fear or excitement, overconfidence, and loss aversion.

This paper examines how external and internal influences affect individual investor decision-making and explores how investors can better control emotional responses when making financial decisions. By analyzing behavioral investing and real market examples, this study investigates the role of psychological biases, market signals, and external pressures in shaping investor behavior.

External factors like interest rates and market manipulation shape the information investors react to, but internal factors like emotions and biases can influence decisions just as much, so understanding both is key to making better financial decisions.

Literature Review

The modern increase in the number of individual investors participating in financial markets has attracted attention to the importance of external and internal factors in investment decisions. External factors such as interest rates, inflation, and market volatility have a considerable impact on investment decisions. For example, changes in monetary policy, such as a rise in interest rates, affect stock prices negatively due to increased borrowing costs and a reduction in expected returns in the future. Inflation also creates uncertainty for investors because it reduces their purchasing power and changes their expectations for returns in real terms, causing investors to react by selling stocks if they believe the stock will lower in the future or they will move the money to safer assets such as bonds or cash. Market volatility and price movements also create uncertainty for investors, making the environment so complex that decisions have to be made under pressure.

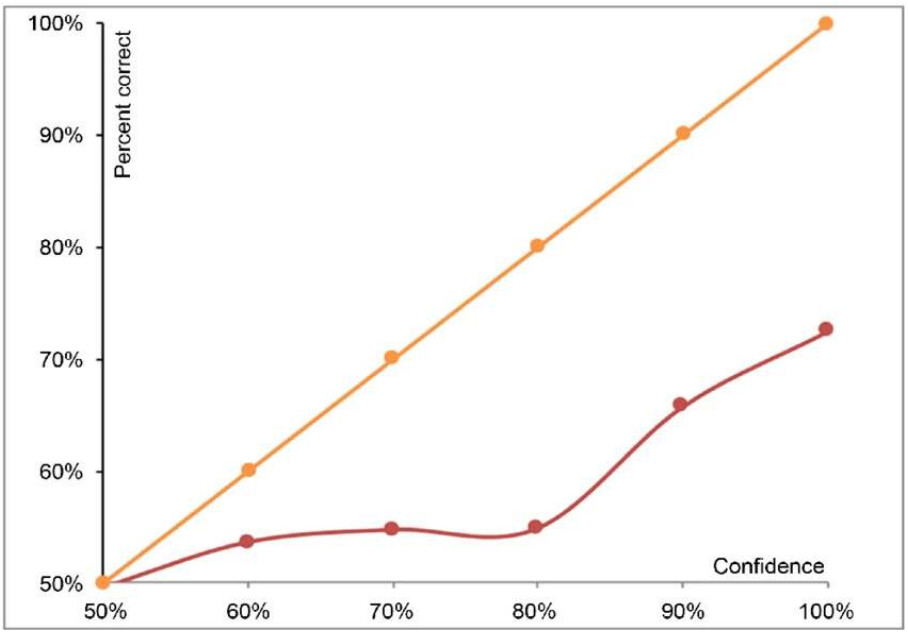
However, research has also shown that, apart from external factors, internal cognitive biases also have a critical role to play in altering investor behavior. Behavioral finance literature has consistently shown that loss aversion, overconfidence, and mental framing are critical factors that contribute to irrational behavior, and reduce investor performance.

Investors make **systematic judgment errors**, which are based on heuristics such as mental shortcuts, that they use for financial decisions (Khare & Kapoor, 2023). The errors end up costing investors lots of time and money in the long run. When investors make mental shortcuts, they often end up forgetting or breaching their personal investment strategy, or even basic investing knowledge, because they believe they are “smarter” than the market. Some of the judgment errors result in outcomes such as the following:

Individuals **overestimate their ability** to predict market outcomes, which ultimately results in excessive trading and increased risk-taking (Atif Sattar, 2020). This overestimation of their abilities is called **overconfidence bias**, which results in making too many judgment errors. Overconfidence bias can develop within an individual in 2 ways. The first way is generally with

beginners, when they believe they understand the way the market works from research and watching videos, but they have never made an actual investment. In turn, these beginner investors will make irrational decisions, and oftentimes forget a majority of the information they learned about investing beforehand, causing them to lose money. Due to this, these investors may give up immediately as they had poured so many hours into learning investing, and lost their money immediately. The second way investors may gain overconfidence is when they get lucky with multiple investments consecutively, rather than through genuine analysis of the specific stock and/or market. This leads to the investor gaining a false sense of hope, and becoming delusional about the fact that they believe they are genuinely good at investing. In the long run, this will hurt the investor as they will believe their "strategy" is actually good, and they will continue to follow this strategy, and in turn, lose money. In Figure 1, data collected by Da Silva (2019) compares a perfect confidence-accuracy line to an actual dataset of real people demonstrating how confident they were compared to how accurate they were. The straight, orange line on the graph represents the 45° identity line, or in other words, the "perfect calibration line". This line represents a perfect scenario, where if a person is 60% confident, then they are 60% correct. The red, curvy line represents the overconfidence bias, showing that if someone were to be 80% "confident" that they were right, in reality, they were just about 55% accurate. This connects directly back to overconfidence bias in investing because it proves that gut feelings of certainty cannot be trusted, and investors who feel "very confident" about a trade are often the ones most likely to be wrong.

Figure 1 Confidence vs. Accuracy in Investor Decision-Making



Note. This graph illustrates the relationship between investor confidence and decision accuracy, showing that confidence often exceeds actual correctness. Adapted from *The Influence of Psychotropic Drugs on Investor Overconfidence*, by Marcia Zindel, Gabriel Deltrejo Ribeiro, Carlos H. Rocha, and Sergio Da Silva 2019, Open Access Library Journal, 3, 3. https://www.scirp.org/journal/paperinformation?paperid=92908

One of the most studied cognitive biases in behavioral finance is loss aversion, where an investor experiences a stronger emotional response to losses than to equivalent gains. Research has consistently shown that this asymmetry in emotional response drives irrational decision-making across two distinct behavioral patterns. The first is the tendency for investors to hold onto losing stocks for far too long, driven by the hope that the stock will eventually recover and turn a profit. Rather than cutting their losses early, these investors remain emotionally attached to the position, and in doing so, end up losing significantly more money than they originally would have. The second behavioral pattern caused by loss aversion is the premature selling of winning stocks. Out of fear that a currently profitable stock may reverse course and result in a loss, investors will exit their position too early, ultimately leaving significant gains on the table. A study by Bouteska and Regaieg (2020) examined this exact issue on a large scale, analyzing 154 companies using 6,777 quarterly observations of US industrial and services companies over the period from 2006 to 2016. The researchers wanted to see if loss aversion could actually show up in the broader economic performance of entire companies and industries. What they found was that loss aversion bias negatively affects the economic performance of companies across both the industrial and services sectors. This means that the irrational, fear-driven decisions being made by countless individual investors weren't just hurting those individuals personally, they were collectively impactful enough to influence real economic outcomes at the company level. In other words, this wasn't just a small-scale psychological quirk affecting a handful of people, it was a widespread pattern with measurable consequences across an entire decade of market data. In the long run, this imbalance that loss aversion brings will consistently hurt the investor's portfolio, as decisions are being driven by emotion and fear rather than rational analysis of the stock market. As this study shows, the ripple effects of these emotionally driven decisions extend far beyond any one person's account, shaping the performance of companies and industries as a whole.

Another bias is confirmation bias, which is the tendency for investors to actively seek out and prioritize information that supports their pre-existing beliefs about a stock or market, while simultaneously ignoring or dismissing information that contradicts their viewpoint. A study by Ko and Huang (2010) took a closer look at how confirmation bias actually plays out among real investors. They found that when investors already had a positive outlook on a stock, they were far more likely to seek out and trust information that supported that view, even when there were clear warning signs pointing in the opposite direction. What made this especially interesting is that the investors in the study weren't doing this on purpose. Most of them genuinely believed they were making well-informed, balanced decisions, when in reality they were unconsciously filtering out anything that challenged their existing opinion. This is what makes confirmation bias so dangerous compared to other biases, because at least with overconfidence or loss aversion, the effects can sometimes be noticed after the fact. With confirmation bias, the investor never even sees the full picture to begin with, meaning they can feel completely justified in a bad decision right up until it goes wrong. This matters in investing because the stock market is constantly throwing out new information, and being able to update your views based on that

information is one of the most important skills an investor can have. When confirmation bias takes over, that ability essentially shuts down. The investor stops learning and starts just looking for reasons to feel right, which over time leads to poor decisions, missed warning signs, and avoidable losses that could have been prevented if they had simply been willing to consider the other side.

This selective consumption of information ultimately leads to poorly informed investment decisions, as the investor is essentially operating within an echo chamber of their own making. In the long run, this will hurt the investor as they will continue to double down on positions that may be fundamentally flawed, blinded by the one-sided information they have chosen to consume, and in turn, suffer significant financial losses that could have otherwise been avoided through a more balanced and objective analysis of the available data.

Another bias that affects investors is the anchoring bias, which is the tendency for an investor to over-rely on a specific piece of information, such as the price they originally bought a stock for, and use that as the main reference point for every decision they make going forward. Rather than evaluating a stock based on where it stands today and where it is realistically headed, the investor keeps mentally tying everything back to that original number. This causes them to make decisions based on information that is no longer relevant to the current state of the market, which, more often than not, leads to poor and unprofitable outcomes. For example, an investor buys a stock at $100 per share. The stock drops to $60, but based on current market conditions and analyst data, the realistic future value of that stock is closer to $50. A rational investor would look at that information and either cut their losses or avoid adding to their position, but an investor experiencing anchoring bias can't move past that original $100 price tag. In their mind, the stock needs to get back to $100 before they should do anything, so they hold on and even buy more, convinced it will bounce back simply because that's what they paid for it. The anchor has nothing to do with the stock's actual value anymore, but it completely controls how the investor thinks and acts. According to SmartAsset (2024), this bias is incredibly common and often goes unnoticed because anchoring feels like logical thinking on the surface. Basing decisions on a concrete number feels rational, but the problem is that the number being used as a reference point is outdated and emotionally driven rather than grounded in current market reality. This makes anchoring bias particularly damaging over time, because the investor isn't just making one bad decision, they are repeatedly making decisions through a distorted lens, compounding their losses while convinced they are simply waiting for things to go back to "normal."

Studies also show that some people view investing as a form of gambling rather than a long-term strategy. A report from A study published in the *Journal of Economic Psychology* by Shang, Duan, and Lu (2021) indicates that this has led investors to not fully comprehend the risks involved, causing them to become overly concerned with short-term gains rather than focusing on the bigger picture. When an investor views investing as a form of gambling, they become overly emotional about the process rather than composing and following a disciplined strategy. This is also a form of loss aversion, as they grow more preoccupied with avoiding losses than with making sound decisions for the future. The result is a pattern of irrational

choices, negative financial outcomes, and a narrower scope, in terms of not having when it comes to long-term planning. However, the fundamental difference between gambling and investing is the ability to control probability. In gambling, no strategy can change the outcome of a given bet. The statistics are fixed against the player from the start, and the house always maintains an edge. In roulette, for example, the green square exists to tilt the odds in the house's favor. On a standard American roulette wheel, there are 38 possible outcomes, but a winning bet on red or black only pays out at 1:1. This gives the house roughly a 5.26% edge on every spin. No amount of strategy or pattern recognition changes that edge. The probabilities are fixed and the outcomes are random. Investing operates on a different foundation. While markets carry risk and short-term outcomes can be unpredictable, investors have access to tools, data, and strategies that influence their probability of success over time. Diversifying a portfolio, analyzing earnings reports, understanding sector performance, and managing risk through asset allocation are all ways an investor can actively improve their odds. Unlike roulette, the market is not designed to make the investor lose. Historically, the broader market has trended upward over long periods of time, rewarding those who approach it with patience and a clear strategy. When investors begin treating their portfolio like a casino table, chasing quick wins, reacting emotionally to every swing, and abandoning their strategy at the first sign of a loss, they remove the advantages that separate investing from gambling entirely.

Discussion

Each and every bias there is are connected to each other in some way or form, however, because of this, investors can practice combatting one bias and end up fighting multiple biases at the same time. This happens because all of these biases share a common root: they are all shortcuts the brain takes when making decisions.

Confirmation bias is dangerous because investors naturally gravitate towards information that already agrees with what they believe, which causes them to miss warning signs or red flags about the trade. To combat confirmation bias, investors should build a rule into their strategy that is to seek out opposing information before making any final decisions. This means that before making a trade, investors should find and understand one piece of information that may lead to the failure of the trade. By implementing this into the strategy, investors become unbiased, and begin to understand the full picture. Over time, this practice rewires how investors approach trading stocks, as it makes it harder for confirmation bias to influence their own decisions before making a trade.

To combat **overconfidence bias**, the investor should set strict rules on how much they are willing to risk on each trade, and never go beyond that limit no matter how certain they are about the outcome. Overconfidence bias is dangerous in the sense that the more success they have based on their own judgement, the more they begin to trust their judgement over data, which leads to bigger and riskier investments. By sticking to a strategy that caps risk and requires concrete evidence/proof of success, the investor is no longer investing based on gut feeling and pure judgment, but on what the numbers and data actually say. Overall, this strategy keeps

overconfidence from taking over by keeping emotions and assumptions out of the picture, and purely relying on strategy and evidence for each and every trade.

To combat these biases on a deeper level, investors should make it a habit to seek out information that challenges their current views before making any decisions. This means actively looking for reasons why an investment might fail, not just reasons why it should succeed. Doing this directly combats confirmation bias, but it also weakens anchoring bias as well, because the investor is now forced to consider a wider range of information, rather than just information that feels comforting instead.

Another way to fight these biases is to follow a set strategy every single time, rather than rely on gut feelings. This means deciding ahead of time what a buy or sell would look like for that individual stock, only investing when those conditions are met, and never breaking those rules no matter how something feels in the moment. When investors follow a strict strategy, they remove a lot of room biases needed to grow, because decisions are already made before emotions and assumptions even have time to get involved, and are long-term proven to grow in wealth. Over time, sticking to a strict strategy helps combat against all biases, because no emotions and/or assumptions are being considered at all.

While these strategies can reduce the influence of biases like confirmation bias and overconfidence, they cannot fully eliminate them. Behavioral finance research suggests that cognitive biases are deeply rooted and can resurface under stress or high-stakes situations, even in disciplined investors. A strict strategy can help in calm market conditions, but during periods of sharp volatility or unexpected news, the same emotional pressures that create these biases in the first place can also be strong enough to break discipline down. This does not mean these strategies are ineffective; even a partial reduction in biased decision-making can meaningfully improve long-term investment outcomes. However, it does suggest that discipline alone may not be enough, and that investors may benefit from additional structural safeguards, such as predetermined stop-loss orders or outside accountability, to reinforce their strategy when willpower is most likely to fail.

Conclusion

Investing is often oversimplified and seen as something as simple as “buy low, sell high”, however, it is much more complicated than that. External factors like interest rates, inflation, and market volatility all put pressure on investors. However, internal factors, the emotions and biases built into the human brain, often matter even more when it comes to investment outcomes. Overconfidence bias makes investors think they are better at predicting the market than they really are. Loss aversion causes investors to hold onto losing stocks for too long while selling winning stocks too early. Confirmation bias makes investors only pay attention to information that agrees with what they already believe. Anchoring bias keeps investors focused on old prices instead of looking at what a stock is actually worth today. These biases are especially dangerous because they do not act alone. Since they all come from the same problem, the brain taking shortcuts under pressure, fighting one bias often means fighting several at the same time.

However, by building simple habits, investors can fight these biases. These habits, such as looking for information that challenges their own beliefs, setting strict limits on risk, and sticking to a plan instead of reacting emotionally, can reduce how much control these biases have over investor decisions. The difference between a disciplined investor and someone who treats the market like a casino comes down to preparation and self-awareness. Unlike gambling, where the odds are always against the player, investing rewards people who understand their own thinking and build habits to protect against their own mistakes. As more everyday people start investing, understanding how outside pressures and inside emotions work together will only become more important, both for individual investors and for the market as a whole. Moving forward, greater emphasis on financial literacy and behavioral awareness, whether through investor education programs, employer-sponsored resources, or simply personal research, could help new investors recognize these biases before they take hold. As access to investing continues to expand, teaching people not just what to invest in, but how their own minds can work against them, may prove just as important as any market strategy.

Works Cited

Arnsten, Amy F. T. "Stress Signalling Pathways That Impair Prefrontal Cortex Structure and Function." Nature Reviews Neuroscience 10 (2009): 410-22. https://doi.org/10.1038/nrn2648.

Atif Sattar, Muhammad, Muhammad Toseef, and Muhammad Fahad Sattar. "Behavioral Finance Biases in Investment Decision Making." International Journal of Accounting, Finance and Risk Management 5, no. 2 (2020): 69-75. https://doi.org/10.11648/j.ijafrm.20200502.11.

Bouteska, Ahmed, and Boutheina Regaieg. "Loss Aversion, Overconfidence of Investors and Their Impact on Market Performance Evidence from the US Stock Markets." Journal of Economics, Finance and Administrative Science 25, no. 50 (2020): 451-78. https://doi.org/10.1108/JEFAS-07-2017-0081.

Khare, Trupti, and Sujata Kapoor. "Behavioral Biases and the Rational Decision-Making Process of Financial Professionals: Significant Factors That Determine the Future of the Financial Market." Journal of Advances in Management Research 21, no. 1 (2024): 44-65. https://doi.org/10.1108/JAMR-03-2023-0086.

Ko, K. Jeremy, and Zhijian Huang. "Persistence of Beliefs in an Investment Experiment." Social Science Research Network. Last modified February 15, 2010. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1571689.

Shang, Xuesong, Hebing Duan, and Jingyi Lu. "Gambling versus Investment: Lay Theory and Loss Aversion." Journal of Economic Psychology 84 (2021): article 102367. https://doi.org/10.1016/j.joep.2021.102367.

SmartAsset. "What Is Anchoring Bias in Investing?" Last modified August 2, 2024. https://smartasset.com/investing/anchoring-bias-in-investing.

Zindel, Marcia, Gabriel Deltrejo Ribeiro, Carlos H. Rocha, and Sergio Da Silva. "The Influence of Psychotropic Drugs on Investor Overconfidence." Open Access Library Journal 3 (2019): 3. https://www.scirp.org/journal/paperinformation?paperid=92908.