Boston University Undergraduate Economics Journal

Spring 2026 Issue V

UNDERGRADUATE ECONOMICS JOURNAL

2024

UNIVERSITAS BOSTONIENSIS : CONDITA MDCCCXXXIX :

The content published in this undergraduate economics journal at Boston University is solely intended for educational and informational purposes. The views and opinions expressed in the articles, essays, and other submissions belong to their respective authors and do not necessarily reflect the views of Boston University, its faculty, staff, or administration. The accuracy and completeness of the information presented within these pages are not guaranteed, and readers are encouraged to conduct further research and consult professional advisors as needed.

Meet the Team

Spring 2026 Issue IV

Ellie Wang
Editor-in-Chief

Ajay Raman
Copy Editor

Nandini Jha
Head Senior Editor

Matthew Morris
Head Layout Editor

Lina Santos
Senior Editor

Dahsom Lim
Senior Editor

Cole Pinto
Senior Editor

Mohammad Aiman Ashad
Senior Editor

Nicole Sim
Senior Editor

Sara Asdhir
Junior Editor

Madi Kim
Junior Editor

Abigail Lam
Layout Editor

BUUEJ Editing Team Spring 2026

Boston University Undergraduate Economics Journal

Table of Contents

Articles

A Time Series Analysis of U.S. Housing Starts

By Yi Li

This paper uses monthly data from January 1985 to June 2025 to examine the dynamics and predictability of U.S. housing starts. On the univariate side, ARIMA and ARFIMA models are augmented with asymmetric GARCH specifications to account for volatility clustering, leverage effects, and potential long-memory behavior. The optimal orders for the models are ARMA(0,1) and GARCH(1,1). Complementing the univariate analysis, the paper estimates a multivariate VAR model incorporating housing starts, residential construction employment, builder sentiment, and related housing-market indicators. The VAR(4) specification shows that mortgage rates remain exogenous in the system, and that residential construction employment has a significant positive effect on sentiment that lasts for two months.

I. Introduction

Total housing starts are a key housing market indicator that measures the annualized pace of new residential construction initiated within a given month, encompassing both single-family and multifamily units. Because residential construction is closely linked to the broader economic activity through channels such as employment, household wealth, and credit demand, trends in housing starts can offer an important gauge of macroeconomic momentum. This relationship is particularly salient in the aftermath of major economic disruptions, such as the 2007–2008 Subprime Mortgage Crisis or the 2020 COVID-19 pandemic, when shifts in housing market dynamics can signal emerging patterns of economic recovery or deterioration.

Indeed, **?** argues that housing investment-housing starts and their changes in particular-consistently leads business cycle turning points. Motivated by this predictive import, the paper aims to employ time series methodologies to uncover the behavior of housing starts and, in turn, their implications for the housing market and the macroeconomy.

The structure of the paper is as follows. Section 2 reviews the existing literature on the housing market and related time series econometric approaches. Section 3 provides an overview of the data and their selection process. Section 4 goes through the econometric methodology used, including preliminary tests like unit roots and structural breaks. Section 5 discusses the limitations of this paper and future improvements.

II. Literature Review

Dua et al. [1999] use Bayesian Vector Autoregressions (BVARs) to forecast U.S. home sales and to evaluate whether different leading indicators improve forecast accuracy. They find that the model, including building permits, consistently gives the most accurate forecasts, and housing starts are a close substitute. However, the approach also has several limitations. Estimating the VAR in levels raises concerns about nonstationarity and long-run interpretation. Moreover, while the results are informative for prediction, the framework remains largely atheoretical, offering limited insight into the structural economic mechanisms underlying housing market dynamics.

Fullerton Jr and West [1998] evaluate the forecast accuracy of regional housing start models by comparing structural econometric forecasts for Florida and its six largest metropolitan areas with univariate ARIMA models and random walk benchmarks, using quarterly forecasts over the period 1986–1995. A key contribution of the paper is its explicit focus on regional and metropolitan-level forecast evaluation, a dimension that has received substantially less attention than national housing market forecasting in the existing literature. As housing markets have become increasingly segmented across the country, the paper highlights the importance of regional perspectives for understanding housing dynamics and assessing the practical usefulness of forecasting models.

Parallel to developments in econometric modeling, a growing strand of literature explores machine learning techniques for forecasting housing-related variables. Studies that rely exclusively on econometric approaches may suffer from misspecification when volatility clustering, nonlinearity, or regime changes are present, whereas purely machine-learning forecasts often overlook time-series properties such as unit roots, cointegration, and conditional heteroskedasticity. Joshi [2019] uses monthly U.S. housing starts data from 1976 to 2018 to systematically compare a broad set of econometric models with several machine learning approaches. The study finds that machine learning models substantially outperform traditional time-series methods in terms of out-of-sample forecast accuracy. However, similar to Dua et al. [1999], the analysis does not explicitly account for structural breaks, which may affect both model estimation and forecast performance in a long sample spanning multiple housing cycles and policy regimes.

Complementing existing literature, this paper conducts a careful assessment of unit roots, structural breaks, and seasonality before proceeding to modeling. This ensures that the models are based on economically sound procedures rather than pure assumptions. The paper models housing starts as both a univariate and multivariate process. The univariate forecasting employs hybrid ARIMA–GARCH and ARFIMA–GARCH specifications to capture persistence, asymmetric volatility, and long-memory behavior. The multivariate VAR in the first difference examines dynamic interactions and shock transmission across key housing-market indicators in a manner consistent with their underlying time series properties.

III. Data Description

We use monthly data of all variables from various sources in the range 1985-2025. The variables are: Total Housing Starts, Residential Construction Jobs, Completed New Homes, 30-Year Fixed-Rate Mortgage, Housing Affordability Index, and Housing Market Index.[1] Some of the data are transformed while others are not. From here onward, lTotalHStarts refers to the natural logarithm of Total Housing Starts, lResConJobs refers to that of Residential Construction Jobs, and lComNewHomes refers to that of Completed New Homes. FRM refers to a 30-Year Fixed-Rate Mortgage, HAI refers to the Housing Affordability Index, and HMI refers to Housing Market Index. All of the data obtained are not seasonally adjusted.

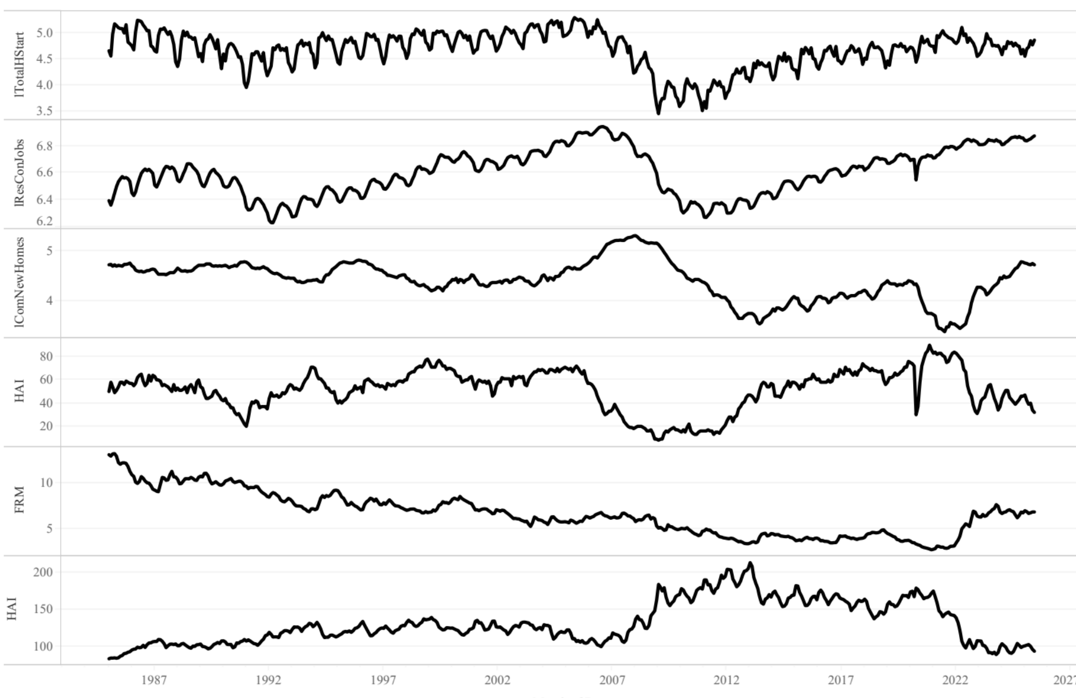


Figure 1: Comovements between Variables

From Figure 1, each variable seems to have at least one structural break, necessitating the need of structural break tests. lTotalHStarts and lResConJobs also seem to exhibit seasonality, which needs to be adjusted after taking unit root and structural break tests. In addition, the top three panels display clear comovement, motivating cointegration tests to determine whether a VECM specification is warranted.

[1]A more detailed overview of the data can be found in Table 2 of the appendix.

IV. Econometric Methodology

A. Testing for Unit Roots

As seen in Figure 1, all of the data seem to have either a deterministic or a stochastic trend, indicating that the series are not stationary. A unit root test is then required to determine the source of this non-stationarity because it can inform us on whether the shocks are transitory (for deterministic trend) or permanent (for stochastic trend). The four common types of tests are: the Augmented Dickey-Fuller (ADF) test, the Dickey-Fuller Generalized Least Squares (DF-GLS) test, the Phillips-Perron (PP) test, and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test. Each test has its strengths and weaknesses. For instance, the ADF test accounts for autocorrelation of the error term when the original DF test ignores it. However, adding lagged differences not only changes the regression, but also leads to loss of degree of freedom. The PP test keeps the original regression but replaces OLS standard errors with Heteroskedasticity and Autocorrelation Consistent standard errors. This solves the degree of freedom problem, but the PP test suffers severe size distortion when there is a large negative moving average unit root. As such, all four tests are implemented to ensure a more comprehensive analysis.

The results of the unit root tests are summarized in Table 3. For $lTotalHStarts$, only the PP test indicates stationarity, while the remaining tests provide insufficient evidence to reject the unit root null at the 5% significance level. For lResconJobs, the DF-GLS test rejects the unit root hypothesis at the 5% level, but the rest of the tests fail to do so. For lComNewHomes, both ADF and DF-GLS strongly reject the null hypothesis. By contrast, for FRM, HAI, and HMI, all tests unanimously fail to reject the unit root null. To make the final verdict, we need to take into account structural breaks, as all series seem to have at least one structural break based on visual inspection. Perron [1989] shows that when there is a one-time break in the data generating process, standard unit root tests have little power to reject the null hypothesis. Therefore, it is necessary to conduct the structural break test before differencing.

B. Unit Root Tests with Structural Breaks – Model AO-A

Following Perron's additive outlier (AO) framework, we allow for a one-time level shift at an unknown break date. The series is first detrended by regressing it on a linear time trend and a break dummy. The detrended residuals are then used in an augmented Dickey–Fuller regression that includes break dummies to test for a unit root. The equation is as follows:

$$\tilde{y}_t = \alpha\tilde{y}_{t-1} + \sum_{j=0}^{k} d_j D(T_B) + \sum_{i=1}^{k} \gamma_i y + \epsilon_t, \tag{1}$$

where

$$D(T_B)_t = \begin{cases} 1, & t = T_B + 1 \\ 0, & \text{otherwise} \end{cases}$$

T_B is the break date. DU_t is the level dummy used in the detrending equation. $D(T_B)_t$ is the impulse dummy that captures the one-time AO shock at the break. Its lags $D(T_B)$ are included because a break in the AO model creates an instantaneous shift in the observation, and we must partial out the effect of that shock from the autoregressive regression. Otherwise, the ADF regression would treat the break as a unit-root effect and would be biased towards non-rejection. k represents the ADF augmentation to remove short-run serial correlation.

There are two variables that appear to have a break in the level: lTotalHStarts and lResConJobs. The t-statistic is computed as $t = \frac{\hat{\alpha}-1}{se(\hat{\alpha})}$ for the coefficient on the first lag of the regression residual. The resulting test statistic is then compared with the critical values reported in Table IV.B of Perron [1989]. The results are all summarized in Table 5. Based on these results, neither variable provides sufficient evidence to reject the unit root null, implying that differencing is appropriate.

C. Unit Root Tests with Structural Breaks – Model AO-B & AO-C

The Zivot-Andrews test is biased and has severe size distortion when there is a break in the trend of the series. Therefore, we should avoid using the Zivot-Andrews test for Model AO-B (break in the trend of a trending series) and AO-C (break in both the level and the trend of a trending series). As a result, the Perron and Yabu [2009] test is used to determine the break date, and the Perron and Yabu [2009] test is used for the unit root test. The Perron and Yabu [2009] test improves upon Perron [1989], which treats the break as exogenous, by allowing for an unknown break date. This addresses the problem of endogeneity and potential misspecification of the break date. The Perron and Yabu [2009] test allows for a break under both the null and the alternative hypotheses, which solves the problem of size distortion when there is a break in the null hypothesis.

The results of the tests are also summarized in Table 5. lTotalHStarts and lResConJobs are also tested using the AO-C model, as the trends of both series exhibit slight changes following the structural break, even though these changes are not particularly pronounced. The criteria for lag selection are based on the Modified Akaike information criterion (MAIC) and the Modified Bayesian information criterion (MBIC). Both criteria are used to provide a more comprehensive picture of the test. Combining the evidence from conventional unit root tests and unit root tests allowing for structural breaks, we conclude that lTotalHStarts, lResConJobs, and HMI contain unit roots and therefore require differencing. Despite the fact that both ADF and DF-GLS strongly reject the null for lComNewHomes, greater weight is placed on the structural break test results.

For FRM, conventional unit root tests suggest the presence of a unit root, whereas the AO-B model indicates trend stationarity. Given that standard tests tend to interpret changes in trend as permanent shifts and are biased toward non-rejection of the null, the structural break test results are again preferred. Finally, for HAI, the MAIC and MBIC criteria lead to conflicting conclusions: MAIC supports a unit root, while MBIC suggests trend stationarity. In light of this ambiguity, a common approach would be to estimate two alternative specifications, one using the differenced series and the other using a detrended series, and assess their relative performance. Yet, after detrending the seasonally adjusted HAI, the ACF displays very slow decay, indicative of near–unit-root persistence, while the PACF exhibits a pronounced spike at one for the first lag. Together, these patterns suggest that the series may still contain a unit root. This interpretation is reinforced by formal unit root tests, which fail to reject the null hypothesis. Consequently, we proceed by first-differencing the series rather than treating it as trend-stationary.

D. Seasonality and Seasonal Adjustment

Many time series data, macroeconomic indicators in the housing market in particular, exhibit seasonal patterns over the year. Removing the seasonal noise allows for a clearer view of the long-term trend and meaningful comparisons between consecutive periods. However, it should be noted that the series should be seasonally adjusted after unit root and structural break tests. Perron and Vogelsang [1993] show that seasonal adjustment can alter the sampling distribution of unit root test statistics and lead to size distortion or reduction in power. Before deseasonalizing, the ACF and PACF of the series are plotted to confirm seasonality, and the results indicate multiplicative models with seasonal patterns.

To handle seasonality, a common method is to use the X-13ARIMA-SEATS (hereafter X-13), which is a software developed by the U.S. Census Bureau. The X-13 offers two alternative approaches to deseasonalization. One is through the X-11 program, which decomposes the series, either additively or multiplicatively, into trend, seasonal, and irregular components. It applies weighted moving averages to smooth out seasonal patterns without relying on any statistical models. The other approach is the Signal Extraction in ARIMA Time Series (SEATS) method, which uses an estimated ARIMA model to statistically extract the underlying seasonal components of the series. When the ARIMA model is correctly specified, SEATS has the advantage of accommodating evolving seasonal patterns more effectively.

The X-11 filter-based program is the most widely adopted procedure, as it is the product of decades of research by scholars in the National Bureau of Economic Research and has been launched by the Census since the 1960s. Nonetheless, the SEATS method is preferred in this paper because (1) it produces a lower Q-statistic, indicating lower residual autocorrelation and (2) the distribution shape of the histogram is more symmetric[2].

E. ARIMA-E-GARCH vs. ARFIMA-E-GARCH

After seasonal adjustment, we initially implemented the ARIMA(0,1,1) model, since it has the lowest AIC and BIC among other alternatives. Using the Ljung-Box Test, we find that the squared residual of the model rejects the white noise hypothesis and hence indicates an ARCH effect[3]. This allows for estimation of the GARCH model, which is oftentimes preferred because it is a more parsimonious model that requires fewer lags than the ARCH one. Below are the information criteria for all possible GARCH models up to order 2.

Table 1: Information Criteria for all Estimated GARCH(p,q) Models

p	q	AIC	BIC
1	0	-1211.47	-1198.91
1	1	-1229.31	-1212.58
1	2	-1229.26	-1208.34
2	0	-1225.81	-1209.07
2	1	-1228.28	-1207.36
2	2	-1227.20	-1202.09

Information criteria strongly favor a GARCH(1,1) model, leading to an ARIMA(0,1,1)–GARCH(1,1) bench mark model. Diagnostic tests on standardized squared residuals suggest that the model effectively accounts for conditional heteroskedasticity. Nevertheless, the model delivers relatively weak in-sample fit. To reconcile this discrepancy, we test for asymmetric volatility and find strong evidence of leverage effects, whereby negative shocks generate larger increases in volatility than positive shocks. The equation is:

$$s_t^2 = a_0 + a_1 D_1 + a_2 D_1 e_1 + a_3(1 - D_1)e_1 + u_t \tag{2}$$

where

$$D_1 = \begin{cases} 1, & \text{if } e < 0, \\ 0, & \text{otherwise,} \end{cases}$$

Engle-Ng **?** developed this sign-bias test to check whether the variance equation correctly captures asymmetry in volatility. D_1 is a dummy for negative shocks, essentially capturing the sign of the bias. The interaction term between D_1 and e_1 reflects the size of the negative bias, and that between $1 - D_1$ and e_1 shows the size of the positive bias. The housing market typically manifests leverage effects, and the result of the regression corroborates the claim. The interaction term of the negative bias has p-value of 0.00, which is more significant than that of the positive bias with p-value close to 0.04. In addition, the coefficient rejects the null hypothesis. This regression formally proves that volatility in housing starts responds asymmetrically to shocks, so a symmetric GARCH model is misspecified, motivating the use of asymmetric volatility models such as E-GARCH

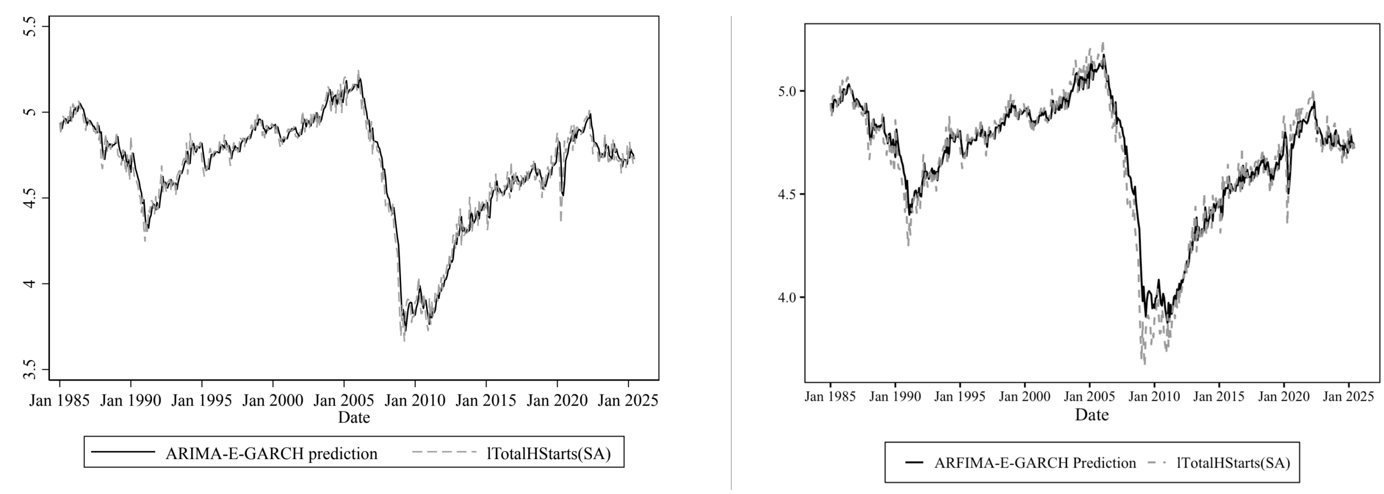
[2]The SARIMA model was also used to remove seasonal patterns, but it does so poorly because there is remaining seasonality. Both the X-11 and SEATS methods completely remove seasonality.

[3]Lags 5, 10, 20, and 40 are tested, all of which have p-values of 0.

or GJR-GARCH.[4]

The ARIMA-E-GARCH and the ARIMA-GJR-GARCH models are tested, and the results are summarized in Table 6. Based on the information criteria, both AIC and BIC favor the ARIMA-E-GARCH model. Post-estimation diagnostic tests show that there is no remaining linear serial correlation in the mean equation and that there is no remaining conditional heteroskedasticity. In addition to the traditional ARIMA-type model, we also experiment with the AutoRegressive Fractionally Integrated Moving Average (ARFIMA) model. ARIMA models can only capture short-range dependence, whereas the impacts of an event tend to persist over a long time in the housing market. The ARFIMA model, on the other hand, is able to capture long-term dependence. In addition, the ACF of the seasonally adjusted series exhibits a very slow hyperbolic decay.[5]

Using the same orders of AR(0), MA(1), ARCH(1), and GARCH(1) as the ARIMA-GARCH-type models, we test ARFIMA-E-GARCH and ARFIMA-GJR-GARCH models. The results are also included in Table 6. ARFIMA-E-GARCH seems to be better than ARFIMA-GJR-GARCH because the leverage effect parameter is highly significant at the 1% level for the former, while that is insignificant for the latter. The remaining parameters do not display materially different levels of statistical significance. The figure below illustrates the fitted values implied by each model alongside the actual observed values.



(a) ARFIMA-E-GARCH Mode (b) ARIMA-E-GARCH Model

Figure 2: Comparison Plots of Fitted vs. Actual Values

The results indicate that the ARIMA–E–GARCH model provides a noticeably better in-sample fit than the ARFIMA–E–GARCH specification. However, the strong in-sample fit of the ARIMA-E-GARCH model should be interpreted with caution, as it may reflect overfitting and thus weaker out-of-sample performance.

A univariate ARIMA-GARCH framework treats housing starts as an isolated process and therefore cannot capture the dynamic transmission of shocks across related housing-market indicators. A VAR model, on the other hand, addresses this limitation by jointly modeling multiple endogenous variables. This multi-variate structure enables the analysis of dynamic interdependencies, lead–lag relationships, and system-wide responses to structural shocks, which are central to understanding housing market dynamics.

[4] The SARIMA model was also used to remove seasonal patterns, but it does so poorly because there is remaining seasonality. Both the X-11 and SEATS methods completely remove seasonality.

[5] Another motivation for exploring the ARFIMA model is that not only does the MA coefficient of the ARIMA-E-GARCH exhibits a relatively negative value of -0.47, but the ACF and PACF plots of the first-differenced, seasonally adjusted series show occasional negative spikes at different lags. Although these are not prima facie evidence of over-differencing, they do hint at a mild over-differencing. The optimal order of fractional differencing suggested by the rugarch library in R is close to 0.5, which also seems to hint at the possibility of fractional differencing.

F. Vector Autoregression (VAR) Model

A univariate ARIMA-GARCH framework treats housing starts as an isolated process and therefore cannot capture the dynamic transmission of shocks across related housing-market indicators. A VAR model. on the other hand, addresses this limitation by jointly modeling multiple endogenous variables. This multi-variate structure enables the analysis of dynamic interdependencies, lead–lag relationships, and system-wide responses to structural shocks, which are central to understanding housing market dynamics.

VAR Model Diagnostic Tests The Hannan and Quinn information criterion (HQIC) and the Schwarz's Bayesian information criterion (SBIC) both select a model with two lags. When all variables are included in the VAR model, HAI appears to play a limited role in the system. Apart from a marginal Granger-causality effect on HMI at the 10% level, HAI does not significantly increase the predictive accuracy of other variables in the system. Accordingly, to improve parsimony, we exclude HAI from the VAR and re-estimate the model with the remaining variables. SBIC, HQIC, and AIC choose lag 2, 3, and 4 respectively. Although SBIC favors a more parsimonious specification, we still adopt a VAR(4) model, as it adequately captures the dynamic structure of the monthly data while mitigating the risk of underfitting.[6]

The VAR(4) specification is evaluated using a series of standard diagnostic tests. First, Lagrange–multiplier tests for residual autocorrelation indicate that the VAR(4) substantially improves upon lower-order specifications. The null hypothesis of no residual autocorrelation cannot be rejected at the first lag, while only marginal evidence against the null is found at the second lag. Overall, the VAR(4) captures the serial dependence in the data reasonably well, particularly relative to the VAR(3), which exhibits strong residual autocorrelation. Second, Wald lag-exclusion tests strongly reject the joint exclusion of each lag from the system, indicating that all four lags are jointly relevant for the VAR as a whole. While the importance of higher-order lags varies across individual equations, lag four plays a meaningful role in key equations such as housing starts and builder sentiment. Finally, normality tests based on the Jarque–Bera statistic reject the null of Gaussian residuals for all equations. Further decomposition reveals significant skewness and excess kurtosis, especially for construction employment. Such departures from normality are common in macroeconomic and housing data and likely reflect large asymmetric shocks during crisis periods.

Before moving on to further analysis, the interpretation of the regression results can be helpful in understanding the economic context of the model. The results can be found in Table 7. Housing starts exhibit stronginternal persistence, with lags two through four remaining positive and highly significant, indicating that shocks to housing activity propagate gradually rather than dissipating immediately. This slow adjustment is consistent with the institutional features of residential construction, including planning, financing, and regulatory delays. Housing starts also exert a large and positive effect on residential construction employment at short horizons, implying that builders adjust labor demand in response to changes in housing activity rather than anticipating them. The housing market index responds particularly strongly to housing starts at the first lag, suggesting that builder sentiment is revised upward following realized strength in housing activity. This reflects confidence reinforcement based on observed market conditions instead of forward-looking expectations alone. In contrast, completed new homes do not respond significantly to housing starts within the VAR horizon, which is economically plausible given the lengthy construction process. Finally, mortgage rates remain largely exogenous to housing-sector variables, consistent with their determination by broader financial and monetary conditions.

Granger Casuality The results of the Granger causality are presented in Table 8. For housing starts, only construction jobs granger-causes it at the 5% level, yet the system jointly granger-causes housing starts at the

[6] We also experiment with both the VAR(2) and VAR(3) model, but they fail the normality diagnostics. The null is also strongly rejected at both lag 1 and lag 2 using the LM test, meaning that there are significant residual autocorrelations remaining in both models.

1% level. HMI seems to be endogenous, as all variables in the system granger-cause it. Construction jobs are also highly endogenous because only completed new homes don't granger-cause it. The rest all granger-causes it at the 1% level. Although mortgage rates are mostly exogenous, including lags of housing starts do improve the predictive accuracy of mortgage rates. Based on the results, the order of variables from most exogenous to least exogenous is: FRM → Housing Starts → Residential Construction Jobs → Completed New Homes for Sale → HMI. The order of the variables can then be used to construct OIRF.

Orthogonalized Impulse Response Function When shocks in an economy impact the variables contemporaneously, the standard IRF provides very little information about the impacts. Thus, OIRF is used to keep track of the impact of exogenous changes from one variable to another. Figure 3 illustrates all OIRFs with meaningful impacts. The top-left panel shows that a contractionary mortgage-rate shock leads to an immediate negative response of sentiment. The impact does not persist and gradually reverts to zero. The top-right panel shows that a standard deviation shock in housing completions does not strongly move sentiment. The effects are transitory. However, for the bottom-left panel, a standard deviation shock in employment immediately leads to a 1.8 percentage boost in builder sentiment, and there is a sharp decline in about two months. Lastly, one standard deviation shock to housing starts initially raises 0.5 percentage sentiment, followed by a delayed negative correction, and eventually converges to zero.

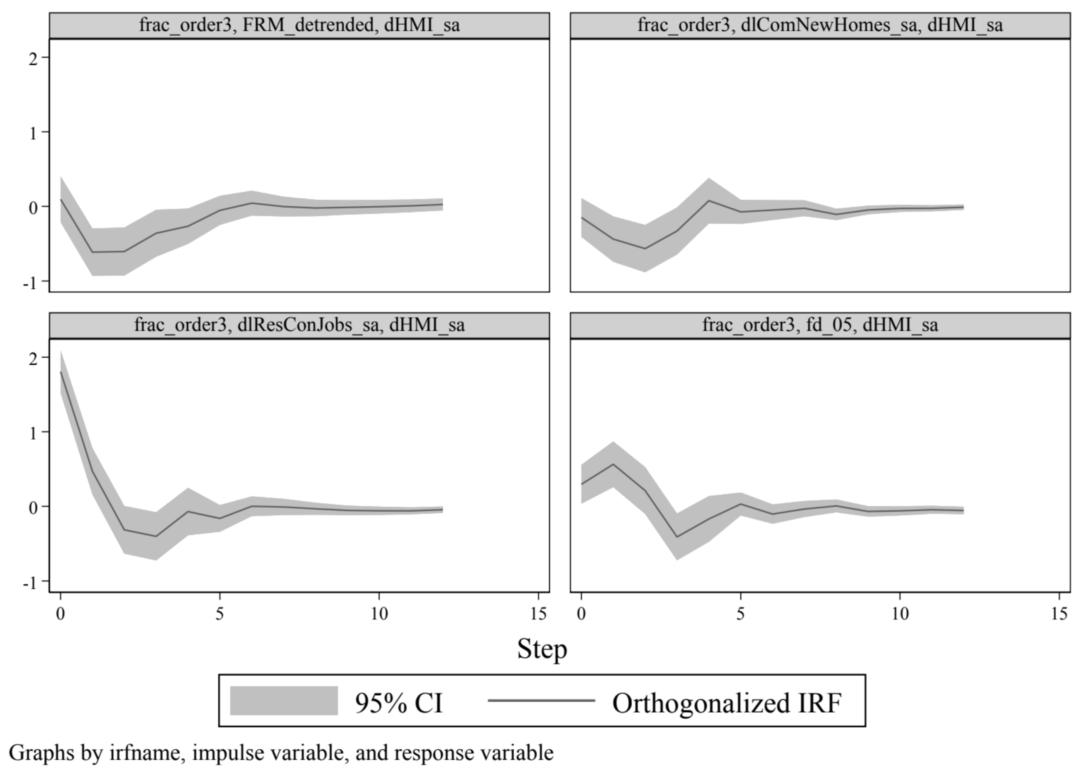


Figure 3: OIRF

Forecast Error Variance Decomposition The forecast error variance decomposition (FEVD) explains the amount of variation in one variable that is attributable to shocks in other variables in the system. In other words, it measures the relative importance of each shock. Table 9 indicates that innovations in residential construction employment account for the largest share of uncertainty in housing market sentiment, explaining roughly one quarter of the forecast error variance of HMI across horizons. Mortgage rate shocks and housing starts contribute more modestly, each explaining about 4–6% of the variance at medium to long horizons, while completed new homes play a comparatively minor role. These results suggest that short- and medium-run fluctuations in builder sentiment are primarily driven by labor market conditions rather than financing costs or realized housing supply, consistent with the forward-looking and cost-sensitive nature of builders' expectations.

V. Discussion

Several limitations of this study suggest directions for future research. First, the analysis allows for at most one structural break in unit root testing, even though housing markets may experience multiple regime shifts over long samples. Allowing for only a single structural break when multiple breaks are present may still bias unit root tests toward nonrejection, as unmodeled regime shifts can be absorbed into the stochastic trend and mistakenly interpreted as persistent shocks. This concern is particularly relevant for long housing-market samples that span multiple crises and policy regimes. Future research could use tests that allow for multiple structural breaks. Relatedly, some housing-market variables may exhibit fractional integration, where shocks decay slowly but eventually dissipate. In such cases, integer differencing may overdifference the data, while standard unit root tests lack power to distinguish between true unit roots and long-memory processes.

Second, The Jarque-Bera normality test overwhelmingly rejects the null hypothesis of multivariate normality for the residuals across the entire system and for every individual equation. The strong rejection is driven by skewness and leptokurtosis. These features suggest that Gaussian innovations may be inappropriate and that extensions such as VAR–GARCH models with heavy-tailed error distributions, including the Student's t distribution, could provide a better representation of housing-market dynamics.

Finally, incorporating additional forward-looking macro-financial variables—such as credit conditions or financial stress indicators—may further enhance the explanatory and forecasting power of the model. Moreover, if long-run equilibrium relationships exist among housing-market variables, the analysis could be extended using a Vector Error Correction Model (VECM) to jointly capture short-run dynamics and long-run cointegrating relationships.

VI. Conclusion

The univariate analysis shows that housing starts exhibit pronounced volatility clustering and asymmetric responses to shocks, motivating the combination of ARIMA and asymmetric GARCH models. The ARIMA E-GARCH provides an outstanding in-sample fit, while the ARFIMA-E-GARCH gives a more realistic fit. The comparison highlights the trade-off between flexibility and overfitting, underscoring the need for caution when interpreting strong in-sample performance as evidence of superior forecasting ability.

The multivariate VAR analysis provides further insights into the dynamic structure of the housing market. Housing starts display substantial internal persistence, with shocks propagating gradually over time rather than dissipating immediately. Residential construction employment emerges as a key driver of both housing activity and builder sentiment, accounting for a sizable share of forecast error variance in sentiment. In contrast, mortgage rates appear largely exogenous to housing-sector variables at short horizons, consistent with their determination by broader financial and monetary conditions. These findings emphasize the central role of labor-market dynamics in shaping short- and medium-run housing-market fluctuations.

References

Francis X Diebold and Glenn D Rudebusch. On the power of Dickey-Fuller tests against fractional alternatives. *Economics Letters*, 35(2):155–160, 1991. doi: 10.1016/0165-1765(91)90163-f.

Pami Dua, Stephen M Miller, and David J Smyth. Using leading indicators to forecast U.S. home sales in a Bayesian Vector Autoregressive framework. *The Journal of Real Estate Finance and Economics*, 18(2): 191–205, 1999. doi: 10.1023/a:1007718725609.

Federal Reserve Bank of St. Louis. 30-year fixed rate mortgage average in the United States, 2025. URL `https://fred.stlouisfed.org/series/MORTGAGE30US`.

Thomas M Fullerton Jr and Carol A Taylor West. Regional econometric housing start forecast accuracy in Florida. *Review of Regional Studies*, 28, 1998. doi: 10.52324/001c.8902.

Eric Ghysels and Pierre Perron. The effect of seasonal adjustment filters on tests for a unit root. *Journal of Econometrics*, 55(1-2):57–98, 1993. doi: 10.1016/0304-4076(93)90004-o.

Sudiksha Joshi. Time series analysis and forecasting of the US housing starts using econometric and machine learning models. Mtholyoke.edu, 2019. URL `https://ida.mtholyoke.edu/items/297a1244-e8da-481f-a396-0d57ec1c9cce`.

Dukpa Kim and Pierre Perron. Unit root tests allowing for a break in the trend function at an unknown time under both the null and alternative hypotheses. *Journal of Econometrics*, 148(1):1–13, 2009. doi: 10.1016/j.jeconom.2008.08.019.

National Association of Home Builders. Housing market index (HMI). URL `https://www.nahb.org/news-and-economics/housing-economics/indices/housing-market-index`. Wells Fargo partnership.

Pierre Perron. The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, 57(6):1361–1401, 1989. doi: 10.2307/1913712.

Pierre Perron and Timothy Vogelsang. A note on the asymptotic distributions of unit root tests in the additive outlier model with breaks. *Brazilian Review of Econometrics*, 13(2), 1993. URL `https://ideas.repec.org/a/sbe/breart/v13y1993i2a2981.html`.

Pierre Perron and Tomoyoshi Yabu. Testing for shifts in trend with an integrated or stationary noise component. *Journal of Business & Economic Statistics*, 27(3):369–396, 2009. doi: 10.2307/27799090.

Christoph Sax and Dirk Eddelbuettel. Seasonal adjustment by X-13ARIMA-SEATS in R. *Journal of Statistical Software*, 87(11), 2018. doi: 10.18637/jss.v087.i11.

U.S. Bureau of Labor Statistics. Industries at a glance: Construction: NAICS 23, 2015. URL `https://www.bls.gov/iag/tgs/iag23.htm`.

U.S. Census Bureau. New residential construction press release. URL `https://www.census.gov/construction/nrc/current/index.html`. Accessed: 2026-04-16.

VII. Appendix

Table 2: Data Sources

Variable	Source	Brief Description
Total Housing Starts	U.S. Census Bureau	A measure of new residential construction (Single-Family + Multifamily), [in numbers]
Completed New Homes	U.S. Census Bureau	Dwellings that are never occupied but are ready for occupancy, [in numbers]
Residential Construction Jobs	Bureau of Labor Statistics	Total number of workers in the industry, including maintenance, [in numbers]
30-Year Fixed-Rate Mortgage	FRED Database	A fixed-rate home loan with a 30-year repayment term, [in percent]
Housing Affordability Index	National Association of Realtor	Measure if a typical family's income is enough to qualify for a mortgage, [in index]
Housing Market Index	National Association of Home Builders / Wells Fargo	A monthly survey of homebuilders that tracks builders' sentiment, [in index]

Table 3: Conventional Unit Root Tests on Level

Variable	Lag Selection [1]	ADF [2]	DF-GLS	PP	KPSS [3]
lTotalHStarts	15	-2.618	-2.696*	-4.177***	0.293***
lResConJobs	17	-3.022	-3.058**	-2.177	0.256***
lComNewHomes	12	-4.006***	-3.974***	-1.664	0.180**
FRM	3	-2.218	-0.828	-1.983	0.517***
HAI	17	-1.625	-1.615	-1.644	0.406***
HMI	17	-2.776	-2.802*	-2.445	0.254***

[1] The lag selection criterion is based on the Ng-Perron sequential t-test.
[2] ADF critical values for are -3.981 (1%), -3.421 (5%), and -3.130 (10%).
[3] KPSS test is biased towards rejection with too few lags, and it requires the same lag across variables for comparability. Thus, a lag of 10 is chosen. Note that the null and alternative hypotheses for KPSS is the reverse of the other three tests.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 4: Conventional Unit Root Tests after Differencing

Variable	Lag Selection [1]	ADF [2]	DF-GLS	PP	KPSS [3]
dlTotalHStarts	14	-3.978***	-0.633	-19.777***	0.0782
dlResConJobs	16	-3.226**	-0.727	-12.773***	0.129
dlComNewHomes	17	-4.572***	-4.243***	-13.593***	0.107
dFRM	17	-5.088***	-1.928*	-15.023***	0.374*
dHAI	16	-4.510***	-3.411***	-14.812***	0.333
dHMI	16	-4.398***	-0.385	-18.758***	0.0696

[1] The lag selection criterion is based on the Ng-Perron sequential t-test.
[2] ADF critical values for are -3.443 (1%), -2.871 (5%), and -2.570 (10%).
[3] The lags for the KPSS test is also 10 here. Note that the null and alternative hypotheses for KPSS is the reverse of the other three tests.
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Unit Root Tests with Structural Break (Model AO-A & AO-B & AO-C)

Variable	Model [1]	Break Date [2]	Lag Selection [3]	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
lTotalHStarts	AO-A	Nov 2007 (obs: 275)	15 (Ng-Perron)	-3.31	-4.45	-3.76	-3.46
lTotalHStarts	AO-C	Nov 2007 (obs: 275)	10 (MAIC) 10 (MBIC)	-2.5979	-4.88	-4.24	-3.95
lResConJobs	AO-A	Jul 2008 (obs: 283)	17 (Ng-Perron)	-2.98	-4.45	-3.76	-3.46
lResConJobs	AO-C	Dec 2008 (obs: 288)	9 (MAIC) 8 (MBIC)	-1.6128 -1.6703	-4.88	-4.24	-3.95
lComNewHomes	AO-C	Mar 2011 (obs: 315)	3 (MAIC) 1 (MBIC)	-3.1587 -2.5568	-4.88	-4.24	-3.95
FRM	AO-B	Sep 2020 (obs: 429)	3 (MAIC) 2 (MBIC)	-4.5909*** -4.8584***	-4.28	-3.72	-3.43
HAI	AO-C	Dec 2008 (obs: 288)	7 (MAIC) 1 (MBIC)	-2.7714 -4.3726**	-4.88	-4.24	-3.95
HMI	AO-C	May 2006 (obs: 257)	2 (MAIC) 1 (MBIC)	-2.7976 -3.2260	-4.90	-4.24	-3.96

Notes: The critical values for Model AO-A can be found in Table IV.B of Perron (1989), those for Model AO-B can be found in Table 1 of the Perron-Vogelsang paper (1993), and those for Model AO-C can be found in Table VI.B of Perron (1989)
[1] lTotalHStarts and lResConJobs are tested by different models because the models can't be easily determined based on the graphs
[2] λ (time of break relative to sample size) for each variable is 0.6, 0.6, 0.6, 0.9, 0.6, and 0.5 respectively
[3] Model AO-A uses Ng-Perron Sequential t-test for lag selection. Model AO-B and AO-C use MAIC and MBIC, both of which are included because they diverge widely for one of the variables (HAI)
*** p<0.01, ** p<0.05, * p<0.1

Appendix B. ARIMA-GARCH and AFRIMA-GARCH models

Table 6: ARIMA / ARFIMA Volatility Model Estimates

	ARIMA-E-GARCH	ARFIMA-E-GARCH	ARIMA-GJR-GARCH	ARFIMA-GJR-GARCH
AIC	-1233.05	-2.3690	-1214.26	-2.3729
BIC	-1207.95	-2.3001	-1193.34	-2.3040
Unconditional Mean	0.0011	4.9351***	0.0008	4.9321***
L1.ma	-0.4706***	0.0662	-0.4647***	0.0596
L1.(e)arch	-0.1191**	0.2948***	0.4291***	0.1213**
L1.earch_a	0.4302***	-0.0732*	–	–
L1.egarch	0.7345***	0.9202***	–	–
L1.tarch	–	–	-0.2821**	0.1889
Constant	-1.4335***	-0.4221**	0.0036***	0.0008*
ARFIMA	–	0.5	–	0.5

Notes: EGARCH and GJR-GARCH models estimated with ARIMA and ARFIMA mean specifications.

[1] The parameter is L1.arch for GJR-GARCH-type model and is L1.earch for E-GARCH-type model.

*** p<0.01, ** p<0.05, * p<0.1

Appendix C. VAR model

Table 7: Vector Autoregression Results (VAR(4), HAI Excluded)

	(1) fdlTotalHStarts_sa	(2) dlComNewHomes_sa	(3) dlResConJobs_sa	(4) dHMI_sa	(5) FRM_detrended
L.fdlTotalHStarts_sa	0.0387	0.0236	0.0309***	8.508***	0.417***
	(0.0484)	(0.0200)	(0.00722)	(2.399)	(0.132)
L2.fdlTotalHStarts_sa	0.185***	-0.00808	0.0188***	4.070*	-0.211
	(0.0490)	(0.0203)	(0.00730)	(2.427)	(0.133)
L3.fdlTotalHStarts_sa	0.183***	-0.00662	0.00673	-6.079**	-0.216
	(0.0487)	(0.0201)	(0.00726)	(2.411)	(0.132)
L4.fdlTotalHStarts_sa	0.192***	0.0190	0.00485	-2.713	0.0551
	(0.0490)	(0.0203)	(0.00730)	(2.427)	(0.133)
L.dlComNewHomes_sa	-0.141	0.218***	-0.00333	-15.44***	0.0835
	(0.110)	(0.0456)	(0.0164)	(5.464)	(0.300)
L2.dlComNewHomes_sa	-0.0267	0.177***	-0.00597	-15.36***	0.102
	(0.113)	(0.0469)	(0.0169)	(5.619)	(0.309)
L3.dlComNewHomes_sa	0.0914	0.0771*	0.0116	-4.115	0.159
	(0.113)	(0.0468)	(0.0168)	(5.599)	(0.307)
L4.dlComNewHomes_sa	0.0391	0.0621	-0.0279*	7.246	0.147
	(0.110)	(0.0454)	(0.0164)	(5.434)	(0.298)
L.dlResConJobs_sa	1.043***	-0.127	-0.125**	16.65	-0.814
	(0.370)	(0.153)	(0.0551)	(18.32)	(1.006)
L2.dlResConJobs_sa	0.308	0.449***	0.00609	-45.55**	1.202
	(0.372)	(0.154)	(0.0554)	(18.41)	(1.011)
L3.dlResConJobs_sa	0.0299	0.0746	0.117**	-33.47*	-0.560
	(0.371)	(0.154)	(0.0553)	(18.39)	(1.010)
L4.dlResConJobs_sa	0.280	0.148	0.157***	32.14*	0.467
	(0.350)	(0.145)	(0.0521)	(17.32)	(0.951)
L.dHMI_sa	0.00132	-0.000359	-0.000320**	0.00788	0.00171
	(0.00106)	(0.000438)	(0.000158)	(0.0525)	(0.00288)
L2.dHMI_sa	0.000587	-0.000519	-0.000155	-0.0779	-0.00363
	(0.00103)	(0.000427)	(0.000154)	(0.0511)	(0.00281)
L3.dHMI_sa	-0.00141	0.000478	-0.000388***	-0.0174	0.00412
	(0.000999)	(0.000413)	(0.000149)	(0.0495)	(0.00272)
L4.dHMI_sa	0.000347	-0.000411	-0.000210	-0.155***	-0.00269
	(0.000970)	(0.000401)	(0.000145)	(0.0480)	(0.00264)
L.FRM_detrended	0.00453	0.0217***	-0.00157	-3.094***	1.379***
	(0.0168)	(0.00694)	(0.00250)	(0.831)	(0.0456)
L2.FRM_detrended	-0.0295	-0.0216*	0.00119	1.662	-0.660***
	(0.0282)	(0.0117)	(0.00420)	(1.396)	(0.0767)
L3.FRM_detrended	-0.000700	0.0218*	-0.00384	0.313	0.316***
	(0.0284)	(0.0117)	(0.00423)	(1.404)	(0.0771)
L4.FRM_detrended	0.0164	-0.0183***	0.00152	0.643	-0.0995**
	(0.0170)	(0.00705)	(0.00254)	(0.844)	(0.0463)
Observations	481	481	481	481	481

Notes: Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 8: Granger Causality Wald Tests (VAR(4), HAI Excluded)

Equation	Excluded	χ^2	df	Prob > χ^2
fdlTotalHStarts_sa	dlComNewHomes_sa	2.2025	4	0.699
fdlTotalHStarts_sa	dlResConJobs_sa	9.541	4	0.049
fdlTotalHStarts_sa	dHMI_sa	3.9266	4	0.416
fdlTotalHStarts_sa	FRM_detrended	7.879	4	0.096
fdlTotalHStarts_sa	ALL	39.424	16	0.001
dlComNewHomes_sa	fdlTotalHStarts_sa	2.8488	4	0.583
dlComNewHomes_sa	dlResConJobs_sa	11.789	4	0.019
dlComNewHomes_sa	dHMI_sa	4.0442	4	0.400
dlComNewHomes_sa	FRM_detrended	19.288	4	0.001
dlComNewHomes_sa	ALL	49.204	16	0.000
dlResConJobs_sa	fdlTotalHStarts_sa	32.951	4	0.000
dlResConJobs_sa	dlComNewHomes_sa	3.6431	4	0.456
dlResConJobs_sa	dHMI_sa	15.97	4	0.003
dlResConJobs_sa	FRM_detrended	13.714	4	0.008
dlResConJobs_sa	ALL	61.545	16	0.000
dHMI_sa	fdlTotalHStarts_sa	18.839	4	0.001
dHMI_sa	dlComNewHomes_sa	24.001	4	0.000
dHMI_sa	dlResConJobs_sa	13.969	4	0.007
dHMI_sa	FRM_detrended	24.963	4	0.000
dHMI_sa	ALL	98.674	16	0.000
FRM_detrended	fdlTotalHStarts_sa	14.318	4	0.006
FRM_detrended	dlComNewHomes_sa	1.4173	4	0.841
FRM_detrended	dlResConJobs_sa	3.5294	4	0.473
FRM_detrended	dHMI_sa	4.2712	4	0.371
FRM_detrended	ALL	22.281	16	0.134

Notes: Each row reports a Wald test of the joint significance of the excluded variable's lagged coefficients in the indicated equation (VAR(4)). The row "ALL" tests the joint significance of all other endogenous variables' lags in that equation.

Table 9: Forecast Error Variance Decomposition (FEVD)

	Impulse: FRM_detrended			Impulse: dlComNewHomes_sa			Impulse: dlResConJobs_sa			Impulse: fdlTotalHStarts_sa		
Step	fevd	Lower	Upper	fevd	Lower	Upper	fevd	Lower	Upper	fevd	Lower	Upper
0	0	0	0	0	0	0	0	0	0	0	0	0
1	0.000783	-0.004214	0.00578	0.001873	-0.004658	0.008404	0.276068	0.208106	0.34403	0.007448	-0.005572	0.020468
2	0.029782	-0.000419	0.059144	0.016575	-0.004702	0.037852	0.269522	0.202858	0.336185	0.03144	0.002591	0.060288
3	0.05441	0.013466	0.095354	0.038733	0.0057	0.071766	0.259622	0.19446	0.324785	0.032711	0.003638	0.061783
4	0.061246	0.018374	0.104118	0.044763	0.009965	0.079561	0.260544	0.195228	0.325859	0.04307	0.009293	0.076848
5	0.064722	0.021742	0.107703	0.044193	0.009591	0.078796	0.255236	0.190459	0.320014	0.044073	0.009331	0.078814
6	0.064758	0.022	0.107516	0.044452	0.010117	0.078787	0.256391	0.191908	0.320874	0.044035	0.009338	0.078732
7	0.064683	0.021923	0.107443	0.044472	0.010278	0.078666	0.255584	0.191247	0.319921	0.044618	0.009834	0.079401
8	0.064673	0.021923	0.107423	0.044505	0.010305	0.078705	0.255547	0.191221	0.319873	0.044695	0.009888	0.079501
9	0.064632	0.021919	0.107345	0.045235	0.010749	0.07972	0.255336	0.191052	0.31962	0.044648	0.009875	0.079421
10	0.064594	0.021896	0.107291	0.045352	0.010797	0.079908	0.255334	0.191067	0.319601	0.044923	0.010086	0.079761
11	0.06456	0.02188	0.107239	0.045375	0.01082	0.07993	0.255432	0.191162	0.319702	0.045138	0.010227	0.080048
12	0.064536	0.021875	0.107197	0.045399	0.010855	0.079943	0.255566	0.191308	0.319824	0.045257	0.010314	0.080199

Notes: FEVD reports the share of forecast error variance of *dHML_sa* attributed to each orthogonalized shock. Lower/Upper are the reported confidence bounds from `irf table fevd`.

Exploring the Impact of Women in National Parliaments on Government Effectiveness

By Shu Shi

This paper examines whether greater women's parliamentary representation improves government effectiveness. Using an unbalanced panel of up to 217 countries (1998-2024), I estimate OLS and fixed-effects regressions relating women's parliamentary seats to the Government Effectiveness index, controlling for GDP, trade, female labor force participation, education spending, fertility, tertiary enrollment, urbanization and additional nonlinear/interaction specifications. Simple OLS shows positive correlation ($\beta = 0.023$, $p < 0.01$), but with country and year fixed effects, the coefficient becomes essentially zero ($\beta = 0.001$, $p = 0.42$). The results provide no evidence that increasing women's representation is associated with short-run improvements in aggregate governance quality.

I. Introduction

Despite recent improvements, gender parity in political institutions remains exclusive. In 2025, women held only 27.2 percent of seats in national parliaments worldwide, with the Americas averaging 35.3 percent, Europe 31.8 percent, and the Middle East and North Africa only 18.1 percent UN Women [2025]. Women served as heads of state in just 11.9 percent of countries and as heads of government in 8.3 percent of countries UN Women [2025]. This persistent under-representation raises an important question: beyond fairness, is greater women's representation actually associated with more efficient governments? This paper will examine whether a higher proportion of women in national parliaments is associated with more effective governments. Meanwhile, existing empirical research has largely focused on corruption issues or specific policy domains, while causal evidence regarding whether increasing the proportion of female legislators can improve overall governance remains inconclusive. My study adds to this debate by examining the broad quality of government across a large sample of countries and years.

Thus, I hypothesize that an increase in the proportion of female parliamentary representatives will be associated with improvements in government effectiveness, but this association may depend on the specific conditions of the institutional and policy environment-particularly the effect may differ with a country's level of social investment, proxied by government spending on education. In high-investment settings, baseline government capacity and service provision may already be stronger, leaving less room for marginal improvements from additional representation. By contrast, in lower-investment settings, additional female legislators may be associated with greater orientation toward human capital and basic service delivery, potentially correlating with greater gains in effectiveness.

Using fixed-effects regressions with socio-economic controls on an unbalanced panel of up to 217 countries from 1998 to 2024, I relate the percentage of parliamentary seats held by women to the Government Effectiveness Index. My findings suggest that in the preferred fixed-effects specification, a 10-percentage-point increase in women's parliamentary share is associated with less than a 0.01-point change in Government Effectiveness, though the effect is modest and the coefficient is not statistically significant at conventional levels (10%, 5%, 1%). The paper is structured as follows. Section 2 reviews the related literature on women's political representation and governance outcomes. Section 3 describes the data and variables. Section 4 outlines the theoretical model and empirical framework. Section 5 presents the regression results and robustness checks. Section 6 concludes and discusses policy implications, while Section 7 addresses limitations and future studies.

II. Literature ??eview

A substantial body of literature in political science and economics examines whether women introduce distinct preferences into politics and whether these differences can enhance governance. Swamy et al. Swamy et al. [2001] find that women are more likely than men to exhibit prosocial behavior, honest character, and concern for the public good, and to demonstrate greater intolerance toward corruption.

At the cross-country level, Dollar, Fisman, and Gatti Dollar et al. [2001] document a negative relationship between women's representation in national legislatures and perceived corruption using a sample of over 100 countries and cross-sectional regression analysis. Importantly, this evidence is primarily correlational since countries that elect more women also tend to differ in income, institutions, and social norms, which may confound the association. While corruption is one dimension of governance, my paper focuses on government effectiveness, which captures the quality of public services and the capacity for policy implementation. Evidence directly linking women's representation to this broader outcome remains limited. Dirir Dirir [2022] studies 12 MENA countries over 2012-2021 and estimates panel models, including fixed-effects and random-effects specifications, and finds a positive coefficient on women in parliament for government effectiveness within that regional sample. However, the scope is narrow in both geography and time, and the set of covariates differs from a global panel setting, so its external validity is uncertain.

A second strand of research studies specific policy outcomes rather than aggregate governance indices. Bhalotra and Clots-Figueras (2014) address selection into electing women by estimating a fixed-effects IV model that uses close elections between male and female candidates as a source of quasi-random variation in women's representation. They find that higher women's representation reduces neonatal mortality and discuss mechanisms such as improved public health provision. However, their identification relies on Indian state-level close elections, so the findings may not generalize to national parliaments or to institutional contexts outside India. An et al. (2025) exploit reserved-seat quota adoption and report that reserved seats for women in African parliaments enhance public access to improved drinking water, with particularly significant effects when quota proportions are higher. Yet the analysis is limited to African countries with reserved-seat quotas, and drinking water access is a narrow outcome that may not reflect broader governance capacity. Together, these studies use credible identification to show that increases in women's political representation can affect specific public goods and service-delivery outcomes. Whether these domain-specific improvements translate into broad aggregate measures of government effectiveness remains an open empirical question, which my global panel approach is designed to examine.

To reduce omitted-variable bias, I include a standard set of controls commonly used in cross-country governance regressions. Prior work shows that governance quality is strongly correlated with baseline development and human capital, motivating controls such as GDP per capita and education-related measures [Hall and Jones, 1999]. The cross-country "quality of government" literature also emphasizes that structural characteristics and institutional environment are systematically related to government performance [La Porta et al., 1999]. In addition, studies of corruption and governance highlight the role of economic openness, motivating a control for trade exposure Treisman [2000]. Accordingly, my main controls include GDP per capita, education/human-capital proxies (government expenditure on education, tertiary school enrollment), trade openness, and structural variables such as urbanization, fertility rate, and female labor force participation.

III. Data and Descriptive Statistics

This paper uses an unbalanced cross-country panel constructed from the World Bank's World Development Indicators (WDI). The data contains annual observations from 1998 to 2024 for up to 217 countries, although coverage differs by series and year. The sample covers all countries with available data in WDI, including both high-income OECD members and low-income developing countries. Depending on the variable, the number of countries–year observations ranges from roughly 3,300 to almost 5,800, so the sample becomes smaller in specifications that include controls with more missing values. Variable descriptions are provided in Table 1, and summary statistics are reported in Table 2 in the Appendix.

The dependent variable is Government Effectiveness (GovEffect). Higher values indicate better quality of public services, a more independent civil service, better policy formulation and implementation, and stronger credibility of the government's commitments. In my sample, GovEffect has a mean of −0.03, a standard deviation of about 1.0, and a range from −2.44 to +2.47, indicating large differences in governance across countries and over time. The main explanatory variable is Women in Parliament (WomenParl), defined as the percentage of seats held by women in the national parliament. On average, women hold 18.6 percent of seats, but the share varies from 0 (e.g., Yemen in recent years) to about 63.8 percent (e.g., Rwanda's Chamber of Deputies). This substantial variation across countries and years is key to identifying the relationship between women's representation and government effectiveness. The regressions control for several time-varying country characteristics. GDP per capita in constant 2015 US dollars (*GDPpc*) captures the level of economic development and ensures comparability over time by removing inflation effects. Female labour force participation ($LPFfemale$) measures the percentage of women aged 15 and older who are economically active, and female tertiary school enrollment ($TerEnroll$) proxies women's educational attainment. Additional controls include the total fertility rate (*Fertility*), the urban population share (*Urban*), $Trade$, and government expenditure on education as a share of GDP (*GovEdu*). In some specifications, I also include the interaction term $WomenParl \times GovEdu$ and the quadratic term $WomenParl^2$.

IV. Econometric Methodology

A. *Model Specification*

To study whether a higher share of women in national parliaments is positively correlated with better government effectiveness, I start from the following population regression model:

$$\text{GovEffect}_{it} = \beta_0 + \beta_1\text{WomenParl}_{it} + \beta_2\text{GDPpc}_{it} + \beta_3\text{TerEnroll}_{it} + \beta_4\text{LPFfemale}_{it} + \beta_5\text{Fertility}_{it} \tag{1}$$

where i indexes countries and t years. GovEffect_{it} is the Government Effectiveness index, and WomenParl_{it} is the percentage of parliamentary seats held by women.

$$X_{it} = (\text{LPFfemale, GDPpc, Urban, Trade, Fertility, TerEnroll, GovEdu}) \tag{2}$$

is a vector of time-varying controls. α_i are country fixed effects that capture time-invariant differences across countries, and γ_t are year fixed effects that capture global shocks or common trends. ϵ_{it} is an error term.

The rationale for including country fixed effects is that many determinants of government effectiveness, such as colonial history, legal origin, geographic factors, and deep-rooted institutional norms, differ across countries but change little over time. Without country fixed effects, the OLS estimates would confound the effect of women's representation with these persistent cross-country differences. By absorbing all time-invariant heterogeneity, the fixed-effects estimator identifies β_1 solely from within-country changes. This indicates whether a given country's government effectiveness improves in years when that country's share of women in parliament rises, relative to its own historical average and the global trend. This means the fixed-effects estimates capture short-run effects of changes in representation, rather than long-run cross-country associations.

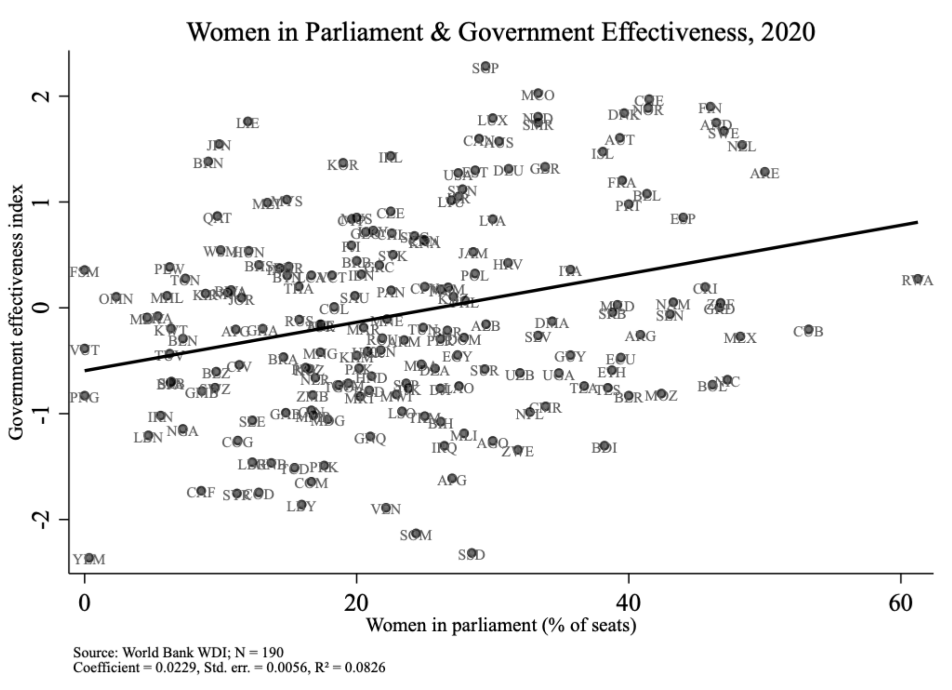
B. *Expected Coefficients*

The coefficient of interest is β_1. I expect $\beta_1 > 0$: if female legislators place greater emphasis on public goods, human capital, and anti-corruption policies, then increasing their representation in parliament should enhance government service quality and policy implementation effectiveness. At the same time, I allow β_1 to be small or statistically uncertain once other determinants of governance are controlled for, since many factors that affect the quality of government also affect women's chances of being elected. The controls are included to reduce omitted-variable bias. *GDPpc* captures the level of economic development since richer countries tend to have stronger bureaucratic capacity and better public services, so I expect $\beta_2 > 0$. $LPFfemale$ and $TerEnroll$ proxy for broader women's empowerment, which should be positively related to both women's political representation and government effectiveness, suggesting $\beta_3 > 0, \beta_4 > 0$. $Fertility$ is usually associated with more traditional gender norms and fewer resources per child, so I expect higher fertility to be linked to weaker state capacity, implying $\beta_5 < 0$. $Urban$ and $Trade$ reflect structural features of the economy; more urban and more open economies may face stronger demands for efficient public services but may also experience governance challenges, so the signs of β_6 and β_7 are a priori ambiguous. Finally, government expenditure on education as a share of GDP (*GovEdu*) captures a key dimension of social investment that is likely to improve governance, so I expect $\beta_8 > 0$.

V. Preliminary Analysis

Figure 1 illustrates the cross-sectional relationship between women's parliamentary representation and government effectiveness for 190 countries with available data in a single year (the remaining countries lack complete data for 2020). I use 2020 as a recent year with broad data coverage across countries, which makes the pattern easy to visualize. The scatter plot reveals a positive correlation, with the fitted line showing a coefficient of 0.0229 (also reported in Table 3, Model 1 in the Appendix). Despite this overall positive trend, substantial cross-country variation exists. For instance, among countries with approximately 40% female parliamentary representation, Denmark (DNK) and Austria (AUT) achieve high government effectiveness scores above 1.5, while countries like North Macedonia (MKD) record lower scores around 0. Similarly, countries with minimal female representation below 10% exhibit divergent governance outcomes, with Yemen (YEM) scoring near -2.4 while Micronesia (FSM) maintains a positive score around 0.3. This variation suggests that while women's parliamentary participation is associated with improved government effectiveness, the relationship is mediated by other institutional, economic, and cultural factors that vary across countries.

Figure 1: Relationship Between Women in Parliament and Government Effectiveness, Year 2020



VI. Results and Empirical Analysis

Table 3 in the Appendix reports seven regression models examining the relationship between women's parliamentary representation and government effectiveness. Models 1-3 use ordinary least squares (OLS), with Model 1 showing a simple cross-section for 2020 and Models $2-3$ using the full panel with controls. Models 4-7 add country and year fixed effects to exploit within-country variation over time. The coefficient on $WomenParl$ is consistently positive but declines sharply in magnitude once controlled for other determinants of governance and country-specific characteristics.

Model 1 estimates a simple regression of Government Effectiveness on $WomenParl$ using 190 countries in 2020. The coefficient is 0.0229 with a standard error of 0.00545, significant at the 1 percent level. In other words, a one-percentage-point increase in women's parliamentary share is associated with a 0.0229-point higher Government Effectiveness score. This is a small but meaningful cross-sectional correlation: countries with more women in parliament tend to have more effective governments. However, this specification omits income, demographics, and education, so the correlation likely reflects confounding factors rather than a causal effect. Models 2 and 3 add control variables to the OLS regressions. In Model 2, I include $LPFfemale$, $GDPpc$, $Urban$, $Trade$, $Fertility$, and $TerEnroll$ using 2,528 country-year observations. The adjusted R^2 rises to 0.737, showing that these variables explain most of the cross-country variation in governance. The coefficient on $WomenParl$ falls to 0.00298, an 87 percent reduction relative to Model 1, but remains positive and statistically significant. A 10-percentage-point increase in women's representation now corresponds to only a 0.03-point (about 0.03 standard deviations) improvement in government effectiveness. This sharp decline indicates that much of the raw correlation reflects differences in development, demographics, and education rather than women's representation.

Model 3 restricts the sample to the 1,993 observations with data on government education spending ($GovEdu$). $GovEdu$ enters with a positive and precisely estimated coefficient of 0.0476, indicating that countries investing more in education tend to have more effective governments. Once $GovEdu$ is included, the coefficient

on WomenParl drops further to 0.00136 and becomes statistically indistinguishable from zero. This pattern suggests that part of the positive OLS correlation in earlier models reflects omitted education-related institutional investments rather than women's representation. After controlling for GovEdu, there is no clear evidence of an independent partial correlation between women's parliamentary share and Government Effectiveness in the cross-section.

Models 4-7 introduce country and year fixed effects to control for unobserved time-invariant country characteristics and common global trends. Model 4 is a parsimonious fixed-effects model that includes only WomenParl as a regressor. The coefficient is 0.00473 and statistically significant at the 5 percent level. Conditional on country and year fixed effects, a 10-percentage-point increase in women's parliamentary share is associated with a 0.047-point improvement in government effectiveness. The adjusted R-squared is only 0.007, indicating that changes in women's representation explain less than 1 percent of the within-country variation in governance quality over time.

Model 5 tests for nonlinearity by adding WomenParl^2. The linear term remains positive and significant, but the quadratic term is negative (–0.000142) and statistically insignificant. This provides no strong evidence for a U-shaped or inverted U-shaped relationship. Given the statistical insignificance of the squared term and the added complexity it introduces, I proceed with linear specifications in Models 6 and 7.

Model 6 explores whether the relationship depends on education spending by including the interaction term WomenParl×GovEdu. The coefficient on WomenParl is 0.00608, and the interaction term is −0.00119; both are statistically significant at the 5 percent level. This negative interaction suggests that the positive association between women's parliamentary representation and government effectiveness is less pronounced in countries that already allocate a high share of GDP to education. One interpretation is that, in contexts with strong educational institutions, the marginal association between female representation and governance quality is weaker, perhaps because other mechanisms to ensure effective government are already in place. Alternatively, it may reflect diminishing returns: countries that invest heavily in education may have already achieved gender parity in education and political participation, so further increases in female representation have less room to improve governance through that channel. However, the adjusted R-squared is only 0.004, so these interaction patterns are modest and should not be over-interpreted.

Model 7 presents my preferred specification, including country fixed effects, year fixed effects, and the full set of controls (LPFfemale, GDPpc, Urban, Trade, Fertility, TerEnroll, GovEdu). The coefficient on WomenParl is 0.000883 with a standard error of 0.00209 - essentially zero and not statistically significant. A 10-percentage-point increase in women's parliamentary share corresponds to less than a 0.01-point change in Government Effectiveness, which is under 1 percent of a standard deviation. Most control variables are also imprecisely estimated in this fixed-effects setting, which is common when both the dependent variable and many regressors change slowly over time. The one clear exception is Trade, which enters with a negative and statistically significant coefficient of –0.00146 (standard error 0.000730). This unexpected sign may reflect that greater trade openness increases governance challenges - such as regulatory complexity or exposure to corruption pressures - or it may indicate omitted variable bias. I retain the controls in Model 7 to reduce potential omitted variable bias, even though most of them do not individually reach significance. The adjusted R-squared in Model 7 is 0.015, meaning the model explains about 1.5 percent of the within-country variation in Government Effectiveness. The low R-squared values across fixed-effects models are not a sign of model failure; rather, they reflect that governance institutions are highly persistent, so annual within-country variation is limited. Most of the variation in Government Effectiveness is between countries rather than within countries over time. I treat Model 7 as the preferred specification because it combines fixed effects with key time-varying controls, prioritizing reduced omitted-variable bias over precision. Given the limited within-country variation and measurement noise in WGI, the estimates are most informative for ruling out large short-run effects in the global panel.

Overall, the evidence shows a consistent pattern. Simple OLS models reveal a strong positive correlation between women's parliamentary representation and government effectiveness, but this correlation largely reflects cross-country differences in income, education, and demographics. Once these factors are controlled for and fixed effects are used to focus on within-country changes, the estimated effect becomes economically negligible and statistically insignificant. In the preferred fixed-effects specification (Model 7), the coefficient on WomenParl is less than one-thousandth of a point and not distinguishable from zero. These results provide no evidence that increasing women's parliamentary representation generates measurable short-run improvements in aggregate government effectiveness, although they also do not suggest any harmful effects and rule out large positive effects in this global panel.

VII. Conclusion and Policy Implications

This paper finds no evidence that increasing women's parliamentary representation produces measurable short-run gains in aggregate governance quality. With country and year fixed effects, the coefficient on women's representation becomes essentially zero ($\beta = 0.001$, $p = 0.42$). A 10-percentage-point increase in women's representation corresponds to less than 0.01 standard deviations in governance quality, which is economically negligible. These results suggest that increasing women's parliamentary representation is not associated with measurable short-run gains in aggregate governance quality. For policymakers, gender quotas may still be associated with improvements in specific outcomes, such as health or corruption, but they should not be expected to quickly transform overall government effectiveness. The null result likely reflects governance's slow-moving nature and the broad, perception-based measurement of the Government Effectiveness index. For instance, Rwanda's share of women in the Chamber of Deputies rose from 48.8% in 2003 to 63.8% in 2013 (IPU Parline, 2024), yet its WGI Government Effectiveness estimate increased from -0.67 to -0.43 (a change of 0.24 points on a scale of -2.5 to 2.5). This modest shift relative to the dramatic change in representation illustrates that governance institutions respond slowly, and a 26-year panel may not be long enough to detect the full effects of increased representation.

VIII. Limitations and Further Research

Several limitations of this study should be kept in mind when interpreting the results and suggest directions for future work. First, the dependent variable, Government Effectiveness, changes slowly over time and is measured with error. The WGI documentation itself stresses that most year-to-year movements are small relative to the confidence intervals. Because the fixed-effects estimates rely only on within-country changes, limited time variation, and perception-based measurement error can push estimated coefficients toward zero and make them imprecise. This likely contributes to the small and often insignificant effects in the fixed-effects models. Second, missing data and sample selection may affect the results. Adding controls such as tertiary enrollment, female labor force participation, and government education spending reduces the number of country–year observations, and countries with weaker statistical capacity are more likely to drop out of the sample. If these missing observations are systematically related to both women's representation and governance, the estimates may be biased. Third, the paper uses a single measure of women's political power, which does not fully capture women's actual roles in the political system or distinguish between elected and reserved seats.

The models also combine all countries without allowing effects to vary by regime type, even though previous research suggests women's influence on governance is stronger in democracies (Esarey & Chirillo, 2013).

Future research could: (i) develop more comprehensive measures of women's political influence across branches and levels of government; (ii) account for different effects based on regime type and region; (iii) integrate global panel data with subnational-level data for analysis; and (iv) employ causal analysis methods based on quota reforms. These improvements could offer clearer evidence of when and how women's representation enhances government quality.

References

J. An, T. Wang, B. Chen, P. Gong, C. Lin, and N. C. Stenseth. Impact of women's political empowerment through gender quotas on improved drinking water access in Africa. *Proceedings of the National Academy of Sciences*, 122(30), 2025. doi: 10.1073/pnas.2511977122.

S. Bhalotra and I. Clots-Figueras. Health and the Political Agency of Women. *American Economic Journal: Economic Policy*, 6(2):164–197, 2014. doi: 10.1257/pol.6.2.164.

Kadir Aden Dirir. Evaluating indicators contributing to government effectiveness in 12 MENA regions between 2012 to 2022: employing a longitudinal data analysis. *Journal of Global Politics and Current Diplomacy*, 10(2):19–41, 2022. URL `https://journal.centruldedic.ro/wp-content/uploads/2023/02/Article-2-JGPCD_2022-2.pdf`.

D. Dollar, R. Fisman, and R. Gatti. Are women really the "fairer" sex? Corruption and women in government. *Journal of Economic Behavior & Organization*, 46(4):423–429, 2001. doi: 10.1016/s0167-2681(01)00169-x.

J. Esarey and G. Chirillo. "Fairer Sex" or Purity Myth? Corruption, Gender, and Institutional Context. *Politics & Gender*, 9(04):361–389, 2013. doi: 10.1017/s1743923x13000378.

R. E. Hall and C. I. Jones. Why do Some Countries Produce So Much More Output Per Worker than Others? *The Quarterly Journal of Economics*, 114(1):83–116, 1999. doi: 10.1162/003355399555954.

IPU Parline. Rwanda — Chamber of Deputies — Data on women, 2024. URL `https://data.ipu.org/parliament/RW/RW-LC01/data-on-women/`. Global data on national parliaments.

R. La Porta, F. Lopez-de Silanes, A. Shleifer, and R. Vishny. The Quality of Government. *Journal of Law, Economics, and Organization*, 15(1):222–279, 1999. doi: 10.1093/jleo/15.1.222.

A. Swamy, S. Knack, Y. Lee, and O. Azfar. Gender and corruption. *Journal of Development Economics*, 64(1): 25–55, 2001. doi: 10.1016/s0304-3878(00)00123-1.

D. Treisman. The causes of corruption: a cross-national study. *Journal of Public Economics*, 76(3):399–457, 2000. doi: 10.1016/s0047-2727(99)00092-4.

UN Women. Women in politics: 2025, March 2025. URL `https://www.unwomen.org/en/digital-library/publications/2025/03/women-in-politics-map-2025`.

World Bank. World Development Indicators, 2025. URL `https://databank.worldbank.org/source/world-development-indicators`.

IX. Appendix

Table 1: Variable Description

GovEffect	Government Effectiveness Score (index ranging from -2.5 to 2.5)
WomenParl	Proportion of seats held by women in national parliaments (%)
Fertility	Fertility rate, total (births per woman)
GDPpc	GDP per capita (constant 2015 US$)
GovEdu	Government expenditure on education, total (% of GDP)
LPFfemale	Labor force participation rate, female (% of female population ages 15+)
TerEnroll	School enrollment, tertiary (% gross)
Trade	Trade (% of GDP)
Urban	Urban population (% of total population)
WomenParl*GovEdu	Interaction term of government expenditure on education and proportion of seats held by women in national parliaments
WomenParl2	Squared proportion of seats held by women in national parliaments (%)

Table 2: Summary Statistics

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
WomenParl	4,915	18.60	11.99	0	63.75
GovEffect	4,783	-0.0286	0.995	-2.440	2.470
GDPpc	5,511	15,370	22,454	219.7	224,582
TerEnroll	3,261	39.48	28.63	0.116	166.7
LPFfemale	5,040	50.34	15.67	4.907	87.83
Fertility	5,642	2.830	1.502	0.586	7.840
Urban	5,805	58.76	24.25	7.830	100
Trade	4,761	89.52	57.71	2.462	863.2
GovEdu	3,558	4.359	1.971	0	16.39

Table 3: Regression Results

VARIABLES	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4	(5) Model 5	(6) Model 6	(7) Model 7
WomenParl	0.0229***	0.00298***	0.00136	0.00473**	0.0116**	0.00608**	0.000883
	(0.00545)	(0.000899)	(0.00110)	(0.00191)	(0.00571)	(0.00307)	(0.00209)
LPFfemale		0.00414***	0.00309***				-0.00234
		(0.000650)	(0.000737)				(0.00393)
GDPpc		2.67e-05***	2.78e-05***				4.48e-06
		(9.33e-07)	(1.06e-06)				(4.74e-06)
Urban		0.00322***	0.00211***				0.0103
		(0.000706)	(0.000770)				(0.00721)
Trade		0.000695**	0.00104***				-0.00146*
		(0.000292)	(0.000344)				(0.000730)
Fertility		-0.156***	-0.148***				0.0515
		(0.00968)	(0.0111)				(0.0644)
TerEnroll		0.00312***	0.00274***				0.000283
		(0.000645)	(0.000788)				(0.00135)
GovEdu			0.0476***			0.0309**	0.00192
			(0.00831)			(0.0149)	(0.0126)
WomenParl2					-0.000142		
					(0.000125)		
WomenParl*GovEdu						-0.00119**	
						(0.000543)	
Constant	-0.594***	-0.492***	-0.587***	-0.0556*	-0.102**	-0.0700	-0.335
	(0.138)	(0.0680)	(0.0778)	(0.0336)	(0.0457)	(0.0783)	(0.516)
Observations	190	2,528	1,993	4,333	4,333	3,016	1,993
Adjusted R-squared	0.078	0.737	0.748	0.007	0.012	0.004	0.015
Country FE	NO	NO	NO	YES	YES	YES	YES
Year FE	NO	NO	NO	YES	YES	YES	YES
Number of Country				193	193	189	150

Robust standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Exploring the Relationship Between Male Unemployment and Male Suicide Mortality Rates Across Countries

By Ekabhat Charnnarong and Rundong Yang

This paper investigates the relationship between male unemployment and suicide a balanced cross-country panel dataset of 143 countries 2001 to 2020. The presence of substantial cross-country heterogeneity in baseline suicide rates and labor market institutions complicates cross-country comparisons. To tackle the issue, country and yearfixed-effects to focus on within-country changes over time. The analysis controls for GDP per capita (PPP), alcohol consumption, internet use, and urbanization. AA log-linear specification is used to model the right-skewed distribution of suicide mortality, hence allowing percentage-based interpretations. Cross-sectional regressions fail to show a statistically significant link between male unemployment and suicide rates. However, after adding country-fixed-effects, the relationship becomes positive and statistically significant. In the preferred two-way fixed-effects log-linear model, a one percentage-point increase in male unemployment is linked to a 1.3 percentage increase in male suicide rates. This result is robust to different model specifications. Although the models explain a small part of the total variation in suicide rates, the results indicate that labor market distress is associated with higher male suicide mortality.

I. Introduction

Globally, suicide remains a major public health problem. Suicide rates vary across nations, with men accounting for a large of suicide deaths. Identifying the factors associated with suicide among men is crucial to policymakers and public health officials. Labor market conditions, especially unemployment, are often cited in the existing literature. However, there is insufficient empirical evidence of a strong cross-country relationship. Several links exist between unemployment and suicide risk, including financial, diminished social status and sense of purpose–as employment often constitutes a core source of identity and social standing, and psychological distress, including depression and anxiety. These mechanisms indicate that unemployment can affect individuals in two ways: loss of income disruption of identity, decreased social integration/connectedness. Empirical research on this relationship has yielded mixed results: some studies demonstrate a positive association between these states, whereas others find weak or non-significant effects when macroeconomic variables are controlled for. This study aims to contribute to the this relationship by conducting a longitudinal analysis a balanced panel of 143 countries from 2001 to 2020. By emphasizing within-country variation over time, this approach captures stronger longitudinal relationships rather than weaker cross-sectional patterns. A major obstacle is substantial cross-country heterogeneity in baseline suicide rates, labor market institutions social safety nets, and reporting practices. Because of this, cross-sectional analyses may obscure within-country dynamics and understate labor market shocks. ongitudinal models better capture how suicide rates adjust to labor market shocks over time. The purpose of this study is to investigate whether an increase in the number of unemployed men in a country is related to a higher male suicide. This examination of cross-sectional data generally does not establish a statistically significant relationship between male unemployment and male suicide rates. However, after controlling for country-fixed-effects, the analysis finds a positive and statistically significant relationship. Using the preferred two-way fixed-effects log-linear model, the estimate indicates that a one-percentage-point increase in male unemployment is associated with an increase of approximately 1.3 percent in male suicide rates. Although the models account for only a small share of the total variation in suicide rates, the consistency of the unemployment coefficient across specifications suggests that labor market conditions are a meaningful correlate of male suicide-a finding with direct implications for public health policy and future research.

II. Literature Review

Numerous studies that worsening economic conditions, particularly rising unemployment, associated with a higher suicide rate. According to the findings of Reeves et al. [Reeves et al., 2014], one of the most comprehensive studies on this topic that examined 20 EU states during the Great Recession, found that with every percentage-point increase in the number of male unemployment, there was a 0.94% increase male suicide rates. The research demonstrated that the effects were concentrated among men aged 25–54 years and that the socio-political context critical; specifically, countries with stronger social safety nets and lower income inequality showed a more attenuated response.

The use of active labor-market programs and high levels of social capital appear to have significantly diminished the strength of the relationship between unemployment and suicide. [Stuckler et al., 2009] examined 26 European countries and documented that rising unemployment was associated with increased suicide rates and other mortality outcomes driven by social stress. Critically, the magnitude of this relationship varied depending on the policy responses of each nation. In countries the implementation of active labor-market policies, the increase in during recession were less pronounced compared to countries that did not invest in these types of programs. In addition to their empirical analyses, [Chang et al., 2013] synthesize worldwide research into suicide during times of economic recession and find that a reliable increase in the rate of suicide occurs during times of labor-market decline, with males disproportionately affected.

Building on this, [Sher, 2005] and [Norström and Rossow, 2016] document that alcohol abuse is itself associated with elevated suicide risk, and that periods of economic stress tend to increase reliance on alcohol as a coping mechanism. This suggests that unemployment may operate through multiple pathways, including both direct financial strain and alcohol-mediated psychological distress. Building on this, [Sher, 2005] and [Norström and Rossow, 2016] document that alcohol abuse is itself associated with elevated suicide risk, and that periods of economic stress tend to increase reliance on alcohol as a coping mechanism. This suggests that unemployment may operate through multiple pathways, including both direct financial strain and alcohol-mediated psychological distress.

Furthermore, many studies utilize suicide rates in logarithmic form n level. [Chen et al., 2009] and [Neumayer, 2003] note that reduce right-skewness and allow coefficients to be interpreted as proportional changes. This is particularly useful when comparing countries with substantially different baseline. Collectively, these studies indicate that existing research consistently that rising unemployment is associated with higher male suicide hough the strength of this relationship depends on broader context.

III. Data Description

Table 1: Variable Descriptions for Mortality Analysis

Variable	**Description**
SuicideRate	Male suicide mortality rate (deaths per 100,000 males)
Unemp	Male unemployment rate
GDP	GDP per capita (PPP, constant international dollars)
Alc	Alcohol consumption (liters per capita, age 15+)
Internet	Internet users (% of population)
Urban	Urban population as a share of total population
Unemp x Alc	Interaction term between unemployment rate and alcohol consumption

This paper uses a balanced cross-country panel of 143 countries from 2001 to 2020, comprising 2,860 country-year observations. The dependent variable is the suicide mortality rate measured as deaths per 100,000 males. Standardized statistical data provided by WHO include standardized adjustments for differences in the levels of under- and over-reporting across countries, thereby improving comparability.

Table 2: Summary Statistics

Variables	(1) N	(2) mean	(3) sd	(4) min	(5) max
Unemp	2,860	7.231	5.498	0.0450	36.97
SuicideRate	2,860	15.05	12.26	0	93.16
GDP	2,860	23,721	24,501	785.8	145,591
Internet	2,860	37.39	30.64	0.000289	100
Alc	2,860	9.482	6.683	0	31.05
Urban	2,860	58.63	22.15	8.461	100

The $maleunemploymentrate$ (%) is the independent variable. Four control variables are included to reduce omitted variable bias. GDP per capita is expressed in dollars at purchasing power parity (PPP). $Internetusers$ (% of the total population) used to capture social connectedness and access to information, since digital access may mitigate isolation-driven distress. $Alcoholconsumption$ is defined as the total liters of pure ethanol consumed by individuals aged 15 and older, expressed as total liters per capita, and is included because heavy drinking is associated with elevated suicide risk. $Urbanpopulation$ is expressed as a percentage of the total population; it is included to control for settlement patterns and differential access to mental health and social services. An interaction term for unemployment and alcohol consumption created to determine if alcohol consumption moderates the relationship between unemployment and male suicide mortality.

IV. Empirical Model

A. Regression Model

To estimate the relationship between male unemployment and male suicide , the following population regression model was constructed using both country and year-fixed-effects:

$$\text{(1)} \quad \ln(\text{Male Suicide Mortality Rate}_{it}) = \beta_0 + \beta_1 \text{Male Unemployment Rate}_{it} + \beta_2(\text{Unemployment}_{it} \times \text{Alcohol consumption}_{it}) + \gamma X_{it} + \alpha_i + \delta_t + \epsilon_{it}$$

where i indexes countries and t years. The dependent variable, $\ln(\text{Male Suicide Mortality Rate}_{it})$, is the natural logarithm of the male suicide mortality rate. Male Unemployment Rate_{it} is the Male Unemployment Rate, the primary explanatory variable, and an indicator of labor-market

The implied primary regression model is:

$$\ln(SuicideRate_{it}) = \beta_1 Unemp_{it} + \beta_2 Alc_{it} + \beta_3(Unemp_{it} \times Alc_{it}) + \beta_4 GDP_{it} + \beta_5 Internet_{it} + \beta_6 Urban_{it} + \alpha_i + \delta_t + \epsilon_{it}$$

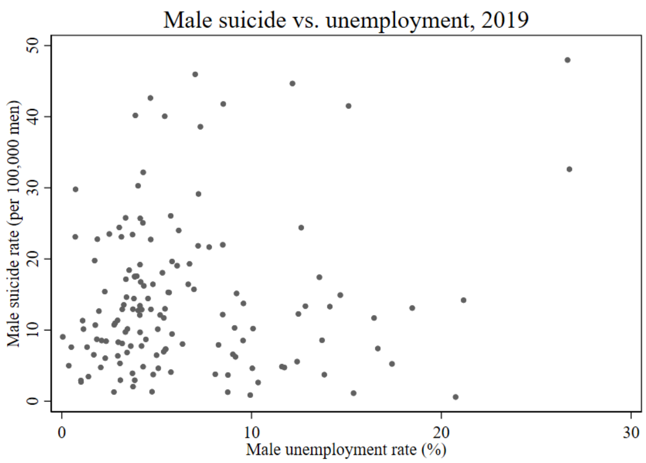
where $\ln(SuicideRate_{it})$ is the logarithmic form of the male suicide mortality rate (Dependent Variable), and $Unemp_{it}$ is the male unemployment rate, which causes distress among men. Alc_{it} denotes alcohol consumption and enters both as a control and through the interaction term ($UnempAlc_{it}$), which tests whether alcohol moderates the effect of unemployment on suicide rates. GDP_{it} (GDP per capita, PPP-adjusted) controls for overall economic development and differences in income level. $Internet_{it}$ (percent of internet users in population) and $Urban_{it}$ (share of the population living in urban areas) represents social connectedness, information availability, settlement patterns, and health care across countries. Fixed effects (α_i) were included in the models to control for time-invariant characteristics specific to each country, including culture, long-term institutional differences, and baseline suicide risk. Fixed effects for each year (δ_t) were included in the models to capture global shocks that affected all countries simultaneously. Standard errors were clustered by country for these models. From the literature review, it is expected that $\beta_1 > 0$, implying that increases in male unemployment is associated with higher male suicide rates. As the dependent variable is in logarithmic form, β_1 can be interpreted as the approximate percentage change in the male suicide mortality rate associated with a one-percentage-point increase in the male unemployment rate, holding constant GDP, alcohol consumption, internet usage, and urbanization, as well as country and year-fixed-effects. The inclusion of α_i and δ_t indicates that identification derives from within-country changes over time, thereby eliminating time-invariant cross-country differences. Specifications including the interaction term $\beta_3 > 0$ if alcohol amplifies the unemployment–suicide link, is expected.

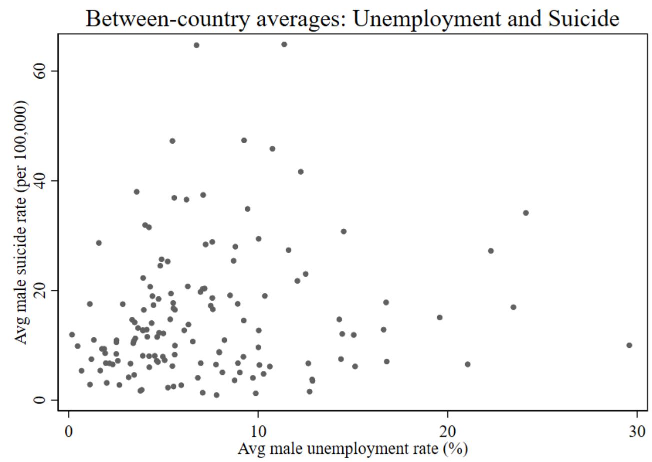
B. Descriptive Evidence

Figures 1 and 2 illustrate why cross-sectional comparisons are insufficient and motivate the use of country and time-fixed effects. Figure 1 shows the cross-sectional relationship between Male suicide mortality rates and Male unemployment rates in 2019. It displays a weak positive correlation between unemployment and suicide rates among males across multiple countries. However, there is quite a lot of diversity in suicide rates despite similar unemployment.

Figure 2 plots each country's long-running averages, plotting the average male suicide rate against the average male unemployment rate for each country. The figure highlights the magnitude of persistent differences across countries, arising from time-invariant features such as cultural context, institutions, and reporting practices, which dominate between-country variance. Since these country-specific features are difficult to measure but may be correlated with the relationship in question, the empirical model employs country and year-fixed-effects.

In addition to these elements, an interaction term (Alcohol Consumption Rate × Unemployment Rate) was also included between the alcohol consumption rate and the unemployment rate. The rationale for this interaction is supported by previous research, such as Sher, L. [Sher, 2005] and Norström, T., & Rossow, I. [Norström and Rossow, 2016], establishing a relationship between alcohol abuse and suicide, as well as designating that economic stress increases reliance on alcohol to cope.





V. Results

Table 1 presents the regression results. Column 1, the results of a simple cross-sectional OLS regression ($Y = \beta_0 + \beta_1(\text{Unemployment}) + \epsilon$) applied to 2019 data. The unemployment coefficient is positive but not statistically significant (p = 0.241). The model also explains very little of the variation in suicide, with an adjusted R-squared = 0.011. This is most likely the result of the dominance of cross-country heterogeneity, as previously shown in Figure 1. Columns 2 and 3 extend the cross-sectional model by using the full panel data and introducing country-fixed-effects ($Y_{it} = \alpha_i + \beta_1(\text{Unemployment}_{it}) + \gamma(\text{Controls}_{it}) + \epsilon_{it}$). The analysis finds that the association between male unemployment and suicide rate remains positive and becomes statistically significant. In both models, the coefficient is significant at the 1% level, and, with the introduction of control variables in model 3, the within R-squared increases substantially (from 0.04 to 0.158). Column 4 introduces the logarithmic form of the suicide mortality rate to the model. Column 5 introduces an interaction term between unemployment and alcohol consumption. While the unemployment coefficient remains positive and statistically significant, the interaction term is small and insignificant. This result suggests that alcohol consumption does not meaningfully alter the relationship, contrary to the proposed model, which is supported by the literature linking the two. Column 6 adds year-fixed-effects, and the estimated relationship remains positive and statistically significant. In the two-way fixed-effects model, the coefficient remains positive and significant (coef = 0.0132, $p < 0.01$).

These results may be attributable to substantial cross-country heterogeneity; therefore, longitudinal statistical techniques should be used to analyze within-country changes over time. In the fixed-effects specification,

Table 3: Main Regression Results

VARIABLES	(1) OLS 2019	(2) FE Y/X	(3) FE Y/X + C	(4) FE lnY + C	(5) FE lnY + Int	(6) FE lnY + YrFE
unemployment_rate	0.281	0.303***	0.230***	0.0125***	0.0172***	0.0132***
	(0.239)	(0.0641)	(0.0524)	(0.00339)	(0.00602)	(0.00347)
gdp_ppp_percapita			-0.000170***	-7.05e-06***	-7.16e-06***	-8.09e-06***
			(5.44e-05)	(1.95e-06)	(1.96e-06)	(2.06e-06)
internet_users_percent			-0.0391**	-0.000871	-0.000873	-0.00136
			(0.0156)	(0.000784)	(0.000784)	(0.000969)
alcohol_consumption_liters			0.316*	0.00807	0.0105*	0.00711
			(0.173)	(0.00641)	(0.00619)	(0.00661)
unemp_alcohol					-0.000388	
					(0.000420)	
urban_population_percent			0.193**	0.00118	0.00143	-0.000711
			(0.0781)	(0.00400)	(0.00402)	(0.00484)
Constant	12.42***	12.86***	4.546	2.361***	2.321***	2.479***
	(1.485)	(0.463)	(4.993)	(0.248)	(0.249)	(0.296)
Observations	143	2,860	2,860	2,858	2,858	2,858
Adjusted R-squared	0.011	0.040	0.158	0.074	0.075	0.078
FE_ctry	NO	YES	YES	YES	YES	YES
FE_year	NO	NO	NO	NO	NO	YES
ClustSE	NO	YES	YES	YES	YES	YES
Number of country		143	143	143	143	143

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

higher unemployment rates within a country are associated with an increase in the suicide mortality rate. Logarithmic specifications of suicide rates reduce the influence of extreme values and allow for a ratio-of-a-values interpretation. When using this preferred model to examine the relationship between male unemployment and male suicide rates, the model estimates that a one-percentage-point increase in male unemployment is associated with an approximately 1.3 percent increase in male suicide rates. An explanation that could account for this association is that as unemployment rises, an individual's financial burden increases, occupational status declines, and the extent to which an individual identifies with their job decreases. Collectively, these mechanisms have been explored in the existing literature and provide some insight into why fluctuations in the labor market correlate with fluctuations in suicide rates. Comparing the functional forms shows that although the level model has higher within-$\mathbf{R^2}$ values than the logarithmic specification, the logarithmic specification yields residuals that follow expected patterns and offer more meaningful interpretations. Nonlinear and interaction-effect tests provide no evidence of curvature or heterogeneity after including fixed effects in the models.

VI. Conclusion

In conclusion, this study examines the relationship between male unemployment and male suicide rates across a balanced cross-country panel of 143 countries from 2001 to 2020. The log-linear regression models show a consistent relation between male unemployment levels and suicide. In the model with fixed effects, a one percentage point increase in male unemployment is associated with an increase of approximately 1.3 percent in male suicide rates. The models explain only a small portion of the total variation in male suicide rates. However, the consistency of the unemployment coefficient across the various regression models provides evidence that labor-market distress is an important correlate for male suicide. The interaction between male unemployment and alcohol consumption was not statistically significant after adjusting for fixed effects. This suggests that the within-country relationship between male unemployment and suicide rates is not substantially altered by alcohol consumption. These findings suggest that policies targeting male unemployment — particularly active labor-market programs, income-support schemes, and job-retention initiatives during downturns — may have downstream benefits for public mental health by reducing the 1.3 percent per percentage-point rise in male suicide rates associated with unemployment identified here. More generally, these findings indicate that labor-market conditions warrant consideration when developing public health strategies. However, other societal and mental health factors also play a significant role in determining male suicide rates within a state.

VII. Limitations and Future Research

First, although the regression models control a variety of economic, technological, demographic, and social variables, there remains the possibility of omitted variable bias. There may be factors influencing unemployment and suicide rates that were not included in the analysis. For instance, institutional quality, access to mental health services, cultural attitudes toward suicide, and social safety-net generosity. Country-fixed-effects may help minimize bias from unobservable characteristics that remain constant over time. However, there may still be bias from time-varying institutional changes that affect the estimates. Second, the explanatory power of the models is relatively low, indicating that there are likely additional psychological, social, and health-related factors influencing male suicide rates that are harder to measure and hence were not included in this analysis. Third, the analysis is based on aggregate national-level data, which masks potential differences across subpopulations, including age groups, regions, and socioeconomic status levels, within each country. Therefore, the relationship between unemployment and suicide may vary extensively across different subpopulations. Future research can enhance this work by using specific data on mental health and social support networks, institutional quality assessments, and other factors that affect mental health support. Doing so should also help account for all unaccounted-for data that may affect the results of this analysis. The study could also be disaggregated by age and region, which could help pinpoint where groups are more vulnerable to the adverse effects of the current economic crisis. Finally, additional studies using effective methods for establishing causal relationships could improve the ability to identify the true causal impact of unemployment on higher rates of suicide. Through these methods, researchers will be better able to fully describe how the work environment affects individuals' mental health.

References

Shu-Sen Chang, David Stuckler, Paul Yip, and David Gunnell. Impact of 2008 global economic crisis on suicide: time trend study in 54 countries. *BMJ*, 347:f5239, 2013. doi: 10.1136/bmj.f5239.

Joe Chen, Yun Jeong Choi, and Yasuyuki Sawada. How is suicide different in japan? *Japan and the World Economy*, 21(2):140–150, 2009. doi: 10.1016/j.japwor.2008.10.001.

Eric Neumayer. Are socioeconomic factors valid determinants of suicide? Controlling for national cultures of suicide with fixed-effects estimation. *Social Science & Medicine*, 57(6):1073–1082, 2003. doi: 10.1016/S0277-9536(02)00485-2.

Thor Norström and Ingeborg Rossow. Alcohol consumption as a risk factor for suicidal behavior: A systematic review of associations at the individual and at the population level. *Archives of Suicide Research*, 20(4): 489–506, 2016. doi: 10.1080/13811118.2016.1158678.

Aaron Reeves, David Stuckler, Martin McKee, David Gunnell, Shu-Sen Chang, and Sanjay Basu. Economic shocks, resilience, and male suicides in the great recession: cross-national analysis of 20 eu countries. *European Journal of Public Health*, 25(3):404–409, 2014. doi: 10.1093/eurpub/cku168.

Leo Sher. Alcohol consumption and suicide. *QJM: An International Journal of Medicine*, 99(1):57–61, 2005. doi: 10.1093/qjmed/hci146.

David Stuckler, Sanjay Basu, Marc Suhrcke, Adam Coutts, and Martin McKee. The public health effect of economic crises and alternative policy responses in europe: an empirical analysis. *The Lancet*, 374(9686): 315–323, 2009. doi: 10.1016/S0140-6736(09)61124-7.

Immigrant Income Under Stricter Enforcement: Evidence from the Trump Administration and Sanctuary States

By Yitong Wang

This paper examines whether stricter immigration enforcement under the Trump administration affected the relative income trajectories of immigrants in the United States. Using American Community Survey data from 2015–2019, I employed linear, difference-in-differences, and triple difference models to compare immigrant and native earnings before and after 2017, across sanctuary and non-sanctuary states. Results indicate that while a persistent immigrant income penalty remains, immigrant incomes grew modestly more than those of natives after 2017, slightly narrowing the gap. However, state-level sanctuary policies did not meaningfully alter these relative gains. These findings challenge the assumption that sanctuary policies provide economic insulation and suggest that changes in immigrant income may be driven by broader labor market dynamics or compositional factors rather than policy alone, highlighting the importance of policies that directly target economic mechanisms.

I. Introduction

This paper investigates whether the stricter immigration enforcement policies implemented during Donald Trump's first presidential term influenced the relative economic standing of immigrants in the United States. While much of the political and scholarly debate has focused on the legal and social repercussions of these policies, systematic empirical evidence on their economic impact, particularly on immigrant wages, remains limited. Moreover, the literature has yet to establish whether sanctuary policies meaningfully moderate these economic effects, a question this paper examines using a triple-difference framework. The 2017–2020 period marked a pronounced shift in federal enforcement, characterized by expanded deportation priorities, increased workplace raids, and the promotion of state-level cooperation through programs such as 287(g). These changes raised widespread concern that heightened enforcement would exacerbate economic vulnerability among immigrant communities, potentially widening the long-standing income gap between immigrants and native-born workers. However, whether these policies ultimately worsened, reshaped, or even partially offset immigrant economic outcomes remains an open question, as labor market adjustments and policy environments may produce countervailing effects or mask underlying vulnerabilities by mitigating larger potential losses.

A key dimension of this debate involves the role of state-level sanctuary policies, which limit cooperation with federal immigration authorities. Proponents argue that such policies may shield immigrants from the economic dislocations of federal enforcement, while skeptics question whether non-cooperation alone can meaningfully alter labor market outcomes. This paper directly tests these competing expectations by asking: Did stricter federal enforcement under Trump affect immigrant income relative to native income? And did sanctuary states moderate this effect?

To answer these questions, the analysis draws on American Community Survey data from 2015–2019 and employs a sequence of empirical strategies, beginning with simple comparisons, then incorporating demographic controls, and finally implementing difference-in-differences and triple difference models. This approach isolates changes in immigrant-native income gaps after 2017 and evaluates whether sanctuary policies meaningfully altered these trajectories.

The results reveal two central findings. First, despite heightened enforcement, immigrant incomes grew modestly more than those of natives after 2017, slightly narrowing the income gap. Second, this relative improvement did not differ significantly between sanctuary and non-sanctuary states. While sanctuary policies may fulfill important non-economic functions, they did not provide detectable economic insulation in the face of federal enforcement. These findings challenge the assumption that sanctuary status meaningfully buffers immigrant earnings and suggest that policies focused solely on non-cooperation may be insufficient to shape economic outcomes, highlighting the need for more proactive, economically integrated approaches to support long-term immigrant wellbeing.

II. Literature Review

Throughout U.S. history, immigrants have played a central role in driving economic growth and contributing to the welfare of native-born populations by expanding the labor force, increasing productivity Blau and Mackie [2017], and fostering innovation that raises overall living standards Peri [2013]. As of 2025, more than 50 million immigrants reside in the United States, yet immigrants continue to face social and political hostility, particularly during periods of restrictive federal policy Kramer and Passel [2025].

A substantial body of economic literature has documented the existence of an immigrant wage gap, whereby immigrants tend to earn less than native-born workers upon arrival. Early work by Barry R. Chiswick Chiswick [1978] suggests that this gap narrows over time as immigrants acquire host-country human capital, such as language skills and labor market experience. However, subsequent research by George J. Borjas Borjas [1985, 1994] demonstrates that this process of earnings convergence is not uniform, emphasizing the role of cohort quality and broader labor market conditions in shaping immigrant wage trajectories. Together, this literature establishes the immigrant income gap as a persistent but dynamic phenomenon, providing an important foundation for examining how policy changes, such as heightened immigration enforcement, may influence immigrant earnings outcomes.

The Trump administration, beginning on January 20, 2017, marked a significant shift in U.S. immigration enforcement (President Donald J. Trump). During this period, federal policy initiatives included expanding 287(g) partnerships, terminating the Deferred Action for Childhood Arrivals (DACA) and Deferred Action for Parents of Americans (DAPA) programs, and creating a specialized Deportation Task Force (Pierce et al., 2018). These measures reflected a broader agenda aimed at strengthening federal immigration enforcement and reducing unauthorized presence in the country.

The Labor Market Effects of Immigration Enforcement finds that heightened enforcement reduces employment among low-skilled noncitizen workers, indicating that stricter immigration policies can directly suppress labor market opportunities and earnings for immigrants (East et al., 2018). These studies suggest that policy interventions targeting immigrants not only shape legal status but also influence economic outcomes such as income, employment stability, and occupational mobility.

Given these documented effects, analyzing the impact of the Trump administration's first-term immigration policies provides a valuable empirical foundation for understanding how stricter enforcement shapes immigrant income. This question remains highly relevant today, as immigration enforcement continues to intensify in the second Trump administration, with ongoing policy changes and increased scrutiny shaping immigrant economic conditions.

III. Data & Methodology

A pooled cross-sectional dataset on the individual level was used for this research paper. All data used in this study was pulled and constructed from the Integrated Public Use Microdata Series (IPUMS). From IPUMS, I used the American Community Survey (ACS), an ongoing survey conducted by the US Census Bureau with a myriad of different economic, social and demographic characteristics of the US population, from 2015-2019. Incorporating data before 2017, the beginning of the first Trump administration, establishes a baseline for comparing immigrant income prior to the implementation of stricter immigration enforcement. Trump's first term ended in 2021, but I decided to not include data from 2020-2021 which were heavily distorted by COVID-19 and therefore cannot provide an accurate reflection of Trump's effect on immigrant income. The sample begins in 2015 to focus on a period when labor market conditions had largely stabilized following the Great Recession, with evidence suggesting that wage growth, particularly for lower-income workers, began to strengthen around this time, providing a more consistent pre-treatment baseline Powell [2021].

For all ACS samples used in this study, I constructed a harmonized dataset containing the following core variables: `survey year`, `statefip`, `sex`, `age`, `race`, birthplace (`bpld`), `yrsusa1`, and total personal income (`inctot`). Using the `bpld` variable, I generated an indicator for immigrant status, coded as $Status = 1$ if respondent born outside the United States and $Status = 0$ if respondent born in the United States. Because the study examines outcomes across multiple survey years, I adjusted all income values for inflation. Specifically, I applied the `CPI99` adjustment factor provided by IPUMS ACS to construct *cpi99income*, which expresses all income values in constant 1999 dollars, thereby ensuring comparability of earnings across the pre- and post-policy periods.

A key component of the empirical strategy is the construction of policy-relevant temporal and geographic indicators. First, I created a *post2017* variable equal to 1 for observations from survey years after the implementation of President Trump's interior immigration enforcement policies in January 2017, and 0 for observations from earlier years. This variable captures the period during which stricter federal immigration enforcement was active and serves as the temporal dimension of the difference-in-differences framework. Second, I constructed a *nonsanctuary* variable based on ACS *statefip* codes, classifying Florida, Texas, and Georgia as high-enforcement, non-sanctuary states. These states were selected because they exhibited greater cooperation with federal immigration authorities during the study period, with the classification of sanctuary and non-sanctuary states defined in detail in the methodology section below.

To control for demographic and assimilation-related differences in labor market outcomes, several individual -level characteristics were included. I used the *sex* variable to construct a dummy variable, *Female*, to identify whether the respondent is male or female. *Age* reports their age in years at the time of interview. The variable *yrsusa1* measures the length of time foreign-born respondents have resided in the United States, which is essential for capturing variation in economic assimilation among immigrant workers. Together, these variables form the foundational dataset for estimating the impact of stricter immigration enforcement on immigrant income.

Table 1: Variable Description

cpi99income	Real income in constant 1999 dollars, adjusted for inflation using CPI99
immigrant	Dummy variable indicating whether the respondent was born outside the U.S. (1 = immigrant, 0 = non-immigrant).
Female	Dummy variable for sex (1 = female, 0 = male).
age	Respondent's age in years as of last birthday.
race	Categorical variable indicating race; multiple categories included, with each category estimated separately in regressions.
yrsusa1	Number of years the respondent has lived in the United States (0 for U.S.-born).
post2017	Dummy variable indicating observations after 2017 ($1 = 2017–2019, 0 = 2015–2016$).
nonsanctuary	Dummy variable indicating non-sanctuary state membership (1 = non-sanctuary, 0 = sanctuary)

Table 2: Summary Statistics

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
age	1.595e+07	41.51	23.68	0	97
yrsusa1	1.595e+07	3.127	10.09	0	95
immigrant	1.595e+07	0.127	0.333	0	1
post2017	1.595e+07	0.605	0.489	0	1
cpi99income	1.329e+07	27,962	42,165	-8,998	1.163e+06
nonsanctuary	1.595e+07	0.204	0.403	0	1
others	1.595e+07	0.233	0.423	0	1
female	1.595e+07	0.511	0.500	0	1

Table 2 presents descriptive statistics for the sample. The mean real income is approximately $27,962, with a standard deviation of $42,165, indicating substantial variation in earnings across individuals. The wide range, from $-8,998$ to over 1.16 million, suggests that a small number of high-income outliers may influence average income, which could motivate the use of log-transformed income or robust regression methods in further analysis. The mean value of the immigrant dummy is 0.127, indicating that roughly 13 percent of the sample are foreign-born, which allows for a meaningful comparison between immigrant and U.S.-born individuals in subsequent models. About 20 percent of respondents live in non-sanctuary states, which is relevant for analyses comparing policy environments. Demographic variables, including age, gender, and others, show variation consistent with a diverse sample and will serve as key control variables in regression models. Overall, these descriptive statistics suggest that there is sufficient variation across income, immigrant status, and demographic characteristics to examine how policy and demographic factors affect earnings outcomes.

IV. Methodology

I began the empirical analysis by estimating a simple linear regression model at the national level to establish association between immigrant status and real income. This baseline model serves as an initial diagnostic because it reveals the raw income differences between foreign-born and U.S.-born individuals before any additional factors are considered. This specification is expanded by incorporating individual-level demographic controls, including sex, age, and years spent in the United States. Adding these variables allows the model to account for observable characteristics that differ systematically between immigrants and non-immigrants, which improves both the accuracy and interpretability of the estimated income gap.

To evaluate whether immigrant income changed following Trump's term in 2017, I implement a difference-in-differences design. For this purpose, I constructed a binary indicator called post2017, which equals 1 for observations collected after the 2017 policy changes (which includes years 2017-2019) and 0 for observations from earlier years (which includes years from 2015-2016). By interacting immigrant status with the post2017 indicator, the model captures the differential change in income for immigrants compared with non-immigrants after the shift in federal enforcement priorities. This approach helps isolate time-related shocks that may have affected immigrants differently and supports a causal interpretation under the assumption of parallel pre-treatment trends, which requires that, in the absence of the 2017 policy change, immigrant and native incomes would have evolved at similar rates over time. However, this assumption cannot be fully tested in this study, as the data include only two pre-treatment years (2015–2016), limiting the ability to assess pre-trends empirically.

In the final stage of the analysis, I expand the empirical framework to a triple-difference model in order to incorporate geographic variation in state-level immigration policy environments. In addition to identifying non-sanctuary states, I constructed a comparison group consisting of major sanctuary states, specifically California, Washington, Oregon, and New York. These states publicly limited cooperation with federal immigration authorities and therefore experienced a substantially different policy context than the non-sanctuary states included in the study. The nonsanctuary variable identifies Florida, Texas, and Georgia as states with higher levels of cooperation with federal immigration enforcement, while the sanctuary states serve as the control group.

There is no universally accepted federal definition of a "sanctuary state," as such policies are more commonly adopted at the city or county level, and full statewide designations are rare. For this study, California, Oregon, Washington, and New York were selected as sanctuary states because each enacted statewide laws or executive directives that limited cooperation with federal immigration enforcement. Oregon became the first sanctuary state in 1987, California passed Senate Bill 54 in 2017 (Senate Bill No. 54, 2017), Washington issued Executive Order 17-01 in 2017 with additional protections codified in the 2019 Keep Washington Working Act (Inslee, 2017), and New York implemented Executive Order 170 in 2017 (Cuomo, 2017).

Similarly, no standard exists for defining non-sanctuary states. To construct a treatment group, I relied on the number of 287(g) agreements, which deputize local officers to enforce federal immigration law. Under Trump, the program expanded significantly (Kolker, 2021). States with the most agreements at the end of Trump's first presidency—Florida (46), Texas (26), Georgia (6)—were classified as non-sanctuary (*Delegation of Immigration Authority section 287(g) immigration and nationality act — ICE*), reflecting strong alignment with federal enforcement priorities. This classification captures meaningful variation in state-level immigration enforcement during the first Trump administration.

By interacting immigrant status, the post2017 indicator, and the nonsanctuary variable, the triple-difference model measures how immigrant income changed over time in non-sanctuary states relative to both non-immigrants and residents of sanctuary states. This approach allows the analysis to separate the effects of federal policy changes from the influence of state-level enforcement regimes. It also provides a more nuanced assessment of how differences in state cooperation with federal immigration authorities shaped economic outcomes for immigrants during the period of intensified enforcement.

Together, the sequence of models, including the baseline regression, the controlled specification, the difference-in-differences design, and the triple-difference model, forms a comprehensive empirical strategy for assessing the impact of stricter immigration policies on immigrant earnings.

V. Model

$$
\begin{aligned}
\text{cpi99income}_{it} =& \beta_0 + \beta_1 \text{immigrant}_{it} + \beta_2 \text{post2017}_t + \beta_3 \text{nonsanctuary}_i \\
&+ \beta_4(\text{immigrant}_{it} \times \text{post2017}_t) \\
&+ \beta_5(\text{immigrant}_{it} \times \text{nonsanctuary}_i) \\
&+ \beta_6(\text{post2017}_t \times \text{nonsanctuary}_i) \\
&+ \beta_7(\text{immigrant}_{it} \times \text{post2017}_t \times \text{nonsanctuary}_i) \\
&+ \beta_8 \text{sex}_i + \beta_9 \text{race}_i + \beta_{10} \text{age}_{it} + \beta_{11} \text{yearsinUS}_{it} + \epsilon_{it}
\end{aligned}
\tag{1}
$$

The model estimates the effects of immigrant status, post-2017 timing, and state-level sanctuary policy on real income (cpi99income) using a triple-difference specification. It includes interactions between immigrant status, post-2017, and non-sanctuary states to capture differential effects across groups and time. Control variables include sex, age, years in the U.S., and race. The coefficient for race β is bolded because multiple categories are included, each with its own estimated effect relative to the reference group.

VI. Results

Before beginning my regression analysis, I first constructed a graph to illustrate the mean income for immigrants versus non-immigrants from 2015 to 2019. The visualization reveals a persistent income gap between the two groups throughout this period, with non-immigrants consistently earning higher average incomes than immigrants. Both groups experienced gradual income growth over the five-year span, but the disparity between them remained relatively stable, providing initial descriptive context for the more rigorous econometric analysis that follows.

Model 1 presents a simple linear regression estimating the unconditional association between immigrant status and real income. Without demographic or socioeconomic controls, immigrants earn, on average, $530 less than natives , a difference significant at the 1

Model 2 adds individual-level demographic controls, including sex, age, race, and years in the United States. With these controls, the estimated income gap for immigrants widens to approximately $1,534, suggesting that part of the baseline difference in Model 1 was masked by demographic composition. Race coefficients show meaningful patterns: Black and African American respondents earn significantly less than Whites, while Chinese, Japanese, and other Asian or Pacific Islander respondents earn substantially more. The adjusted R-squared rises to 0.053, reflecting that demographic factors explain a meaningful portion of income variation, though substantial variation remains unexplained.

Table 3: Regression Results

VARIABLES	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4
immigrant	-530.2***	-1,534***	-2,041***	-3,308***
	(34.38)	(53.12)	(65.20)	(110.3)
Female		-14,367***		
		(23.01)		
age		263.6***		
		(0.476)		
Race (Base: White)				
Black/African American		-10,017***		
		(26.85)		
American Indian/Alaska Native		-11,863***		
		(69.20)		
Chinese		5,655***		
		(123.7)		
Japanese		5,266***		
		(246.6)		
Other Asian/Pacific Islander		3,678***		
		(72.52)		
Other race, nec		-10,904***		
		(39.55)		
Two major races		-5,185***		
		(67.63)		
Three or more major races		-4,203***		
		(178.5)		
Years in the U.S.		39.12***		
		(1.746)		
post2017			989.1***	1,710***
			(24.44)	(65.94)
immi_post2017			891.6***	719.2***
			(67.76)	(124.3)
nonsanctuary				-3,572***
				(69.13)
immi_nonsanc				77.90
				(140.6)
post2017_nonsanc				-848.5***
				(89.64)
immi_post2017_nonsanc				-62.42
				(183.7)
Constant	28,039***	24,059***	23,471***	25,946***
	(12.39)	(26.55)	(30.21)	(64.85)
Observations	13,285,006	13,285,006	13,285,006	5,225,738
Adjusted R-squared	0.000	0.053	0.053	0.055
Controls	No	Yes	Yes	Yes

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

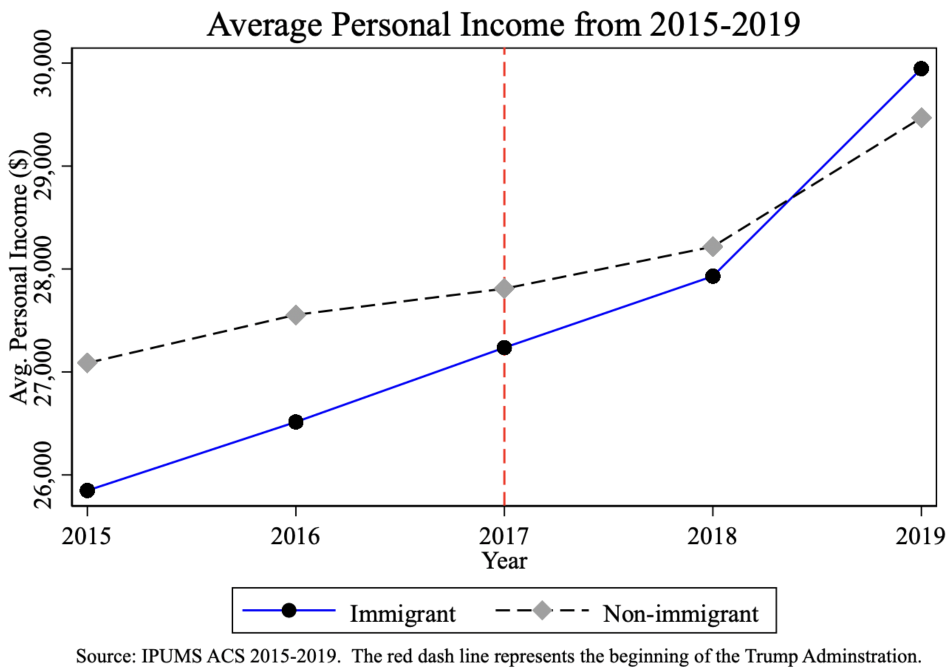


Figure 1: Average Personal Income from 2015-2019

Model 3 introduces the post-2017 indicator and its interaction with immigrant status in a Difference-in-Differences (DiD) design. The post-2017 coefficient shows that average incomes rose by roughly $1,710 nationwide, regardless of immigration status. The DiD interaction indicates that immigrants experienced an additional $719 increase relative to natives, narrowing the income gap modestly. This suggests that stricter federal enforcement did not reduce immigrant incomes relative to natives. One possible explanation is that lower-earning or more vulnerable immigrants exited the labor force or became underrepresented in survey data following increased enforcement, mechanically raising the average income of the remaining immigrant sample. The adjusted R-squared remains 0.053, as demographic controls are unchanged.

Model 4 employs a Triple Difference (DDD) approach to examine state-level differences, comparing non-sanctuary states (Florida, Texas, Georgia) with sanctuary states (California, Oregon, Washington, New York). This specification is estimated on a restricted subsample of these seven states, reducing the sample size from approximately 13.3 million to 5.2 million observations. The baseline immigrant penalty is about $3,308. Post-2017, average incomes rise by roughly $1,710 nationwide, and immigrant incomes grow an additional $719 relative to natives, narrowing the gap. The triple-difference coefficient $immi_post2017_nonsanc$ is small and statistically insignificant, indicating that post-2017 relative gains for immigrants do not differ meaningfully by state sanctuary status.

In summary, Model 4 shows that immigrants earned less than natives on average, but their incomes grew faster than those of natives after 2017. Non-sanctuary states experienced slower overall post-2017 income growth, yet the relative gains for immigrants do not differ significantly between sanctuary and non-sanctuary states, suggesting that federal enforcement shaped immigrant economic outcomes similarly across policy environments.

VII. Discussion

This study examined the economic impact of stricter federal immigration enforcement during the Trump administration on immigrant earnings in the United States. Using difference-in-differences and triple difference models, the analysis reveals several important insights. First, immigrant incomes grew modestly more than those of natives after 2017, slightly narrowing the persistent income gap. This suggests that the relative income gap narrowed during this period, indicating that stricter enforcement did not translate into worse relative economic outcomes for immigrants. However, this improvement should be interpreted with caution, as it may reflect underlying labor market dynamics or compositional changes rather than a direct effect of policy. Second, state-level sanctuary policies did not meaningfully moderate this relative economic trajectory, indicating that federal enforcement shaped immigrant economic outcomes similarly across policy environments.

The triple difference analysis highlights that the relative post-2017 income gains of immigrants did not differ significantly between sanctuary and non-sanctuary states. Consistent with Model 4, the analysis shows that non-sanctuary states experienced slower overall income growth after 2017, while immigrants' incomes grew faster than those of natives on average. Most importantly, state sanctuary policy did not significantly influence immigrants' relative post-2017 gains. This challenges the expectation that sanctuary policies, by limiting cooperation with federal immigration authorities, would provide an economic buffer against enforcement shocks. Instead, the economic trajectory of immigrants relative to natives remained broadly similar regardless of state-level sanctuary status. This finding suggests that differences in state-level cooperation with federal immigration authorities were not sufficient to produce measurable differences in immigrant income outcomes.

These findings carry important implications for immigration and labor policy. While sanctuary policies provide critical protections against deportation and foster trust between immigrant communities and local institutions, they appear insufficient on their own to alter relative income trends in the short term. At the same time, the observed improvement in immigrant incomes relative to natives suggests that economic outcomes may be shaped by broader labor market dynamics or compositional changes, rather than by sanctuary policies themselves. This implies that policies focused solely on non-cooperation with federal enforcement may be insufficient to influence economic outcomes. Enhancing economic integration for immigrant communities may require complementary measures, such as enforceable labor protections, targeted investments in education, workforce development, and social services that more directly engage with labor market mechanisms.

In conclusion, this study demonstrates that although federal enforcement can shape immigrant labor market outcomes, the presence of sanctuary policies alone does not significantly change the short-term income trajectory of immigrants relative to natives. These results suggest that understanding immigrant economic well-being requires looking beyond enforcement and legal protection frameworks to the broader economic forces and policy tools that influence labor market participation and earnings. Taken together, these results suggest that the economic effects of immigration policy are not solely determined by enforcement intensity or legal protection, but by whether policies directly engage with the underlying mechanisms that shape labor market participation and earnings.

VIII. Limitations and Future Research Directions

The study has several limitations that qualify its conclusions. First, the analysis covers a relatively short period (2015–2019), excluding 2020–2021 due to COVID-19 disruptions, which limits the ability to assess longer-term trends. Second, the data primarily capture documented immigrants, leaving the experiences of undocumented populations, who may be most vulnerable to enforcement shocks, unobserved.

Another limitation lies in the selection of sanctuary and non-sanctuary states. For this analysis, sanctuary states were defined as California, Oregon, Washington, and New York, while non-sanctuary states included Florida, Texas, and Georgia. This represents a small subset of U.S. states and may not capture the full variation in sanctuary policy implementation. The selection process was challenging because there is no universally accepted definition of "sanctuary" or "non-sanctuary" status. Non-sanctuary states were identified based on the prevalence of 287(g) agreements during the Trump administration's first term, but this may not fully reflect all forms of cooperation with federal enforcement. Although not exhaustive, the selected states provide meaningful insight into how policy variation may shape immigrant economic outcomes. The observed patterns can inform broader discussions, as these states include some of the largest and most diverse immigrant populations in the U.S., making the results relevant for understanding general trends.

Future research could expand the set of states examined, incorporate more measures of sanctuary policy implementation, and extend the time frame to assess longer-term economic outcomes. Including additional controls such as education, occupation, and industry, as well as capturing the experiences of undocumented immigrants, would further strengthen the evidence base. In particular, this paper helps lay the foundation for future research on immigrant income during Trump's second presidency, when 287(g) programs were heavily utilized, offering an opportunity to study the economic effects of expanded federal-local enforcement cooperation while informing strategies to support immigrant economic well-being.

REFERENCES

Francine D. Blau and Christopher D. Mackie. *The Economic and Fiscal Consequences of Immigration*. National Academies Press, Washington, DC, 2017.

George J. Borjas. Assimilation, changes in cohort quality, and the earnings of immigrants. *Journal of Labor Economics*, 3(4):463–489, 1985. doi: 10.1086/298065.

George J. Borjas. The economics of immigration. *Journal of Economic Literature*, 32(4):1667–1717, 1994.

Barry R. Chiswick. The effect of americanization on the earnings of foreign-born men. *Journal of Political Economy*, 86(5):897–921, 1978. doi: 10.1086/260717.

Stephanie Kramer and Jeffrey Passel. What the data says about immigrants in the u.s., August 2025. URL `https://www.pewresearch.org/short-reads/2025/08/21/key-findings-about-us-immigrants/`.

Giovanni Peri. Migration: The economic benefits of immigration. Technical report, UC Berkeley: Center for Latin American & Caribbean Studies, 2013.

Jerome H. Powell. Speech by chair powell on getting back to a strong labor market, February 2021. URL `https://www.federalreserve.gov/newsevents/speech/powell20210210a.htm`.

The Effect of U.S. Female Manager Proportion on the Gender Wage Gap Within the Finance, Insurance, and Real Estate Industry

By Wanming Zhang

Reducing gender wage gaps has been a widely debated topic in modern economies. This research uses data from 2003 to 2024 to examine whether the proportion of female managers affects the gender wage gap in the U.S. finance, insurance, and real estate (FIRE) industry. Using a panel dataset of sub-industries and a fixed effects model with sub-industry and year fixed effects—while controlling for average age, education, working hours, and private-sector employment—the analysis shows that an increase in the proportion of female managers is associated with a significant narrowing of the gender wage gap.

I. Introduction

The gender wage gap has long been a key focus in labor economics and public policy, especially in the financial industry. Within finance, career hierarchies are relatively rigid and promotion pathways are strict, indicating that the gender composition of management has a systemic influence on the industry's overall wage structure (Bertrand et al., 2019). As global discussions on gender equality and gender quotas intensify, the effectiveness of policies promoting gender equality has drawn significant attention. Among these, whether increasing the proportion of female managers can reduce gender wage disparities and foster broader wage equity remains underexplored empirically. In this context, this paper poses the central question: Does the proportion of female managers in the U.S. finance, insurance, and real estate (FIRE) industry from 2003 to 2024 influence the industry-level gender wage gap? More specifically, does an increase in female managers within the FIRE industry narrow the gender wage gap, thereby reflecting a more equitable compensation structure? Current literature provides both theoretical foundations and analytical motivation for this research. A study synthesizes Norwegian board gender quotas, noting that while increasing women's representation on the board of public limited liability companies does not necessarily have direct positive effects on broader female employees or young women in business, this policy did significantly reduce the gender wage gap within boards (Bertrand et al., 2019). Despite the limited spillover effects observed beyond boardrooms, it is reasonable to hypothesize that the proportion of female managers may exhibit a negative correlation with the gender wage gap. To test this hypothesis, this paper constructs a sub-industry and year panel dataset covering the period from 2003 to 2024, segmented across six sub-industries: banking, savings institutions, credit agencies, security brokerage, insurance, and real estate. This study employs a fixed effects model (incorporating sub-industry and year fixed effects) to control for industry-specific characteristics that do not change over time and common macroeconomic trends. It further includes key control variables influencing compensation structures - average age, proportion with a college degree or higher, average weekly working hours, and the share of private-sector salaried employees - to reduce potential omitted variable bias. The findings reveal a significant negative linear correlation between the main variables, as sub-industries with higher female managerial representation tend to exhibit smaller gender wage disparities. In conclusion, this paper highlights the critical role of female leadership in paving the way for advancing wage equity.

II. Literature Review

Whether the proportion of women in management positions can narrow the gender wage gap has been a highly focused research topic in recent years. Previous literature primarily explored the potential impact of female managers on gender wage equity from various perspectives, such as institutional reforms, organizational management structures, and different leadership styles. Despite differences in research contexts and data frameworks, the studies mentioned above collectively formed the theoretical foundation of this research and identified key perspectives for analyzing gender wage structures within the U.S. FIRE industry.

First, Bertrand et al. (2019) used Norway's 2003 board gender quota policy as an experimental background to assess whether increasing female representation on boards improved internal gender disparities. The findings showed that this policy significantly increased the observable human capital characteristics of newly appointed female directors, including higher educational attainment and stronger earnings ranks, and narrowed the gender wage gap within boards. However, this institutional change did not produce a noticeable "trickle-down" effect - it neither improved the wages of female employees more broadly within firms nor significantly expanded advancement opportunities for young women in business. This conclusion points to an important limitation, indicating that whether the proportion of female managers at the industry level can influence the overall gender wage gap requires further empirical study.

Second, Theodoropoulos et al. (2022) directly examined the relationship between the proportion of female managers and the gender wage gap at the workplace level. Using UK employer-employee matching data, they found that higher proportions of female managers were associated with smaller gender wage gaps, which were effectively eradicated once female managers accounted for more than 60% of workplace management, a scenario observed in roughly one fifth of workplaces. The study further indicated that when female managers possessed greater authority in compensation decisions, they were more likely to raise female employees' wages, thereby narrowing the gender wage gap. Unlike Bertrand et al. (2019), this study provided direct evidence at the organizational level, suggesting that an increase in the proportion of female managers not only reflected structural changes but might also influence gender wage structures through specific decision-making mechanisms. These results suggest that managerial gender composition may influence wage outcomes through decision-making authority.

Third, Tate and Yang [2015] conducted a study on the gender wage gap in the U.S. context. Their study exploited involuntary worker displacement caused by U.S. plant closures as an identification strategy to address endogenous job mobility and isolate wage changes following forced job transitions. They found that displaced women experienced significantly larger wage losses than men when moving to new firms. However, this gender gap was substantially smaller when the hiring firm was led by women, suggesting that female leadership might influence wage outcomes not only through direct compensation decisions but also through broader organizational norms and workplace culture. The effect was strongest among women displaced from male-led plants and from less competitive industries. This finding is particularly relevant for the U.S. FIRE industry, which is typically characterized by high competition and performance-based compensation. This evidence suggests examining whether female leadership continues to play a similarly significant role in reducing the gender wage gap within the highly competitive U.S. FIRE industry.

In summary, existing literature broadly supported the potential of female managers to enhance gender wage equity. Bertrand et al. [2019] demonstrated that institutional quotas primarily improved senior-level representation with limited spillover effects beyond boardrooms; Theodoropoulos et al. [2022] focused primarily on the UK context; while Tate and Yang [2015] focused on less competitive industries in the U.S. Addressing these study gaps, this paper examines whether the proportion of female managers influences gender wage gaps at the industry level using sub-industry panel data from the U.S. finance, insurance, and real estate (FIRE) industry during 2003–2024. This provides new empirical evidence for understanding the macro-level impact of female leadership in advancing gender pay equity.

III. Data Description

The data used in this research are from IPUMS CPS, covering individual worker records employed in the Finance, Insurance, and Real Estate (FIRE) industry in the U.S. from 2003 to 2024. This study interprets panel data at the sub-industry and year level, including six sub-industries: banking, savings institutions, credit agencies, security brokerage, insurance, and real estate, which together contain 132 observations (6 sub-industries × 22 years). After cleaning the individual-level data by excluding observations with missing values and restricting the sample to workers with positive weekly earnings and valid working hours, this research calculated key analytical variables: (1) Dependent Variable-Gender Wage Gap: Based on the standard definition, calculated using the formula

$$GenderWageGap_{it} = \frac{MedianMaleEarnings_{it} - MedianFemaleEarnings_{it}}{MedianMaleEarnings_{it}} \tag{1}$$

for the median weekly earnings of male versus female by sub-industry and year. (2) Independent Variable-Female Manager Proportion: Based on the OCC2010 occupational codes, managerial positions were defined as series 10-430 [Flood et al., 2025]. Subsequently, the proportion of female managers among all managers was calculated for each sub-industry and year combination. Additionally, this research constructed four control variables: average age (AvgAge), proportion with a college degree or higher (CollegeDegree+), average weekly working hours (AvgWeeklyHrs), and proportion of private sector employees (PrivateSecProp), defined based on IPUMS

CPS class-of-worker codes 21–23, which correspond to private wage and salary workers [Flood et al., 2025]. These variables were used to control for differences in labor force composition. Since the analysis is conducted at the sub-industry-year level, individual-level variables such as age and working hours were aggregated using averages. Average age captures variation in workforce experience, while the proportion with a college degree or higher reflects differences in human capital. Average weekly working hours account for variations in labor intensity, and the proportion of private sector employees controls for differences in employment structure. The descriptions for all variables are presented in Table 1.

Table 1: Variable Descriptions

Variable	**Description**
GenderWageGap	Gender wage gap, computed as $\frac{MedianMaleEarnings_{it} - MedianFemaleEarnings_{it}}{MedianMaleEarnings_{it}}$; unit: ratio (typically 0–1).
FemaleManagerProp	Proportion of female managers in each sub-industry and year; unit: ratio (0–1).
FemaleManagerProp2	Squared proportion of female managers.
AvgAge	Average age of workers; unit: years.
CollegeDegree+	Proportion of workers with a bachelor's degree or higher; unit: ratio (0–1).
AvgWeeklyHrs	Average weekly working hours; unit: hours/week.
PrivateSecProp	Proportion of private-sector employees; unit: ratio (0–1).
FemaleManagers*AvgAge	Interaction between female manager proportion and average age; unit: ratio*years.

Table 2 reports descriptive statistics for all regression variables, including sample size, mean, standard deviation, minimum, and maximum values. It is notable that the FIRE industry exhibited several structural characteristics within the given timeframe. First, the mean average gender wage gap is 0.333, indicating that women's median weekly wages were approximately 33% lower than men's, revealing persistent gender wage inequality within the industry. Notably, the minimum value of -0.132 indicates that in some sub-industry and year combinations, women's median weekly earnings exceeded those of men, reflecting meaningful variation across sub-industries over the study period. Second, the average proportion of female managers is approximately 0.495. On average, about half of managerial positions are held by women, though variations exist across sub-industries, as shown in the relationship graph in the next section. Third, FIRE employees generally possess high educational attainment, with an average of 47.3% holding a bachelor's degree or higher, aligning with the industry's demand for specialized skills and qualifications. Furthermore, the proportion of private-sector salaried employees is exceptionally high (mean 0.973), indicating near-total dominance by private institutions and reinforcing the market-driven nature of the industry's operational structure. These aggregate statistical characteristics not only help in understanding the FIRE industry's labor force composition but also provide essential contextual and interpretive frameworks for subsequent fixed effects regression analyses.

Table 2: Summary Statistics

GenderWageGap	132	0.333	0.146	-0.132	0.702
FemaleManagerProp	132	0.495	0.094	0.291	0.733
AvgAge	132	42.447	2.456	35.545	47.688
CollegeDegree+	132	0.473	0.148	0.238	0.868
AvgWeeklyHrs	132	40.823	1.632	36.441	45.220
PrivateSecProp	132	0.973	0.023	0.863	1.000

IV. Model

This research aims to examine whether the proportion of female managers affects the gender wage gap in the U.S. finance, insurance, and real estate (FIRE) industry. Accordingly, a panel fixed effects model is constructed at the sub-industry and year level. The dependent variable is the gender wage gap, which is a continuous variable typically ranging from 0 to 1. The independent variable is the proportion of female managers, also a continuous variable ranging from 0 to 1. If female managers improve gender wage equity, a negative linear relationship is expected between the proportion of female managers and the gender wage gap. To estimate this relationship, the following model is used:

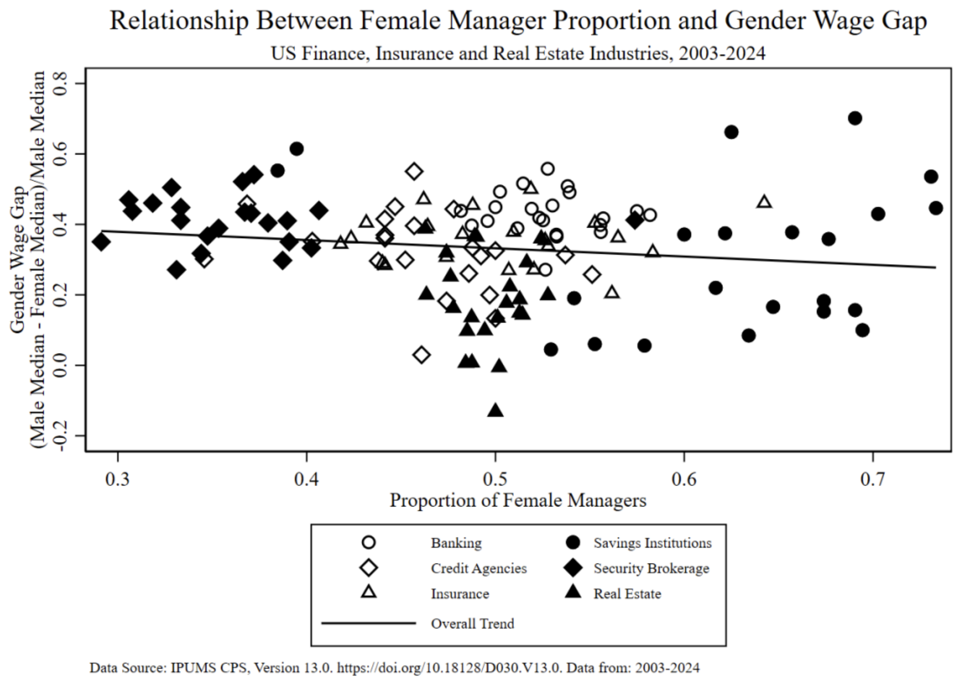
$$\begin{aligned} GenderWageGap_{it} &= \beta_0 + \beta_1 FemaleManagerProp_{it} + \beta_2 AvgAge_{it} + \beta_3 CollegeDegree{+}_{it} \\ &+ \beta_4 AvgWeeklyHrs_{it} + \beta_5 PrivateSecProp_{it} + \alpha_i + \lambda_t + \upsilon_{it} \end{aligned} \tag{2}$$

$GenderWageGap_{it}$ denotes the gender wage gap in sub-industry i in year t, serving as the continuous dependent variable. $FemaleManagerProp_{it}$ represents the independent variable, indicating the proportion of female managers among all managerial positions. $AvgAge_{it}$, $CollegeDegree{+}_{it}$, $AvgWeeklyHrs_{it}$, and $PrivateSecProp_{it}$ represent four control variables that help adjust for differences in workforce age structure, education level, working hours, and private sector share to reduce omitted variable bias. α_i denotes the sub-industry fixed effects, controlling for sub-industry characteristics that do not change over time but may influence wage structures; δ_t represents the year fixed effects, controlling for common shocks across periods such as macroeconomic fluctuations, industry cycles, and institutional environments. ϵ_{it} is the error term. This model was chosen primarily because the gender wage gap is a continuous indicator, making it suitable for estimation via a linear model. Secondly, considering that industry structure, skill requirements, and labor force composition may systematically influence wage levels, incorporating average age, education level, average weekly working hours, and the proportion of the private sector can effectively reduce omitted variable bias. Third, given the long-term structural differences across sub-industries and the shared economic cyclical effects across years, incorporating sub-industry and year fixed effects is essential. Fixed effects models enhance the interpretability of estimates, allowing coefficients to better reflect relationships.

Additionally, this research estimated alternative models including a quadratic term ($FemaleManagerProp^2$) to test for potential nonlinear relationships, and an interaction term ($FemaleManagers \times AvgAge$) to examine whether the effect of female manager proportion depends on average age. However, neither the quadratic term nor the interaction term was statistically significant, and these extended models did not provide a better fit to the data. Consequently, the research focuses on the linear fixed effects panel data model as the primary analytical framework.

To help readers understand the relationship between independent and dependent variables intuitively, this graph shows scatter plots with overall trend lines illustrating the relationship between the proportion of female managers and the gender wage gap. Different shapes in the scatter plot further highlight the distribution characteristics across various sub-industries. The figure shows an overall negative linear correlation between the

Figure 1: Relationship Between Female Manager Proportion and Gender Wage Gap, U.S. FIRE Industry, Year 2003-2024



two variables, providing preliminary visual evidence supporting the expected coefficient sign (< 0). The model and graph reflect the core hypothesis of this research that within the FIRE industry, a higher proportion of female managers correlates with a smaller gender wage gap.

V. Empirical Analysis and Results

Table 3 presents key models and empirical results from this research.

Column (1) reports a basic OLS model using 2018 cross-sectional data as a pre-pandemic cross-sectional reference point. Given the limited number of sub-industries, this model serves only as a preliminary reference. Columns (2) - (4) introduce sub-industry fixed effects and control variables using the full panel (2003–2024). Across these models, the coefficient on FemaleManagerProp remains statistically insignificant. Additional models incorporating a quadratic term and an interaction term were estimated to test for potential nonlinear relationships and to examine whether the effect of female manager proportion on the gender wage gap depends on average age. However, these terms were not statistically significant. The results change notably when year fixed effects are introduced in column (5). After controlling for both sub-industry heterogeneity and time-specific macroeconomic shocks, the coefficient on FemaleManagerProp becomes statistically significant at the 5% level. Columns (6) and (7) present models with fewer control variables. Since the coefficient on PrivateSecProp is statistically insignificant and the descriptive statistics in Table 2 show minimal variation in private sector employment within the FIRE industry, this variable is excluded in column (6). After removing PrivateSecProp, the coefficient on FemaleManagerProp becomes significant at the 1% level. Meanwhile, AvgAge becomes statistically insignificant. Therefore, after excluding AvgAge in column (7), the coefficient on FemaleManagerProp is -0.437 and statistically significant at the 1% level, with an adjusted R^2 of 0.077. Comparing all seven models, the final model demonstrates a statistically significant negative linear correlation between the proportion of female

Table 3: Female Manager Proportion and Gender Wage Gap in US FIRE Industry

VARIABLES	**(1) OLS(2018)**	**(2) FE**	**(3) FE**	**(4) FE**	**(5) FE**	**(6) FE**	**(7) FE**
FemaleManagerProp	-0.897	-0.229	-3.899	-1.598	-0.436**	-0.429***	-0.437***
	(0.830)	(0.137)	(2.159)	(2.686)	(0.113)	(0.103)	(0.101)
FemaleManagerProp2			3.582				
			(1.900)				
AvgAge	0.000878	0.00132		-0.0169	0.00359*	0.00340	
	(0.00371)	(0.00285)		(0.0361)	(0.00150)	(0.00180)	
CollegeDegree+	-0.263	-0.280	-0.240	-0.316	-0.324	-0.311	
	(0.264)	(0.220)	(0.293)	(0.447)	(0.430)	(0.432)	
AvgWeeklyHrs	-0.00644	-0.00419	-0.00655	-0.00564	-0.00536	-0.00573	
	(0.0126)	(0.0116)	(0.0132)	(0.0108)	(0.0101)	(0.00990)	
PrivateSecProp	-0.149	-0.295	-0.110	0.118			
	(0.230)	(0.289)	(0.251)	(0.439)			
FemaleManagers*AvgAge				0.0330			
				(0.0632)			
Constant	0.786*	0.942	1.888***	1.639	0.679	0.788*	0.939**
	(0.367)	(0.585)	(0.361)	(1.832)	(0.397)	(0.352)	(0.304)
Observations	6	132	132	132	132	132	132
Adjusted R-squared	-0.078	0.006	0.058	0.000	0.061	0.070	0.077
Number of ind1990	6	6	6	6	6	6	
Sub-Industry FE	YES	YES	YES	YES	YES	YES	
Year FE	NO	NO	NO	YES	YES	YES	

*Note: Robust standard errors in parentheses; Significance: *** p<0.01, ** p<0.05, * p<0.1*

managers and the gender wage gap within the U.S. FIRE industry during 2003-2024. This finding provides new empirical evidence for understanding the role of female leaders in advancing wage equity at the industry level.

VI. Summary and Conclusion

This study analyzes the relationship between the proportion of female managers and the gender wage gap in the U.S. FIRE industry from 2003 to 2024. Based on panel data covering six sub-industries, and after incorporating sub-industry and year fixed effects along with key control variables, the final regression model demonstrates a significant negative linear relationship between the proportion of female managers and the gender wage gap. Holding other factors constant, a 0.1 increase in the proportion of female managers is associated with a gender wage gap that is approximately 0.0437 smaller. This implies that, within the industry-level structural context, a higher proportion of female managers correlates with a smaller gender wage gap. It also suggests that higher female representation in management is associated with both statistically and economically significant improvements in compensation fairness. This finding complements and extends existing literature. Unlike studies focusing on policy reforms in other countries or less competitive industries, this research provides new industry-level evidence from the more competitive U.S. financial industry. It suggests that the association between female managerial representation and wage outcomes extends beyond the firm level to the broader industry level. From a policy perspective, the findings highlight the importance of advancing women into management roles. Industry-specific leadership development programs, corporate-level promotion system reforms, and macro-level gender equality policies may collectively increase the proportion of female leaders and generate spillover effects that improve equity. Moreover, in highly market-driven environments like the financial industry, the significant negative correlation suggests that gender equality may be advanced not only through mandatory quotas but also through structural changes in management composition and career development pathways. Overall, this research provides new empirical evidence and data support for understanding how female leaders promote gender wage equity at the industry level. It should be noted that the adjusted R^2 of 0.077 indicates that the model explains a relatively small share of the overall variation in the gender wage gap, suggesting that other factors beyond those included here likely contribute to wage disparities within the industry.

VII. Limitations and Future Research

This research has several limitations. First, constrained by the IPUMS CPS data structure, the FIRE industry includes only six sub-industries, resulting in a relatively small sample size for the panel data, which may affect the precision of the estimation results. Second, the gender wage gap indicator is constructed using industry-level median earnings. While this approach ensures comparability across sub-industries and years, it does not allow for an examination of wage disparities across different income segments or occupational positions within industries. For example, it remains unclear whether female executives have significantly improved the wage structure for women in lower-to-middle income segments. Third, the absence of firm-level and management-tier data limits the ability to examine how the impact of female leadership may vary across different levels of managerial authority (such as middle versus senior managers). Without detailed information on organizational hierarchy, it is difficult to assess whether female leaders at different ranks exert varying degrees of influence on wage-setting practices. Future research could integrate employer-employee matching data, firm-level management structure data, or additional policy shock scenarios to explore the mechanisms through which female managers influence wage outcomes for specific groups (for example, entry-level female employees or recent college graduates). Such approaches would help clarify whether the observed correlation operates through changes in compensation decisions, hiring practices, or broader organizational dynamics, thereby providing a more comprehensive understanding of the role of female leaders in advancing wage equity.

References

Marianne Bertrand, Sandra E. Black, Sissel Jensen, and Adriana Lleras-Muney. Breaking the glass ceiling? The effect of board quotas on female labour market outcomes in Norway. *The Review of Economic Studies*, 86(1): 191–239, 2019. URL `https://www.jstor.org/stable/26611511`.

Sarah Flood, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, Daniel Backman, Annie Breton, Grace Cooper, Julia A. R. Drew, Seth Richards, David Van Riper, and Katherine C. W. Williams. IPUMS CPS: Version 13.0 [data set], 2025.

Geoffrey Tate and Liu Yang. Female leadership and gender equity: Evidence from plant closure. *Journal of Financial Economics*, 117(1):77–97, 2015. doi: 10.1016/j.jfineco.2014.01.004.

Nikolaos Theodoropoulos, John Forth, and Alex Bryson. Are women doing it for themselves? Female managers and the gender wage gap. *Oxford Bulletin of Economics and Statistics*, 84(6):1329–1355, 2022. doi: 10.1111/obes.12509.

The Effect of Tobin's q Ratio on Corporate Investment: Evidence from the U.S. Firm Panel Data Set

By Duc Thanh (Timothy) Nguyen

This paper examines the effect of Tobin's q on corporate investment using a panel of 1,962 U.S. firms. Pooled OLS models show an investment elasticity of 0.24, which is robust to financial controls. Notably, non-NYSE/AMEX firms are more sensitive to q than listed firms. Fixed-effects models yield a higher elasticity of 0.28, with Hausman tests favoring fixed over random effects. A fixed-effects IV model produces the largest elasticity (0.34), though tests fail to reject the exogeneity of q. Overall, Tobin's q has a positive, meaningful causal effect on investment, though firm heterogeneity and financial variables remain significant factors.

I. Introduction

Corporate investment plays a central role in shaping productivity, long-run growth, and firm-level competitiveness. When firms decide how much to invest, they compare the cost of installing new capital with the value that financial markets place on their existing assets. Tobin's q ratio–the market value of the firm relative to the replacement cost of its assets–provides a natural summary of these expectations. If financial markets correctly anticipate future profitability, then a higher Tobin's q should be associated with higher investment, while firms with low q should cut back.

A large empirical literature has tested this prediction by estimating "q-models" of investment, but the evidence is mixed. Many studies find a positive and statistically significant relationship between investment and Tobin's q, yet the explanatory power of q alone is often modest, and investment also appears to depend on variables such as cash flow, leverage, and R&D. At the same time, concerns have been raised that average Tobin's q may be endogenous: it may reflect not only fundamentals, but also time-varying risk, mispricing, or feedback from investment itself. These issues make it difficult to interpret simple correlations between q and investment as evidence of a strong causal effect.

This paper revisits the effect of Tobin's q on corporate investment using a detailed panel dataset of U.S. corporations that includes information on investment, market value, assets, cash flow, debt, R&D, and advertising. Rather than modeling stock-market valuation as the outcome, I treat Tobin's q, constructed as a value-to-assets ratio, as the key regressor and study how firm-level investment responds to changes in q once I control for financial variables and unobserved firm characteristics. The central research question is: To what extent does Tobin's q have a causal effect on firms' investment-to-assets ratios, once we control for firm characteristics, financial variables, and time-invariant unobserved heterogeneity?

To answer this question, I estimate a sequence of models that progressively impose more structure and address more sources of bias. I begin with simple pooled regressions of investment on q, then add additional firm-level controls and interactions. I then exploit the panel nature of the data by estimating fixed-effects models that control for time-invariant firm heterogeneity and common year shocks. Finally, I use instrumental variables derived from present-value-of-dividends measures to address potential endogeneity of Tobin's q. This approach allows for a systematic comparison between ordinary least squares, fixed-effects, and fixed-effects IV estimates of the q-investment relationship.

The rest of the paper is organized as follows. Section 2 reviews the theoretical and empirical literature on Tobin's q, investment, and financing constraints. Section 3 describes the U.S. firm panel dataset and presents summary statistics. Section 4 sets out the econometric framework and modelling strategy. Section 5 reports the empirical results, and Section 6 discusses their economic interpretation, limitations, and implications. Section 7 concludes.

II. Literature Review

Tobin's q-theory of investment predicts that firms invest until the market value of an extra unit of capital equals its replacement cost, so investment should rise with the ratio of market value to replacement cost. Hayashi shows that under restrictive conditions - constant returns to scale and a particular form of adjustment costs - marginal q equals average q, which can be measured as a market-value-to-assets ratio. This result justifies using average Tobin's q in empirical investment equations, but it also highlights that violations of these assumptions, as well as measurement error in capital and market value, can weaken the link between average q and true investment opportunities.

A common empirical strategy in this literature is to regress an investment ratio on Tobin's q and additional controls such as cash flow. For example, Blundell et al. [1992] estimate panel regressions of the form

$$\frac{I_{it}}{K_{it-1}} = \alpha_i + \lambda_t + \beta q_{it} + \gamma \frac{CF_{it}}{K_{it-1}} + u_{it} \quad (1)$$

using a large panel of manufacturing firms and finding the coefficient on q is positive and statistically significant, but economically modest, and that including cash flow improves the fit of the model. Bond et al. [2003] use similar investment equations for firm-level panels from several European countries and also conclude that Tobin's q has a limited ability to explain investment dynamics once other financial variables are included. Relative to these studies, my preferred fixed-effects estimates in Section 5 suggest a somewhat larger elasticity of investment with respect to Tobin's q–about 0.28 in the within-firm OLS model and about 0.34 when Tobin's q is instrumented – but like the previous literature, my models also have fairly low R^2 and indicate that q is only one among several determinants of investment.

The dataset used in this paper comes from Hall and Hall [1993], who study the "value and performance" of U.S. corporations by comparing stock-market valuations with the present discounted value of dividends. Their focus is on explaining valuation gaps using variables such as investment, R&D, and leverage, rather than on estimating an explicit investment equation. In contrast, this paper takes Tobin's q (constructed as a value-to-assets ratio) as the key regressor and examines how investment responds to changes in q in a firm-level panel. Following the financing-constraints literature, I include cash flow, debt, and R&D and advertising intensities as controls, and I address concerns about the endogeneity of q using fixed effects and instruments derived from Hall and Hall's present-value-of-dividends measures. This design allows a reassessment of how large and robust the effect of Tobin's q on corporate investment is in the U.S. firm panel dataset.

A. *Theoretical Foundations*

The theoretical basis for using Tobin's q as a determinant of investment originates with Tobin [1969] and Brainard and Tobin [1968], who argued that firms should invest when the market valuation of capital exceeds its replacement cost. The formal link between marginal and average q was established by Hayashi, who demonstrated that under constant returns to scale and convex adjustment costs, the observable average q is a sufficient statistic for investment. However, Abel and Blanchard [1986] showed that when adjustment costs are more general or when firms face irreversibility constraints, the gap between marginal and average q can be substantial, introducing measurement error into empirical investment equations.

Subsequent theoretical work has refined these insights. Abel and Eberly [1994] develop a model with irreversible investment and show that the option value of waiting generates a wedge between q and the investment threshold, implying that average q may systematically overstate investment incentives when uncertainty is high. Caballero and Engel [1999] introduce heterogeneous firms with lumpy adjustment and demonstrate that aggregate investment dynamics can differ markedly from the smooth adjustment implied by standard q models. These theoretical developments underscore the importance of controlling for firm heterogeneity and using panel methods, as I do in this paper.

B. *Empirical Evidence on Tobin's q and Investment*

The empirical literature on q and investment is extensive but yields mixed conclusions regarding the strength of the relationship. Early work by Summers [1981] found that q models explain relatively little of the variation in aggregate investment, a finding that has persisted in many subsequent studies. Blundell et al. [1992] estimate panel regressions using a large sample of UK manufacturing firms and report a positive but economically modest coefficient on q, with cash flow providing additional explanatory power. Bond et al. [2003] extend this analysis to firm-level panels from Belgium, France, Germany, and the United Kingdom and also find that q has limited ability to explain investment dynamics once financial variables are included.

A key methodological concern in this literature is measurement error in Tobin's q. Erickson and Whited [2000] demonstrate that standard proxies for q are noisy measures of true marginal q and that this measurement error can severely attenuate the estimated coefficient in investment regressions. They propose higher-order moment estimators to address this problem and find substantially larger q coefficients after correction. Cummins et al. [2006] use analysts' earnings forecasts as an alternative measure of fundamentals and report a much stronger relationship between investment and q when measurement error is reduced. These findings motivate my use of instrumental variables in Section 5, where I employ the present discounted value of dividends and asset-pricing terms from Hall and Hall [1993] to address potential attenuation bias.

C. Financial Constraints and Investment

A central debate in the investment literature concerns the role of financial constraints. Fazzari et al. [1988] argue that the sensitivity of investment to cash flow, after controlling for q, reflects financing frictions: firms that face higher costs of external finance rely more heavily on internal funds. This interpretation was challenged by Kaplan and Zingales [1997], who show that the relationship between financial constraints and investment-cash flow sensitivity is not monotonic, casting doubt on the use of cash-flow coefficients as indicators of financing constraints.

The financial constraints literature has since developed in several directions. Whited [1992] demonstrates that debt capacity and credit market imperfections can generate investment-cash flow sensitivity even in well-specified q models. Gilchrist and Himmelberg [1995] use the fundamental component of q estimated from a VAR framework and find that cash flow retains explanatory power after removing measurement error in q, supporting the financing-constraints interpretation. More recently, **?** propose the SA index based on firm size and age as a more reliable measure of financial constraints than earlier classification schemes. In my analysis, I include cash flow, leverage, and an NYSE/AMEX listing dummy as controls to capture different dimensions of financial frictions and test whether the q-investment relationship varies with firms' likely access to external capital markets.

III. Data Description And Summary Statistics

A. Data Source and Sample

The data for this paper come from the firm-level panel compiled by Hall and Hall [1993], "The Value and Performance of U.S. Corporations," downloaded from Bronwyn Hall's data archive. The dataset contains accounting and stock-market information for U.S. corporations identified by the first six digits of their CUSIP code and followed annually from 1960 to 1991.

In the Stata implementation, I construct a firm identifier from the CUSIP code and declare the data as a panel with firm and year as the two dimensions. This yields an unbalanced panel with 1,962 firms and 27,566 firm-year observations in total. Firms enter and exit the sample at different dates and are observed for varying numbers of years, reflecting listing and delisting on U.S. stock exchanges. My estimation sample drops observations with missing values in the variables used in the baseline specification (investment, Tobin's q, cash flow, debt, R&D, advertising, and sales). All variables used in the paper come from the same web source (Hall and Hall's pstar dataset); any additional transformations, such as logs and interactions, are constructed by me in Stata. I also construct two time dummies to capture the major oil-price shock episodes in the sample. The variable $d7374$ equals 1 for observations in 1973 and 1974 and 0 otherwise. The variable $d7981$ equals 1 for observations in 1979, 1980, and 1981 and 0 otherwise. These indicators are designed to absorb the common shifts in investment associated with the 1973–74 and 1979–81 oil crises and the associated recessions. In the pooled regressions without full year fixed effects, I include these oil-crisis dummies as additional controls.

B. Summary Statistics

For detailed summary statistics and figures, see Appendix 1–4. The estimation sample contains 27,566 firm-year observations. Average investment intensity is about 0.10 of beginning-of-year assets (sd ≈ 0.08, range 0–2.38). Average Tobin's q is 1.62 but with a very wide range (0-407), and annual sales average about 730

Table 1: Core Values Used in the Analysis

Variable	**Description**	**Type**	**Units**	**Role in analysis**	**Min.**	**Max.**
ln_inva	Log investment / beginning-of-year assets	Continuous	Log of ratio	Dependent variable	-8.047	0.867
ln_vala	Log Tobin's q (total market value / assets)	Continuous	Log of ratio	Key regressor	-8.047	6.009
cfa	Cash flow / assets	Continuous	Ratio (share of assets)	Control: internal finance	-5.726	17.785
debta	Long-term debt / assets	Continuous	Ratio (share of assets)	Control: leverage	-0.009	9.545
rnda	R&D / assets	Continuous	Ratio (share of assets)	Control: R&D intensity	0.000	12.853
adva	Advertising / assets	Continuous	Ratio (share of assets)	Control: advertising intensity	0.000	6.496
ln_sales	Log annual sales	Continuous	Log of million USD	Control: firm size	-6.908	11.179
nyseamex	1 if listed on NYSE or AMEX, 0 otherwise	Dummy (0/1)	Indicator	Control; also used in interaction with ln_vala	0.000	1.000
inva	Investment / beginning-of-year assets	Continuous	Ratio (share of assets)	Underlying variable used to form ln(inva)	0.000	2.380
vala	Tobin's q: total market value / assets	Continuous	Ratio (no units)	Underlying variable used to form ln_vala and vala2	0.000	407.094
sales	Annual sales	Continuous	Million USD	Underlying variable used to form ln(sales)	0.001	71643.38
vala2	Tobin's q squared	Continuous	Ratio squared	Used in quadratic specification	0.000	166000.0
ln_vala_nyse	Log Tobin's q $\times$ NYSE/AMEX dummy	Continuous	Log of ratio	Interaction term	-8.047	6.009
pstar	Present discounted value of dividends	Continuous	Million USD	Instrument for log Tobin's q in IV	0.026	539.690
h0	Asset-pricing term h0	Continuous	Index	Instrument for log Tobin's q in IV	0.000	94.311
h1	Asset-pricing term h1	Continuous	Index	Instrument for log Tobin's q in IV	0.000	84.524
oil7374	Oil-crisis dummy: 1973–1974	Dummy (0/1)	Indicator	Time dummy for first oil-crisis period	0.000	1.000
oil7981	Oil-crisis dummy: 1979–1981	Dummy (0/1)	Indicator	Time dummy for second oil-crisis period	0.000	1.000

million USD, again with substantial dispersion. The log-transformed variables ln(inva), ln(vala), and ln(sales) have much tighter distributions and are used in the main regressions.

The extreme observations come from only a handful of firm-year pairs. The lowest investment-to-assets ratio in the estimation sample is 0.00032 for firm 636635 (CUSIP 636635) in 1977, while the highest value, 2.38, occurs for firm 250568 (CUSIP 250568) in 1990. Similarly, Tobin's q ranges from essentially zero - 0.00032 for firm 265720 (CUSIP 265720) in 1969 - to about 407.09 for firm 911843 (CUSIP 911843) in 1990. These extreme observations motivate the use of log transformations and robustness checks, but they represent only a tiny fraction of the 27,566 firm-year observations in the panel.

Appendix 2 shows that log Tobin's q has a roughly symmetric interquartile range around zero but a number of extreme low and high outliers. Appendix 4 plots $ln(inva)$ against $ln(vala)$ and reveals a clear but noisy positive slope, consistent with q-theory. The correlation matrix in Appendix 3 confirms a moderate correlation between *inva* and *vala* (about 0.33), strong comovement between cash flow and advertising intensity, and reasonably low correlations between *vala* and the proposed instruments, while the instruments are correlated with Tobin's q, as desired. The moderate but clearly positive association between $ln(inva)$ and $ln(vala)$ in Appendix 4 supports using a log–log specification as one of the main functional forms in Section 4.

IV. Econometric Framework And Modelling

A. Goal and Main Variables

The goal of this paper is to estimate the causal effect of a firm's stock-market valuation, measured by Tobin's q, on its investment intensity. I use a panel of U.S. corporations indexed by $i = 1, ...,$ N and years indexed by t = 1,..., N. The dependent variable is the investment–to–assets ratio - $inva_{it}$ defined as investment during fiscal year t divided by beginning-of-year assets for firm i. The key explanatory variable is Tobin's q, $vala_{it}$ defined as the ratio of the firm's market value to the book value of its assets.

I also include a set of control variables that may affect investment and may be correlated with Tobin's q: cfa (cash flow/assets), *debta* (long-term debt/assets), *rnda* (R&D/assets), *adva* (advertising/assets), *sales* (later used in logs as a size proxy), and *nyseamex* (a dummy equal to 1 if the firm is listed on NYSE or AMEX and 0 otherwise).

In addition, in the pooled regressions I include two time-dummy variables that capture the major oil-price shock episodes in the sample: $d7374$ (dummy equal to 1 for observations in 1973–1974, 0 otherwise), $d7981$ (dummy equal to 1 for observations in 1979–1981, 0 otherwise).

The coefficient on $vala_{it}$ will be interpreted as the change in the firm's investment intensity when Tobin's q increases, holding the other variables constant.

B. Simple Regression and Functional Form

I begin with a very simple pooled regression that relates the investment-to-assets ratio to Tobin's q and ignores both the panel structure and the other firm-level controls:

Model 1 (M1):

$$inva_{it} = \alpha + \beta vala_{it} + \upsilon_{it} \tag{2}$$

Here υ_{it} is an error term that captures all remaining determinants of investment. Under the assumption $E[\epsilon_{it}|vala_{it}] = 0$, ordinary least squares (OLS) yields an unbiased estimate of β. I expect $\beta > 0$: when the market values a firm highly relative to its replacement cost, investment opportunities are more attractive and firms should invest more. Model (M1) provides a useful benchmark but is likely too simple, because it ignores potential non-linearities, omitted firm characteristics, and the panel nature of the data.

To examine whether a linear specification is appropriate, I plot investment against Tobin's q and add a fitted line. If the relationship appears curved, I estimate richer functional forms. One possibility is a quadratic in Tobin's q, which allows the marginal effect of q on investment to be stronger at low or moderate values of q and then flatten at higher values. Economically, this would be consistent with firms responding sharply when valuation first signals profitable opportunities, but less strongly once the most attractive projects are already being funded.

Model 2 (M2):

$$inv_{it} = \beta_0 + \beta_1 q_{it} + \beta_2 q_{it}^2 + \epsilon_{it} \tag{3}$$

A negative coefficient β_2 would indicate that the marginal effect of Tobin's q on investment declines at high values of q. Because both $inv/assets$ and q are positive for most firm-year observations, I also consider a log–log specification:

Model 3 (M3):

$$\log(inv/assets)_{it} = \beta_0 + \beta_1 \log(q_{it}) + e_{it} \tag{4}$$

In this model, β_1 is an elasticity: it gives the percentage change in the investment-to-assets ratio associated with a 1 percent change in Tobin's q. I compare the goodness of fit of models (M1) – (M3), using measures such as R^2 and the root mean squared error, and I select the functional form that fits best and is economically sensible. That chosen form becomes the baseline specification for the rest of the analysis. In all simple regressions I compute heteroskedasticity-robust standard errors.

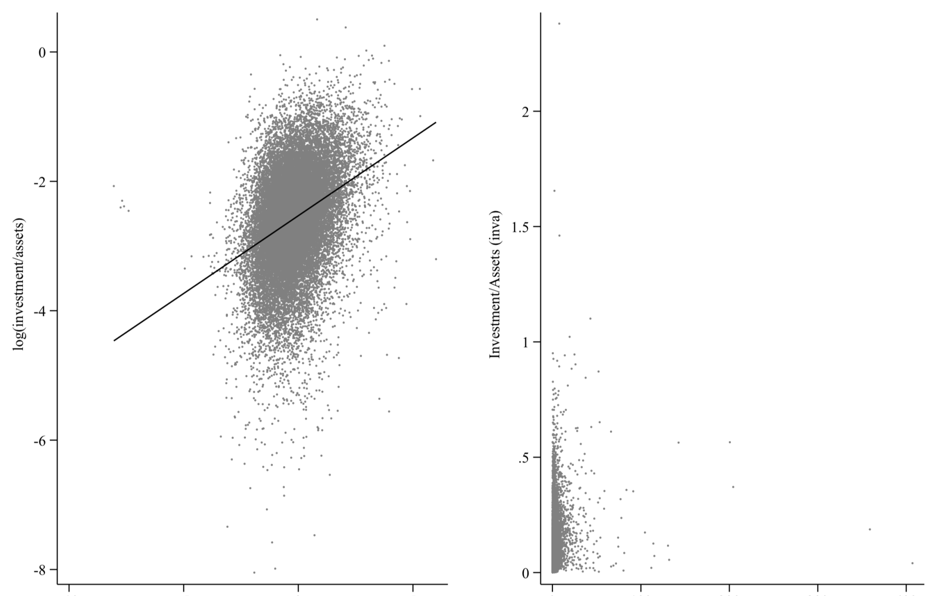


Figure 1: Investment vs Tobin's Q: Raw vs Log-Log Comparison

Figure 1 plots log(investment/assets) against log(Tobin's q) for all firm–year observations, with a fitted regression line. The cloud of points slopes upward, indicating a clear positive association between q and investment intensity. However, the relationship is quite noisy: most observations are concentrated in a dense oval around the center, with substantial vertical dispersion and a few extreme outliers at very low q. This pattern suggests that q helps explain investment but that additional controls and a flexible (log–log) functional form are needed to capture the underlying relationship more accurately.

C. Multiple Regression with Controls and Interaction

The next step is to extend the preferred simple model by adding firm-level controls that may influence investment and are plausibly correlated with Tobin's q. Using the log–log form as the baseline, the pooled multiple-regression specification is

Model 4 (M4):

(5)

$$\text{inva}_{it} = \beta_0 + \beta_1 \ln(\text{vala})_{it} + \beta_2 \text{cfa}_{it} + \beta_3 \text{debta}_{it} + \beta_4 \text{rnda}_{it} + \beta_5 \text{adva}_{it} + \beta_6 \ln(\text{sales})_{it} + \beta_7 \text{oil7374}_t + \beta_8 \text{oil7981}_t + \epsilon_{it}$$

Comparing β_1 in (M4) with the coefficient on $\log(q_{it})$ in the simple log–log model (M3) reveals the extent of omitted-variable bias in the simple regression. If the coefficient on Tobin's q changes substantially after adding cash flow, leverage, R&D, advertising, size, and the oil-crisis dummies, this indicates that these variables were confounding the relationship between valuation and investment. The coefficients γ_1 and γ_2 capture average shifts in investment intensity during the 1973-74 and 1979-81 oil-crisis periods relative to other years. In these pooled multiple regressions I continue to use heteroskedasticity-robust standard errors.

The responsiveness of investment to Tobin's q may differ between firms that are listed on NYSE/AMEX and firms that are not. To allow for this possibility, I add an interaction between $\log(q_{it})$ and the NYSE/AMEX dummy:

Model 5 (M5):

$$\begin{aligned} \ln(\text{inva})_{it} = & \gamma_0 + \gamma_1 \text{vala}_{it} + \gamma_2 \text{nyseamex}_{it} + \gamma_3 (\text{vala}_{it} \times \text{nyseamex}_{it}) \\ & + \gamma_4 \text{cfa}_{it} + \gamma_5 \text{debta}_{it} + \gamma_6 \text{rnda}_{it} + \gamma_7 \text{adva}_{it} \\ & + \gamma_8 \ln(\text{sales})_{it} + \gamma_9 \text{oil7374}_t + \gamma_{10} \text{oil7981}_t + \epsilon_{it} \end{aligned} \tag{6}$$

In specification (M5), the elasticity of investment with respect to Tobin's q for firms that are not listed on NYSE/AMEX (that is, $N_{it} = 0$) is equal to β_1. For listed firms ($N_{it} = 1$), the elasticity is $\beta_1 + \beta_2$. A test of the null hypothesis $H_0 : \beta_2 = 0$ shows whether the response to Tobin's q differs systematically between listed and non-listed firms.

D. Panel Data and Fixed Effects

The dataset follows each firm over several years, which means that unobserved time-invariant characteristics, such as long-run management quality or corporate culture, may affect both Tobin's q and investment. If these characteristics are not taken into account, the pooled regressions may suffer from bias.

To address this concern, I estimate a fixed-effects model that includes a separate intercept for each firm and a separate intercept for each year:

Model 6 (M6):

$$\ln(\text{inva}_{it}) = a_i + \lambda_t + \theta_1 \ln(\text{vala}_{it}) + \theta_2 \text{cfa}_{it} + \theta_3 \text{debta}_{it} + \theta_4 \text{rnda}_{it} + \theta_5 \text{adva}_{it} + \theta_6 \ln(\text{sales}_{it}) + \epsilon_{it} \tag{7}$$

The term α_i is a firm fixed effect that captures all time-invariant differences between firms, and δ_t is a year fixed effect that captures shocks common to all firms in year t, such as macroeconomic conditions. In particular, the year dummies absorb the effects of aggregate events like the 1973-74 and 1979-81 oil crises. Because D_{73-74} and D_{79-81} are just specific combinations of the year dummies, including them in the fixed-effects models would be redundant; in practice, they are perfectly collinear with λ_t and are dropped by the software. The fixed-effects estimator therefore uses within-firm changes over time, with year fixed effects capturing all common macro shocks, to identify δ_1.

For comparison, I also estimate a random-effects version of (M6). A Hausman test is then used to decide whether the random-effects assumptions are plausible; if the test rejects random effects in favour of fixed effects, I treat the fixed-effects estimates as my main specification. In all panel models I cluster standard errors at the firm level to allow for arbitrary serial correlation within firms.

E. Endogeneity of Tobin's q and Instrumental Variables

Even after controlling for fixed firm and year effects and adding observed covariates, Tobin's q may remain endogenous in the investment equation. Investment shocks today may influence future stock prices and therefore future q, and time-varying investment opportunities or risk factors that are not fully captured by the controls may affect both q and investment. Measurement error in Tobin's q can also bias the estimated coefficient towards zero.

To address these issues, I use a fixed-effects instrumental-variables (IV) approach. I treat q as an endogenous regressor and use instruments that are strongly related to Tobin's q but plausibly affect current investment only through their impact on q. In this dataset, I use the present discounted value of dividends, denoted $pstar_{it}$, and two additional variables derived from the asset-pricing relation, denoted $h0_{it}$ and $h1_{it}$, as instruments.

In the first stage, I model the endogenous regressor q as a function of the instruments, the observable firm characteristics, and firm and year fixed effects. The fixed-effects first-stage regression is

$$\ln(\text{vala}_{it}) = a_i + \lambda_t + \pi_1 \text{pstar}_{it} + \pi_2 \text{h0}_{it} + \pi_3 \text{h1}_{it} + \pi_4 \text{cfa}_{it} + \pi_5 \text{debta}_{it} + \pi_6 \text{rnda}_{it} + \pi_7 \text{adva}_{it} + \pi_8 \ln(\text{sales}_{it}) + \nu_{it} \tag{8}$$

Here α_i and λ_t are firm and year fixed effects, and v_{it} is the first-stage error term. The coefficients on $pstar_{1it}$, $h0_{it}$, and $h1_{it}$ capture how the present value of dividends and the pricing terms shift Tobin's q; I assess instrument relevance using the first-stage F-statistic on these three variables.

The IV fixed-effects model has the same "second-stage" form as the fixed-effects regression in (M6), but $\ln(vala_{it})$ is now instrumented:

Model 7 (M7):

$$\ln(\text{inva}_{it}) = a_i + \lambda_t + \theta_1 \ln(\text{vala}_{it}) + \theta_2 \text{cfa}_{it} + \theta_3 \text{debta}_{it} + \theta_4 \text{rnda}_{it} + \theta_5 \text{adva}_{it} + \theta_6 \ln(\text{sales}_{it}) + \epsilon_{it} \tag{9}$$

In practice, $ln(vala_{it})$ is first regressed on the instruments, the controls, and the fixed effects, and the predicted component $\hat{q}$ is then used in equation (M7). I check instrument strength using the first-stage F-statistic reported by the software and, when there are more instruments than endogenous variables, I perform an over-identification test to assess whether the instruments as a group are consistent with the model assumptions. The coefficient θ_1 in model (M7) is my main estimate of the causal effect of Tobin's q on corporate investment. In the results section, I compare this IV fixed-effects estimate with the OLS and fixed-effects estimates to show how accounting for endogeneity changes the conclusions.

V. Results

Table 2: Investment / Assets and Tobin's Q: Pooled OLS, Fixed Effects, and FE-IV

Variable	**(1) Linear inva–q**	**(2) Log–log**	**(3) Pooled + controls + interaction**	**(4) FE (firm, year FE)**	**(5) FE–IV (ln_vala IV)**
ln_vala	—	0.240***	0.277***	0.280***	0.339***
		(0.004)	(0.011)	(0.014)	(0.059)
vala	0.003***	—	—	—	—
	(0.001)				
ln_vala × NYSE/AMEX	—	—	-0.040***	—	—
			(0.011)		
cfa	—	—	0.206***	0.051	-0.011
			(0.035)	(0.045)	(0.077)
debta	—	—	0.259***	0.123***	0.133***
			(0.022)	(0.026)	(0.029)
rnda	—	—	0.220***	0.133***	0.116*
			(0.058)	(0.045)	(0.063)
adva	—	—	-0.950***	-0.063	-0.011
			(0.079)	(0.115)	(0.135)
ln_sales	—	—	0.079***	0.168***	0.176***
			(0.003)	(0.020)	(0.021)
NYSE/AMEX dummy	—	—	-0.147***	—	—
			(0.011)		
oil7374	—	—	0.207***	—	—
			(0.015)		
oil7981	—	—	0.088***	—	—
			(0.012)		
Constant	0.094***	-2.531***	-2.956***	-3.044***	-3.089***
	(0.001)	(0.004)	(0.020)	(0.190)	(0.195)
N	27,566	27,566	27,566	27,566	27,566
R^2 / within R^2	0.031	0.108	0.162	0.120	0.117
RMSE	0.079	0.711	0.689	0.545	0.546

Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Model (1) uses investment/assets in levels; models (2) - (5) use log investment/assets.

A. *Baseline Specification And Functional Form*

Appendix 5 reports three simple pooled regressions of investment intensity on Tobin's q. In column (1), the coefficient on Tobin's q is 0.003 with a robust standard error of 0.001, so higher q is significantly associated with higher investment intensity, but the fit is weak ($R^2 = 0.031$, RMSE = 0.079 in levels). In column (2), adding a quadratic term raises R^2 only to 0.062, so allowing for curvature in levels brings little improvement. The log–log model in Appendix 5, column (3), performs best: the elasticity of investment with respect to q is 0.240, meaning a 1% increase in q is associated with about a 0.24% increase in investment intensity. The R^2 rises to 0.108 (RMSE = 0.711 in logs), a noticeable gain relative to the level specifications. Residual plots for M3 (Appendix 6) show residuals roughly symmetric around zero and without strong systematic patterns, which supports using the log-log form as the baseline.

I also ran diagnostic tests on these simple models. For the preferred log–log specification (M3), the RESET test reports $F(3, 27, 561) = 21.37$ ($p < 0.001$), indicating that some non-linearities or omitted variables likely remain, while the Breusch–Pagan test gives $\chi^2(1) = 3.16$ ($p = 0.075$) and White's test $\chi^2(2) = 246.50$ ($p < 0.001$). For the level specifications M1 and M2, both RESET and Breusch–Pagan/White tests are even more extreme - for example, for M1 the BP statistic is $\chi^2(1) = 27,510.83$ and White's $\chi^2(2) = 1,267.44$ (both $p < 0.001$) - confirming that those models fit the data poorly and exhibit strong heteroskedasticity. Given this evidence and the large cross-sectional variation in the data, I treat heteroskedasticity as likely and therefore report heteroskedasticity-robust standard errors for all pooled regressions and firm-clustered standard errors for panel models in the rest of the paper.

B. Multiple Regression with Controls

Table 2, column (3), reports the pooled log–log specification with controls, oil-crisis dummies, the NYSE/AMEX dummy, and the interaction between ln(vala) and NYSE/AMEX. In that specification, the coefficient on ln(vala) is 0.277 (s.e. 0.011), so for firms not listed on NYSE/AMEX a 10 percent increase in Tobin's q is associated with about a 2.8 percent increase in the investment-to-assets ratio. The same column also reports positive coefficients on oil7374 (0.207) and oil7981 (0.088), implying higher investment intensity during those years conditional on q and the other controls. Model fit improves relative to the simple log–log regression: Table 2 reports $\mathrm{R}^{\mathbf{2}} = 0.162$ and RMSE = 0.689 in column (3), compared with 0.108 and 0.711 in column (2).

The control coefficients in Table 2, column (3), are economically sensible. Cash flow is 0.206 (s.e. 0.035), debt is 0.259 (0.022), R&D intensity is 0.220 (0.058), and firm size is 0.079 (0.003), while advertising intensity is negative at -0.950 (0.079). I therefore continue to report heteroskedasticity-robust standard errors for the pooled regressions. Multicollinearity also appears limited: Appendix 7 reports a mean VIF of 1.83 and a maximum VIF of 3.53 for the pooled specification with the interaction term.

C. Heterogeneous Effects and Interaction

Table 2, column (3), also shows a negative interaction coefficient on ln_vala × NYSE/AMEX of -0.040 (s.e. 0.011). This means that being listed on NYSE/AMEX lowers the elasticity of investment with respect to Tobin's q by about 0.04 relative to non-listed firms. The implied elasticity is 0.277 for non-listed firms and 0.237 for NYSE/AMEX firms, so investment remains positively related to q in both groups but is less sensitive among listed firms.

The other coefficients in Table 2, column (3), continue to look reasonable: cash flow, leverage, R&D intensity, and firm size are positively associated with investment, advertising is negative, and the NYSE/AMEX dummy itself is -0.147 (s.e. 0.011), suggesting lower average investment intensity for listed firms conditional on Tobin's q and the other controls. Appendix 7 again reports modest VIFs for this specification (mean 1.83; max 3.53), so multicollinearity does not appear severe. As in earlier pooled models, I interpret these coefficients using heteroskedasticity-robust standard errors.

D. Panel Data: Fixed and Random Effects

To exploit the panel structure of the data, I next estimate fixed-effects (FE) and random-effects (RE) versions of the baseline log–log model with year dummies. The FE regression uses within-firm variation over time and controls for any time-invariant firm characteristics through firm-specific intercepts. The within R^2 is 0.120, only slightly lower than the pooled multiple regression, indicating that a meaningful share of the variation in log investment intensity is explained by within-firm movements in Tobin's q and the controls. The coefficient on log Tobin's q is 0.280 (robust standard error 0.014, t $\approx$ 20.2), so even after conditioning on firm fixed effects, year dummies, cash flow, leverage, R&D, advertising, and log sales, higher market valuation is strongly associated with higher investment intensity. The coefficients on debt, R&D, and firm size remain positive and statistically

Table 3: Panel Data Regressions: Fixed vs. Random Effects (Dependent Variable: ln_inva)

Variable	**(1) Fixed effects**	**(2) Random effects**
ln_vala	0.280	0.281
	(0.0069)	(0.0063)
cfa	0.051	0.112
	(0.0200)	(0.0187)
debta	0.123	0.149
	(0.0152)	(0.0145)
rnda	0.133	0.171
	(0.0298)	(0.0281)
adva	-0.063	-0.400
	(0.0744)	(0.0633)
ln_sales	0.168	0.090
	(0.0086)	(0.0052)
Constant	-3.044	-2.772
	(0.232)	(0.231)
Year FE	Yes	Yes
Firm FE	Yes	No
Number of firms	1,962	1,962
Observations	27,566	27,566
Within R-squared	0.120	0.116
Overall R-squared	0.107	0.151
Hausman test (FE vs RE)	$\chi^2(37) = 310.02$, p=0.000	—

significant, while the cash-flow effect becomes small and imprecise once firm fixed effects are included. An F-test of the joint significance of the firm effects, F(1,961,25,567)=9.40 with $p < 0.001$, strongly rejects the hypothesis that all firm intercepts are equal, providing clear evidence that unobserved, time-invariant firm heterogeneity matters. The RE model produces a very similar elasticity of investment with respect to Tobin's q (0.281 with standard error 0.006), but the Hausman test comparing FE and RE, $X^2(37)$=310.02 with $p < 0.001$, rejects the null hypothesis that the RE estimator is consistent. This indicates that the unobserved firm effects are correlated with Tobin's q and other regressors, so the RE assumptions are not credible in this setting. I therefore treat the FE specification with firm-clustered standard errors as my preferred panel model and use it as the benchmark for interpreting the causal effect of Tobin's q on corporate investment in the remainder of the paper.

E. Instrumental Variables / Endogeneity Of Tobin's Q

In this subsection I treat log Tobin's q as endogenous and estimate a fixed-effects IV model using the present discounted value of dividends (pstar) and the asset-pricing terms h0 and h1 as instruments. The FE–IV regression (Table 4) is estimated with firm and year fixed effects and firm-clustered standard errors. The IV estimate of the elasticity of investment with respect to Tobin's q is 0.339 (robust s.e. 0.059, z $\approx$ 5.7, p $<$ 0.001, 95% CI $\approx$ [0.223, 0.455]). This is noticeably larger than the corresponding FE OLS estimate (about 0.28), suggesting that ignoring endogeneity - through measurement error or feedback from investment to market value - tends to understate the responsiveness of investment to valuation.

Most control coefficients remain economically sensible. Leverage (debta) and firm size (ln_sales) are still positive and precisely estimated: a one-unit increase in debta is associated with about a 0.13 increase in log investment/assets (s.e. 0.029), and a 1% increase in sales raises investment intensity by roughly 0.18% (coefficient 0.176, s.e. 0.021). By contrast, cash flow, R&D, and advertising lose significance once Tobin's

Table 4: Fixed-Effects IV Regression: Log Investment/Assets

Variable	**Coefficient**	**Robust s.e.**
ln_vala	0.339	(0.059)
cfa	-0.011	(0.077)
debta	0.133	(0.029)
rnda	0.116	(0.063)
adva	-0.011	(0.135)
ln_sales	0.176	(0.021)
Constant	-3.089	(0.195)
Observations	27,566	
Number of Firms	1,962	
Within R-squared	0.117	
Overall R-squared	0.115	
Wald χ^2	4,549.75	

Notes Firm and year fixed effects included. Robust standard errors clustered by firm are reported in parentheses.

q is instrumented, which is consistent with part of their apparent effect in the OLS and FE models being absorbed by the corrected valuation term. The within $\mathbf{R^2}$ of the FE–IV model ($\approx$ 0.117) is very close to the FE OLS value, and the regressors are strongly jointly significant (Wald $X^2(37) \approx$ 4,550, $<$ 0.001), indicating that the IV specification still explains a substantial share of within-firm investment variation while correcting for endogeneity in Tobin's q.

As a check on the IV strategy, I estimated a pooled 2SLS version of the model and ran standard instrument-validity tests (Appendix 8). The first-stage regression shows that the instruments are strongly relevant: the joint F-statistic for pstar and h1 is 187.1 (partial $R^2 \approx 0.034$, p $<$ 0.001). The overidentification test is less comforting: the score $X^2(1) = 4.79$ (p = 0.029) rejects the null that both instruments are perfectly exogenous, suggesting that at least one may be weakly correlated with the investment disturbance. At the same time, the endogeneity tests for ln_vala give $X^2(1) = 1.83$ (p = 0.18) and F(1, 27,527) = 1.72 (p = 0.19), so I cannot reject exogeneity of Tobin's q. I therefore view the FE–IV estimates as a robustness check that yields a somewhat larger elasticity of investment with respect to q, rather than a clearly preferred baseline relative to the FE-OLS specification.

F. Goodness-of-fit and F-tests

Across specifications, the models have moderate explanatory power. In the pooled regressions shown in Table 2, the R^2 rises from 0.108 in the simple log–log model (column (2)) to 0.162 in the pooled specification with controls and interaction (column (3)). In the panel models, Table 2 reports within R^2 values of 0.120 for the fixed-effects model (column (4)) and 0.117 for the FE–IV model (column (5)), indicating that a meaningful share of within-firm variation in investment intensity is explained by Tobin's q and the controls. Table 5 also shows that the regressors are jointly highly significant across specifications. These results imply that the added controls, interaction term, and fixed effects improve fit while leaving the core positive relationship between q and investment intact.

G. Oil-crises Dummies and Robustness

The sample includes two major oil-price shocks (1973–74 and 1979–81). To check whether these episodes drive the results, I re-estimate the pooled log–log models including dummies for each crisis period.

Table 5: Summary of F-Tests

Model	Description	Test	Stat.	d.f.	p-val	Interpretation
M3	Simple log–log pooled OLS	All slopes = 0	F = 3345.6	$(1, 27564)$	$< .001$	$\ln(vala)$ is highly jointly significant.
M4	Pooled OLS with controls	Robust F-test	F = 575.09	$(6, 27559)$	$< .001$	All regressors are jointly significant in the multiple regression.
M5	Pooled OLS + NYSE interaction	Robust F-test	F = 473.82	$(8, 27557)$	$< .001$	Extended model with interaction remains strongly significant.
M6	Firm FE (with year dummies)	Cluster-robust	F = 42.32	$(37, 1961)$	$< .001$	Covariates/time effects are significant after absorbing firm FE.
FE test	Firm fixed effects (non-robust)	$u_i = 0$ for all i	F = 9.40	$(1961, 25567)$	$< .001$	Reject H_0 of no firm heterogeneity; FE is warranted over Pooled OLS.

Adding these dummies to the pooled specification does not materially change the estimated elasticity of investment with respect to Tobin's q. In the pooled specification reported in Table 2, column (3), both oil dummies are positive, indicating higher average investment intensity in 1973–74 and 1979–81 conditional on q and the other controls, but the main ln(vala) coefficient remains positive and precisely estimated.

The same conclusion holds for the interaction model (M5): the q elasticities for NYSE/AMEX and non-listed firms remain similar to the baseline, and the difference between them stays large and significant.

In the fixed-effects and fixed-effects IV models, year fixed effects already absorb these shocks, so adding oil-crisis dummies is redundant and does not change the results. Overall, the findings are robust to controlling for major oil-price shocks and are not driven by a few extreme macroeconomic episodes.

VI. Discussion & Extenstion

The estimates across models tell a consistent story: investment intensity is positively related to Tobin's q, but the magnitude of this link depends on how much heterogeneity and endogeneity I control for. In the simple pooled log–log regression shown in Table 2, column (2), the elasticity of investment with respect to q is 0.240. In the pooled specification with controls and the NYSE/AMEX interaction, reported in Table 2, column (3), the elasticity for non-listed firms is 0.277, while for listed firms it is 0.237 after accounting for the -0.040 interaction term. The R^2 also rises from 0.108 to 0.162 between columns (2) and (3), so Tobin's q remains central even after conditioning on standard firm-level determinants.

Allowing the elasticity to differ by listing status reveals economically meaningful heterogeneity. As Table 2, column (3), shows, the estimated elasticity is 0.277 for firms not listed on NYSE/AMEX and 0.237 for listed firms. This pattern suggests that investment at smaller or less prominent firms is more sensitive to changes in valuation, consistent with the idea that these firms face tighter financial frictions and react more strongly when market value improves.

The panel estimates highlight the importance of unobserved firm characteristics. Once I include firm and year fixed effects (M6), the elasticity rises to about 0.280 (s.e. ≈ 0.014), and the Hausman test strongly rejects random effects in favor of fixed effects ($X^2(37) \approx 310$, $p < 0.001$). This indicates that pooled OLS understates the responsiveness of investment to q because it ignores time-invariant firm traits—such as management quality, technology, or corporate culture—that are correlated with both valuation and investment.

The IV results provide an upper-bound perspective on causality. In the fixed-effects IV model (M7), instrumenting log q with pstar, h0, and h1 yields an elasticity of about 0.339 (s.e. $\approx$ 0.059). In a complementary pooled 2SLS regression with year dummies, the elasticity is about 0.220 (s.e. $\approx$ 0.027), and the first-stage statistics indicate strong instrument relevance: the partial R-squared on ln_vala is about 0.034, and the robust first-stage F is approximately 187.1 ($p < 0.001$). At the same time, the overidentification test rejects the joint exogeneity of the instruments at the 5 percent level ($X^2(1) \approx 4.79$, $p \approx 0.029$), while the endogeneity tests do not reject the null that ln_vala is exogenous (p $\approx$ 0.18–0.19). This pattern suggests that the IV estimates should be interpreted cautiously: they likely bracket the true causal effect from above, while the fixed-effects estimate around 0.28 is a more conservative and credible central value.

Several extensions could deepen the analysis. A first extension would be to allow sector-specific elasticities - interacting ln_vala with industry dummies - to test whether high-tech, R&D-intensive, or regulated sectors react more strongly to valuation shocks than traditional manufacturing or utilities. A second extension would be to explore non-linearities around the theoretical benchmark q = 1, for example by estimating models that allow the elasticity to differ when q is below, near, or well above one, to see whether firms only respond strongly once they are clearly "in the money." A third extension would be to move toward a dynamic panel framework that includes lagged investment and lagged q, capturing adjustment costs and expectations more explicitly and connecting the reduced-form results more closely to formal q-theory. These extensions would all build directly on the existing specification while offering sharper tests of how and when Tobin's q translates into real investment.

VII. Conclusion

In the preferred fixed-effects OLS specification with firm and year dummies and financial controls (Model M6), the elasticity of investment with respect to log Tobin's q is about 0.28 (s.e. $\approx$ 0.014). In the corresponding fixed-effects IV model (Model M7), where log Tobin's q is instrumented by the present discounted value of dividends and the asset-pricing terms h0and h1, the elasticity rises to about 0.34 (s.e. $\approx$ 0.059). With an average Tobin's q of roughly 1.62 in the sample, this IV elasticity of 0.34 implies a semi-elasticity of about 0.21 in a levels specification, i.e. a coefficient of approximately 0.21 on Tobin's q in levels in a regression of $ln(investment/assets)$ on q.

Taken together, these estimates imply that a 1 percent increase in Tobin's q is associated with roughly a 0.25 - 0.30 percent increase in investment intensity, depending on the specification, while a one-unit increase in qaround its mean (for example, from 1.5 to 2.5) corresponds to about a 0.21 increase in log investment/assets. This directly answers the central research question posed in the introduction: Tobin's q does have a causal effect on firms' investment-to-assets ratios, but the magnitude is moderate rather than dominant.

Economically, the results support the core intuition of q-theory: firms that the market values more highly relative to their asset base tend to invest more. At the same time, the modest size of the elasticity and the significant roles of cash flow, leverage, R&D, advertising, and firm size indicate that Tobin's q is only one of several channels through which financial conditions shape real investment. Compared with the existing literature - such as Blundell et al. (1992) and Bond et al. (2003), who typically report smaller coefficients on q in similar investment equations - my preferred fixed-effects and FE–IV estimates suggest a somewhat stronger, but still far from overwhelming, sensitivity of investment to valuation.

The diagnostic and instrument-validity tests point to heteroskedasticity and raise some concerns about instrument exogeneity: first-stage F-statistics indicate that the instruments are strongly relevant, but the overidentification test rejects perfect exogeneity at conventional levels, while endogeneity tests do not reject exogeneity of Tobin's q. For this reason, the FE–IV estimates with log q and their implied levels interpretation are best viewed as upper bounds, with the fixed-effects OLS elasticity around 0.28 serving as a more conservative central estimate.

Overall, the evidence from the Hall and Hall panel data confirms that Tobin's q is a useful summary measure of investment opportunities but does not fully determine corporate investment behaviour. Future work that incorporates sectoral heterogeneity, non-linearities around $q = 1$, and dynamic adjustment (for example, through lagged investment and valuation terms) could further clarify when market-based signals are most informative and how they interact with technological constraints and financial frictions in shaping firms' investment decisions.

References

Andrew B Abel and Olivier J Blanchard. The present value of profits and cyclical movements in investment. *Econometrica*, 54(2):249–273, 1986.

Andrew B Abel and Janice C Eberly. A unified model of investment under uncertainty. *American Economic Review*, 84(5):1369–1384, 1994.

Richard Blundell, Stephen Bond, Michael Devereux, and Fabio Schiantarelli. Investment and Tobin's Q: Evidence from company panel data. *Journal of Econometrics*, 51(1–2):233–257, 1992.

Stephen Bond, Julie Ann Elston, Jacques Mairesse, and Benoit Mulkay. Financial factors and investment in Belgium, France, Germany, and the United Kingdom: A comparison using company panel data. *Review of Economics and Statistics*, 85(1):153–165, 2003.

William C Brainard and James Tobin. Pitfalls in financial model building. *American Economic Review*, 58(2): 99–122, 1968.

Ricardo J Caballero and Eduardo MRA Engel. Explaining investment dynamics in U.S. manufacturing: A generalized (S, s) approach. *Econometrica*, 67(4):783–826, 1999.

Jason G Cummins, Kevin A Hassett, and Stephen D Oliner. Investment behavior, observable expectations, and internal funds. *American Economic Review*, 96(3):796–810, 2006.

Timothy Erickson and Toni M Whited. Measurement error and the relationship between investment and q. *Journal of Political Economy*, 108(5):1027–1057, 2000.

Steven M Fazzari, R Glenn Hubbard, and Bruce C Petersen. Financing constraints and corporate investment. *Brookings Papers on Economic Activity*, 1988(1):141–195, 1988.

Simon Gilchrist and Charles P Himmelberg. Evidence on the role of cash flow for investment. *Journal of Monetary Economics*, 36(3):541–572, 1995.

Charles J Hadlock and Joshua R Pierce. New evidence on measuring financial constraints: Moving beyond the KZ index. *Review of Financial Studies*, 23(5):1909–1940, 2010.

Bronwyn H Hall and Robert E Hall. The value and performance of U.S. corporations. *Brookings Papers on Economic Activity*, 1993(1):1–50, 1993.

Fumio Hayashi. Tobin's marginal Q and average Q: A neoclassical interpretation. *Econometrica*, 50(1).

Steven N Kaplan and Luigi Zingales. Do investment-cash flow sensitivities provide useful measures of financing constraints? *Quarterly Journal of Economics*, 112(1):169–215, 1997.

James H Stock and Mark W Watson. *Introduction to econometrics*. Pearson, 4 edition, 2019.

Lawrence H Summers. Taxation and corporate investment: A q-theory approach. *Brookings Papers on Economic Activity*, 1981(1):67–140, 1981.

James Tobin. A general equilibrium approach to monetary theory. *Journal of Money, Credit and Banking*, 1 (1):15–29, 1969.

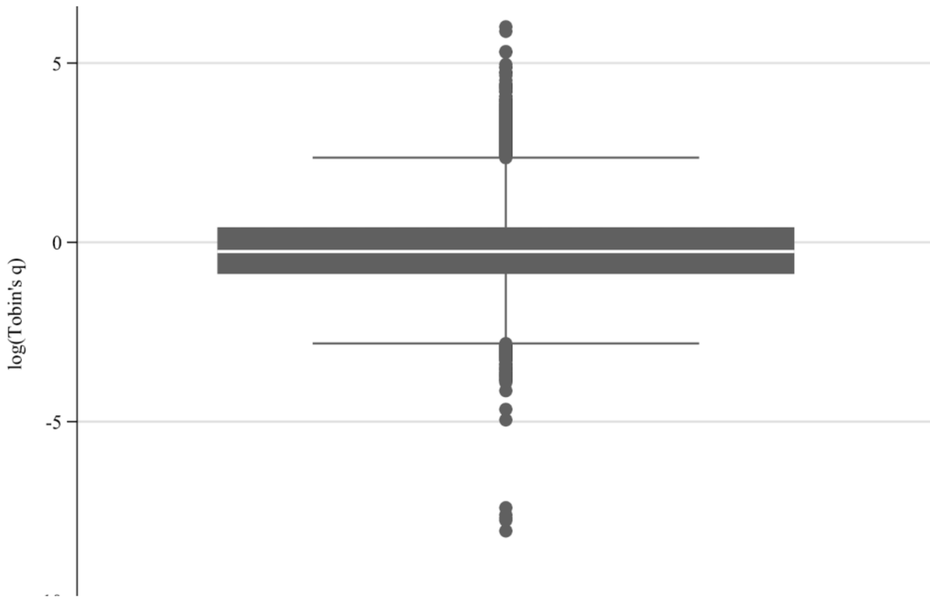
Toni M Whited. Debt, liquidity constraints, and corporate investment: Evidence from panel data. *Journal of Finance*, 47(4):1425–1460, 1992.

VIII. Appendix

Appendix 1: Summary statistics for main variables (estimation sample, N = 27,566)

Variable	**Description**	**N**	**Mean**	**Std. dev.**	**Min**	**Max**
inva	Investment / beginning-of-year assets (investment intensity)	27,566	0.099	0.081	0.000	2.380
vala	Tobin's q: total market value / book value of assets	27,566	1.619	5.348	0.000	407.094
cfa	Cash flow/assets	27,566	0.248	0.331	-5.726	17.785
debta	Long-term debt / assets (leverage)	27,566	0.276	0.300	-0.009	9.545
rnda	R&D / assets (R&D intensity)	27,566	0.045	0.163	0.000	12.853
adva	Advertising/assets	27,566	0.027	0.105	0.000	6.496
sales	Annual sales (million USD)	27,566	729.995	2,729.344	0.001	71,643.380
nyseamex	1 if firm listed on NYSE or AMEX, 0 otherwise	27,566	0.697	0.459	0.000	1.000
ln_inva	Log of investment / assets (for positive inva only)	27,566	-2.578	0.753	-8.047	0.867
ln_vala	Log of Tobin's q (for positive vala only)	27,566	-0.195	1.030	-8.047	6.009
ln_sales	Log of annual sales (million USD)	27,566	4.897	1.770	-6.908	11.179

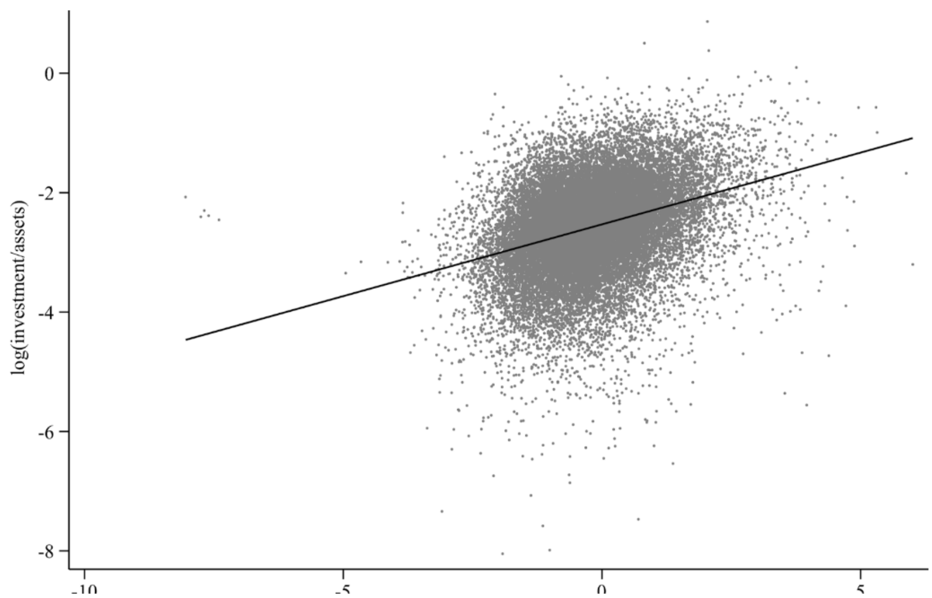
Appendix 2: Box Plot of Log Tobin's q



Appendix 3: Correlation matrix (Pearson correlations; main variables and instruments, estimation sample, N = 27,566)

Variable	**ln_inva**	**ln_vala**	**cfa**	**debta**	**rnda**	**adva**	**ln_sales**	**nyseamex**	**pstar**	**h0**	**h1**
ln_inva	1.000	0.329	0.169	0.042	0.143	-0.002	0.070	-0.091	0.055	-0.029	0.028
ln_vala	0.329	1.000	0.431	-0.122	0.322	0.170	-0.183	-0.172	0.050	-0.053	0.055
cfa	0.169	0.431	1.000	-0.093	0.096	0.603	0.012	-0.120	0.066	-0.044	-0.003
debta	0.042	-0.122	-0.093	1.000	0.015	0.034	-0.023	-0.005	-0.064	-0.069	-0.070
rnda	0.143	0.322	0.096	0.015	1.000	0.067	-0.177	-0.159	-0.018	-0.135	-0.091
adva	-0.002	0.170	0.603	0.034	0.067	1.000	0.019	-0.050	0.036	-0.092	-0.072
ln_sales	0.070	-0.183	0.012	-0.023	-0.177	0.019	1.000	0.373	0.410	-0.081	0.030
nyseamex	-0.091	-0.172	-0.120	-0.005	-0.159	-0.050	0.373	1.000	0.132	0.285	0.249
pstar	0.055	0.050	0.066	-0.064	-0.018	0.036	0.410	0.132	1.000	-0.159	-0.217
h0	-0.029	-0.053	-0.044	-0.069	-0.135	-0.092	-0.081	0.285	-0.159	1.000	0.786
h1	0.028	0.055	-0.003	-0.070	-0.091	-0.072	0.030	0.249	-0.217	0.786	1.000

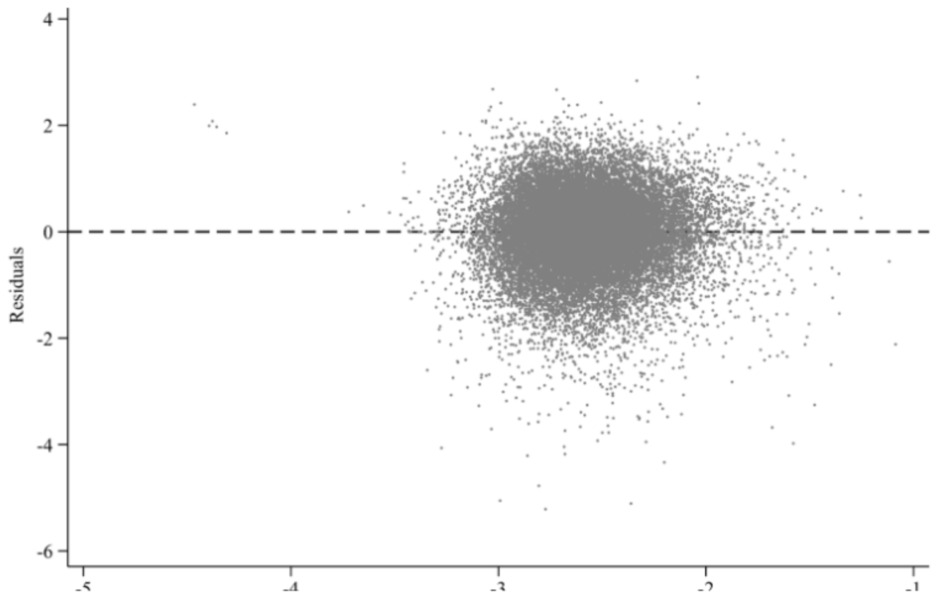
Appendix 4: Scatterplot of log investment/assets and log Tobin's q

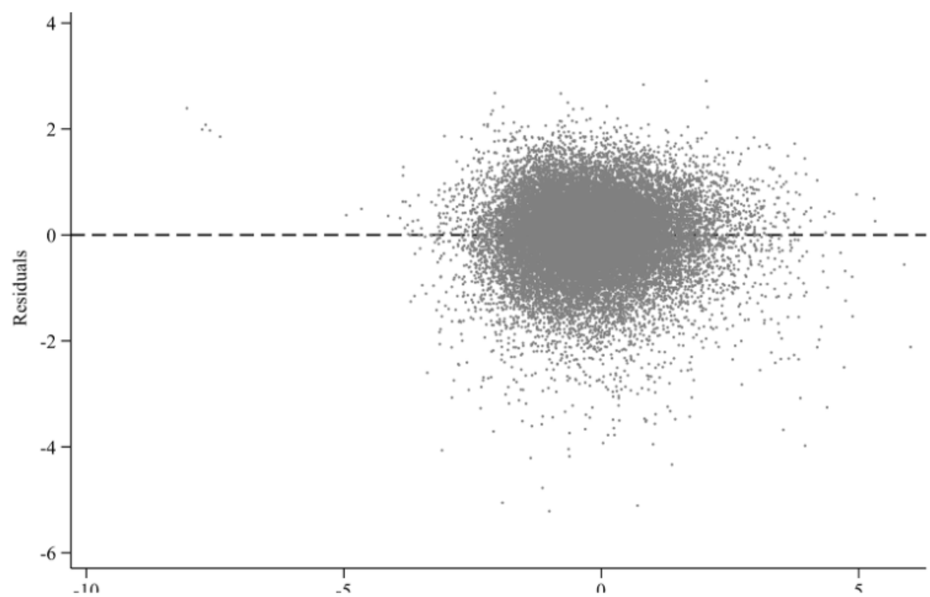


Appendix 5: Simple pooled regressions of investment intensity on Tobin's q in three alternative functional forms: (1) Linear, (2) Quadratic, and (3) Log–log

Variable	**(M1) Linear**	**(M2) Quadratic**	**(M3) Log–log**
vala	0.003	0.006	—
	(0.001)	(0.000)	
vala2	—	-0.000	—
		(0.000)	
ln_vala	—	—	0.240
			(0.005)
Constant	0.094	0.089	-2.531
	(0.001)	(0.001)	(0.004)
N	27,566	27,566	27,566
R-squared	0.031	0.062	0.108
RMSE	0.079	0.078	0.711

Appendix 6: Graphs of Residuals and Predicted Values





Appendix 7. Instrument validity tests for Tobin's q (pooled IV model)

Model	Specification (short label)	Variables used for VIF	Max VIF	Mean VIF
(1) M1	Linear pooled OLS (inva on vala)	vala	1.00	1.00
(2) M2	Quadratic pooled OLS (inva on vala, vala^2)	vala, vala2	2.57	2.57
(3) M3	Log–log pooled OLS (ln_inva on ln_vala)	ln_vala	1.00	1.00
(4) M4	Pooled log–log + controls + oil dummies	ln_vala, cfa, debta, rnda, adva, ln_sales	1.94	1.37
(5) M5	Pooled log–log + controls + NYSE/AMEX interaction + oil dummies	ln_vala, ln_vala_nyse, cfa, debta, rnda, adva, ln_sales, nyseamex	3.53	1.83
(6) M6	FE model (firm, year FE)	ln_vala, cfa, debta, rnda, adva, ln_sales, year dummies	—	—
(7) M7	IV model – first-stage for ln_vala	pstar, h0, h1, cfa, debta, rnda, adva, ln_sales, year dummies	—	—

Notes: Variance inflation factors (VIFs) are computed for pooled OLS specifications (M1–M5). In specifications (M6) and (M7), VIFs are not reported as the within-transformation and instrument structure create mechanical collinearity with limited within-firm variation. For the IV model (M7), instrument relevance is evaluated via the first-stage F-statistic in Appendix 8[cite: 293].

Appendix 8. Instrument validity tests for Tobin's q (pooled IV model)

Test	Name	Null hypothesis	Statistic	p-value	Interpretation
First-stage relevance	Instrument relevance for ln_vala	Instruments (pstar, h1) jointly irrelevant in first stage	F(2, 27,527) = 187.08	0.000	Strong evidence that instruments predict ln_vala; no weak-instrument problem.
Overidentification test	Score χ^2 J-test	Instruments jointly exogenous	$\chi^2(1) = 4.79$	0.0286	Null of perfect exogeneity rejected at 5%; at least one instrument may be mildly invalid.
Endogeneity of ln_vala	Score χ^2 test	Tobin's q is exogenous	$\chi^2(1) = 1.83$	0.176	Cannot reject exogeneity of ln_vala.
Endogeneity of ln_vala	Regression F-test	Tobin's q is exogenous	F(1, 27,527) = 1.72	0.189	Cannot reject exogeneity; IV not statistically required.

Spring 2026 Issue V Volume II