Nonlinearities and Heterogeneity in Firms Response to Aggregate Fluctuations: What Can We Learn From Machine Learning?

Marco Errico — Bank of Italy

Simone Pesce — Boston College

Luigi Pollio* — UMBC

This Draft: February 9, 2025
First Draft: June 27, 2024

Abstract

Firms respond heterogeneously to aggregate fluctuations, yet standard linear models impose restrictive assumptions on firm sensitivities. Applying the Generalized Random Forest to U.S. firm-level data, we document strong nonlinearities in how firm characteristics shape responses to macroeconomic shocks. We show that nonlinearities significantly lower aggregate responses, leading linear models to overestimate the economy's sensitivity to shocks by up to 1.7 percentage points. We also find that larger firms, which carry disproportionate economic weight, exhibit lower sensitivities, leading to a median reduction in aggregate economic sensitivity of 52%. Our results highlight the importance of accounting for nonlinearities and firm heterogeneity when analyzing macroeconomic fluctuations and the transmission of aggregate shocks.

JEL Codes: D22, E32, C14, E5

Keywords: Firm Heterogeneity, Firm Sensitivity, Machine Learning, Aggregate Fluctuations, Monetary Policy, Business Cycle, Uncertainty, Oil Shock.

*Contact: Luigi Pollio, lpollio1@umbc.edu; Marco Errico, marco.errico@esterni.bancaditalia.it; Simone Pesce, simone.pesce@bc.edu.

We thank Michele Lenza and all the participants at the 2024 CFE-CMStatistics Conference, 2024 SNDE-IIF Symposium, Bank of Italy, and ECB for their comments and suggestions. Part of the research was conducting while Simone Pesce was visiting the European Central Bank. The views expressed here do not necessarily reflect those of the Bank of Italy or the Euro-System.

1 Introduction

Firms do not respond uniformly to aggregate fluctuations and shocks. Some adjust sharply to changes in GDP growth and interest rates, while others remain largely unaffected. Studying the cross-sectional heterogeneity in firm sensitivity to aggregate fluctuations and its underlying drivers provides insights into the dynamics of aggregate outcomes across different phases of the economic cycle (Cooley and Quadrini, 2006; Buera and Moll, 2015). Prior research suggests that firm responses to aggregate shocks depend linearly on their underlying characteristics, such as size and default risk (Gertler and Gilchrist, 1994; Ottonello and Winberry, 2018). However, more recent evidence indicates that these relationships may be nonlinear (Crouzet and Mehrotra, 2020; Paranhos, 2024). The extent to which nonlinearities in the relationship between firm sensitivities and their underlying characteristics matters for both firm and macroeconomic outcomes remains an open question.

In this paper, we study the role of nonlinearities and heterogeneity in firm response to aggregate fluctuations using a nonparametric machine learning approach. While a heterogeneour linear panel regression model can capture systematic heterogeneity–such as differences in sensitivity based on firm characteristics such as size, leverage, and industry–it imposes linearity in how characteristics impact firm sensitivity, ruling out nonlinearities and complex interactions among characteristics. By using a machine learning approach, we can estimate how firms respond to aggregate fluctuations as a function of a large set of characteristics without imposing restrictive parametric assumptions on the underling mapping. These nonlinearities may be crucial for understanding firm-level heterogeneity in sensitivities, offering new insights into the transmission mechanisms of aggregate shocks and their macroeconomic implications.

We employ the Generalized Random Forest (GRF, henceforth) model by Athey et al. (2019) to analyze how U.S. firms respond to aggregate fluctuations. Using firm-level quarterly Compustat data spanning from 1990 to 2019, we estimate the firm-level responses of key firm outcomes: sales, investment, debt issuance, and market value, as functions of balance-sheet characteristics and across multiple sources of aggregate fluctuations. We focus on four key sources of aggregate fluctuations that are extensively studied in the literature: business cycle fluctuations (Crouzet and Mehrotra, 2020); and three major exogenous shocks: monetary policy shocks (Bauer and Swanson, 2023), uncertainty shocks (Jurado et al., 2015), and oil price shocks (Känzig, 2021). Finally, we model firms' sensitivity to aggregate fluctuations using a set of financial and non-financial characteristics that are widely examined in prior research, including leverage, liquidity, distance to default, share of short-term debt, size, return on assets, sales volatility, and industry scope.

We provide evidence of strong nonlinearities in how balance-sheet characteristics influence conditional firm sensitivities across all outcome variable-aggregate shock pairs we study. While the average firm sensitivities are statistically identical between GRF and the linear panel model (LPM, henceforth), we find substantial differences in higher-order moments. The standard deviation of firm sensitivities estimated by the GRF model is 50% lower than that of the LPM, while kurtosis (excess tail risk) in GRF is 20% lower. This suggests that the nonlinear model captures more moderate and accurate patterns of heterogeneity, whereas the LPM misspecifies the distribution of heterogeneity across firms despite providing a reasonable first-order approximation. Using machine learning tools such as accumulated local effects, we qualitatively show that the marginal effect of each balance-sheet characteristic on firm-evel sensitivities is not constant but exhibits kinks, U-shaped, or inverted U-shaped patterns, ultimately rejecting the linearity assumption embedded in the LPM. Additionally, using Friedman's H-statistic, we find that between 10% and 40% of the total effect of each characteristic on firm outcomes is mediated through its interaction with other characteristics, with firm size playing a particularly prominent role.

We evaluate the role of individual firm characteristics in shaping firms' sensitivities to aggregate shocks and their heterogeneity. One advantage of machine learning approaches is their ability to mitigate the curse of dimensionality, allowing us to analyze a large set of characteristics while automatically detecting their relative importance. Using the absolute mean Shapley value of each characteristic, we assess its quantitative importance for the firm responses to aggregate shocks. We find that size is the dominant factor in explaining firms' sensitivity to aggregate shocks. However, more broadly, no single characteristic overwhelmingly dominates, reinforcing the importance of incorporating multiple characteristics to explain firm-level sensitivities. We also assess the contribution of each characteristic to heterogeneity in firms' sensitivities by measuring the depth-weighted frequency of splits where the characteristic is used. Our results indicate that heterogeneity is driven by multiple characteristics and that the ranking of characteristics varies significantly across aggregate shock-outcome variable pairs. For example, firm size, along with other non-financial characteristics such as industry scope, plays a dominant role in explaining firms' sensitivity to business cycle fluctuations and uncertainty shocks, whereas default risk and other financial characteristics are significantly more important in shaping heterogeneity in responses to monetary policy shocks.

Motivated by such evidence, we pursue two additional questions. First, do nonlinearities in firm sensitivities matter at the aggregate level? While we document evidence of nonlinearities at the firm level, they may not be quantitatively relevant at the aggregate level. Second, how much does firm heterogeneity influence aggregate responses? Given the highly unequal distribution of firm weights in the economy, the sensitivity of larger firms disproportionately shape macroeconomic outcomes.

We begin by proposing a theory of aggregation that links firm-level responses to macroeconomic outcomes by weighting firms' sensitivities to shocks according to their contribution in the economy. The aggregate response to macroeconomic shocks depends not only on individual firm reactions but also on their relative size in the economy. In our framework, aggregate fluctuations are driven by two components: the average firm-level response to a given shock, and the covariance between firm sensitivities and their economic weight. The first captures how firms, on average, react to shocks, while the second reflects whether firms with greater economic importance exhibit systematically different sensitivities. A positive covariance implies that more sensitive firms hold greater weight, amplifying aggregate fluctuations, whereas a negative covariance suggests that less sensitive firms dominate, dampening macroeconomic volatility. We use this decomposition to quantify the role of firm heterogeneity in shaping aggregate outcomes.

We show that models ignoring nonlinearities tend to overestimate the economy's sensitivity to business cycle fluctuations and shocks. By comparing aggregate responses estimated using the GRF with those from a standard LPM, we quantify the macroeconomic impact of nonlinearities. In particular, the GRF model consistently yields lower aggregate response estimates, causing linear models to overstate the economy's responsiveness to aggregate shocks. The quantitative discrepancies are substantial. For example, while a LPM predicts that a 1% increase in GDP leads to a 2.4% rise in aggregate sales and a 5.4% increase in stock market value within a year, accounting for nonlinearities reduces these estimates by approximately 0.3 and 0.2 percentage points, respectively. A similar pattern holds for contractionary monetary policy shocks: the GRF model predicts a 6.6% smaller drop in stock market prices and a more subdued response of aggregate investment. The primary driver of these differences is the covariance term, as the LPM consistently estimates larger covariance effects than GRF. This suggests that incorporating nonlinearities weakens the relationship between firms' sensitivities and their economic weight, ultimately dampening aggregate fluctuations.

Lastly, we find that heterogeneity in firm sensitivity dampens business cycle fluctuations and the aggregate response to shocks. We quantify this effect by measuring the contribution of the covariance term to the overall aggregate response. Our results show that larger firms, which tend to have lower absolute sensitivities, systematically reduce the impact of shocks. Specifically, we estimate that their presence lowers the aggregate response of sales by 6% and investment by 53% to business cycle fluctuations while amplifying the stock market response by approximately 24%. A similar pattern emerges for uncertainty and monetary policy shocks–where, despite a strongly negative unweighted average firm response, the aggregate effect is muted due to lower sensitivities among firms with greater economic weight. This dampening is particularly strong for investment, where the covariance term fully offsets the average firm response to these shocks. These findings highlight that aggregate fluctuations are shaped not just by the average firm response but also by the interaction between firm sensitivities and their economic relevance, underscoring the importance of accounting for firm heterogeneity in macroeconomic analysis.

We further explore the quantitative role of heterogeneity in firm sensitivities across several dimensions. Using a rolling window regression framework, we show that the decomposition between mean and covariance terms remains stable over time, suggesting no significant composition effects over the past three decades. We then decompose the role of heterogeneity into within- and across-industry margins, finding that both contribute equally to dampening aggregate fluctuations. Finally, we examine the relative importance of financial and non-financial firm characteristics by constructing counterfactual scenarios that hold one type of heterogeneity constant while allowing the other to vary. We find that abstracting from heterogeneity in non-financial characteristics leads to larger and statistically significant deviations from the benchmark aggregate response compared to shutting down heterogeneity in financial characteristics. This suggests a stronger covariance between firms with large economic weight and variations in non-financial characteristics, further emphasizing their role in shaping aggregate dynamics.

The remainder of the paper is organized as follows. Section 2 provides information on the methodologies we use, Linear Panel Model and random forest based on GRF, and on the Montecarlo exercise. Section 3 presents the data used in the empirical application and the key results on firm-level sensitivities and their drivers. Section 4 proposes an aggregation theory and presents the results on the aggregate implications. Section 5 concludes.

Literature. This paper contributes to the rapidly growing literature applying machine learning techniques to economic analysis. Machine learning offers advantages both for the estimation of conditional average treatment effects and causal inference in high-dimensional settings (Athey and Imbens, 2017; Varian, 2014) and for predicting outcomes to improve targeting and forecasting (Mullainathan and Spiess, 2017). Our work relates to the estimation of conditional average treatment effects using machine learning; however, few studies have applied machine learning techniques to examine firm-level heterogeneous sensitivity to aggregate shocks and macroeconomic fluctuations more broadly.[1] [2] [3] The closet study to our work is Paranhos (2024), which examines the relationship between firms' default risk and the effectiveness of monetary policy transmission to investment

[1]Estimating conditional average treatment effects using machine learning is more common on the consumers and household side rather than on the firm side. For instance, Belloni et al. (2017) estimates the effect of 401(k) eligibility and participation on accumulated assets using local quantile treatment effects. Khazra (2021) explores the heterogeneity of house price elasticity of consumption using micro panel data via GRF (Athey et al., 2019), finding that neglecting local heterogeneities in elasticity leads to overestimating the total consumption response during housing market booms and busts.

[2]The forecasting advantages of machine learning have been explored in macroeconomics in relation to inflation forecast, with Paranhos (2021) and Nakamura (2005) both using neural networks to predict future inflation.

[3]Machine learning is more widely used in finance and asset prices; for instance, Freyberger et al. (2020) and Gu et al. (2020) use machine learning techniques to predict stock market returns and asset risk premiums, respectively, accounting for non-linearities and many characteristics.

decisions, by generalizing standard local projection methods nonparametrically. Differently from their work, we apply random forest models to study firm heterogeneity in sensitivity to multiple aggregate shocks and multiple firm outcomes, incorporating a high-dimensional firm characteristic space. Our findings highlight the strong quantitative role of interactions among characteristics in shaping firm-level sensitivities, indicating that multiple characteristics jointly drive heterogeneity in firm-level outcomes.

Our work also contributes to the literature that studies the heterogeneity in firm-level sensitivity to aggregate shocks and its determinants. A non-exhaustive list of important contributions includes Ottonello and Winberry (2018), Jeenas (2018), Gertler and Gilchrist (1994) and Jungherr et al. (2024), that study the roles of leverage and distance to default, liquidity, size, and debt maturity for the response of firm investment to monetary policy shocks, respectively; Gürkaynak et al. (2022), who investigate how liquidity and leverage influence the response of market value; Gertler and Gilchrist (1994) and Crouzet and Mehrotra (2020) examine respectively the role of size and industry scope for the response of sales and investments to monetary policy and business cycle fluctuations; Covas and Haan (2011) and Begenau and Salomao (2019) study debt issuance by firm size over the business cycle. Prior research typically examines a single balance-sheet characteristic at a time and imposes linearity in how firm characteristics influence responses to aggregate shocks. In contrast, we depart from the standard linear panel regression approach and apply machine learning methods to incorporate a large set of firm characteristics simultaneously.[4] We show that the heterogeneity in firm sensitivities is highly non-linear with strong interactions among characteristics, underscoring the importance of a comprehensive analysis with a high dimensional characteristic space.

Moreover, the macroeconomic literature has devoted limited attention to heterogeneity in firm-level sensitivity to uncertainty and oil shocks, despite these being key drivers of macroeconomic fluctuations (Christiano et al., 2014; Känzig, 2021). For instance, Alfaro et al. (2024) shows that aggregate financial frictions amplify the negative effects of uncertainty shocks on investment, sales, and debt issuance, while Kumar et al. (2023) finds that the impact of uncertainty on sales and investment depends on firm size, using an RTC design. In the case of oil shocks, Narayan and Sharma (2011) and Tsai (2015) study how a firm's market value reacts to oil shocks depending on its size and industry scope. We expand this literature by providing novel evidence on the heterogeneity in firm-level sensitivity to uncertainty and oil shocks, its determinants, and its aggregate impact.

Lastly, our paper contributes to the growing literature on the macroeconomic implications of firm-level heterogeneity–including differences in size, leverage, industrial sector, and debt maturity structure–for aggregate fluctuations and the transmission of shocks. Prominent contributions in this area include Cooley and Quadrini (2001), Cooley and Quadrini (2006), Buera and Moll (2015),

[4] For data limitation, we do not consider additional firm characteristics such as paying dividends (Farre-Mensa and Ljungqvist, 2016) or firm age (Cloyne et al., 2018).

Crouzet (2018), Ottonello and Winberry (2018), Deng and Fang (2022), and Krusell et al. (2023), among others. Unlike most earlier contributions, which rely on quantitative macroeconomic models, we develop an aggregation framework in the spirit of Crouzet and Mehrotra (2020) and leverage the estimated distribution of firm-level sensitivities from the random forest model to decompose the aggregate effect into a mean and a covariance term. [5] Our findings on the dampening effects of firm-level heterogeneity on the transmission of shocks to macroeconomic aggregates provide a more general perspective than previous studies, demonstrating that firm-level heterogeneity systematically weakens the aggregate response to macroeconomic fluctuations. Moreover, differently from previous studies, we assess the quantitative role of the heterogeneity driven by the non-linear relationships between firm sensitivity and underlying characteristics.

2 Methodology

Our objective is to analyze how firms' balance-sheet characteristics affect the sensitivity of their outcomes to aggregate fluctuations and the heterogeneity of these responses. We employ the Generalized Random Forest algorithm, introduced by Athey et al. (2019), to estimate heterogeneous firm-level responses and compare its performance to that of a standard linear panel regression model. To assess their quantitative accuracy, we conduct Monte Carlo simulations under various data-generating process scenarios.

2.1 Linear Panel Regression

Consider an empirical setting where we observe the outcome variables and characteristics of a set of firms, indexed by i, over multiple consecutive periods, indexed by t. The outcome variable of interest, $Y_{i,t}$, represents firm-level performance measures such as sales growth, investment, or other key indicators. Let W_t denote the treatment effect or a source of aggregate fluctuation that is common to all firms. Firm-level characteristics, $X_{i,t-1}$, can influence the sensitivity of $Y_{i,t}$ to W_t.

To estimate the heterogeneous response of firms' outcomes $(Y_{i,t})$ to an aggregate shock (W_t), conditional on a set of firm-level characteristics $(X_{i,t-1})$ we estimate the following heterogeneous linear panel regression using OLS:

$$Y_{i,t} = \alpha + \beta_0 \cdot W_t + \beta_1 \cdot X_{i,t-1} + \beta_2' \left(W_t \cdot X_{i,t-1}\right) + \epsilon_{i,t}, \tag{1}$$

[5] Chang et al. (2024a), Chang et al. (2024b), and Lenza and Savoia (2024) offer an alternative approach based on functional VARs and heterogeneous VARs, which integrate aggregate variables with cross-sectional distributions to study their dynamic interactions. In contrast, our approach uses machine learning techniques to estimate firm-level sensitivities to aggregate shocks, which we then aggregate in a bottom-up framework to assess macroeconomic implications.

where $\epsilon_{i,t}$ is an i.i.d. error term, and firm characteristics are predetermined at $t-1$. The parameter vector of interest, β_2, captures how firms' sensitivity to aggregate shocks varies with their characteristics. The marginal effect of the aggregate shock W_t on firm outcomes is given by $\beta_0 + \beta_2' X_{i,t-1}$, which depends linearly on $X_{i,t-1}$. Equation (1) provides a standard econometric framework for estimating heterogeneity in firms' responsiveness to aggregate shocks.

2.2 A Brief Description of a Generalized Random Forest

Machine learning methods provide a flexible approach to estimating heterogeneous sensitivities, allowing for a potentially complex, high-dimensional characteristic space and non-linear relationships in the marginal effects. Specifically, the GRF algorithm developed by Athey et al. (2019), enables the nonparametric estimation of the following model:

$$Y_{i,t} \;=\; b\big(X_{i,t-1}\big)\,\cdot W_t \;+\; \varepsilon_{i,t} \quad , \quad \beta(x) \;=\; \mathbb{E}\big[b\big(X_{i,t-1}\big)\,\big|\,X_{i,t-1}=x\big], \tag{2}$$

where $\epsilon_{i,t}$ is an i.i.d. error term, b is a flexible function of firms' characteristics, and $\beta(x)$ is the average conditional effect of the aggregate shock W_t on the outcome $Y_{i,t}$ for firms with characteristics equal to x. The latter, which is our object of interest, is determined as follows in the GRF algorithm:

$$\hat{\beta}(x) = \frac{\sum_{i=1}^{n} \alpha_i(x)\left(W_i - \bar{W}_\alpha\right)\left(Y_i - \bar{Y}_\alpha\right)}{\sum_{i=1}^{n} \alpha_i(x)\left(W_i - \bar{W}_\alpha\right)}, \tag{3}$$

where, $\alpha_i(x)$ is a weight determined by the causal forest, $\bar{W}_\alpha = \sum_{i=1}^{n} \alpha_i(x)\,W_i$ is a weighted average of the shock, and $\bar{Y}_\alpha = \sum_{i=1}^{n} \alpha_i(x)\,Y_i$ is a weighted average of the outcome.

The GRF algorithm estimates $\beta(x)$ in two steps: first, it constructs a forest of decision trees designed to partition the data in a way that maximizes heterogeneity in firms' responses to aggregate shocks, and second, it estimates the conditional average treatment effect (CATE) using a locally weighted regression approach. In the first stage, GRF builds a collection of honest and adaptive decision trees that recursively split the data based on firm characteristics $X_{i,t-1}$. Unlike standard regression trees, which minimize prediction errors, GRF partitions the sample to maximize heterogeneity in firms' sensitivity to the aggregate shock W_t. The algorithm is considered "honest" because it uses one subsample to determine optimal splits and a separate subsample to estimate treatment effects within each leaf, thereby mitigating overfitting. Each tree is constructed by selecting a random subsample of the data, and splits are determined by optimizing a criterion that prioritizes variation in the estimated treatment effects rather than differences in outcome levels alone. In other words, the algorithm selects splits by maximizing the expected heterogeneity in treatment effects across partitions, typically based on the variance of W_t within candidate splits. Once the forest is grown, it provides a data-driven partitioning of the firm characteristic space, grouping together firms that exhibit similar estimated sensitivity to aggregate fluctuations.

In the second stage, GRF estimates the heterogeneous treatment effect $\beta(x)$ by aggregating information across trees. For a given firm with characteristics $X_{i,t-1} = x$, the algorithm identifies neighboring firms that frequently appear in the same leaves across multiple trees. The estimated treatment effect is then computed using a local linear regression, where each observation is assigned a weight $\alpha_i(x)$ based on how often it appears in the same terminal node as the hypothetical firm x. These weights, which are determined by the structure of the causal forest, ensure that $\hat{\beta}(x)$ is locally smoothed and not overly sensitive to a single partition. Finally, using these weights, GRF estimates $\beta(x)$ via a weighted regression of firm outcomes $Y_{i,t}$ on aggregate shocks W_t, ensuring that identification relies on variation in W_t within locally homogeneous subgroups. The algorithm further regularizes estimation by tuning the minimum leaf size, selecting the optimal number of trees, and controlling for variance in the estimated treatment effects.

Advantages of GRF. The GRF algorithm in Equation (2) offers key advantages over the standard linear panel regression in Equation (1), making it particularly well-suited for estimating the heterogeneous effects of aggregate shocks. The GRF agnostic approach to the function $b\big(X_{i,t-1}\big)$ allows it to account for non-linear, flexible relationships in the marginal effects of shocks, accommodating a complex, high-dimensional balance-sheet characteristic space. The linear panel regression model assumes that firms' characteristics linearly influence the heterogeneity in their sensitivity to aggregate shocks. However, this linearity assumption may be restrictive and could lead to misspecification if the nonlinear component of heterogeneity is significant. While the linear panel regression model can incorporate more complex forms of heterogeneity by including polynomial terms in the firm-level characteristics, it remains a parametric approach that requires the econometrician to take a stance of the unknown forms of non-linearities, making the LPM vulnerable to errors from model misspecification. In contrast, GRF explores the covariate space non-parametrically, adaptively detecting intricate relationships without requiring a pre-specified form. Moreover, GRF can efficiently handle a high-dimensional characteristics space, automatically putting more weight on the most important covariates. This feature of the GRF mitigates the curse of dimensionality inherent in models with large sets of covariates and interactions. Enumerating all possible pairwise (or higher-order) interactions in a linear model quickly leads to over-parameterization and multicollinearity, while GRF adaptively partitions the data, freeing the researcher from having to manually specify functional forms or interactions. Thus, the linearity and parametric features of the linear panel regression model become more restrictive in the presence of high-dimensional characteristics space.[6]

[6]However, GRF's flexibility comes with trade-offs, such as the potential loss of precision in smaller samples and reliance on careful hyperparameter tuning. When the true relationship between covariates and the conditional effect of shocks is linear–or can be sufficiently well captured by a modest set of polynomial terms–a linear panel regression model may perform comparably to GRF. We illustrate the relative performance of

2.3 Monte Carlo simulation

We conduct a Monte Carlo simulation to compare the precision of the GRF and a linear panel regression in estimating heterogeneous responses to aggregate fluctuations. We assume several underlying data-generating processes, incorporating both linear and nonlinear relationships in the conditional effects, with multiple covariates driving the heterogeneity. An econometrician seeking to understand how firms respond to aggregate shocks as a function of their balance-sheet characteristics does not observe the true data-generating process. Instead, they estimate the conditional effects using either a linear panel regression model, as specified in Equation (1), or the GRF algorithm, as described in Equation (2).

Data generating process We generate synthetic data to replicate the econometric setting used in the empirical application studies below. We assume that the simulated economy consists of 6000 firms, indexed by i, over $T = 20$ periods. We denote with $X^j_{i,t}$ denote the j-th characteristic of firm i at time t, where $j = 1, \ldots, 6$. Each covariate follows an independent autoregressive process with a persistence of 0.9, and shocks drawn from a standard normal distribution with mean zero and unit variance. We assume a relatively high value of persistence to be consistent with the balance sheet characteristics in the empirical application. The aggregate shock, W_t, is also drawn from a standard normal distribution. We assume that the outcome variable for firm i at time t, $Y_{i,t}$, depends on the firm's characteristics and the aggregate shock according to the following specification:

$$Y_{i,t} = W_t + \sum_{j=1}^{J} X^j_{i,t} + F\Big(\{X^j_{i,t}\}_{j=1}^{J'}\Big) \cdot W_t + \varepsilon_{i,t}, \qquad \varepsilon_{i,t} \sim N(0,1), \tag{4}$$

where $\varepsilon_{i,t}$ is an independent and identically distributed noise term drawn from a normal distribution with mean zero and variance normalized to one. The aggregate shock, W_t, propagates to $Y_{i,t}$ differently across firms, depending on a subset of firm characteristics, $\{X^j_{i,t}\}_{j=1}^{J'}$. The function $F\Big(\{X^j_{i,t}\}_{j=1}^{J}\Big)$ governs the heterogeneity in firms' responses to aggregate fluctuations. Without loss of generality, we model heterogeneity as a function of the contemporaneous realization of $X^j_{i,t}$, given that W_t is independently drawn by construction and the covariates evolve solely based on their own history.

We consider three scenarios for the function F to evaluate the performance of a linear panel regression and the GRF under different data-generating processes: (i) linear, (ii) non-linear, and (iii) threshold-based. The corresponding data-generating processes are specified as follows:

the models in the Monte Carlo simulation exercise below.

i. Linear:

$$F\left(\{X_{i,t}^{j}\}_{j=1}^{J'}\right) \;=\; \sum_{j=1}^{J'} X_{i,t}^{j}.$$

ii. Non-linear:

$$F\left(\{X_{i,t}^{j}\}_{j=1}^{J'}\right) = \begin{cases} \sum_{j=1}^{J'} X_{i,t}^{j} + \alpha_1 \sum_{j=1}^{J'} X_{i,t}^{j,2}, & \text{Quadratic} \\ \sum_{j=1}^{J'} X_{i,t}^{j} + \alpha_1 \sum_{j=1}^{J'} \sum_{k=j+1}^{J'} X_{i,t}^{j} \cdot X_{i,t}^{k}. & \text{Interactions} \end{cases}$$

iii. Threshold-based:

$$F\left(\{X_{i,t}^{j}\}_{j=1}^{J'}\right) = \sum_{j=1}^{J'} \left(\alpha_1 \mathbb{1}_{X_{i,t}^{j}>0} + \alpha_2 \mathbb{1}_{X_{i,t}^{j}\leq 0}\right) \cdot X_{i,t}^{j}.$$

The heterogeneous effect is estimated using both a linear panel regression and the GRF. To assess the models' ability to recover the true heterogeneity as the dimensionality of the characteristic space increases, we vary the number of covariates relevant for the heterogeneity, J', up to six. The models are evaluated using standard statistical metrics, including Root Mean Squared Error (RMSE), average bias, and explained heterogeneity.[7]

Results. Table 1 shows that GRF consistently outperforms a linear panel regression in capturing heterogeneous sensitivities, particularly when the heterogeneity is nonlinear or exhibits complex patterns. In cases of linear heterogeneity, GRF performs similarly to a correctly specified OLS, provided that the number of characteristics driving the heterogeneity is small. However, as the dimensionality of the heterogeneity increases, GRF's precision declines due to the limitations imposed by the dataset's size. Figure 11 in Appendix C compares the relationship between predicted and true sensitivities for a data-generating process with a single variable driving the heterogeneity. The results show that the loss of precision of the GRF is primarily concentrated in the tails of the distribution, where data are sparse. The advantages of GRF are especially pronounced in nonlinear data-generating processes, where a linear panel regression is misspecified and fails to fully capture

[7]The average bias of an estimator is defined as the expected deviation of the estimator from the true parameter value, averaged over multiple simulation runs. Explained heterogeneity is measured as the ratio of the variance of the predicted treatment effects to the variance of the true treatment effects. A value close to one indicates that the model effectively captures the variability in the true treatment effect, while a value near zero suggests poor performance in identifying heterogeneity. Values greater than one may signal overfitting, where the model captures noise rather than the underlying structure.

Table 1: Results Monte Carlo simulation

	Linear Panel Model			GRF		
DGP of Heterogeneity	Bias	RMSE	Explained	Bias	RMSE	Explained
Panel A: Variables relevant for heterogeneity J′ = 1						
Linear	0.01	0.01	1.00	0.01	0.10	1.00
Quadratic	0.18	6.84	0.00	0.01	0.48	0.98
Threshold	0.02	0.66	0.73	0.01	0.08	0.99
Panel B: Variables relevant for heterogeneity J′ = 3						
Linear	0.01	0.01	1.00	0.01	0.43	0.91
Quadratic	0.55	12.10	0.00	0.03	2.40	0.88
Interaction	0.05	8.33	0.00	0.02	1.58	0.82
Threshold	0.05	1.57	0.85	0.02	0.87	0.86
Panel C: Variables relevant for heterogeneity J′ = 6						
Linear	0.01	0.01	1.00	0.03	1.55	0.64
Quadratic	1.15	17.70	0.00	0.16	6.05	0.70
Interaction	0.13	18.78	0.00	0.08	8.88	0.35
Threshold	0.14	1.74	0.90	0.02	1.41	0.73

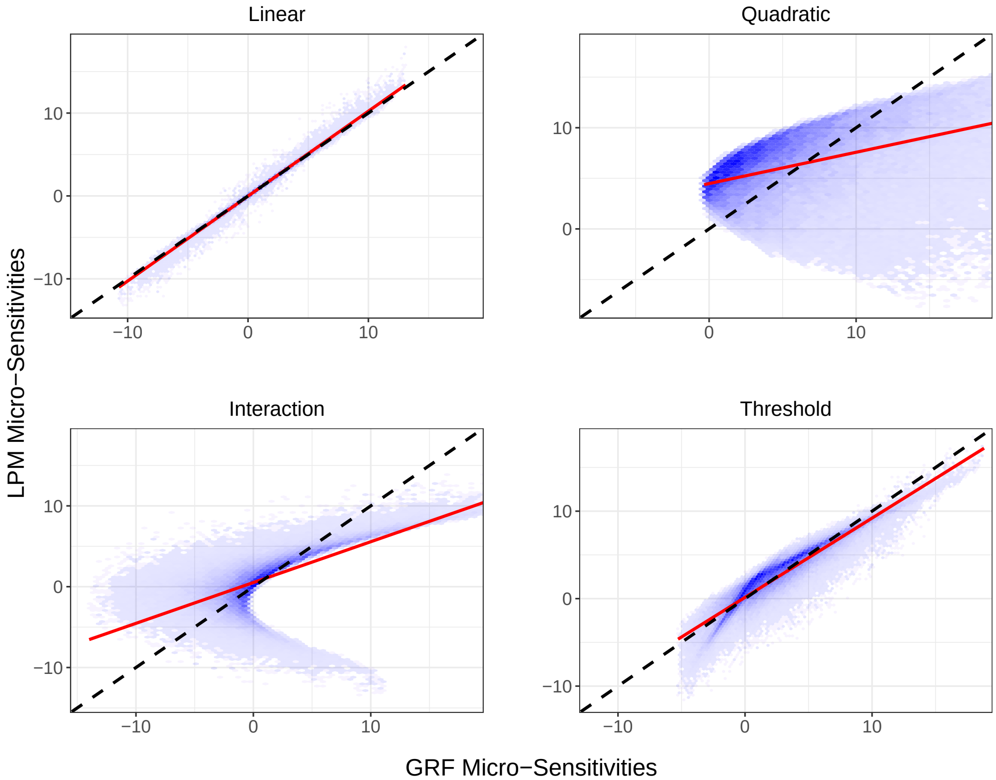
Notes: The table compares the performance of a linear panel regression and the GRF in estimating heterogeneous sensitivities across different data-generating processes and levels of heterogeneity complexity. The evaluation metrics include absolute average bias (Bias), root mean squared error (RMSE), and the proportion of variance in true heterogeneity explained by each model (Explained). Panel A, B, and C report results for a setting where only one, three, and six characteristics drives heterogeneity, respectively. We assume $\alpha_1 = 0.5$ for quadratic and interaction heterogeneity, and set α_1 and α_2 of 0.5 and 1.5 for threshold-based heterogeneity. Results are averaged over 10 simulations of a panel comprising 6,000 firms observed over 20 periods.

heterogeneity. Although both methods experience some loss of precision as the number of covariates, J', increases – reflected in higher RMSE and lower explained variance – GRF remains more robust in high-dimensional settings, effectively capturing more intricate patterns of heterogeneity.

A direct comparison between sensitivities estimated by a linear panel regression and GRF provides a useful diagnostic tool to detect misspecification due to nonlinear heterogeneity in the data.[8] Figure 1 compares the sensitivities estimated by both methods in a Monte Carlo simulation with three sources of heterogeneity ($J' = 3$). When the true data-generating process is linear, the estimates from both models align closely along the 45-degree line. However, in scenarios with nonlinear or threshold-based heterogeneity, the correlation between the two weakens, and the distribution of predicted sensitivities becomes more dispersed. Depending on the scenario, the differences between the sensitivities estimated by the two methods can be as large as 100% or even exhibit opposite signs, underscoring the strong misspecification bias introduced by imposing

[8]This is not a formal test but rather a graphical check that suggests the presence of unmodeled nonlinear heterogeneity in the estimated model.

Figure 1: Comparison of sensitivities on simulated data



Notes: The figure presents predicted firm-level sensitivities from a Monte Carlo simulation across four scenarios: linear (top left), nonlinear quadratic (top right), nonlinear interaction (bottom left), and threshold-based (bottom right). The x-axis represents firm-level sensitivity estimates from the GRF, while the y-axis shows estimates from the linear panel regression. Each point corresponds to a simulated firm-time observation. The black dashed line represents the 45-degree reference line, while the red solid line depicts the fitted linear regression. Results are based on a single simulation of a panel with 6,000 firms observed over 20 periods. We assume a $\alpha_1 = 0.5$ for quadratic and interaction heterogeneity, and set α_1 and α_2 of 0.5 and 1.5 for threshold-based heterogeneity. The data-generating process assumes that three characteristics ($J' = 3$) drive heterogeneity.

linearity in firm-level sensitivities.

3 Application to U.S. Firms

We apply the GRF algorithm to examine how firm outcomes respond to aggregate fluctuations based on a high-dimensional set of observed balance-sheet characteristics, using firm-level data from the U.S. over the period 1990–2019. We begin by describing the data, followed by an analysis of the estimated firm-level sensitivities and a comparison with those obtained from a linear panel regression.

3.1 Data and measurement

Our primary data source is the quarterly Compustat dataset, which provides comprehensive financial statement information for publicly listed companies in the U.S. We merge firm-level data with a set of aggregate variables and shocks commonly used in the literature. The final dataset includes 220,259 firm-quarter observations spanning from 1990 Q1 to 2019 Q4. These dates align with those of the aggregate variables in the panel, excluding the COVID-19 period. Additionally, all variables are deflated using the implied price index of gross value added in the U.S. non-farm business sector. Below, we provide a brief overview of the primary firm-level balance-sheet variables and the measurement of aggregate variables. Additional details on variable construction and data cleaning are provided in Appendix B.

Firm-level data. Our empirical analysis utilizes two sets of firm-level variables. The goal is to estimate the heterogeneous sensitivity of four firm outcome variables: annual real sales growth, debt issuance (measured by the one-year percentage change in short- and long-term debt), market value growth, and the investment rate (measured as the one-year percentage change in capital stock using the perpetual inventory method). The second set consists of eight firm-level balance-sheet characteristics, which we categorize into two groups: financial and non-financial variables. Non-financial characteristics include firm size (measured by the logarithm of total assets), industry scope (captured by NAICS 5-digit industry codes), ten-years sales volatility, and firm profitability (measured by return on assets, ROA). Financial characteristics include the liquidity ratio (cash-to-total assets), leverage ratio (total debt-to-total assets), distance to default (Merton, 1974)), and debt liquidity (measured by the proportion of short-term debt to total debt). These firm-level balance-sheet characteristics have been widely used in the literature to study the heterogeneity in the transmission of aggregate fluctuations onto firm outcomes.[9] Appendix B presents selected summary statistics and histograms of all firm-level variables used in the empirical analysis. Notably, Table 5 in Appendix B reports the pairwise correlation between all independent variables, showing that, although some correlation exists among firm-level characteristics, they provide distinct information along different dimensions.

[9]For instance, Ottonello and Winberry (2018), Cloyne et al. (2018), and Jeenas (2018) study the role that distance to default, leverage and liquidity play in the transmission of monetary policy shocks to investment, respectively. Similarly, Alfaro et al. (2024) studies the effects of uncertainty on firms' financial variables such as liquidity and leverage, while Crouzet and Mehrotra (2020) focuses on how size and industry scope impact the response to business cycle fluctuations.

Aggregate fluctuations. We investigate the sensitivity of firm outcomes to the following aggregate fluctuations: business cycles, macroeconomic uncertainty shocks, monetary policy shocks, and oil price shocks.[10]. Business cycle fluctuations are proxied by the annual percentage change in real GDP following Crouzet and Mehrotra (2020) Monetary policy shocks are measured using interest rate surprises around Federal Reserve announcements, identified using high-frequency variations in the 3-month federal funds rate futures, and cleaned of past aggregate fluctuations as in Bauer and Swanson (2023). Uncertainty shocks are exogenous change in macroeconomic uncertainty, as measured in Jurado et al. (2015). Oil price shocks are proxied with high-frequency changes in oil supply expectations around OPEC announcements from Känzig (2021). To normalize the size of the shocks, we use them as instruments for a set of endogenous variables. Using the exogenous variables as instrument imposes a unit effect normalization of the shocks in terms of a one-unit change in the endogenous variable (Stock and Watson, 2018). Specifically, we use the one-year percentage change in the one-year government bond yield for monetary policy shocks, the one-year change in the oil price index for oil price shocks, and the volatility index for uncertainty shocks. Figure 9 in Appendix B presents the time series of the aggregate fluctuations used in the paper.

Estimation details. We estimate the sensitivity of four firm outcome variables to each aggregate shock, conditional on eight firm-level balance-sheet characteristics, considering a total of 16 scenarios. For each outcome variable-aggregate shock pair, we estimate the GRF model in Equation (2) and the LPM in Equation (1) via OLS. Since the outcome variables are constructed as one-year percentage changes, we lag the balance-sheet characteristics by four periods in the empirical application.

In the GRF model, we set the number of trees in the forest to 2,000, with equal weighting. We use honest splitting for sub-sample partitioning, allocating 50% of the data to build each tree and ensuring a minimum of five observations per tree leaf. Observations are clustered at the firm level with equal weight, so firms with more observations receive greater weight, thereby reducing the influence of entry and exit. Splitting is allowed across all characteristics, with the tuning parameter controlling the maximum imbalance of a split set at 0.05. For the LPM, we estimate the equation using OLS. We include the interaction between the aggregate shock, W_t , and industry scope, while absorbing the level of industry scope to reduce computational burden.

In both models, we do not include time fixed effects, as our objective is to estimate the average unconditional effects of aggregate shocks on firm outcomes.[11] We also omit firm fixed effects because

[10] We examine oil price shocks for two reasons. First, they provide a clear and distinct example of exogenous inflation changes driven by supply factors. Second, oil price shocks have gained increasing importance in the macroeconomic literature, particularly following the Covid-19 pandemic.

[11] If macroeconomic confounding factors are a concern, macroeconomic variables can be partialled out before estimation.

our primary interest is in assessing the role of industry scope – which is constant at the firm level – in driving and explaining the response of outcome variables and their heterogeneity across firms. In a LPM, demeaning variables at the firm level neutralizes all variation in industry scope, allowing for the estimation of its heterogeneous effect but not its average effect. However, demeaning is not feasible in the GRF model, as the algorithm operates on variables in levels. Thus, we opt to include industry scope for comparability purposes across models while effectively partialling out other (potentially unobserved) firm-level heterogeneity.

Finally, the data are "centered" before GRF estimation takes place: this step involves differencing out the effect of the firm-level characteristics on the outcome variables. This is done to ensure that the GRF model captures the effect of the aggregate shocks on the outcome variables, conditional on the firm-level characteristics, rather than the effect of the firm-level characteristics themselves. The LPM does not require this step, as the inclusion of firm-level characteristics in the model already partials out their effect on the outcome variables. We effectively estimate the random forest on centered variables $\tilde{Y}_i = Y_i - \hat{y}_i^{(-i)}(X_i)$ and $\tilde{W}_i = W_i - \hat{w}_i^{(-i)}(X_i)$, where $\hat{y}_i^{(-i)}(X_i)$ and $\hat{w}_i^{(-i)}(X_i)$ are leave-one-out estimates of marginal expectations, computed without the i-th observation.[12]

3.2 Documenting non-linearities in firms' sensitivities

We compare the firm-level sensitivities estimated using the GRF from Equation (2) with those obtained from the linear panel regression model featuring only linear heterogeneity in balance-sheet characteristics as in Equation (1). We document the presence of strong non-linearities in how balance-sheet characteristics influence the marginal effect estimated using GRF, which are overlooked by the LPM. In doing so, we leverage a combination of quantitative and qualitative machine learning tools and standard statistical testing.

Comparison with linear model. Table 2 shows that the Generalized Random Forest and the LPM yield similar estimates of the mean firm-level sensitivity across all outcome variables and aggregate shocks, but they diverge substantially in higher-order moments. The signs and magnitudes of the average conditional effects align with economic intuition and are consistent with findings in the existing literature.[13] While the average conditional effects are statistically identical between GRF and the LPM, higher-order moments (i.e., standard deviation, skewness, and

[12] Athey et al. (2019) note that the performance of the forests can be improved by this procedure, and that the estimator $\hat{\beta}(x)$ is more robust to confounding effects. Chernozhukov et al. (2018) also apply a similar orthogonalization procedure.

[13] For example, both GRF and the LPM estimate that a 1% increase in GDP is associated, on average, with a 2.1% increase in firms' sales, closely aligning with the 3% reported by Crouzet and Mehrotra (2020) using QFR establishment-level data.

Table 2: Summary statistics of estimated firm-level sensitivities

	GRF				Linear Panel Model			
Outcome variable	Mean	St. Dev.	Skewness	Kurtosis	Mean	St. Dev.	Skewness	Kurtosis
Panel A: Business Cycle								
Sales	2.15	0.71	0.45	2.64	2.16	1.61	0.51	2.68
Market Value	4.24	1.40	0.06	2.34	4.38	2.86	0.11	2.85
Investment	0.91	0.60	0.22	2.55	0.90	1.09	0.18	3.04
Debt	1.26	1.00	0.01	2.46	1.27	1.93	0.20	3.39
Panel B: Monetary Policy								
Sales	1.31	3.33	0.30	2.65	1.08	5.59	0.44	3.30
Market Value	-9.17	7.99	0.30	2.28	-9.87	11.26	-0.14	3.19
Investment	-0.86	2.27	-0.22	2.66	-1.11	3.51	0.04	3.00
Debt	-1.05	3.78	-0.08	2.81	-0.95	8.26	0.57	3.63
Panel C: Uncertainty								
Sales	-0.22	0.12	-0.11	2.41	-0.22	0.25	-0.42	3.64
Market Value	-1.29	0.26	-0.18	2.52	-1.30	0.53	-0.14	2.99
Investment	-0.09	0.11	-0.41	2.61	-0.09	0.20	0.13	2.97
Debt	-0.07	0.16	-0.32	2.74	-0.10	0.39	0.10	3.59
Panel D: Oil Price								
Sales	-0.02	0.06	0.02	2.75	-0.02	0.14	0.10	4.42
Market Value	-0.03	0.17	-0.18	2.53	-0.01	0.34	-0.30	3.21
Investment	-0.04	0.05	-0.27	2.66	-0.04	0.09	0.02	3.51
Debt	-0.07	0.10	-0.24	3.08	-0.07	0.21	-0.27	3.81

Notes: The table presents the summary statistics of the estimated firm-level sensitivities obtained from the GRF and the linear panel regression model across different outcome variables and shocks. Metrics include the mean, standard deviation, skewness, and kurtosis for each method. Panels A through D correspond to business cycle fluctuations, monetary policy, uncertainty, and oil price shock, respectively for all outcome variables analyzed.

kurtosis) of the distribution of firm-level sensitivities exhibit significant differences between the two methods. Specifically, the distribution of sensitivities estimated using the LPM exhibits, on average, 50% greater dispersion and 20% higher kurtosis than GRF. This suggests that GRF captures more moderate and precise patterns of heterogeneity, while the LPM provides a good first-order approximation but amplifies extreme values due to its rigid functional form and potential overfitting in high-dimensional spaces. By capturing nonlinear interactions, machine learning provides a more stable and realistic characterization of firm responses, demonstrating that nonlinearities not only affect individual firms but also shape the overall distribution of responses at the macro level.

These differences in higher-order moments result in substantial firm-level deviations between the sensitivities estimated by GRF and the LPM, despite their strong overall correlation. Figure 2 compares the individual firm-level sensitivities estimated by GRF to those obtained from the linear panel regression. Across all cases, the sensitivities estimated by GRF and the LPM exhibit a strong positive correlation, as indicated by the red linear fit trend, suggesting that both methodologies capture similar patterns in firm-level sensitivities to aggregate shocks. While sensitivities cluster around the 45-degree line in central regions, firm-level deviations between GRF and the linear model are substantial, particularly in the tails of the distribution and for firms with extreme balance-sheet characteristics. These deviations, which can be as large as 100% in magnitude and even opposite in sign, suggest that balance-sheet characteristics influence the conditional effect in complex and nonlinear ways that the LPM fails to capture.[14]

Accumulate Local Effects. An Accumulated Local Effects (ALE) plot is a visualization tool used to illustrate the relationship between one or more features and the predicted outcome of a machine learning model. It helps interpret the marginal effect of a feature on model predictions while accounting for interactions and correlations with other features.[15] We use ALE plots to assess whether the marginal effect of each balance-sheet characteristic on firm-level sensitivities varies with the level of the characteristic itself, providing evidence of potential nonlinearities in firm responses.

Figure 13 in Appendix C presents ALE plots for each balance-sheet characteristic across outcome variable-aggregate shock pairs. These plots reveal strong nonlinearities in most cases. While the estimated relationships are monotonic in some instances, ALE plots frequently exhibit pronounced concavities or convexities, suggesting that a linear specification may be an inadequate approximation–particularly in the tails of the distribution of firm characteristics. For example, distance to default often exhibits a kink around values of five or ten, beyond which the marginal effect flattens out. Similarly, the effect of firm size frequently follows an S-shaped pattern, where marginal effects are strongest for firms near the center of the size distribution. Other characteristics, such as cash holdings, sales volatility, and ROA, display U-shaped or inverted U-shaped patterns, further rejecting linearity. Only in a few cases we observe approximately linear marginal effects–for instance, leverage appears to have an approximately constant effect on the sensitivity of market value to uncertainty shocks.

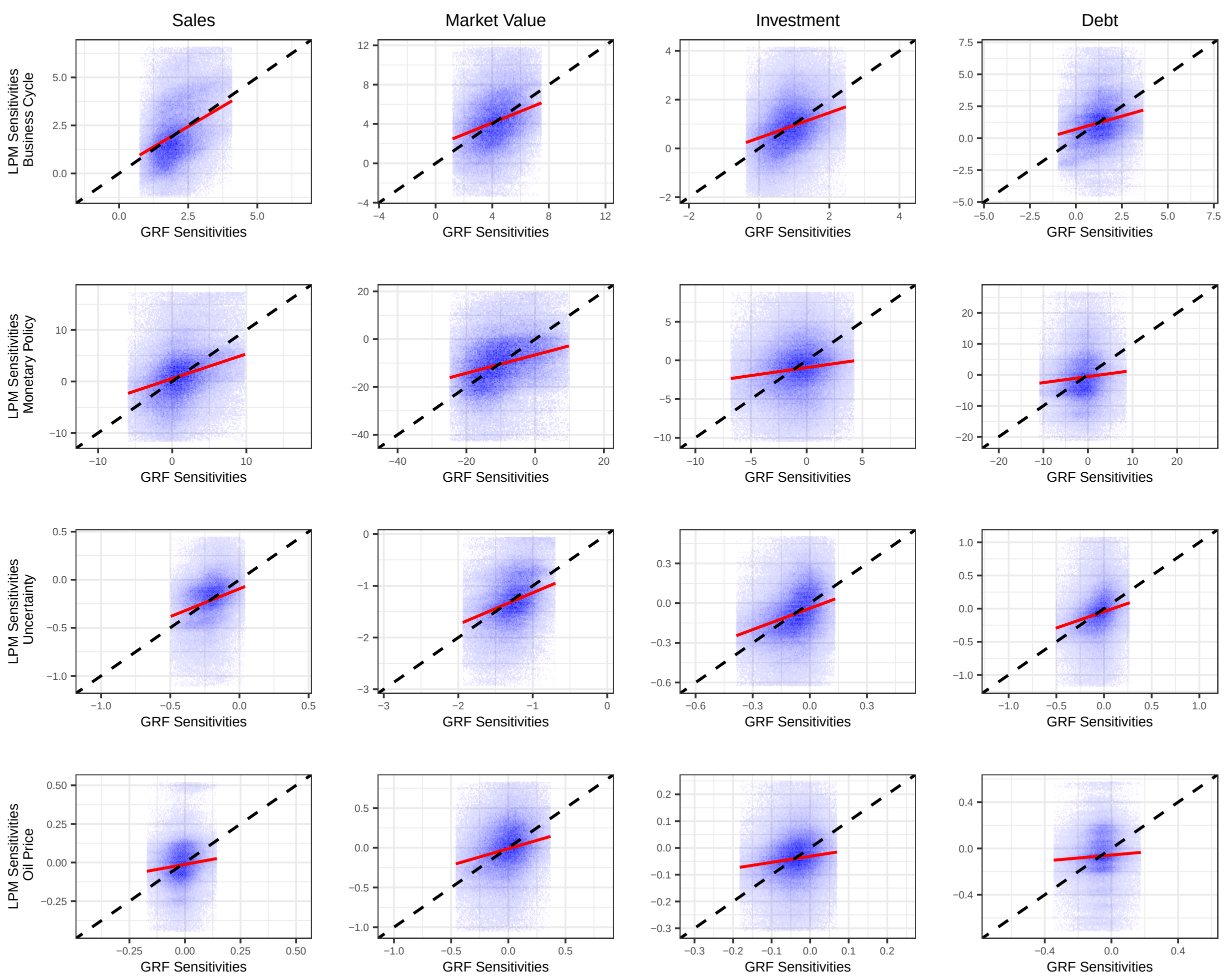
Role of interactions between characteristics. We show that interactions between firms' characteristics strongly influence firms' sensitivities to aggregate shocks. The linearity assumption embedded in the LPM rules out any non-linearity in which combinations of characteristics jointly influence firms' sensitivities by interacting with each other.

We quantify the strength of these interactions for the GRF estimated sensitivities leveraging machine learning tools. Specifically, we rely on the Friedman's H-statistic, which is a measure used

[14] Figure 12 in Appendix C reports the distribution of errors, defined as the percentage deviation between GRF and linear panel sensitivities, for each aggregate shock-outcome variable pair.

[15] ALE plots provide a more reliable alternative to the commonly used Partial Dependence Plot (PDP). A key assumption underlying PDPs is that the analyzed features are independent of others, which may not hold in empirical applications. In contrast, ALE plots compute local effects within intervals, conditioning on the joint distribution of other features and thereby allowing for correlations.

Figure 2: Comparison of sensitivities on actual data



Notes: The figure compares firm-level sensitivities to aggregate fluctuations estimated using the GRF and a linear panel regression. The LPM estimates are derived by regressing each firm's outcome variable on the aggregate shock, with interactions between the shock and all firm-level characteristics. Each subplot represents a specific aggregate shock - outcome variable pair. The columns correspond to the four outcome variables (sales, market value, debt, and investment), while the rows represent the four aggregate shocks (business cycle, uncertainty, monetary policy, and oil price shocks). The dashed black line represents the 45-degree line, indicating perfect alignment between the two estimates, while the red line shows the fitted relationship between the GRF and LPM sensitivities. Firm-level sensitivities are trimmed at the 1.5% level on both tails.

to quantify the degree of interaction between characteristics in a predictive model. It evaluates whether the joint effect of two or more characteristics on the model's output is significantly greater than the sum of their individual effects. The statistic is computed by comparing the variance in the model's predictions explained by the interaction between characteristics with the total variance explained by the characteristics. The statistic has the desirable property of ranging from zero to one, where zero indicates purely additive effects with no interaction between characteristics, and one indicates that characteristics only affect the model jointly. We consider two cases: a total interaction measure evaluating a characteristic's interaction with all other characteristics in the model, and a two-way interaction measure assessing the interaction between two characteristics.[16]

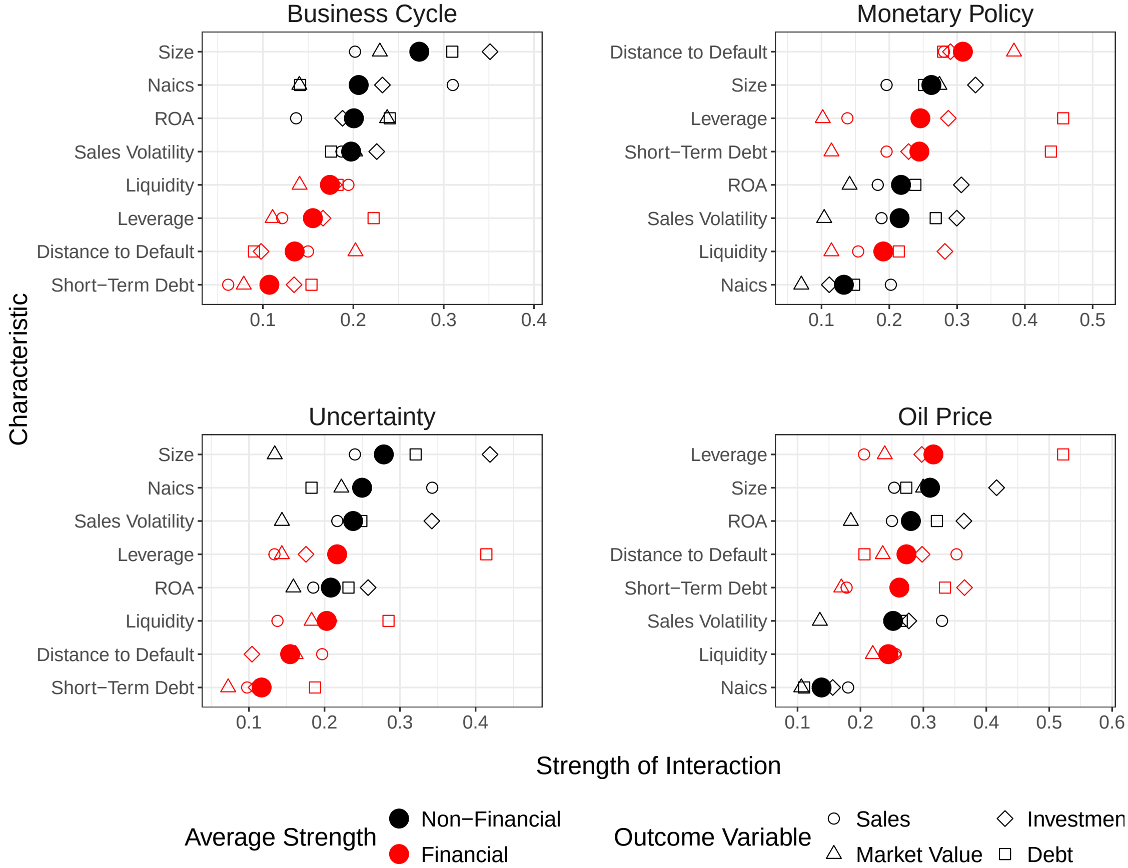
Figure 3 shows that interactions among characteristics are quantitatively relevant and strongly influence firms' sensitivities to aggregate shocks. We measure the strength of interactions that each characteristic has with all other characteristics together for each outcome variable - aggregate shock pair. On average, interactions can represent up to 40% of the variance in the outcome variable explained by a given characteristic, underscoring the importance of this form of non-linearities. Firm size is the characteristic with the highest or second highest H-statistic across all aggregate shocks, indicating that a large portion of its relevance for firms' sensitivities derives from its influence on the effect of other characteristics. Non-financial characteristics exhibit stronger interactions than financial characteristics for business cycle fluctuations and uncertainty shocks, while the ranking is more balanced between financial and non-financial characteristics for monetary policy and oil price shocks. Notably, on average across all cases, firms' debt and investment choices are the outcome variables that exhibit the strongest influence from interactions among characteristics.

Figure 14 in Appendix C evaluates the role of pairwise interactions among firms' characteristics, extending the insights from the joint interaction case above. We focus on the ten most significant characteristic pairs for each outcome variable-aggregate shock pair, presenting the average strength of interactions across outcome variables for each aggregate shock. We find that monetary policy and oil price shocks exhibit stronger pairwise interactions, while business cycle fluctuations and uncertainty shocks have fewer and more moderate interactions, indicating that interactions are more diffuse among characteristics in the latter case. Interestingly, we also find that financial and non-financial characteristics interact with each other, highlighting the importance of including both sets of characteristics in the analysis. For instance, firm size strongly interacts with many, often all, other characteristics, in line with its strong quantitative relevance in the joint interaction case.

Statistical tests for non-linearity. We complement the machine learning tools by formally testing whether the relationship between the conditional effect estimated using GRF and firms' characteristics is linear. In the LPM, the implied conditional effect of an aggregate shock on firms'

[16]Formally, the H-statistic is defined as $H_j^2 = \frac{\sum_{i=1}^n \left[\hat{f}(x^{(i)}) - PD_j(x_j^{(i)}) - PD_{-j}(x_{-j}^{(i)})\right]^2}{\sum_{i=1}^n \hat{f}^2(x^{(i)})}$, where $\hat{f}(x^{(i)})$ is the prediction function for observation i, and $PD_j(x_j^{(i)})$ and $PD_{-j}(x_{-j}^{(i)})$ are the partial dependence functions that depend on characteristic j and all features except the j-th characteristic, respectively. The statistic can be easily extended to the pairwise case, where the 2-way partial dependence function replaces the prediction function.

Figure 3: Strength of interactions



Notes: This plot presents the strength of interaction between firm characteristics for each aggregate shock. We measure the strength of interaction of each characteristic using the Friedman's H-statistic against all other characteristics. Each panel corresponds to a specific shock (e.g., business cycle, uncertainty, monetary policy, or oil price shock). The characteristics on the x-axis are ordered by their average strength of interaction within each aggregate shock, with filled points representing the average strength of interaction for each characteristic. "Financial" characteristics are depicted in red, while "Non-Financial" characteristics are shown in black. Unfilled shapes overlay the interaction strength for individual outcome variables: circles represent sales, triangles represent market value, squares represent debt, and diamonds represent investment. The x-axis reports the interaction strength, where a value of 0.01 corresponds to 1%.

outcomes is linear in firms' characteristics, i.e. $b(X_{i,t-1}) = \beta_0 + \sum_{j \in J} \beta_j \cdot X^j_{i,t-1}$, where J is the set of characteristics. We test whether the estimated GRF sensitivities, $\widehat{\beta(X_{i,t-1})}$, are linear in the characteristics, leveraging three different statistical measures commonly used in testing for linearity: the estimated degrees of a Generalized Additive Model (GAM henceforth), and the Harvey-Collier and Regression Specification Error Test (RESET henceforth) tests.

We estimate a GAM of the firms' sensitivities on firms' characteristics. In a GAM, a univariate dependent variable depends linearly on unknown smooth functions of some predictor variables.

Formally, this translates into estimating the following GAM: $\widehat{\beta(X_{i,t-1})} = \sum_{j \in J} f_j(X^j_{i,t-1})$, where J is the set of characteristics and f_j is a smooth function of characteristic j.[17] The effective degrees of freedom estimated by the GAM for each smooth function f_j can be interpreted as a proxy for the degree of non-linearity in the relationship between dependent and predictor variables: an EDF around one indicates a linear relationship, while an EDF larger than one indicates a non-linear relationship. The last column of Table 6 in Appendix C reports the minimum estimated degree of freedom across characteristics for each outcome variable - aggregate shock pair. In all cases, the minimum EDF is around six, well above the threshold value of one, indicating the presence of strong non linearities in firms' characteristics, in line with partial dependence analysis.

As an alternative, we run the RESET to check for misspecification in a linear OLS regression of $\widehat{\beta(X_{i,t-1})}$ onto the complete set of firms' characteristics as explanatory variables. The test adds higher-order terms or interaction terms of the independent variables to the regression. If these added terms are statistically significant, it suggests that the model may be misspecified. Columns (3) and (4) of Table 6 report the test statistics and the corresponding p-values for each outcome variable - aggregate shock pair, respectively. Also, in this case, linearity is rejected as no outcome variable - aggregate shock pair accepts the null hypothesis of correct model specification.

Lastly, the Harvey-Collier test for linearity involves a t-test on the mean of the recursive residuals between dependent and independent variables, which should be equal to zero under the null hypothesis that their relationship is linear. We perform the test for each aggregate shock-outcome variable pair by testing the linearity between the firm-level sensitivities estimated using GRF and firms' characteristics. Formally, we consider a linear OLS regression of $\widehat{\beta(X_{i,t-1})}$ onto the complete set of firms' characteristics as explanatory variables The first two columns of Table 6 report the test statistics and the corresponding p-values, respectively. As expected, linearity is strongly rejected, in line with previous statistical measures.

3.3 Relevance of characteristics for heterogeneity

We use traditional and modern machine learning tools to evaluate the role of firm characteristics in shaping firms' responses to aggregate shocks and their heterogeneity across firms. Unlike traditional parametric models, GRF estimates firm-level sensitivities without imposing a predetermined functional form. GRF efficiently manages this complexity by automatically assigning greater weight to the most relevant covariates. This approach allows for a more precise assessment of how firm characteristics drive variation in sensitivity across firms. Crucially, the use of a high-dimensional characteristic space does not compromise the interpretability of the results.

[17] We exclude industry scope from the set of characteristics because it is unreasonable to assess whether the conditional effect is non linear in 5-digit NAICS. We include 5-digit NAICS fixed effects to control for heterogeneity in industry scope.

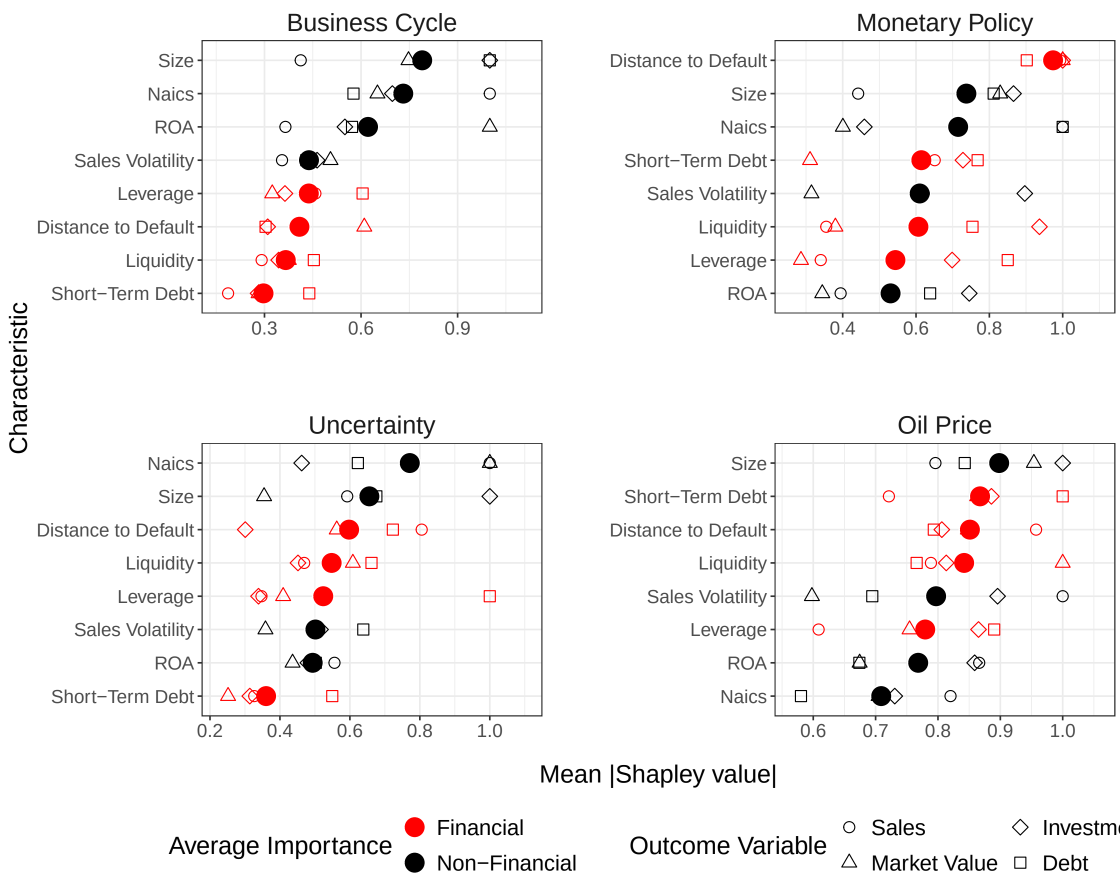
Heterogeneity in characteristics' relevance We quantify the marginal impact of a characteristic on firms' sensitivities using Shapley values, a game-theoretic approach for attributing the contribution of individual features to a machine learning model's predictions. Shapley values measure a feature's marginal contribution by computing the difference in predictions with and without the feature across all possible subsets, averaging these contributions over all subsets. Given the computational requirements, we compute Shapley values for each characteristic in all outcome variable-aggregate shock pairs over a grid of 100 points corresponding to the characteristic's percentiles. To quantify the average importance of a characteristic to firms' sensitivities, we follow standard practice and compute the mean absolute value of the estimated Shapley values over the hundred points. The intuition is that each Shapley value is a force that either increases or decreases the model's output; therefore, characteristics with large absolute Shapley values are relatively more important. We normalize the importance of each characteristic so that it is equal to one for the characteristic with the highest mean absolute value in each given outcome variable - aggregate shock pair.

Figure 4 shows that size is the leading feature in explaining firms' sensitivity to aggregate shock, being the most relevant characteristic for the response to business cycle fluctuations and uncertainty shocks, and the second most relevant for monetary policy and oil price shocks. Distance to default is the most relevant characteristic in explaining firms' sensitivity to monetary policy shock. Importance measures are dispersed, but the distribution of importance measures does not exhibit a strong skewness. In other words, we do not see many cases where one characteristic has an overwhelming effect on firms' sensitivity relative to all the other characteristics. This indicates that, on average, the importance of each characteristic, relative to the most important characteristic, is comparable in magnitude, underscoring the importance of including multiple characteristics to explain firm-level sensitivities to aggregate shocks. We find that non-financial characteristics are overwhelmingly more important than financial characteristics for sensitivity to business cycle fluctuations, while the relative importance of the two sets of characteristics is ambiguous for the other shocks.

Heterogeneity in firms' sensitivity We also assess the importance of each characteristic in driving heterogeneity in firms' sensitivities.[18] Specifically, we measure the contribution of each characteristic to heterogeneity by analyzing its role in the moment function, which is derived from the proportion of splits associated with the characteristic of interest. In a random forest framework,

[18]We formally test for the presence of heterogeneity in conditional average effects across firms using the machine-learning based Chernozhukov et al. (2018) test. Appendix A provides details on the construction of the test. Figure 15 in Appendix C reports the coefficients and the corresponding p-values of the heterogeneity in treatment effects test. In almost all outcome variable - aggregate shock pairs, the Chernozhukov et al. (2018) test strongly supports the presence of heterogeneity in sensitivities across firms.

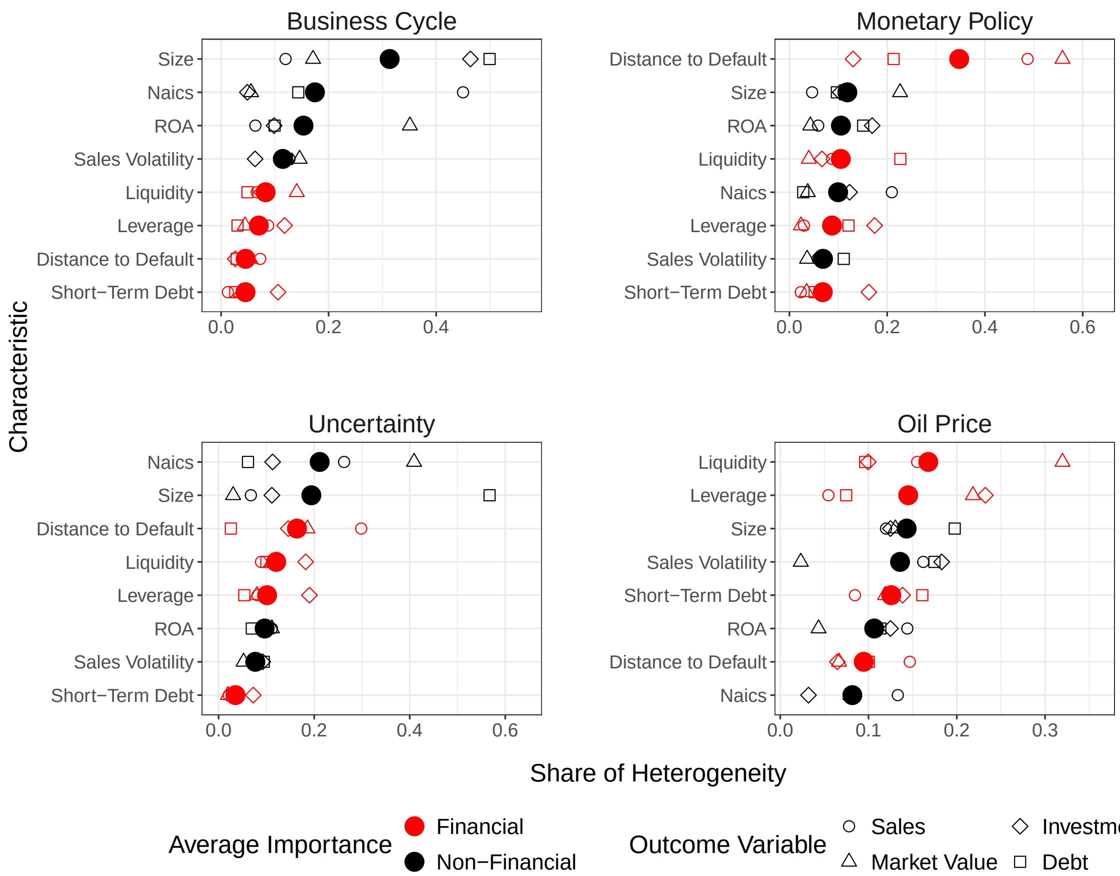
Figure 4: Marginal impact of characteristics on firms' sensitivity - Shapley values



Notes: This plot visualizes Shapley value-based importance of various characteristics across different shocks and outcome variables. Each panel corresponds to a specific aggregate shock (e.g., business cycle, uncertainty, monetary policy, or oil price shock). The x-axis orders characteristics by their mean absolute Shapley value, capturing their marginal contribution to firms' sensitivities. We compute Shapley values for each characteristic in all outcome variable-aggregate shock pairs over a grid of 100 points corresponding to the characteristic's percentiles. We compute the mean absolute value of the estimated Shapley values over the hundred points. We normalize importance by scaling each characteristic to the highest mean absolute Shapley value within each outcome variable - aggregate shock pair, setting the maximum to one. Filled points represent the average across outcome variables for each characteristic, with financial characteristics in red and non-financial characteristics in black. Unfilled shapes overlay the importance for individual outcome variables: circles represent sales, triangles represent market value, squares represent debt, and diamonds represent investment. A value of 0.01 corresponds to 1%.

the importance of a characteristic is measured as the depth-weighted frequency of splits where the characteristic is used. This metric provides an intuitive interpretation of how much of the variation in sensitivities is attributable to each firm characteristic. We compute this measure for each characteristic across all outcome variable-aggregate shock pairs, allowing us to decompose the sources of heterogeneity in firm responses to aggregate fluctuations.

Figure 5: Characteristics importance for heterogeneity



Notes: This plot visualizes the share of heterogeneity explained by each characteristic across different shocks and outcome variables. The share of heterogeneity explained by each characteristics is computed as the depth-weighted frequency of splits in the forest where the characteristic is used. Each panel corresponds to a specific shock (e.g., business cycle, uncertainty, monetary policy, or oil price shock). The characteristics on the x-axis are ordered by their average importance share within each shock, with filled points representing the average importance share of each characteristic. "Financial" characteristics are depicted in red, while "Non-Financial" characteristics are shown in black. Unfilled shapes overlay the importance share for individual outcome variables: circles represent sales, triangles represent market value, squares represent debt, and diamonds represent investment. The x-axis shows the importance share, where a value of 0.01 corresponds to 1% of total heterogeneity.

Figure 5 shows that heterogeneity is dispersed across many characteristics. On average, most characteristics contribute between 10% and 20% to the overall heterogeneity in firms' sensitivities, with only a few instances exceeding 50%. This suggests that firm-level heterogeneity is not driven by a single characteristic but rather by a broad set of attributes. Furthermore, the ranking of characteristics varies significantly across aggregate shock-outcome variable pairs. For example, firm size, along with other non-financial characteristics such as industry scope, plays a dominant role in explaining firms' sensitivity to business cycle fluctuations and uncertainty shocks. In contrast, financial variables are more relevant for monetary policy and uncertainty shocks, with distance to default explaining nearly 60% of the heterogeneity in monetary policy shocks, while cash holdings and leverage account for large shares of heterogeneity in response to uncertainty shocks.

We find that these measures of importance are strongly correlated, despite there being no ex-ante reason for such a correlation. Figure 16 in Appendix C illustrates the relationship between the share of heterogeneity and the Shapley-based characteristic relevance, after controlling for aggregate shocks, outcome variables, and characteristic fixed effects. The two measures exhibit a positive correlation, suggesting that the characteristics influencing firms' sensitivities on average are also those driving heterogeneity across firms. Additionally, we find a positive correlation between the strength of interactions and Shapley-based relevance, indicating that a characteristic's importance for the outcome variable depends significantly on its interaction with other characteristics.

3.4 Comparison with the literature

Figure 5 also offers insights in line with previously studied cases in the literature. First, related to the literature on business cycles, we find that non-financial characteristics collectively emerge as the primary drivers of firms' sensitivity to business cycle fluctuations, accounting for 86.5% of the heterogeneity in investment responses and 76% in sales. Among non-financial characteristics, industry plays the most significant role in explaining sales growth variability, contributing 45% to its heterogeneity. Meanwhile, firm size is particularly relevant for understanding variations in investment and debt issuance, contributing approximately 50% during periods of economic booms and recessions. In contrast, heterogeneity in stock market price responses at the micro level is more closely linked to firms' overall profitability, which accounts for 35% of its variation. Notably, financial characteristics are the least important in explaining firms' sensitivities to business cycles across all variables considered. These findings align with Crouzet and Mehrotra (2020), who highlight that demand conditions and industry scope are critical determinants of firms' heightened sensitivity to business cycle fluctuations, while financial characteristics play a more limited role.

Second, Figure 5 suggests two facts related to the literature on monetary policy. First, firms' default risk emerges as the most significant variable explaining heterogeneity in firms' sensitivity to identified monetary policy shocks. Specifically, we find that distance-to-default accounts for more than 50% of the heterogeneous response of firms' stock market prices and sales growth to monetary policy shocks, whereas it shares the first place along with liquidity in explaining firms' investment responses to monetary policy shocks (contributing 23% and 21%, respectively). These findings align with the competing empirical results of Ottonello and Winberry (2018) and Jeenas

(2018), who emphasize the role of distance-to-default and the importance of liquidity in studying firms' investment responses to monetary policy.[19]

Second, our findings stand in contrast to the emphasis placed on sectoral heterogeneity in the monetary policy literature. Notably, **?** and **?** highlight the role of production networks, price rigidity, and input-output linkages in amplifying the real effects of monetary policy shocks across sectors. Their findings suggest that sectors with higher price flexibility and stronger network connections contribute disproportionately to the transmission of monetary policy. While our results do not dismiss the relevance of sectoral mechanisms at a macroeconomic level, they highlight that firm-specific characteristics are the primary drivers of heterogeneous responses at the micro level. Along the same lines, **?** show that monetary policy shocks generate significant heterogeneity in stock returns across sectors, driven by production network linkages. We find that sectoral heterogeneity is almost insignificant in explaining heterogeneity in stock market returns.

Third, Figure **??** shows that non-financial characteristics, particularly size and industry, are relevant for the heterogeneity in the response of investment and stock market prices to changes in uncertainty (57% and 41%, respectively). In relation to the literature studying the role of financial positions in uncertainty shocks (Alfaro et al., 2024), we find that financial characteristics, particularly distance-to-default, leverage, and liquidity, are relevant only for the response of debt issuance and, to a certain extent, for the response of sales, which are not the focus of their work.

Lastly, relatively few studies have explored the effect of oil shocks and their heterogeneity across firms. We contribute to the literature by showing that financial variables, such as leverage and liquidity, are key for the heterogeneity in the response of market value to oil shocks, with an importance share of 32% and 22%, respectively. In contrast, real outcome variables (sales, investment, and debt) exhibit a more balanced division between financial and non-financial characteristics and a more dispersed allocation across characteristics, suggesting that these cases require a more comprehensive analysis.

4 Aggregate Implications of Firms' Heterogeneity

This section studies the aggregate implications of the heterogeneity and non-linearity in firms' sensitivity to aggregate shocks. We first propose a theory of aggregation to compute the response of any aggregate variable to aggregate fluctuations by aggregating firm-level individual responses. Then, using the aggregation theory and the estimated sensitivities, we assess the contributions of non-linearity and heterogeneity in firms' sensitivity to the overall response of the aggregate

[19] Additionally, debt issuance in response to monetary policy shocks, which is less studied in the literature, is largely driven by leverage (17%) and short-term debt (16%). This finding aligns with the idea that firms preserving financial capacity are more likely to adjust their debt positions.

economy.

4.1 A simple theory of aggregation

Consider a set I_t of firms continuing to operate between t and $t-1$. Let G_t and $g_{i,t}$ denote the aggregate and the firm-level response of variable Y_t following an aggregate shock W_t, respectively:

$$G_t = \frac{Y_t}{Y_{t-1}} \qquad g_{i,t} = \frac{Y_{i,t}}{Y_{i,t-1}}. \tag{5}$$

Let ω_{it-1} be the share of Y_{t-1} accounted for by firm i:

$$\omega_{it-1} = \frac{Y_{i,t-1}}{Y_{t-1}} \qquad \text{where } Y_{t-1} = \sum_{i \in I_t} Y_{i,t-1}. \tag{6}$$

It follows that we can write the aggregate response of variable Y_t to an aggregate shock at time t as:

$$G_t = \sum_{i \in I_t} \omega_{i,t-1} g_{i,t}. \tag{7}$$

Importantly, in our setting, we can easily construct the aggregate response G_t using the firm-level responses from the estimated models as $\widehat{g_{i,t}} = \widehat{\beta\,(x)} W_t$ and construct the corresponding shares from our dataset.

The aggregation in Equation (7) highlights that both firm-level sensitivities and shares matter for the aggregate response. In fact, we can write Equation (7) to achive the following decomposition:

$$G_t = \overline{g_t} + \mathbb{C}\mathrm{ov}(w_{i,t-1}, g_{i,t}), \tag{8}$$

where the first term is the unweighted average sensitivity across firms, $\frac{1}{|I_t|} \sum_{i \in I_t} g_{i,t}$, and the second term is the covariance between firm sensitivity and firms' importance in the aggregate, $\sum_{i \in I_t} \left(\omega_{i,t-1} - \frac{1}{|I_t|} \right) (g_{i,t} - \overline{g_t})$. The first term captures how, on average, firms respond to aggregate fluctuations without considering their relative importance in the economy. The second term reflects how heterogeneity in firms' sensitivities interacts with the heterogeneity in their weights. A positive covariance indicates that firms with higher sensitivities tend to have greater relative importance in the aggregate, amplifying the aggregate response. Conversely, a negative covariance suggests that firms with lower sensitivities are more influential, dampening the overall aggregate response.

We use Equations (7) and (8) to construct and decompose aggregate responses into an average term and a covariance term. We begin by constructing a measure of the average aggregate effects of a shock on a given observable using micro-level data. First, we calculate the aggregate response to a shock, $\hat{G}_t$, by weighting the predicted firm-level sensitivities, $\hat{g}_{i,t}$, estimated using the GRF algorithm, by their relative importance in the aggregate. This importance is proxied by each firm's share of a total outcome measure, such as sales or assets, $w_{i,t-1}$.

We construct the aggregate response G_t using Equation (7) and estimate the average aggregate effect of a shock using the following time-series regression:

$$\hat{G}_t = \alpha + \gamma W_t + \epsilon_t, \tag{9}$$

where the coefficient γ reflects the average aggregate effect of a 1% aggregate shock on an outcome. This coefficient captures the effect of both the average sensitivity of firms and the interaction between firm-level heterogeneity and their weights. We then separate the contributions of the average and covariance terms by regressing the two terms in Equation (8) on the aggregate shock in a time-series regression like Equation (9).

We use this theory of aggregation and relative decomposition to quantify the impact of non-linearities in aggregate fluctuations and the role of firm-level heterogeneity at the aggregate level.

4.2 The aggregate role of non-linearities in sensitivity

We show that non-linearities in firm-level sensitivities are not only prevalent at the micro level but also significantly influence the aggregate response of outcomes to macroeconomic fluctuations. While Section 3 documents substantial non-linearities in firms' sensitivity to aggregate fluctuations due to balance-sheet characteristics, their macroeconomic relevance depends on the distribution of weights and their correlation with firm sensitivities. If the sensitivity of firms with larger weights is not particularly affected by the presence of non-linearities, then the firm-level non-linearities may not fully translate into aggregate fluctuations. To evaluate their aggregate impact, we construct the economy-wide response of sales, market value, investment and debt using firm-level sensitivities estimated from both GRF and the LPM. We then compare the average aggregate effect of a shock, γ from Equation (9) across methods, to assess the role of non-linearities and higher-order interactions in shaping macroeconomic dynamics. We collect and report the estimated effects with relative standard errors in Table 3.

Table 3 shows that the differences between the average aggregate response estimated via GRF and LPM are statistically significant and economically relevant, indicating that non-linearities in firms' sensitivities play a crucial role in shaping the aggregate response to macroeconomic fluctuations. We find that, in most of cases, the non-linearity bias in the average aggregate response is negative, i.e. GRF tends to estimate a lower aggregate response. Panel A suggests that firm-level non-linearities significantly dampen the response of sales and stock market prices to business cycle fluctuations. While a LPM predicts that a 1% increase in GDP leads to a 2.4% rise in aggregate

Table 3: Comparing average aggregate response

	GRF		Linear Panel Model		
Outcome variable	Coefficient	StD. Error	Coefficient	StD. Error	Difference
Panel A: Business Cycle					
Sales	2.07	0.03	2.41	0.04	-0.341***
Market Value	5.22	0.06	5.42	0.08	-0.199**
Investment	0.43	0.01	0.23	0.04	0.203***
Debt	0.02	0.03	0.15	0.05	-0.134**
Panel B: Monetary Policy					
Sales	1.04	0.11	0.84	0.26	0.201
Market Value	-16.78	0.19	-17.97	0.40	1.195***
Investment	0.07	0.03	-0.06	0.15	0.131
Debt	-0.99	0.07	0.77	0.27	-1.757***
Panel C: Uncertainty					
Sales	-0.18	0.01	-0.22	0.00	0.038***
Market Value	-1.13	0.01	-1.08	0.02	-0.045**
Investment	0.01	0.00	0.08	0.01	-0.071***
Debt	0.05	0.00	0.05	0.02	-0.002
Panel D: Oil Price					
Sales	0.01	0.00	0.06	0.00	-0.048***
Market Value	0.04	0.01	0.09	0.01	-0.047***
Investment	-0.02	0.00	0.03	0.00	-0.051***
Debt	-0.05	0.00	-0.04	0.01	-0.006

Notes: The table presents, for each outcome variable - aggregate shock pair, the estimated average aggregate response from Equation (9) using GRF and LPM, along with their respective standard errors. Coefficients are estimated using the time-series regression in Equation (9), using the aggregate response series from Equation (7). Panels A through D correspond to business cycle fluctuations, monetary policy, uncertainty, and oil price shocks, respectively, for all analyzed outcome variables. We also report the statistical significance of the differences at the following levels: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

sales and a 5.4% increase in stock market value, accounting for non-linearities reduces these estimates by approximately 0.3 and 0.2 percentage points, respectively. Panels B, C, and D reinforce this result by showing a similar dampening effect across macroeconomic shocks. In response to a contractionary monetary policy shock that raises interest rates by 1 percentage point (Panel B), the GRF model estimates a substantially smaller decline in stock market prices, nearly 7% less than the linear model prediction. Similarly, the aggregate investment response is near zero in the GRF model but about 0.1 percentage point more negative in the linear model. A similar pattern emerges in response to uncertainty and oil price shocks (Panels C and D, respectively), where the GRF model predicts a more muted decline in investment. These findings suggest that non-linearities play a key role in shaping aggregate investment responses to macroeconomic shocks.

Figure 17 in Appendix C shows that the primary driver of the differences in the average aggregate effects obtained from the two methods is the difference in the covariance terms. The mean components from GRF and the LPM are nearly identical and statistically indistinguishable, confirming that both methods estimate similar average sensitivities. However, the covariance terms – which capture the interaction between the distribution of sensitivities and firms' weights – differ significantly and explain most of the discrepancies in aggregate responses. Specifically, the covariance terms in the LPM tend to be higher than the GRF ones, suggesting that accounting for nonlinearities weakens the relationship between firms' sensitivity and weight when the covariance is positive or causes them to move further in opposite directions when the covariance is already negative

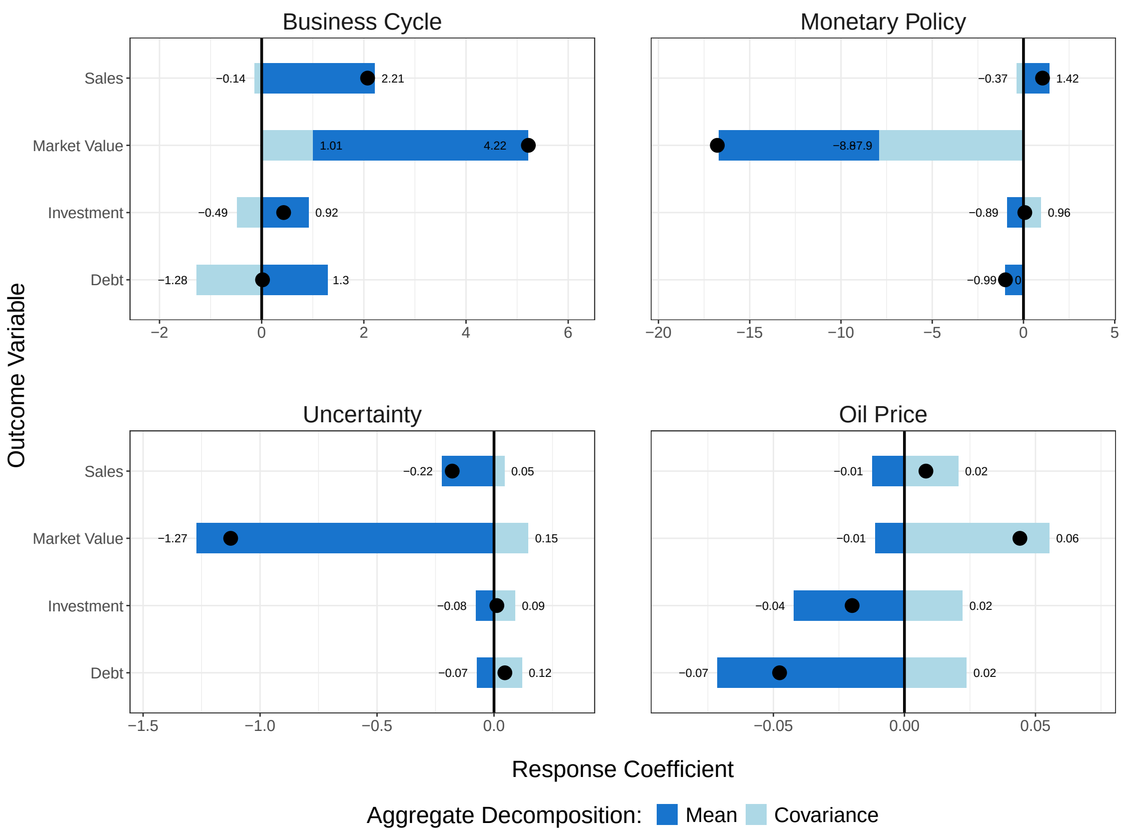
4.3 The aggregate role of heterogeneity in sensitivity

We apply the mean-covariance decomposition from Equation (8) to quantify the role of heterogeneity in shaping the average aggregate effect of aggregate fluctuations. Figure 10 in Appendix B shows that the distribution of firms' weights, ω, is highly unequal, with a small number of firms accounting for a disproportionately large share. This concentration can have significant macroeconomic implications when firm-level sensitivities to aggregate shocks are heterogeneous. If firms with larger shares systematically exhibit higher or lower sensitivity to shocks, their disproportionate weight in the economy may amplify or dampen aggregate fluctuations. By isolating the covariance term in Equation (8), we quantify the extent to which this heterogeneity influences the overall aggregate response. These findings are crucial for understanding the distributional consequences of shocks and the role that dominant firms play in shaping macroeconomic dynamics.

Figure 6 shows that the covariance term dampens the effect of aggregate shocks, highlighting the important role of firm-level heterogeneity in shaping the average aggregate response. While the unweighted average firm response (i.e., mean term) to aggregate shocks is substantial and aligns with economic intuition, the covariance term consistently exhibits the opposite sign of the mean effect, thereby dampening the overall response. This dampening effect arises because firms with larger shares, ω, exhibit lower absolute sensitivities to aggregate shocks. As a result, their disproportionate weight in the economy moderates the overall response, stabilizing fluctuations in economic expansions and contractions and reducing aggregate volatility. This pattern holds across all cases studied, except for the stock market's reaction to business cycle and monetary policy shocks, where the covariance term amplifies the aggregate response.

Quantitatively, the covariance term plays a significant but heterogeneous role across outcome variables and shocks. In response to business cycle fluctuations, it dampens the aggregate response of sales and investment by approximately 6% and 53%, respectively, while amplifying the response

Figure 6: Decomposition of average aggregate responses



Notes: The figure illustrates the decomposition of aggregate responses into mean and covariance terms for each outcome variable and aggregate shock. Bars represent the contributions of the mean and covariance terms, while the black point denotes the total average aggregate response. We estimate Equation (9) using the mean and covariance terms in Equation (8) as dependent variable, $\hat{G}_t$. The mean and covariance terms are constructed using benchmark set of firm-level sensitivities estimated with the GRF algorithm.

of stock market prices by about 24%. These effects become even more pronounced for specific exogenous shocks. We find that following a monetary policy shock, the covariance term amplifies the stock market response by nearly 89% and fully offsets the unweighted average firms' investment response. Similarly, after an unexpected increase in uncertainty, while the unweighted average firm-level response is large and negative, the aggregate effect is often muted as firms that contribute more to the economy exhibit lower sensitivities. In some cases, the heterogeneity in firm-level sensitivity is stronger than the mean effect, driving the direction of the overall aggregate response. This is relevant in the case of oil price shocks on sales and market value, which negatively affect most firms on average but may result in a positive aggregate response due to the disproportionate influence of firms with greater economic weight.

Over time We show that the estimated average aggregate responses remain stable over the period considered. To assess whether these responses are driven by specific time periods, we estimate the time-series framework in Equation (9) using a five-year rolling window. Figure 18 in Appendix C reports the estimated coefficients along with their decomposition into mean and covariance terms. The results indicate that, overall, the average aggregate responses exhibit substantial stability over time, with both the mean and covariance components remaining relatively constant. The only exception is market value, which shows a slight increase in aggregate sensitivity, particularly to business cycle fluctuations and oil price shocks. This increase is primarily driven by a rising mean effect rather than changes in the composition of firms and their sensitivities, suggesting an overall increase in firms' average stock price sensitivity to cyclical and supply shocks. In some cases, such as the response of investment to business cycle fluctuations and uncertainty shocks, the stability of the average aggregate response masks offsetting dynamics between the mean and the covariance terms: a decline in the covariance term, reflecting a weaker correlation between firms' shares and sensitivities, is accompanied by a change in the mean sensitivity of similar magnitude but opposite sign.

Within and across sector heterogeneity We show that both within-sector and across-sector heterogeneity equally contribute to the dampening of the aggregate response due to firms' heterogeneity. To illustrate this, we consider a counterfactual scenario where the sensitivity of each firm is set to the median sensitivity of all firms within the same sector for a given quarter, where sectors are defined as 5-digit NAICS industries. We then construct a counterfactual aggregate response using the aggregation theory and the counterfactual sensitivities. Re-estimating the time-series framework in Equation (9), we obtain a counterfactual average aggregate response that accounts only for across-sector variation in firms' sensitivities. Comparing these counterfactual average aggregate response coefficients and their decomposition into mean and covariance terms with those obtained in the benchmark case using the full set of firms' sensitivities, we can assess the relative importance of within-sector and across-sector heterogeneity in firms' sensitivities.

Figure 20 in Appendix C shows that accounting only for sectoral heterogeneity reduces by half the effect of firms' heterogeneity on the average aggregate response. Not surprisingly, the mean effects estimated when setting firms' sensitivities equal to the median sensitivity with each sector are quantitatively and statistically identical to the benchmark case, as the average effect is usually well approximated by the average sensitivity across firms. However, the covariance term estimated in the counterfactual case is approximately half of the covariance term estimated in the benchmark case across all scenarios. This indicates that firms' heterogeneity due to both within-sector and across-sector variation equally contributes to the dampening effect of the covariance term in Figure 6. In other words, sectors with larger economic shares exhibit lower sensitivities in absolute terms, but firms with larger shares in each sector also exhibit lower sensitivities relative to the sectoral average. Both margins of heterogeneity are equally significant in shaping the aggregate response to shocks, underscoring the importance of accounting for both dimensions of heterogeneity.

Heterogeneity in financial and non-financial characteristics To estimate the role of heterogeneity in characteristics for aggregate dynamics, we estimate the aggregate response of the outcome variables under alternative distributions of financial and non-financial firm characteristics. We compare the aggregate response when financial characteristics are held constant at the quarter median – allowing non-financial heterogeneity to fully operate – and vice versa. Figure 19 in Appendix C shows that abstracting from the heterogeneity in non-financial characteristics generates relatively larger and statistically significant departures from the aggregate response of the benchmark case than abstracting from the heterogeneity in financial characteristics. The greater role of non-financial characteristics is observed not only when non-financial characteristics are overwhelmingly relevant for the heterogeneity in firm-level responses, but also in aggregate shock-outcome variable pairs in which the role of financial characteristics is predominant. For instance, as shown in Section 3, the heterogeneity in non-financial characteristics impacts the aggregate response more than the heterogeneity in financial characteristics in the response of the investment and debt to the business cycle, where the share of importance of non-financial characteristics is overwhelming (86% and 67%, respectively). However, the role of the heterogeneity in non-financial characteristics remains stronger even when the share of importance of financial characteristics exceeds 60%, such as in the response of market value and investment to monetary policy (65% and 61%, respectively). The reason is that the aggregate response depends on how the distribution of firms' shares correlates with the underlying distribution of characteristics and sensitivities. While the mean term of the aggregate response does not change when abstracting from either financial or non-financial characteristics, most of the adjustment comes from the covariance term. This indicates the presence of a stronger covariance between firms with large weight in the economy and the heterogeneity in the underlying non-financial characteristics.

5 Conclusions

This paper highlights the importance of understanding firm-level sensitivity to aggregate shocks, the factors driving this sensitivity, its heterogeneity, and its implications for macroeconomic dynamics. Leveraging the Generalized Random Forest model, we uncover substantial nonlinear heterogeneity in firms' responses to economic shocks – features that traditional linear models fail to capture. At the firm level, we show that characteristics such as size play a critical role in shaping these sensitivities, with strong nonlinearities and interactions driving heterogeneity across firms.

At the macro level, we demonstrate that these firm-level nonlinearities reduce the average aggregate response of the economy to aggregate shocks. Moreover, heterogeneity in firm sensitivities systematically dampens the aggregate response, suggesting that larger firms tend to exhibit lower sensitivity to shocks. Our findings underscore the necessity of employing advanced statistical models – such as machine learning – to accurately characterize firm heterogeneity and its aggregate implications. These insights have important implications for policymakers seeking to understand how the distribution of firm characteristics affects the effectiveness of monetary interventions and the magnitude of business cycle fluctuations. Future research can extend this framework by examining cross-country differences in firm sensitivities, incorporating international linkages, and assessing how heterogeneous firm responses shape the global transmission of economic shocks.

References

Alfaro, I., N. Bloom, and X. Lin (2024): "The finance uncertainty multiplier," *Journal of Political Economy*, 132, 000–000.

Athey, S. and G. W. Imbens (2017): "The state of applied econometrics: Causality and policy evaluation," *Journal of Economic perspectives*, 31, 3–32.

Athey, S., J. Tibshirani, and S. Wager (2019): "Generalized Random Forests," *The Annals of Statistics*.

Bauer, M. D. and E. T. Swanson (2023): "A reassessment of monetary policy surprises and high-frequency identification," *NBER Macroeconomics Annual*, 37, 87–155.

Begenau, J. and J. Salomao (2019): "Firm financing over the business cycle," *The Review of Financial Studies*, 32, 1235–1274.

Belloni, A., V. Chernozhukov, I. Fernandez-Val, and C. Hansen (2017): "Program evaluation and causal inference with high-dimensional data," *Econometrica*, 85, 233–298.

Buera, F. J. and B. Moll (2015): "Aggregate implications of a credit crunch: The importance of heterogeneity," *American Economic Journal: Macroeconomics*, 7, 1–42.

Chang, M., X. Chen, and F. Schorfheide (2024a): "Heterogeneity and Aggregate Fluctuations," *Journal of Political Economy*.

——— (2024b): "On the Effects of Monetary Policy Shocks on Earnings and Consumption Heterogeneity," .

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernandez-Val (2018): "Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India," *National Bureau of Economic Research*.

Christiano, L. J., R. Motto, and M. Rostagno (2014): "Risk shocks," *American Economic Review*, 104, 27–65.

Cloyne, J., C. Ferreira, M. Froemel, and P. Surico (2018): "Monetary policy, corporate finance and investment," Tech. rep., National Bureau of Economic Research.

Cooley, T. F. and V. Quadrini (2001): "Financial markets and firm dynamics," *American economic review*, 91, 1286–1310.

——— (2006): "Monetary policy and the financial decisions of firms," *Economic Theory*, 27, 243–270.

Covas, F. and W. J. D. Haan (2011): "The cyclical behavior of debt and equity finance," *American economic review*, 101, 877–899.

Crouzet, N. (2018): "Aggregate implications of corporate debt choices," *The Review of Economic Studies*, 85, 1635–1682.

Crouzet, N. and N. R. Mehrotra (2020): "Small and large firms over the business cycle," *American Economic Review*, 110, 3549–3601.

Deng, M. and M. Fang (2022): "Debt maturity heterogeneity and investment responses to monetary policy," *European Economic Review*, 144, 104095.

Farre-Mensa, J. and A. Ljungqvist (2016): "Do measures of financial constraints measure financial constraints?" *The review of financial studies*, 29, 271–308.

Freyberger, J., A. Neuhierl, and M. Weber (2020): "Dissecting characteristics nonparametrically," *The Review of Financial Studies*, 33, 2326–2377.

Gertler, M. and S. Gilchrist (1994): "Monetary policy, business cycles, and the behavior of small manufacturing firms," *The Quarterly Journal of Economics*, 109, 309–340.

Gu, S., B. Kelly, and D. Xiu (2020): "Empirical asset pricing via machine learning," *The Review of Financial Studies*, 33, 2223–2273.

Gürkaynak, R., H. G. Karasoy-Can, and S. S. Lee (2022): "Stock market's assessment of monetary policy transmission: The cash flow effect," *The Journal of Finance*, 77, 2375–2421.

Jeenas, P. (2018): "Firm balance sheet liquidity, monetary policy shocks, and investment dynamics," in *Technical Report*, Working paper.

Jungherr, J., M. Meier, T. Reinelt, and I. Schott (2024): "Corporate debt maturity matters for monetary policy," *International Finance Discussion Paper*.

Jurado, K., S. C. Ludvigson, and S. Ng (2015): "Measuring uncertainty," *American Economic Review*, 105, 1177–1216.

Känzig, D. R. (2021): "The macroeconomic effects of oil supply news: Evidence from OPEC announcements," *American Economic Review*, 111, 1092–1125.

Khazra, N. (2021): "Heterogeneities in the House Price Elasticity of Consumption," .

Krusell, P., C. Thürwächter, and J. Weiss (2023): "Upwardly Mobile: The Response of Young vs. Old Firms to Monetary Policy," Tech. rep., Working Paper.

Kumar, S., Y. Gorodnichenko, and O. Coibion (2023): "The effect of macroeconomic uncertainty on firm decisions," *Econometrica*, 91, 1297–1332.

Lenza, M. and E. Savoia (2024): "Do we need firm data to understand macroeconomic dynamics?" Tech. rep., Sveriges Riksbank Working Paper Series.

Merton, R. C. (1974): "On the pricing of corporate debt: The risk structure of interest rates," *The Journal of finance*, 29, 449–470.

Mullainathan, S. and J. Spiess (2017): "Machine learning: an applied econometric approach," *Journal of Economic Perspectives*, 31, 87–106.

Nakamura, E. (2005): "Inflation forecasting using a neural network," *Economics Letters*, 86, 373–378.

Narayan, P. K. and S. S. Sharma (2011): "New evidence on oil price and firm returns," *Journal of Banking & Finance*, 35, 3253–3262.

Ottonello, P. and T. Winberry (2018): "Financial heterogeneity and the investment channel of monetary policy," Tech. rep., National Bureau of Economic Research.

Paranhos, L. (2021): "Predicting inflation with neural networks," .

——— (2024): "How Do Firms Financial Conditions Influence the Transmission of Monetary Policy A Non-parametric Perspective," .

Stock, J. H. and M. W. Watson (2018): "Identification and estimation of dynamic causal effects in macroeconomics using external instruments," *The Economic Journal*, 128, 917–948.

Tsai, C.-L. (2015): "How do US stock returns respond differently to oil price shocks pre-crisis, within the financial crisis, and post-crisis?" *Energy Economics*, 50, 47–62.

Varian, H. R. (2014): "Big data: New tricks for econometrics," *Journal of economic perspectives*, 28, 3–28.

Appendix

A Theoretical Details

A.1 Generalized Random Forest - Algorithm

The GRF relies heavily upon the Random Forests (RF) models, since they both perform random split selection and sub-sampling. To this extent, GRF augments the methodology of RF by allowing the estimated parameters to be a weighted average of predictions, and not a pure simple average as performed in RF.

Formally, the objective of RF models is to estimate the expected value of an outcome $Y_{i,t}$, conditional on covariates $X_{i,t}$ for a given data generating process: $\mu(x) = \mathbb{E}[Y_{i,t}|X_{i,t} = x]$. The GRF aims to estimate the following moment condition:

$$\mathbb{E}[\psi_{\theta(x),\nu(x)}(O)_{i,t}|X_{i,t} = x] = 0 \quad \forall x \in \mathcal{X}, \text{and } i = 1, \ldots, n, \;\; t = 1, \ldots, T \tag{10}$$

where $O_{i,t}$ contains the set of observables, both dependent and covariates variables described in the previous section, as well as the set of exogenous shocks (W_t) that we focus on; $X_{i,t}$ represents the set of auxiliary covariates, while $\nu(x)$ is an optional nuisance parameter. Our focus is to estimate the elasticity $\hat{\theta}(x)$ for each dependent variable-shock pair, as function of all covariates.

The GRF model fits the empirical version of condition 10 by minimizing the weighted moment condition:

$$(\hat{\theta}(x), \hat{\nu}(x)) \in \;\text{argmin}_{\theta,\nu} \left\{ |\sum_{i=1}^{n} \alpha_i(x)\psi_{\theta,\nu}(O_{i,t})|_2 \right\} \tag{11}$$

The main additional feature of the GRF comes from the weighting function $\alpha_i(x)$: this aims to find firms with similar elasticities - depending on their characteristics $X_{i,t}$ - and associate higher weights to them. The algorithm developed by Athey et al. (2019) grows a set of B trees and defines $L_b(x)$ as the training set falling in the same "leaf" as x.

$$\alpha_{bi}(x) = \frac{\mathbf{1}(\{X_i \in L_b(x)\})}{|L_b(x)|}, \qquad \alpha_i(x) = \frac{1}{B}\sum_{b=1}^{B} \alpha_{bi}(x)$$

By bootstrapping the dataset and growing random forests, the methodology allows estimating the parameters of interest defined on many dimensions, in contrat with linear models (e.g. OLS). The interpretation of the estimated parameters $\hat{\theta}(x)$ is of a conditional local average treatement of the elasticity for a given shock.

We further estimates the average effect in the causal forests via estimates of the average partial effect, i.e. $\mathbb{E}[\text{Cov}(W_t, Y_{i,t})/\text{Var}(W_t|X_{i,t})]$. These average effects are reported in Figure 2.

A.2 Chernozhukov et al. (2018) test for heterogeneity

The test creates two synthetic variables, C_i and D_i:

$$C_i = \bar{\beta}(W_i - \hat{W}_i),$$
$$D_i = (\hat{\beta}^{cf} - \bar{\beta})(W_i - \hat{W}_i),$$

where the former uses only the average treatment effect while the latter is the prediction that takes into account the heterogeneity as predicted by the casual forest. The test consists in running the following regression of residuals in treatment on C_i and D_i:

$$Y_i - \hat{Y}_i = \gamma C_i + \delta D_i \tag{12}$$

The null hypothesis of the test is $\delta = 0$, which indicates that the casual forest does not capture any heterogeneity. In line with the evidence on the CV, we find that we can reject the null hypothesis of no heterogeneity in treatment effects for almost all aggregate shock-outcome variable pairs.

B Construction of the dataset and cleaning

B.1 Firm-level variables

We construct the firm-level variables in the Compustat database following standard practices. Outcome variables are calculated as a 1-year percentage growth using the Haltiwanger formula. Nominal sales are represented by the variable *saleq* in Compustat. The market value of the firm is the stock price (*prccq*) multiplied by the number of outstanding shares (*cshoq*). The investment rate is the 1-year change in capital stock, with capital stock equal to the book value of capital calculated using the perpetual inventory method. The initial value of a firm's capital stock is measured as the earliest available entry of *ppegtq*, and we then iteratively construct it from *ppentq*. Debt issuances are the percentage change in total debt, calculated as the sum of debt in current liabilities (*dlcq*) and long-term debt (*dlttq*). Inventories are represented by the variable *invtq* in Compustat. Independent variables are always expressed in levels. Leverage is calculated as the ratio of debt in current liabilities (*dlcq*) and long-term debt (*dlttq*) to total assets (*atq*). The cash ratio is the ratio of cash and short-term investments (*cheq*) to total assets (*atq*). Sales growth volatility is the standard deviation of firms' real sales growth in a 10-year rolling window. Distance to default is calculated for each firm using the algorithm in Merton (1974). The short-term debt ratio is the ratio of current debt (*dlcq*) to total debt. Size is the log of total assets (*atq*). Return on assets is the ratio of net income (*niq*) to total assets. Finally, industry scope is proxied with industry classification based on the NAICS-5 industry digit. All the independent variables, with the exception of industry classification, are yearly averaged before cleaning.

Additionally, to compute variables in real terms, we deflate capital stock, sales, and total assets using the implied price index of gross value added in the U.S. non-farm business sector.

B.2 Sample selections and cleaning

The sample period is 1990Q1 to 2019Q4. We perform the following cleaning steps:

i) We keep only US-based firms, $fic_{i,t} =$"USA".

ii) To avoid firms with strange production functions, drop regulated utilities and financial companies, we drop all firm-quarters for which the 4-digit sic code is in the range [4900,5000) or [6000,7000).

iii) To get rid of years with extremely large values for acquisitions to avoid the influence of large mergers, we drop all firm-quarters for which the value of acquisitions $acq_{i,t}$ is greater than 5% of total assets $atq_{i,t}$.

iv) We drop all firm-quarters for which the measurement of Total Assets $atq_{i,t}$, Sales $saleq_{i,t}$, Property, Plant and Equipment (Net) $ppentq_{i,t}$, Cash and Short-Term Investments $cheq_{i,t}$, Debt in Current Liabilities $dlcq_{i,t}$, Total Long-Term Debt $dlttq_{i,t}$, Total Inventories $invtq_{i,t}$ are missing or negative.

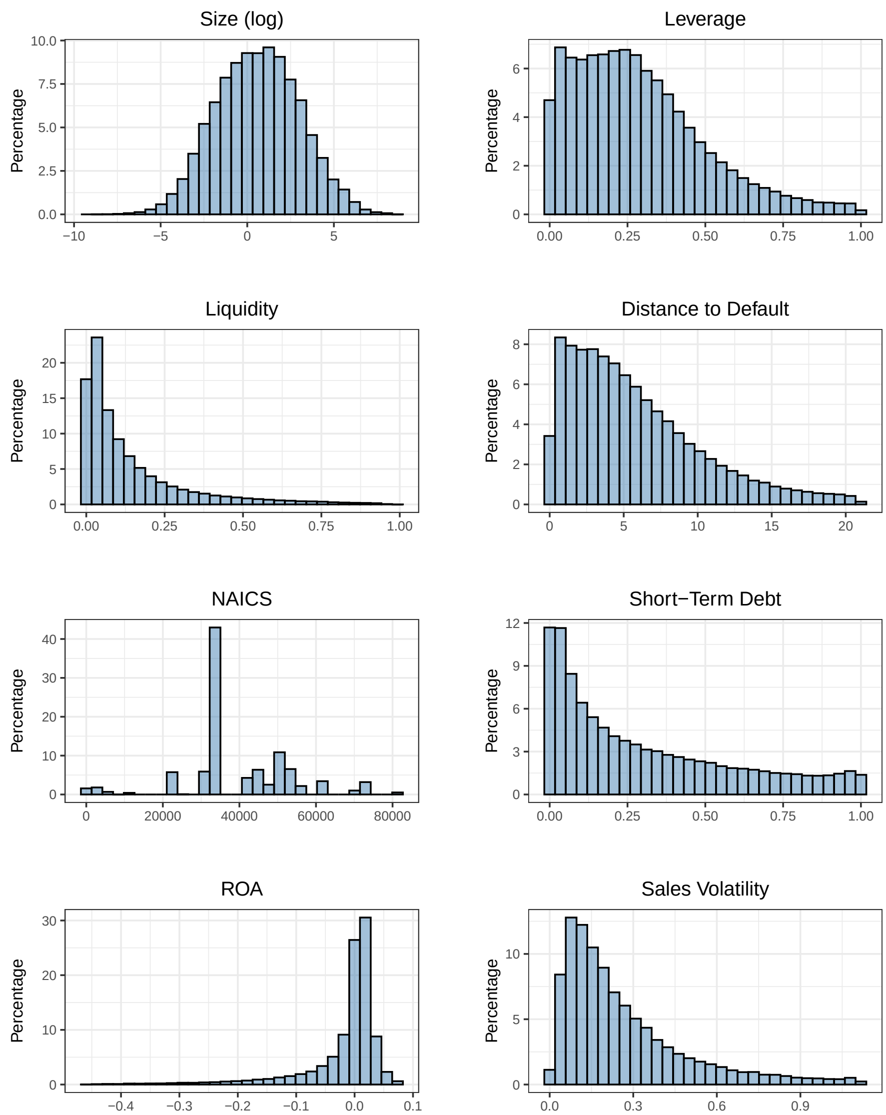
v) We drop all firm-quarters before a firm's first observation of Property, Plant, and Equipment (Gross) $ppegtq_{i,t}$.

Before estimating the models, we trim the variables at the top 1.5% level when the variables are strictly positive, and we trim 1.5% on both sides if the variables can also be negative. To reduce the number of missing values in the GRF, we linearly interpolate each independent variable after completing all cleaning steps.

We further group variables by type, distinguishing between financial and non-financial characteristics. Financial variables include leverage, liquidity, distance to default, and short-term debt. Non-financial variables include size, sales growth volatility, return on assets, and industry classification at the 5-digit NAICS level.

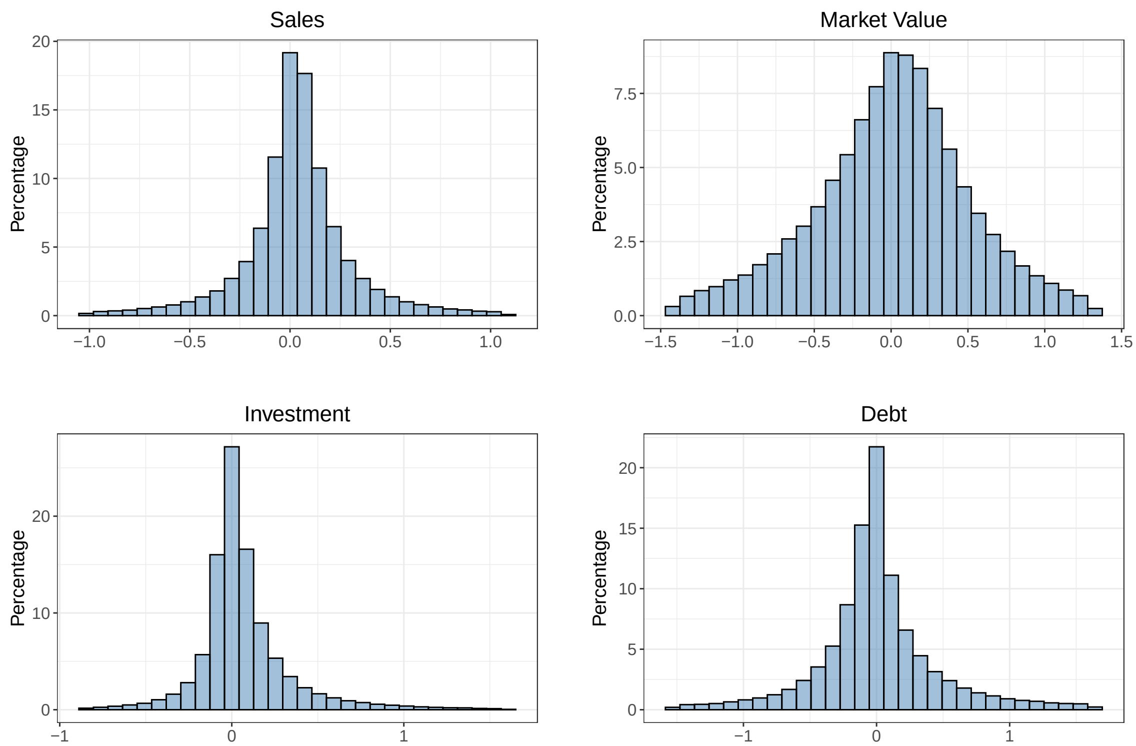
B.3 Distribution of firms' outcome and features

Figure 7: Distribution of the independent variables



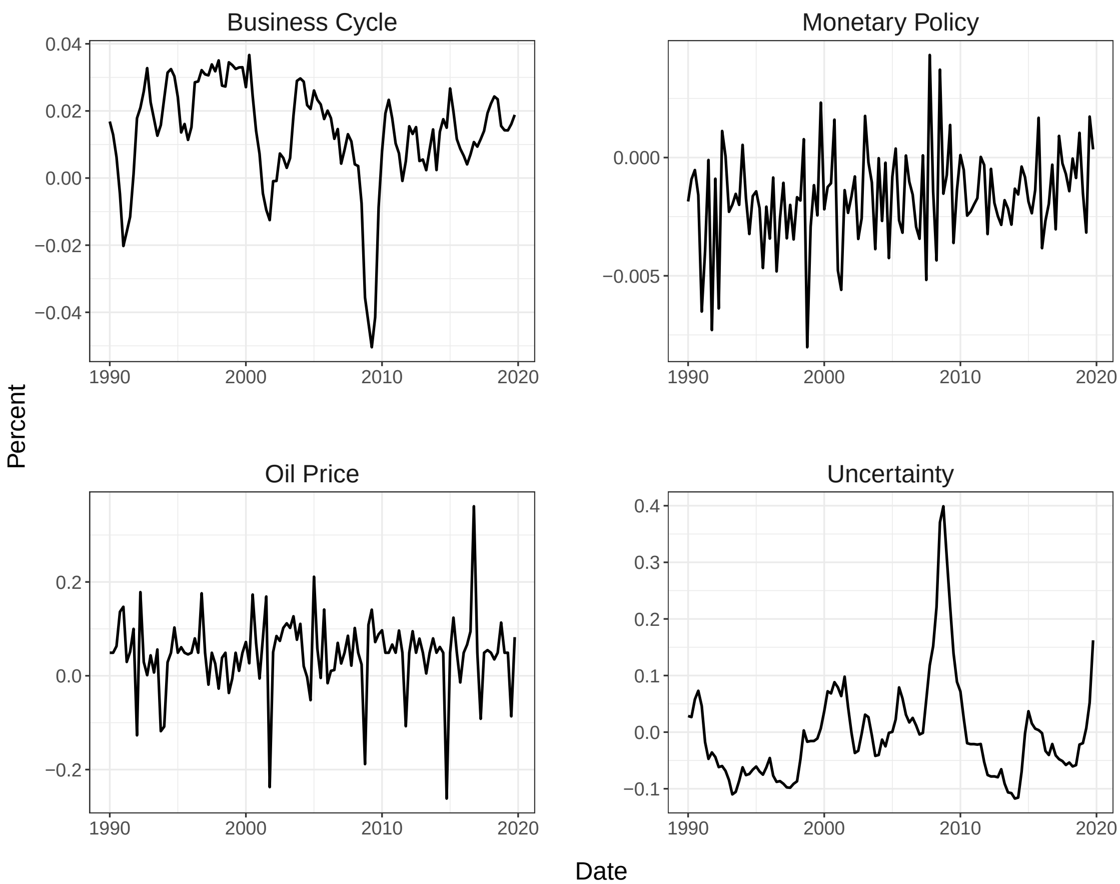
Notes: The figure shows the distribution of firm-quarter balance-sheet characteristics used as independent variables in the empirical application. The data are from quarterly Compustat, spanning from 1990-Q1 to 2019-Q4. Variables are trimmed at the 98.5th percentile and then linearly interpolated before the empirical application. Additional details on variable construction and data cleaning are provided in Appendix A.

Figure 8: Distribution of the dependent variables



Notes: The figure shows the distribution of firm-quarter balance-sheet characteristics used as dependent variables in the empirical application. The data are from quarterly Compustat, spanning from 1990-Q1 to 2019-Q4. Growth rates are annual and they are calculated using the Haltiwanger formula. Variables are trimmed at the 1.5th and 98.5th percentile before being used in the empirical application. Units of measurement are in percentage points, where 0.01 represents 1%. Additional details on variable construction and data cleaning are provided in Appendix A.

Figure 9: Time series of the aggregate fluctuations



Notes: The figure shows the time-series of the aggregate fluctuations and shocks used in the empirical application. Units of measurement are in percentage points, where 0.01 represents 1%. Additional information on the variable construction can be found in Appendix A.

B.4 Summary statistics and correlation matrix

Table 4: Summary statistics

	Statistics							
Variable	Mean	Median	St. Dev.	Min	Max	IQR	Skewness	Obs.
Panel A. Characteristics								
Size	0.63	0.64	2.41	-9.36	8.61	3.40	-0.01	448856
Leverage	0.29	0.26	0.21	0.00	1.00	0.28	0.90	339760
Liquidity	0.14	0.07	0.17	0.00	0.99	0.16	2.04	363361
Distance to Default	5.76	4.74	4.45	0.00	21.03	5.84	1.05	336085
Short-Term Debt	0.30	0.20	0.29	0.00	1.00	0.43	0.89	443857
ROA	-0.02	0.01	0.07	-0.46	0.08	0.04	-2.76	437471
Sales Volatility	0.27	0.20	0.23	0.00	1.12	0.25	1.52	378874
Panel B. Outcome								
Sales Growth	0.04	0.03	0.27	-1.03	1.07	0.22	-0.04	239625
Market Value Growth	0.01	0.03	0.51	-1.43	1.32	0.61	-0.19	214285
Investment Rate	0.07	0.02	0.28	-0.86	1.60	0.20	1.45	418937
Debt Rate	0.01	-0.02	0.45	-1.53	1.64	0.35	0.32	227627

Notes: The first panel contains the summary statistics for quarterly balance-sheet firm characteristics used as independent variables. The second panel contains the summary statistics for the outcome variables. The data are from quarterly Compustat, covering 1990Q1-2019Q4. All dependent variables are trimmed at the 1.5th and 98.5th percentiles, while independent variables are trimmed at the 98.5th percentile when positive. Independent variables are linearly interpolated after cleaning steps. Units of measurement of the outcome variables are in percentage points, where 0.01 represents 1%. Additional information on variable construction can be found in Appendix A.

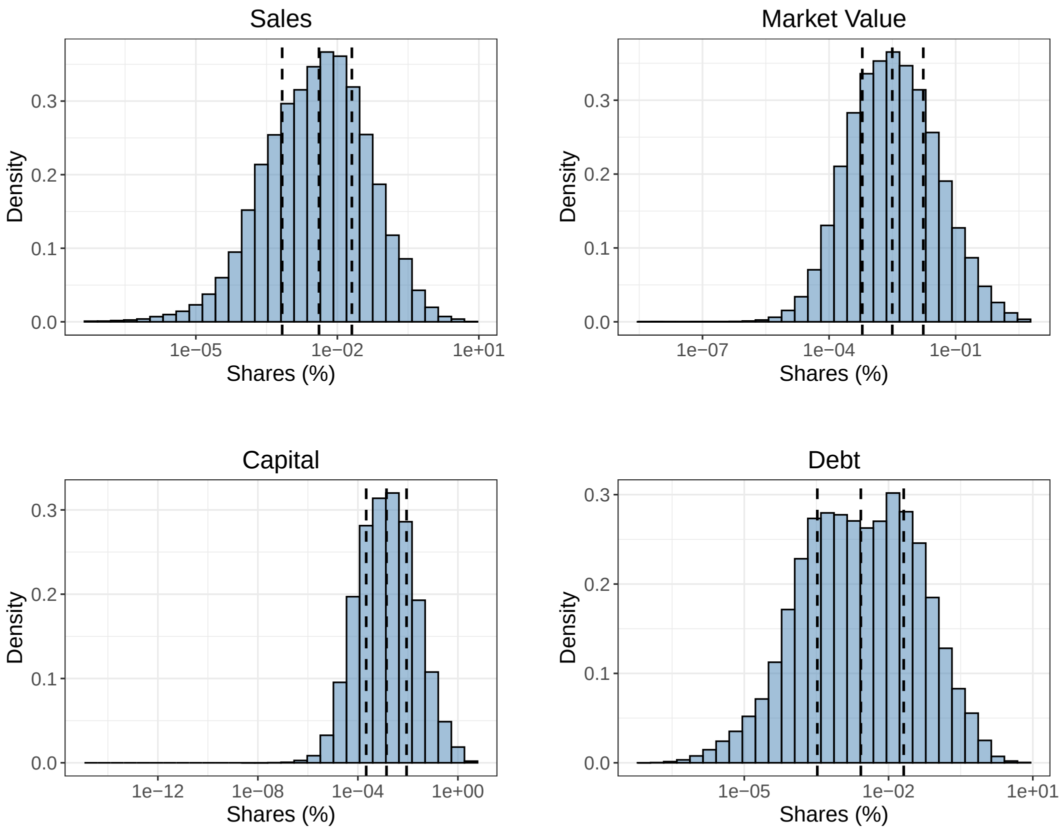
Table 5: Pairwise correlation matrix of balance sheet characteristics

Variable	Size	Leverage	Liquidity	Distance to Default	Short-Term Debt	ROA	Sales Volatility
Size	1.00	0.07	-0.19	0.33	-0.41	0.41	-0.40
Leverage	0.07	1.00	-0.31	-0.37	-0.22	-0.09	0.02
Liquidity	-0.19	-0.31	1.00	0.14	0.13	-0.26	0.31
Distance to Default	0.33	-0.37	0.14	1.00	-0.05	0.28	-0.23
Short-Term Debt	-0.41	-0.22	0.13	-0.05	1.00	-0.21	0.20
ROA	0.41	-0.09	-0.26	0.28	-0.21	1.00	-0.43
Sales Volatility	-0.40	0.02	0.31	-0.23	0.20	-0.43	1.00

Notes: The table contains the pairwise correlation statistics for quarterly firm balance-sheet characteristics used as independent variables. The data are from quarterly Compustat, covering 1990Q1-2019Q4. All independent variables are trimmed at the 98.5th percentile when positive. Independent variables are linearly interpolated after cleaning steps. Additional information on variable construction can be found in Appendix A.

B.5 Distribution of firms' shares

Figure 10: Distribution of the shares of outcome variables

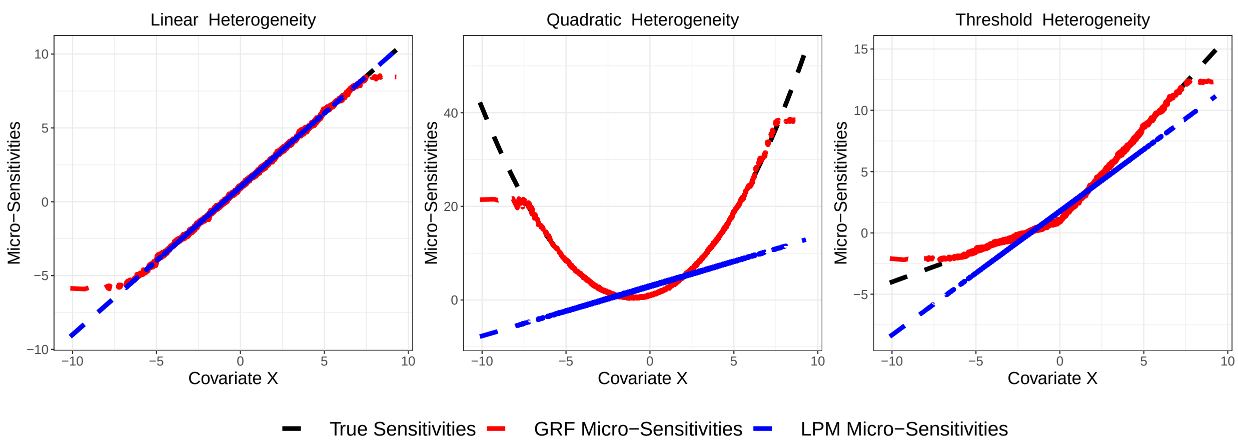


Notes: This figure presents the distribution of firm-level shares across different outcome variables. The x-axis represents the firm-level share on a log scale, while the y-axis denotes the density. The vertical lines indicate the first, second, and third quartiles of the distribution.

C Additional Figures and Tables - Firm level

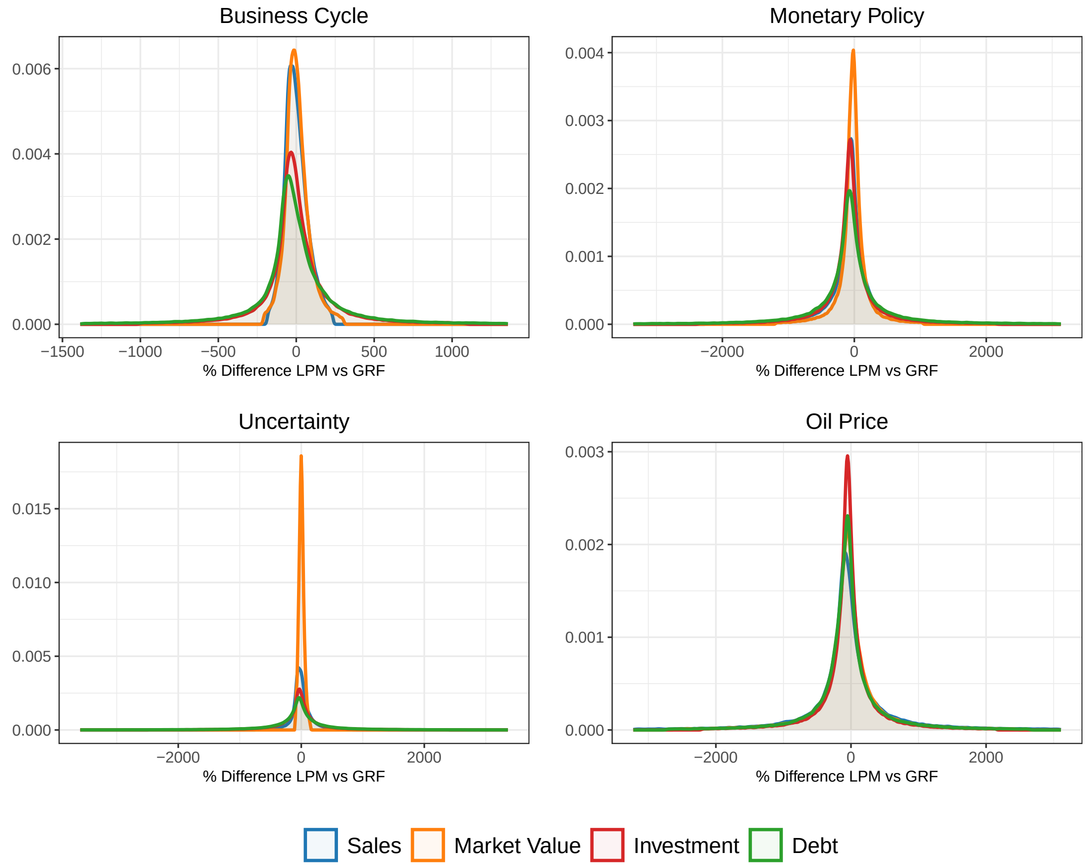
C.1 Comparing LPM and GRF sensitivities

Figure 11: Estimated sensitivities on simulated data



Notes: This figure illustrates the relationship between estimated sensitivities and the covariate X across different data-generating processes and one characteristic relevant for the heterogeneity ($J' = 1$). The sensitivities are estimated using the GRF and the LPM via OLS, and compared to the the true underlying heterogeneity. The black line represents the true sensitivities, while the red and blue lines correspond to GRF and LPM estimates, respectively. Results are based on a single simulation of a panel with 6,000 firms observed over 20 periods.

Figure 12: Density of differences between GRF and LPM sensitivities



Notes: The figure presents the kernel density estimates of the percentage difference between LPM and GRF firm-level sensitivities across four dependent variables: Sales, Market Value, Debt, and Investment. The x-axis represents the percentage difference between LPM and GRF estimates, calculated as $(\text{LPM}/\text{GRF} - 1) \times 100$. Each panel corresponds to a specific aggregate shock: business cycle, uncertainty, monetary policy, and oil price. The densities highlight the distribution of deviations for each dependent variable, with colors indicating the specific variable. Differences are trimmed at 2.5% on both sides.

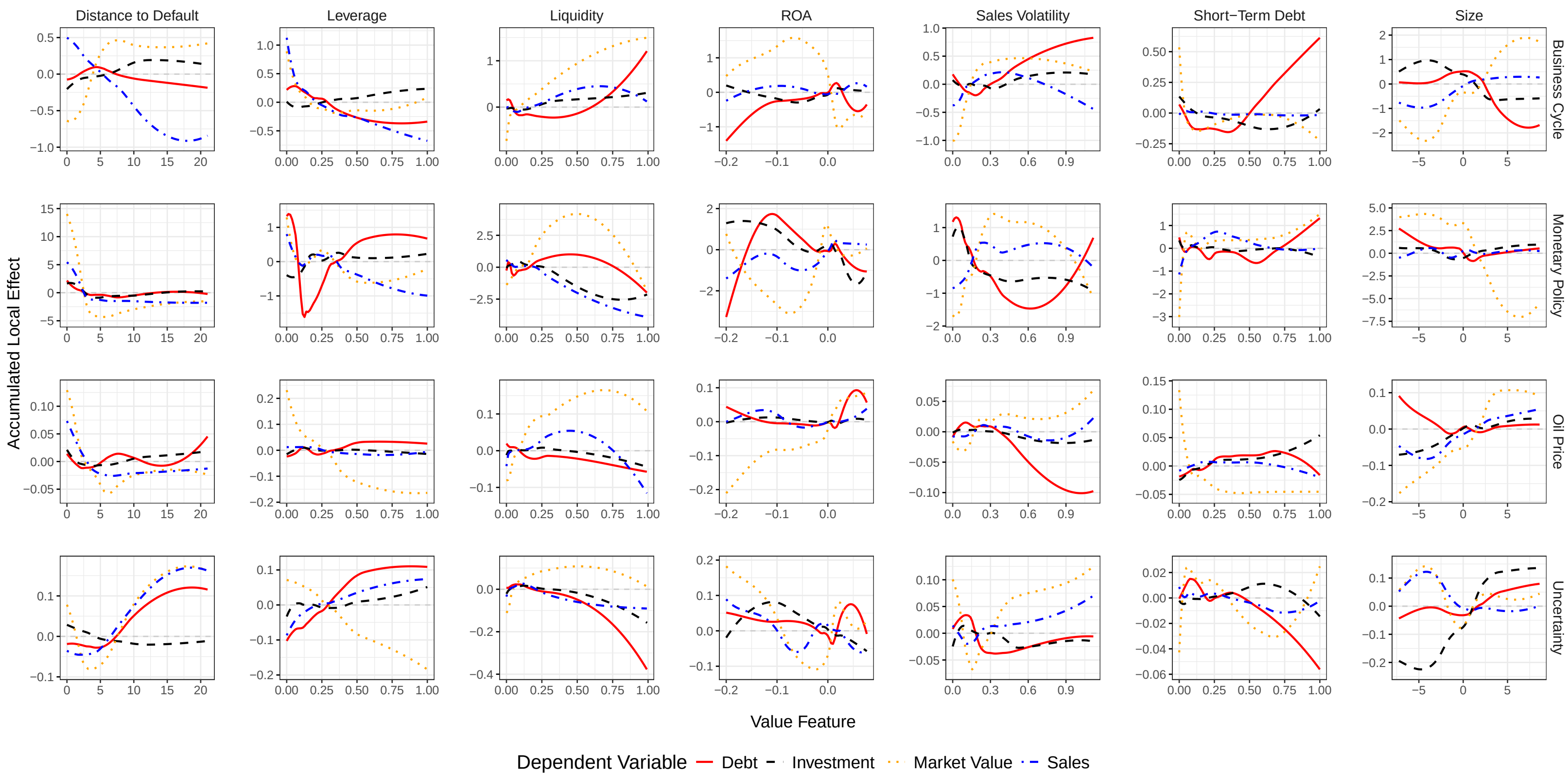
C.2 Non-linearities and heterogeneity

Table 6: Statistical test for non-linearity

	Harvey-Collier Test		RESET Test		GAM
Outcome variable	Statistic	P-Value	Statistic	P-Value	Min EDF
Panel A: Business Cycle					
Sales	1.95	0.05	604.78	0.00	7.64
Market Value	14.03	0.00	2439.42	0.00	7.87
Investment	8.81	0.00	2554.48	0.00	7.02
Debt	7.01	0.00	3786.25	0.00	5.73
Panel B: Monetary Policy					
Sales	8.59	0.00	10479.23	0.00	7.41
Market Value	18.93	0.00	5384.36	0.00	7.00
Investment	3.91	0.00	243.82	0.00	7.51
Debt	4.58	0.00	193.67	0.00	7.59
Panel C: Uncertainty					
Sales	15.01	0.00	669.88	0.00	6.64
Market Value	14.69	0.00	1928.16	0.00	7.86
Investment	14.02	0.00	3156.06	0.00	6.10
Debt	5.85	0.00	289.32	0.00	6.88
Panel D: Oil Price					
Sales	4.19	0.00	492.47	0.00	7.27
Market Value	9.30	0.00	2939.43	0.00	7.46
Investment	1.11	0.27	152.68	0.00	7.66
Debt	2.52	0.01	47.78	0.00	7.35

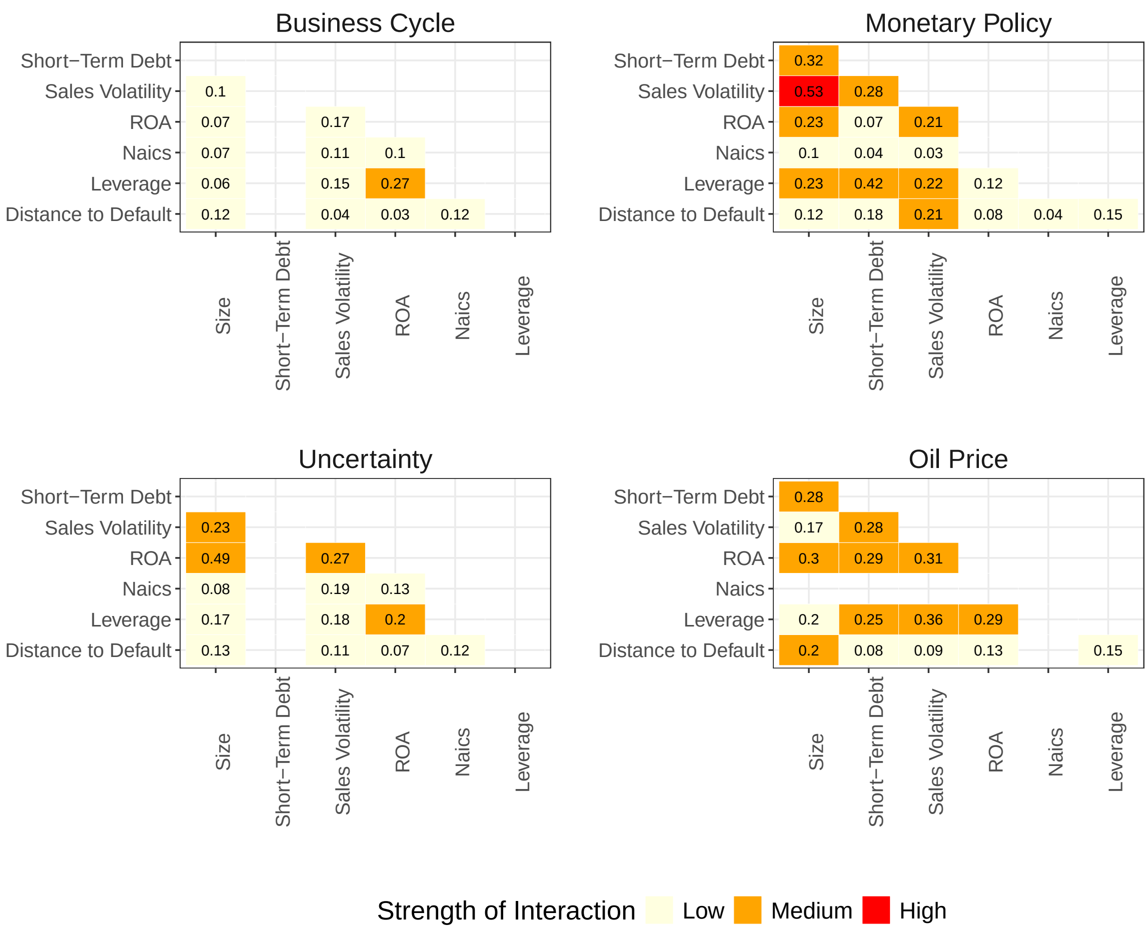
Notes: The table reports the results of three different linear specification tests between covariates and the conditional average sensitivities produced by GRF for each outcome variable across four shocks. We assess the linearity of the conditional effect of an aggregate shock on firms' outcome in firms' characteristics, i.e. $b(X_{i,t-1}) = \beta_0 + \sum_{j \in J} \beta_j \cdot X^j_{i,t-1}$, where J is the set of characteristics. The null hypothesis of Harvey-Collier Test and the RESET Test is that the model is linear. For both tests, we report the test statistics and p-value of the test. We estimate a GAM model that includes all characteristics. For each characteristic we estimate a the effective degrees of freedom (EDF). We report the minimum effective degrees of freedom among characteristics in each outcome variable - aggregate shock. Results are presented for debt, investment, market value, and sales under each aggregate shock (business cycle, uncertainty, monetary policy, and oil price).

Figure 13: Accumulated Local Effects



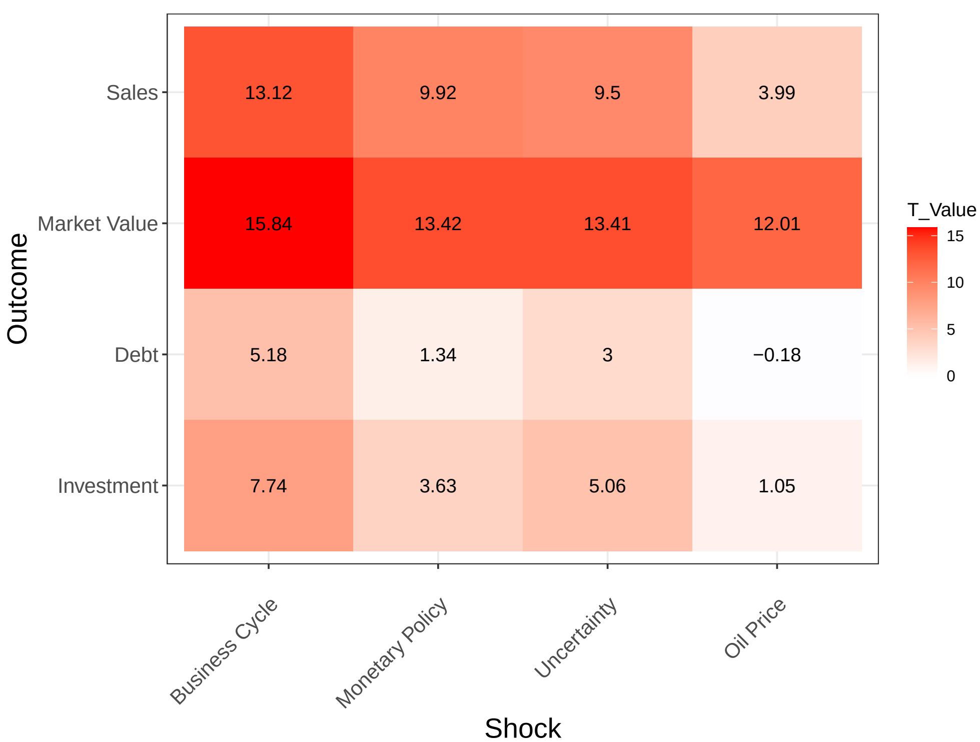
Notes: The figure presents Accumulated Local Effects (ALE) estimated for each firm characteristic across different aggregate shock-outcome variable pairs. Each row corresponds to an aggregate shock. Each column to a balance-sheet characteristics. The solid red lines represent debt, the black dash lines represent investment, the dot orange lines represent market value, and the blue dash-dot lines represent sales. The x-axis represents the percentile distribution of the characteristic, while the y-axis shows the estimated difference in firm sensitivity relative to the average firm sensitivity.

Figure 14: Pairwise strength of interactions



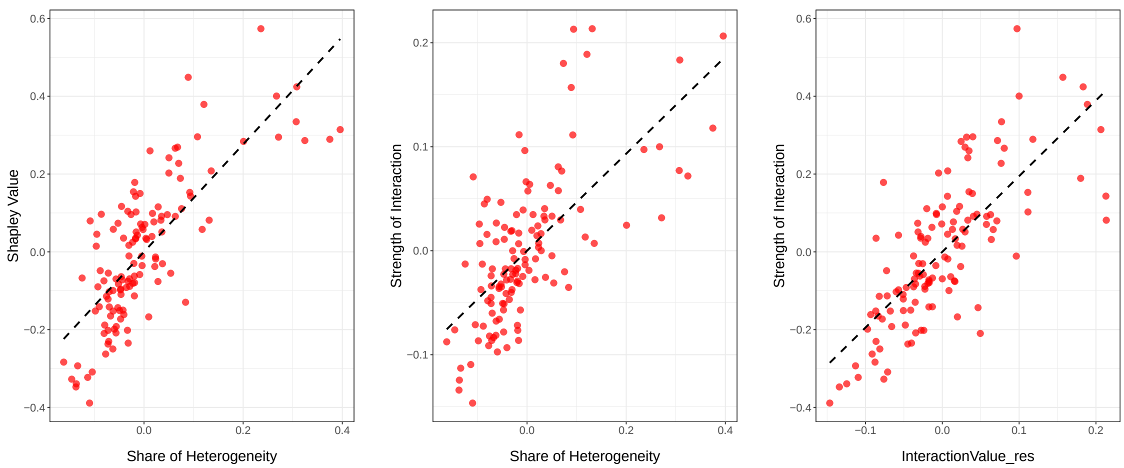
Notes: This heatmap visualizes the average pairwise strength of interaction between firm characteristics for each aggregate shock. We measure the strength of interaction of each pair of characteristic using the pairwise Friedman's H-statistic. Each panel corresponds to a specific shock (e.g., business cycle, uncertainty, monetary policy, and oil price). For each pair of characteristics, interaction values are averaged across outcome variables. For each outcome variable - aggregate shock pair, we consider the ten strongest pairwise interactions. Interaction strength is categorized into three ranges: low (0–0.2), medium (0.2–0.5), and high (0.5+). The ranges are determined based on commonly observed thresholds in machine learning literature and are tailored to highlight meaningful variation in the dataset. The x-and y-axes denote the interacting characteristics, and the color scale indicates the strength of the interaction. The diagonal is omitted as it represents the self-interaction of a characteristic, which is not defined in this context.

Figure 15: Test for heterogeneity in sensitivity



Notes: The table reports the t-statistic of the Chernozhukov et al. (2018) test for each aggregate shock - outcome variable pair. An absolute t-statistic value below 1.648 indicates no particular degree of heterogeneity, while a value above the threshold of 1.648 suggests a statistical high level of heterogeneity in firm sensitivity at a 90% confidence interval. Appendix A provides additional details on the test.

Figure 16: Correlation between importance measures

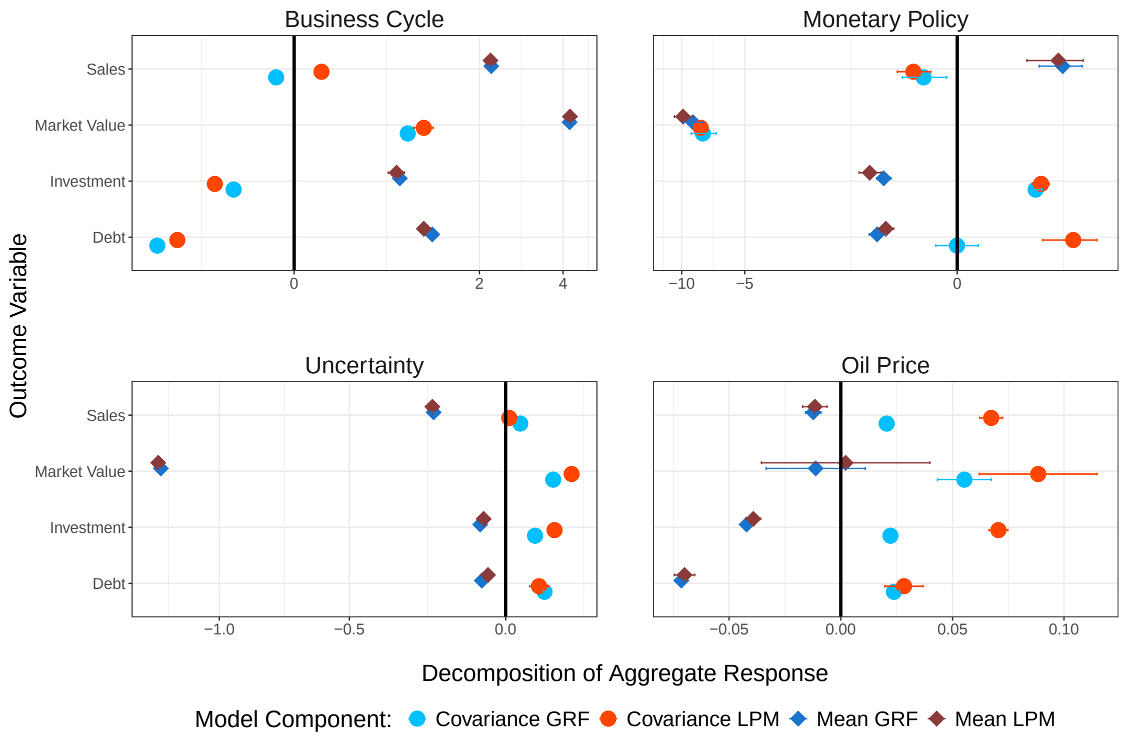


Notes: The figure reports the scatter plots between the share of heterogeneity explained by each characteristics and the Shapley-based measure of relevance (left panel) and the strength of interaction measure (central panel), and between the strength of interaction measure and the Shapley-based measure of relevance (right panel). The share of heterogeneity explained by each characteristics is computed as the depth-weighted frequency of splits in the forest where the characteristic is used. We compute Shapley values for each characteristic in all outcome variable-aggregate shock pairs over a grid of 100 points corresponding to the characteristic's percentiles. We compute the mean absolute value of the estimated Shapley values over the hundred points. We normalize importance by scaling each characteristic to the highest mean absolute Shapley value within each outcome variable - aggregate shock pair, setting the maximum to one. We measure the strength of interaction of each characteristic using the Friedman's H-statistic against all other characteristics. In all cases, we absorb aggregate shocks, outcome variables, and characteristic fixed effects. Black dashed lines represent a linear fit.

D Additional Figures and Tables - Aggregate

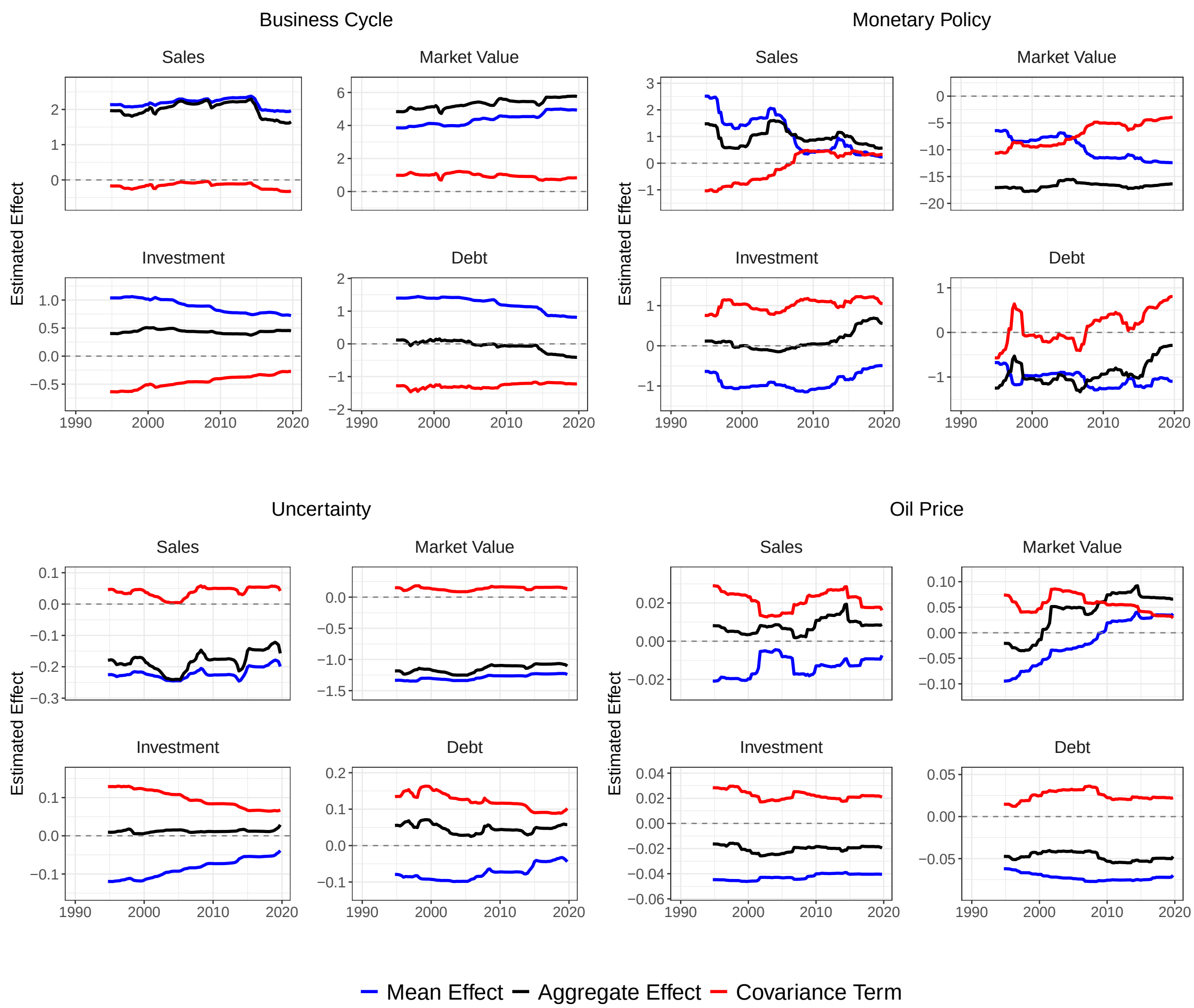
D.1 Mean-covariance decomposition across models

Figure 17: Comparison Mean - Covariance Decomposition



Notes: The figure decomposes the average aggregate response into mean and covariance components for both GRF and linear panel model. Each point represents the estimated coefficient from Equation (9), with the mean term capturing the average firm-level sensitivity (squared markers) and the covariance term reflecting the interaction between firm shares and sensitivities (circle markers). Blue markers denote estimates from the GRF model, while red markers correspond to estimates from the LPM. Error bars indicate 95% confidence intervals based on robust standard errors.

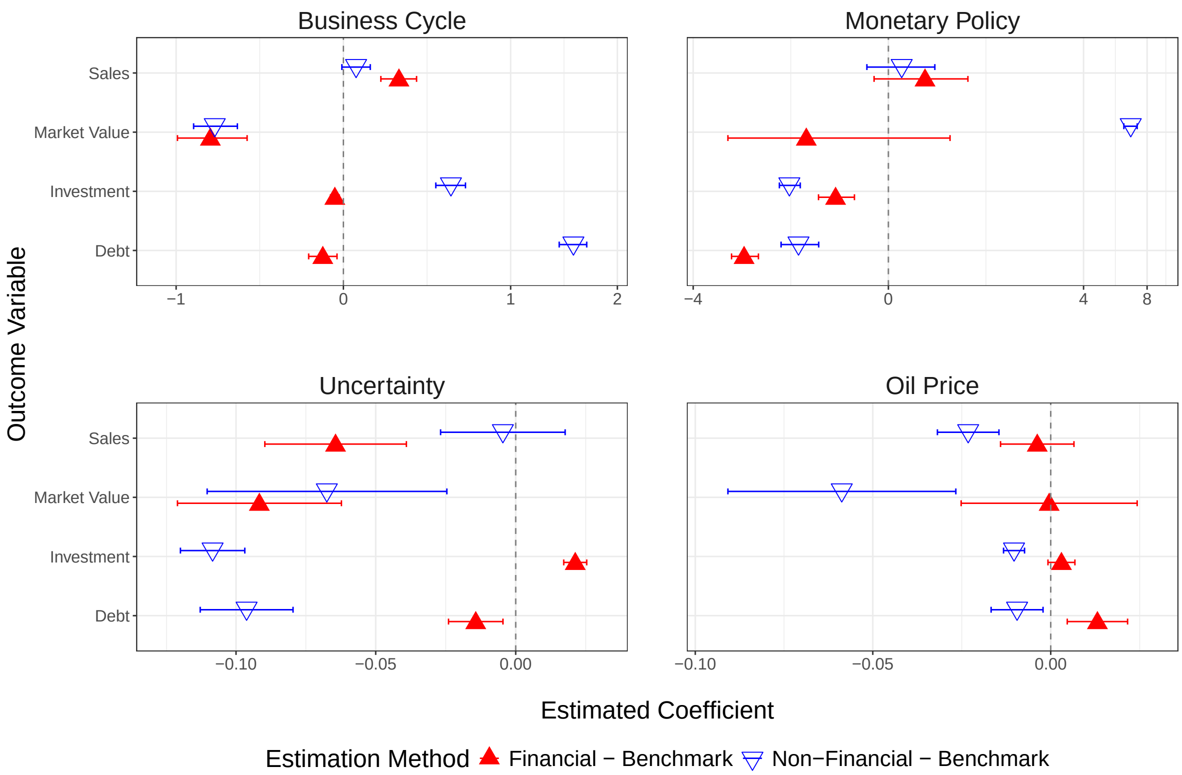
Figure 18: Aggregate Response Decomposition Over Time



Notes: The figures illustrate the mean and covariance decomposition of the average aggregate response across all outcome variable - aggregate shock pairs, utilizing a five-year rolling window version of Equation (9). Specifically, we estimate the time-serie model with the mean and covariance components, as defined in Equation (8), serving as the dependent variable Z_t. Each point in the time series represents the corresponding coefficient estimate, derived from a sample ending at the respective quarter and spanning the preceding five years. The mean and covariance components are calculated based on the benchmark set of firm-level sensitivities.

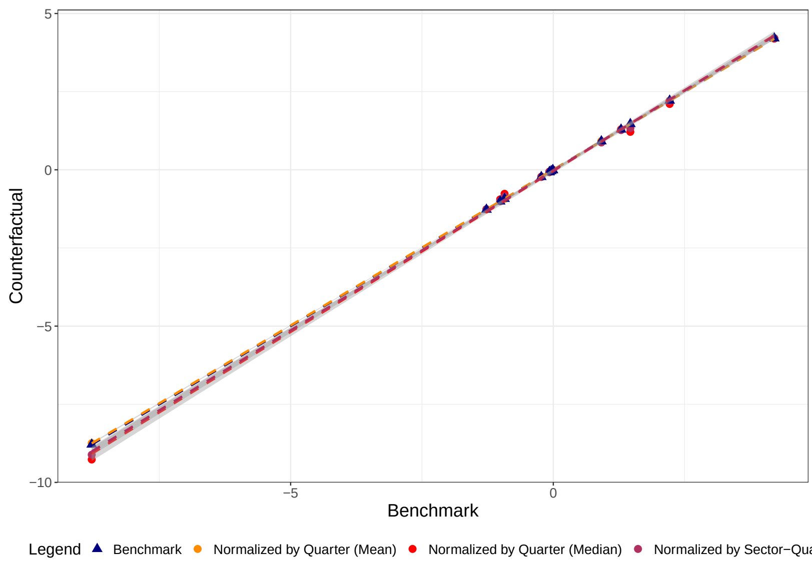
D.2 Heterogeneity in sensitivities

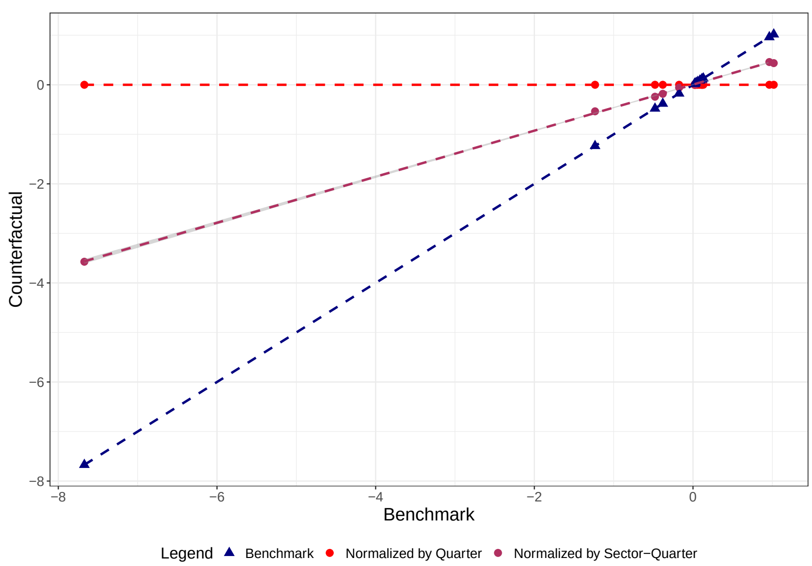
Figure 19: Role of financial and non-financial heterogeneity



Notes: The figure illustrates the contribution of financial and non-financial heterogeneity to the average aggregate response. To isolate their respective roles, we construct two counterfactual firm-level sensitivities and aggregate series: one where financial characteristics are fixed at quarter median while non-financial characteristics vary, and another where non-financial characteristics are fixed while financial characteristics vary. We then compute the aggregate response by weighting these sensitivities by firms' shares and estimate the average aggregate response using the time-series model in Equation (9). The red and blue triangles indicate respectively the differences between the counterfactual aggregate responses based on financial and non-financial characteristics relative to the benchmark average aggregate response. Standard error are clustered at quarterly level.

Figure 20: Within and across sector heterogeneity





Notes: This figure presents a comparison of firm-level sensitivity estimates under different counterfactual scenarios. The benchmark sensitivity is computed using the original firm-level estimates, while the counterfactual sensitivities are obtained by normalizing firm responses across different dimensions. In the left panel, we compare the covariance components across the benchmark, normalized by quarter, and normalized by sector-quarter specifications. The right panel displays the corresponding comparison for the mean effect. The fitted lines represent linear approximations of the relationship between the benchmark and counterfactual estimates. A lower covariance or mean effect in the counterfactual scenarios indicates that firm-level heterogeneity plays a significant role in shaping aggregate responses.