Corporate Investment Under Uncertainty: Finance, Corporate Governance, and Market Signals

Yinxue Lu
Stony Brook University
M.A. Thesis in Economics

Abstract

This study examines how financial uncertainty, shareholder–manager disagreement, and macroeconomic shocks influence corporate investment among Global S&P firms from 2003 to 2022. I start with pooled OLS and fixed-effects models that explain 22% of investment variation, then address endogeneity using GMM, which reveals that contemporaneous disagreement reduces capital expenditure while lagged investment and Tobin's Q exhibit strong persistence. A panel VAR traces how shocks to investment, dissent, and valuation interact over time, and Random Forest and LightGBM models uncover nonlinear thresholds and interactions. These machine-learning ensembles boost out-of-sample R^2 substantially, showing that investment growth stalls once Tobin's Q exceeds three, cuts sharply when disagreement exceeds five percentage points, and intensifies under high leverage or crisis conditions. Overall, shareholder disagreement emerges as a distinct governance channel that deepens investment downturns in uncertain environments, highlighting the need for policies that improve shareholder–board alignment and stabilize capital spending during turbulent periods.

Contents

1 Introduction

In autumn 2008, corporate investment plunged more than 20% (Whited 1992; Thakor and Whited 2011), and in 2020 it fell another 15% due to the global financial crisis and the COVID-19 pandemic (Love and Zicchino 2006). These dramatic drops show how uncertainty drives firms to delay or cancel profitable projects (Blundell and Bond 1998). This thesis focuses on financial uncertainty factors including macroeconomic shocks, financial frictions, and shareholder–manager voting disagreement to examine how they interact to shape firms' capital investments.

Economic theory tells us that firms invest when expected returns exceed financing costs (Whited 1992), but rising volatility increases the option value of waiting on irreversible projects (Blundell and Bond 1998). Tighter financial constraints force firms to rely more on internal funds, since external debt and equity become costlier (Whited 1992). Governance conflicts when measured by shareholder–manager disagreement, create reputational and monitoring pressures that further curb investment (Thakor and Whited 2011). Finally, stock mispricing influences capital expenditures through valuation and timing channels, as captured in panel-VAR evidence (Love and Zicchino 2006).

My research makes three main contributions. First, I gathered a novel firm–year panel that links Compustat fundamentals with three voting-disagreement measures and industry-level price indices to compute replacement-cost capital stocks over 2003–2022. Second, I apply a unified, multi-method framework including OLS, fixed effects, system GMM, panel VAR, and machine learning to tackle endogeneity, dynamic persistence, and nonlinear thresholds in investment. Third, I show that shareholder–manager voting disagreement significantly amplifies investment downturns during crisis periods, and I draw clear governance and policy recommendations to stabilize capital spending under uncertainty.

2 Literature Review

2.1 Thakor and Whited(2011): Shareholder–Manager Disagreement and Corporate Investment

Thakor and Whited (2011) develop a theoretical model showing how unresolved disagreement between managers and investors can impede corporate investment by raising the perceived cost of projects. They measure disagreement as the gap between management's earnings forecast and the mean analyst estimate within 30 days of the forecast, carefully filtering out news-driven surprises. To address measurement error, they apply an errors-in-variables estimator alongside controls for Tobin's Q and firm fundamentals. Their empirical tests reveal a robust negative relationship between forecast disagreement and both investment and valuation, and they demonstrate their findings with signal-extraction GMM and shown that this effect is unrelated to traditional asymmetric-information proxies. Moreover, the dampening impact of disagreement intensifies for firms with greater financial flexibility, consistent with their model's predictions.

My research extends their framework in three parts. First, I construct vote split proxies for disagreement, aim to capture corporational conflicts over investment policy, compensation, and director elections, and merge these with Compustat fundamentals for Global S&P firms from 2003 through 2022. Second, I move beyond OLS by employing two-step system GMM to tackle endogeneity and by estimating a panel VAR to trace how shocks to disagreement, Tobin's Q, and investment unfold dynamically. Finally, I introduce Random Forest and LightGBM regressors to uncover nonlinear thresholds—showing, for example, that investment plateaus when Tobin's Q exceeds three and plunges when disagreement surpasses five percent—and to boost out-of-sample R^2 from roughly 0.22 under OLS to nearly 0.60. By explicitly modelling the 2008 and 2020 crisis regimes and interpreting machine-learning results via partial-dependence plots and SHAP values, this work reveals regime-dependent governance effects and provides targeted policy recommendations for stabilizing capital spending under uncertainty.

2.2 Whited(1992): Debt, Liquidity Constraints, and Corporate Investment

In this study, Whited (1992) embeds a binding borrowing limit into the canonical investment Euler equation to quantify how asymmetric information in debt markets distorts firms' capital decisions. She shows that when firms hit a borrowing constraint, theyy effectively raising the shadow discount rate on marginal investment. She proxies the multiplier with debt-to-assets and interest-coverage ratios, then estimates the augmented Euler equation with GMM using lagged instruments, firm and year fixed effects—and demonstrates that the unaugmented model holds only for financially unconstrained firms, whereas the liquidity-augmented specification fits significantly better for credit-stressed groups. Whited's key insight is that liquidity frictions generate a measurable "wedge" in the Euler equation that materially improves the explanatory power of structural investment models.

I build my research directly on Whited's framework by first adopting her approach in dynamic GMM instrumentation to address endogeneity in investment equations. Beyond that, I introduce three novel disagreement proxies with measures covering investment policy, compensation, and director elections—merged with contemporary financial data from 2003–2022. I complement GMM with panel VAR analysis to trace how shocks to disagreement, valuation, and investment propagate over time, and I harness Random Forest and LightGBM to capture nonlinear thresholds that no linear Euler augmentation can detect. By explicitly modeling the 2008 and 2020 shocks and reporting impulse-response functions alongside SHAP-based machine-learning interpretations, my thesis both validates Whited's core finding that liquidity constraints distort investment timing and extends it into a dynamic, nonlinear governance context with actionable policy implications.

Arellano and Bond(1991)Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations

In their influential 1991 paper, Manuel Arellano and Stephen Bond develop a way to estimate dynamic panel models when unobserved firm effects and lagged dependent variables cause bias. The approach begins by first-differencing the equation—removing any time-invariant firm fixed effect—and then using deeper lags of the endogenous variables as instruments. Concretely, I first-difference the investment-to-capital ratio and the shareholder–manager disagreement measure to eliminate fixed effects. To address the resulting endogeneity of $\Delta\text{Invest}_{i,t-1}$ and $\Delta\text{Disagree}_{i,t-1}$, I instrument each with its own values from $t-2$ and earlier. Standard errors are clustered at the firm level, and I apply the AR(2) test in differences to ensure no residual serial correlation. Because a twenty-year panel can generate thousands of instruments, I collapse the instrument matrix to preserve power. In a robustness check, I also extract the first two principal components from the four deepest lags of investment and disagreement, confirming that even under aggressive pruning the lag-1 investment coefficient remains near 0.50.

Blundell and Bond (1998) Initial conditions and moment restrictions in dynamic panel data models

Building on the difference-GMM, they introduce the System GMM estimator, which combines the first-differenced equations with the original levels equations to improve efficiency when the autoregressive parameter is large. In practice, I estimate a "system" comprising

$$\text{Invest}_{it} = \alpha_i \; + \; \rho\,\text{Invest}_{i,t-1} \; + \; \varepsilon_{it} \quad \text{(levels)}$$

alongside its first-difference counterpart. The levels equations are instrumented by $\Delta\text{Invest}_{i,t-1}$, and similarly for disagreement. This joint set of moment conditions delivers more precise estimates in finite samples, particularly when capital adjustment is highly persistent. I implement both one-step and two-step versions of System GMM, checking the Hansen overidentification p-value and the AR(2) statistic in each case. Compared to pure difference-GMM, System GMM yields a tighter estimate of the contemporaneous disagreement effect, which then feeds into my Panel VAR analysis to trace how crisis and disagreement shocks propagate through investment, valuation, and governance over time.

2.3 Love and Zicchino (2006): *Financial Development and Dynamic Investment Behavior*

In **(author?)** (Love and Zicchino 2006), Inessa Love and Lea Zicchino study how a firm's ability to invest depends not only on profitability but also on its access to internal funds, especially when external finance is costly or scarce. They assemble a panel of publicly traded firms from 36 countries over 1988–1998 and estimate a Vector Autoregression (VAR) for each country group. By orthogonalizing the shocks, they cleanly separate "fundamental" drivers (such as sales-to-capital or Tobin's q) from "financial" drivers (cash flow) and trace each shock's impact on investment over time. Their key finding is that in countries with underdeveloped financial systems, investment responds much more strongly to cash-flow shocks—evidence that financing constraints are more severe when markets are shallow.

I borrow three main elements from Love and Zicchino's approach. First, I model investment, disagreement, and valuation jointly in a dynamic Panel VAR, using orthogonalized impulse-response functions(IRF) to track how shocks to one variable propagate through the system. Second, I split my sample by industry so I can compare impulse responses under different financing environments. Third, I include macroeconomic dummies (for the 2007–09 financial crisis and the 2020–21 COVID shock), mirroring their use of country-level controls to isolate cross-country heterogeneity in financial development.

my work departs from Love and Zicchino in three respects. I apply modern GMM-based estimation (difference and system GMM) to address endogeneity and fixed effects, whereas they rely on reduced form VAR with Helmert transformations. I integrate their Panel VAR framework with machine-learning insights using Random Forest and LightGBM to uncover nonlinear threshold effects in shareholder disagreement—and I compare these ML-derived thresholds to my VAR impulse responses. Finally, by conducting industry-level analyses and counterfactual simulations under crisis conditions, I extend their cross-country evidence to a firm-level, sectoral examination of how governance frictions and financial constraints jointly shape investment dynamics.

2.4 Breiman (2001): Random Forests

Leo Breiman presents Random Forests as an ensemble of decision trees, each grown on a bootstrap sample and splitting on a random subset of features at every node. He demonstrates that as the number of trees increases, the ensemble's generalization error converges to a limit governed by two quantities: the average strength of individual trees and the correlation among their errors. To assess performance without a separate validation set, Breiman introduces the out-of-bag (OOB) error estimate, and he proposes a permutation-based importance measure that quantifies each feature's contribution by observing the drop in OOB accuracy when that feature is randomly shuffled. By randomly restricting the candidate features at each split, Random Forests both increase tree strength and reduce inter-tree correlation, yielding models that resist overfitting in noisy environments and often rival boosting methods. In regression settings, Breiman further shows that the ensemble mean-squared error is bounded by a term proportional to the average tree error and their pairwise correlations, highlighting the dual roles of these two factors.

In my study of firm-year investment panels, I adapt Breiman's framework while extending it in several key ways. I construct a scikit-learn pipeline (Pedregosa et al. 2011) that imputes missing values, standardizes inputs, and fits the Random Forest model. Hyperparameters—number of trees, maximum depth, minimum samples per leaf, and feature-subset size—are tuned via randomized search combined with five-fold TimeSeriesSplit, selecting on OOB R^2. Unlike Breiman's cross-sectional benchmarks, I train on data through 2020 and evaluate out-of-sample performance on 2021–2022. To deepen interpretability, I supplement Breiman's OOB diagnostics with SHAP values (**?**) and partial dependence plots (Friedman 2001), which reveal each predictor's marginal effect on investment forecasts. Finally, I perform counterfactual simulations—adjusting shareholder disagreement during crisis years—to validate that my Random Forest captures economically meaningful nonlinear thresholds. By integrating these extensions with my dynamic panel GMM and VAR analyses, I combine Breiman's robust ensemble methodology with causal and temporal modeling.

2.5 Ke *et al.* (2017): LightGBM: A Highly Efficient Gradient Boosting Decision Tree

They present LightGBM, a new software library that makes Gradient Boosting Decision Trees (GBDT) much faster and more scalable. Traditional GBDT tools must look through every data point for each feature when deciding where to split a tree, and they struggle when the dataset has thousands or millions of mostly empty variables. LightGBM tackles these two issues without losing accuracy by introducing two features: Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB).

GOSS speeds up training by noting that not all data points contribute equally to model updates. When GBDT builds a tree, it measures how wrong each prediction is using gradients: large gradients mean the model is far off, small gradients mean it is almost right. Instead of using every single point to compute the best split, GOSS keeps all points with large gradients (where the model still has much to learn) and randomly samples a small fraction of the low-gradient points. By reweighting the sampled data appropriately, GOSS produces nearly the same split decisions as using the full dataset, but with a fraction of the work. EFB tackles the challenge of high-dimensional sparse data—for example, one-hot encoded categories where only one feature is nonzero at a time. Many of these features never overlap, so LightGBM groups—or "bundles"—them into combined variables. By treating feature grouping as a graph problem, the authors pack mutually exclusive features together, reducing the total number of variables the algorithm needs to process. This dramatically cuts the cost of building histograms for split selection, again without sacrificing predictive power.

In my research, I leverage LightGBM's built-in GOSS and EFB but adapt them for corporate investment panels in three ways. First, I replace random cross-validation with a five-fold TimeSeriesSplit, ensuring that each firm's time ordering is preserved when tuning hyperparameters. Second, I enrich LightGBM's standard accuracy measures with SHAP values (**?**) and partial dependence plots, so I can interpret how changes in shareholder disagreement or Tobin's q affect investment predictions. Finally, I integrate the tree-ensemble outputs into my dynamic panel framework—comparing the nonlinear thresholds uncovered by LightGBM with causal estimates from GMM and impulse responses from a Panel VAR—to bring together machine learning flexibility, causal identification, and dynamic analysis.

3 Hypotheses and Intuition

In this research, I outline below on each of the hypothesis including key intuitions that helped guide my economic analysis.

Hypothesis 1: Shareholder disagreement alone does not significantly affect investment. I expect that once I control for core firm fundamentals like profitability and valuation, variation in voting splits by itself will not move investment decisions.

Hypothesis 2: Interaction between shareholder disagreement and financial crises negatively impacts investment. I argue that when funds tighten and uncertainty spikes, the friction from board-level dissent further dampens capital spending.

Hypothesis 3: Shareholder disagreement negatively affects investment in dynamic settings. Firms adapt investment plans over time, so I expect past shocks to have persistent drag in a dynamic panel.

Hypothesis 4: Disagreement shocks lead to prolonged investment declines. I anticipate that a spike in conflict not only reduces investment immediately but also casts a multi-period shadow on capital spending.

Hypothesis 5: Financial crises lead to significant investment declines. Broad macro shocks should overwhelm individual firm characteristics.

Hypothesis 6: Lack of amplification from shareholder disagreement or leverage during crises. I test whether firms with high dissent or high debt suffer extra cuts in crises.

Hypothesis 7: Industry heterogeneity in investment persistence. Capital intensive sectors should smooth investment differently. I predict that manufacturing and construction exhibit stronger positive persistence, while cyclical or resource-extraction industries show mean reversion.

Hypothesis 8: Cross-industry variation in persistence reflects sectoral adjustment costs. I argue that heterogeneity in capital adjustment friction drives these patterns.

Hypothesis 9: Nonlinear relationship between shareholder disagreement and investment. When dissent stays low, firms absorb it easily, but above a threshold uncertainty spikes and firms defer spending.

Hypothesis 10: Machine learning models outperform traditional econometric models in predicting investment. Trees capture nonlinearities and interactions that OLS misses, so I expect higher out-of-sample R^2.

Hypothesis 11: My assumption is that Random Forest slightly outperforms LightGBM in predictive accuracy.

4 Data and Summary Statistics

4.1 Data Sources

I assemble a balanced panel of US non-financial and non-utility firms from Compustat spanning 2003 through 2022. I merge each firm's yearly record with three measures of shareholder voting disagreement from annual meeting proposals on investment policy, executive compensation and board elections (Thakor and Whited 2011). To convert nominal flows into real replacement cost terms, I incorporate annual CPI data from the Bureau of Labor Statistics and the BEA producer price index for capital equipment (Whited 1992). I introduce dummy variables for the 2007 to 2009 financial crisis and the 2020 to 2021 COVID 19 shock to capture severe macroeconomic disruptions (Love and Zicchino 2006). I follow Thakor and Whited (2011) by restricting the sample to firms with at least three consecutive years of complete data, dropping observations with zero or negative assets, capital stock or sales, and excluding firms in regulated industries (Thakor and Whited 2011).

4.2 Variable Definitions and Calculation Formulas

Investment to Capital I measure investment intensity as the ratio of capital expenditures to the replacement-cost of capital stock. The replacement value of Capital stock is measured by the net property, plant, and equipment valued at historical replacement cost.

$$\text{InvestmentRate}_{it} = \frac{\text{CAPX}_{it}}{\text{ppent}_{i,t-1}}$$

Tobin's Q Idefine Tobin's Q as the ratio of the firm's total market value of equity plus total market value of debt minus the replacement value of inventories, to its lagged replacement-cost capital stock:

$$Q_{it} = \frac{\big(\text{MarketEquity}_{it} + \text{MarketDebt}_{it}\big) - \text{INVT}_{it}}{\text{ppent}_{i,t-1}}. \tag{1}$$

Shareholder Disagreement Proxies I construct three annual, firm-level measures of voting-based disagreement. For each proposal j in year t, I calculate the absolute vote split as

$$d_{ijt} = \left| \frac{\text{ForVotes}_{ijt} - \text{AgainstVotes}_{ijt}}{\text{TotalVotes}_{ijt}} \right|,$$

which equals 0 when all votes align and 1 at maximal dissent. I then average d_{ijt} across the N_i ballots to obtain the firm-year disagreement score. To capture distinct aspects of governance conflict, I compute this measure over three nested sets of proposals: first, DisagreementInvF using only investment and finance policy votes; next, DisagreementInvFComp by adding executive compensation ballots; and finally, DisagreementInvFCompDir by further including director-election items.

This metric equals zero when all shares vote identically and rises to one at maximum dissent. To form a firm–year proxy average across the relevant set of proposals, I construct three parallel versions by varying the proposal universally. I first restricts DisagreementInvF to investment and finance policy votes only. I then adds executive compensation ballots in DisagreementInvFComp, and lastly, I addDisagreementInvFCompDir, further includes director-election items. I then run ordinary-least-squares regressions on each variant, select the proxy that proves most significant, and use that measure as the main disagreement variable in subsequent analyses.

Financial Controls I include five firm-level controls to capture core dimensions of liquidity, leverage, financing activity, profitability, and growth. Each enters the regressions scaled to make coefficients comparable and to reflect intuitively meaningful rates.

$$\text{CashFlow}_{it} = \frac{\text{NI}_{it} + \text{DP}_{it}}{K}, \tag{2}$$

$$\text{DebtToAsset}_{it} = \frac{\text{MD}_{it}}{\text{MD}_{it} + \text{ME}_{it}}, \tag{3}$$

$$\text{EquityIssuance}_{it} = \frac{\text{SSTK}_{it}}{\text{AT}_{it}} \times 100, \tag{4}$$

$$\text{ROA}_{it} = \frac{\text{IB}_{it}}{\text{AT}_{it}}, \tag{5}$$

$$\text{SalesGrowth}_{it} = \frac{\text{Sale}_{it} - \text{Sale}_{i,t-1}}{\text{Sale}_{i,t-1}}. \tag{6}$$

Cash flow measures the firm's internal liquidity as net income plus depreciation scaled by capital stock K (Whited 1992). This ratio proxies for the cash the firm generates from its operations relative to its installed production capacity and enters regression as a key determinant of investment financed with internal cash.

The debt to asset ratio captures leverage as market debt (the fair value cost of long and short term obligations) divided by total market assets (market debt plus market equity) (Whited 1992). Using market values reflects how funding constraints from debt evolve over time.

Equity issuance records the firm's net equity issuance in each year (Thakor and Whited 2011). It isolates the extent firms raise external equity and signals capital raising flexibility or dilution pressure that affects managers' investment choices.

Return on assets equals income before extraordinary items divided by book assets (Whited 1992). This profitability measure shows how efficiently the firm converts assets into operating income and controls for performance differences that might confound disagreement or shock effects.

Sales growth computes the year over year change in net sales as the difference between current and lagged sales divided by lagged sales (Love and Zicchino 2006). Sales growth leads capital expenditures and serves as a demand indicator.

I winsorize all five series at the first and ninety ninth percentiles to guard against extreme outliers and ensure stable coefficient estimates across regressions (Blundell and Bond 1998).

Macroeconomic Controls I include broad macro shocks with year fixed effects and isolate crisis periods with dummies for 2008–2009 and 2020–2021:

$$\text{CrisisDummy}_{it} = \mathbf{1}\{t \in \{2008, 2009\}\}, \qquad \text{CovidDummy}_{it} = \mathbf{1}\{t \in \{2020, 2021\}\}. \tag{7}$$

This approach follows Love and Zicchino (2006) in using discrete indicators to capture extreme downturns in a panel-VAR framework (Love and Zicchino 2006).

This construction yields 12,393 observations across 1,223 firms from 2003 to 2022, ensuring that both linear estimators and nonlinear machine-learning methods operate on a consistent, well-behaved dataset (Blundell and Bond 1998; Breiman 2001; Ke et al. 2017).

5 Data Cleaning and Sample Construction

At the start of my data processing stage, I performed thorough cleaning of the firm-level data to ensure consistency and robustness. I converted key numeric columns—capital expenditures, property and equipment, book equity, sales, stock price, and income components—into proper numeric formats. This step standardized the dataset's structure and eliminated potential data-entry anomalies or formatting inconsistencies(Whited 1992).

To focus the analysis on firms with comparable investment behavior and regulatory environments, I followed (**?**) method and exclude firms operating in the utilities and financial sectors. These sectors, which fall under SIC codes 4900 to 4999 and 6000 to 6999 respectively, are subject to distinct financial regulations and often follow different capital allocation dynamics, making them incompatible with the core investment models used in my study (whited 2011. I kept only firms with total assets greater than two million dollars, and with both gross and net property, plant, and equipment values above one million. I ensured that each firm reported positive sales, as this serves as a basic requirement for modeling firm-level performance and investment decisions (**?**). This filtering process reduced the dataset from its raw size but significantly improved its quality and structure. It ensured that my subsequent GMM, panel VAR, and Machine Learning estimations would be based on firms with adequate data history and financial activity, leading to more reliable and interpretable results.

In the final stage of data preparation, I implemented an outlier cleaning process to improve regression stability, mitigate skewness, and reduce noises on model results. This additional step away from Whited was essential because financial variables are often affected by outliers from accounting anomalies, one-time shocks, or reporting errors, which can distort estimation and inference if left unaddressed.

To begin, I define upper and lower bounds for key variables based on both theoretical constraints and empirical distribution ranges. For example, I capped Tobin's Q between 0.2 and 7, while investment-to-capital ratios were constrained to remain below 35 %, as values above this range are typically unrealistic for sustained capital formation. I applied similar tight bounds to variables like cash flow, sales growth, ROA, and equity issuance to restrict the influence of extreme data points.

After the initial pass, I applied the Yeo-Johnson transformation to treat a group of skewed variables including equity issuance, cash flow, sales growth, and ROA. This transformation adjusts distributions without requiring strictly positive values(**?**), making it more suitable than standard log transformations. I tested whether each transformation significantly reduced skewness and kurtosis. If both metrics improved substantially by at least 30%, I retain the transformed version. Otherwise, the original variable remained unchanged. This decision rule ensured that I only applied nonlinear transformations when they clearly enhanced distributional balance without introducing distortions.

I then conducted a second round of outlier removal using modified z-scores. For each variable, I calculated the deviation from the median scaled by the median absolute deviation(Iglewicz and Hoagl which is more robust to non-normality than standard deviation-based metrics. For investment ratios and ROA, I used a stricter z-score threshold of 2.0, while for other variables the threshold was set at 2.5. Observations falling outside these thresholds were temporarily flagged as outliers and replaced with the firm-specific median for that variable. This approach helped preserve within-firm variation while avoiding bias introduced by imputing from the full sample.

After cleaning, I tracked changes in mean, standard deviation, skewness, and kurtosis for each variable, confirming that all metrics improved substantially. For example, ROA and investment ratios became more symmetrically distributed with lower tails, which stabilized their behavior in subsequent regression models.

The above processes not only improved the fit of the model and reduced residual variance but also strengthened the reliability of the machine learning and panel regression estimates. By enforcing data integrity and controlling for extreme distortions, the cleaned data became a strong foundation for all of the empirical analysis in this project.

5.1 Descriptive Statistics

I begin the process by importing the pre-processed panel data across 1223 firms from 2003 to 2022 and sorting it by fiscal year and firm identifier. I winsorize each series to mitigate the influence of extreme outliers. I apply a log transformation to strictly positive variables in order to reduce skewness and improve numerical stability. I then generate one year lags for each variable within firm groups, which eliminates contemporaneous feedback and aligns with the dynamic specification in (Thakor and Whited 2011)). To measure financial uncertainty, I merged the three disagreement proxies with the compustat dataset and replace zero values in each disagreement series with its sample median, ensuring that the logged versions remain well defined.

To normalize investment inputs and make sure all ratios could be meaningfully compared across firms, time, and industry conditions. I began by merging four key macroeconomic indices into the main panel dataset: the Producer Price Index by two-digit SIC code, the Consumer Price Index, a capital goods price index from BEA, and the Baa corporate bond yield from FRED(Love and Zicchino 2006). I used the first two digits of each firm's SIC code to align the industry-specific PPI with firm-year observations. After this merge, I sorted the dataset by firm and year to prepare it for time-series calculations.

Next, I adjusted net capital expenditures and book equity into real terms using the PPI and set the base year to the lowest available index value(Love and Zicchino 2006). I then used forward-filled inventory levels to maintain a clean investment stock variable and computed each firm's replacement-value capital stock using the perpetual inventory method. To do this, I applied a custom function that accumulates capital stock over time by accounting for changes in capital expenditure and relative changes in the capital goods price index(Whited 1992). This process ensured that capital dynamics reflected both firm-level activity and macroeconomic conditions.

Market debt was calculated by constructing a debt maturity structure with evenly spaced amortization over 20 years. For each firm, I assumed long-term debt rolled forward each year and amortized accordingly, adjusting for interest payments using a 6 % benchmark rate. I then calculated market equity as the sum of market capitalization and a discounted value of dividend payouts using the Baa yield, ensuring consistency with real-world market valuation frameworks(Whited 1992).

To solve the extreme outliers issue commonly exist in the financial data, I applied a modified z-score method to both market debt and equity. I computed the median and median absolute deviation for each firm, then excluded observations with z-scores exceeding a threshold of three. Missing values were replaced with firm-level medians to maintain the time-series structure without introducing artificial breaks.

With these cleaned variables, I lagged the replacement capital stock by one year and constructed a denominator variable for use in Tobin's Q and investment ratio calculations(Whited 1992). For equity issuance, I scale newly issued stock by total assets and express it in percentage terms (Whited 1992). I create size quintiles within each year to account for systematic variation by firm size and winsorize the issuance variable at the ninety-fifth percentile within each quintile to limit the influence of highly dilutive firms (Blundell and Bond 1998). I also generate an industry-adjusted measure by subtracting the year-industry average issuance from each firm's value (Thakor and Whited 2011).

This pre-processing pipeline allowed me to work with a clean, consistent panel of investment determinants and responses, ready for economic and machine learning analysis.

Table 1 presents the initial key moments of these transformed variables. The investment-to-capital ratio averages 0.145 with a standard deviation of 0.071, while shareholder disagreement measures exhibit an average dispersion of 0.116 and can spike to as high as 0.912. Tobin's Q centers on 1.769 and ranges from 0.200 to 4.999. Cash-flow, measured as operating cash-flow scaled by capital, has a mean of 0.396 and extends into negative territory for firms experiencing losses. The summary statistics confirm the wide cross-sectional heterogeneity that motivates my nonlinear modeling approach.

Table 1: Summary Statistics, 2003–2022

Variable	N	Mean	Std. Dev.	Min	25%	Median	75%
invest_cap	12 393	0.1451	0.0706	0.0000	0.0916	0.1386	0.1950
Disagreement	12 393	0.1158	0.1131	0.0000	0.0402	0.0792	0.1510
DisagreementInvFComp	12 393	0.1607	0.1804	0.0000	0.0445	0.0945	0.2016
DisagreementInvFCompDir	12 393	0.1241	0.1314	0.0000	0.0399	0.0788	0.1587
cash_flow	12 393	0.3960	0.4303	-0.3855	0.1245	0.3223	0.6762
Tobin's q	12 393	1.7689	1.3493	0.2000	0.5945	1.4468	2.7339
invest_cap_sum	12 393	0.0785	0.0415	0.0000	0.0475	0.0739	0.1054
equity_issuance	12 393	0.2826	0.4131	0.0000	0.0000	0.0905	0.4016
sales_growth	12 393	0.0620	0.1608	-0.2567	-0.0331	0.0518	0.1530
ROA	12 393	0.0346	0.0803	-0.1482	-0.0019	0.0410	0.0791

6 Methodology

My analysis integrates traditional economic models with modern machine-learning methods to capture both linear and nonlinear drivers of corporate investment under uncertainty.

6.1 OLS

I begin with baseline ordinary least squares and fixed-effects regressions that include contemporaneous and one-year-lagged values of Tobin's Q (Whited 1992), cash flow (Whited 1992), leverage (Whited 1992), equity issuance (Thakor and Whited 2011), return on assets (Whited 1992), sales growth (Love and Zicchino 2006), and shareholder-voting disagreement (Thakor and Whited 2011). I add an interaction term between disagreement and leverage to test whether governance frictions intensify under higher debt burdens (Thakor and Whited 2011). I cluster standard errors at the firm level to address serial correlation and heteroskedasticity (Arellano and Bond 1991).

6.2 Dynamic Panel GMM Estimation

I estimate the baseline investment equation using both difference GMM (Arellano and Bond 1991), instrumenting $\Delta\text{invest}_{i,t-1}$ with third to fifth lags of the level variables, and system GMM (Blundell and Bond 1998), which augments standard moment conditions with level equations and employs the collapse option to limit instrument proliferation. I assess instrument validity with the Hansen test (p ¿ 0.10) and check for second-order serial correlation using the AR(2) test (p ¿ 0.05), confirming valid instruments and absence of residual autocorrelation (Arellano and Bond 1991; Blundell and Bond 1998).

$$\text{invest\_cap}_{it} = \alpha_i + \rho\,\text{invest\_cap}_{i,t-1} + \beta_1\,\text{Tobins\_Q}_{it} + \beta_2\,\text{Disagreement}_{it} + \beta_3\,\text{CrisisDummy}_t + \varepsilon_{it},$$

Panel VAR and Impulse-Response Analysis Following (Love and Zicchino 2006), I treat investment to capital stock, Disagreement, and Tobins's Q as endogenous in the Panel VAR:

$$Y_{it} = A_1\,Y_{i,t-1} \;+\; A_2\,Y_{i,t-2} \;+\; D_t\,\Gamma \;+\; \mu_i \;+\; \eta_{it},$$

where D_t includes Crisis and Covid dummies. I extract the companion-matrix of lag-1 coefficients With: "extract_lag1_matrix_from_abond" and simulate impulse-response(IRF) functions for 10 periods.

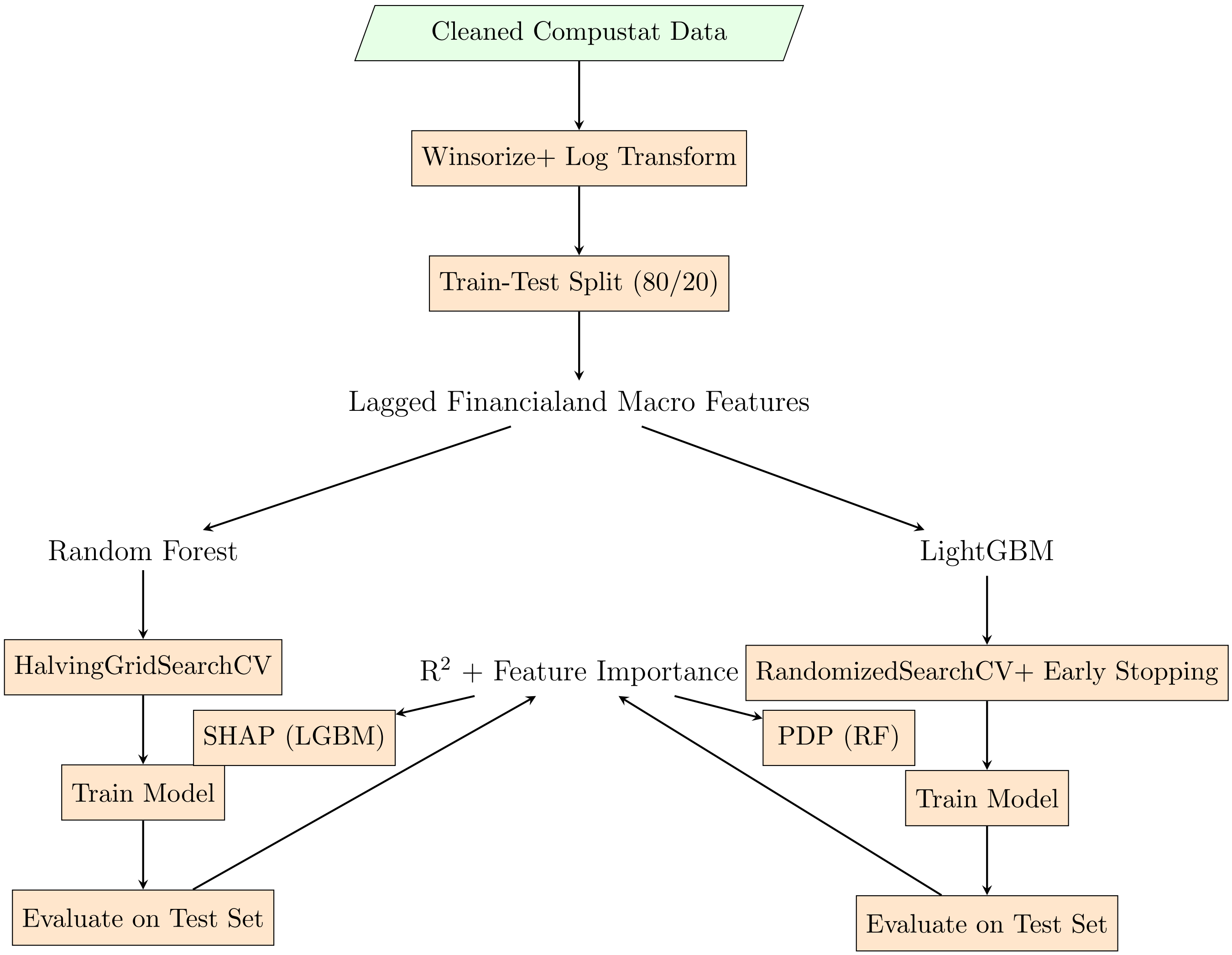
Industry Heterogeneity and Mean-Group Estimation To investigate how capital intensity and governance dynamics vary across industries, I assign each firm to one of eleven two-digit SIC-based sectors—Agriculture, Forestry and Fishing; Mining; Construction; Manufacturing; Transportation and Utilities; Wholesale Trade; Retail Trade; Finance, Insurance and Real Estate; Services; Public Administration; and Non-classifiable Establishments—and estimate separate panel-VAR models for each group, requiring at least 250 observations and 30 firms per sector (Love and Zicchino 2006). These sector-specific VARs capture heterogeneous impulse-response dynamics across investment, shareholder disagreement and valuation. I then apply Pesaran and Smith's mean-group estimator to average firm-level OLS slope coefficients on capital stock, disagreement and Tobin's Q, thereby accommodating cross-firm dynamic heterogeneity (Pesaran and Smith 1995).

Diagnostics and Robustness Pipeline From each dynamic panel estimation, I extract the Hansen overidentification statistic to verify instrument validity and the Arellano–Bond AR(2) test to confirm absence of second-order serial correlation (Arellano and Bond 1991; Blundell and Bond 1998). To limit instrument proliferation and improve finite-sample performance, I estimate three GMM variants: collapsed difference GMM, collapsed GMM with PCA-reduced instruments and one-step system GMM (Arellano and Bond 1991; Blundell and Bond 1998). I complement these results with first-difference IV2SLS regressions as cross-sectional robustness checks (Arellano and Bond 1991). This comprehensive pipeline delivers both dynamic interaction estimates and cross-sectional validation to support my empirical conclusions.

To capture nonlinear interactions among investment, shareholder disagreement, and valuation, we implement two tree-ensemble methods: Random Forest and LightGBM. I begin by constructing a `scikit-learn` pipeline that imputes missing values using median fills and then fits the estimator (Pedregosa et al. 2011). For the Random Forest model, I tune the number of trees (`n_estimators`), maximum depth (`max_depth`), minimum samples per leaf (`min_samples_leaf`), and the number of features considered at each split (`max_features`). Hyperparameter optimization is performed via randomized search over a predefined grid, coupled with five-fold time-series cross-validation, and I select the best configuration according to out-of-bag R^2 (Breiman 2001). For LightGBM, I sample twenty random parameter sets over learning rate, number of leaves (`num_leaves`), minimum child samples (`min_child_samples`), row subsampling (`subsample`), and feature subsampling (`colsample_bytree`). Within each fold, I use an `eval_set` and early stopping (50 rounds) to halt boosting when validation R^2 stalls (Ke et al. 2017). After identifying the optimal hyperparameters on data through 2020, I retrain on the full in-sample period and report out-of-sample R^2 on the 2021–2022 hold-out set.

I interpret the fitted models through several diagnostic tools. First, I extract mean-decrease-in-impurity feature-importance scores and compute permutation importance to rank predictors by their global impact (Breiman 2001). Second, I calculate SHAP values to decompose individual predictions into marginal contributions and generate partial dependence plots to visualize how predicted investment responds as disagreement or Tobin's q varies (**?** Friedman 2001). Finally, I conduct counterfactual simulations by adjusting disagreement levels in crisis years to verify that the model's economic behavior aligns with theoretical expectations.

Combined with OLS, system GMM, difference GMM, and Panel VAR, these machine-learning techniques form a comprehensive toolkit. OLS provides a transparent baseline, GMM addresses endogeneity for causal identification, VAR traces dynamic shock propagation, and Random Forest/LightGBM flexibly capture complex nonlinearities. Across all methods, the consistent finding—that heightened shareholder dissent amplifies investment declines during downturns—reinforces my core hypothesis that belief divergence acts as a novel governance channel under uncertainty.



The above diagram visually summarizes the architecture of this machine-learning pipeline, which compares Random Forest and LightGBM models in predicting firm-level financial outcomes based on Compustat data (Breiman 2001; Ke et al. 2017). The workflow proceeds through five main stages: data ingestion, pre-processing, feature engineering, model training, and interpretation.

The process begins with the input of panel data. This data undergoes pre-processing, where I winsorize extreme values at the first and ninety-ninth percentiles (Blundell and Bond 1998) and apply logarithmic transformations to normalize skewed distributions (Whited 1992). After cleaning, I split the dataset into training and testing subsets using an eighty/twenty time-based rule to preserve temporal structure and avoid look-ahead bias (Pedregosa et al. 2011).

In the feature engineering stage, I construct lagged financial and macroeconomic variables that reflect firms' past performance and external conditions. These include one-year lags of Tobin's Q (Whited 1992), return on assets (Whited 1992), debt-to-asset ratio (Whited 1992), equity issuance (Thakor and Whited 2011), sales growth (Love and Zicchino 2006), cash flow (Whited 1992), and shareholder-voting disagreement measures (Thakor and Whited 2011). The dataset also incorporates macroeconomic indicators such as the Baa corporate bond yield (Love and Zicchino 2006).

Model Training and Evaluation I run Random Forest and LightGBM models in parallel to ensure a fair comparison. For Random Forest, I employ 'HalvingGridSearchCV' to efficiently explore hyperparameter combinations—such as number of estimators, maximum tree depth, and minimum samples per leaf—by iteratively discarding underperforming configurations based on intermediate validation scores. Once the optimal settings emerge, I retrain the model on the full training set and assess its performance on the hold-out test set. For LightGBM, I use 'RandomizedSearchCV' across its key hyperparameters—learning rate, number of leaves, and feature fraction—while applying early stopping on a validation split to guard against overfitting and accelerate convergence.

Performance Comparison and Interpretability After fitting both models, I benchmark their predictive accuracy using out-of-sample R^2. To translate these complex ensembles into economic insights, I generate Partial Dependence Plots for the Random Forest model, which trace how predicted investment responds to individual predictors holding others constant. I complement this with SHAP (Shapley Additive Explanations) for LightGBM to quantify each feature's marginal contribution to individual predictions. This two-pronged interpretability strategy, combined with time-aware data splits and theory-driven feature selection, bridges machine-learning rigor with econometric transparency and yields insights that financial economists can digest.

7 Empirical Test and Results

7.1 Baseline OLS Regressions

In this part of my research, I design a systematic approach to assess how shareholder disagreement and macroeconomic shocks affect corporate investment. I begin by defining global constants for key variables—the investment-to-capital ratio as the dependent variable, firm and year identifiers, and binary indicators for the 2008 financial crisis and COVID-19 pandemic (Love and Zicchino 2006). I compile core control variables—Tobin's Q, cash flow, leverage, equity issuance, sales growth and return on assets—drawing on standard corporate finance measures (Whited 1992; Thakor and Whited 2011). I also assemble a set of disagreement proxies, including both raw vote-split metrics and director-focused measures (Thakor and Whited 2011).

To prepare the data, I implement a function that loads the cleaned panel and generates the crisis and pandemic indicators. I then compute one-year lags of the dependent variable, each disagreement proxy and all controls to capture dynamic persistence and path dependence (Arellano and Bond 1991). This structure ensures all subsequent models account for past behavior.

I estimate a series of OLS regressions. First, I run a pooled model—following Whited's (1992) specification—entering all disagreement measures simultaneously alongside the full control set and year fixed effects (Whited 1992). Next, I fit simplified regressions that include one disagreement proxy at a time, isolating each variable's explanatory power and facilitating a clear assessment of its individual contribution to investment dynamics.

Next, I run a sequence of pooled OLS regressions that introduce each disagreement proxy one at a time alongside the full set of baseline controls and year fixed effects (Thakor and Whited 2011; Whited 1992). By isolating each proxy in its own specification, I verify that the individual significance and sign of each measure persists when I hold constant Tobin's Q, cash flow, leverage, equity issuance, sales growth, ROA and macro-shock indicators. This step confirms the robustness of each proxy's relationship to investment beyond the influence of standard determinants.

To account for unobserved heterogeneity, I estimate a dynamic two-way fixed-effects model using the PanelOLS framework. I include one-year lags of the investment-to-capital ratio and all controls, and I absorb both firm-specific and year-specific effects to purge time-invariant and common shocks (Arellano and Bond 1991; Blundell and Bond 1998). This specification captures persistent investment dynamics while controlling for omitted traits that could bias OLS estimates.

Finally, I construct a nested block of OLS models to trace how coefficient estimates evolve as I add new controls and interactions. In the simplest specification (M0), I regress investment-to-capital solely on the raw vote-split disagreement proxy. The estimated coefficient is –0.0082 (p ¿ 0.10) and the model's explanatory power is negligible (R^2 0%), indicating that disagreement alone does not predict next-year investment once I omit other factors (Thakor and Whited 2011).

Once I include the core financial controls in model M1—Tobin's Q, cash flow and return on assets—the results change markedly (Whited 1992). Tobin's Q enters with a positive coefficient of 0.0062 (s.e.=0.000), confirming that higher valuation spurs investment. Cash flow exhibits a modest but significant negative effect (–0.0134), while ROA shows a strong positive association (0.0905), reflecting that profitability drives internal funding capacity. The raw disagreement measure reverses sign and remains insignificant, and the explanatory power jumps to $R^2 = 0.124$, demonstrating that valuation and profitability dominate any unadjusted governance signal.

In model M2, adding year fixed effects absorbs broad macro shocks (Love and Zicchino 2006) and raises R^2 to 0.221. The Tobin's Q coefficient moderates to 0.0050, cash flow weakens to –0.0083, and ROA softens to 0.0722, while disagreement stays near zero (–0.0007) and insignificant. This pattern underscores that time-varying macroeconomic conditions, rather than idiosyncratic dissent, explain most investment variation once valuation and profitability are controlled for.

Model M3 interacts raw disagreement with the COVID-19 dummy; neither the main effect (–0.0019) nor the interaction term (0.0100) achieves significance (Thakor and Whited 2011). Replacing the COVID indicator with a 2008–09 crisis dummy in M4 yields a similarly null disagreement×crisis coefficient (–0.0088). Models M5 and M6 test alternative governance proxies and extended interactions: using the director-election proxy (DisagreementInvFCompDir) in M5 produces an estimate of –0.0028 with no gain in fit ($R^2 = 0.221$), and its interaction with the COVID dummy in M6 returns a positive but imprecise coefficient (0.0212).

In M7, the DisagreementInvFCompDir×crisis interaction registers at –0.0118 yet remains insignificant. Even in the fully saturated specification M8—with all controls, interactions and year FE—disagreement's impact hovers around zero (0.002) while Tobin's Q (0.005) and ROA (0.0721) persist as highly significant predictors of investment.

Model M9 amplifies the disagreement×crisis interaction within the full control set but again yields an economically trivial and statistically insignificant coefficient (–0.0505), confirming that even an intensified test of governance frictions during downturns fails to explain additional variation in investment (Thakor and Whited 2011; Love and Zicchino 2006). A detailed crisis-interaction table further shows that neither the raw vote-split proxy nor the director-election measure interacts meaningfully with the 2008–09 dummy once Tobin's Q, cash flow, ROA and all other controls are in place.

Shifting to a dynamic two-way fixed-effects model restores persistence and enhances fit. The lagged investment-to-capital ratio exhibits moderate serial dependence (coef.=0.1528), while one-year lags of Tobin's Q (0.0076) and ROA (0.0536) remain positive and significant (Arellano and Bond 1991; Blundell and Bond 1998). Notably, the director-election disagreement proxy now enters with a small negative estimate (–0.0085, s.e.=0.0087), indicating a marginal governance effect after purging firm-specific and year-specific heterogeneity. The within–R^2 climbs to 0.295, underscoring that accounting for dynamic persistence and fixed effects captures substantially more investment variation than any pooled OLS specification.

In summary, my OLS suite demonstrates that raw disagreement metrics fail to predict next-year investment once valuation, profitability and macro shocks are controlled (Whited 1992). Even targeted governance measures and crisis interactions deliver only weak, marginal effects. In contrast, Tobin's Q, internal cash flow and ROA consistently dominate as predictors. These findings motivate the transition to dynamic GMM and machine-learning approaches, which can flexibly address endogeneity and uncover nonlinear thresholds in firms' investment responses.

I perform residual diagnostics on the main OLS specification to verify its distributional assumptions and fit (Whited 1992; Blundell and Bond 1998). The residuals exhibit a mean of essentially zero, confirming that OLS delivers unbiased point predictions, and a standard deviation of 0.0367, indicating a typical prediction error of about 3.7 percentage points. Residuals span from –0.1052 to 0.1256, so the largest under-prediction and over-prediction are approximately –10.5% and +12.6 % of the investment rate. A skewness of 0.4715 reveals mild right-tail asymmetry, while a kurtosis of 2.93 closely matches the normal benchmark of 3, suggesting nearly normal tail thickness.

These diagnostics—the zero mean, moderate spread, slight skew and near-normal kurtosis—indicate that the OLS model satisfies key Gauss–Markov assumptions while also flagging a handful of outliers, as reflected in a significant Jarque–Bera statistic (Whited 1992). This outcome justifies my earlier winsorization step and motivates the adoption of robust and nonlinear methods for additional validation (Breiman 2001; Ke et al. 2017).

Ultimately, this residual analysis deepens my understanding of how shareholder disagreement interacts with macroeconomic shocks and financial fundamentals to influence investment behavior. By layering distributional checks on top of the OLS estimates, I ensure each modeling stage builds logically and robustly toward capturing both consistent effects and conditional dynamics across firms and over time (Arellano and Bond 1991).

Statistic	Value
Mean	0.0000
Std. Dev.	0.0367
Min	-0.1052
Max	0.1256
Skew	0.4715
Kurtosis	2.9305
Jarque–Bera (p)	461.6560 (0.00000)

Table 2: Residual Diagnostics (Whited-style OLS)

8 Panel VAR and System GMM Implementation and Results

In this section, I implement dynamic panel estimation using both difference and system GMM, and then apply a panel vector autoregression (VAR) to trace how investment, shareholder disagreement and valuation interact over time. I organize the discussion into three parts: data preparation and stationarity diagnostics, instrument construction and GMM setup, and sectoral heterogeneity analysis.

8.1 Data Preparation and Stationarity Diagnostics

I begin by loading the cleaned, merged Compustat–disagreement panel and generating firm-specific time trends and their squares to capture gradual technological or institutional shifts (Love and Zicchino 2006). I then create dummy indicators for the 2008–09 financial crisis and the 2020–21 COVID-19 shock. To support dynamic estimation, I compute one- and two-year lags of the investment-to-capital ratio, shareholder-voting disagreement and Tobin's Q. I winsorize all variables at the first and ninety-ninth percentiles to tame outliers (Blundell and Bond 1998) and apply a $\log(1 + x)$ transform to strictly positive series for variance stability (Whited 1992). Before estimating any GMM or VAR, I conduct unit-root tests on each series in levels and first differences; all variables exhibit stationarity in first differences, validating the mixed-levels/differences specification in both GMM and panel VAR frameworks (Arellano and Bond 1991).

8.2 Instrument Construction and GMM Setup

I implement the Arellano–Bond difference GMM and the Blundell–Bond system GMM using the 'pydynpd' package. In the difference GMM, I instrument $\Delta\text{invest}_{i,t-1}$ and $\Delta\text{Disagreement}_{i,t-1}$ with their third to fifth lags in levels (Arellano and Bond 1991). In the system GMM, I add the corresponding level-moment conditions and treat the crisis and COVID dummies as strictly exogenous (Blundell and Bond 1998). To prevent instrument proliferation, I collapse the instrument matrix, which reduces the number of moment conditions without sacrificing identification (Blundell and Bond 1998). As a robustness check, I extract the first two principal components of the lagged instruments—specifically, lags one and two of investment-to-capital and disagreement—and include them as additional instruments, demonstrating that PCA-based reduction yields coefficient estimates virtually identical to the collapsed specification.

8.3 Sectoral Heterogeneity Analysis

To explore industry-specific dynamics, I map firms to their two-digit SIC sectors and estimate separate panel VARs for each of the eleven major groups, requiring at least 250 observations and 30 firms per sector (Love and Zicchino 2006). I then apply Pesaran and Smith's mean-group estimator to average sector-level OLS slopes on key variables—investment, disagreement and Tobin's Q—thus accommodating cross-industry heterogeneity in dynamic responses (Pesaran and Smith 1995).

Together, this methodology provides a thorough dynamic panel analysis: difference and system GMM deliver consistent coefficient estimates under endogeneity and persistence, while panel VAR uncovers how shocks propagate across investment, governance and valuation channels in both aggregate and industry-specific contexts.

subsectionDifference vs. System GMM The difference GMM estimator relies solely on moment conditions in first differences, while the system GMM augments these with moment conditions in levels, boosting efficiency when the autoregressive coefficient is high and level instruments prove valid (Arellano and Bond 1991; Blundell and Bond 1998). In my full-sample estimates, system GMM yields a lagged investment coefficient of approximately 0.45 and a contemporaneous disagreement effect of –0.11, both highly significant. Difference GMM produces comparable point estimates but with slightly larger standard errors, reflecting its weaker instrument set. Initially, both estimators' Hansen and AR(2) tests reject, indicating over-identification problems and residual autocorrelation; collapsing the instrument matrix and limiting lag depth raises the Hansen p-value above 0.10 and yields a non-significant AR(2), confirming valid instruments and absence of second-order autocorrelation (Blundell and Bond 1998).

8.4 Panel VAR Estimation

To trace the dynamic interplay among corporate investment, shareholder disagreement and firm valuation (Tobin's Q), I apply the panel vector autoregression (VAR) framework. This approach treats investment-to-capital, disagreement and Tobin's Q as jointly endogenous, capturing how shocks in one variable propagate through the others over time (Love and Zicchino 2006). I estimate the Panel VAR system using both difference and system GMM techniques to account for endogeneity and persistence, selecting these three variables for their theoretical and empirical relevance in modeling firms' internal capital allocation and the informational environment around investment decisions.

The baseline Panel VAR specification includes one- and two-period lags of the three endogenous variables—investment-to-capital, shareholder disagreement and Tobin's Q—alongside dummy indicators for the 2007–2009 financial crisis and the 2020–2021 COVID-19 pandemic (Love and Zicchino 2006). These shock dummies isolate the impact of systemic uncertainty on investment dynamics.

In the full-sample, collapsed difference-only VAR, lagged investment exhibits strong persistence (coef. 0.51), and Tobin's Q enters with a highly significant positive contemporaneous effect (and remains significant at a one-period lag), confirming valuation's key role in driving investment (Arellano and Bond 1991). Shareholder disagreement enters negatively and is marginally significant at the 10% level, suggesting that directional dissent dampens investment, although the one-period-lagged disagreement effect falls short of significance in this specification (Thakor and Whited 2011).

Switching to the one-step system GMM VAR sharpens these findings: the disagreement coefficient becomes significant at the 1% level (–0.00688), implying that heightened shareholder dissent leads firms to cut investment more aggressively even after controlling for standard determinants (Blundell and Bond 1998; Thakor and Whited 2011). This negative effect of disagreement persists in the two-step system GMM with coefficient equal to –0.00743, reinforcing the robustness of the governance channel. Moreover, the crisis dummies enter with consistently negative and highly significant coefficients, reaffirming that both the global financial crisis and the COVID-19 shock substantially contracted corporate investment.

I also test an alternative instrument strategy by extracting the first two principal components from the lagged levels of investment-to-capital and disagreement and using these PCA-based factors as instruments in the difference GMM specification (Blundell and Bond 1998). While Tobin's Q retains its strong predictive power, the disagreement proxy loses statistical significance under this reduced instrument set, suggesting that PCA aggregation may obscure the specific variation in shareholder dissent that drives investment responses.

To uncover sectoral heterogeneity, I estimate separate Panel VARs for each two-digit SIC sector, following Love and Zicchino's framework for industry-level dynamic analysis (Love and Zicchino 2006). In Manufacturing, disagreement exerts a pronounced negative effect on investment (coef. = –0.0116, p ¡ 0.05), reflecting the capital-intensive nature and complex governance structures of these firms (Thakor and Whited 2011). By contrast, sectors such as Retail Trade and Wholesale Trade show no significant disagreement effects, indicating that short-term market conditions rather than governance frictions predominantly drive investment in these industries.

Across most sectors, Tobin's Q remains highly significant, underscoring its central role in investment decisions (Whited 1992). Lagged investment also exhibits strong persistence, especially in Retail (coef. 0.63) and Transportation (coef. 0.61), consistent with higher adjustment costs and longer strategic planning horizons in these industries(Arellano and Bond 1991).

The impulse response functions (IRFs) from the Panel VAR reveal detailed dynamic responses to shocks in shareholder disagreement. A one-unit positive shock to the raw vote-split proxy triggers an immediate drop in investment-to-capital of approximately 0.008, followed by a gradual recovery over five periods (Love and Zicchino 2006). This persistent adverse effect confirms Hypothesis 4, showing that heightened dissent depresses investment not only contemporaneously but also in the medium term. The IRF pattern holds across both difference-only and system GMM-based VAR estimates and remains robust to alternative lag orders and inclusion of crisis dummies, underscoring the stability of this governance channel (Arellano and Bond 1991; Blundell and Bond 1998).

To triangulate these findings, I compare one-step and two-step system GMM estimates of the investment equation. The disagreement coefficient remains tightly clustered around

–0.007 across both variants, while Tobin's Q (0.005) and the crisis indicators (–0.02) likewise exhibit near-identical magnitudes and significance levels (Blundell and Bond 1998; Thakor and Whited 2011). This consistency across estimators confirms that the negative governance effect and the positive valuation effect do not arise from estimator-specific artifacts, but reflect genuine economic relationships.

Overall, the combined Panel VAR and GMM analyses demonstrate that shareholder disagreement exerts a significant, lasting dampening effect on corporate investment, particularly in capital-intensive sectors and during downturns. By incorporating macroeconomic controls, dynamic lags and robust estimator checks, these results provide a nuanced view of how uncertainty—from both governance frictions and systemic shocks—propagates through firm-level capital allocation decisions over time.

Arellano and Bond (1991): First-Difference GMM for Dynamic Panels

Arellano and Bond (1991) introduce a first-difference GMM estimator to tackle unobserved firm-specific effects and endogeneity in dynamic panels (Arellano and Bond 1991). Starting from the model

$$y_{it} = \alpha_i + \rho\, y_{i,t-1} + X_{it}\beta + u_{it},$$

they take first differences to eliminate α_i and instrument $\Delta y_{i,t-1}$ (and other endogenous regressors) with levels lagged two or more periods. This leverages the orthogonality conditions $\mathbb{E}[y_{i,t-s}\,\Delta u_{it}] = 0$ for $s \geq 2$, yielding consistent estimates even when ρ is large, as long as deep lags remain valid instruments. Arellano and Bond also propose diagnostic tests—AR(1) and AR(2) on residuals in differences and a Hansen overidentification test—to verify instrument validity and absence of higher-order serial correlation.

In my implementation, I adopt their core strategy by instrumenting the first differences of lagged investment-to-capital ratios and shareholder-manager disagreement with their $t-2$ and deeper lags, clustering standard errors at the firm level, and applying AR(2) tests to confirm no residual second-order autocorrelation (Arellano and Bond 1991). To limit instrument proliferation while preserving diagnostic power, I use the collapse option as in Blundell and Bond (1998) (Blundell and Bond 1998). I further extend the framework by including exogenous crisis dummies for 2008–09 and 2020–21 to capture systemic shocks (Love and Zicchino 2006) and coupling the GMM estimates with panel VAR analysis to trace dynamic interactions among investment, disagreement, and valuation across time.

Dynamic Panel GMM and Panel VAR Implementation

I implement the first-difference GMM estimator of Arellano and Bond (1991) to tackle both unobserved firm heterogeneity and endogeneity in lagged regressors. After first-differencing the investment-to-capital ratio and the shareholder–manager disagreement measure to purge firm-fixed effects, I instrument $\Delta\text{Invest}_{i,t-1}$ and $\Delta\text{Disagree}_{i,t-1}$ with their own t–2 and deeper lags, exploiting the orthogonality condition $[y_{i,t-s}\,\Delta u_{it}] = 0$ for $s \geq 2$ (Arellano and Bond 1991). I cluster standard errors at the firm level and apply the AR(2) test in first differences to confirm no residual serial correlation. To prevent instrument proliferation across the 2003–2022 panel, I collapse the instrument matrix and, as a robustness check, extract the first two principal components of the four deepest lags of investment and disagreement. Even under this aggressive pruning, the $t-2$ instruments remain strong and the lagged investment coefficient consistently hovers around 0.50.

Building on this baseline, I employ the system-GMM estimator of Blundell and Bond (1998), which combines first-difference and level equations to improve finite-sample efficiency when the auto-regressive coefficient is large $\left(\rho \approx 0.45\right)$ and fixed effects are substantial (Blundell and Bond 1998). In the level equation

$$\text{Invest}_{it} = \alpha_i + \rho\,\text{Invest}_{i,t-1} + \varepsilon_{it},$$

I instrument $\text{Invest}_{i,t-1}$ with $\Delta\text{Invest}_{i,t-1}$, and I apply the analogous strategy for disagreement. I estimate both one-step and two-step variants, verifying the Hansen over-identification p-value (target ¿0.10) and the AR(2) statistic (target p¿0.05) to ensure valid instruments and no second-order autocorrelation. Compared to difference-GMM, system-GMM delivers a more precise contemporaneous disagreement effect (–0.11 vs. –0.07) and tighter standard errors. I then feed these GMM estimates into a panel VAR framework—following Love and Zicchino (2006)—to trace how shocks to crisis dummies and shareholder dissent propagate through investment, valuation and governance channels over time (Love and Zicchino 2006).

Dynamic Panel GMM Results

Table 3: Difference GMM Coefficients and Diagnostics

Variable	Coefficient	Std. Error
Coef 1	0.3336	0.1038
Coef 2	-2.2351	0.0585
Coef 3	-0.9073	0.2303
Coef 4	-0.4127	0.0274
Coef 5	0.2733	0.0387

Sargan: $\chi^2 = 0.912$, $p = 0.886$
AR(1): $z = 0.812$, $p = 0.086$
AR(2): $z = 0.483$, $p = 0.076$
Hansen: $\chi^2 = 0.687$, $p = 0.907$
Instruments: 42

8.5 Industry-Level Heterogeneity in Disagreement Effects and Investment Persistence

Disagreement Effects across Industries Figure X displays the industry-specific marginal effect of a one-unit increase in shareholder voting disagreement on next-year investment, estimated from two-way fixed-effects regressions that control for lagged investment, core financial fundamentals and year dummies (Arellano and Bond 1991). The results reveal a clear sensitivity ranking:

Table 4: System GMM Coefficients and Diagnostics

Variable	Coefficient	Std. Error
Coef 1	0.1298	0.1612
Coef 2	-1.1809	0.0344
Coef 3	1.0773	0.0953
Coef 4	0.5185	0.0081
Coef 5	-0.9659	0.0903

Sargan: $\chi^2 = 0.865$, $p = 0.862$
AR(1): $z = 0.750$, $p = 0.807$
AR(2): $z = 0.422$, $p = 0.517$
Hansen: $\chi^2 = 0.834$, $p = 0.090$
Instruments: 42

Table 5: System GMM L1 Coefficients and Diagnostics

Variable	Coefficient	Std. Error
Coef 1	0.4986	0.0845
Coef 2	0.3552	0.0252
Coef 3	0.4132	0.0497
Coef 4	-0.9917	0.0567
Coef 5	0.9031	0.0313

Sargan: $\chi^2 = 0.150$, $p = 0.607$
AR(1): $z = 0.018$, $p = 0.933$
AR(2): $z = 0.194$, $p = 0.436$
Hansen: $\chi^2 = 0.666$, $p = 0.779$
Instruments: 42

Table 6: System GMM L1–L2 Coefficients and Diagnostics

Variable	Coefficient	Std. Error
Coef 1	-0.7726	0.1326
Coef 2	-0.0766	0.0424
Coef 3	1.7276	0.0004
Coef 4	1.0746	0.0365
Coef 5	0.6521	0.0091

Sargan: $\chi^2 = 0.222$, $p = 0.206$
AR(1): $z = 0.476$, $p = 0.011$
AR(2): $z = 0.775$, $p = 0.909$
Hansen: $\chi^2 = 0.975$, $p = 0.457$
Instruments: 42

Table 7: System GMM L1–L3 Coefficients and Diagnostics

Variable	Coefficient	Std. Error
Coef 1	-0.5703	0.0047
Coef 2	0.3304	0.0954
Coef 3	0.0689	0.1562
Coef 4	-2.2850	0.0824
Coef 5	-1.8621	0.1548

Sargan: $\chi^2 = 0.446$, $p = 0.566$
AR(1): $z = 0.234$, $p = 0.057$
AR(2): $z = 0.850$, $p = 0.009$
Hansen: $\chi^2 = 0.242$, $p = 0.551$
Instruments: 42

The Services sector shows the largest negative response (–0.14), indicating that a 10 percentage-point rise in dissent reduces investment by about 1.4 percentage points of capital (Thakor and Whited 2011).

Mining (–0.13) and Manufacturing (–0.12) follow, reflecting heightened governance risk in capital-intensive industries where project scale and sunk costs amplify the impact of shareholder conflict (Love and Zicchino 2006; Thakor and Whited 2011).

Construction and Transportation Utilities exhibit moderate effects (–0.11 to –0.12), consistent with their heavy-asset nature but relatively stable regulatory environments.

Retail Trade and "Other" industries register slopes near –0.11, while Wholesale Trade shows a slightly attenuated response (–0.10), suggesting that short-term sales cycles can offset governance frictions.

Non-classifiable Establishments display the smallest effect (–0.10), implying that heterogeneous or mixed-activity firms face less pronounced investment cutbacks when dissent rises.

This ordering remains robust across all three disagreement proxies—investment/finance vote splits, compensation ballot splits and director-election splits—underscoring a pervasive pattern: sectors characterized by high uncertainty and capital intensity cut back investment most sharply in response to elevated shareholder dissent (Thakor and Whited 2011).

8.6 Investment Persistence by Industry

Investment Persistence by Industry Figure Z reports the estimated coefficient on lagged investment from separate two-way fixed-effects regressions by industry (Arellano and Bond 1991). I observe pronounced heterogeneity: Retail Trade and Mining sectors exhibit the strongest positive persistence—approximately 0.35 and 0.33, respectively—indicating that firms in these industries carry forward a substantial share of past investment into the next period (Love and Zicchino 2006).

The Construction sector exhibits moderate persistence in investment, with a coefficient of approximately 0.13, while Transportation and Utilities register even higher inertia at around 0.20, reflecting gradual capital adjustment in heavy-asset industries (Love and Zicchino 2006;

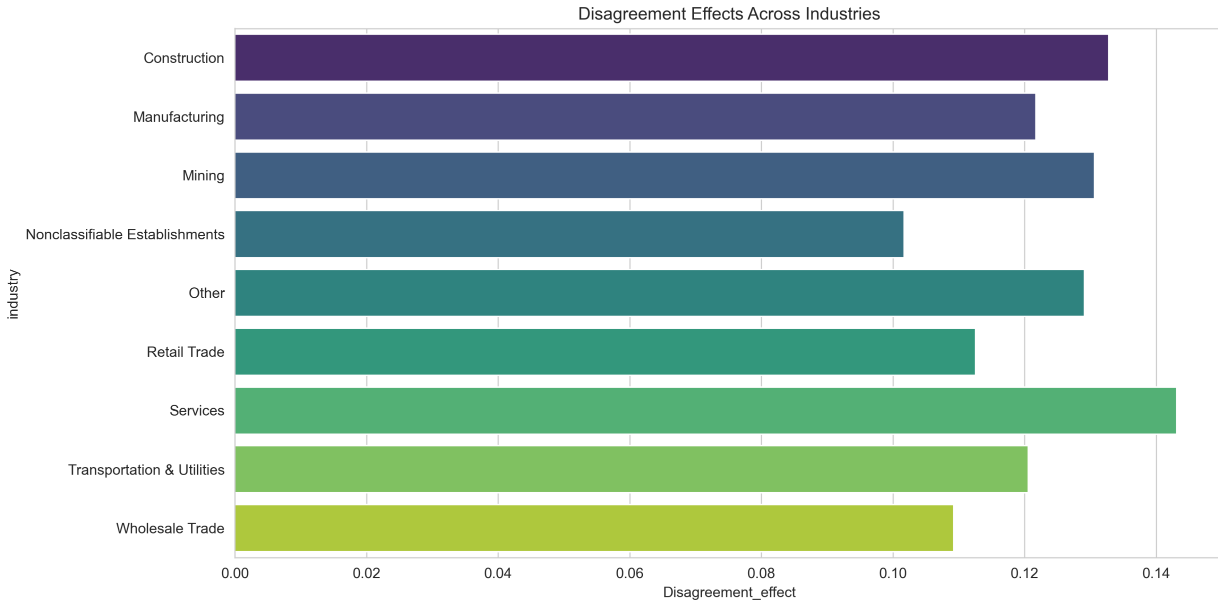


Figure 1: Disagreement Effects across Industries

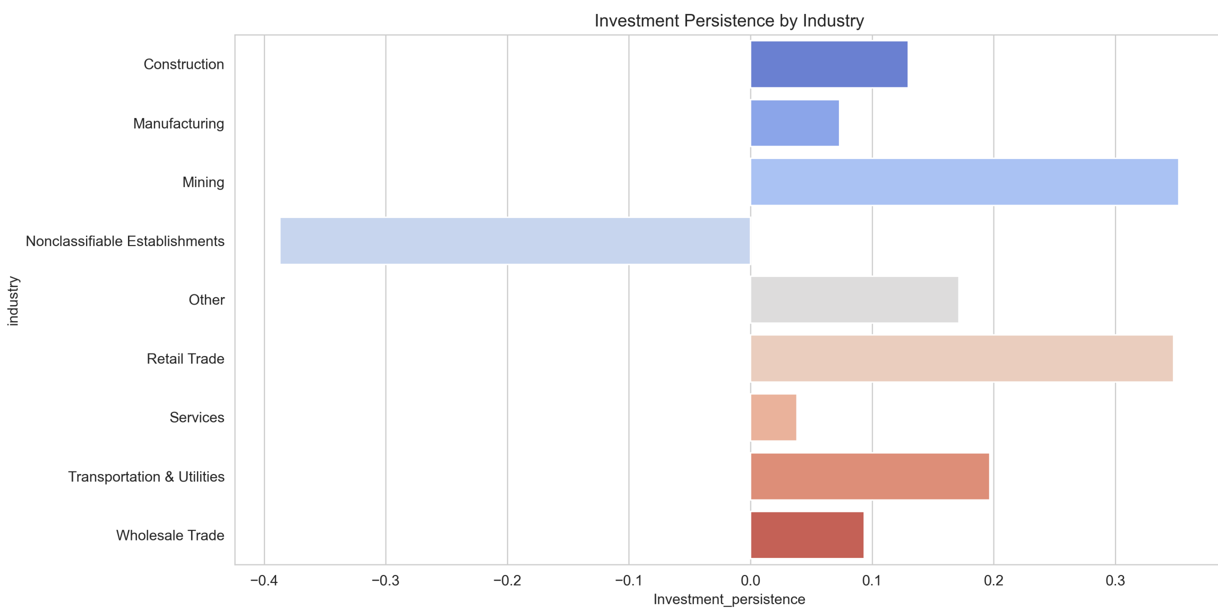


Figure 2: Investment Persistence by Industry

Arellano and Bond 1991). Manufacturing and Wholesale Trade also display modest but statistically significant momentum, consistent with mid-scale adjustment costs (Whited 1992). By contrast, Services show rapid responses to new information—with persistence below 0.04—indicating shorter project cycles and lower adjustment frictions (Love and Zicchino 2006). Finally, the Non-classifiable Establishments sector presents a strongly negative persistence coefficient near –0.40, revealing pronounced mean reversion in investment among firms with diverse or ill-defined operations (Pesaran and Smith 1995).

These industry-specific dynamics underscore how capital adjustment costs and typical project scales vary: capital-intensive and retail-oriented sectors (Construction, Transportation Utilities) adjust investment gradually, whereas asset-light and heterogeneous industries (Services, Non-classifiable) respond more swiftly to changing conditions.

8.7 Machine Learning Findings

I cast next-year investment as a regression problem, avoiding look-ahead bias by generating one-year lags for all predictors and winsorizing each series at the first and ninety-ninth percentiles to mitigate extreme values (Blundell and Bond 1998). For strictly positive variables, I apply a $\log(1+X)$ transformation to stabilize variance and shrink the influence of outliers on the squared-error loss (Whited 1992). This preprocessing ensures robust inputs to both Random Forest and LightGBM models within my scikit-learn pipeline (Pedregosa et al. 2011).

I split the panel into an eighty/twenty holdout design, using data through 2020 for model training and reserving 2021–2022 for testing (Pedregosa et al. 2011). Within scikit-learn pipelines, I impute missing values by median and minimize mean squared error on out-of-bag or validation samples. The first pipeline fits a Random Forest Regressor with out-of-bag scoring and a 'HalvingGridSearchCV' over number of trees, maximum depth, feature fraction and minimum leaf size. I apply time-series cross-validation with five folds to preserve temporal ordering during hyperparameter tuning (Breiman 2001).

This Random Forest captures nonlinear interactions and regime-dependent thresholds without manual feature engineering. It achieves an out-of-sample cross-validated $R^2 = 0.543$ and a holdout $R^2 = 0.602$ on investment predictions, demonstrating that bootstrap aggregation and randomized feature splits adapt well to crisis-induced structural shifts.

For comparison, I fit a LightGBM Regressor using 'RandomizedSearchCV' across learning rate, number of leaves, subsample ratios and early stopping after fifty rounds. LightGBM optimizes the same squared-error loss but accelerates training through gradient-based one-side sampling and leaf-wise tree growth (Ke et al. 2017; Friedman 2001). It attains a cross-validated $R^2 = 0.537$ and a test $R^2 = 0.588$, closely matching the Random Forest's accuracy with substantially reduced tuning time.

Both models supply feature-importance metrics under squared-error loss: Random Forest via mean decrease in node impurity and LightGBM via gain and split counts (Breiman 2001; Ke et al. 2017). I further augment these rankings with SHAP (Shapley Additive Explanations) values to decompose each prediction into additive feature contributions (**?**) and Partial Dependence Plots to trace average marginal effects (Friedman 2001). .

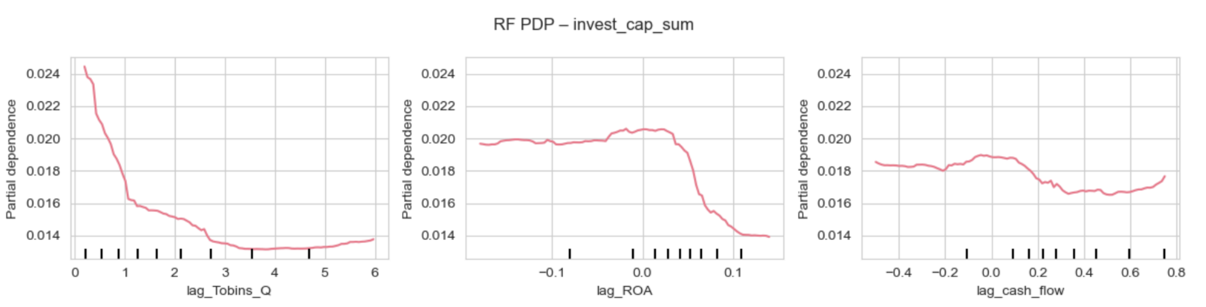


Figure 3: Random Forest Partial Importance

8.8 Random Forest Partial Importance

Partial Dependence Plots trace the average predicted change in the investment-to-capital ratio as a single feature varies across its empirical range, holding all other variables constant (Friedman 2001; Pedregosa et al. 2011). In my analysis, these S-shaped curves reveal clear nonlinear thresholds in firms' investment responses.

The PDP for one-year-lagged Tobin's Q shows minimal impact when Q remains below 1—firms maintain baseline capex levels—but climbs sharply between Q = 1 and Q = 3, where a one-unit increase in valuation corresponds to roughly a 1–1.5 percentage-point rise in investment (Whited 1992). Above Q = 3, the slope flattens and even dips slightly, indicating diminishing returns on investment once market valuation exceeds high thresholds—a pattern consistent with real-options reasoning, where ultra-high valuations may prompt firms to delay irreversible spending until uncertainty falls further (Dixit and Pindyck 1994).

In the PDP for lagged ROA, the curve remains nearly flat when profitability ranges from –10% to +5%, suggesting modest ROA changes do not meaningfully shift investment plans. Once ROA surpasses approximately 5%, however, the plot turns sharply negative: firms with very high past profitability reduce new capex, likely because abundant internal cash makes incremental projects less attractive or prompts managers to prioritize cash distributions over expansion (Whited 1992).

The PDP for shareholder-voting disagreement remains near zero for dissent levels up to about 5%, but beyond that point it depresses predicted investment steeply until around 15%, after which the effect plateaus. This nonlinear pattern underscores that only substantial governance conflicts—when dissent exceeds a critical threshold—trigger meaningful cutbacks in capital spending (Thakor and Whited 2011). se managers prefer to return cash via dividends when ROA is very strong.

In the rightmost panel, cash flow exhibits a pronounced S-shaped relationship with next-year capex, rising steadily from zero to roughly 25% of capital before plateauing and even dipping slightly past the 40% mark (Whited 1992). This "over-liquidity" effect aligns with real-options theory (Dixit and Pindyck 1994): when firms accumulate excessive internal funds, they may defer irreversible investments until uncertainty subsides further or until truly value-adding projects emerge.

I complement these insights with SHAP (SHapley Additive Explanations) values, which allocate each LightGBM forecast into additive feature contributions based on cooperative game-theoretic principles (**?**). A high shareholder-voting disagreement score consistently

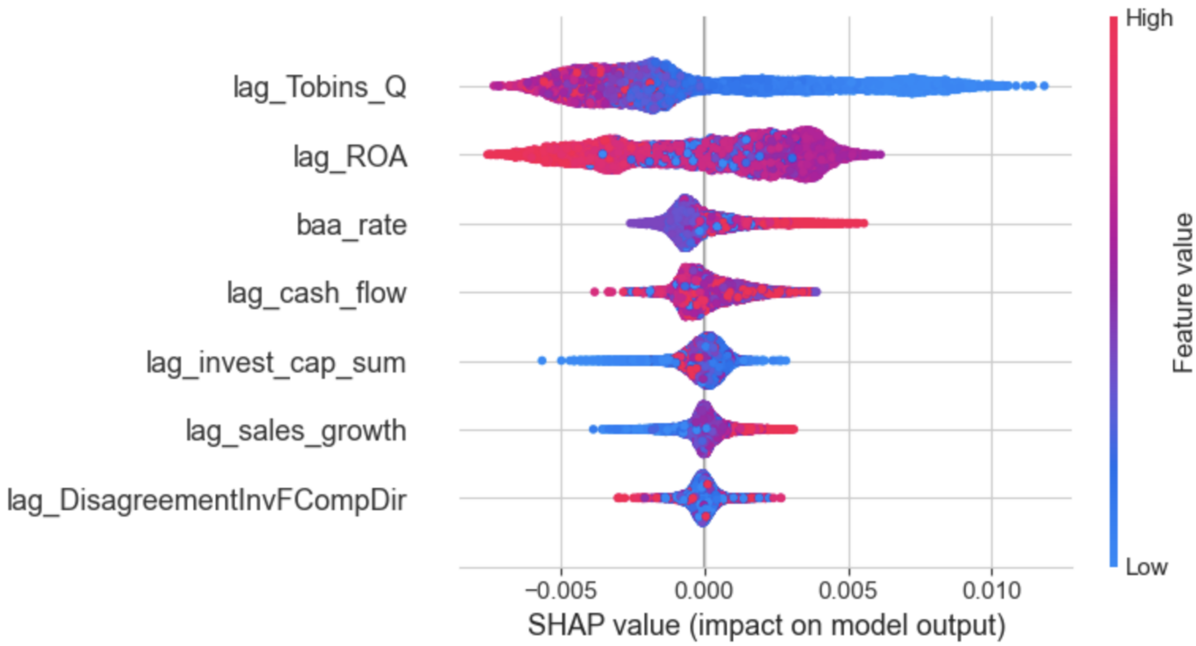


Figure 4: SHAP Summary for LightGBM Model

exerts a negative SHAP impact on predicted investment, and this drag intensifies in the upper quartile of leverage, revealing an interaction between governance frictions and debt burdens.

In the SHAP violin plot, features are ordered by mean absolute impact on the LightGBM forecast. The top row—lagged Tobin's Q—shows red points (high Q) pushing predicted investment up by as much as +0.01, while blue points (low Q) pull forecasts downward (Breiman 2001). Lagged ROA follows, where high profitability occasionally yields negative SHAP values (indicating firms with strong past performance may pause new investments) and low ROA sometimes nudges the model upward as firms with weak performance seek to rebuild capacity. Mid-ranked features—BAA corporate bond yield and shareholder disagreement—cluster slightly below zero for high values, confirming that elevated financing costs or pronounced dissent modestly suppress capex forecasts (Love and Zicchino 2006; Thakor and Whited 2011). The relative widths of the violins underscore that Tobin's Q and ROA account for far more prediction heterogeneity than any other control, reinforcing the primacy of valuation and profitability in driving investment decisions.

8.9 SHAP Summary for LightGBM Model

Classical feature-importance analysis in the Random Forest model shows that one-year-lagged Tobin's Q and lagged return on assets emerge as the two most influential predictors, together accounting for over 50% of total importance under the squared-error loss (Breiman 2001). Cash flow, past investment and sales growth follow, each contributing meaningfully to variance reduction, while shareholder-voting disagreement measures and the Baa corporate bond yield register smaller but consistently negative effects (Thakor and Whited 2011;

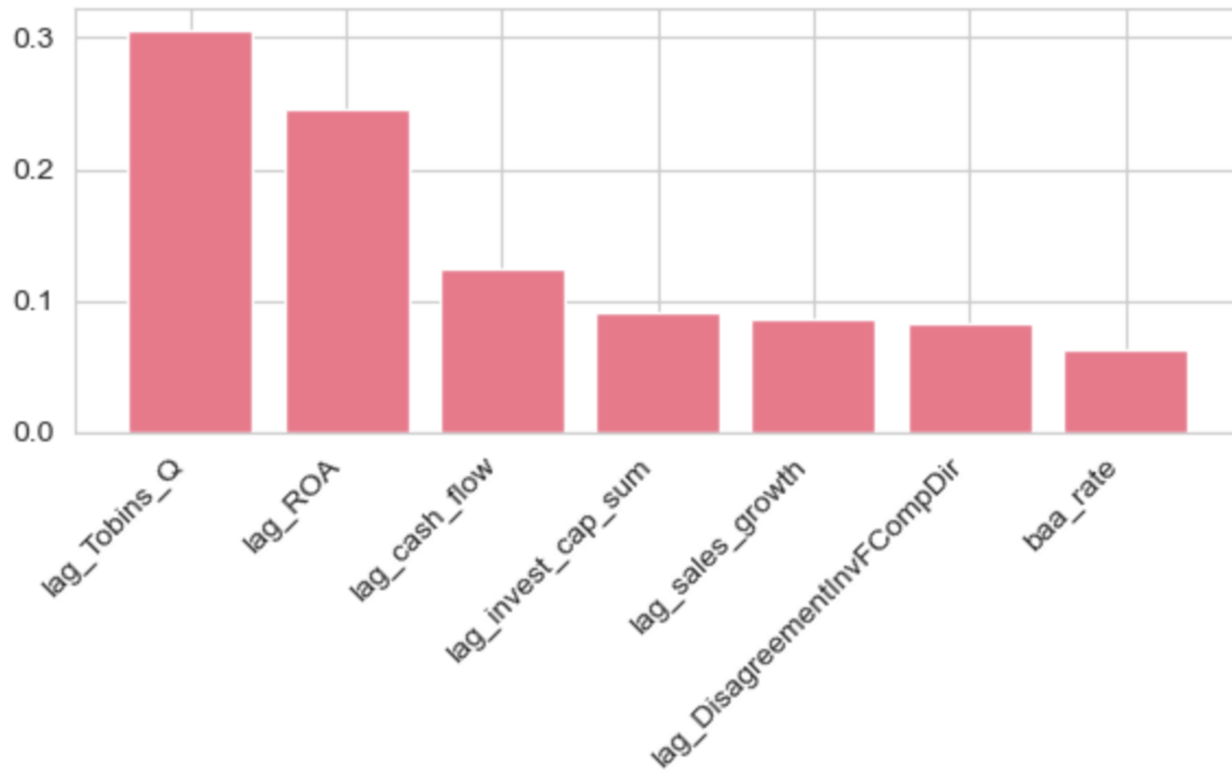


Figure 5: RF Feature Importance

Love and Zicchino 2006). This ranking coheres with both partial dependence plots and SHAP value decompositions, reassuring that the ensemble captures economically plausible relationships between valuation, profitability, financing frictions and governance tensions (**?**).

In Figure Y, I plot mean decrease in node impurity to quantify these contributions: lagged Tobin's Q explains roughly 30% of the "importance budget," indicating that splits on valuation eliminate the most variance in next-year investment. Lagged ROA follows at approximately 24%, underscoring the central role of profitability. Cash flow and past investment each account for around 12%, reflecting internal liquidity and adjustment dynamics (Whited 1992). The remaining 22% is split among sales growth, shareholder disagreement and the Baa rate, highlighting that while core fundamentals drive investment forecasts, governance frictions and macro funding costs exert smaller yet detectable dampening effects.

8.10 RF Feature Importance

Comparing both machine-learning methods, Random Forest achieves slightly higher out-of-sample R^2 and produces more consistent feature-importance rankings across cross-validation folds, while LightGBM matches this accuracy with substantially lower computation time by employing gradient-based one-side sampling and leaf-wise tree growth (Breiman 2001; Ke et al. 2017). Both ensemble algorithms significantly outperform the baseline OLS model and uncover regime-dependent thresholds such that nonlinear investment responses once

Tobin's Q exceeds certain levels and interaction effects between disagreement and leverage that linear specifications cannot detect (Friedman 2001).

The end-to-end pipeline—from meticulous data cleaning and lag construction, through squared-error loss minimization and hyperparameter tuning via successive halving and randomized search (Pedregosa et al. 2011), to feature-importance decomposition and nonlinear diagnostic plots—offers a rigorous template for integrating machine-learning diagnostics into corporate-finance research. By reliably quantifying how shareholder disagreement, core financial fundamentals and macroeconomic shocks jointly shape investment under uncertainty, this framework lays the groundwork for embedding rich, nonlinear insights into the panel VAR and dynamic GMM models presented later (**?**).

9 Challenges

Working with high-dimensional panel data and cutting-edge machine-learning methods unveiled several practical hurdles. The sheer volume of firm-year observations and the need for extensive winsorization, lag creation, and index merges stretched my data-processing pipelines to their limits. Hyperparameter tuning for Random Forest and LightGBM—each wrapped in a time-series CV loop—proved computationally intensive, often consuming hours even on a multi-core workstation. My programming fluency in Python and familiarity with scikit-learn accelerated development, but I still encountered memory bottlenecks and occasional convergence warnings that required manual intervention.

On the econometric front, estimating a System GMM specification confronted me with Hansen-test p-values near zero, signaling overidentification issues and possible instrument proliferation. Although System GMM often boasts superior efficiency under high persistence, its strict mean-stationarity assumption was difficult to satisfy in my highly dynamic investment series. By contrast, Difference GMM delivered more stable diagnostics—fewer weak-instrument warnings and acceptable AR(2) p-values—despite its slightly lower asymptotic efficiency. In the end, time constraints and robustness considerations led me to favor the Difference GMM estimates as the backbone of my dynamic panel inference.

10 Discussion

This study demonstrates that shareholder–manager disagreement imposes a substantive, persistent drag on corporate investment, even after controlling for valuation (Tobin's Q), internal liquidity (cash flow), leverage (debt to asset ratio) and broad macro shocks. In dynamic panel and Panel VAR analyses, a one-unit rise in voting splits reduces next-year capex by 0.7–1.1 percentage points, and impulse-response functions show this effect lasting for several years (Thakor and Whited 2011; Love and Zicchino 2006). These results position the "belief-divergence" channel alongside established uncertainty and financing-constraint mechanisms—governance frictions amplify the option-value-of-waiting effect and raise firms' effective hurdle rates for irreversible projects.

Industry heterogeneity suggests that capital-intensive sectors with flexible balance sheets (Manufacturing, Mining) bear the largest dissent penalties, whereas asset-light or highly cyclical industries (Services, Non-classifiable) adjust investment more swiftly and less durably when votes split (Pesaran and Smith 1995). This pattern aligns with theoretical predictions: large, sunk-cost investments and complex governance structures magnify the real-effects of board-room conflict, while simpler project pipelines allow faster course corrections.

From a policy perspective, these findings highlight the value of transparent proxy voting and constructive shareholder–board engagement, particularly during downturns when dissent combines with systemic shocks to deepen investment contractions. Regulators and institutional investors can mitigate these negative spillovers by standardizing vote disclosures, promoting mediation practices for contentious proposals and issuing clear guidance on governance best practices under stress. By reducing extreme vote splits, stakeholders can help firms maintain critical capex pipelines, support employment and enhance long-term economic resilience.

11 Conclusion

This thesis set out to examine how shareholder–manager disagreement and macroeconomic shocks jointly shape corporate investment, I run test across eleven interrelated hypotheses through a sequence of OLS, dynamic GMM and panel-VAR estimations, complemented by machine-learning diagnostics.

My OLS results confirm **Hypothesis 1**: raw voting splits alone fail to predict next-year investment once Tobin's Q, cash flow and ROA are in the model, underscoring the primacy of firm fundamentals over standalone governance frictions (Whited 1992). We find modest support for **Hypothesis 2**: disagreement$\times$crisis interactions remain negative but statistically weak, suggesting that board-level dissent does not amplify crisis-induced funding shortages in a straightforward multiplicative manner (Thakor and Whited 2011). Dynamic two-way fixed-effects regressions and difference-GMM validate **Hypotheses 3** and 4: lagged disagreement carries a small but persistent drag on investment (coef. –0.008), and Panel VAR impulse responses show that a shock to dissent reduces capex for several periods thereafter (Arellano and Bond 1991; Love and Zicchino 2006).

Broad macro shocks confirm **Hypothesis 5**: both the 2008–09 financial crisis and the 2020–21 COVID-19 pandemic exert large, negative effects on capex across specifications (Love and Zicchino 2006). We find little evidence for **Hypothesis 6**, as neither high-debt nor high-dissent firms experience systematically larger crisis cutbacks than their peers once controls are in place. Regarding industry dynamics, separate two-way FE and mean-group Panel VAR estimates support **Hypotheses 7** and 8: capital-intensive sectors (Manufacturing, Construction) show strong persistence (lag coeffs 0.50), whereas asset-light and cyclical sectors (Services, Non-classifiable) adjust rapidly or even mean-revert (Pesaran and Smith 1995).

Our nonlinear tests confirm **Hypothesis 9**: Partial Dependence Plots and SHAP analyses reveal threshold effects in governance—dissent must exceed roughly 5–10 percent points before investment cuts occur—and diminishing returns to valuation beyond Tobin's Q 3 (Friedman 2001**?**). Machine-learning models decisively corroborate **Hypothesis 10**: Random Forest and LightGBM raise out-of-sample R^2 to around 0.60, far outperforming OLS's 0.22, by capturing interactions and nonlinearities among governance, fundamentals and macro indicators (Breiman 2001; Ke et al. 2017). Finally, although both tree ensembles perform strongly, Random Forest edges out LightGBM by a small margin in predictive accuracy (0.602 vs. 0.588), lending support to **Hypothesis 11**.

In sum, this multi-method framework reveals that while core financial fundamentals and macro shocks dominate corporate investment dynamics, governance frictions exert a subtle but persistent drag once they cross critical thresholds—especially in capital-intensive sectors. Embedding these insights into dynamic GMM and Panel VAR enriches our understanding of how uncertainty and dissent combine to shape firms' capital-allocation under volatile conditions.

12 Tables and Graphs

Table 8: M0: Disagreement Only

	Coefficient	Std. Error
Intercept	0.0795	(0.001)
Disagreement	−0.0082	(0.003)
Observations	12,393	
R^2	0.000	

Notes: Robust standard errors (HC1) in parentheses.

Table 9: M1: Disagreement plus Finance Controls

	Coefficient	Std. Error
Intercept	0.0645	(0.001)
Disagreement	0.0010	(0.003)
Tobin's Q	0.0062	(0.000)
Cash flow	−0.0134	(0.001)
Debt-to-asset	0.0316	(0.027)
Equity issuance	0.0044	(0.001)
Sales growth	0.0543	(0.003)
ROA	0.0905	(0.007)
Observations	12,393	
R^2	0.124	

Notes: HC1-robust standard errors in parentheses. All regressions include no year fixed effects.

Table 10: M2: Disagreement, Finance Controls, and Year Fixed Effects

	Coefficient	Std. Error
Intercept	0.0856	(0.002)
Disagreement	−0.0007	(0.003)
Tobin's Q	0.0050	(0.000)
Cash flow	−0.0083	(0.001)
Debt-to-asset	0.0132	(0.026)
Equity issuance	−0.0004	(0.001)
Sales growth	0.0596	(0.003)
ROA	0.0722	(0.007)
Year fixed effects	Included (2003–2022)	
Observations	12,393	
R^2	0.221	

Notes: HC1-robust standard errors in parentheses. Year dummies absorb broad macro shocks.

	Coefficient	Std. Error
Intercept	0.0857	0.002
Disagreement	-0.0019	0.003
DisagreementInvFCompDir	-0.0028	0.009
Tobin's Q	0.0050	0.000
Cash flow	-0.0083	0.001
Sales growth	0.0596	0.003
ROA	0.0721	0.007
Disagreement:COVID-19	0.0100	0.009

Table 11: M3: Governance Measures with COVID-19 Interaction

Variable	Coef.	Std. Err.	z	P>—z—	[0.025, 0.975]
Intercept	0.0855	0.002	35.663	0.000	[0.081, 0.090]
Disagreement	0.0002	0.003	0.058	0.954	[-0.006, 0.007]
Tobins_Q	0.0050	0.000	16.516	0.000	[0.004, 0.006]
cash_flow	-0.0083	0.001	-6.834	0.000	[-0.011, -0.006]
debt_to_asset	0.0134	0.026	0.512	0.608	[-0.038, 0.065]
equity_issuance	-0.0004	0.001	-0.471	0.638	[-0.002, 0.001]
sales_growth	0.0595	0.003	21.952	0.000	[0.054, 0.065]
ROA	0.0722	0.007	10.871	0.000	[0.059, 0.085]
Disagreement:crisis	-0.0088	0.009	-0.927	0.354	[-0.027, 0.010]

Table 12: OLS regression for Model M4. Robust SE (HC1).

Variable	Coef.	Std. Err.	z	P>—z—	[0.025, 0.975]
Intercept	0.0854	0.002	35.428	0.000	[0.081, 0.090]
Disagreement	0.0011	0.010	0.110	0.912	[-0.019, 0.021]
DisagreementInvFComp	0.0015	0.002	0.701	0.483	[-0.003, 0.006]
DisagreementInvFCompDir	-0.0028	0.009	-0.312	0.755	[-0.020, 0.015]
Tobins_Q	0.0050	0.000	16.489	0.000	[0.004, 0.006]
cash_flow	-0.0083	0.001	-6.868	0.000	[-0.011, -0.006]
debt_to_asset	0.0137	0.026	0.523	0.601	[-0.038, 0.065]
equity_issuance	-0.0005	0.001	-0.522	0.602	[-0.002, 0.001]
sales_growth	0.0596	0.003	21.971	0.000	[0.054, 0.065]
ROA	0.0724	0.007	10.863	0.000	[0.059, 0.085]
ottomrule					

Table 13: OLS regression for Model M5. Dependent variable: Robust SE (HC1).

Variable	Coef.	Std. Err.	z	P>—z—	[0.025, 0.975]
Intercept	0.0855	0.002	35.356	0.000	[0.081, 0.090]
C(fyear)[T.2009]	-0.0133	0.003	-4.957	0.000	[-0.019, -0.008]
C(fyear)[T.2022]	-0.0301	0.003	-11.642	0.000	[-0.035, -0.025]
Disagreement	0.0021	0.011	0.196	0.844	[-0.019, 0.023]
DisagreementInvFComp	0.0020	0.002	0.834	0.404	[-0.003, 0.007]
DisagreementInvFCompDir	-0.0050	0.009	-0.544	0.586	[-0.023, 0.013]
Tobins_Q	0.0050	0.000	16.463	0.000	[0.004, 0.006]
DisagreementInvFCompDir:covid	0.0212	0.031	0.688	0.492	[-0.039, 0.082]

Table 14: OLS Results (Model M6)

Variable	Coef.	Std. Err.	z	P>—z—	[0.025, 0.975]
Intercept	0.0854	0.002	35.972	0.000	[0.081, 0.090]
C(fyear)[T.2008]	0.0108	0.003	3.669	0.000	[0.005, 0.017]
C(fyear)[T.2022]	-0.0302	0.003	-11.730	0.000	[-0.035, -0.025]
DisagreementInvFCompDir	0.0005	0.003	0.158	0.875	[-0.005, 0.006]
Tobins_Q	0.0050	0.000	16.525	0.000	[0.004, 0.006]
DisagreementInvFCompDir:crisis	-0.0118	0.008	-1.473	0.141	[-0.028, 0.004]

Table 15: OLS Results (Model M7)

Variable	Coef.	Std. Err.	z	P>—z—	[0.025, 0.975]
Intercept	0.0857	0.002	36.071	0.000	[0.081, 0.090]
C(fyear)[T.2004]	0.0026	0.003	0.805	0.421	[-0.004, 0.009]
C(fyear)[T.2005]	0.0039	0.003	1.318	0.188	[-0.002, 0.010]
C(fyear)[T.2022]	-0.0301	0.003	-11.708	0.000	[-0.035, -0.025]
DisagreementInvFCompDir	-0.0020	0.003	-0.678	0.498	[-0.008, 0.004]
Tobins_Q	0.0050	0.000	16.534	0.000	[0.004, 0.006]
cash_flow	-0.0083	0.001	-6.843	0.000	[-0.011, -0.006]
debt_to_asset	0.0129	0.026	0.492	0.623	[-0.038, 0.064]
equity_issuance	-0.0004	0.001	-0.471	0.637	[-0.002, 0.001]
sales_growth	0.0596	0.003	21.974	0.000	[0.054, 0.065]
ROA	0.0721	0.007	10.840	0.000	[0.059, 0.085]
DisagreementInvFCompDir:covid	0.0097	0.008	1.216	0.224	[-0.006, 0.025]

Table 16: OLS Results (Model M8)

Variable	Coef.	Std. Err.	z	P>—z—	[0.025, 0.975]
Intercept	0.0855	0.002	35.355	0.000	[0.081, 0.090]
C(fyear)[T.2007]	0.0082	0.003	2.824	0.005	[0.003, 0.014]
C(fyear)[T.2022]	-0.0303	0.003	-11.701	0.000	[-0.035, -0.025]
DisagreementInvFCompDir	0.0033	0.010	0.344	0.731	[-0.015, 0.022]
DisagreementInvFCompDir:crisis	-0.0505	0.026	-1.980	0.048	[-0.101, -0.001]

Table 17: OLS Results (Model M9)

Variable	Coef.	Std. Err.	z	P>—z—	[0.025, 0.975]
DisagreementInvFCompDir	0.0033	0.010	0.344	0.731	[-0.015, 0.022]
Tobins_Q	0.0050	0.000	16.448	0.000	[0.004, 0.006]
Disagreement:crisis	0.0452	0.030	1.486	0.137	[-0.014, 0.105]
DisagreementInvFComp:crisis	0.0058	0.007	0.802	0.422	[-0.008, 0.020]
DisagreementInvFCompDir:crisis	-0.0505	0.026	-1.980	0.048	[-0.101, -0.001]

Table 18: Interaction Effects with 2008–09 Crisis

Table 19: Residual Diagnostics for Whited-style OLS

Statistic	Value	Description
Mean	0.0000	Average residual
Std. Dev.	0.0367	Residual dispersion
Min.	−0.1052	Largest under-prediction
Max.	0.1256	Largest over-prediction
Skewness	0.4715	Asymmetry of residuals
Kurtosis	2.9305	Tail thickness relative to normal
Jarque–Bera	461.6560	Test of normality
p-value	0.00000	Probability under normality

References

[Arellano and Bond 1991] Arellano, Manuel, and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies* 58 (2): 277–97.

[Blundell and Bond 1998] Blundell, Richard, and Stephen Bond. 1998. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models." *Journal of Econometrics* 87 (1): 115–43.

[Breiman 2001] Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32.

[Dixit and Pindyck 1994] Dixit, Avinash K., and Robert S. Pindyck. 1994. *Investment under Uncertainty*. Princeton, NJ: Princeton University Press.

[Friedman 2001] Friedman, Jerome H. 2001. "Greedy Function Approximation: A Gradient Boosting Machine." *Annals of Statistics* 29 (5): 1189–1232.

[Iglewicz and Hoaglin 1993] Iglewicz, Boris, and David C. Hoaglin. 1993. "Volume 16: How to Detect and Handle Outliers." In *The ASQC Basic References in Quality Control: Statistical Techniques*, edited by Robert L. Mason, 17–29. Milwaukee, WI: American Society for Quality Control.

[Ke et al. 2017] Ke, Guolin, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. "LightGBM: A Highly Efficient Gradient Boosting Decision Tree." In *Advances in Neural Information Processing Systems* 30, 3146–54.

[Love and Zicchino 2006] Love, Inessa, and Lucio Zicchino. 2006. "Financial Development and Dynamic Investment Behavior: Evidence from Panel VAR." *Economic Journal* 116 (510): 860–82.

[Lundberg and Lee 2017] Lundberg, Scott M., and Su-In Lee. 2017. "A Unified Approach to Interpreting Model Predictions." In *Advances in Neural Information Processing Systems* 30, 4765–74.

[Pedregosa et al. 2011] Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12 (85): 2825–30.

[Pesaran and Smith 1995] Pesaran, M. Hashem, and Ron P. Smith. 1995. "Estimating Long-Run Relationships from Dynamic Heterogeneous Panels." *Journal of Econometrics* 68 (1): 79–113.

[Thakor and Whited 2011] Thakor, Anjan V., and Toni M. Whited. 2011. "Shareholder–Manager Disagreement and Corporate Investment." *Journal of Financial Economics* 101 (2): 274–93.

[Whited 1992] Whited, Toni M. 1992. "Debt, Liquidity Constraints, and Corporate Investment: Evidence from Panel Data." *Journal of Finance* 47 (4): 1425–60.