Market Power, Expectations, and Asset Prices*

Kevin Ren† Dalton Rongxuan Zhang‡

First Version: July 25, 2025
This Version: November 18, 2025
Click here for the latest version

Abstract

Recent asset pricing research highlights subjective expectations of long-run profit growth as a key driver of return variation. We show that these expectations—and their deviations from full-information rational expectations—systematically depend on firms' market power. Using the U.S. Census Longitudinal Business Database combined with Compustat/CRSP and IBES analyst forecast data, we find that firms with low total market power is overreact the most, while firms with high total market power overreact the least. We also document that firm-level overreactions are an order of magnitude larger than overreaction at the aggregate level. A macro-finance model incorporating these heterogeneities reproduces core asset pricing moments—high excess returns, volatility, and predictable reversals—and shows that distorted expectations, interacting with heterogeneous market power, generate procyclical misallocation and amplify macro-financial fluctuations.

*We would like to thank our advisors George-Marios Angeletos, Lawrence Christiano, Nicolas Crouzet, Martin Eichenbaum, Diego Känzig, Dimitris Papanikolaou, Matthew Rognlie, and Bryan Seegmiller for their guidance and support. We would also like to thank Efraim Benmelech, Vivek Bhattacharya, Joel David, Maarten De Ridder, Nicola Gennaioli, Erik Hurst, Gaston Illanes, Rafael La Porta, Kiminori Matsuyama, Konstantin Milbradt, Cristoforo Pizzimenti, Raul Guarini Riva, Jonathan Rothbaum, Kunal Sangani, Lawrence Schmidt, Aditya Soenarjo, Adi Sunderam, and Brandon Zborowski for helpful discussions and comments. The U.S. Census Bureau has reviewed this data product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data used to produce this product (Data Management System (DMS) number: P-7503840 (P-7518691), Disclosure Review Board (DRB) approval numbers: CBDRB-FY25-SEHSD003-016, CBDRB-FY25-0271, CBDRB-FY25-0359).

†Kellogg School of Management, Northwestern University. kevin.ren@kellogg.northwestern.edu

‡Department of Economics, Northwestern University. r.zhang@u.northwestern.edu

"The single most important decision in evaluating a business is pricing power. If you've got the power to raise prices without losing business to a competitor, you've got a very good business."
— Warren Buffett[1]

1 Introduction

Recent research in finance and macroeconomics has moved beyond the fully rational benchmark to incorporate behavioral biases such as overreaction and distorted beliefs. In asset pricing, this shift has also brought new attention to subjective cash-flow expectations as a key driver of returns, whereas earlier work emphasized discount-rate movements. A growing body of evidence shows that overreaction varies across asset classes and industries, with important implications for macro-finance dynamics. In this paper, we examine how overreaction varies at the firm level with respect to the key source of those cash flows, market power, and study the implications of the joint distribution of overreaction and market power for macro-finance outcomes.

We answer this question empirically and quantitatively. We construct a dataset merging CRSP/Compustat, the U.S. Census Longitudinal Business Database (LBD) restricted-use microdata, and IBES equity-analyst earnings-growth forecasts. We use these data to estimate the key parameters of a simple model of diagnostic expectations. These parameters map directly into forecast-error and belief-revision regressions, which allow us to identify them from reduced-form evidence. We then use these estimates to calibrate a simple macro-finance model with heterogeneity in both market power and overreaction, allowing us to examine its fit with the data and its implications for key macroeconomic and financial variables.

Empirically, we document that firms with low market power exhibit the greatest overreaction and most persistent belief distortions, while firms with high market power exhibit the least. These patterns are validated in return-predictability tests: low market power firms display economically and statistically significant return reversals, whereas high market power firms display much weaker reversals, consistent with the belief distortions. We also find that overreaction at the firm level is substantially larger than overreaction at the aggregate level. Quantitatively, the model with heterogeneity in market power and diagnostic expectations generates key macro-finance moments, including elevated excess returns, predictable return reversals, the correct cross-sectional patterns. Furthermore, the model can generate procyclical misallocation which arises from the joint distribution of market power and overreaction.

We categorize firms into three types: high total market power, baseline, and low total market power. We measure total market power using the total wedge, which is defined as the product of the price markup (product market power) and wage markdown (labor market power). Using a novel approach that exploits stock-level cross-sectional variation in overreaction we can then document empirically that expectations are systematically distorted and co-vary with market power.

[1] Buffett (2010), interview with the Financial Crisis Inquiry Commission.

To examine if overreaction varies across market power, we study if forecast-error predictability and belief revisions vary across firms with different levels of market power. We regress realized forecast errors of long-term earnings growth (LTG) onto the lagged beliefs and the most recent belief revision of LTG and in a separate specification we regress belief revisions onto lagged beliefs. These regressions allow us to measure the magnitude and persistence of belief distortions. Under rational expectations, forecast errors should be unpredictable and belief revisions should mirror the true statistical process of the forecasted variable. We find that firms with low total market power exhibit more predictable forecast errors and significantly slower belief updating than implied by the true statistical process of earnings growth. Conversely, firms with high total market power display more attenuated forecast-error predictability and faster belief revisions that are still too slow under rational expectations, but notably faster than for low total market power firms.

We quantify these patterns by jointly estimating the parameters of a simple diagnostic expectations model across firm types following Bordalo et al. (2024*a*). This specification of diagnostic expectations provides closed-form mappings between the regression coefficients and the key behavioral parameters, allowing us to directly interpret the empirical moments. We find that low total market power firms exhibit the strongest overreaction and the most persistent belief distortions, whereas high total market power firms display weaker but still significant overreaction and the least persistence; baseline firms fall between these two extremes. Firm-level overreaction is also substantially larger than the aggregate-level overreaction documented in prior research. For example Bordalo et al. (2024*a*) find an overreaction parameter estimate of 0.8 for the aggregate stock market using aggregate expectations whereas we find estimates ranging from 5 to 8 using individual-level expectations.

Furthermore, we find cross-sectional evidence on return dynamics that is consistent with these the empirical findings and our structural estimates. Low total market power firms experience sharper return reversals following new information, while high total market power firms exhibit weaker and shorter-lived reversals. These return patterns mirror the cross-sectional differences in forecast errors and belief revisions. These results align closely with our structural estimation, which reinforces the link between market power, expectations, and return behavior.

The empirical evidence raises a natural question: can a simple macro-finance model that embeds the observed heterogeneity in expectations and market power reproduce key asset-pricing and macroeconomic facts? We address this by constructing a model that matches three core empirical features: (i) high excess returns and return volatility, (ii) predictable return reversals, and (iii) realistic cross-sectional patterns across firms. Taking as given the distribution of overreaction and market power estimated in the data, we also ask what these features imply for aggregate asset prices and real outcomes.

In the model, firms differ along two key dimensions reflecting our empirical design and results. The first is heterogeneous market power, capturing variation in both product and labor markets. The second is heterogeneous overreaction, which varies systematically with market power and is characterized by differences in the magnitude and persistence of belief distortions as documented empirically. For tractability, the model is set in partial equilibrium and features no discount-rate variation. As a result, asset-return dynamics arise solely from variation in cash flows and their subjective expectations. This design directly tests the viability of the emerging view that fluctuations in subjective cash-flow expectations, rather than discount rates, account for a large share of observed return dynamics.

Our model successfully replicates key asset-pricing moments that the rational expectations benchmark fails to match. Despite its simple structure and fixed discount rates, it generates high excess returns and return volatility purely through variation in subjective cash-flow expectations. The impulse response functions (IRFs) reveal predictable return reversals following positive productivity shocks, consistent with the empirical evidence on overreaction and subsequent mean reversion in beliefs and valuations. Decomposing the return responses shows that both heterogeneity in market power and heterogeneity in overreaction contribute to these dynamics, though the majority of the variation is driven by overreaction heterogeneity. By contrast, the rational expectations model cannot reproduce these predictable reversals.

To understand the underlying mechanism, we examine the model's real-side dynamics. Diagnostic expectations induce stronger and more persistent real responses, producing hump-shaped patterns in output and profits that qualitatively match empirical macroeconomic IRFs (e.g. Auclert, Rognlie and Straub, 2020, 2024). Capital investment, in particular, exhibits a pronounced boom-bust cycle that serves as the main real driver of both elevated return volatility and predictable reversals. In contrast, the rational expectations model generates much flatter and less persistent dynamics, underscoring the amplifying role of belief distortions in shaping both asset prices and real activity that qualitatively match empirical evidence.

Beyond asset pricing moments, understanding the allocation of capital across firms is central to evaluating the broader macroeconomic consequences of belief distortions. Our model shows that the interaction between distorted expectations and heterogeneous market power generates misallocation not only in the steady state but also in response to aggregate shocks. Moreover, the correlation between overreaction, productivity, and market power determines the direction of the effect: if firms that overreact more also have lower market power, misallocation is procyclical; if the opposite correlation holds, it becomes countercyclical. Following a positive productivity shock, the model with diagnostic expectations and heterogeneous firms exhibits a peak relative productivity loss of -0.33% compared to the efficient allocation. By contrast, a rational expectations model with heterogeneous firms generates a smaller peak loss of -0.25%. The diagnostic-expectations model implies more than a 30% larger misallocation. These results underscore the amplifying role of belief distortions when combined with firm-level market power.

To further isolate the role of heterogeneous belief distortions, we consider a model with diagnostic expectations and heterogeneous market power but without heterogeneity in expectation parameters. This specification yields a slightly smaller peak misallocation of -0.32%, suggesting that most of the amplification arises from the presence of diagnostic expectations itself, while heterogeneity in belief distortions adds a modest incremental effect. Conversely, when firm heterogeneity is shut down and diagnostic expectations are applied to a representative firm, the relative misallocation disappears entirely. This highlights the importance of heterogeneous market power in generating misallocation: firms with low market power respond more aggressively to shocks—both in fundamentals and beliefs—while high market power firms adjust more slowly. Under diagnostic expectations, this imbalance is amplified as investors overreact to firm-level news, further distorting investment flows and increasing misallocation over the cycle.

The results of this paper highlight the importance of the interaction between market power and behavioral distortions. While each has been studied extensively in isolation, their interaction has received little attention despite its central role in shaping both asset-pricing dynamics and real outcomes. We show that overreaction varies systematically with market power and that a model incorporating this joint heterogeneity can replicate key asset-pricing and macroeconomic moments that a rational-expectations benchmark fails to match without introducing additional structural complexity. Moreover, our framework demonstrates that the interaction between heterogeneous market power and heterogeneous diagnostic expectations is essential to generate procyclical misallocation. Heterogeneous overreaction amplifies the misallocation effects of market power over the business cycle, and the direction of the misallocation depends critically on the covariance between overreaction and market power.

Related Literature. The growing shift in incorporating behavioral biases in general in finance and macroeconomics has led to the development of expectations formation such as diagnostic expectations. Diagnostic expectations is a generalization of rational expectations that generates overreaction. There is also a shift to study the effects of cash-flow news and expectations for asset pricing. Since the seminal findings of Shiller (1981) and Campbell and Shiller (1988), the literature has emphasized discount rate variation as the main source of aggregate asset return fluctuations.[2] Studies such as De La O and Myers (2021) and Bordalo et al. (2024*b*,*a*) show that investors' subjective expectations of future cash flows can explain a large share of aggregate stock returns. They also show that expectations deviate systematically from full-information rational expectations (FIRE) and exhibit overreaction. We add to this literature by examining firm-level overreaction and firm-level returns as well as study how it varies by market power. We discuss below in more detail how this paper contributes to three core strands of literature.

This paper contributes to a growing literature that studies and incorporates diagnostic expectations in finance and macroeconomics (e.g., Bordalo, Gennaioli and Shleifer, 2018; Bordalo et al., 2019, 2020, 2024*b*,*a*; Gennaioli, Ma and Shleifer, 2016; Maxted, 2024; Bianchi, Ilut and Saijo, 2024; Ilut and Valchev, 2023; De La O and Myers, 2021; De la O and Myers, 2024; Delao, Han and Myers,

[2]See Campbell and Ammer (1993) and Cochrane (2011) for a detailed review. Vuolteenaho (2002) shows that at the firm-level cash flow news is a more significant driver of returns even under rational expectations.

2025). We introduce a cross-sectional perspective by estimating firm-level overreaction and relating it to observable firm characteristics, namely the degree and type of market power. Firms with low market power exhibit stronger overreaction, while those with high market power exhibit weaker responses. We incorporate these patterns into a stylized macro-finance model with heterogeneous firms and diagnostic expectations, which generates excess return volatility, return predictability, and real-side volatility in a standard production setting. This approach links belief distortions not only to asset pricing moments but also to variation in firm behavior and aggregate productivity dynamics.

Our paper also contributes to a strand of literature that examines the relationship between market power and asset pricing (e.g., Hou and Robinson, 2006; Hoberg and Phillips, 2010; Gu, 2016; Bustamante and Donangelo, 2017; Iraola and Santos, 2017; Kogan, Papanikolaou and Stoffman, 2020; Corhay, Kung and Schmid, 2020; Corhay, Li and Tong, 2022; Seegmiller, 2023; Corhay, Kung and Schmid, 2025). In particular, Hoberg and Phillips (2010) show that in competitive industries high valuations and investment are followed by lower cash flows and abnormal stock returns and that analyst estimates are positively biased. In more less competitive industries however, Hoberg and Phillips (2010) find that these relations are weak and not statistically significant. These results suggest that participants in competitive industries to not fully internalize competition dynamics on cash flows and stock returns leading to these predictable errors. Furthermore, Rhodes-Kropf and Viswanathan (2004), Rhodes–Kropf, Robinson and Viswanathan (2005), and Shleifer and Vishny (2003) show that both the stock market and managers misprice firms and that leads to merger waves. These results also imply they misinterpret the competitive environment. Together, these findings show that the strength of belief distortions vary systematically with firms' competitive environments. We build on this work by incorporating both diagnostic expectations and firm-level characteristics documented by Ren and Zhang (2025) to study the joint interaction of behavioral biases, market power, and asset prices at the firm-level. In our model, belief distortions alone generate sizable excess returns, return volatility, and return predictability, even without variation in discount rates. Moreover, introducing heterogeneous firm-level market power amplifies the effects of diagnostic expectations on asset prices and gives rise to more substantial procyclical misallocation over the business cycle.

Finally, this paper contributes to the literature that studies market power in the macroeconomic context (e.g., Syverson, 2011; De Loecker, Eeckhout and Unger, 2020; Farhi and Gourio, 2018; Gutiérrez and Philippon, 2017; Autor et al., 2020; Peters, 2020; De Ridder, 2024; Atkeson and Burstein, 2008; Traina, 2018; Crouzet and Eberly, 2021; Helpman and Niswonger, 2022; Yeh, Macaluso and Hershbein, 2022; Berger, Herkenhoff and Mongey, 2022; Crouzet and Eberly, 2023; Akcigit and Ates, 2023). This literature typically examines the rise and heterogeneity of firm-level market power, their implications for productivity, labor market outcomes, and investment, and their role in driving long-run macroeconomic trends such as declining labor shares and reduced business dynamism. We contribute to this literature by incorporating the interaction of diagnostic expectations and asset prices with market power, and show that diagnostic expectations can amplify procyclical misallocation resulting from heterogeneous market power.

Outline. Section 2 describes the datasets used as well as how price markups and wage markdowns are estimated. Section 3 presents the model of diagnostic expectations and the empirical evidence of firm-level heterogeneity. Section 4 discusses the comparative statics of the diagnostic expectations model's empirical predictions and estimates the model's parameters. Section 5 presents a macro-finance model that incorporates the calibrated diagnostic expectations. Section 6 conducts various quantitative exercises with the macro-finance model. Finally, Section 7 concludes.

2 Data and Market Power Estimation

The datasets we utilize build on the firm-level panel developed by Ren and Zhang (2025). We augment it with the Institutional Brokers' Estimate System (IBES) Unadjusted U.S. Summary Statistics file, which contains analyst forecasts for U.S. publicly traded firms. The base dataset from Ren and Zhang (2025) is an annual panel that merges the Longitudinal Business Database (LBD), which a restricted-use business establishment dataset from the U.S. Census Bureau, with the annual CRSP/Compustat Merged database. Then Ren and Zhang (2025) proceed to estimate price markups and wage markdowns using this dataset and construct a firm-year panel of firm-level market power. We extend this dataset in two ways: first, by merging in year-end IBES observations to construct an annual panel; and second, by creating a monthly panel that retains the IBES frequency and links in firm-level market power and characteristics at the annual level. We describe these datasets in detail in Section 2.1 and summarize the estimation approach in Section 2.2. Additional details on data construction and estimation are provided in Appendix B.

2.1 Data Sources

The IBES dataset is a monthly panel containing analyst forecasts for U.S. publicly traded firms (Center for Research in Security Prices, 2025). The key variable of interest in our analysis is the median analyst forecast of a firm's earning per share (EPS) long-term growth (LTG). The LTG is defined as the "expected annual increase in operating earnings over the company's next full business cycle. These forecasts refer to a period of between three to five years." The LTG variable is available beginning in December 1981. We follow the procedure of Bordalo et al. (2024*a*) to construct the monthly firm-level panel of median LTG forecasts.[3] We also follow their procedure to construct aggregate-level LTG measures.

There are potential concerns with using analyst forecasts such as they may be distorted by agency conflicts or may not accurately reflect the expectations of investors or firms. First, to mitigate

[3]We thank Rafael La Porta and Nicola Gennaioli for assistance with the replication package.

the influence of idiosyncratic biases and outliers, we use the median forecast across analysts for each firm. Moreover, prior research suggests that, despite predictable forecast errors, sell-side analysts exert substantial effort and have strong incentives to produce accurate forecasts. van Binsbergen, Han and Lopez-Lira (2023) show that analyst forecasts contain valuable information for predicting future earnings, even relative to a machine-learning algorithm trained on an extensive set of publicly available variables. Grennan and Michaely (2020) provide complementary evidence that analysts actively gather firm-specific information such as by asking questions during earnings calls and meeting with management and institutional investors. More broadly, Kothari, So and Verdi (2016) survey the literature linking analyst forecasts and asset pricing and conclude that reputational concerns provide analysts with strong motivation to issue credible and accurate forecasts.

A second concern is that analyst forecasts may not represent the expectations of investors or firms. Empirical evidence suggests otherwise. Bordalo et al. (2019) document that analysts not only learn about firm fundamentals based on past performance but also share investors' overreaction to new information, as reflected in the joint dynamics of returns, expectations, and realized earnings growth. De La O and Myers (2021) show that short-term analyst earnings expectations closely track short-term cash-flow growth. Finally, Gennaioli, Ma and Shleifer (2016) and Bordalo et al. (2024*b*) find that CFO expectations co-move with analyst expectations and display similar deviations from rational expectations, including overreaction. Taken together, these findings suggest that analyst forecasts provide an informative and behaviorally consistent proxy for the beliefs that shape both firm decisions and market prices.

The next key data source is the LBD, which is an establishment-level annual census of the U.S. non-farm private sector (United States Census Bureau, 2022). We collapse the LBD to the firm-level using the Census-provided firm identifiers. The version we use has coverage from 1976 to 2019. Chow et al. (2021) provide more information on the construction of the LBD. We merge the firm-level LBD to the annual Compustat/CRSP merged dataset using a crosswalk provided by Lawrence Schmidt. The LBD provides reliable measures of firm-level employment and payroll, which are often missing or inconsistently reported in Compustat (variables EMP and XLR, respectively). The merged CRSP/Compustat-LBD dataset allows us to estimate price markups and wage markdowns while retaining detailed financial accounts data. This merged panel also includes various standard macroeconomic and financial time series that are publicly available.

We construct two merged panels for analysis. The first is an annual panel, created by taking December values from the IBES dataset and merging them to the panel from Ren and Zhang (2025). The second is a monthly panel that retains the IBES monthly structure and merges in firm-level variables, such as market power measures, by matching on firm and year. Table 1 reports summary statistics for key variables in our final annual-level sample. For comparison, Table A1 in Appendix A.2 presents the summary statistics for the full Compustat sample over the same period. Firms in our merged sample tend to be larger on average. In addition, LTG values are

Table 1: Final Sample Summary Statistics (Annual Level)

Variable	Mean (1)	SD (2)	P10 (3)	P25 (4)	Median (5)	P75 (6)	P90 (7)	Obs. (8)
Log Sales	19.820	2.042	17.210	18.350	19.780	21.220	22.550	69,500
Log COGS	19.330	2.104	16.630	17.820	19.310	20.770	22.100	69,500
Log SGA	18.210	1.936	15.780	16.800	18.120	19.500	20.810	69,500
Log Wage Bill	18.070	1.938	15.590	16.700	18.080	19.390	20.610	69,500
Log Employment	7.179	2.006	4.605	5.793	7.211	8.561	9.801	69,500
Log Physical Capital	18.160	2.358	15.160	16.470	18.070	19.770	21.340	69,500
Log Intangible Capital	18.260	2.000	15.770	16.810	18.140	19.540	20.960	69,500
Log Total Assets	19.670	2.097	16.990	18.130	19.570	21.100	22.470	69,500
Log Market Cap	19.020	2.397	16.000	17.230	18.920	20.700	22.170	69,000
Labor Share VA	0.668	0.391	0.298	0.470	0.641	0.789	0.956	69,500
LTG	0.157	0.100	0.080	0.105	0.150	0.200	0.250	35,500

Notes: This table presents the summary statistics of the final annual-level sample used in the analysis. The sample ranges from 1977 to 2019. All nominal variables are deflated using the BEA's GDP Price Deflator. Column (1) reports the mean, and Column (2) reports the standard deviation. Columns (3) to (7) report the 10th percentile, 25th percentile, median, 75th percentile, and 90th percentile, respectively. Column (8) reports the number of observations. All figures are rounded in accordance with U.S. Census disclosure requirements. Other than the LTG variable, this table is directly replicated from Ren and Zhang (2025).

generally high (with a sample mean of 15.7%), consistent with the findings of Chan, Karceski and Lakonishok (2003), who document that analyst long-term growth forecasts on average tend to be overly optimistic.

2.2 Estimation of Price Markups and Wage Markdowns

In this section we briefly describe how price markups and wage markdowns are estimated, which are our measures of product market power and labor market power, respectively.[4] We define the total wedge as the product of the two and use this as our measure of total market power. Price markups and wage markdowns are estimated following the methods of Hall (1988), De Loecker and Warzynski (2012), Dobbelaere and Mairesse (2013) and Yeh, Macaluso and Hershbein (2022). The procedure consists of two steps: we first estimate production functions to recover firm-level output elasticities, and then use ratio estimators to compute price markups and wage markdowns.

[4]See Appendix B and Ren and Zhang (2025) for a more comprehensive discussion of this procedure as well as the associated caveats. See also Traina (2018), Bond et al. (2021), Yeh, Macaluso and Hershbein (2022), and De Ridder, Grassi and Morzenti (2025) for more discussions on the caveats of production function estimation and the use of the ratio estimators.

We define firm-level price markups $\mu_{i,t}$ and wage markdowns $\nu_{i,t}$ as follows

$$\mu_{i,t} = \frac{\mathrm{P}_{i,t}}{\mathrm{MC}_{i,t}}, \tag{1}$$

$$\nu_{i,t} = \frac{\mathrm{MRPL}_{i,t}}{\mathrm{W}_{i,t}}, \tag{2}$$

where the price markup $\mu_{i,t}$ is the ratio between the price of output $P_{i,t}$ and its marginal cost $\mathrm{MC}_{i,t}$ and the wage markdown $\nu_{i,t}$ is the ratio between the marginal revenue product of labor $\mathrm{MRPL}_{i,t}$ to the wage paid $W_{i,t}$. While we do not directly observe prices, marginal costs, or marginal revenue products, we can relate Equations (1) and (2) to the ratio of output elasticities and cost shares through the firm's first-order optimality conditions.[5] We can estimate output elasticities with the data available and cost shares are directly observable with financial accounts data.

We now formally define two input types used in the estimation procedure. Definition 1 introduces a *flexible* input f, which is used for the estimation of price markups and wage markdowns. Definition 2 defines a *monopsonistic* input l, which satisfies all of the flexible input conditions except that it may be subject to monopsony power and is therefore not used for the markup estimation but are used for the markdown estimation.

Definition 1 (Flexible Input). *An input f is considered flexible if it satisfies the following conditions:*

1. *It is not subject to adjustment costs.*
2. *It is not subject to monopsony or oligopsony power.*
3. *It is chosen statically by the firm in each period.*
4. *The production function is twice continuously differentiable in input f.*
5. *It is used solely for the production of output.*

Definition 2 (Monopsonistic Input). *An input l is considered monopsonistic if it satisfies all the conditions of a flexible input except that it may be subject to monopsony or oligopsony power.*

Given Definitions 1 and 2 we can now define the ratio estimators for price markups and wage markdowns used in our analysis with Proposition 1.

Proposition 1 (Ratio Estimators). *Suppose there exists at least one input f that is flexible and at least one input l that is monopsonistic. Then firm i's price markup and markdown on input l are given by*

$$\mu_{i,t} = \frac{\theta^f_{i,t}}{\alpha^f_{i,t}}, \tag{3}$$

$$\nu^l_{i,t} = \frac{\theta^l_{i,t}}{\alpha^l_{i,t}} \frac{\alpha^f_{i,t}}{\theta^f_{i,t}}, \tag{4}$$

[5]We define the cost share as the total expenditure of the input divided over total revenue.

where $\theta^j_{i,t}$ denotes the output elasticity of input j and $\alpha^j_{i,t}$ is the cost share of revenue of input j.

Proof. See Appendix B.2 for the proof. □

We estimate the production function estimation procedure developed by De Loecker and Warzynski (2012), which relies on a proxy variable approach and builds upon the methods of Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg, Caves and Frazer (2015). The estimation consists of two stages, the first stage recovers expected output and an error term and the second stage estimates the production function parameters given the productivity process. We estimate a separate production function for each two-digit NAICS (NAICS2) industry in the sample.

We utilize four inputs to estimate the production function: materials, labor, physical capital, and intangible capital. Materials serve as the flexible input and are defined as the sum of cost of goods sold (COGS) and selling, general, and administrative expenses (XSGA), minus the wage bill from the LBD and other fixed costs such as rent. Labor is directly observed from the LBD. Physical capital is constructed using a standard forward-iteration capitalization method, and intangible capital is constructed using the same method following the approach of Eisfeldt and Papanikolaou (2013) and Peters and Taylor (2017). Our choice of materials as the flexible input is motivated by the critique in Traina (2018), who argue that COGS alone does not fully capture flexible input expenditures for markup estimation.[6]

Using these inputs, we implement the first stage of the estimation procedure by regressing the log deflated sales $y_{i,t}$ onto a specified functional form. In general we estimate

$$y_{i,t} = \phi_t(\mathbf{x}_{i,t}, \mathbf{z}_{i,t}) + \varepsilon_{i,t}, \tag{5}$$

where $\mathbf{x}_{i,t}$ is a vector of inputs, $\mathbf{z}_{i,t}$ is a vector of controls such as time fixed effects, $\varepsilon_{i,t}$ is an error term, and

$$\phi_t(\mathbf{x}_{i,t}, \mathbf{z}_{i,t}) = \mathcal{P}_k(\mathbf{x}_{i,t}) + h_t(m_{i,t}, k_{i,t}, \mathbf{z}_{i,t}),$$

where $\mathcal{P}_k(\cdot)$ is a k-order polynomial function and $h_t(\cdot,\cdot,\cdot)$ is the control function. We implement the first stage by regressing $y_{i,t}$ onto a second-order polynomial of $\mathbf{x}_{i,t}$ with year fixed effects via OLS. This yields the predicted log output $\hat{\phi}_{i,t}$ and predicted residuals $\hat{\varepsilon}_{i,t}$, which are utilized in the second stage.

The second stage addresses the endogeneity arising from unobserved firm-level productivity, which is assumed to be Hicks-neutral and denoted $\omega_{i,t}$ (in logs). Since productivity not directly observed, regression coefficients from the first stage cannot be interpreted as the production function parameters. We assume that the production function is translog. Let $f(\mathbf{x}_{i,t}; \hat{\boldsymbol{\beta}})$ be the log production function evaluated at the input bundle $\mathbf{x}_{i,t}$ given the production function parameter

[6] See Ren and Zhang (2025) for a more detailed discussion.

vector $\hat{\boldsymbol{\beta}}$. Then the implied log firm-level productivity is given by

$$\omega_{i,t}(\hat{\boldsymbol{\beta}}) = \hat{\phi}_{i,t} - f(\mathbf{x}_{i,t}; \hat{\boldsymbol{\beta}}).$$

We assume that firm-level productivity follows a stationary AR(1) given by

$$\omega_{i,t}(\hat{\boldsymbol{\beta}}) = (1-\rho_\omega)\bar{\omega} + \rho_\omega \omega_{i,t-1}(\hat{\boldsymbol{\beta}}) + \xi_{i,t}(\hat{\boldsymbol{\beta}}),$$

where $\rho_\omega \in (0,1)$ is the persistence parameter, $\bar{\omega}$ is the unconditional mean, and $\xi_{i,t}(\hat{\boldsymbol{\beta}})$ is a mean zero shock process.

We estimate $\hat{\boldsymbol{\beta}}$ using generalized methods of moments (GMM) with the following moment condition

$$\mathbb{E}\left[\xi_{i,t}(\hat{\boldsymbol{\beta}}) \cdot \tilde{\mathbf{z}}_{i,t}\right] = \mathbf{0}, \tag{6}$$

where $\tilde{\mathbf{z}}_{i,t}$ denotes a vector of instruments. The instruments include the lagged materials and labor inputs as well as the current physical capital and intangible capital inputs along with their interactions and second-order terms. The identification of the parameters relies on a timing assumption. The productivity shock $\xi_{i,t}$ is observed by firms at the beginning of period t but unobservable to the econometrician. Firms make contemporaneous decisions on materials and labor, while capital stocks are predetermined. As such, the productivity shock is assumed to be uncorrelated with the current capital stock and lagged input choices.

Given the production function's estimated parameters $\hat{\boldsymbol{\beta}}$, we compute firm-level output elasticities. Under the translog specification, output elasticities are linear functions of the input bundle and are straightforward to compute. As previously noted, cost shares are directly observable from the financial data. With both components, we can implement the ratio estimators to recover price markups and wage markdowns.[7] With firm-level price markups and wage markdowns estimated, we now proceed to the discussing and presenting some stylized facts about market power and analyst expectations.

Using the estimated firm-level price markups and wage markdowns, we construct indicator variables that equal one if a firm falls within the top decile of price markups or, separately, the top decile of wage markdowns within its NAICS2–year cell. These measures allow us to capture the cross-sectional heterogeneity most relevant for our analysis.

The central empirical patterns documented by Ren and Zhang (2025), which we use to discipline the calibration of the model in Section 5, are as follows:

1. The total wedge (defined as the product of price markups and wage markdowns and a measure of total market power) is positively related to firm size and productivity, but negatively

[7] We apply the correction method of De Loecker and Warzynski (2012) and De Loecker, Eeckhout and Unger (2020) to the cost share $\hat{\alpha}^j_{i,t} \equiv \alpha^j_{i,t} \exp(-\hat{\varepsilon}_{i,t})$ where $\alpha^j_{i,t}$ is the computed cost share of input j and $\hat{\varepsilon}_{i,t}$ is the residual from the first stage in Equation (5). This approach removes any output variation not related to variables impacting input demand and market characteristics.

related to the labor share.

2. Firms with high total market power have low price markups and high wage markdowns whereas firms with low total market power have the opposite.

3. Price markups are negatively related to firm size and productivity, but positively related to the labor share.

4. Wage markdowns are positively related to firm size and productivity, but negatively related to the labor share.

5. Price markups and wage markdowns are negatively correlated.

Regression results supporting these stylized facts are reported in Tables A2 to A4 of Appendix A.2. These tables reproduce the univariate regressions with NAICS2–year fixed effects from Ren and Zhang (2025).

3 Diagnostic Expectations and the Asymmetry of Market Power

This section presents a model of diagnostic expectations following Bordalo et al. (2024*a*), and uses this framework to analyze how behavioral responses vary with firm-level market power. The rest of the section is structured as follows: Section 3.1 outlines the model of expectations and returns, and Section 3.2 presents the empirical results.[8]

3.1 Model of Diagnostic Expectations and Returns

To analyze how belief distortions shape asset prices, we begin with the decomposition of log returns following Campbell and Shiller (1987, 1988). The log realized return $r_{i,t+1}$ between periods t to $t+1$ is given by

$$r_{i,t+1} = \alpha p_{i,t+1} + (1-\alpha) d_{i,t+1} - p_{i,t} + k, \tag{7}$$

where $p_{i,t+1}$ is the log stock price of i at $t+1$, $d_{i,t+1}$ is the log dividend paid at $t+1$, and $k < 0$ and $\alpha \in (0,1)$ are constants from the log linearization. Rearranging Equation (7) in terms of the log price-to-earnings (P/E) ratio yields

$$p_{i,t} - e_{i,t} = k + (e_{i,t+1} - e_{i,t}) - r_{i,t+1} + (1-\alpha)\hat{d}_{i,t} + \alpha(p_{i,t+1} - e_{i,t+1}), \tag{8}$$

where $e_{i,t}$ is log earnings at $t+1$ and $\hat{d}_{i,t} \equiv d_{i,t} - e_{i,t}$ is the log dividend payout ratio. Following De La O and Myers (2021), we assume $1-\alpha \approx 0$ and drop the payout ratio term for simplicity. We

[8] We repurpose some notation from the previous section where appropriate; otherwise, notation retains its prior definitions.

iterate forward Equation (8) and impose a transversality condition $\lim_{s\to\infty}\alpha^{t+s}\tilde{\mathbb{E}}_t\left[p_{i,t+s}-e_{i,t+s}\right]=0$ to obtain

$$p_{i,t}-e_{i,t}=\frac{k}{1-\alpha}+\sum_{s\geq 0}\alpha^s\tilde{\mathbb{E}}_t\left[g_{i,t+1+s}\right]-\sum_{s\geq 0}\alpha^s\tilde{\mathbb{E}}_t\left[r_{i,t+1+s}\right], \tag{9}$$

where $g_{i,t+1+s}\equiv e_{i,t+1+s}-e_{i,t+s}$ is the earnings growth rate and $\tilde{\mathbb{E}}_t\left[\cdot\right]$ denotes subjective expectations at time t that may not necessarily be rational. Equation (9) can be rearranged to express realized returns

$$r_{i,t+1}=\tilde{\mathbb{E}}_t\left[r_{i,t+1}\right]+\sum_{s\geq 0}\alpha^s\Delta\tilde{\mathbb{E}}_{t+1}\left[g_{i,t+1+s}\right]-\sum_{s\geq 1}\alpha^s\Delta\tilde{\mathbb{E}}_{t+1}\left[r_{i,+1+s}\right], \tag{10}$$

where $\Delta\tilde{\mathbb{E}}_{t+1}\left[x\right]\equiv\tilde{\mathbb{E}}_{t+1}\left[x\right]-\tilde{\mathbb{E}}_t\left[x\right]$ is the revision in beliefs. Following De La O and Myers (2021) and Bordalo et al. (2024*a*), we assume a constant discount rate and focus on distortions in expectations about future earnings growth. Under this assumption, return predictability arises from belief revisions rather than from time-varying discount rates. This motivates the central role of expectations over future earnings growth, which we describe in more detail.

We model earnings growth as an exogenous AR(1) process:

$$g_{i,t+1}=\rho_g g_{i,t}+\xi_{i,t+1}, \tag{11}$$

where $\rho_g\in(0,1)$ is the true persistence of earnings growth and $\xi_{i,t+1}$ is an independent and identically distributed innovation with mean zero and variance σ^2_ξ. We assume that $\xi_{i,t+1}$ consists of two uncorrelated mean-zero Gaussian components: a tangible news shock $\tau_{i,t+1}$ that is observed in $t+1$ and an intangible news shock $\kappa_{i,t}$ that is observed in t. That is, $\xi_{i,t+1}=\tau_{i,t+1}+\kappa_{i,t}$ and $\sigma^2_\xi=\sigma^2_\tau+\sigma^2_\kappa$ where σ^2_τ is the variance of tangible news and σ^2_κ is the variance of intangible news. Both $\tau_{i,t+1}$ and $\kappa_{i,t}$ are assumed to be independently distributed across firms and over time.

We model belief formation over earnings growth using the diagnostic expectations framework of Bordalo et al. (2024*a*), building on the general theory of representativeness developed in Gennaioli and Shleifer (2010) and Bordalo, Gennaioli and Shleifer (2018). Under diagnostic expectations, agents overweight states that are more representative relative to prior beliefs. Specifically, the subjective density used to forecast one-period-ahead earnings growth is given by

$$f^\theta_t(g_{i,t+1})=f(g_{i,t+1}\mid g_{i,t},\kappa_{i,t})\left(\frac{f(g_{i,t+1}\mid g_{i,t},\kappa_{i,t})}{f(g_{i,t+1}\mid \mathbb{E}_{t-1}\left[g_{i,t}\right],\kappa_{i,t-1})}\right)^\theta\frac{1}{Z}, \tag{12}$$

where $Z\in\mathbb{R}_{++}$ is a constant that ensures the density integrates to 1 over its support and $\theta\geq 0$ governs the intensity of representativeness.[9] When $\theta=0$ agents use the correct distribution and

[9] Gennaioli and Shleifer (2010) formalize representativeness as follows: consider the distribution of a trait T in group G, with true distribution $f(T=t\mid G)$. The representativeness of trait $T=t$ for group G is defined as the likelihood ratio $f(T=t\mid G)/f(T=t\mid G^c)$, where G^c is a comparison group. A trait is more representative if it is relatively more frequent in G than in G^c. Due to limited memory, more representative types are more easily recalled and thus

thus have rational expectations. Conversely, when $\theta > 0$ agents overweight representative states and underweight unrepresentative states.

Since we are interested in agents' expectations not only one period ahead but across multiple future periods $s \geq 1$, we extend the diagnostic expectations formulation in Equation (12) following Bordalo, Gennaioli and Shleifer (2018). The distorted subjective density used to forecast s-period-ahead earnings growth is

$$f_t^\theta(g_{i,t+s}) = f(g_{i,t+s} \mid g_{i,t}, \kappa_{i,t}) R_{i,s,t}^\theta, \tag{13}$$

$$R_{i,s,t} = \frac{1}{Z} \prod_{n \geq 1} \left[\frac{f(g_{i,t+s} \mid \mathbb{E}_{t+1-n}\left[g_{i,t}\right], \kappa_{i,t+1-n})}{f(g_{i,t+s} \mid \mathbb{E}_{t-n}\left[g_{i,t}\right], \kappa_{i,t-n})} \right]^{\gamma_n}, \tag{14}$$

where Z is a normalizing constant and γ_n governs how much weight agents place on changes in prior beliefs at different lags. The term $R_{i,s,t}$ captures the representativeness of the forecasted outcome $g_{i,t+s}$ based on how surprising it appears relative to past expectations. When $\gamma_n = 0$ for all $n > 1$, the model reduces to a memoryless version in which only the most recent belief revision matters.

To operationalize the general diagnostic expectations framework, we adopt a specific form of memory decay in belief formation. Following Bordalo et al. (2024*a*), we assume that memory weights decline geometrically: $\gamma_n = \gamma^{n-1}$ for some $\gamma \in (0, 1)$. This structure implies that agents gradually adjust their beliefs over time, placing greater weight on more recent revisions. It captures sluggish updating while remaining analytically tractable.

Under this assumption, the diagnostic forecast of future earnings growth simplifies considerably. Using results from Bordalo, Gennaioli and Shleifer (2018), the distorted expectation is given by

$$\tilde{\mathbb{E}}_t\left[g_{i,t+s}\right] = \mathbb{E}_t\left[g_{i,t+s}\right] + \theta \sum_{n \geq 1} \gamma^{n-1} \left(\mathbb{E}_{t+1-n}[g_{i,t+s}] - \mathbb{E}_{t-n}[g_{i,t+s}]\right), \tag{15}$$

where the second term captures accumulated overreaction to past forecast revisions. Given the AR(1) structure of earnings growth, we can further collapse this expression into a closed-form distortion process, which we state formally below.

In particular, we set $\gamma = \rho_\zeta / \rho_g$, where $\rho_\zeta < \rho_g$ is the persistence of belief distortions. This assumption nests the sluggish diagnostic expectations framework of Bordalo et al. (2024*a*) and allows us to simplify the diagnostic forecast in Equation (15) into a closed-form expression. Proposition 2 formalizes this result.

Proposition 2 (Simplified Diagnostic Expectations)**.** *Suppose agents form expectations using the distorted distribution defined by Equations* (13) *and* (14)*, with memory weights* $\gamma_n = \gamma^{n-1}$ *and* $\gamma = \rho_\zeta / \rho_g$*,*

overweighted in agents' beliefs.

where $\rho_\zeta < \rho_g$. Then the s-period-ahead forecast of earnings growth satisfies:

$$\tilde{\mathbb{E}}_t\left[g_{i,t+s}\right] = \mathbb{E}_t\left[g_{i,t+s}\right] + \rho_g^{s-1}\zeta_{i,t}, \tag{16}$$

where the rational forecast is $\mathbb{E}_t[g_{i,t+s}] = \rho_g^{s-1}(\rho_g g_{i,t} + \kappa_{i,t})$, and $\zeta_{i,t}$ is a belief distortion evolving according to

$$\zeta_{i,t} = \rho_\zeta \zeta_{i,t-1} + u_{i,t}, \tag{17}$$

with the expectations shock given by

$$u_{i,t} = \theta(\rho_g \tau_{i,t} + \kappa_{i,t}). \tag{18}$$

Proof. See Appendix C.1 for the proof. □

The belief distortion $\zeta_{i,t}$ captures systematic deviations from rational expectations, driven by overreaction to recent news. Its persistence ρ_ζ governs the dynamics of the belief distortion. When $\rho_\zeta < \rho_g$, distortions decay more quickly than fundamentals, implying that LTG revisions $\Delta\tilde{\mathbb{E}}_t[g_{i,t+s}]$ are negatively autocorrelated. In other words, excess optimism or pessimism tends to revert.

The expectation shock $u_{i,t} = \theta(\rho_g \tau_{i,t} + \kappa_{i,t})$ represents the distorted response to new information received at time t. It is the immediate update to beliefs, and reflects agents' over- or underreaction to both the tangible news $\tau_{i,t}$ and intangible news $\kappa_{i,t}$ arriving in period t.[10] The parameter θ governs the direction and magnitude of distortion: if $\theta = 0$, expectations are rational and agents respond proportionally to news; if $\theta > 0$, representative news is overweighted, leading to exaggerated revisions.

The combination of Equations (10) and (16) yields the empirical tests and predictions of Bordalo et al. (2024*a*), which we extend to the firm-level and augment to allow for heterogeneity by characteristics. In particular, we examine whether belief distortions vary systematically with firm characteristics such as product and labor market power. While the baseline model assumes common values of the distortion persistence ρ_ζ and reaction parameter θ across firms, we test empirically whether these parameters differ by market power. We now state the original empirical specifications from Bordalo et al. (2024*a*) and link them to the model's structural parameters.

We consider three main empirical specifications adapted from Bordalo et al. (2024*a*). First, we examine earnings growth forecast errors as a function of belief revisions and past expectations, which is given by

$$g_{i,t+s} - \tilde{\mathbb{E}}_t\left[g_{i,t+s}\right] = \beta_0^{\text{FE}} + \beta_1^{\text{FE}}\Delta\tilde{\mathbb{E}}_t\left[g_{i,t+s}\right] + \beta_2^{\text{FE}}\tilde{\mathbb{E}}_{t-1}\left[g_{i,t+s}\right] + \varepsilon_{i,t+s}. \tag{19}$$

If there is overreaction, i.e. $\theta > 0$, then $\beta_1^{\text{FE}} < 0$ and $\beta_2^{\text{FE}} < 0$. Under rational expectations ($\theta = 0$) then $\beta_1^{\text{FE}} = \beta_2^{\text{FE}} = 0$.

[10] Note that tangible news needs to be adjusted by ρ_g given the timing assumption of when it affects growth.

Second, we test for belief reversion with

$$\Delta\tilde{\mathbb{E}}_t\left[g_{i,t+s}\right] = \beta_0^{\text{Rev}} + \beta_1^{\text{Rev}}\tilde{\mathbb{E}}_{t-1}\left[g_{i,t+s}\right] + \varepsilon_{i,t}. \tag{20}$$

If belief distortions are less persistent than the underlying fundamentals ($\rho_\zeta < \rho_g$), then $\beta_1^{\text{Rev}} < 0$, implying that long-term growth expectations revert over time. Furthermore, if agents have rational expectations then $\rho_\zeta = \rho_g - 1$ otherwise under diagnostic expectations $\rho_\zeta \neq \rho_g - 1$. The coefficients $\beta_1^{\text{FE}}, \beta_2^{\text{FE}}, \beta_1^{\text{Rev}}$ in the regressions in Equations (19) and (20) have close-form expressions, which are provided in Appendix C.2.

Finally, return predictability arises from systematic belief errors and distorted responses to news. The return regression implied by the model is given by

$$r_{i,t+1} = r_i + \left(\frac{1-\alpha\rho_\zeta}{1-\alpha\rho_g}\right)\mathbb{E}_t\left[g_{i,t+1} - \tilde{\mathbb{E}}_t[g_{i,t+1}]\right] + \left(\frac{1+\alpha\theta}{1-\alpha\rho_g}\right)\tau_{i,t+1} + \left(\frac{\alpha(1+\theta)}{1-\alpha\rho_g}\right)\kappa_{i,t+1}. \tag{21}$$

Overreaction implies that both LTG revisions and lagged expectations negatively predict forecast errors and realized returns, while forecast errors themselves positively predict future returns. This behavior also implies that stock returns exhibit predictable reversals.

We extend these regressions by allowing the coefficients to vary with firm-level product and labor market power. Heterogeneity in these coefficients is consistent with heterogeneity in the structural parameters θ and ρ_ζ as well as ρ_g. We detail our empirical implementation and interaction specifications in Section 3.2.

3.2 Empirical Results

We proceed by estimating Equations (19) to (21), allowing for heterogeneity across firm types defined by market power. We begin with the forecast error regression, estimated using the following specification:

$$g_{i,t+s} - \tilde{\mathbb{E}}_t\left[g_{i,t+s}\right] = \sum_{j\in D}\left(\beta_{1,j}^{\text{FE}}\Delta\tilde{\mathbb{E}}_t\left[g_{i,t+s}\right] + \beta_{2,j}^{\text{FE}}\tilde{\mathbb{E}}_{t-1}\left[g_{i,t+s}\right]\right) \times \mathbf{1}\{i \in j\} + Z_{i,t}^{\intercal}\gamma + \varepsilon_{i,t+s}, \tag{22}$$

where D denotes the set of firm types (all firms, low total market power firms, high total market power firms), and $Z_{i,t}$ includes fixed effects and the uninteracted indicator variables.[11] Low total market power firms are defined as those in the bottom 10% of total wedges within their NAICS2 industry-year. High total market power firms are defined analogously but for the top 10%.[12]

Table 2 reports results across four specifications. Column (1) includes interactions with low

[11]Unlike Bordalo et al. (2024*a*), we do not standardize LTGs and instead keep them in levels, following the comments of De la O and Myers (2024).

[12]Throughout the rest of the text we will use "market power" interchangeably with "total market power" unless otherwise noted.

Table 2: 5-Year Forecast Error Predictability (NAICS2 × Month FE)

	Forecast Error (5-Year)		
	(1)	(2)	(3)
LTG Revision	-0.825	-0.826	-0.829
	(0.026)	(0.026)	(0.027)
LTG Revision × Low Market Power	0.037		0.038
	(0.107)		(0.108)
LTG Revision × High Market Power		0.015	0.018
		(0.086)	(0.087)
LTG (Lag 1)	-0.823	-0.848	-0.848
	(0.031)	(0.031)	(0.033)
LTG (Lag 1) × Low Market Power (Lag 1)	-0.017		0.004
	(0.058)		(0.059)
LTG (Lag 1) × High Market Power (Lag 1)		0.163	0.164
		(0.067)	(0.068)
NAICS2 × Month FE	Yes	Yes	Yes
Firm FE	No	No	No
Observations	246,000	246,000	246,000

Notes: This table presents regressing 5-year realized forecast error against LTG revisions and lagged LTG along with interactions. All specifications include NAICS2 × month fixed effects. Column (1) presents the specification with only an interaction for low total market power firms. Column (2) presents the specification with only an interaction for high total market power firms. Column (3) presents the specification with interactions with both low total market power and high total market power firms. Standard errors are reported in parentheses and are two-way clustered by firm and year. All figures are rounded in accordance with U.S. Census disclosure requirements.

total market power firms only; Column (2) with high total market power firms only; Column (3), our preferred specification, includes both. Results are consistent across specifications.[13]

As in Bordalo et al. (2024*a*), we find strong evidence of overreaction: the baseline coefficients on LTG levels and revisions are negative and highly significant across all specifications. For instance, the coefficient on LTG revision is -0.829 in Column (3), and the coefficient on lagged LTG is -0.848. Under rational expectations, these coefficients would be statistically indistinguishable from zero. The interaction terms reveal important heterogeneity. For low total market power firms, the interaction terms on both LTG revision (0.038) and LTG level (0.004) are small and statistically insignificant, suggesting that forecast error predictability is similar to baseline firms. In contrast, high total market power firms exhibit significantly attenuated overreaction: the interaction coefficient on lagged LTG is 0.164, partially offsetting the baseline overreaction. The interaction

[13]Table A5 in Appendix A.2 presents the results including a firm fixed effect for robustness. Tables A6 and A7 present the results using the three-year forecast error instead using the same set of specifications for robustness.

Table 3: LTG Revision Predictability (NAICS2 $\times$ Year FE)

	LTG Revision		
	(1)	(2)	(3)
LTG (Lag 1)	-0.496	-0.494	-0.504
	(0.039)	(0.039)	(0.039)
LTG (Lag 1) $\times$ Low Market Power (Lag 1)	0.112		0.117
	(0.060)		(0.061)
LTG (Lag 1) $\times$ High Market Power (Lag 1)		0.038	0.047
		(0.047)	(0.047)
NAICS2 FE	No	No	No
NAICS2 $\times$ Year FE	Yes	Yes	Yes
Firm FE	No	No	No
Observations	30,000	30,000	30,000

Notes: This table presents regressing future LTG revisions against lagged LTG along with interactions. All specifications include NAICS2 $\times$ year fixed effects. Column (1) presents the specification with only an interaction for low total market power firms. Column (2) presents the specification with only an interaction for high total market power firms. Column (3) presents the specification with interactions with both low total market power and high total market power firms. Standard errors are reported in parentheses and are two-way clustered by firm and year. All figures are rounded in accordance with U.S. Census disclosure requirements.

terms on LTG revisions for high total market power (0.018) is statistically insignificant, however. These results imply that while analysts overreact on average, this overreaction is less pronounced for firms with high market power.

We next examine heterogeneity in the predictability of LTG revisions. Specifically, we estimate the following regression:

$$\Delta\tilde{\mathbb{E}}_t\left[g_{i,t+s}\right] = \sum_{j\in D} \beta_{1,j}^{\text{Rev}} \tilde{\mathbb{E}}_{t-1}\left[g_{i,t+s}\right] \times \mathbf{1}\{j \in D\} + Z_{i,t}^{\intercal}\gamma + \varepsilon_{i,t}, \tag{23}$$

where the dependent variable is the revision in long-term earnings growth expectations, and the key regressor is the lagged level of the LTG, interacted with firm type indicators. Table 3 presents the results, following a similar format as earlier specifications. Across all columns, we find that LTG revisions are significantly negatively related to their lagged values, with baseline coefficients around -0.50, consistent with predictable mean reversion in expectations.[14] This finding holds across all firm types.

However, we observe heterogeneity in the degree of reversion. The interaction term for low total

[14]Table A8 in Appendix A.2 presents the results with only a NAICS2 fixed effect for robustness.

Table 4: 5-Year Return Predictability (NAICS2 × Month FE)

	Future Return (5-Year)		
	(1)	(2)	(3)
LTG Revision	-0.377	-0.470	-0.453
	(0.072)	(0.066)	(0.077)
LTG Revision × Low Market Power	-0.338		-0.267
	(0.236)		(0.246)
LTG Revision × High Market Power		0.473	0.457
		(0.124)	(0.130)
LTG (Lag 1)	-0.863	-0.975	-0.941
	(0.103)	(0.105)	(0.108)
LTG (Lag 1) × Low Market Power (Lag 1)	-0.426		-0.355
	(0.224)		(0.224)
LTG (Lag 1) × High Market Power (Lag 1)		0.643	0.612
		(0.201)	(0.202)
NAICS2 × Month FE	Yes	Yes	Yes
Firm FE	No	No	No
Observations	340,000	340,000	340,000

Notes: This table presents regressing 5-year future returns against LTG revisions and lagged LTG along with interactions. All specifications include NAICS2 × month fixed effects. Column (1) presents the specification with only an interaction for low total market power firms. Column (2) presents the specification with only an interaction for high total market power firms. Column (3) presents the specification with interactions with both low total market power and high total market power firms. Standard errors are reported in parentheses and are two-way clustered by firm and year. All figures are rounded in accordance with U.S. Census disclosure requirements.

market power firms is positive and statistically significant, with a coefficient of 0.112 in Column (1) and 0.117 in Column (3). This attenuates the overall reversion, indicating that belief updates are more persistent for low total market power firms. In contrast, the interaction terms for high total market power firms (0.047) is positive but statistically insignificant. These results suggest that while all firms exhibit mean reversion in LTG expectations, the process is slower and less reactive for firms with low total market power. Beliefs about low total market power firms adjust more sluggishly, consistent with more persistent narratives or extrapolation dynamics.

Finally, we turn to the relationship between expectations and realized outcomes by estimating return predictability and examining if there are predictable return reversals. Specifically, we estimate:

$$\sum_{k=1}^{s} r_{i,t+k} = \sum_{j \in D} \left(\beta_{1,j}^{\text{Return}} \Delta \tilde{\mathbb{E}}_t \left[g_{i,t+s} \right] + \beta_{2,j}^{\text{Return}} \tilde{\mathbb{E}}_{t-1} \left[g_{i,t+s} \right] \right) \times \mathbf{1}\{i \in j\} + Z_{i,t}^{\intercal} \gamma + \varepsilon_{i,t+s}, \quad (24)$$

Table 5: 5-Year Return Predictability (NAICS2 × Month FE; Instrumented)

	Future Return (5-Year)		
	(1)	(2)	(3)
Predicted Forecast Error	0.643	0.681	0.670
	(0.109)	(0.108)	(0.109)
Predicted Forecast Error × Low Market Power	0.151		0.126
	(0.158)		(0.159)
Predicted Forecast Error × High Market Power		-0.281	-0.274
		(0.133)	(0.133)
NAICS2 × Month FE	Yes	Yes	Yes
Firm FE	No	No	No
Observations	244,000	244,000	244,000

Notes: This table presents regressing 5-year future returns against predicted five-year forecast errors along with interactions. All specifications include NAICS2 × month fixed effects. Column (1) presents the specification with only an interaction for low total market power firms. Column (2) presents the specification with only an interaction for high total market power firms. Column (3) presents the specification with interactions with both low total market power and high total market power firms. Standard errors are reported in parentheses and are two-way clustered by firm and year. All figures are rounded in accordance with U.S. Census disclosure requirements.

where the outcome is the cumulative realized return over the next five years. This specification mirrors Equation (25) but focuses on the realized return response to expectations and their revisions. Following Bordalo et al. (2024*a*), we also estimate a version using predicted forecast errors as a regressor:

$$\sum_{k=1}^{s} r_{i,t+k} = \sum_{j\in D} \beta_{1,j}^{\text{Return}} \hat{\text{FE}}_{i,t} \times \mathbf{1}\{i \in j\} + Z_{i,t}^{\intercal}\gamma + \varepsilon_{i,t+s}, \tag{25}$$

where $\hat{\text{FE}}_{i,t}$ is the fitted value from the forecast error regression in Column (3) of Table 2. According to the theory outline in Section 3.1, it there is overreaction than stock returns exhibit predictable reversals and the relative strength of the reversal follow the relative strength of overreaction.

Tables 4 and 5 present the results.[15] Consistent with the forecast error analysis, higher LTGs are associated with lower subsequent realized returns and exhibit predictable reversals. In Column (3) of Table 4, the coefficient on the lagged LTG is -0.941, and that on the LTG revision is -0.453, both statistically significant. These findings are consistent with the classic overreaction framework: firms with higher growth expectations tend to underperform, leading to a predictable negative relationship between expectations and future returns.

The interaction terms reveal heterogeneity in return predictability across firm types. For low

[15] Tables A9 to A11 in Appendix A.2 present robustness checks including a firm fixed effect and using three-year returns instead of five-year returns, respectively.

total market power firms, the interactions are small and statistically insignificant, again suggesting that their return dynamics do not systematically differ from baseline firms. In contrast, high total market power firms exhibit attenuated return predictability. For example, the interaction term on lagged LTG for high total market power firms is 0.643 and is significant and offsetting a meaningful portion of the baseline coefficient. These results align with earlier findings: for these firms, expectations are less negatively related to subsequent realized performance, suggesting either less overreaction or stronger realized fundamentals.

Table 5 provides further support by linking predicted forecast errors to future returns. The fitted forecast error is positively associated with realized returns, with a coefficient of 0.643 in Column (1). This implies that firms expected to exhibit more overreaction. That is, firms with high predicted forecast errors tend to deliver stronger future returns. This pattern is consistent with the idea that overly pessimistic expectations about certain firms subsequently reverse, generating excess returns.

Taken together, these findings reinforce the broader interpretation: forecast errors, belief dynamics, and returns are tightly linked in the cross-section. Overreaction is pervasive in the data, but its severity varies systematically with firm-level market power. Firms with high market power exhibit attenuated overreaction, both in expectations and in realized returns. In contrast, firms with low market power do not display meaningful differences in predictability relative to the baseline, but do exhibit more persistent expectations and slower belief reversion. These results underscore the importance of market power when interpreting the informational content of expectations and their implications for asset prices.

4 Calibration of Diagnostic Expectations

We now examine which values of the model parameters introduced in Section 3.1 can rationalize the empirical findings presented in Section 3.2. We also explore the economic intuition behind these patterns, focusing on how different configurations of belief distortions map to the observed heterogeneity. The rest of the section proceeds as follows: Section 4.1 presents various comparative statics exercises to build intuition for the model's predictions as well as understand parameter identification, and Section 4.2 estimates the structural parameters using the empirical moments.

4.1 Comparative Statics and Identification

We focus on two key parameters of the diagnostic expectations model: the reaction parameter θ and the persistence of the belief distortion ρ_ζ. These govern the strength and duration of biased responses to news, and together they shape the dynamic properties of subjective expectations. Given the tight mapping between these parameters and the regression coefficients in Equations (19) and (20), we examine the comparative statics of these coefficients with respect to θ and ρ_ζ. For this initial analysis, we fix the total variance of the expectation shock σ_u^s and the true persistence of earnings growth ρ_g, since variation in θ and ρ_ζ provides a more direct explanation for the empirical

Table 6: Comparative Statics of Regression Coefficients

Reg. Coef.	$\theta = 5.0,\ \rho_\zeta = 0.3$ (1)	$\theta = 5.0,\ \rho_\zeta = 0.2$ (2)	$\theta = 4.0,\ \rho_\zeta = 0.3$ (3)	$\theta = 4.0,\ \rho_\zeta = 0.2$ (4)
β_1^{FE}	-0.833	-0.833	-0.799	-0.799
β_2^{FE}	-0.781	-0.751	-0.740	-0.706
β_1^{Rev}	-0.207	-0.413	-0.199	-0.396

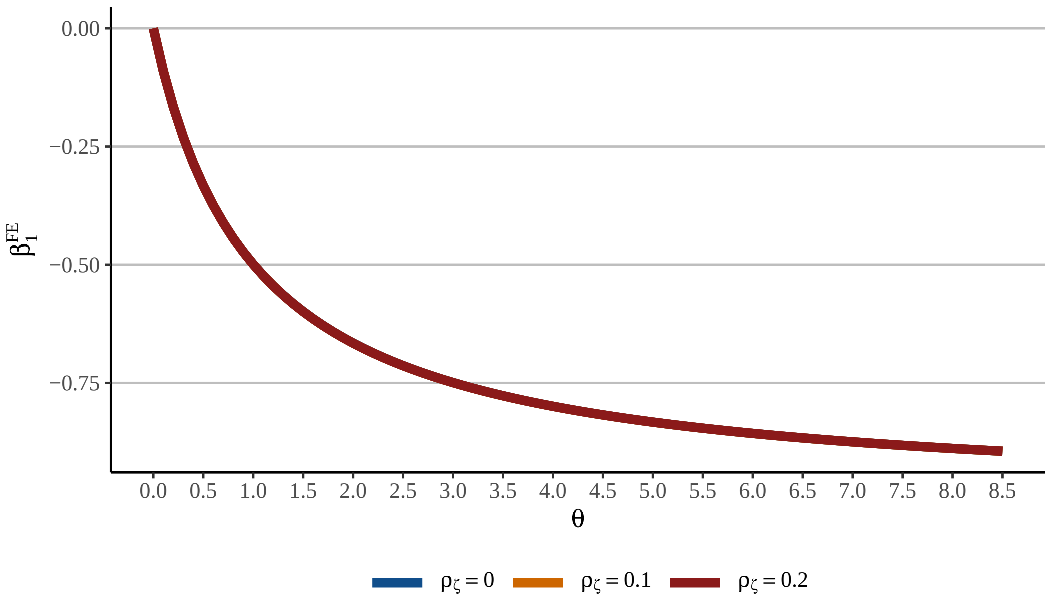
Notes: This table presents a simple comparative statics of β_1, β_2, and γ_1. We set $\rho_g = 0.4$, $\sigma_u = 0.02$, and $s = 5$ in all examples. Each column represents a distinct (θ, ρ_ζ) pair. Column (1) sets $\theta = 5$ and $\rho_\zeta = 0.3$; Column (2) changes ρ_ζ to 0.2 holding θ fixed; Column (3) changes θ to 4 holding $\rho_\zeta = 0.3$ fixed; Column (4) changes both to $\theta = 4$, $\rho_\zeta = 0.2$.

patterns of interest. We elaborate further on this interpretation in the estimation results below, and also examine the comparative statics with respect to ρ_g later in this section.

Table 6 reports how the model-implied regression coefficients vary with the reaction parameter θ and the persistence of the belief distortion ρ_ζ, fixing $\sigma_u = 0.02$ and $\rho_g = 0.4$. We first examine the role of θ by comparing Columns (1) to (3) and Columns (2) to (4), which hold ρ_ζ fixed at 0.3 and 0.2, respectively, while lowering θ from 5.0 to 4.0. Across both comparisons, we observe that the coefficients β_1^{FE} increase from -0.833 to -0.799, and β_2^{FE} from -0.781 to -0.740 in Columns (1) to (3). In Columns (2) to (4), β_1^{FE} increases from -0.833 to -0.799 and β_2^{FE} increases from -0.751 to -0.706. These results indicate that forecast errors become less systematically biased. The revision regression coefficient β_1^{Rev} also becomes less negative, increasing from -0.207 to -0.199 in Columns (1) to (3), and from -0.413 to -0.396 in Columns (2) to (4). These patterns are consistent with the interpretation that a smaller θ reduces the amplification of news, thereby dampening predictable return reversals and slowing the reversion of expectations.

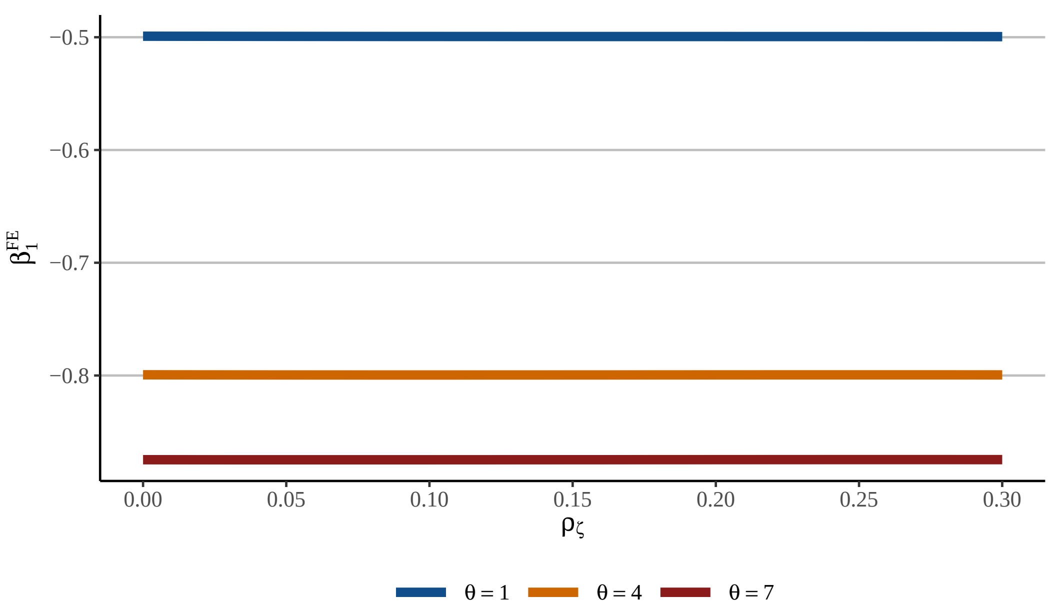
Next, we examine the role of belief distortion persistence by comparing Columns (1) to (2) and Columns (3) to (4), which hold θ fixed while lowering ρ_ζ from 0.3 to 0.2. In both cases, the coefficient β_1^{FE} remains unchanged at -0.833 and -0.799, respectively, while β_2^{FE} increases from -0.781 to -0.751 and from -0.740 to -0.706, indicating that as belief distortions decay more quickly, the magnitude of predictable forecast errors diminishes. The stability of β_1^{FE} reflects the fact that belief distortion decay affects the persistence of biased expectations, but not their initial impact. In other words, ρ_ζ governs how long distorted beliefs continue to influence returns, whereas the on-impact response is determined primarily by θ. As a result, β_1^{FE}, which captures the immediate link between belief revisions and returns, remains largely unchanged, while β_2^{FE}, which reflects the predictive power of lagged expectations, declines in magnitude as ρ_ζ falls. Similarly, the revision regression coefficient β_1^{Rev} becomes substantially more negative from -0.207 to -0.413 in Columns (1) to (2), and from -0.199 to -0.396 in Columns (3) to (4). These results imply a sharper reversion of long-term growth expectations. Together, these changes show that lower ρ_ζ leads to distortions that decay faster.

Figure 1: Effect of Overreaction θ on β_1^{FE}, Across Belief Persistence ρ_ζ



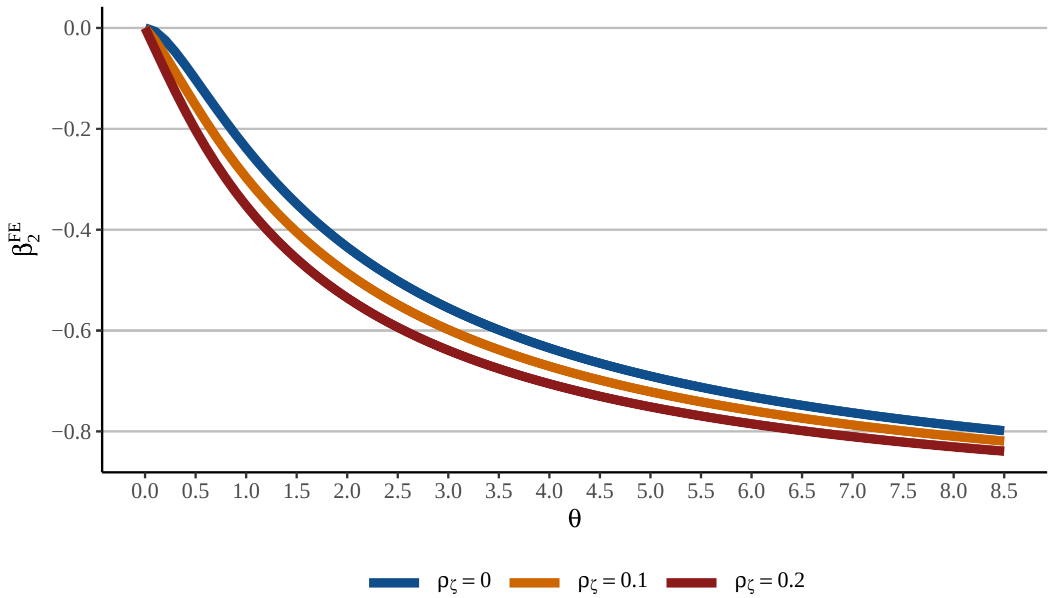
Notes: This figure reports the comparative statics of β_1^{FE} with respect to θ, holding $\rho_g = 0.4$, $\sigma_u = 0.02$, and $s = 5$ fixed. Each line corresponds to a different fixed value of ρ_ζ: the blue line uses $\rho_\zeta = 0$, the orange line uses $\rho_\zeta = 0.1$, and the red line uses $\rho_\zeta = 0.2$.

Figure 2: Effect of Belief Persistence ρ_ζ on β_1^{FE}, Across Overreaction Levels θ



Notes: This figure reports the comparative statics of β_1^{FE} with respect to ρ_ζ, holding $\rho_g = 0.4$, $\sigma_u = 0.02$, and $s = 5$ fixed. Each line corresponds to a different fixed value of θ: the blue line uses $\theta = 1$, the orange line uses $\theta = 4$, and the red line uses $\theta = 7$.

Figure 3: Effect of Overreaction θ on β_2^{FE}, Across Belief Persistence ρ_ζ

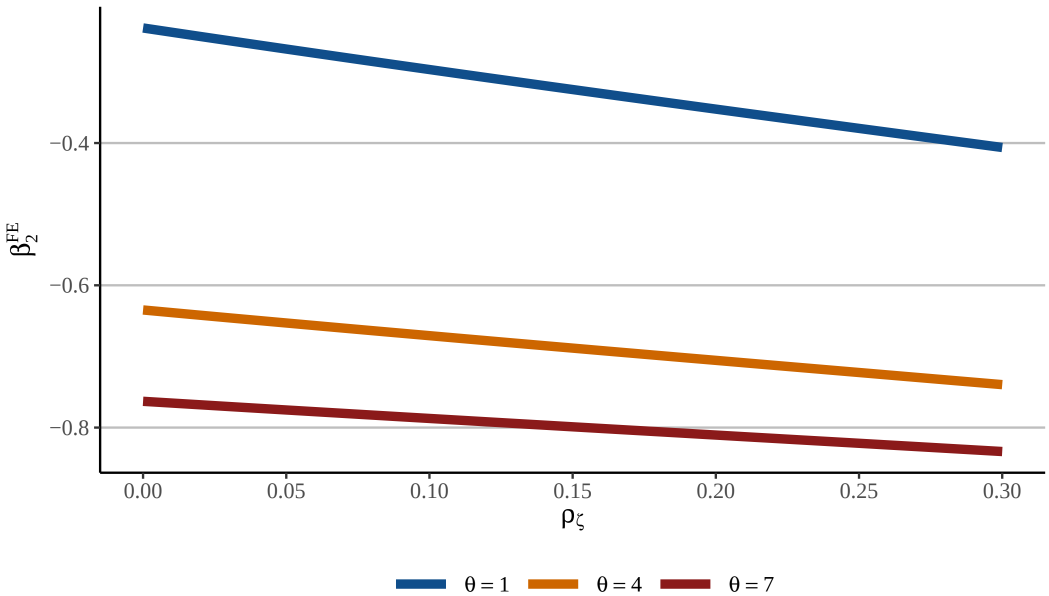


Notes: This figure reports the comparative statics of β_2^{FE} with respect to θ, holding $\rho_g = 0.4$, $\sigma_u = 0.02$, and $s = 5$ fixed. Each line corresponds to a different fixed value of ρ_ζ: the blue line uses $\rho_\zeta = 0$, the orange line uses $\rho_\zeta = 0.1$, and the red line uses $\rho_\zeta = 0.2$.

To visualize how β_1^{FE} varies more continuously with the reaction parameter θ, Figure 1 plots its value over the range $\theta \in [0, 8.5]$ for three fixed values of distortion persistence, $\rho_\zeta \in \{0, 0.1, 0.2\}$, holding all other parameters constant. Similarly, Figure 2 shows how β_1^{FE} varies with $\rho_\zeta \in [0, 0.3]$, for fixed values of $\theta \in \{1, 4, 7\}$. Together, Figures 1 and 2 show that β_1^{FE} is strongly decreasing in θ, while it is essentially unaffected by ρ_ζ. These patterns corroborate the comparative statics in Table 6, confirming that θ governs the strength of on-impact forecast bias, whereas ρ_ζ plays little role. Figures A1 to A4 in Appendix A.1 plot β_1^{FE} against the reaction parameter θ and belief distortion persistence ρ_ζ for various values of the true earnings growth persistence ρ_g, as well as against ρ_g for different fixed values of θ and ρ_ζ. These figures further confirm that only θ meaningfully affects β_1^{FE}; both ρ_ζ and ρ_g have minimal impact.

We now turn to β_2^{FE}, the coefficient on lagged expectations in the forecast error regression in Equation (19), which captures the persistence of belief-driven forecast errors beyond the initial revision. Figures 3 and 4 plot β_2^{FE} against θ for various values of ρ_ζ, and against ρ_ζ for various values of θ, respectively. In both cases, β_2^{FE} decreases as either θ or ρ_ζ increases, consistent with the comparative statics in Table 6 and the model's intuition: stronger or more persistent belief distortions lead to more pronounced and predictable forecast biases. Figures A5 to A8 in Appendix A.1 plot β_2^{FE} against the reaction parameter θ and belief distortion persistence ρ_ζ for various values of the true earnings growth persistence ρ_g, as well as against ρ_g for different fixed values of θ and ρ_ζ.

Figure 4: Effect of Belief Persistence ρ_ζ on β_2^{FE}, Across Overreaction Levels θ



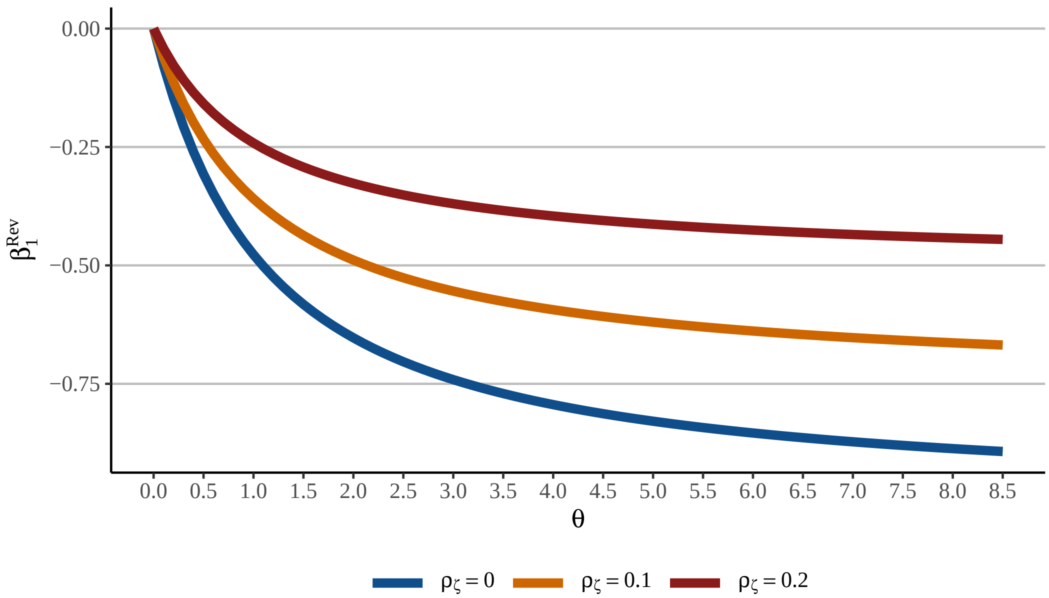
Notes: This figure reports the comparative statics of β_2^{FE} with respect to ρ_ζ, holding $\rho_g = 0.4$, $\sigma_u = 0.02$, and $s = 5$ fixed. Each line corresponds to a different fixed value of θ: the blue line uses $\theta = 1$, the orange line uses $\theta = 4$, and the red line uses $\theta = 7$.

These figures show that ρ_g has a modest positive effect on β_2^{FE}. As the true growth process becomes more persistent, belief distortions that also persist generate forecasts that are less biased over time. In other words, having more persistent fundamentals can help offset the bias in forecasts.

Finally, we examine β_1^{Rev}, the coefficient on lagged LTG expectations in the LTG revision regression in Equation (20), which reflects the strength in which expectations revert. Figures 5 and 6 plot β_1^{Rev} against θ for various values of ρ_ζ, and against ρ_ζ for various values of θ, respectively. Both figures confirm the earlier findings: β_1^{Rev} becomes more negative as θ increases, but less negative as ρ_ζ increases. Stronger overreaction leads to sharper reversals in expectations, while more persistent belief distortions slow the reversion process. Figures A9 to A12 in Appendix A.1 plot β_1^{Rev} against the reaction parameter θ and belief distortion persistence ρ_ζ for various values of the true earnings growth persistence ρ_g, as well as against ρ_g for different fixed values of θ and ρ_ζ. As ρ_g increases β_1^{Rev} becomes more negative, indicating more mean reversion. A more persistence true fundamental means that on average means that the true process mean reverts faster, which on average produces news that eventually reverts the biased expectations.

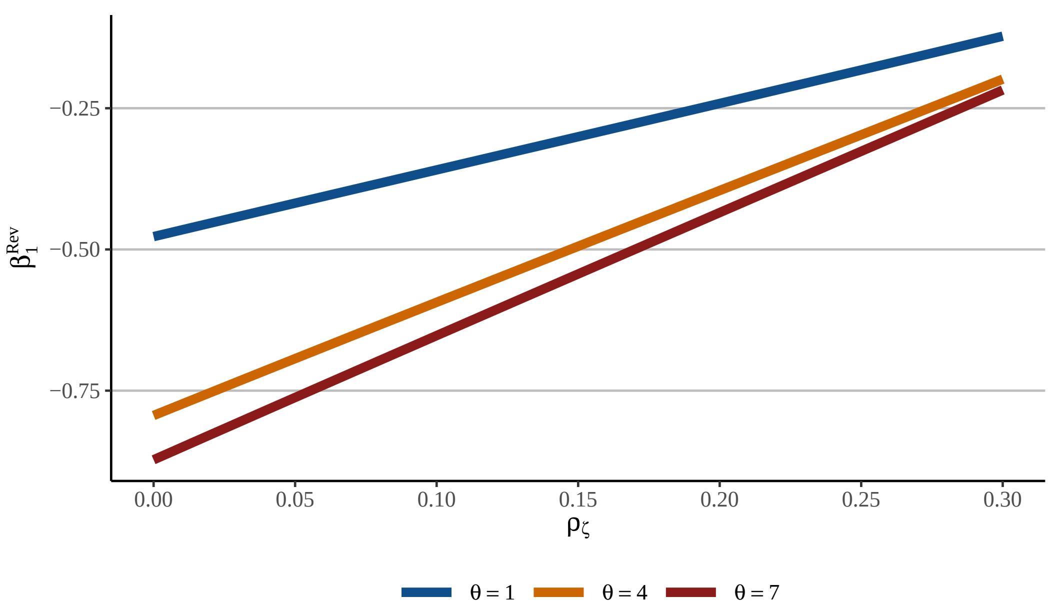
Given these comparative statics results, we now turn to the estimation of the structural parameters using empirical moments. The insights developed above not only illustrate how each parameter influences the regression coefficients, but also clarify which empirical patterns provide identification. These relationships guide and validate our estimation and inform the interpretation of the fitted parameters in the next section.

Figure 5: Effect of Overreaction θ on β_1^{Rev}, Across Belief Persistence ρ_ζ



Notes: This figure reports the comparative statics of β_1^{Rev} with respect to θ, holding $\rho_g = 0.4$, $\sigma_u = 0.02$, and $s = 5$ fixed. Each line corresponds to a different fixed value of ρ_ζ: the blue line uses $\rho_\zeta = 0$, the orange line uses $\rho_\zeta = 0.1$, and the red line uses $\rho_\zeta = 0.2$.

Figure 6: Effect of Belief Persistence ρ_ζ on β_1^{Rev}, Across Overreaction Levels θ



Notes: This figure reports the comparative statics of β_1^{Rev} with respect to ρ_ζ, holding $\rho_g = 0.4$, $\sigma_u = 0.02$, and $s = 5$ fixed. Each line corresponds to a different fixed value of θ: the blue line uses $\theta = 1$, the orange line uses $\theta = 4$, and the red line uses $\theta = 7$.

Table 7: Realized LTG AR1 (NAICS2 × Year FE)

	Realized LTG		
	(1)	(2)	(3)
Realized LTG (Lag 1)	0.363	0.372	0.369
	(0.014)	(0.012)	(0.013)
Realized LTG (Lag 1) × Low Market Power (Lag 1)	0.027		0.021
	(0.033)		(0.033)
Realized LTG (Lag 1) × High Market Power (Lag 1)		-0.056	-0.054
		(0.028)	(0.028)
NAICS2 FE	No	No	No
NAICS2 × Year FE	Yes	Yes	Yes
Firm FE	No	No	No
Observations	22,000	22,000	22,000

Notes: This table presents regressing future LTG revisions against lagged LTG along with interactions. All specifications include NAICS2 × year fixed effects. Column (1) presents the specification with only an interaction for low total market power firms. Column (2) presents the specification with only an interaction for high total market power firms. Column (3) presents the specification with interactions with both low total market power and high total market power firms. Standard errors are reported in parentheses and are two-way clustered by firm and year. All figures are rounded in accordance with U.S. Census disclosure requirements.

4.2 Parameter Estimation

We estimate the parameters $(\theta, \rho_\zeta, \rho_g, \sigma_u)$ separately for three groups of firms: those with low market power, those with high market power, and a residual group comprising the remainder of the sample. Estimation is based on the empirical values of β_1^{FE}, β_2^{FE}, and β_1^{Rev}. To identify ρ_g and σ_u, we supplement these regressions by estimating the AR(1) process for realized long-term earnings growth and by matching the model-implied variance of LTG revisions to its empirical counterpart. Together, this yields five moments for each firm group, which we use to estimate four parameters per group. This results in a system with 15 moments and 12 parameters across the three groups, yielding overidentifying restrictions that allow us to assess model fit.

Tables 7 and 8 report regression estimates used to recover two key empirical moments: the AR(1) coefficient of the real long-term growth (LTG) process, and the cross-sectional standard deviation of LTG revisions, respectively. In Table 7, we compute the realized 5-year average earnings growth rate for each firm and regress it on its lag, interacted with indicators for high-markup and high-markdown firms. We focus on Column (3) to recover estimates of the AR(1) parameter, which captures the persistence of the underlying earnings growth process.

Table 8 exploits the monthly frequency of IBES forecasts to compute firm-level standard deviations of LTG revisions at the annual level. These firm-year measures are merged into the annual

Table 8: LTG Revision SD

	LTG Revision SD			
	(1)	(2)	(3)	(4)
Intercept	0.016			
	(0.001)			
Low Market Power	-0.001	0.000	-0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
High Market Power	-0.003	-0.003	-0.002	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)
NAICS2 FE	No	Yes	No	No
NAICS2 $\times$ Year FE	No	No	Yes	Yes
Firm FE	No	No	No	Yes
Observations	35,500	35,500	35,500	35,500

Notes: This table presents regressing annual LTG revision SD against dummies for low total market power and high total market power firms. Column (1) presents the specification without any fixed effects. Column (2) presents the specification with NAICS2 fixed effects. Column (3) presents the specification with NAICS2 $\times$ year fixed effects. Column (4) presents the specification with NAICS2 $\times$ year and firm fixed effects. Standard errors are reported in parentheses and are two-way clustered by firm and year. All figures are rounded in accordance with U.S. Census disclosure requirements.

panel and regressed on indicators for high-markup and high-markdown firms under various fixed effects specifications. As the results are broadly consistent across specifications, we rely on the estimates from Column (1) for simplicity.

These empirical moments serve as inputs into the model-based estimation. Since the model yields closed-form expressions for all targeted moments, we estimate the structural parameters Θ by minimizing the nonlinear least squares distance between empirical and model-implied moments. Let $m(\Theta)$ denote the vector of model-implied moments and $\hat{m}$ the corresponding empirical estimates. The parameter vector $\hat{\Theta}$ solves

$$\hat{\Theta} = \arg\min_{\Theta} \left(m(\Theta) - \hat{m}\right)^{\intercal} W \left(m(\Theta) - \hat{m}\right), \tag{26}$$

where W is a symmetric, positive semi-definite weighting matrix estimated using a clustered bootstrap procedure, following Horowitz (2001) and Cameron and Miller (2015). Additional details on the estimation procedure are provided in Appendix D.

Table 9 reports the parameter estimates and bootstrap standard errors for each firm group; note that the persistence parameters and variances are estimated at the annual level. The results

Table 9: Estimation of Diagnostic Expectations Parameters

	Baseline		Low Market Power		High Market Power	
	Estimate (1)	Standard Error (2)	Estimate (3)	Standard Error (4)	Estimate (5)	Standard Error (6)
θ	6.649	0.816	8.101	5.014	5.009	2.743
ρ_ζ	0.163	0.008	0.258	0.028	0.135	0.024
ρ_g	0.399	0.012	0.387	0.039	0.304	0.031
σ_u	0.023	0.008	0.001	0.010	0.019	0.046

Notes: This table presents the estimated model parameters $(\theta, \rho_\zeta, \rho_g, \sigma_u)$ using the regression coefficients as moments. Columns (1) and (2) contain the estimate and standard error, respectively for the baseline firm. Columns (3) and (4) contain the estimate and standard error, respectively for the low total market power firms. Columns (5) and (6) contain the estimate and standard error, respectively for the high total market power firms. Standard errors are computed via the Delta method following Hansen (1982), Hansen and Singleton (1982), and Newey and McFadden (1994).

Table 10: Model vs. Data Moment Comparison

	Baseline		Low Market Power		High Market Power	
	Model (1)	Data (2)	Model (3)	Data (4)	Model (5)	Data (6)
β_1^{Rev}	-0.511	-0.504	-0.297	-0.387	-0.462	-0.457
β_1^{FE}	-0.847	-0.829	-0.879	-0.790	-0.815	-0.811
β_2^{FE}	-0.727	-0.848	-0.784	-0.844	-0.681	-0.684
ρ_g	0.399	0.369	0.387	0.390	0.304	0.315
$\text{SD}[\Delta\text{LTG}_{i,t}]$	0.001	0.016	0.000	0.016	0.000	0.013

Notes: This table presents the model implied moments and compares them with the empirical moments. Columns (1) and (2) show the model implied moments and the data moments, respectively, for the baseline case. Columns (3) and (4) show the model implied moments and the data moments, respectively, for the low total market power case. Columns (5) and (6) show the model implied moments and the data moments, respectively, for the high total market power case.

differences across groups, particularly in the overreaction parameter θ and the persistence of belief distortions ρ_ζ. High markup firms exhibit the strongest overreaction, with $\theta = 8.101$, and the most persistent distortions, with $\rho_\zeta = 0.258$. In contrast, high markdown firms display the weakest overreaction ($\theta = 5.009$) and the least persistent distortions ($\rho_\zeta = 0.135$). Baseline firms fall between these extremes on both dimensions. The parameters governing the persistence of the earnings growth process ρ_g and volatility of total news σ_u are more similar across groups, with only modest variation in the persistence of fundamentals.

We report the model-implied moments alongside their empirical counterparts in Table 10. Overall, the estimated parameters generate model moments that closely match the data. As shown in the comparative statics analysis in Section 4.1, since neither ρ_ζ nor ρ_g have a significant impact on β_1^{FE}, the parameter θ is primarily identified by β_1^{FE}. The AR(1) regression of realized LTG determines ρ_g. Conditional on these, the residual variation in β_2^{FE} and β_1^{Rev} largely identifies ρ_ζ. Finally, in the expressions for the regression coefficients, σ_u cancels out in both the numerator and denominator, implying that its level is identified directly from the volatility of LTG revisions.

Given this identification strategy and the model's ability to match observed moments, the estimated values of θ are striking in magnitude. The firm-level overreaction parameters are an order of magnitude larger than prior estimates at the aggregate level, indicating substantial overreaction relative to the size of new information. For example, Bordalo, Gennaioli and Shleifer (2018) estimate $\theta = 0.91$ in the U.S. investment-grade corporate bond market, while Bordalo et al. (2019) report similar values when comparing firms with high versus low LTG expectations in response to aggregate sentiment shocks. At the macroeconomic level, Bordalo et al. (2020) find an average $\theta = 0.5$ across forecast revisions.

A distinctive feature of our analysis is the focus on firm-level expectations, rather than aggregate or portfolio-level dynamics. Conceptually, this parallels the literature on idiosyncratic versus aggregate volatility, where firm-level return volatility far exceeds market-level volatility (e.g., Campbell et al., 2001). In the same way, overreaction in expectations may be more pronounced at the micro level, reflecting firms' exposure to idiosyncratic noise or behavioral biases. More broadly, the literature documents wide variation in the degree of over- or underreaction depending on the level of aggregation, information sets, and identification strategies employed (e.g., Coibion and Gorodnichenko, 2015; Kohlhas and Walther, 2021; Broer and Kohlhas, 2024).[16]

5 Macro-Finance Model with Diagnostic Expectations

In this section, we develop a simple macro-finance model in partial equilibrium that incorporates diagnostic expectations and heterogeneous firms. We investigate whether a simple model with these features can generate standard asset pricing and macroeconomic moments as well as generate predictable return reversals. Then we study potential implications of the model given its fit.

The model features firms that differ in their productivity, market power, and belief distortions, the latter captured through diagnostic expectation parameters. We abstract from household behavior and general equilibrium channels to focus on firm-level dynamics and isolate the effects of distorted beliefs on outcomes such as investment, capital accumulation, and asset returns. The model also allows us to examine macroeconomic distortions such as misallocation, following the spirit of Hsieh and Klenow (2009) and David, Schmid and Zeke (2022). We first describe the environment and firm problem, then turn to the aggregation concept and calibration strategy.

[16] See Angeletos, Huo and Sastry (2021) for a comprehensive review.

5.1 Environment

There is a unit mass of atomistic firms indexed by i, each belonging to one of J discrete types. Let N_j denote the fixed mass of firms of type j, where $\sum_{j=1}^{J} N_j = 1$. Firms are representative within type, and we denote by $j(i)$ the type of firm i. Firms operate as local monopolists in their product markets and local monopsonists in their labor markets, while capital markets are assumed to be competitive. Each firm type is characterized by a fixed tuple of parameters governing technology, market power, and belief distortions.

Given their market power in both product and labor markets, firms face exogenous residual demand and labor supply curves. For simplicity, firms are unlevered and issue a single security. Output is sold to a competitive final goods producer, who aggregates heterogeneous firm-level production into a single final consumption good that serves as the numeraire. Time is discrete and indexed by t.

5.2 Firms

Each firm accumulates physical capital $k_{i,t}$ and hires labor $l_{i,t}$ to produce output $y_{i,t}$ according to the Cobb-Douglas production function

$$y_{i,t} = \omega_{i,t} k_{i,t}^{\alpha} l_{i,t}^{1-\alpha}, \tag{27}$$

where $\omega_{i,t}$ is the firm's Hicks-neutral productivity level, and $\alpha \in (0,1)$ is the capital share. Productivity evolves as a stationary AR(1) process

$$\ln \omega_{i,t} = \rho_{\omega} \ln \omega_{i,t-1} + (1-\rho_{\omega}) \ln \bar{\omega}_{j(i)} + \tau_{i,t}, \tag{28}$$

where $\rho_{\omega} \in (0,1)$ governs persistence, $\bar{\omega}_{j(i)}$ is the type-specific mean productivity, and $\tau_{i,t} \sim \mathcal{N}(0, \sigma_{\tau}^2)$ is an i.i.d. shock.

Firms face exogenous CES residual product demand and labor supply curves

$$y_{i,t} = p_{i,t}^{-\varepsilon_{j(i)}}, \tag{29}$$

$$l_{i,t} = w_{i,t}^{\eta_{j(i)}}, \tag{30}$$

where $p_{i,t}$ is the firm's price, $\varepsilon_{j(i)} > 1$ is the product demand elasticity, $w_{i,t}$ is the offered wage, and $\eta_{j(i)} > 0$ is the labor supply elasticity.[17] These elasticities are type-specific. Equations (29) and (30)

[17] Tables A12 to A17 in Appendix A.2 report estimates of the persistence of price markups, wage markdowns, and the total wedge across various specifications. All results indicate that firm-level market power is highly persistent. For tractability, we therefore assume constant markups and markdowns at the firm level. For more on homothetic aggregators that generalize CES demand and allow for endogenous firm-level variation in price markups and wage markdowns, see Matsuyama and Ushchev (2017) and Matsuyama (2025).

imply constant price markups and wage markdowns given by

$$\mu_{j(i)} = \frac{\varepsilon_{j(i)}}{\varepsilon_{j(i)} - 1}, \tag{31}$$

$$\nu_{j(i)} = \frac{\eta_{j(i)} + 1}{\eta_{j(i)}}. \tag{32}$$

Capital evolves according to the standard law of motion

$$k_{i,t+1} = (1-\delta)k_{i,t} + i_{i,t}, \tag{33}$$

where $\delta \in (0,1)$ is the depreciation rate and $i_{i,t}$ denotes investment. Firms face convex capital adjustment costs given by

$$\phi(i_{i,t}, k_{i,t}) = \frac{\psi}{2}\left(\frac{i_{i,t}}{k_{i,t}} - \delta\right)^2 k_{i,t}, \tag{34}$$

where $\psi > 0$ is a scaling parameter governing the size of adjustment frictions. The firm's current-period dividends are therefore

$$\pi_{i,t} = p_{i,t}y_{i,t} - w_{i,t}l_{i,t} - i_{i,t} - \phi(i_{i,t}, k_{i,t}). \tag{35}$$

The firm's dynamic problem is given by

$$\begin{gathered} V_i(\omega_{i,t}, \zeta_{i,t}, k_{i,t}) = \max_{l_{i,t}, i_{i,t}} \pi_{i,t} + \beta \tilde{\mathbb{E}}_t \left[V_i(\omega_{i,t+1}, \zeta_{i,t+1}, k_{i,t+1}) \right] \\ \text{subject to} \\ y_{i,t} = \omega_{i,t} k_{i,t}^{\alpha} l_{i,t}^{1-\alpha}, \\ y_{i,t} = p_{i,t}^{-\varepsilon_{j(i)}}, \\ l_{i,t} = w_{i,t}^{\eta_{j(i)}}, \\ k_{i,t+1} = (1-\delta)k_{i,t} + i_{i,t}, \end{gathered} \tag{36}$$

where $\tilde{\mathbb{E}}_t[\cdot]$ denotes diagnostic expectations conditional on information at time t, and $\beta \in (0,1)$ is the discount factor. The belief distortion $\zeta_{i,t}$ follows an AR(1) process as in Equation (17), with type-specific persistence $\rho_{\zeta,j(i)}$ and innovation $u_{i,t} = \theta_{j(i)}\rho_\omega \tau_{i,t}$.

Finally, the realized gross unlevered return of firm i in period $t+1$ is

$$R_{i,t+1} = \frac{V_i(\omega_{i,t+1}, \zeta_{i,t+1}, k_{i,t+1})}{V_i(\omega_{i,t}, \zeta_{i,t}, k_{i,t}) - \pi_{i,t}}, \tag{37}$$

and the implied gross risk-free rate is $R_f = \beta^{-1}$.

Following De La O and Myers (2021) and Bordalo et al. (2024*a*), variation in returns is driven entirely by cash flows or changes in beliefs about future cash flows, rather than discount rate variation, as discount rates are constant across firms and over time. Since we relax rational expectations, the Modigliani-Miller conditions of Modigliani and Miller (1958) need not hold, and a simple leverage factor cannot be used to recover equity returns from firm value. We focus on unlevered returns for tractability and to obtain conservative measures of return variation.[18]

5.3 Aggregation

A representative final goods producer aggregates the vector of intermediate outputs $\mathbf{y}_t$ into a single final good Y_t using the production technology:

$$Y_t = F(\mathbf{p}_t, \mathbf{y}_t). \tag{38}$$

This sector is competitive and we impose the accounting identity

$$P_t Y_t = \mathbf{p}_t^\intercal \mathbf{y}_t, \tag{39}$$

where $\mathbf{p}_t$ is the vector of firm-level prices. Without loss of generality, we normalize the price of the final good to one, i.e., $P_t = 1$.

Aggregate inputs are defined as

$$K_t = \int_0^1 k_{i,t}\, di, \tag{40}$$

$$L_t = \int_0^1 l_{i,t}\, di. \tag{41}$$

We define the aggregate wage as the expenditure-weighted average

$$W_t = \frac{1}{L_t} \int_0^1 w_{i,t} l_{i,t}\, di. \tag{42}$$

Given these definitions and the firm-level production function in Equation (27), we define aggregate total factor productivity (TFP) as the residual from the aggregate Cobb-Douglas production function

$$\Omega_t = \frac{Y_t}{K_t^\alpha L_t^{1-\alpha}}. \tag{43}$$

5.4 Calibration

Before evaluating the model's quantitative performance, we outline the calibration strategy. We set $J = 3$ to reflect our empirical strategy: (i) normal firms, (ii) low market power firms, and (iii) high market power firms. The model is solved under two informational assumptions: one with diagnostic expectations, using estimated belief distortion parameters from Section 4.2, and one

[18]See Boldrin, Christiano and Fisher (2001) and Papanikolaou (2011) for examples of this adjustment.

Table 11: Common Parameters

Parameter	Description	Value
α	Input Intensity of Capital	0.3
β	Subject Time Discount Rate	0.96
δ	Depreciation Rate of Capital	0.1
ρ_ω	Productivity Persistence	0.5319
σ_τ	Productivity Innovation SD	0.01
ψ	Capital Adjustment Cost	5

Notes: This table presents the parameter values that are common across all model specifications.

Table 12: Persistence of Productivity (NAICS2 $\times$ Year and Firm FEs)

	Log Labor Productivity			
	(1)	(2)	(3)	(4)
Log Labor Productivity (Lag 1)	0.532	0.536	0.501	0.505
	(0.018)	(0.018)	(0.014)	(0.013)
Log Labor Productivity (Lag 2)			0.084	0.088
			(0.012)	(0.012)
Log Sale (Lag 1)		-0.017		-0.026
		(0.008)		(0.009)
NAICS2 FE	No	No	No	No
NAICS2 $\times$ Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Observations	60,000	60,000	52,500	52,500

Notes: This table presents various specifications regressing the log labor productivity onto its own lags and various controls. All specifications include NAICS2 $\times$ year and firm fixed effects. Column (1) presents an AR(1) specification. Column (2) adds the first lag of log sales as a control. Column (3) estimates an AR(2) specification, and Column (4) adds the first lag of log sales to the AR(2) model. Standard errors are reported in parentheses and are two-way clustered by firm and year. All figures are rounded in accordance with U.S. Census disclosure requirements.

with rational expectations, corresponding to the special case where $\theta_{j(i)} = 0$ for all $i \in [0,1]$.[19] The model is calibrated and solved at the annual level.

Table 11 presents the parameter values that are common across all firm types and model variants. We set the capital intensity in production to $\alpha = 0.3$, consistent with standard values in the macro and growth literature. Under competitive markets, this implies a labor share of 0.7. The discount factor $\beta = 0.96$ corresponds to an annual risk-free rate of approximately 4%, in line with common

[19] Appendix E for the computational method to solve this model numerically.

Table 13: Varying Parameters under Diagnostic Expectations

Parameter	Description	Normal	Low Market Power	High Market Power
θ	Overreaction	6.649	8.101	5.009
ρ_ζ	Belief Distortion Persistence	0.163	0.258	0.135
ε	Product Demand Elasticity	10	8	12
η	Labor Supply Elasticity	4	8	2
$\bar{\omega}$	Productivity Level	2.574	1	7.765
N	Firm Count	2000	250	250

Notes: This table presents the parameters under diagnostic expectations across normal, low total market power, and high total market power firm types.

Table 14: Varying Parameters under Rational Expectations

Parameter	Description	Normal	Low Market Power	High Market Power
θ	Overreaction	0	0	0
ρ_ζ	Belief Distortion Persistence	NA	NA	NA
ε	Product Demand Elasticity	10	8	12
η	Labor Supply Elasticity	4	8	2
$\bar{\omega}$	Productivity Level	2.574	1	7.765
N	Firm Count	2000	250	250

Notes: This table presents the parameters under rational expectations across normal, low total market power, and high total market power firm types.

asset pricing and macroeconomic calibrations. The depreciation rate $\delta = 0.1$ reflects standard estimates of physical capital depreciation.

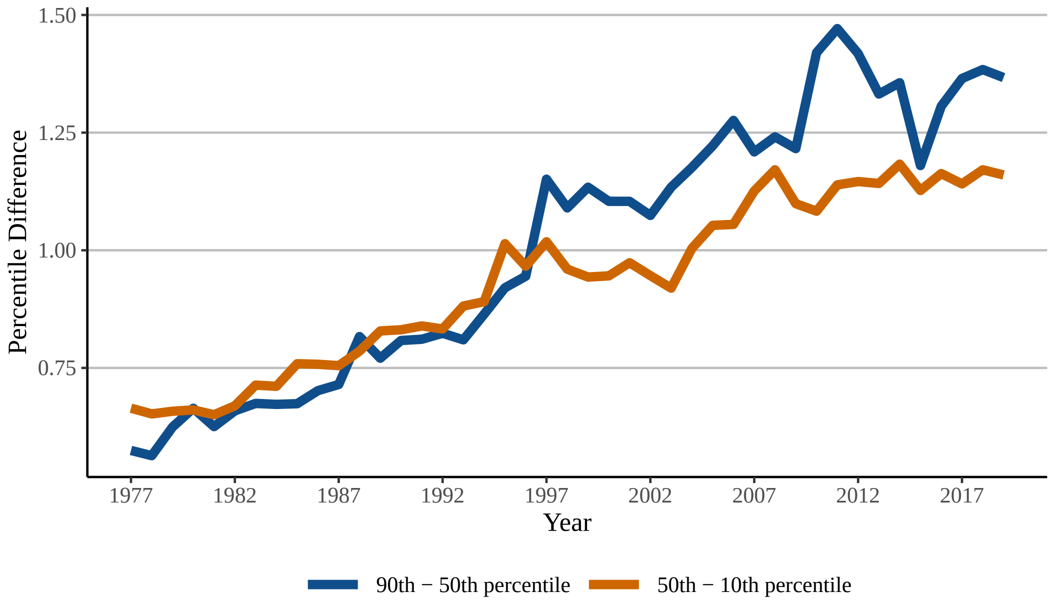
The productivity process is moderately persistent, with $\rho_\omega = 0.5319$, and features innovation volatility $\sigma_\tau = 0.01$, which is consistent with the calibration of Kydland and Prescott (1982) and King, Plosser and Rebelo (1988). Table 12 reports productivity persistence estimates under various fixed effects specifications. We also present additional specifications in Table A18 in Appendix A.2. Our preferred specification includes firm fixed effects to capture differences in unconditional firm-level productivity means. We take the estimate from Column (1) of Table 12, which most closely aligns with the structure of the productivity process in our model.

Finally, the capital adjustment cost parameter $\psi = 5$ is set to smooth investment dynamics and prevent excessive volatility in capital accumulation. This value lies within the range commonly used in models with convex adjustment costs.[20]

Tables 13 and 14 report the parameters that vary across firm types under both diagnostic expectations (DE) and rational expectations (RE). For the DE specification, the belief distortion parameters θ and ρ_ζ are taken from the estimated values in Table 9. The product demand and

[20] See Hayashi (1982); Abel and Eberly (1996, 1999); Jermann (1998); Cooper (2006); Cooper and Haltiwanger (2006); and Bloom (2009) for further discussion on capital adjustment costs and their applications in finance and macroeconomics.

Figure 7: Productivity Cross-Sectional Dispersion



Notes: This figure reports the difference between the 90th percentile and 50th percentile log labor productivity (blue line) and the difference between the 50th percentile and 10th percentile log labor productivity (orange line) across time. All figures are rounded in accordance with U.S. Census disclosure requirements.

labor supply elasticities are chosen to target varying levels of market power across types. These imply price markups and wage markdowns of 1.11 and 1.25 for normal firms, 1.14 and 1.13 for low market power firms, and 1.09 and 1.50 for high market power firms, respectively. These patterns reflect the empirical negative relationship between price markups and wage markdowns, as well as the relative dispersion in market power across firms, as documented in Ren and Zhang (2025).[21]

The type-specific mean productivity levels $\bar{\omega}$ are set to match observed heterogeneity in firm productivity. Specifically, we assign values based on the difference between the 90th and 50th percentiles and the 50th and 10th percentiles of log labor productivity in the year 2000. Figure 7 plots the time series of these cross-percentile differences. We align these relative productivity levels with firm types following the empirical findings of Ren and Zhang (2025), who document that firm-level productivity is positively correlated with wage markdowns and negatively correlated with price markups. The resulting values imply that high market power firms are 7.8 times more productive than low market power firms, and that normal firms are 2.6 times more productive than low market power firms. Finally, we set the relative firm counts across types to reflect our empirical strategy of focusing on firms in the top decile of price markups and wage markdowns, while ensuring that the total firm mass is consistent with the approximate number of publicly traded firms in the United States.

[21] See also Tables A2 and A3 in Appendix A.2 for regression results from Ren and Zhang (2025).

Table 15: Firm-Level Model vs. Empirical Moments

	Baseline Firms		Low Market Power		High Market Power	
Statistic	Model	Data	Model	Data	Model	Data
I/K Ratio	0.100	0.099	0.101	0.036	0.100	0.097
Labor Share	0.504	0.678	0.544	0.765	0.428	0.493
Dividend / Profit Rate	0.304	0.224	0.262	0.199	0.377	0.410
Return (1 yr)	0.098	0.232	0.138	0.304	0.064	0.140

Notes: This table presents the various real and financial moments from the data and the model. Columns (1) and (2) show the moments for the model and data moments, respectively for normal firms. Columns (3) and (4) show the same for low market power firms and Columns (5) and (6) show the same for high market power firms.

6 Quantitative Results

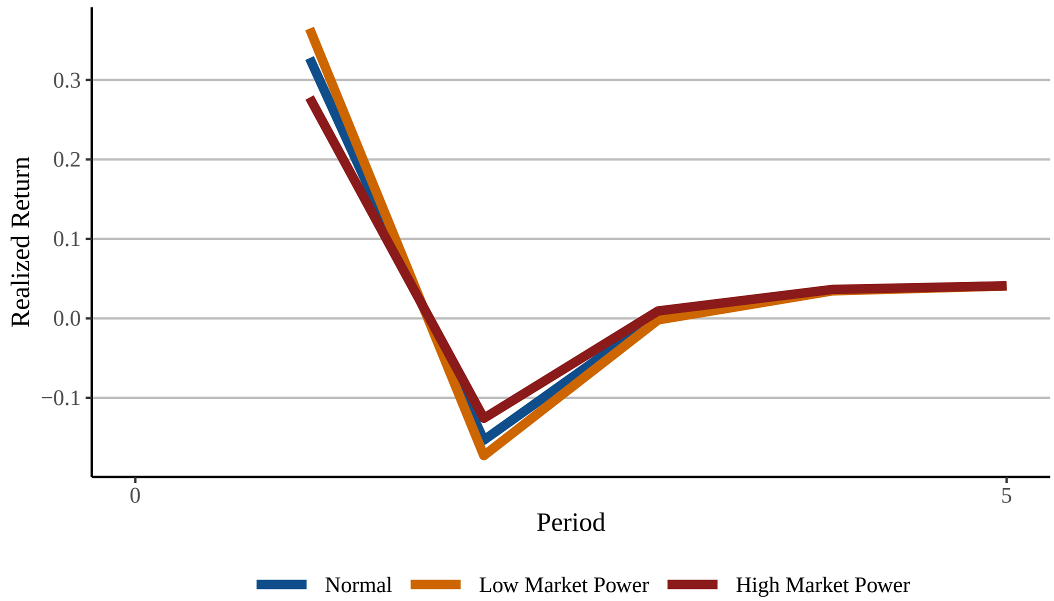
This section presents the quantitative implications of the model described in Section 5, based on the calibration strategies outlined in Section 5.4. We evaluate the model's ability to replicate key moments in the data and examine how the presence of diagnostic expectations alters both firm-level behavior and aggregate dynamics relative to the benchmark rational expectations case. The rest of this section is organized as follows: Section 6.1 focuses on firm-level outcomes and Section 6.2 turns to aggregate dynamics.

6.1 Firm-Level Results

We simulate a panel of 2,500 firms (comprising of 2,000 baseline firms, 250 low market power firms, and 250 high market power firms) using the parameter configurations outlined in Tables 13 and 14. Each firm is simulated over 200 periods (years), and the productivity process is assumed to evolve independently across firms. This abstraction isolates the idiosyncratic component of firm dynamics, allowing us to cleanly identify how diagnostic expectations alter firm-level behavior in the absence of macroeconomic shocks. In Section 6.2, we introduce an aggregate productivity shock and assess the model's predictions under partial equilibrium aggregation.

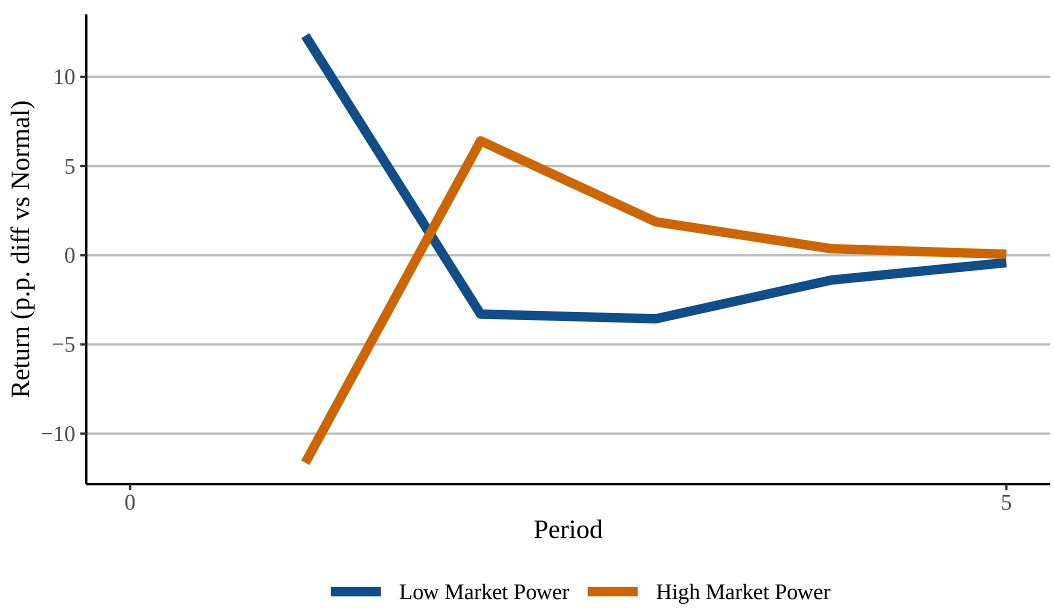
Table 15 shows the moments comparing the data with the model at the firm-level. The model is able to generate the correct cross-sectional patterns for various real variables (investment-to-capital ratio, labor share, dividend/profit rate) as well as returns. The model, is also able to generate substantial returns despite it being an otherwise standard production model. The subjective time discount factor is fixed at $\beta = 0.96$, implying an annual net risk-free rate of approximately 4.2%. This implies that all firm types generate substantial risk premia. While the standard model under rational expectations struggles to generate substantial risk premia, as emphasized by Mehra and Prescott (1985), Hansen and Jagannathan (1991), Jermann (1998), and Boldrin, Christiano and Fisher (2001), under diagnostic expectations it is able to generate these easily.

Figure 8: Impulse Response of Return – Across Firm Types



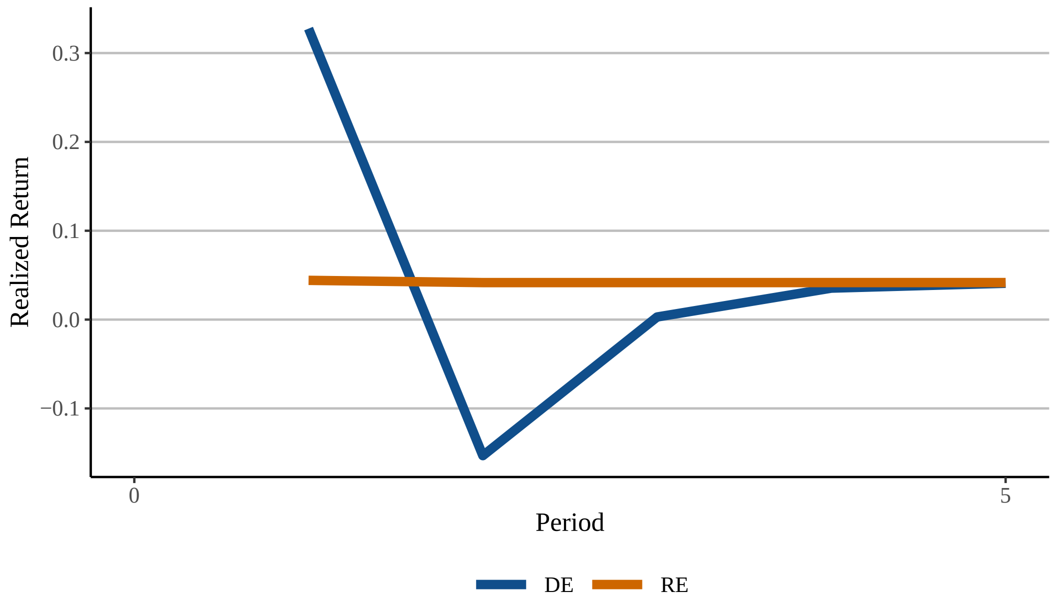
Notes: This figure reports the impulse response function of returns for all firms. The blue line denotes normal firms, the orange line denotes low market power firms, and the red line denotes the high market power firms. The shock is a 1 SD positive productivity shock.

Figure 9: Difference in Returns Across Firm Types



Notes: This figure reports the differences in returns in response to a 1 SD positive productivity shock relative to normal firms. The blue line denotes the difference for low market power firms and the orange line denotes the difference for high market power firms.

Figure 10: Impulse Response of Return – Normal Firms (DE vs. RE)



Notes: This figure reports the impulse response function of returns for normal firms. The blue line denotes IRF under diagnostic expectations and the orange line shows the IRF under rational expectations. The shock is a 1 SD positive productivity shock.

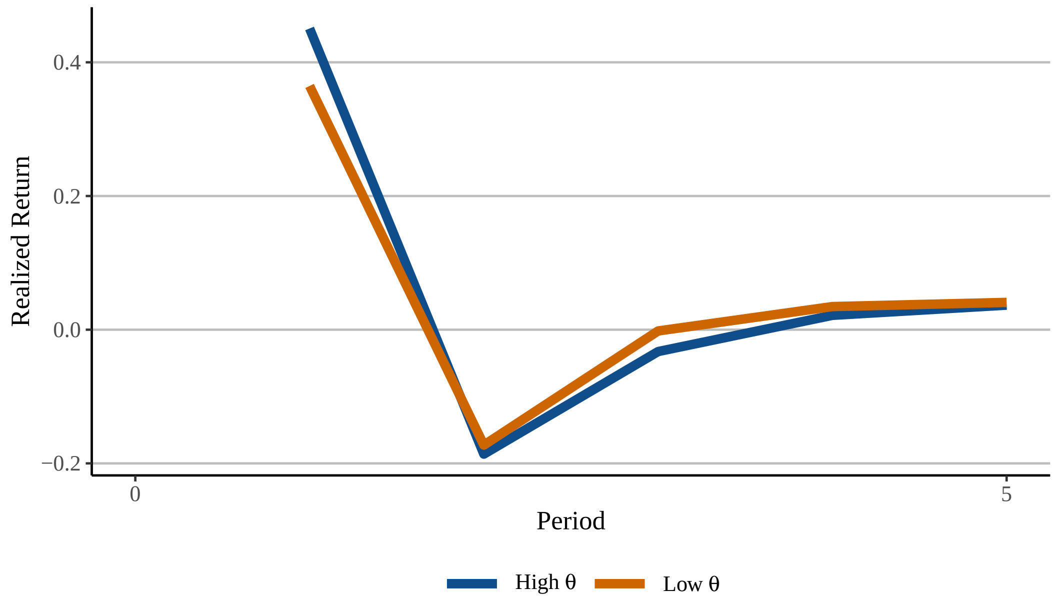
First, we examine the response of returns across all firm types with respect to a positive productivity shock to evaluate whether the model can generate not only high excess returns but also predictable reversals. Figure 8 reports the impulse responses of gross returns across firm types. All firm types display volatile returns and predictable reversals, consistent with the empirical evidence in Section 3.2. Figure 9 plots the difference in returns between low market power and high market power firms relative to normal firms in response to the shock. On impact, realized returns rise sharply, peaking between 20% and 30%, before reversing and reaching troughs between -15% and -10%. Low market power firms exhibit the most pronounced reversal, whereas high market power firms display the mildest, which is also consistent with the data.

Figure 10 presents the IRFs of return for normal firms under both diagnostic expectations (blue line) and rational expectations (orange line).[22] This figure shows that the model with RE cannot generate high excess and volatile returns as expected. Furthermore, it cannot generate a predictable reversal, instead returns are very smooth and are close to the risk-free rate.

We next consider alternative specifications in which either the degree of overreaction is fixed or market power is fixed while overreaction varies. Figure 11 shows the IRF of gross returns for a low market power firm under two values of θ. The blue line corresponds to the estimated level of overreaction, while the orange line reports the response when θ is set to the value estimated for

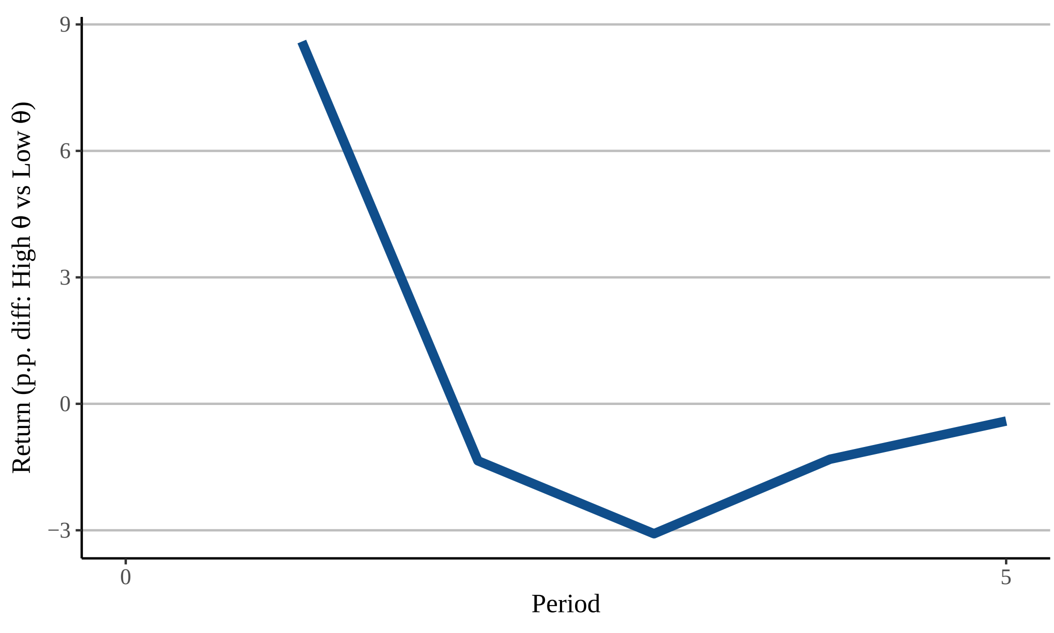
[22] Figures A13 and A14 show the IRFs for returns for low market power and high markdown firms, respectively, in Appendix A.1.

Figure 11: Impulse Response of Return – Low Market Power Firm Different Overreaction Levels



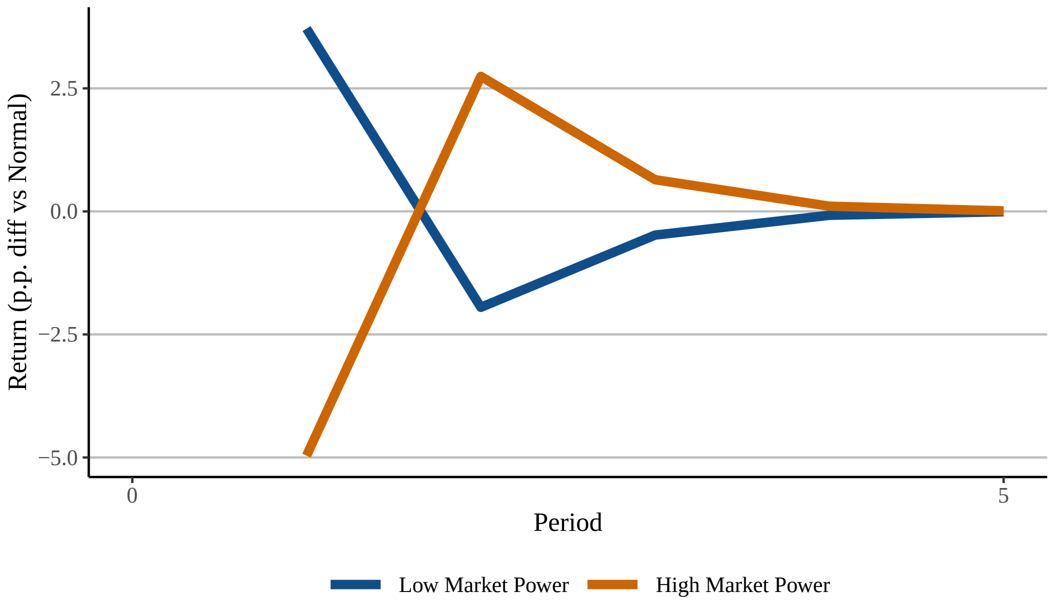
Notes: This figure reports the impulse response function of returns for low market power firms across different levels of overreaction. The blue line denotes the case in which overreaction is set at the high level (the estimated level for low market power firms). The orange line denotes the case in which overreaction is set at the low level (the medium value among the three estimated values).

Figure 12: Difference in Returns Across Firm Types Fixing Market Power



Notes: This figure reports the difference in returns in response to a 1 SD positive productivity shock between high and low overreaction levels for low market power firms.

Figure 13: Difference in Returns Across Firm Types Fixing Overreaction



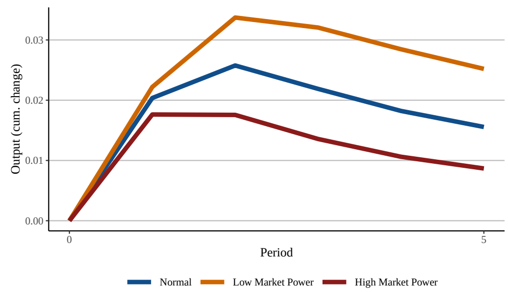
Notes: This figure reports the differences in returns in response to a 1 SD positive productivity shock relative to normal firms while fixing the level of overreaction to that of normal firms for all firms. The blue line denotes the difference for low market power firms and the orange line denotes the difference for high market power firms.

normal firms. Figure 12 shows the difference in returns across these specifications: with higher overreaction, the initial return is 8.6 percentage points higher, while the reversal peaks at 3.1 percentage points lower. Thus, greater overreaction produces more volatile returns and deeper reversals, in line with the data.

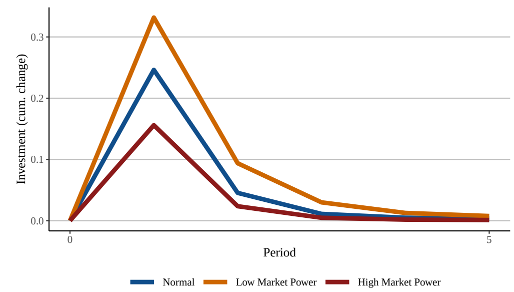
Finally, Figure 13 reports the difference in returns relative to normal firms for low market power and high market power firms when the overreaction parameter θ is fixed at the level estimated for normal firms. This isolates the role of heterogeneity in market power for predictable returns under diagnostic expectations. For low market power firms, the on-impact difference is nearly 4 percentage points, with the reversal peaking at about 2 percentage points. For high market power firms, the on-impact return is almost 5 percentage points lower, while the peak reversal is 2.7 percentage points smaller. These results indicate that differences in market power significantly affect both return levels and predictability. Nevertheless, variation in overreaction remains the more important driver of return dynamics.

Next we examine various real outcomes to understand the model's performance for real variables as well as how those feed into the return dynamics. Figure 14 shows the IRFs across all firm types for various real outcomes. In general we find that low market power firms respond the most and high market power firms respond the least (with the exception of wages). One stark difference is the sharp boom-bust of investment rising from 15% to 35% before crashing. This

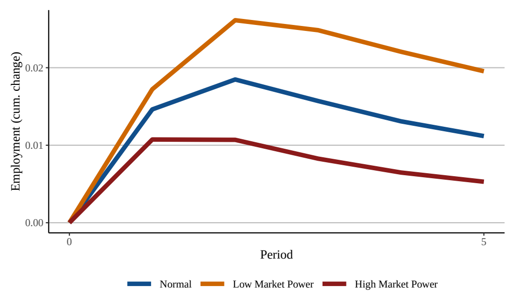
Figure 14: Impulse Response Function: All Firms under DE (1 SD Productivity Shock)



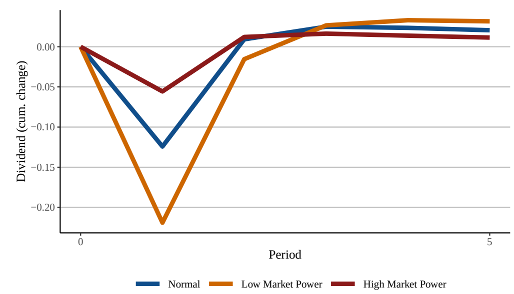
(a) Output



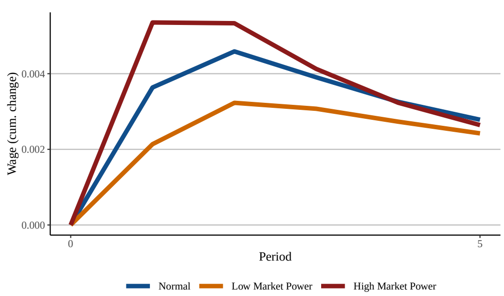
(b) Investment



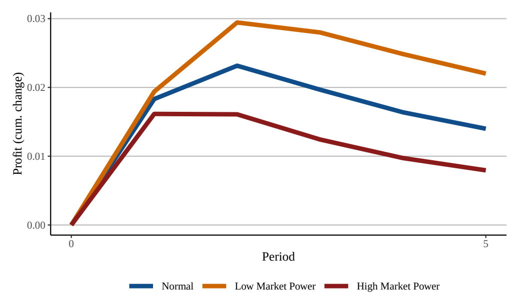
(c) Employment



(d) Dividend



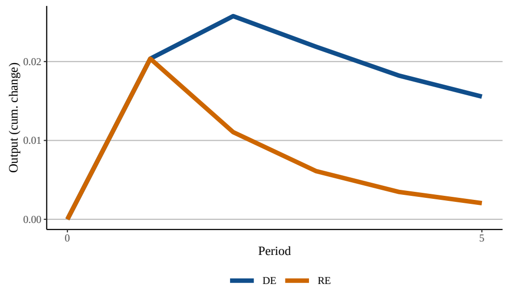
(e) Wage



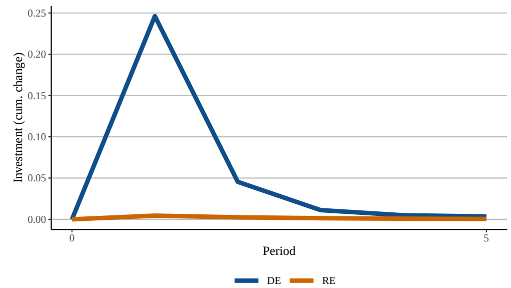
(f) Profit

Notes: This figure reports the impulse response functions for all firms under diagnostic expectations. The blue line represents normal firms, the orange line represents low market power firms, and the red line represents high market power firms. Panel (a) presents output, panel (b) presents investment, panel (c) presents employment, panel (d) presents dividend, panel (e) presents wage, and panel (f) presents profit. The shock is a one standard deviation positive productivity shock.

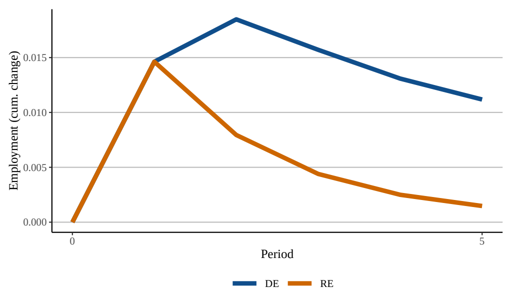
Figure 15: Impulse Response Function: Normal Firms (1 SD Productivity Shock)



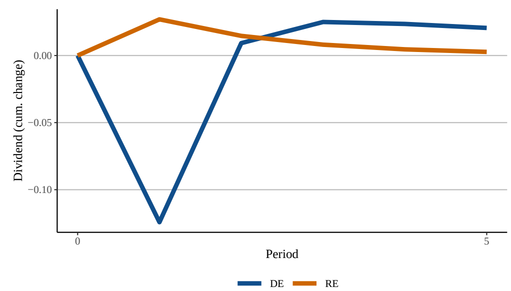
(a) Output



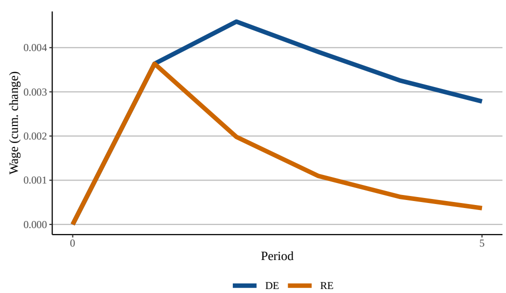
(b) Investment



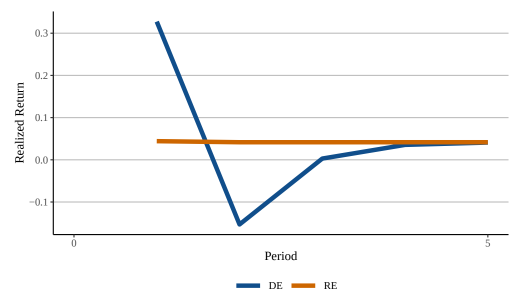
(c) Employment



(d) Dividend



(e) Wage



(f) Profit

Notes: This figure reports the impulse response functions for normal firms under both diagnostic expectations (blue lines) and rational expectations (orange lines). Panel (a) presents output, panel (b) presents investment, panel (c) presents employment, panel (d) presents dividend, panel (e) presents wage, and panel (f) presents profit. The shock is a one standard deviation positive productivity shock.

is driven by diagnostic expectations as firms become overly optimistic and invest too much and have inflated valuations. However, that quickly dissipates but the elevated capital level persists generating the more persistent effect on other real variables such as output or profits.

Figure 15 shows the IRFs for the same real outcomes as Figure 14 but only for normal firms and comparing the IRFs for the models with DE and RE (blue line and orange line, respectively).[23] The

[23]Figures A15 and A16 in Appendix A.1 show the IRFs comparing DE and RE for low market power firms and high market power firms, respectively. Figures A17 to A19 in Appendix A.1 show the IRFs comparing DE and RE for normal,

Table 16: Aggregate-Level Model vs. Empirical Moments

Statistic	Model DE (1)	Model RE (2)	Data (3)
I/K Ratio	0.100	0.100	0.092
Labor Share	0.500	0.500	0.668
Dividend Rate	0.307	0.307	0.240
Return (1 yr)	0.098	0.042	0.131
Return (5 yr)	0.318	0.130	0.656
Return SD (1 yr)	0.365	0.002	0.168
Return SD (5 yr)	0.483	0.006	0.377

Notes: This table presents the aggregated macroeconomic and financial moments for both the model with diagnostic expectations in Column (1) and the model with rational expectations in Column (2). Column (3) contains the empirical moments.

model with DE exhibits more pronounced responses and more hump-shaped dynamics relative to the model with RE, which matches the empirical IRFs found in macroeconomics better (e.g. Auclert, Rognlie and Straub, 2020, 2024). For example, for output the model with DE has a peak response of 2.5% that occurs in the second period after the shock whereas the model with RE has a peak response of 2.0% that occurs in the first period after the shock. Moreover the boom-bust pattern of investment that the model with DE demonstrates explains partly why the model with DE is able to generate the high excess returns and predictable reversals. The hump-shaped dynamics is also qualitatively consistent with the empirically estimated IRFs in macroeconomics.

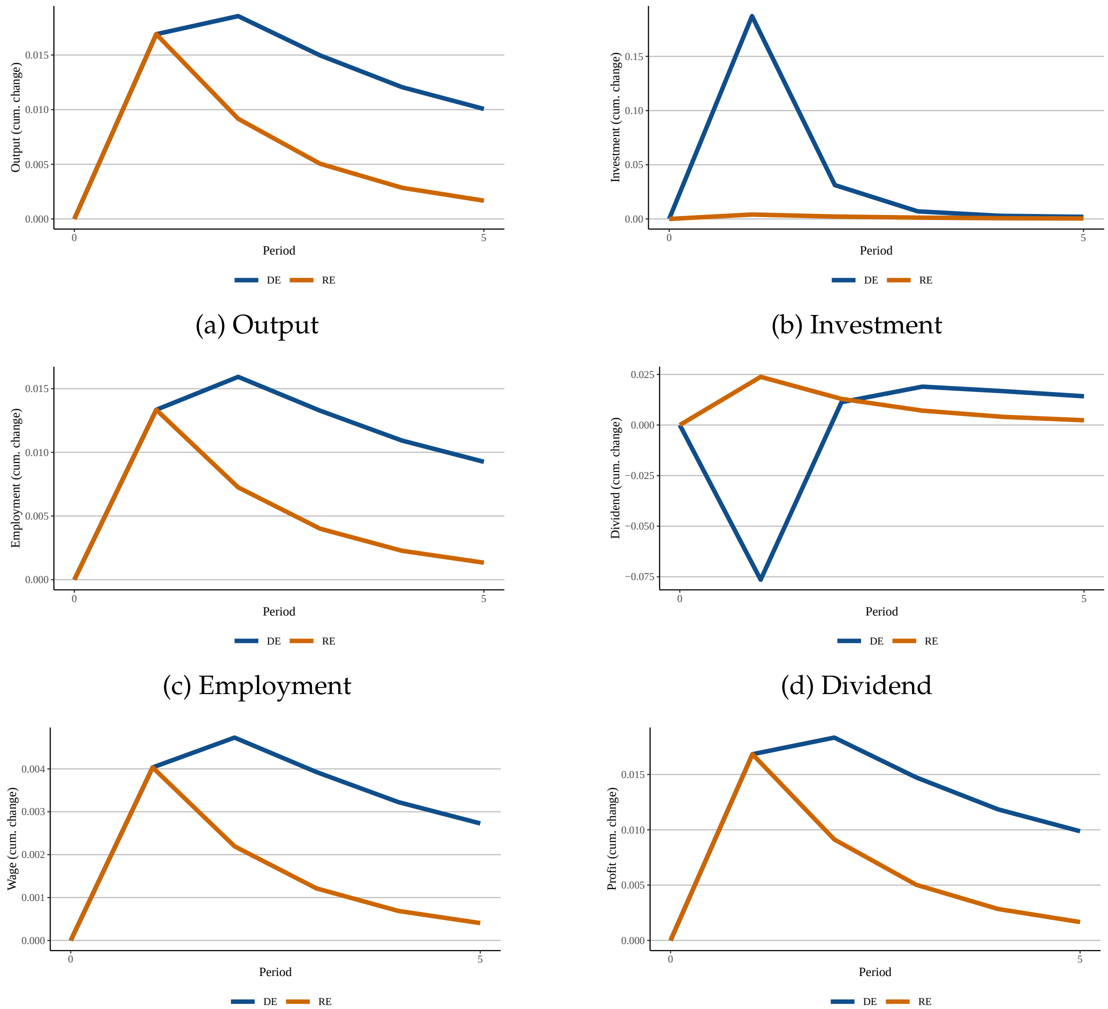
6.2 Aggregate Results

We now examine the aggregate response of the economy to a common productivity shock in partial equilibrium. Aggregated outcomes are constructed using the definitions in Equations (38) to (43), as outlined in Section 5.3. In addition to standard macroeconomic aggregates, we evaluate several measures of resource misallocation, including the dispersion of marginal products of capital (MPK), following the frameworks of Hsieh and Klenow (2009), Restuccia and Rogerson (2008), and David, Schmid and Zeke (2022). This allows us to assess how belief distortions affect not only the dynamics of macro aggregates, but also the efficiency of capital allocation across heterogeneous firms.

Table 16 shows various moments at the aggregate level comparing the data with model the model with diagnostic expectations and the model with rational expectations. In the aggregate,

low market power, and high market power firms, respectively, for a negative productivity shock.

Figure 16: Impulse Response Function: Aggregate (1 SD Productivity Shock)



(a) Output (b) Investment

(c) Employment (d) Dividend

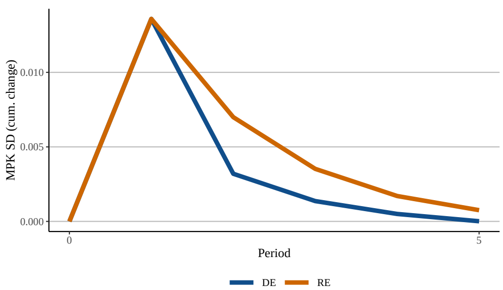
(e) Wage (f) Profit

Notes: This figure reports the aggregate impulse response functions under both diagnostic expectations (blue lines) and rational expectations (orange lines). Panel (a) presents output, panel (b) presents investment, panel (c) presents employment, panel (d) presents dividend, panel (e) presents wage, and panel (f) presents profit. The shock is a one standard deviation positive productivity shock.

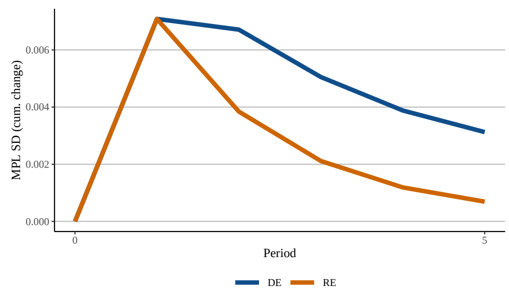
these results confirm that while various real moments on average are similar across both DE and RE, the model with DE can generate substantial excess returns and return volatility whereas the model with RE cannot.

Figures 16 and 17 present the aggregate impulse responses of key macroeconomic variables and misallocation measures, respectively, to a one standard deviation positive aggregate productivity shock. For completeness, Figures A24 and A25 in Appendix A.1 show the corresponding IRFs for a negative shock. As expected, the aggregate responses in Figure 16 broadly mirror the firm-level

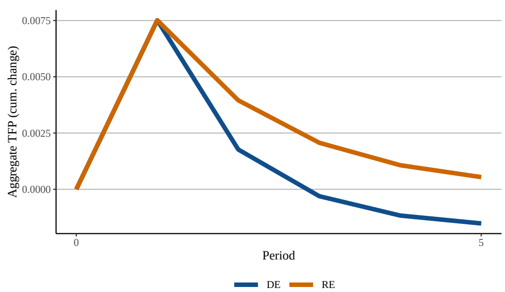
Figure 17: Impulse Response Function: Aggregate Misallocation (1 SD Productivity Shock)



(a) MPK SD



(b) MPL SD



(c) Aggregate Productivity

Notes: This figure reports the aggregate impulse response functions under both diagnostic expectations (blue lines) and rational expectations (orange lines). Panel (a) presents the standard deviation of MPK, panel (b) presents the standard deviation of MPL, and panel (c) presents aggregate productivity.

dynamics discussed earlier, with diagnostic expectations inducing greater amplification and more pronounced hump-shaped adjustments relative to rational expectations. Financial returns under diagnostic expectations, like the firm-level results, exhibit predictable reversals.

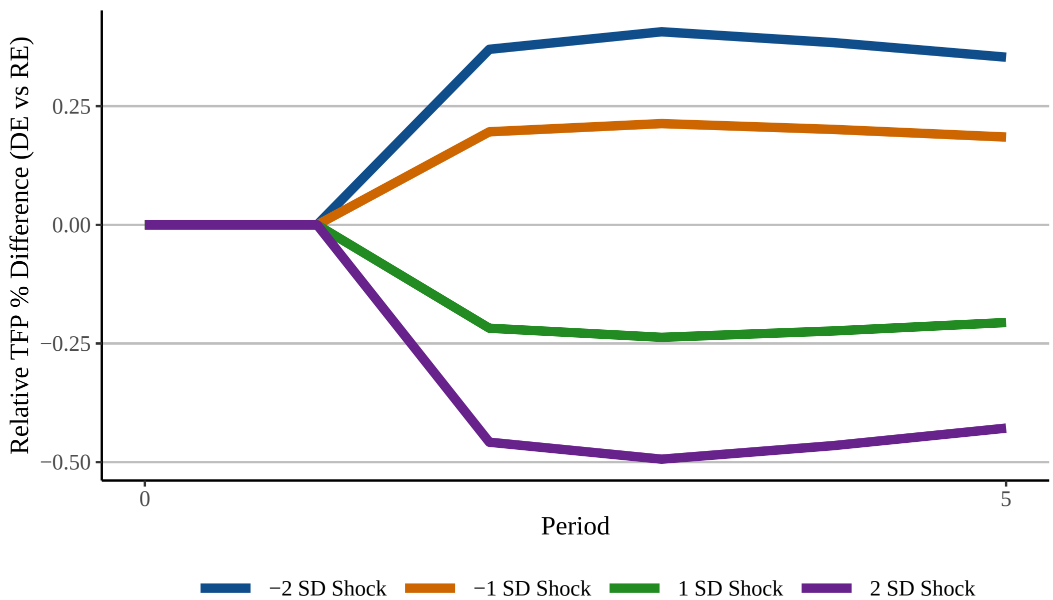
To assess misallocation, Figures 17a and 17b display the evolution of the cross-sectional standard deviation in firms' marginal products of capital (MPK) and labor (MPL), respectively. Figure 17c shows the evolution of aggregate productivity. On impact MPK and MPL dispersion as well as aggregate productivity respond the same since the differences in capital investment only impact static outcomes starting in the second period. Interestingly, MPK dispersion under diagnostic expectations declines more rapidly from its peak than under rational expectations. This occurs because the overreaction in investment driven by extrapolative beliefs pushes MPKs downward across all firms, mechanically compressing the cross-sectional dispersion despite underlying heterogeneity.[24] In this sense, diagnostic expectations may reduce apparent misallocation as measured by MPK dispersion not through improved efficiency but as a mechanical outcome of excessive capital accumulation. In contrast, MPL dispersion does not exhibit the same mechanical compression effect. Instead, dispersion remains elevated for a longer period under diagnostic expectations, reflecting persistent heterogeneity in labor demand responses across firms driven by extrapolative beliefs along with heterogenous market power.

Finally, we turn to aggregate productivity to more directly assess the overall efficiency implications of belief distortions and to compare the extent of misallocation across the two belief regimes. Following the initial impact, aggregate productivity in the model with diagnostic expectations falls below that in the rational expectations benchmark. Notably, under diagnostic expectations, aggregate productivity even declines below its pre-shock level. This decline is driven by the combination of excessive overinvestment induced by extrapolative beliefs and its interaction with heterogeneous market power. Low market power firms despite being the least productive have the lowest total market power and the strongest overreaction to shocks. As a result, a positive productivity shock induces them to expand disproportionately, contributing to a decline in aggregate productivity. In contrast, high markdown firms are the most productive but also have the highest total market power and exhibit the weakest response to the shock. Their relatively muted expansion limits the reallocation of resources toward the most efficient producers, exacerbating misallocation and further depressing aggregate productivity. Importantly, this asymmetric responsiveness is already present under rational expectations due to the dampening effect of market power on firm-level adjustment. The distribution of overreaction across firms amplify this pre-existing pattern by magnifying overreaction where it is least efficient and muting it where greater expansion would be most beneficial, worsening the aggregate efficiency loss.

Figure 18 compares the relative aggregate productivity outcomes under diagnostic and rational expectations across a range of common productivity shocks, varying in both magnitude and direction. The figure reports the percentage difference in aggregate TFP between the two belief

[24] See Figure A26 in Appendix A.1 which shows the firm-level IRFs for MPK and MPL.

Figure 18: Relative Difference in Aggregate TFP



Notes: This figure reports the relative differences in aggregate TFP between the model with diagnostic expectations and the model with rational expectations. The blue line reports the differences under a -2 SD productivity shock from the steady state. The orange line reports the differences under -1 SD productivity shock fromt the steady state. The green and purple lines report the same but for a 1 and 2 SD shock productivity, respectively.

regimes following -2, -1, $+1$, and $+2$ standard deviation shocks. Consistent with the IRF presented in Figure 17c, aggregate productivity under diagnostic expectations is persistently lower following a one standard deviation positive shock (green line), with a peak shortfall of approximately 0.24% that decays slowly over time. For a two standard deviation shock (purple line), the shortfall rises to around 0.49%. In contrast, diagnostic expectations generate productivity gains relative to rational expectations in response to negative shocks, with peak improvements of roughly 0.21% and 0.41% for -1 and -2 standard deviation shocks (blue and orange lines), respectively.

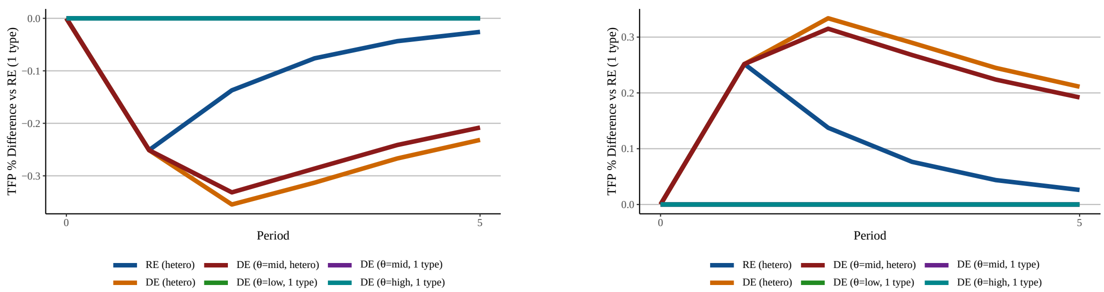
The modest asymmetry in productivity responses reflects both curvature in the future capital policy function and, more importantly, the interaction between belief distortions and heterogeneous market power. Firms with low total market power despite being less productive exhibit stronger overreaction to shocks, while highly productive firms with high market power adjust more modestly. This imbalance limits capital reallocation toward the most efficient producers during expansions, exacerbating misallocation and depressing aggregate productivity. This mechanism and result runs against the total procyclical capital reallocation results and channels documented by Eisfeldt and Rampini (2006) and David, Schmid and Zeke (2022). A more general framework could reconcile these findings with the procyclical reallocation patterns emphasized by Eisfeldt and Rampini (2006) and David, Schmid and Zeke (2022), allowing both forces to operate simultaneously. In such a setting, belief distortions and heterogeneous market power would dampen but not necessarily reverse the procyclicality of capital reallocation, making aggregate misallocation appear countercyclical only under certain conditions.

To further decompose these dynamics, we benchmark additional model variants against a rational expectations economy with homogeneous firms. This baseline matches the steady-state aggregates of the heterogeneous-firm models by appropriately calibrating the representative firm's productivity, price markup, and wage markdown.[25] We then explore four extensions: (i) heterogeneous firms under rational expectations, (ii) heterogeneous firms under diagnostic expectations, (iii) diagnostic expectations that are common across firms but with heterogeneous market power, and (iv) diagnostic expectations with homogeneous firms and varying levels of overreaction. This richer comparison allows us to disentangle the separate roles of diagnostic beliefs, heterogeneous market power, and heterogeneity in overreaction in shaping aggregate misallocation dynamics.

Figure 19 shows the relative difference in aggregate productivity across model variants, benchmarked against the rational expectations model with homogeneous firms. Figures 19a and 19b display the responses to a positive and negative productivity shock, respectively. The blue line corresponds to the rational expectations model with heterogeneous firms, and the orange line to the diagnostic expectations model with heterogeneous firms. In both cases, aggregate productivity falls relative to the baseline following a positive shock, reflecting increased misallocation as less

[25] Following our empirical definitions, we use the revenue-weighted aggregate price markup and wage-bill-weighted aggregate wage markdown.

Figure 19: Relative Difference in Aggregate TFP to RE and One Firm Type



(a) Positive Shock

(b) Negative Shock

Notes: This figure reports the relative differences in aggregate TFP between various models and the model with rational expectations and one firm type. The blue line represents the model with rational expectations with heterogenous firms. The orange line represents the model with diagnostic expectations and heterogenous firms. The red line represents the model with diagnostic expectations that are common but the remaining firm characteristics are varing. The green, purple, and cyan lines represent the model with one type of firm under various specifications of diagnostic expectations.

productive firms with lower market power expand more rapidly.

Under rational expectations with heterogeneous firms, the productivity shortfall peaks immediately at -0.25% before decaying quickly. In contrast, under diagnostic expectations, the peak is larger and more delayed. It reaches -0.33% in the second period, which is consistent with the amplification mechanism discussed earlier. For negative shocks, the rational expectations model exhibits a nearly symmetric improvement in aggregate productivity. The diagnostic expectations model also shows qualitative symmetry, though the magnitudes differ modestly.

The red line depicts a variant with diagnostic expectations and heterogeneous market power, but where all firms share the same intermediate level of overreaction. The resulting peak misallocation is slightly smaller at -0.32%, indicating that while diagnostic beliefs amplify misallocation through market power, additional amplification from heterogeneity in overreaction is relatively modest.

Finally, the green, purple, and cyan lines represent models with diagnostic expectations and homogeneous market power, across different fixed levels of overreaction. In these cases, aggregate productivity closely follows the rational expectations benchmark. Since all firms respond identically, belief distortions do not generate misallocation. This highlights that it is the *interaction* between distorted expectations and heterogeneous firm characteristics, especially market power, that drives the efficiency losses observed in this environment. It is also important to emphasize that the analysis thus far is conducted in *partial equilibrium*, which provides intuition for the misallocation channels but abstracts from general equilibrium adjustments.

In general equilibrium, the implications depend on how aggregate price levels adjust. With a homothetic aggregator, even in the presence of heterogeneous firms, there is no cyclical misallocation: all prices adjust proportionally, which cancels out at the aggregate level.[26] Even when firms overreact to shocks, if that overreaction is homogeneous across all firms, the aggregate price level adjusts proportionally, and the general equilibrium outcome exhibits no cyclical misallocation.

By contrast, when firms differ both in their degree of market power and in the extent of their belief-driven overreaction, aggregate prices no longer adjust in a way that fully offsets firm-level distortions. In this case, heterogeneous distorted expectations and heterogeneous market power jointly generate cyclical efficiency losses. Moreover, the correlation structure between productivity, market power, and belief distortions determines the sign of the effect: if low-productivity, low-market-power firms are those that overreact most, the result is procyclical misallocation, as we find in the data. If instead the opposite correlation holds, the misallocation channel becomes countercyclical. Hence, only when both heterogeneity in market power and heterogeneity in overreaction are present does cyclical misallocation survive in general equilibrium.

We now turn to welfare, where the results should be understood as quantifying the partial equilibrium channels emphasized above. While the general equilibrium logic implies that misallocation vanishes under homothetic aggregation unless both heterogeneous market power and heterogeneous belief distortions are present, the PE analysis provides a clean benchmark to assess the efficiency costs and asymmetries of these mechanisms. We assume that households are endowed with CRRA preferences, given by

$$U_0 = \sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\gamma}}{1-\gamma}, \tag{44}$$

where the coefficient of relative risk aversion $\gamma \in \{1, 2, 4\}$, and aggregate consumption C_t is defined using the accounting identity $C_t = Y_t - I_t$.

To compute the consumption-equivalent adjustment, let $\tilde{U}_{0,T}$ denote the finite-horizon utility from the benchmark consumption path $\{\tilde{C}_t\}_{t=0}^{T}$, generated by the rational expectations model with homogeneous firms given the shock. Let $U_{0,T}$ denote the corresponding utility from an alternative model with consumption path $\{C_t\}_{t=0}^{T}$, for a fixed horizon T. We set $T = 150$. The consumption-equivalent loss or gain λ is defined implicitly by

$$\sum_{t=0}^{T} \beta^t \frac{(C_t(1+\lambda))^{1-\gamma}}{1-\gamma} = \tilde{U}_{0,T}. \tag{45}$$

For all $\gamma > 0$, Equation (45) admits a unique solution for λ. When $\lambda > 0$, households would require a permanent percentage increase in consumption to be indifferent between the alternative and the benchmark model, implying that the benchmark (rational expectations with homogeneous firms) yields higher utility. Conversely, $\lambda < 0$ indicates a welfare gain under the alternative model relative

[26] This applies not only to the canonical CES aggregator but also to more general homothetic forms such as those developed by Matsuyama and Ushchev (2017); Matsuyama (2023, 2025). See also Berger, Herkenhoff and Mongey (2022) for an example of how aggregate shocks are simplified in a general equilibrium setting with homothetic aggregators.

Table 17: Consumption-Equivalent Welfare Gains (Shocks $\pm 1\sigma$)

Regime	$\gamma = 1$, -1σ	$\gamma = 1$, +1σ	$\gamma = 2$, -1σ	$\gamma = 2$, +1σ	$\gamma = 4$, -1σ	$\gamma = 4$, +1σ
RE (hetero)	-0.022	0.022	-0.022	0.022	-0.023	0.021
DE (hetero)	-0.094	0.102	-0.094	0.101	-0.095	0.101
DE (θ=mid, hetero)	-0.087	0.093	-0.087	0.093	-0.087	0.092
DE (θ=low, 1 type)	0.000	0.000	0.000	0.000	-0.000	-0.000
DE (θ=mid, 1 type)	0.000	0.000	-0.000	0.000	-0.000	0.000
DE (θ=high, 1 type)	0.000	0.000	-0.000	0.000	-0.000	-0.000

Notes: This table reports consumption-equivalent welfare gains in percent for $\pm 1\sigma$ productivity shocks relative to the case of rational expectations and homogenous firms. The columns correspond to the specified γ and shock value.

to the benchmark.

Table 17 reports the consumption-equivalent welfare gains or losses expressed in percentage terms relative to the rational expectations model with homogeneous firms, across the various model specifications, shock directions, and values of γ. Welfare losses are generally stable across risk aversion levels. In the rational expectations model with heterogeneous firms, eliminating firm heterogeneity yields a welfare gain equivalent to at least a 0.02% increase in lifetime consumption following a positive shock, and at least a 0.02% loss following a negative shock.[27] These values are broadly comparable to the canonical estimates in Lucas (1987) for the welfare cost of aggregate business cycle fluctuations.

In contrast, the model with diagnostic expectations and heterogeneous firms generates substantially larger welfare losses, at least 0.09% in expansions and -0.10% in recessions, due to the amplified misallocation effects driven by belief distortions. The incremental welfare cost from heterogeneity in overreaction (relative to the case with uniform overreaction) is approximately 0.007 percentage points. While this amplification is significant compared to the benchmark rational expectations model, it remains modest in absolute terms. As expected, there is no welfare difference when diagnostic expectations are applied in a homogeneous firm setting, since misallocation does not arise.

These results indicate that diagnostic expectations, when interacting with heterogeneous firms, particularly those with differing degrees of market power, can generate substantially larger welfare losses or gains over the business cycle due to amplified misallocation, relative to the benchmark of rational expectations with homogeneous firms. While this analysis abstracts from explicit policy design, the findings suggest that accounting for belief distortions and firm heterogeneity is important for understanding the efficiency costs of business cycle fluctuations. In particular, the asymmetric and persistent nature of the misallocation effects highlights the potential value of future work examining how informational frictions and behavioral biases shape aggregate outcomes in

[27] Table A19 shows the consumption-equivalent welfare changes for the case of shocks $\pm 2\sigma$.

environments with market power and capital reallocation dynamics.

7 Conclusion

We study how heterogeneity in firm-level market power relates to heterogeneity in belief distortions and overreaction. This link is motivated by recent evidence that subjective cash-flow expectations explain much of the variation in asset returns, and that market power is a central driver of firm cash flows and profits as well as prior evidence of now a firm's competitive environment is related to its future returns and investment levels. We document empirically that firms with low total market power exhibit the strongest overreaction, while firms with high total market power exhibit the weakest. Excess optimism and pessimism are short-lived on average but more persistent among low total market power firms, and firm-level overreaction is substantially larger than at the aggregate level.

Motivated by these findings, we embed the heterogeneous expectations and their relationship to firm characteristics in a simple macro-finance model in partial equilibrium. The model shows that subjective cash-flow expectations alone without variation in discount rates can generate elevated and more volatile returns within an endogenous production setting. It also produces hump-shaped real dynamics and predictable return reversals following excessive optimism/pessimism, consistent with empirical patterns. A decomposition of returns confirms that heterogeneity in belief distortions drives both the excess and predictable responses, as well as cross-sectional variation in performance. While heterogeneity in market power matters, differences in overreaction are the primary driver of cross-sectional variation in financial returns.

The model further implies that the interaction between diagnostic expectations and heterogeneous market power amplifies procyclical misallocation. Under rational expectations, heterogeneity in market power alone generates misallocation in the steady state. Following an aggregate shock in partial equilibrium, the model also implies procyclical misallocation, magnified by diagnostic expectations through overreaction. The additional amplification due to their negative correlation with market power, though modest, reinforces the effect. In general equilibrium, however, cyclical misallocation survives only when both heterogeneity in market power and heterogeneity in overreaction are present. With a homothetic aggregator, heterogeneous market power is insufficient to generate cyclical misallocation. Even with overreaction, when it is homogeneous, all firms respond with proportional price adjustments that offset distortions exactly in general equilibrium. With heterogeneous overreaction, however, price adjustments are no longer proportional, and the result persists. Moreover, the correlation structure among productivity, market power, and overreaction determines the sign of the effect: if low productivity, low market power firms overreact more, misallocation is procyclical; if the reverse holds, it becomes countercyclical.

While highly stylized, the model replicates several key asset pricing and macroeconomic moments that a comparable rational expectations framework fails to match. At the same time, it generates counterfactual predictions, such as an abrupt dividend decline following a positive productivity shock and Sharpe ratios that remain too low. These limitations suggest that extensions incorporating financial frictions, other adjustment costs, or richer belief dynamics may be necessary to better align the model with the data.

Overall, the results show that belief distortions rooted in diagnostic expectations can amplify firm- and aggregate-level dynamics through their interaction with heterogeneous market power. This mechanism links micro-level belief heterogeneity to macro-level misallocation and welfare costs, offering a unified behavioral channel through which expectations shape both asset prices and real outcomes. Future work embedding these mechanisms in a full general equilibrium framework could further clarify how belief distortions and market power jointly shape business-cycle dynamics and the behavior of asset prices.

References

Abel, Andrew B., and Janice C. Eberly. 1996. "Optimal Investment with Costly Reversibility." *Review of Economic Studies*, 63(4): 581–593.

Abel, Andrew B., and Janice C. Eberly. 1999. "The Effects of Irreversibility and Uncertainty on Capital Accumulation." *Journal of Monetary Economics*, 44(3): 339–377.

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. "Identification Properties of Recent Production Function Estimators." *Econometrica*, 83(6): 2411–2451.

Akcigit, Ufuk, and Sina T. Ates. 2023. "What Happened to US Business Dynamism?" *Journal of Political Economy*, 131(8): 2059–2124.

Angeletos, George-Marios, Zhen Huo, and Karthik A. Sastry. 2021. "Imperfect Macroeconomic Expectations: Evidence and Theory." *NBER Macroeconomics Annual*, 35: 1–86.

Atkeson, Andrew, and Ariel Burstein. 2008. "Pricing-to-Market, Trade Costs, and International Relative Prices." *American Economic Review*, 98(5): 1998–2031.

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub. 2020. "Micro Jumps, Macro Humps: Monetary Policy and Business Cycles in an Estimated HANK Model." NBER Working Paper 26647.

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub. 2024. "The Intertemporal Keynesian Cross." *Journal of Political Economy*, 132(12): 4068–4121.

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen. 2020. "The Fall of the Labor Share and the Rise of Superstar Firms." *Quarterly Journal of Economics*, 135(2): 645–709.

Berger, David, Kyle Herkenhoff, and Simon Mongey. 2022. "Labor Market Power." *American Economic Review*, 112(4): 1147–1193.

Bianchi, Francesco, Cosmin Ilut, and Hikaru Saijo. 2024. "Diagnostic Business Cycles." *Review of Economic Studies*, 91(1): 129–162.

Bloom, Nicholas. 2009. "The Impact of Uncertainty Shocks." *Econometrica*, 77(3): 623–685.

Board of Governors of the Federal Reserve System. 2024. "H.15 Selected Interest Rates (Daily)."

Boldrin, Michele, Lawrence J. Christiano, and Jonas D. M. Fisher. 2001. "Habit Persistence, Asset Returns, and the Business Cycle." *American Economic Review*, 91(1): 149–166.

Bond, Steve, Arshia Hashemi, Greg Kaplan, and Piotr Zoch. 2021. "Some Unpleasant Markup Arithmetic: Production Function Elasticities and Their Estimation From Production Data." *Journal of Monetary Economics*, 121: 1–14.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2018. "Diagnostic Expectations and Credit Cycles." *Journal of Finance*, 73(1): 199–227.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer. 2019. "Diagnostic Expectations and Stock Returns." *Journal of Finance*, 74(6): 2839–2874.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer. 2024*a*. "Belief Overreaction and Stock Market Puzzles." *Journal of Political Economy*, 132(5): 1450–1484.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, Matthew OBrien, and Andrei Shleifer. 2024*b*. "Long Term Expectations and Aggregate Fluctuations." *NBER Macroeconomics Annual*, 38: 311–347.

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer. 2020. "Overreaction in Macroeconomic Expectations." *American Economic Review*, 110(9): 2748–2782.

Broer, Tobias, and Alexandre N. Kohlhas. 2024. "Forecaster (Mis-)Behavior." *Review of Economics and Statistics*, 106(5): 1334–1351.

Buffett, Warren. 2010. "Financial Crisis Inquiry Commission - Interview of Warren Buffett."

Bureau of Economic Analysis. 2024. "National Income and Product Accounts."

Bureau of Labor Statistics. 2024. "Consumer Price Index."

Bustamante, M. Cecilia, and Andres Donangelo. 2017. "Product Market Competition and Industry Returns." *Review of Financial Studies*, 30(12): 4216–4266.

Cameron, A. Colin, and Douglas L. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources*, 50(2): 317–372.

Campbell, John Y., and John Ammer. 1993. "What Moves the Stock and Bond Markets? A Variance Decomposition for Long-Term Asset Returns." *Journal of Finance*, 48(1): 3–37.

Campbell, John Y., and Robert J. Shiller. 1987. "Cointegration and Tests of Present Value Models." *Journal of Political Economy*, 95(5): 1062–1088.

Campbell, John Y., and Robert J. Shiller. 1988. "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors." *Review of Financial Studies*, 1(3): 195–228.

Campbell, John Y., Martin Lettau, Burton G. Malkiel, and Yexiao Xu. 2001. "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk." *Journal of Finance*, 56(1): 1–43.

Center for Research in Security Prices. 2025. "I/B/E/S Guidance."

Chan, Louis K. C., Jason Karceski, and Josef Lakonishok. 2003. "The Level and Persistence of Growth Rates." *Journal of Finance*, 58(2): 643–684.

Chow, Melissa, Teresa C. Fort, Christopher Goetz, Nathan Goldschlag, James Lawrence, Elisabeth Ruth Perlman, Martha Stinson, and T. Kirk White. 2021. "Redesigning the Longitudinal Business Database." U.S. Census Bureau CES 21-08.

Cochrane, John H. 2011. "Presidential Address: Discount Rates." *Journal of Finance*, 66(4): 1047–1108.

Coibion, Olivier, and Yuriy Gorodnichenko. 2015. "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts." *American Economic Review*, 105(8): 2644–2678.

Cooper, Ilan. 2006. "Asset Pricing Implications of Nonconvex Adjustment Costs and Irreversibility of Investment." *Journal of Finance*, 61(1): 139–170.

Cooper, Russell W., and John C. Haltiwanger. 2006. "On the Nature of Capital Adjustment Costs." *Review of Economic Studies*, 73(3): 611–633.

Corhay, Alexandre, Howard Kung, and Lukas Schmid. 2020. "Competition, Markups, and Predictable Returns." *Review of Financial Studies*, 33(12): 5906–5939.

Corhay, Alexandre, Howard Kung, and Lukas Schmid. 2025. "Q: Risk, Rents, or Growth?" *Journal of Financial Economics*, 165: 103990.

Corhay, Alexandre, Jun E. Li, and Jincheng Tong. 2022. "Markup Shocks and Asset Prices." SSRN Working Paper.

Crouzet, Nicolas, and Janice Eberly. 2021. "Intangibles, Markups, and the Measurement of Productivity Growth." *Journal of Monetary Economics*, 124: S92–S109.

Crouzet, Nicolas, and Janice Eberly. 2023. "Rents and Intangible Capital: A Q+ Framework." *Journal of Finance*, 78(4): 1873–1916.

David, Joel M., Lukas Schmid, and David Zeke. 2022. "Risk-Adjusted Capital Allocation and Misallocation." *Journal of Financial Economics*, 145(3): 684–705.

De La O, Ricardo, and Sean Myers. 2021. "Subjective Cash Flow and Discount Rate Expectations." *Journal of Finance*, 76(3): 1339–1387.

De la O, Ricardo, and Sean Myers. 2024. "Which Subjective Expectations Explain Asset Prices?" *Review of Financial Studies*, 37(6): 1929–1978.

Delao, Ricardo, Xiao Han, and Sean Myers. 2025. "The Return of Return Dominance: Decomposing the Cross-Section of Prices." *Journal of Financial Economics*, 169: 104059.

De Loecker, Jan, and Frederic Warzynski. 2012. "Markups and Firm-Level Export Status." *American Economic Review*, 102(6): 2437–2471.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger. 2020. "The Rise of Market Power and the Macroeconomic Implications." *Quarterly Journal of Economics*, 135(2): 561–644.

De Ridder, Maarten. 2024. "Market Power and Innovation in the Intangible Economy." *American Economic Review*, 114(1): 199–251.

De Ridder, Maarten, Basile Grassi, and Giovanni Morzenti. 2025. "The Hitchhiker's Guide to Markup Estimation." Working Paper.

Dobbelaere, Sabien, and Jacques Mairesse. 2013. "Panel Data Estimates of the Production Function and Product and Labor Market Imperfections." *Journal of Applied Econometrics*, 28(1): 1–46.

Donangelo, Andres, François Gourio, Matthias Kehrig, and Miguel Palacios. 2019. "The Cross-Section of Labor Leverage and Equity Returns." *Journal of Financial Economics*, 132(2): 497–518.

Eisfeldt, Andrea L., and Adriano A. Rampini. 2006. "Capital Reallocation and Liquidity." *Journal of Monetary Economics*, 53(3): 369–399.

Eisfeldt, Andrea L., and Dimitris Papanikolaou. 2013. "Organization Capital and the Cross-Section of Expected Returns." *Journal of Finance*, 68(4): 1365–1406.

Farhi, Emmanuel, and François Gourio. 2018. "Accounting for Macro-Finance Trends: Market Power, Intangibles, and Risk Premia." *Brookings Papers on Economic Activity*.

Gennaioli, Nicola, and Andrei Shleifer. 2010. "What Comes to Mind." *Quarterly Journal of Economics*, 125(4): 1399–1433.

Gennaioli, Nicola, Yueran Ma, and Andrei Shleifer. 2016. "Expectations and Investment." *NBER Macroeconomics Annual*, 30: 379–431.

Grennan, Jillian, and Roni Michaely. 2020. "Artificial Intelligence and High-Skilled Work: Evidence from Analysts." Social Science Research Network SSRN Working Paper 3681574.

Gu, Lifeng. 2016. "Product Market Competition, R&D Investment, and Stock Returns." *Journal of Financial Economics*, 119(2): 441–455.

Gutiérrez, Germán, and Thomas Philippon. 2017. "Declining Competition and Investment in the U.S." NBER Working Paper 23583.

Hall, Robert E. 1988. "The Relation between Price and Marginal Cost in U.S. Industry." *Journal of Political Economy*, 96(5): 921–947.

Hansen, Lars Peter. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica*, 50(4): 1029–1054.

Hansen, Lars Peter, and Kenneth J. Singleton. 1982. "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models." *Econometrica*, 50(5): 1269–1286.

Hansen, Lars Peter, and Ravi Jagannathan. 1991. "Implications of Security Market Data for Models of Dynamic Economies." *Journal of Political Economy*, 99(2): 225–262.

Hayashi, Fumio. 1982. "Tobin's Marginal q and Average q: A Neoclassical Interpretation." *Econometrica*, 50(1): 213–224.

Helpman, Elhanan, and Benjamin Niswonger. 2022. "Dynamics of Markups, Concentration, and Product Span." *American Economic Journal: Macroeconomics*, 14(3): 42–81.

Hoberg, Gerard, and Gordon Phillips. 2010. "Real and Financial Industry Booms and Busts." *Journal of Finance*, 65(1): 45–86.

Horowitz, Joel L. 2001. "Chapter 52 - The Bootstrap." In *Handbook of Econometrics*. Vol. 5, , ed. James J. Heckman and Edward Leamer, 3159–3228. Elsevier.

Hou, Kewei, and David T. Robinson. 2006. "Industry Concentration and Average Stock Returns." *Journal of Finance*, 61(4): 1927–1956.

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics*, 124(4): 1403–1448.

Ilut, Cosmin, and Rosen Valchev. 2023. "Economic Agents as Imperfect Problem Solvers." *Quarterly Journal of Economics*, 138(1): 313–362.

Iraola, Miguel A., and Manuel S. Santos. 2017. "Asset Price Volatility, Price Markups, and Macroeconomic Fluctuations." *Journal of Monetary Economics*, 90: 84–98.

Jermann, Urban J. 1998. "Asset Pricing in Production Economies." *Journal of Monetary Economics*, 41(2): 257–275.

King, Robert G., Charles I. Plosser, and Sergio T. Rebelo. 1988. "Production, Growth and Business Cycles: I. The Basic Neoclassical Model." *Journal of Monetary Economics*, 21(2): 195–232.

Kogan, Leonid, Dimitris Papanikolaou, and Noah Stoffman. 2020. "Left Behind: Creative Destruction, Inequality, and the Stock Market." *Journal of Political Economy*, 128(3): 855–906.

Kohlhas, Alexandre N., and Ansgar Walther. 2021. "Asymmetric Attention." *American Economic Review*, 111(9): 2879–2925.

Kothari, S. P., Eric So, and Rodrigo Verdi. 2016. "Analysts' Forecasts and Asset Pricing: A Survey." *Annual Review of Financial Economics*, 8: 197–219.

Kydland, Finn E., and Edward C. Prescott. 1982. "Time to Build and Aggregate Fluctuations." *Econometrica*, 50(6): 1345–1370.

Lehr, Nils H. 2023. "Does Monopsony Matter for Innovation?" Working Paper.

Levinsohn, James, and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *Review of Economic Studies*, 70(2): 317–341.

Lucas, Robert E. 1987. *Models of Business Cycles.* B. Blackwell.

Matsuyama, Kiminori. 2023. "Non-CES Aggregators: A Guided Tour." *Annual Review of Economics*, 15: 235–265.

Matsuyama, Kiminori. 2025. "Homothetic Non-CES Demand Systems with Applications to Monopolistic Competition." *Annual Review of Economics*, 17: 261–292.

Matsuyama, Kiminori, and Philip Ushchev. 2017. "Beyond CES: Three Alternative Classes of Flexible Homothetic Demand Systems." Working Paper.

Maxted, Peter. 2024. "A Macro-Finance Model with Sentiment." *Review of Economic Studies*, 91(1): 438–475.

Mehra, Rajnish, and Edward C. Prescott. 1985. "The Equity Premium: A Puzzle." *Journal of Monetary Economics*, 15(2): 145–161.

Modigliani, Franco, and Merton H. Miller. 1958. "The Cost of Capital, Corporation Finance and the Theory of Investment." *American Economic Review*, 48(3): 261–297.

Newey, Whitney K., and Daniel McFadden. 1994. "Chapter 36 Large Sample Estimation and Hypothesis Testing." In *Handbook of Econometrics*. Vol. 4, , ed. Robert F. Engle and Daniel McFadden, 2111–2245. Elsevier.

Olley, G. Steven, and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica*, 64(6): 1263–1297.

Papanikolaou, Dimitris. 2011. "Investment Shocks and Asset Prices." *Journal of Political Economy*, 119(4): 639–685.

Peters, Michael. 2020. "Heterogeneous Markups, Growth, and Endogenous Misallocation." *Econometrica*, 88(5): 2037–2073.

Peters, Ryan H., and Lucian A. Taylor. 2017. "Intangible Capital and the Investment-q Relation." *Journal of Financial Economics*, 123(2): 251–272.

Ren, Kevin, and Dalton Rongxuan Zhang. 2025. "Price Markups or Wage Markdowns?" Working Paper.

Restuccia, Diego, and Richard Rogerson. 2008. "Policy Distortions and Aggregate Productivity With Heterogeneous Establishments." *Review of Economic Dynamics*, 11(4): 707–720.

Rhodes-Kropf, Matthew, and S. Viswanathan. 2004. "Market Valuation and Merger Waves." *Journal of Finance*, 59(6): 2685–2718.

Rhodes–Kropf, Matthew, David T. Robinson, and S. Viswanathan. 2005. "Valuation Waves and Merger Activity: The Empirical Evidence." *Journal of Financial Economics*, 77(3): 561–603.

Rouwenhorst, K. G. 1995. "Asset Pricing Implications of Equilibrium Business Cycle Models." In *Frontiers of Business Cycle Research.* , ed. Thomas F. Cooley, 294–330. Princeton University Press.

Seegmiller, Bryan. 2023. "Valuing Labor Market Power: The Role of Productivity Advantages." Working Paper.

Shiller, Robert J. 1981. "Do Stock Prices Move Too Much to Be Justified by Subsequent Changes in Dividends?" *American Economic Review*, 71(3): 421–436.

Shleifer, Andrei, and Robert W. Vishny. 2003. "Stock Market Driven Acquisitions." *Journal of Financial Economics*, 70(3): 295–311.

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature*, 49(2): 326–365.

Traina, James. 2018. "Is Aggregate Market Power Increasing? Production Trends Using Financial Statements." SSRN Working Paper.

United States Census Bureau. 2022. "Longitudinal Business Database."

van Binsbergen, Jules H, Xiao Han, and Alejandro Lopez-Lira. 2023. "Man versus Machine Learning: The Term Structure of Earnings Expectations and Conditional Biases." *Review of Financial Studies*, 36(6): 2361–2396.

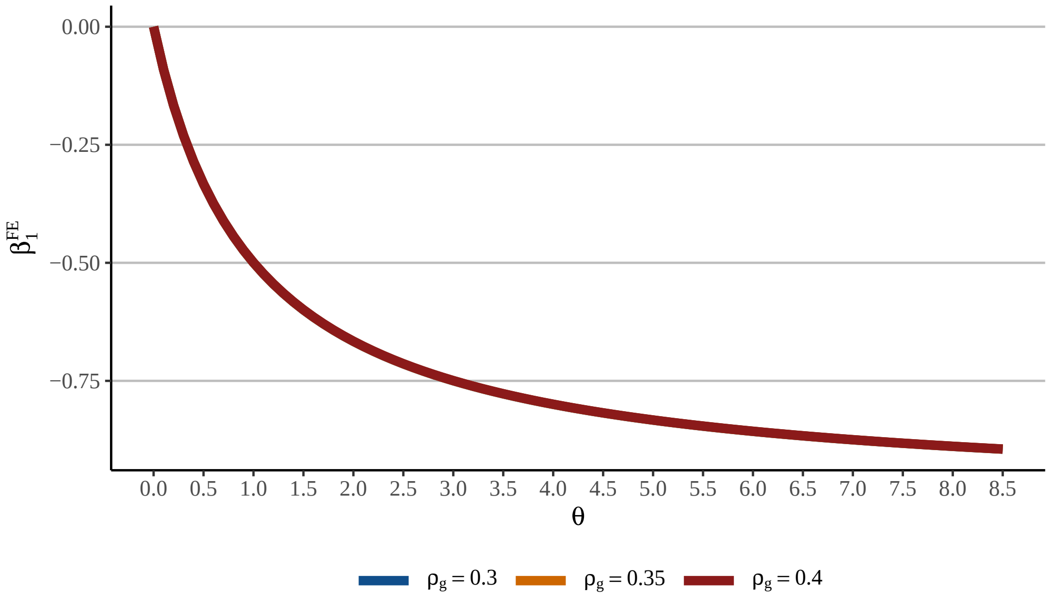
Vuolteenaho, Tuomo. 2002. "What Drives Firm-Level Stock Returns?" *Journal of Finance*, 57(1): 233–264.

Yeh, Chen, Claudia Macaluso, and Brad Hershbein. 2022. "Monopsony in the US Labor Market." *American Economic Review*, 112(7): 2099–2138.

A Additional Figures and Tables

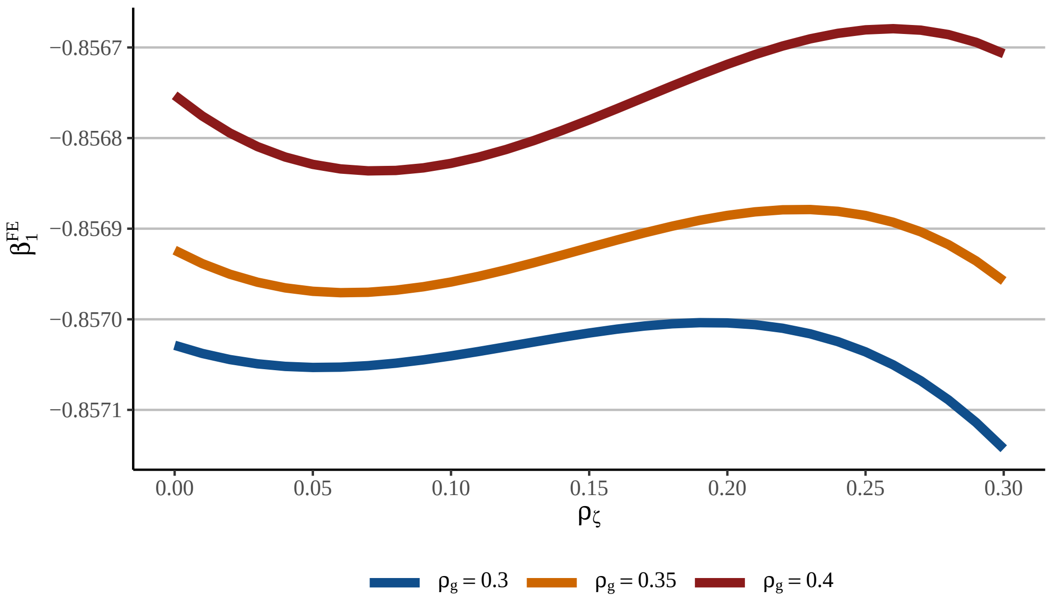
A.1 Figures

Figure A1: Effect of Overreaction θ on β_1^{FE}, Across Growth Persistence ρ_g



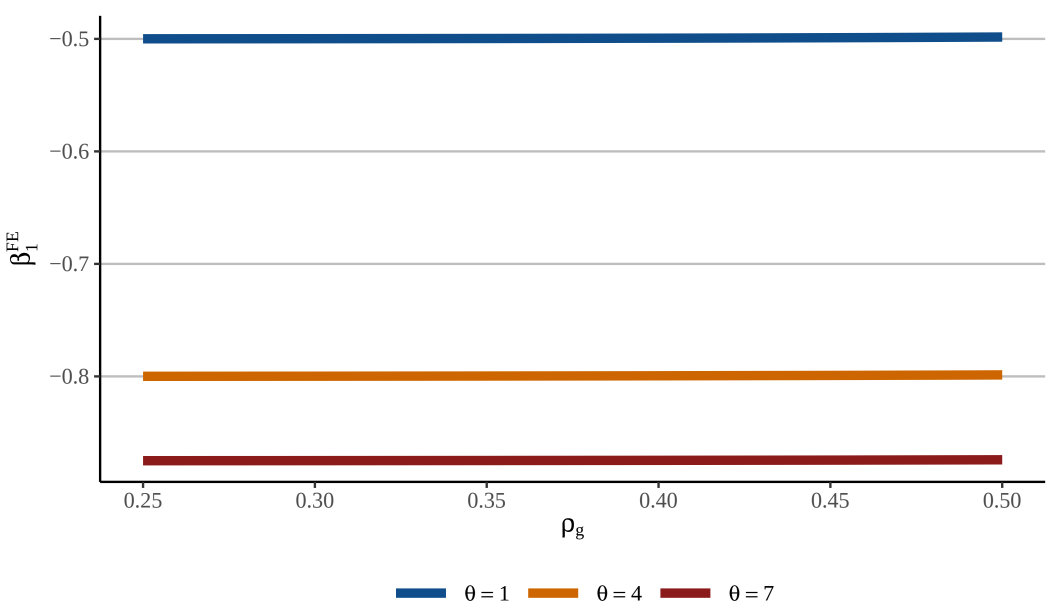
Notes: This figure reports the comparative statics of β_1^{FE} with respect to θ, holding $\rho_\zeta = 0.2$, $\sigma_u = 0.02$, and $s = 5$ fixed. Each line corresponds to a different fixed value of ρ_g: the blue line uses $\rho_g = 0.30$, the orange line uses $\rho_g = 0.35$, and the red line uses $\rho_g = 0.40$.

Figure A2: Effect of Belief Persistence ρ_ζ on β_1^{FE}, Across Growth Persistence ρ_g



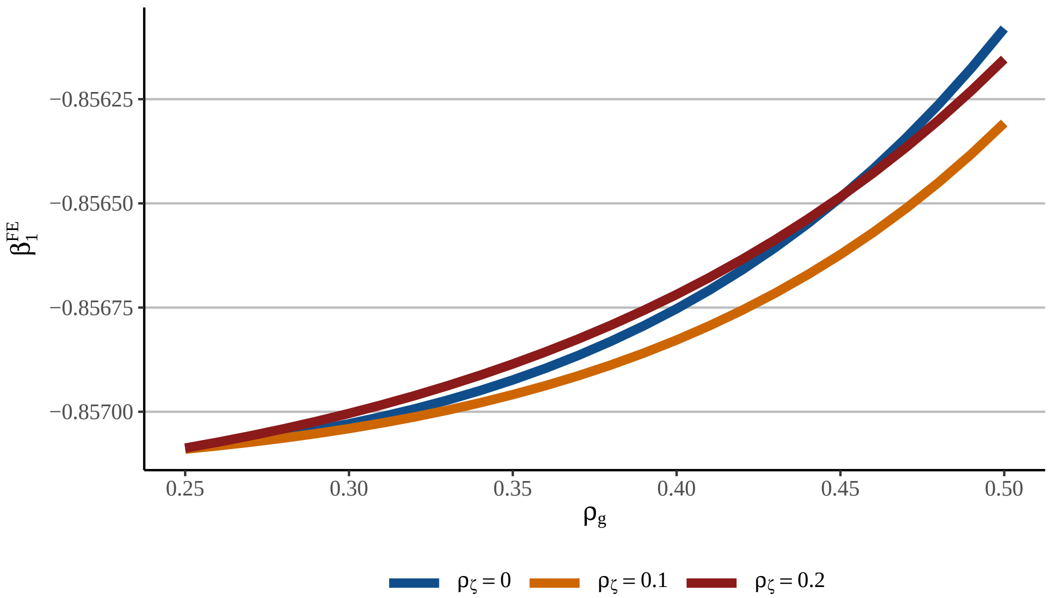
Notes: This figure reports the comparative statics of β_1^{FE} with respect to ρ_ζ, holding $\theta = 6$, $\sigma_u = 0.02$, and $s = 5$ fixed. Each line corresponds to a different fixed value of ρ_g: the blue line uses $\rho_g = 0.30$, the orange line uses $\rho_g = 0.35$, and the red line uses $\rho_g = 0.40$.

Figure A3: Effect of Growth Persistence ρ_g on β_1^{FE}, Across Overreaction θ



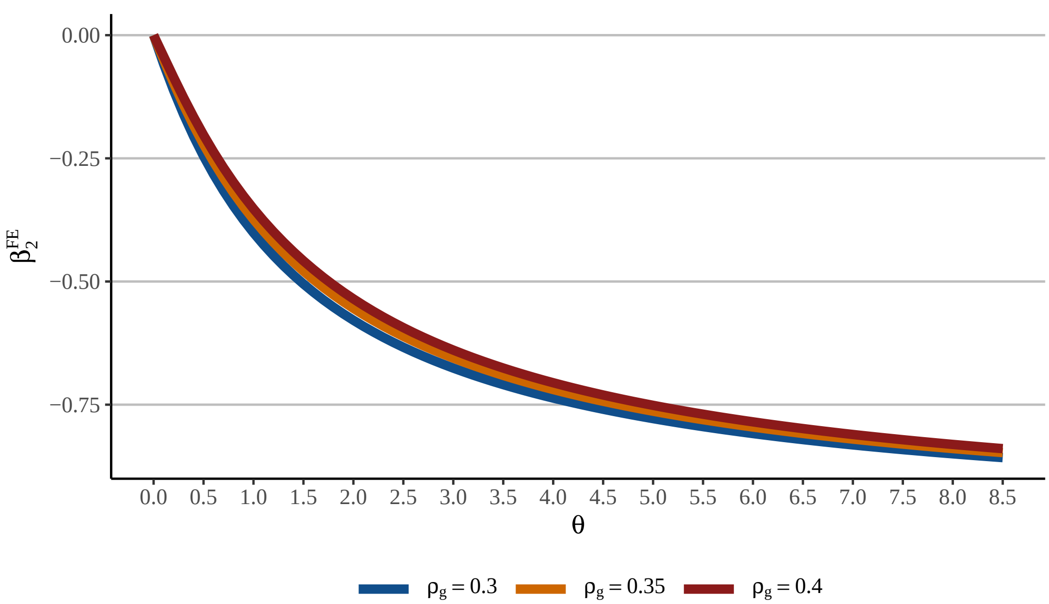
Notes: This figure reports the comparative statics of β_1^{FE} with respect to ρ_g, holding $\rho_\zeta = 0.2$, $\sigma_u = 0.02$, and $s = 5$ fixed. Each line corresponds to a different fixed value of θ: the blue line uses $\theta = 1$, the orange line uses $\theta = 4$, and the red line uses $\theta = 7$.

Figure A4: Effect of Growth Persistence ρ_g on β_1^{FE}, Across Belief Persistence ρ_ζ



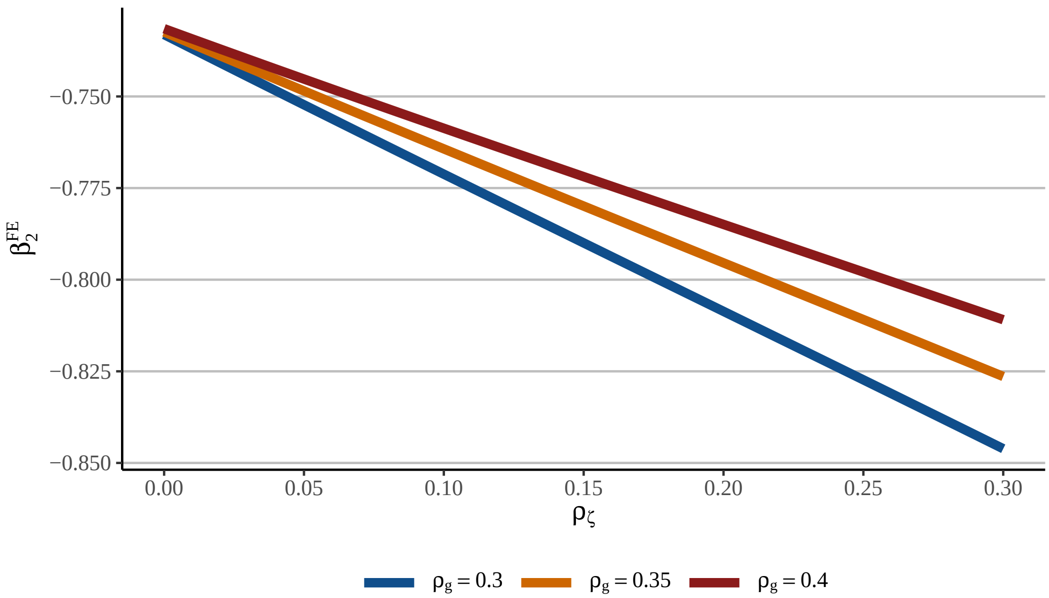
Notes: This figure reports the comparative statics of β_1^{FE} with respect to ρ_g, holding $\theta = 6$, $\sigma_u = 0.02$, and $s = 5$ fixed. Each line corresponds to a different fixed value of ρ_ζ: the blue line uses $\rho_\zeta = 0$, the orange line uses $\rho_\zeta = 0.1$, and the red line uses $\rho_\zeta = 0.2$.

Figure A5: Effect of Overreaction θ on β_2^{FE}, Across Growth Persistence ρ_g



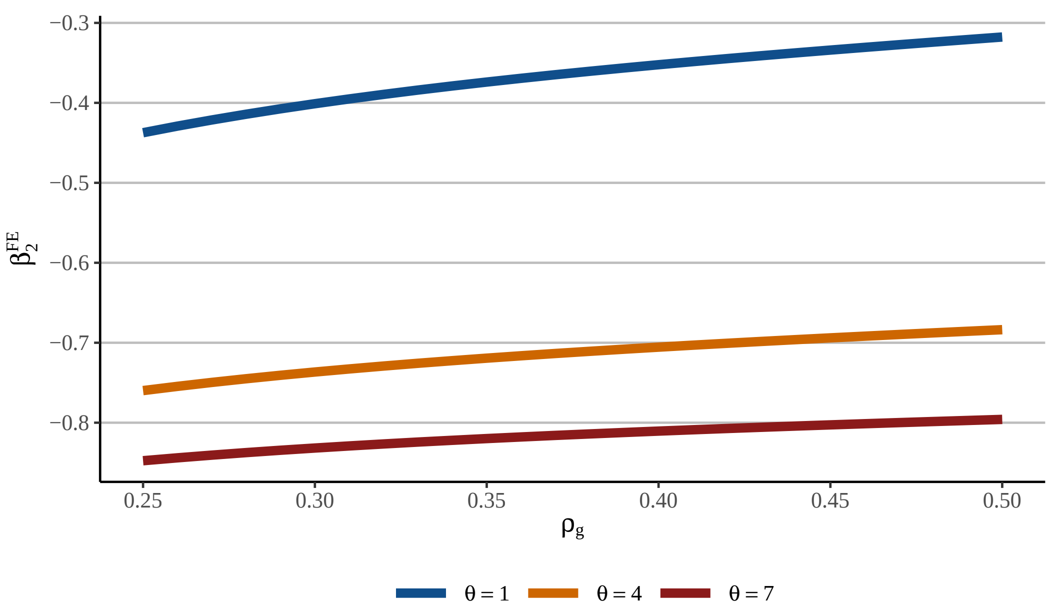
Notes: This figure reports the comparative statics of β_2^{FE} with respect to θ, holding $\rho_\zeta = 0.2$, $\sigma_u = 0.02$, and $s = 5$ fixed. Each line corresponds to a different fixed value of ρ_g: the blue line uses $\rho_g = 0.30$, the orange line uses $\rho_g = 0.35$, and the red line uses $\rho_g = 0.40$.

Figure A6: Effect of Belief Persistence ρ_ζ on β_2^{FE}, Across Growth Persistence ρ_g



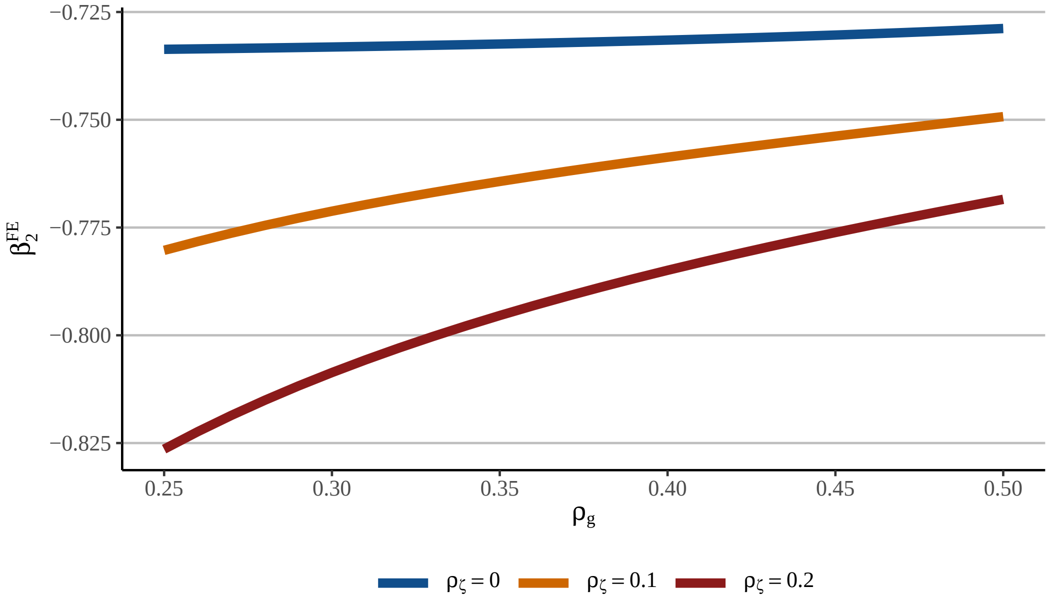
Notes: This figure reports the comparative statics of β_2^{FE} with respect to ρ_ζ, holding $\theta = 6$, $\sigma_u = 0.02$, and $s = 5$ fixed. Each line corresponds to a different fixed value of ρ_g: the blue line uses $\rho_g = 0.30$, the orange line uses $\rho_g = 0.35$, and the red line uses $\rho_g = 0.40$.

Figure A7: Effect of Growth Persistence ρ_g on β_2^{FE}, Across Overreaction θ



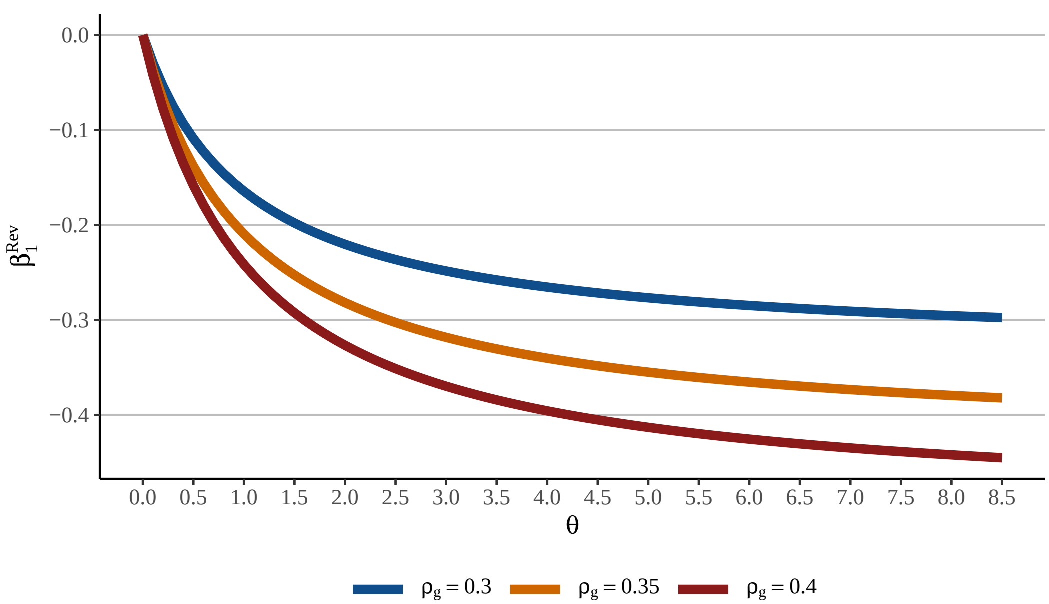
Notes: This figure reports the comparative statics of β_2^{FE} with respect to ρ_g, holding $\rho_\zeta = 0.2$, $\sigma_u = 0.02$, and $s = 5$ fixed. Each line corresponds to a different fixed value of θ: the blue line uses $\theta = 1$, the orange line uses $\theta = 4$, and the red line uses $\theta = 7$.

Figure A8: Effect of Growth Persistence ρ_g on β_2^{FE}, Across Belief Persistence ρ_ζ



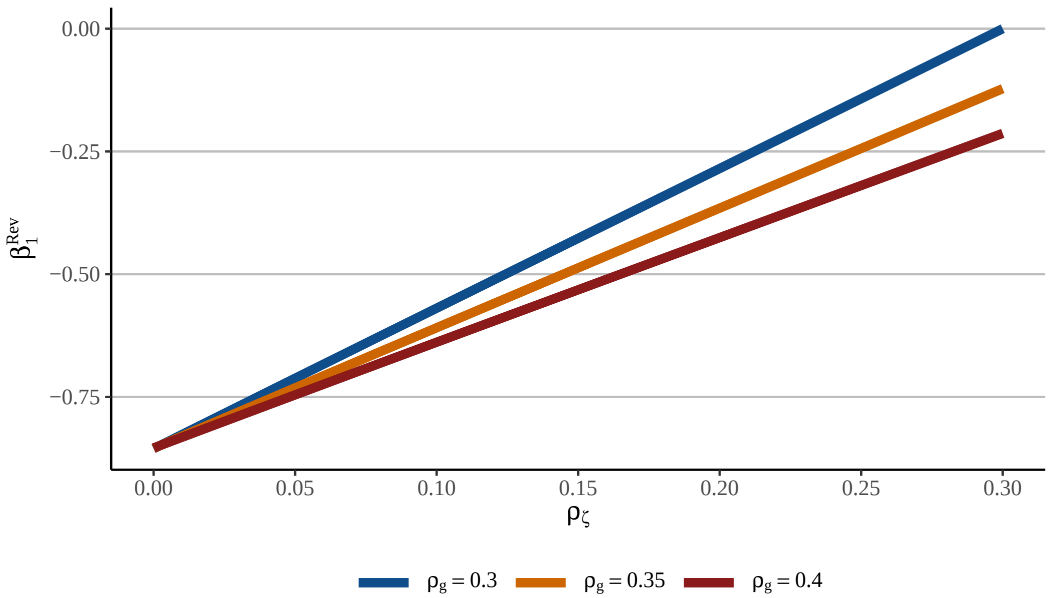
Notes: This figure reports the comparative statics of β_2^{FE} with respect to ρ_g, holding $\theta = 6$, $\sigma_u = 0.02$, and $s = 5$ fixed. Each line corresponds to a different fixed value of ρ_ζ: the blue line uses $\rho_\zeta = 0$, the orange line uses $\rho_\zeta = 0.1$, and the red line uses $\rho_\zeta = 0.2$.

Figure A9: Effect of Overreaction θ on β_1^{Rev}, Across Growth Persistence ρ_g



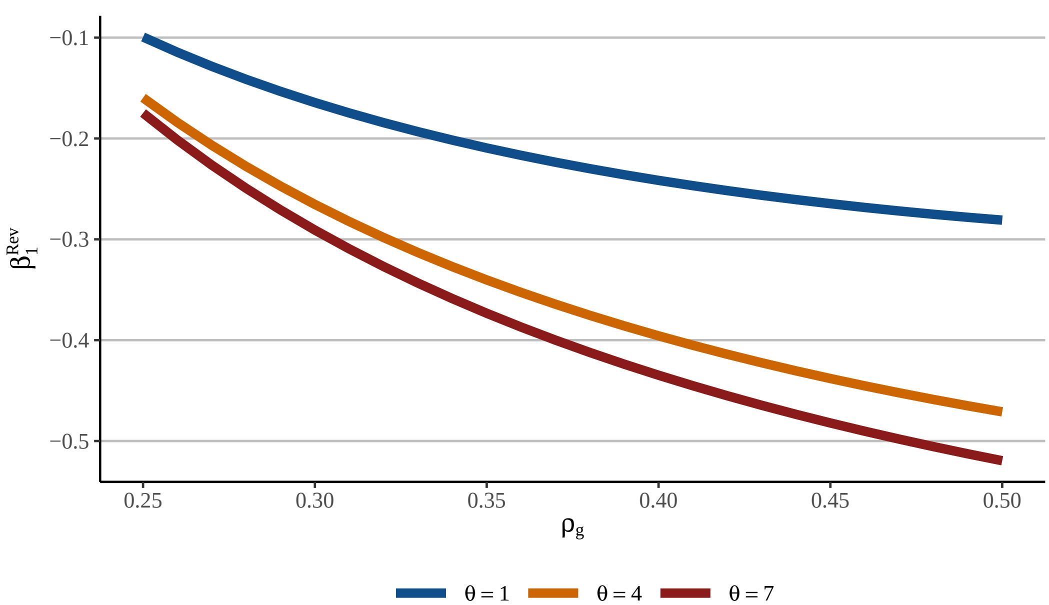
Notes: This figure reports the comparative statics of β_1^{Rev} with respect to θ, holding $\rho_\zeta = 0.2$, $\sigma_u = 0.02$, and $s = 5$ fixed. Each line corresponds to a different fixed value of ρ_g: the blue line uses $\rho_g = 0.30$, the orange line uses $\rho_g = 0.35$, and the red line uses $\rho_g = 0.40$.

Figure A10: Effect of Belief Persistence ρ_ζ on β_1^{Rev}, Across Growth Persistence ρ_g



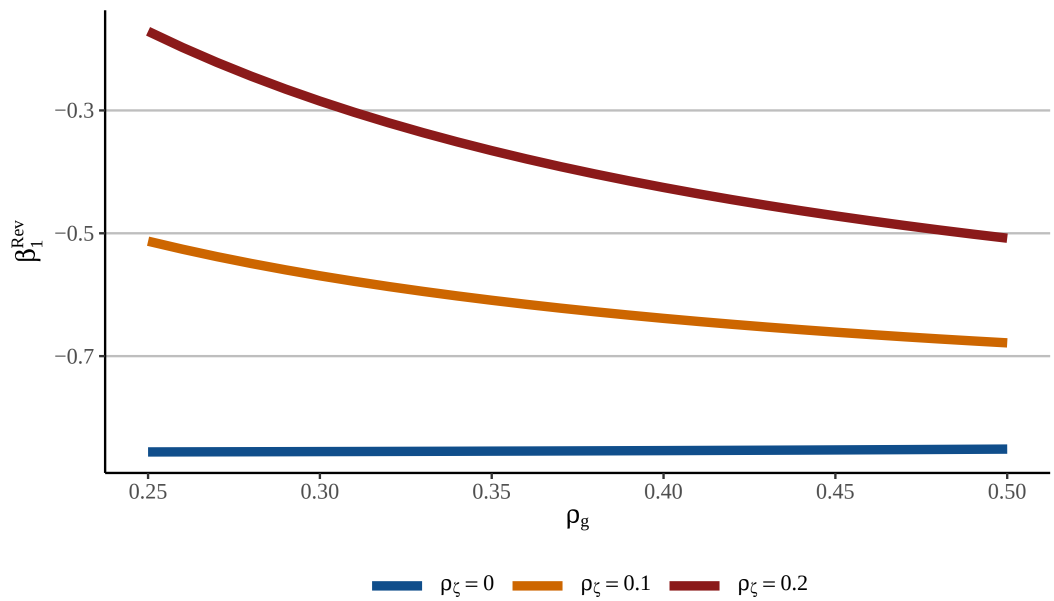
Notes: This figure reports the comparative statics of β_1^{Rev} with respect to ρ_ζ, holding $\theta = 6$, $\sigma_u = 0.02$, and $s = 5$ fixed. Each line corresponds to a different fixed value of ρ_g: the blue line uses $\rho_g = 0.30$, the orange line uses $\rho_g = 0.35$, and the red line uses $\rho_g = 0.40$.

Figure A11: Effect of Growth Persistence ρ_g on β_1^{Rev}, Across Overreaction θ



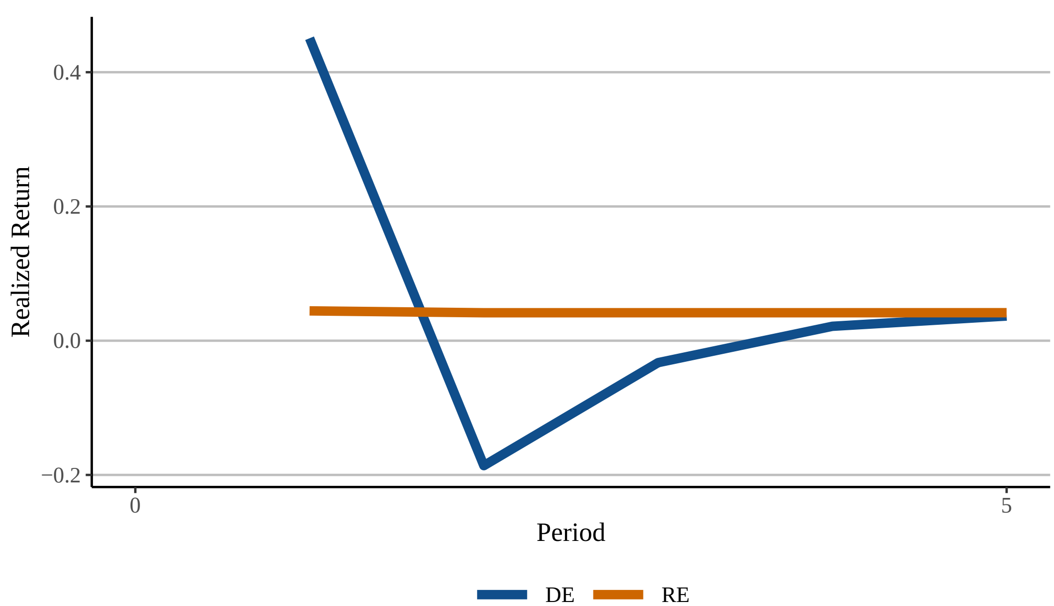
Notes: This figure reports the comparative statics of β_1^{Rev} with respect to ρ_g, holding $\rho_\zeta = 0.2$, $\sigma_u = 0.02$, and $s = 5$ fixed. Each line corresponds to a different fixed value of θ: the blue line uses $\theta = 1$, the orange line uses $\theta = 4$, and the red line uses $\theta = 7$.

Figure A12: Effect of Growth Persistence ρ_g on β_1^{Rev}, Across Belief Persistence ρ_ζ



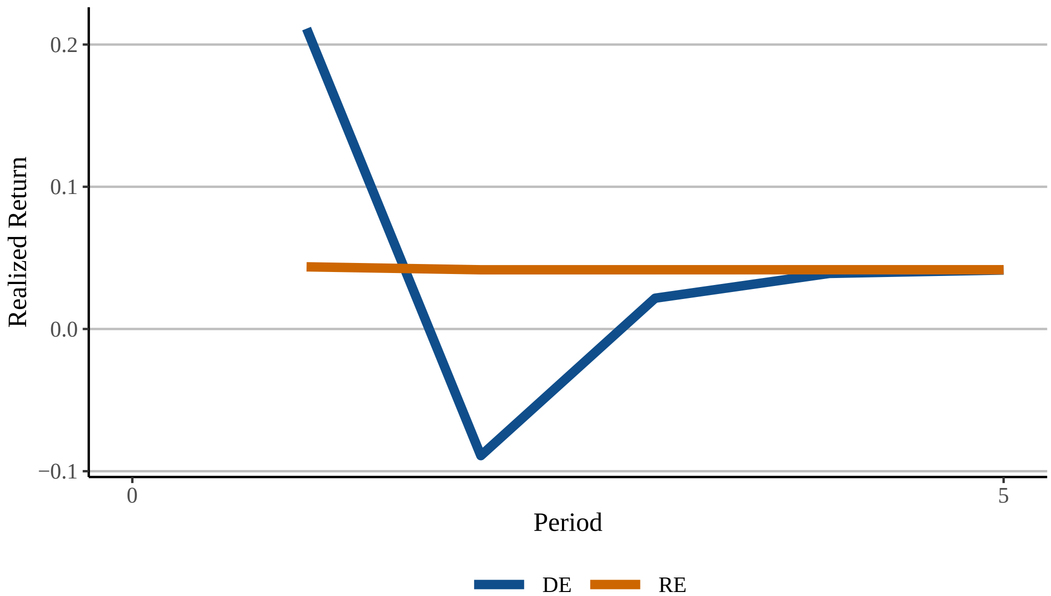
Notes: This figure reports the comparative statics of β_1^{Rev} with respect to ρ_g, holding $\theta = 6$, $\sigma_u = 0.02$, and $s = 5$ fixed. Each line corresponds to a different fixed value of ρ_ζ: the blue line uses $\rho_\zeta = 0$, the orange line uses $\rho_\zeta = 0.1$, and the red line uses $\rho_\zeta = 0.2$.

Figure A13: Impulse Response of Return – Low Market Power Firms (DE vs. RE)



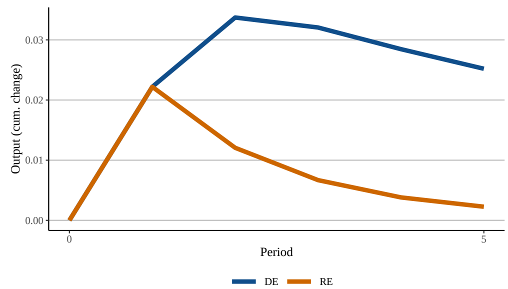
Notes: This figure reports the impulse response function of returns for low market power firms. The blue line denotes IRF under diagnostic expectations and the orange line shows the IRF under rational expectations. The shock is a 1 SD positive productivity shock.

Figure A14: Impulse Response of Return – High Market Power Firms (DE vs. RE)

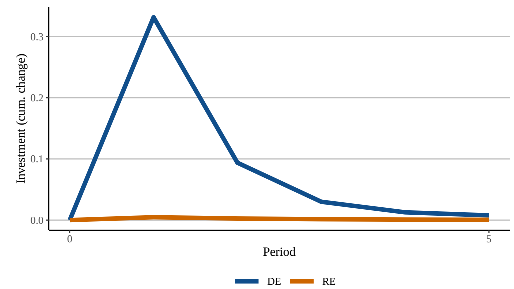


Notes: This figure reports the impulse response function of returns for high market power firms. The blue line denotes IRF under diagnostic expectations and the orange line shows the IRF under rational expectations. The shock is a 1 SD positive productivity shock.

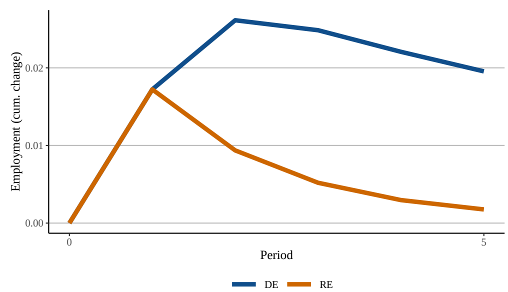
Figure A15: Impulse Response Function: Low Market Power Firms (1 SD Productivity Shock)



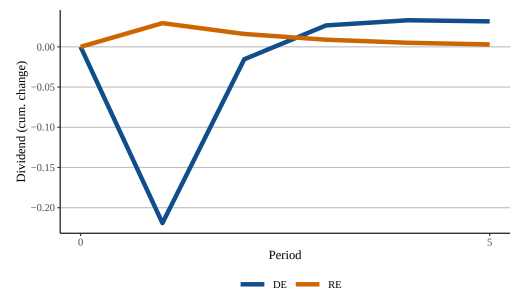
(a) Output



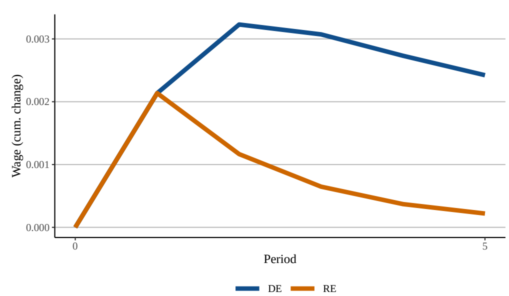
(b) Investment



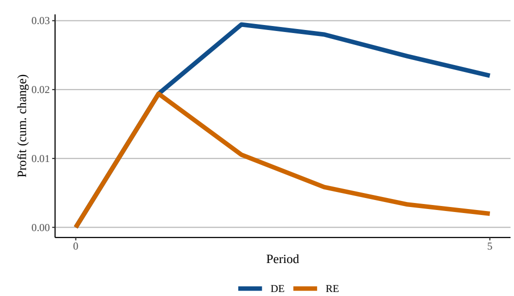
(c) Employment



(d) Dividend



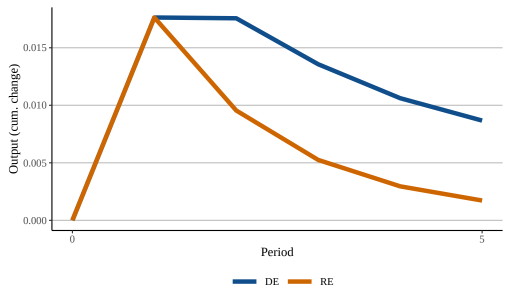
(e) Wage



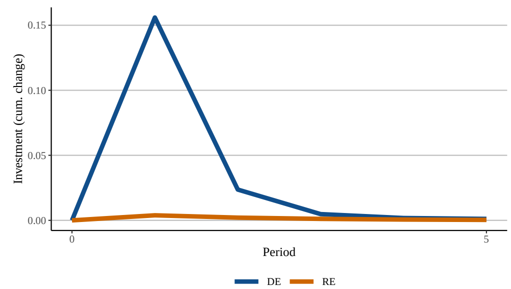
(f) Profit

Notes: This figure reports the impulse response functions for low market power firms under both diagnostic expectations (blue lines) and rational expectations (orange lines). Panel (a) presents output, panel (b) presents investment, panel (c) presents employment, panel (d) presents dividend, panel (e) presents wage, and panel (f) presents profit. The shock is a one standard deviation positive productivity shock.

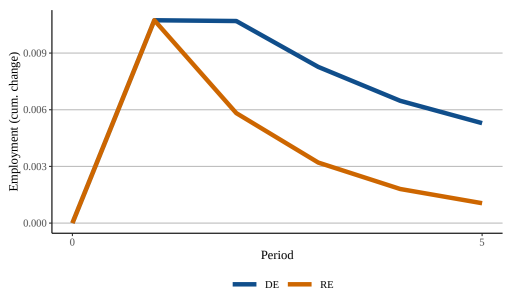
Figure A16: Impulse Response Function: High Market Power Firms (1 SD Productivity Shock)



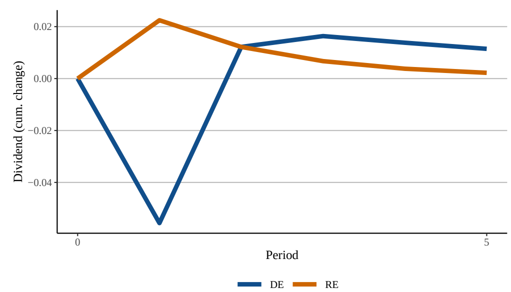
(a) Output



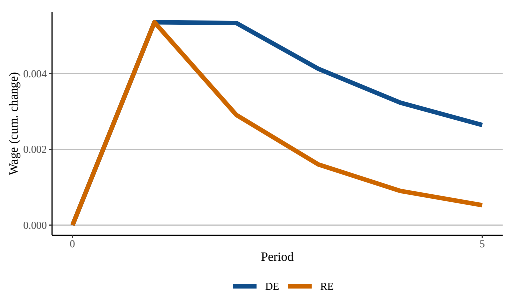
(b) Investment



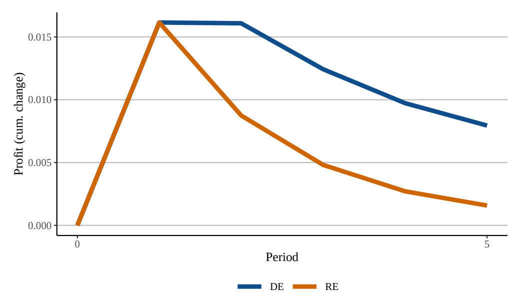
(c) Employment



(d) Dividend



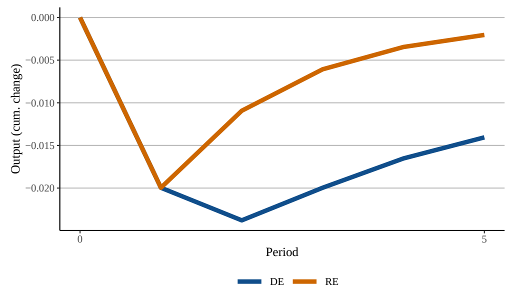
(e) Wage



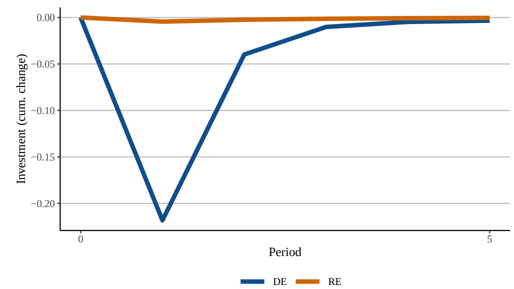
(f) Profit

Notes: This figure reports the impulse response functions for high market power firms under both diagnostic expectations (blue lines) and rational expectations (orange lines). Panel (a) presents output, panel (b) presents investment, panel (c) presents employment, panel (d) presents dividend, panel (e) presents wage, and panel (f) presents profit. The shock is a one standard deviation positive productivity shock.

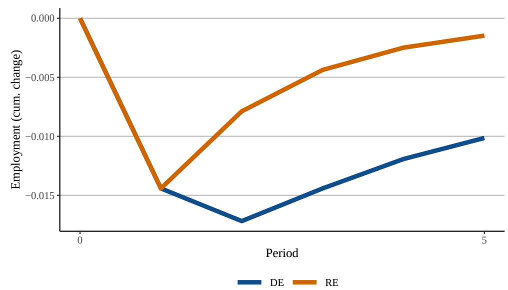
Figure A17: Impulse Response Function: Normal Firms (-1 SD Productivity Shock)



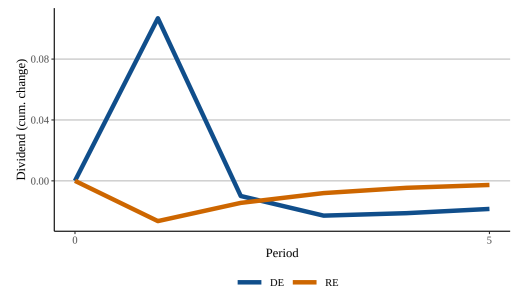
(a) Output



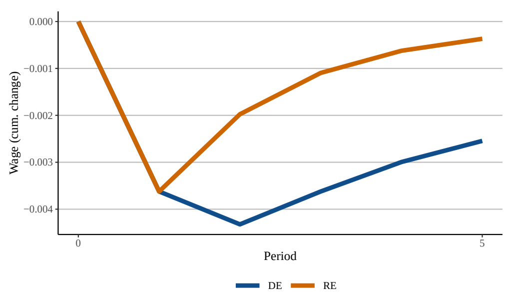
(b) Investment



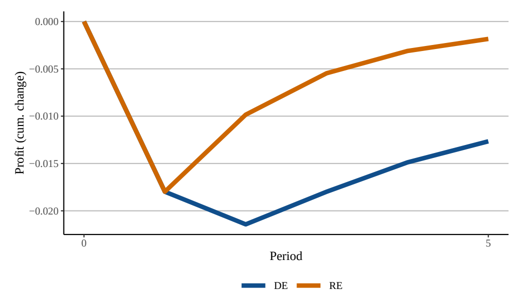
(c) Employment



(d) Dividend



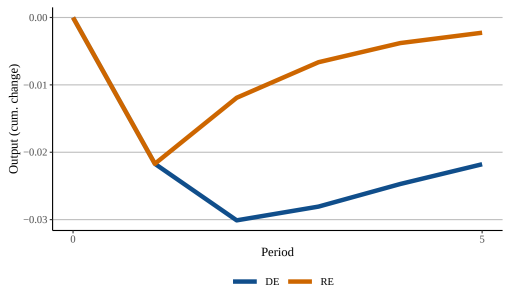
(e) Wage



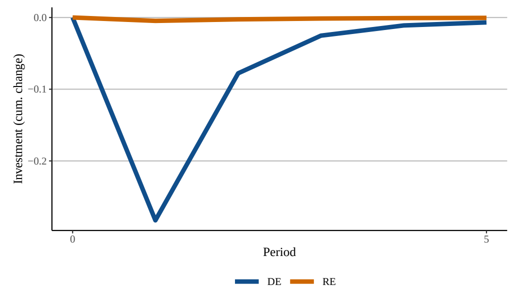
(f) Profit

Notes: This figure reports the impulse response functions for normal firms under both diagnostic expectations (blue lines) and rational expectations (orange lines). Panel (a) presents output, panel (b) presents investment, panel (c) presents employment, panel (d) presents dividend, panel (e) presents wage, and panel (f) presents profit. The shock is a one standard deviation negative productivity shock.

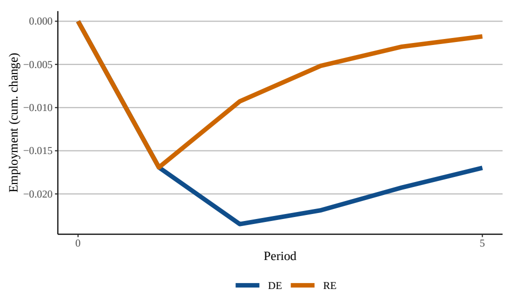
Figure A18: Impulse Response Function: Low Market Power Firms (-1 SD Productivity Shock)



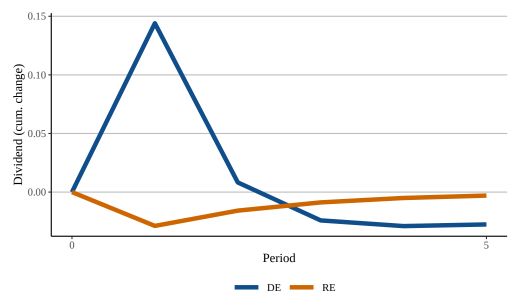
(a) Output



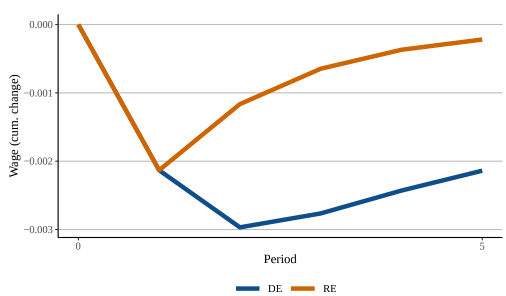
(b) Investment



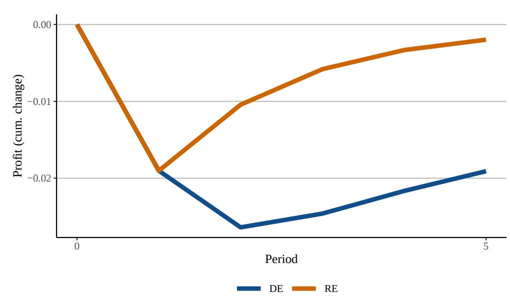
(c) Employment



(d) Dividend



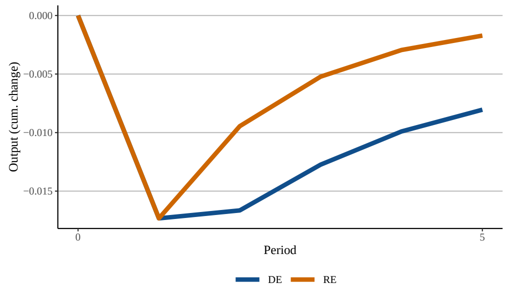
(e) Wage



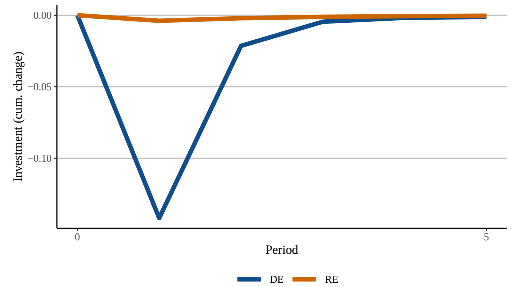
(f) Profit

Notes: This figure reports the impulse response functions for low market power firms under both diagnostic expectations (blue lines) and rational expectations (orange lines). Panel (a) presents output, panel (b) presents investment, panel (c) presents employment, panel (d) presents dividend, panel (e) presents wage, and panel (f) presents profit. The shock is a one standard deviation negative productivity shock.

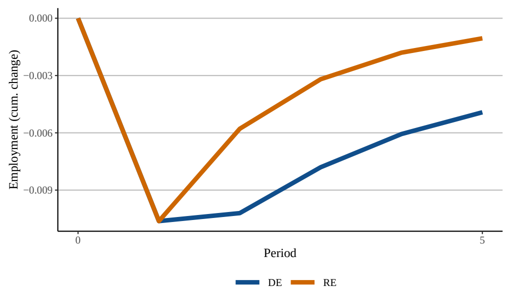
Figure A19: Impulse Response Function: High Market Power Firms (-1 SD Productivity Shock)



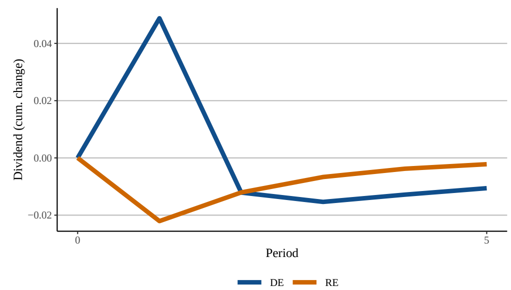
(a) Output



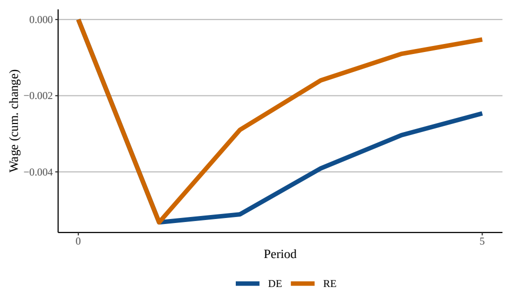
(b) Investment



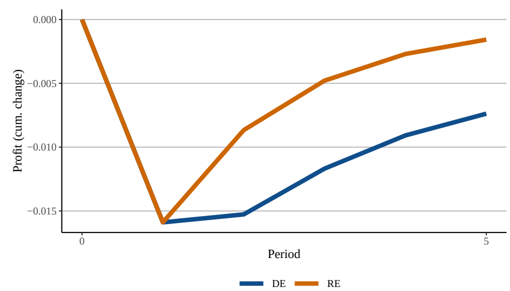
(c) Employment



(d) Dividend



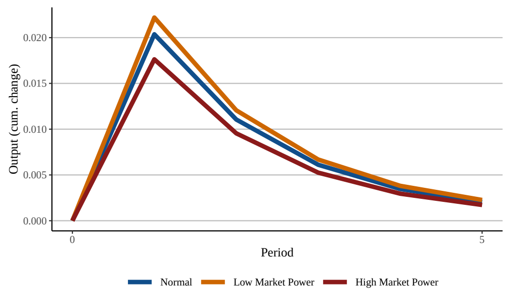
(e) Wage



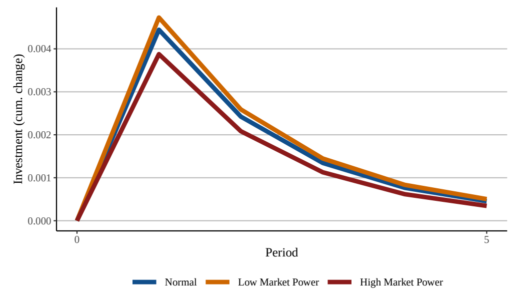
(f) Profit

Notes: This figure reports the impulse response functions for high market power firms under both diagnostic expectations (blue lines) and rational expectations (orange lines). Panel (a) presents output, panel (b) presents investment, panel (c) presents employment, panel (d) presents dividend, panel (e) presents wage, and panel (f) presents profit. The shock is a one standard deviation negative productivity shock.

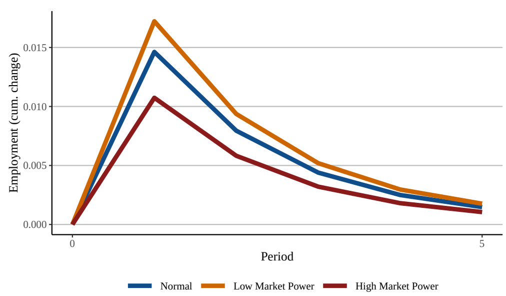
Figure A20: Impulse Response Function: All Firms under RE (1 SD Productivity Shock)



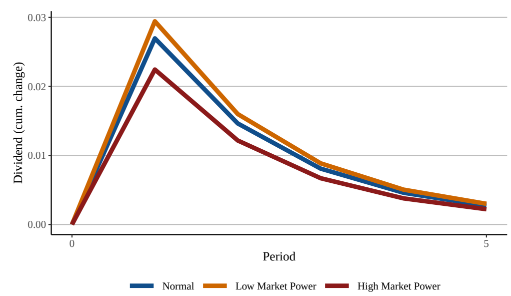
(a) Output



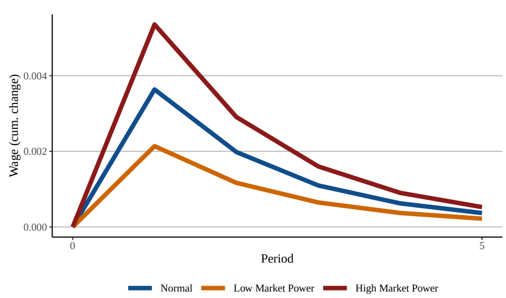
(b) Investment



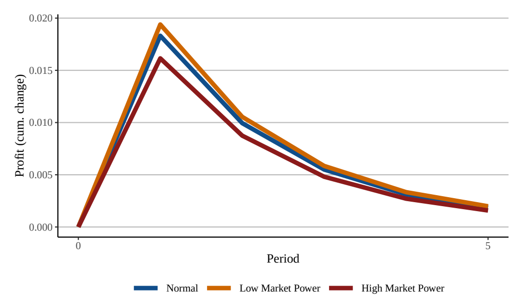
(c) Employment



(d) Dividend



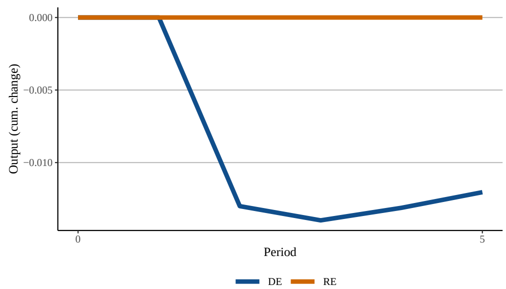
(e) Wage



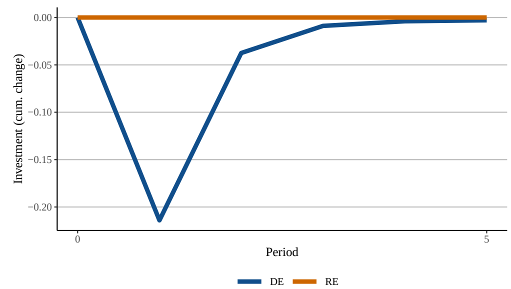
(f) Profit

Notes: This figure reports the impulse response functions for all firms under rational expectations. The blue line represents normal firms, the orange line represents low market power firms, and the red line represents high market power firms. Panel (a) presents output, panel (b) presents investment, panel (c) presents employment, panel (d) presents dividend, panel (e) presents wage, and panel (f) presents profit. The shock is a one standard deviation positive productivity shock.

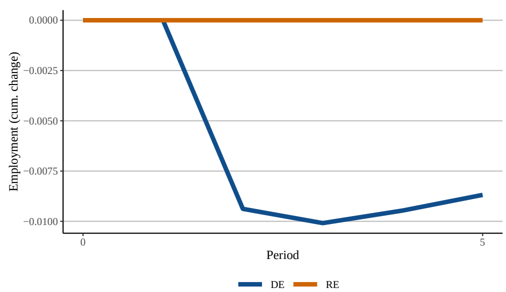
Figure A21: Impulse Response Function: Normal Firms (-1 SD Belief Distortion Shock)



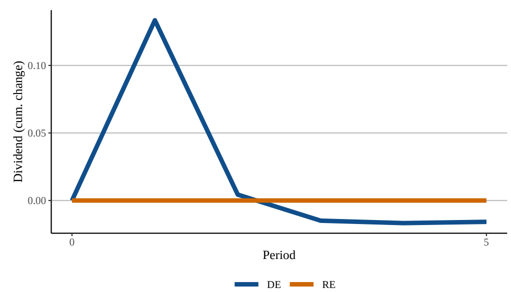
(a) Output



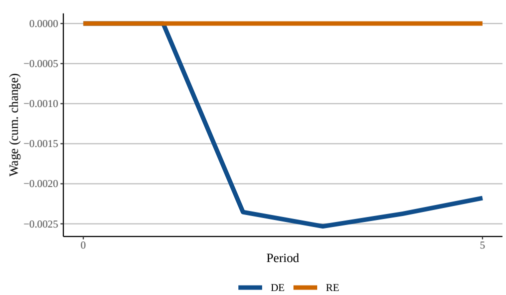
(b) Investment



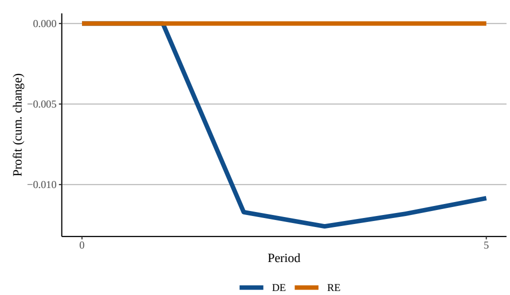
(c) Employment



(d) Dividend



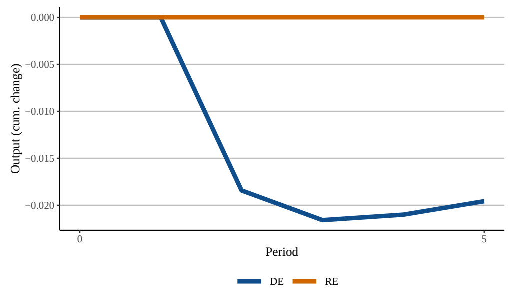
(e) Wage



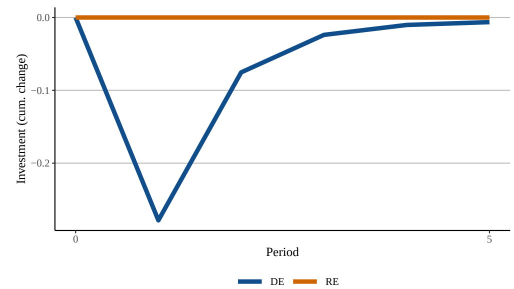
(f) Profit

Notes: This figure reports the impulse response functions for normal firms under both diagnostic expectations (blue lines) and rational expectations (orange lines). Panel (a) presents output, panel (b) presents investment, panel (c) presents employment, panel (d) presents dividend, panel (e) presents wage, and panel (f) presents profit. The shock is a one standard deviation negative belief distortion shock.

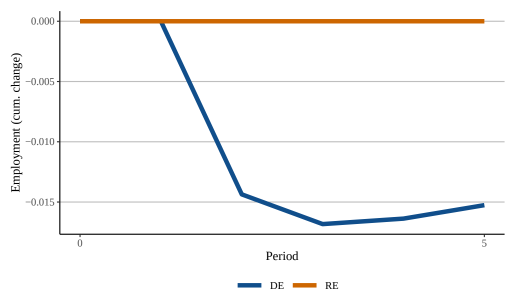
Figure A22: Impulse Response Function: Low Market Power Firms (-1 SD Belief Distortion Shock)



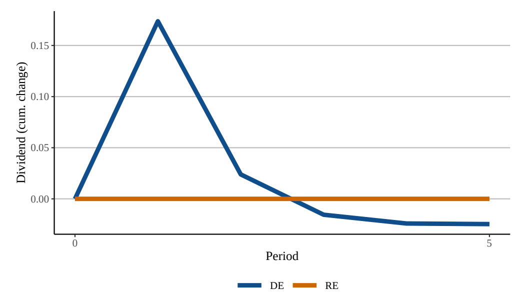
(a) Output



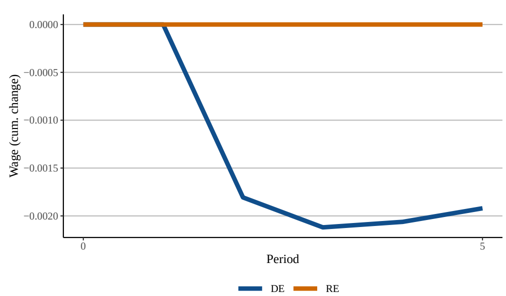
(b) Investment



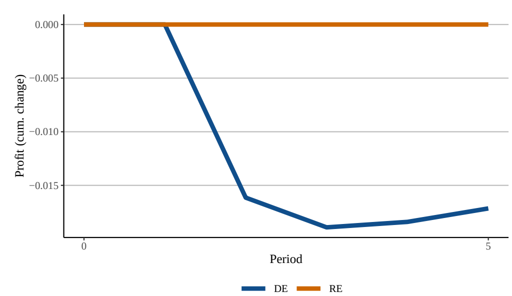
(c) Employment



(d) Dividend



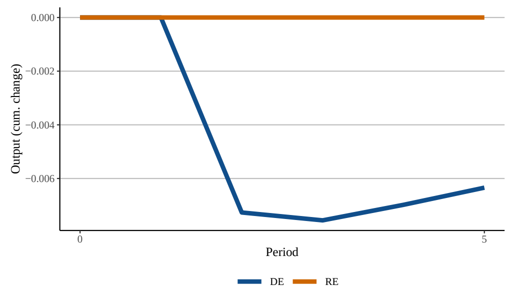
(e) Wage



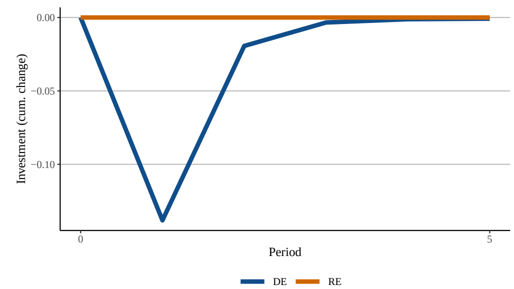
(f) Profit

Notes: This figure reports the impulse response functions for low market power firms under both diagnostic expectations (blue lines) and rational expectations (orange lines). Panel (a) presents output, panel (b) presents investment, panel (c) presents employment, panel (d) presents dividend, panel (e) presents wage, and panel (f) presents profit. The shock is a one standard deviation negative belief distortion shock.

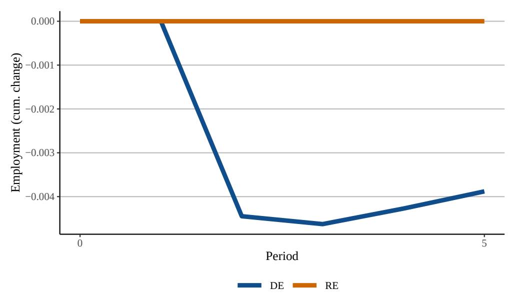
Figure A23: Impulse Response Function: High Market Power Firms (-1 SD Belief Distortion Shock)



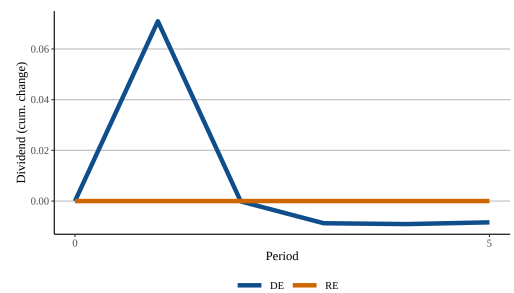
(a) Output



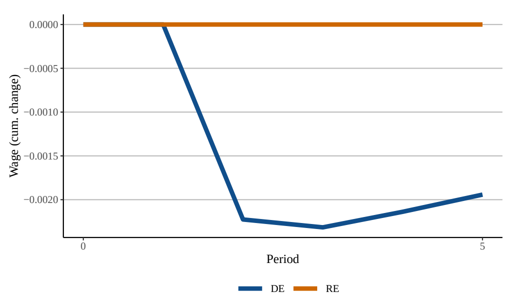
(b) Investment



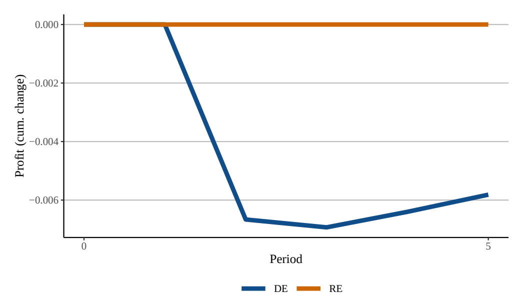
(c) Employment



(d) Dividend



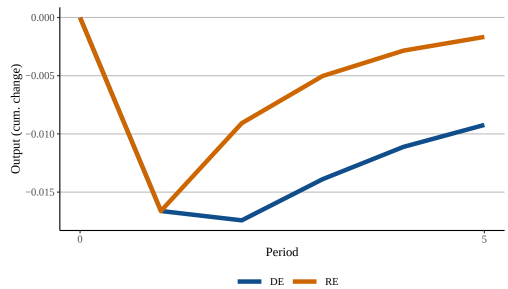
(e) Wage



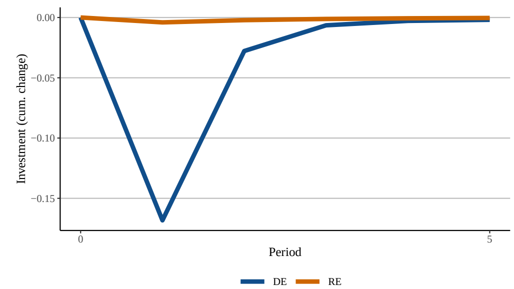
(f) Profit

Notes: This figure reports the impulse response functions for high market power firms under both diagnostic expectations (blue lines) and rational expectations (orange lines). Panel (a) presents output, panel (b) presents investment, panel (c) presents employment, panel (d) presents dividend, panel (e) presents wage, and panel (f) presents profit. The shock is a one standard deviation negative belief distortion shock.

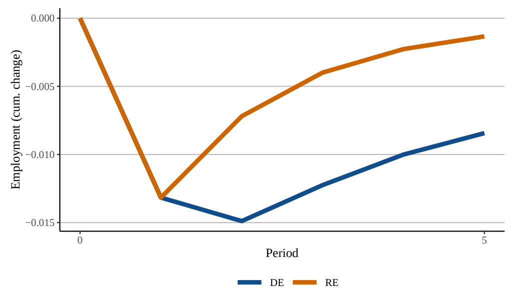
Figure A24: Impulse Response Function: Aggregate (-1 SD Productivity Shock)



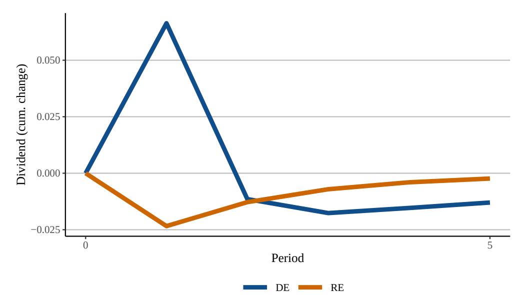
(a) Output



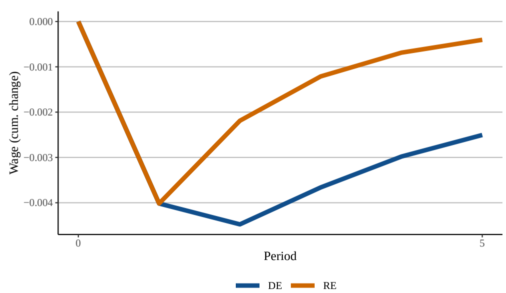
(b) Investment



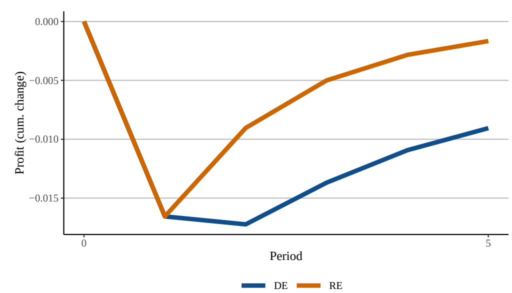
(c) Employment



(d) Dividend



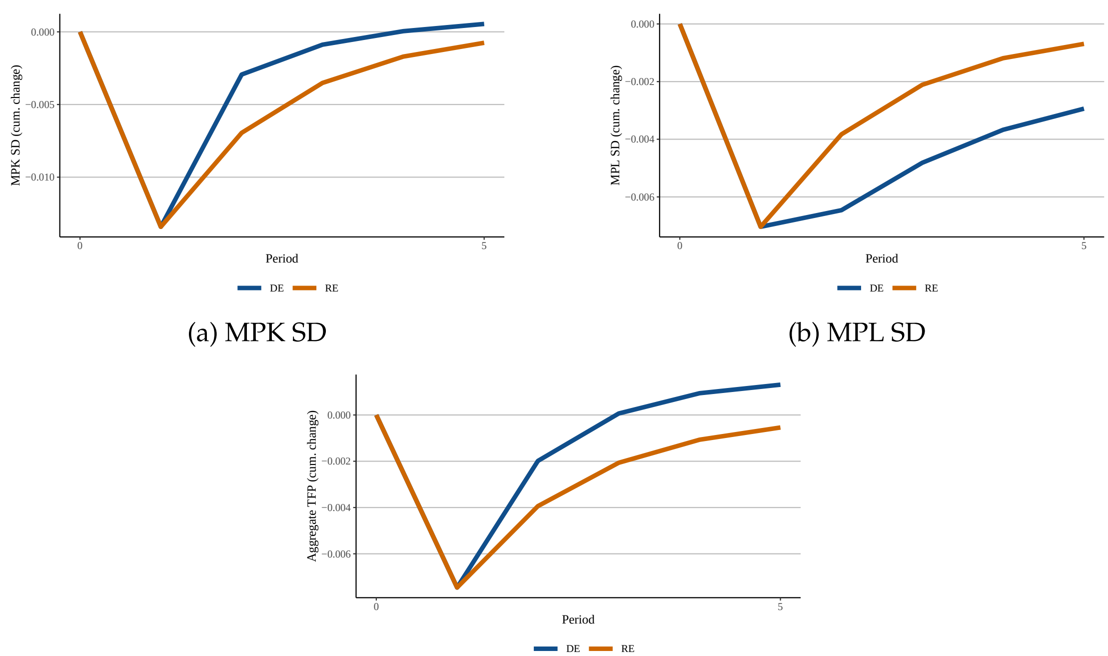
(e) Wage



(f) Profit

Notes: This figure reports the aggregate impulse response functions under both diagnostic expectations (blue lines) and rational expectations (orange lines). Panel (a) presents output, panel (b) presents investment, panel (c) presents employment, panel (d) presents dividend, panel (e) presents wage, and panel (f) presents profit. The shock is a one standard deviation negative productivity shock.

Figure A25: Impulse Response Function: Aggregate Misallocation (-1 SD Productivity Shock)



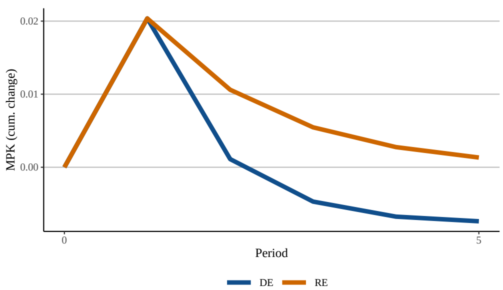
(a) MPK SD

(b) MPL SD

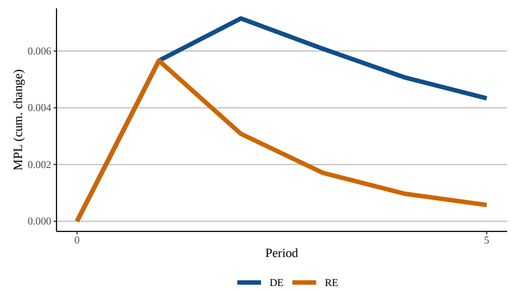
(c) Aggregate Productivity

Notes: This figure reports the aggregate impulse response functions under both diagnostic expectations (blue lines) and rational expectations (orange lines). Panel (a) presents the standard deviation of MPK, panel (b) presents the standard deviation of MPL, and panel (c) presents aggregate productivity.

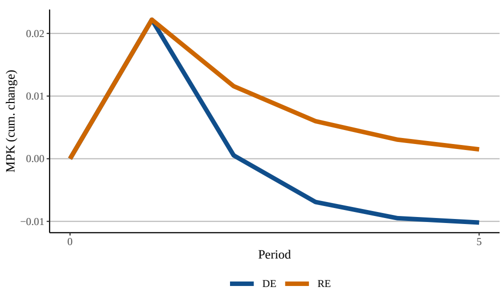
Figure A26: Impulse Response Functions: Firm-Level Marginal Products



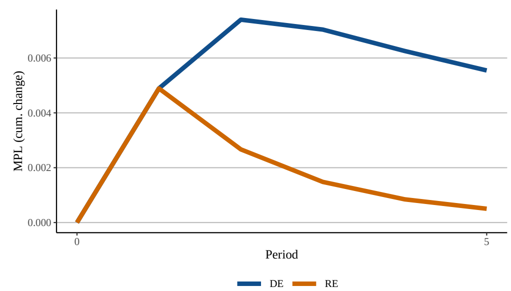
(a) MPK Normal Firms



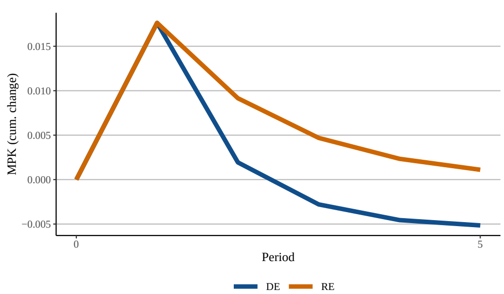
(b) MPL Normal Firms



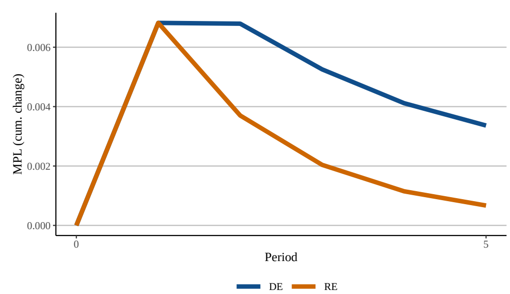
(c) MPK Low Market Power Firms



(d) MPL Low Market Power Firms



(e) MPK High Market Power Firms



(f) MPL High Market Power Firms

Notes: This figure reports the impulse response functions of the marginal product of capital (MPK) and marginal product of labor (MPL) for all firm types. For all panels the blue line indicates the IRF for the case of diagnostic expectations and the orange line indicates the IRF for the case of rational expectations. Panels (a) and (b) show the MPK and MPL, respectively, for normal firms. Panels (c) and (d) show the MPK and MPL, respectively, for low market power firms. Panels (e) and (f) show the MPK and MPL, respectively, for high market power firms.

A.2 Tables

Table A1: Full Compustat Sample Summary Statistics (Annual Level)

Variable	Mean (1)	SD (2)	P10 (3)	P25 (4)	Median (5)	P75 (6)	P90 (7)	Obs. (8)
Log Sales	19.040	2.484	15.900	17.410	19.110	20.720	22.200	151,000
Log COGS	18.490	2.577	15.210	16.790	18.550	20.250	21.770	151,000
Log SGA	17.650	2.098	15.050	16.140	17.550	19.020	20.440	151,000
Log Physical Capital	17.580	2.692	14.190	15.660	17.510	19.440	21.190	151,000
Log Intangible Capital	17.750	2.150	15.080	16.200	17.640	19.130	20.620	151,000
Log Total Assets	19.080	2.396	16.040	17.390	19.050	20.710	22.230	151,000
Log Market Cap	18.640	2.522	15.510	16.810	18.530	20.380	21.980	140,000

Notes: This table presents the summary statistics of the original Compustat sample at the annual level. The sample ranges from 1977 to 2019. All nominal variables are deflated using the BEA's GDP Price Deflator. Column (1) reports the mean, and Column (2) reports the standard deviation. Columns (3) to (7) report the 10th percentile, 25th percentile, median, 75th percentile, and 90th percentile, respectively. Column (8) reports the number of observations. All figures are rounded in accordance with U.S. Census disclosure requirements. This table is directly replicated from Ren and Zhang (2025).

Table A2: Price Markups and Firm Characteristics (CRS)

	Log Price Markup						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log Wage Markdown	-0.104						
	(0.004)						
Log TFPR		0.002					
		(0.015)					
Log Labor Productivity			-0.045				
			(0.004)				
Log Sales				-0.016			
				(0.001)			
Log Wage Bill					-0.001		
					(0.001)		
Profit Share						0.007	
						(0.018)	
Log Labor Share VA							0.044
							(0.004)
NAICS2 FE	No	No	No	No	No	No	No
NAICS2 $\times$ Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	No	No	No	No	No	No	No
Observations	69,500	69,500	69,000	69,500	69,500	69,500	69,500

Notes: This table reports the results from regressing log price markups onto firm characteristics with NAICS2 $\times$ year fixed effects. This table shows the results for markups and markdowns estimated with the CRS restriction. Column (1) examines log wage markdowns, Columns (2) and (3) analyze log TFPR and log labor productivity, Columns (4) and (5) assess log sales and log wage bill, and Columns (6) and (7) evaluate profit share and labor share of value-added. Standard errors are reported in parentheses and are two-way clustered by firm and year. All figures are rounded in accordance with U.S. Census disclosure requirements. This table is reproduced exactly from Ren and Zhang (2025).

Table A3: Wage Markdowns and Firm Characteristics (CRS)

	Log Wage Markdown						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log Price Markup	-2.627						
	(0.057)						
Log TFPR		1.205					
		(0.128)					
Log Labor Productivity			0.236				
			(0.018)				
Log Sales				0.206			
				(0.008)			
Log Wage Bill					0.133		
					(0.007)		
Profit Share						0.699	
						(0.097)	
Log Labor Share VA							-0.423
							(0.026)
NAICS2 FE	No	No	No	No	No	No	No
NAICS2 $\times$ Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	No	No	No	No	No	No	No
Observations	69,500	69,500	69,000	69,500	69,500	69,500	69,500

Notes: This table reports the results from regressing log wage markdowns onto firm characteristics with NAICS2 $\times$ year fixed effects. This table shows the results for markups and markdowns estimated with the CRS restriction. Column (1) examines log price markups, Columns (2) and (3) analyze log TFPR and log labor productivity, Columns (4) and (5) assess log sales and log wage bill, and Columns (6) and (7) evaluate profit share and labor share of value-added. Standard errors are reported in parentheses and are two-way clustered by firm and year. All figures are rounded in accordance with U.S. Census disclosure requirements. This table is reproduced exactly from Ren and Zhang (2025).

Table A4: Total Wedge and Firm Characteristics (CRS)

	Log Total Wedge					
	(1)	(2)	(3)	(4)	(5)	(6)
Log TFPR	1.189					
	(0.124)					
Log Labor Productivity		0.140				
		(0.016)				
Log Sales			0.186			
			(0.009)			
Log Wage Bill				0.128		
				(0.007)		
Profit Share					0.549	
					(0.092)	
Log Labor Share VA						-0.340
						(0.020)
NAICS2 FE	No	No	No	No	No	No
NAICS2 $\times$ Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	No	No	No	No	No	No
Observations	69,500	69,000	69,500	69,500	69,500	69,500

Notes: This table reports the results from regressing log total wedges onto firm characteristics with NAICS2 $\times$ year fixed effects. This table shows the results for markups and markdowns estimated with the CRS restriction. Columns (1) and (2) analyze log TFPR and log labor productivity, Columns (3) and (4) assess log sales and log wage bill, and Columns (5) and (6) evaluate profit share and labor share of value-added. Standard errors are reported in parentheses and are two-way clustered by firm and year. All figures are rounded in accordance with U.S. Census disclosure requirements. This table is reproduced exactly from Ren and Zhang (2025).

Table A5: 5-Year Forecast Error Predictability (NAICS2 $\times$ Year and Firm FEs)

	Forecast Error (5-Year)		
	(1)	(2)	(3)
LTG Revision	-0.896	-0.912	-0.909
	(0.022)	(0.026)	(0.027)
LTG Revision $\times$ Low Market Power	-0.054		-0.045
	(0.103)		(0.104)
LTG Revision $\times$ High Market Power		0.048	0.045
		(0.067)	(0.068)
LTG (Lag 1)	-0.912	-0.948	-0.945
	(0.029)	(0.032)	(0.032)
LTG (Lag 1) $\times$ Low Market Power (Lag 1)	-0.057		-0.033
	(0.062)		(0.062)
LTG (Lag 1) $\times$ High Market Power (Lag 1)		0.166	0.163
		(0.058)	(0.058)
NAICS2 $\times$ Month FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Observations	246,000	246,000	246,000

Notes: This table presents regressing 5-year realized forecast error against LTG revisions and lagged LTG along with interactions. All specifications include NAICS2 $\times$ month and firm fixed effects. Column (1) presents the specification with only an interaction for low total market power firms. Column (2) presents the specification with only an interaction for high total market power firms. Column (3) presents the specification with interactions with both low total market power and high total market power firms. Standard errors are reported in parentheses and are two-way clustered by firm and year. All figures are rounded in accordance with U.S. Census disclosure requirements.

Table A6: 3-Year Forecast Error Predictability (NAICS2 × Month FE)

	Forecast Error (3-Year)		
	(1)	(2)	(3)
LTG Revision	-0.986	-0.983	-0.984
	(0.032)	(0.038)	(0.039)
LTG Revision × Low Market Power	0.040		0.041
	(0.116)		(0.118)
LTG Revision × High Market Power		0.014	0.015
		(0.090)	(0.091)
LTG (Lag 1)	-1.231	-1.242	-1.238
	(0.037)	(0.037)	(0.038)
LTG (Lag 1) × Low Market Power (Lag 1)	-0.054		-0.044
	(0.078)		(0.079)
LTG (Lag 1) × High Market Power (Lag 1)		0.128	0.123
		(0.071)	(0.071)
NAICS2 × Month FE	Yes	Yes	Yes
Firm FE	No	No	No
Observations	272,000	272,000	272,000

Notes: This table presents regressing 3-year realized forecast error against LTG revisions and lagged LTG along with interactions. All specifications include NAICS2 × month fixed effects. Column (1) presents the specification with only an interaction for low total market power firms. Column (2) presents the specification with only an interaction for high total market power firms. Column (3) presents the specification with interactions with both low total market power and high total market power firms. Standard errors are reported in parentheses and are two-way clustered by firm and year. All figures are rounded in accordance with U.S. Census disclosure requirements.

Table A7: 3-Year Forecast Error Predictability (NAICS2 $\times$ Year and Firm FEs)

	Forecast Error (3-Year)		
	(1)	(2)	(3)
LTG Revision	-0.853	-0.849	-0.846
	(0.030)	(0.032)	(0.033)
LTG Revision $\times$ Low Market Power	-0.030		-0.038
	(0.103)		(0.105)
LTG Revision $\times$ High Market Power		-0.067	-0.069
		(0.112)	(0.113)
LTG (Lag 1)	-0.894	-0.910	-0.908
	(0.030)	(0.032)	(0.033)
LTG (Lag 1) $\times$ Low Market Power (Lag 1)	-0.031		-0.021
	(0.073)		(0.073)
LTG (Lag 1) $\times$ High Market Power (Lag 1)		0.086	0.085
		(0.061)	(0.061)
NAICS2 $\times$ Month FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Observations	272,000	272,000	272,000

Notes: This table presents regressing 3-year realized forecast error against LTG revisions and lagged LTG along with interactions. All specifications include NAICS2 $\times$ month and firm fixed effects. Column (1) presents the specification with only an interaction for low total market power firms. Column (2) presents the specification with only an interaction for high total market power firms. Column (3) presents the specification with interactions with both low total market power and high total market power firms. Standard errors are reported in parentheses and are two-way clustered by firm and year. All figures are rounded in accordance with U.S. Census disclosure requirements.

Table A8: LTG Revision Predictability (NAICS2 FE)

	LTG Revision		
	(1)	(2)	(3)
LTG (Lag 1)	-0.483	-0.477	-0.488
	(0.038)	(0.038)	(0.038)
LTG (Lag 1) $\times$ Low Market Power (Lag 1)	0.121		0.124
	(0.064)		(0.064)
LTG (Lag 1) $\times$ High Market Power (Lag 1)		0.009	0.019
		(0.058)	(0.058)
NAICS2 FE	Yes	Yes	Yes
NAICS2 $\times$ Year FE	No	No	No
Firm FE	No	No	No
Observations	30,000	30,000	30,000

Notes: This table presents regressing future LTG revisions against lagged LTG along with interactions. All specifications include NAICS2 fixed effects. Column (1) presents the specification with only an interaction for low total market power firms. Column (2) presents the specification with only an interaction for high total market power firms. Column (3) presents the specification with interactions with both low total market power and high total market power firms. Standard errors are reported in parentheses and are two-way clustered by firm and year. All figures are rounded in accordance with U.S. Census disclosure requirements.

Table A9: 5-Year Return Predictability (NAICS2 × Year and Firm FEs)

	Future Return (5-Year)		
	(1)	(2)	(3)
LTG Revision	-0.371	-0.432	-0.424
	(0.079)	(0.076)	(0.089)
LTG Revision × Low Market Power	-0.218		-0.166
	(0.211)		(0.219)
LTG Revision × High Market Power		0.277	0.270
		(0.143)	(0.149)
LTG (Lag 1)	-0.659	-0.730	-0.706
	(0.106)	(0.113)	(0.116)
LTG (Lag 1) × Low Market Power (Lag 1)	-0.294		-0.254
	(0.198)		(0.199)
LTG (Lag 1) × High Market Power (Lag 1)		0.284	0.262
		(0.186)	(0.186)
NAICS2 × Month FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Observations	340,000	340,000	340,000

Notes: This table presents regressing 5-year future returns against LTG revisions and lagged LTG along with interactions. All specifications include NAICS2 × month and firm fixed effects. Column (1) presents the specification with only an interaction for low total market power firms. Column (2) presents the specification with only an interaction for high total market power firms. Column (3) presents the specification with interactions with both low total market power and high total market power firms. Standard errors are reported in parentheses and are two-way clustered by firm and year. All figures are rounded in accordance with U.S. Census disclosure requirements.

Table A10: 3-Year Return Predictability (NAICS2 × Month FE)

	Future Return (3-Year)		
	(1)	(2)	(3)
LTG Revision	-0.204	-0.245	-0.226
	(0.053)	(0.054)	(0.057)
LTG Revision × High Markup	-0.326		-0.309
	(0.138)		(0.141)
LTG Revision × High Markdown		0.177	0.162
		(0.099)	(0.100)
LTG (Lag 1)	-0.569	-0.638	-0.598
	(0.074)	(0.075)	(0.078)
LTG (Lag 1) × High Markup (Lag 1)	-0.443		-0.420
	(0.158)		(0.161)
LTG (Lag 1) × High Markdown (Lag 1)		0.291	0.259
		(0.166)	(0.165)
NAICS2 × Month FE	Yes	Yes	Yes
Firm FE	No	No	No
Observations	379,000	379,000	379,000

Notes: This table presents regressing 3-year future returns against LTG revisions and lagged LTG along with interactions. All specifications include NAICS2 × month fixed effects. Column (1) presents the specification with only an interaction for low total market power firms. Column (2) presents the specification with only an interaction for high total market power firms. Column (3) presents the specification with interactions with both low total market power and high total market power firms. Standard errors are reported in parentheses and are two-way clustered by firm and year. All figures are rounded in accordance with U.S. Census disclosure requirements.

Table A11: 3-Year Return Predictability (NAICS2 × Year and Firm FEs)

	Future Return (3-Year)		
	(1)	(2)	(3)
LTG Revision	-0.208	-0.239	-0.224
	(0.054)	(0.054)	(0.058)
LTG Revision × Low Market Power	-0.274		-0.259
	(0.133)		(0.136)
LTG Revision × High Market Power		0.101	0.087
		(0.102)	(0.105)
LTG (Lag 1)	-0.448	-0.503	-0.477
	(0.076)	(0.080)	(0.082)
LTG (Lag 1) × Low Market Power (Lag 1)	-0.344		-0.319
	(0.148)		(0.149)
LTG (Lag 1) × High Market Power (Lag 1)		0.213	0.188
		(0.137)	(0.137)
NAICS2 × Month FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Observations	379,000	379,000	379,000

Notes: This table presents regressing 3-year future returns against LTG revisions and lagged LTG along with interactions. All specifications include NAICS2 × month and firm fixed effects. Column (1) presents the specification with only an interaction for low total market power firms. Column (2) presents the specification with only an interaction for high total market power firms. Column (3) presents the specification with interactions with both low total market power and high total market power firms. Standard errors are reported in parentheses and are two-way clustered by firm and year. All figures are rounded in accordance with U.S. Census disclosure requirements.

Table A12: Persistence of Price Markups (NAICS2 $\times$ Year FE)

	Log Price Markup				
	(1)	(2)	(3)	(4)	(5)
Log Price Markup (Lag 1)	0.856	0.849	0.851	0.705	0.703
	(0.009)	(0.010)	(0.010)	(0.013)	(0.013)
Log Price Markup (Lag 2)				0.184	0.183
				(0.012)	(0.012)
Log Wage Markdown (Lag 1)		-0.003	-0.000		0.001
		(0.001)	(0.001)		(0.001)
Log Sale (Lag 1)			-0.002		-0.001
			(0.000)		(0.000)
NAICS2 FE	No	No	No	No	No
NAICS2 $\times$ Year FE	Yes	Yes	Yes	Yes	Yes
Firm FE	No	No	No	No	No
Observations	60,500	60,500	60,500	53,000	53,000

Notes: This table presents various specifications regressing the log price markup onto its own lags and various controls. All specifications include NAICS2 $\times$ year fixed effects. Column (1) presents an AR(1) specification. Column (2) adds the first lag of the log wage markdown as a control. Column (3) builds on Column (2) by further including the first lag of log sales. Column (4) estimates an AR(2) specification, and Column (5) adds the first lag of log sales to the AR(2) model. Standard errors are reported in parentheses and are two-way clustered by firm and year. All figures are rounded in accordance with U.S. Census disclosure requirements.

Table A13: Persistence of Price Markups (NAICS2 $\times$ Year and Firm FEs)

	Log Price Markup				
	(1)	(2)	(3)	(4)	(5)
Log Price Markup (Lag 1)	0.585	0.580	0.579	0.531	0.531
	(0.016)	(0.015)	(0.015)	(0.016)	(0.016)
Log Price Markup (Lag 2)				0.101	0.101
				(0.012)	(0.012)
Log Wage Markdown (Lag 1)		-0.002	-0.002		0.000
		(0.002)	(0.002)		(0.002)
Log Sale (Lag 1)			-0.001		0.000
			(0.002)		(0.002)
NAICS2 FE	No	No	No	No	No
NAICS2 $\times$ Year FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
Observations	60,500	60,500	60,500	53,000	53,000

Notes: This table presents various specifications regressing the log price markup onto its own lags and various controls. All specifications include NAICS2 $\times$ year and firm fixed effects. Column (1) presents an AR(1) specification. Column (2) adds the first lag of the log wage markdown as a control. Column (3) builds on Column (2) by further including the first lag of log sales. Column (4) estimates an AR(2) specification, and Column (5) adds the first lag of log sales to the AR(2) model. Standard errors are reported in parentheses and are two-way clustered by firm and year. All figures are rounded in accordance with U.S. Census disclosure requirements.

Table A14: Persistence of Wage Markdowns (NAICS2 $\times$ Year FE)

	Log Wage Markdown				
	(1)	(2)	(3)	(4)	(5)
Log Wage Markdown (Lag 1)	0.918	0.927	0.890	0.732	0.717
	(0.007)	(0.008)	(0.009)	(0.013)	(0.013)
Log Wage Markdown (Lag 2)				0.210	0.199
				(0.012)	(0.012)
Log Price Markup (Lag 1)		0.088	0.060		0.041
		(0.026)	(0.025)		(0.024)
Log Sale (Lag 1)			0.025		0.021
			(0.003)		(0.003)
NAICS2 FE	No	No	No	No	No
NAICS2 $\times$ Year FE	Yes	Yes	Yes	Yes	Yes
Firm FE	No	No	No	No	No
Observations	60,500	60,500	60,500	53,000	53,000

Notes: This table presents various specifications regressing the log wage markdown onto its own lags and various controls. All specifications include NAICS2 $\times$ year fixed effects. Column (1) presents an AR(1) specification. Column (2) adds the first lag of the log price markup as a control. Column (3) builds on Column (2) by further including the first lag of log sales. Column (4) estimates an AR(2) specification, and Column (5) adds the first lag of log sales to the AR(2) model. Standard errors are reported in parentheses and are two-way clustered by firm and year. All figures are rounded in accordance with U.S. Census disclosure requirements.

Table A15: Persistence of Wage Markdowns (NAICS2 × Year and Firm FEs)

	Log Wage Markdown				
	(1)	(2)	(3)	(4)	(5)
Log Wage Markdown (Lag 1)	0.627	0.621	0.598	0.552	0.529
	(0.018)	(0.019)	(0.020)	(0.017)	(0.019)
Log Wage Markdown (Lag 2)				0.125	0.118
				(0.013)	(0.014)
Log Price Markup (Lag 1)		-0.042	-0.001		-0.014
		(0.039)	(0.039)		(0.038)
Log Sale (Lag 1)			0.070		0.064
			(0.007)		(0.007)
NAICS2 FE	No	No	No	No	No
NAICS2 × Year FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
Observations	60,500	60,500	60,500	53,000	53,000

Notes: This table presents various specifications regressing the log wage markdown onto its own lags and various controls. All specifications include NAICS2 × year and firm fixed effects. Column (1) presents an AR(1) specification. Column (2) adds the first lag of the log price markup as a control. Column (3) builds on Column (2) by further including the first lag of log sales. Column (4) estimates an AR(2) specification, and Column (5) adds the first lag of log sales to the AR(2) model. Standard errors are reported in parentheses and are two-way clustered by firm and year. All figures are rounded in accordance with U.S. Census disclosure requirements.

Table A16: Persistence of Total Wedge (NAICS2 × Year FE)

	Log Total Wedge			
	(1)	(2)	(3)	(4)
Log Total Wedge (Lag 1)	0.923	0.888	0.728	0.708
	(0.008)	(0.009)	(0.012)	(0.012)
Log Total Wedge (Lag 2)			0.220	0.209
			(0.012)	(0.012)
Log Sale (Lag 1)		0.023		0.019
		(0.002)		(0.002)
NAICS2 FE	No	No	No	No
NAICS2 × Year FE	Yes	Yes	Yes	Yes
Firm FE	No	No	No	No
Observations	60,500	60,500	53,000	53,000

Notes: This table presents various specifications regressing the log total wedge onto its own lags and various controls. All specifications include NAICS2 × year fixed effects. Column (1) presents an AR(1) specification. Column (2) adds the first lag of log sales as a control. Column (3) estimates an AR(2) specification, and Column (4) adds the first lag of log sales to the AR(2) model. Standard errors are reported in parentheses and are two-way clustered by firm and year. All figures are rounded in accordance with U.S. Census disclosure requirements.

Table A17: Persistence of Total Wedge (NAICS2 $\times$ Year and Firm FEs)

	Log Total Wedge			
	(1)	(2)	(3)	(4)
Log Total Wedge (Lag 1)	0.627	0.598	0.547	0.525
	(0.018)	(0.019)	(0.017)	(0.017)
Log Total Wedge (Lag 2)			0.135	0.129
			(0.013)	(0.014)
Log Sale (Lag 1)		0.070		0.065
		(0.006)		(0.006)
NAICS2 FE	No	No	No	No
NAICS2 $\times$ Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Observations	60,500	60,500	53,000	53,000

Notes: This table presents various specifications regressing the log total wedge onto its own lags and various controls. All specifications include NAICS2 $\times$ year and firm fixed effects. Column (1) presents an AR(1) specification. Column (2) adds the first lag of log sales as a control. Column (3) estimates an AR(2) specification, and Column (4) adds the first lag of log sales to the AR(2) model. Standard errors are reported in parentheses and are two-way clustered by firm and year. All figures are rounded in accordance with U.S. Census disclosure requirements.

Table A18: Persistence of Productivity (NAICS2 × Year FE)

	Log Labor Productivity			
	(1)	(2)	(3)	(4)
Log Labor Productivity (Lag 1)	0.844	0.832	0.699	0.693
	(0.011)	(0.012)	(0.011)	(0.011)
Log Labor Productivity (Lag 2)			0.194	0.189
			(0.010)	(0.010)
Log Sale (Lag 1)		0.016		0.013
		(0.002)		(0.002)
NAICS2 FE	No	No	No	No
NAICS2 × Year FE	Yes	Yes	Yes	Yes
Firm FE	No	No	No	No
Observations	60,000	60,000	52,500	52,500

Notes: This table presents various specifications regressing the log labor productivity onto its own lags and various controls. All specifications include NAICS2 × year fixed effects. Column (1) presents an AR(1) specification. Column (2) adds the first lag of log sales as a control. Column (3) estimates an AR(2) specification, and Column (4) adds the first lag of log sales to the AR(2) model. Standard errors are reported in parentheses and are two-way clustered by firm and year. All figures are rounded in accordance with U.S. Census disclosure requirements.

Table A19: Consumption-Equivalent Welfare Gains (Shocks $\pm 2\sigma$)

Regime	$\gamma = 1$, -1σ	$\gamma = 1$, +1σ	$\gamma = 2$, -1σ	$\gamma = 2$, +1σ	$\gamma = 4$, -1σ	$\gamma = 4$, +1σ
RE (hetero)	-0.022	0.022	-0.022	0.022	-0.023	0.021
DE (hetero)	-0.094	0.102	-0.094	0.101	-0.095	0.101
DE (θ=mid, hetero)	-0.087	0.093	-0.087	0.093	-0.087	0.092
DE (θ=low, 1 type)	0.000	0.000	0.000	0.000	-0.000	-0.000
DE (θ=mid, 1 type)	0.000	0.000	-0.000	0.000	-0.000	0.000
DE (θ=high, 1 type)	0.000	0.000	-0.000	0.000	-0.000	-0.000

Notes: This table reports consumption-equivalent welfare gains in percent for $\pm 2\sigma$ productivity shocks relative to the case of rational expectations and homogenous firms. The columns correspond to the specified γ and shock value.

B Additional Data Information

In this section, we discuss in more detail how the final dataset and key variables are constructed as well as the proof of Proposition 1 and a discussion of the implementation of the production function estimation. In addition to the LBD and CRSP/Compustat Merged database, we utilize the following publicly available data: BEA NIPA Table 1.1.9 (Bureau of Economic Analysis, 2024), Market Yield on U.S. Treasury Securities at 1-Year Constant Maturity (Board of Governors of the Federal Reserve System, 2024), and the CPI (Bureau of Labor Statistics, 2024). We use Lines 1 and 9 of BEA NIPA Table 1.1.9 to deflate financial statement line items, which correspond to the GDP implicit price deflator and non-residential fixed investment implicit price deflator, respectively. Line 9 is used for physical capital only. The 1-year U.S. Treasury data is used to proxy a one-year nominal risk-free rate and it is deflated by the CPI. We use the annual versions of all these datasets or convert them into annual series through averaging when we obtain them via FRED. The FRED series IDs are provided below.

1. BEA NIPA Table 1.1.9 Line 1 – FRED Series ID: A191RD3A086NBEA
2. BEA NIPA Table 1.1.9 Line 2 – FRED Series ID: A008RD3A086NBEA
3. Market Yield on U.S. Treasury Securities at 1-Year Constant Maturity – FRED Series ID: DGS1
4. CPI – FRED Series ID: CPIAUCSL

B.1 Final Dataset Creation and Key Variable Construction

Here we provide greater detail on how the final dataset is constructed given the raw data files. The final dataset is a merged CRSP/Compustat-LBD firm-level panel. We use the Compustat variable name where applicable. The steps are as follows:

1. Collapse the LBD files from the establishment level to the firm level using the firm-level identifiers
2. Merge the macroeconomic and time series datasets to the Compustat sample
3. Merge the LBD firm-level files to the Compustat sample
4. Replace the any missing values of XRD, XAD, XSGA, and XRENT with 0
5. Drop any observations without up to NAICS4 industry information
6. Generate intangible and physical capital investment and use forward iteration to generate the stock
7. Drop observations in NAICS2 industries 22, 52, 53, 92, and 99
8. Drop observations that are missing or non-positive in SALE, COGS, XSGA, AT, and the created capital stocks variables
9. Create the materials, profits, value-added, and labor productivity variables

10. Trim any observations that are in bottom and top 1% by year of COGS/SALE, XSGA/SALE, and materials/SALE. Also remove observations that are in the top 1% of XRD/SALE by year
11. Winsorize value-added, labor productivity, profit share, and investment rates at the 1% and 99% percentiles by year
12. Keep observations from 1976 onward
13. Run first-stage and second-stage estimation procedures of the production function estimation
14. Merge elasticity estimates to the current merged CRSP/Compustat-LBD panel
15. The firm-level elasticity and cost share estimates are winsorized at the 5% and 95% percentiles since these are used in the ratio estimators to reduce the impact of extreme values
16. Compute the markups and markdowns
17. Compute firm-level TFPR
18. Remove any observations with non-positive markups or markdowns
19. Merge in the replicated panel of Bordalo et al. (2024*a*)

The final datasets are a firm-year panel that ranges from 1980 to 2019 as well as a firm-month panel spanning the same range.

We proceed to discuss how key variables are constructed that are not covered in the main text. It explains the construction of the production function inputs, including physical capital, intangible capital, and materials. Next, it describes the computation of firm value-added, labor productivity, labor share, and profit share. Then we discuss how to compute firm-level output elasticities and TFPR given the production function estimates. Finally, the section explains the methodology for creating the aggregate indices used in the analysis.

Production Function Inputs. The three remaining inputs (physical capital, intangible capital, and materials) are not directly taken from the existing data but rather are constructed using the existing data. Physical capital and intangible capital are both computed using a capitalization approach. Once these are computed, we can then compute materials.

For each firm's physical capital stock, we identify its first observation and initialize its value to the maximum of 0 and the first year's value of PPEGT (gross property, plant, and equipment). Then we take the first difference of PPENT (net property, plant, and equipment) to compute net investment and recursively define the subsequent year's physical capital stock. We deflate physical capital using the nonresidential fixed investment good deflator, which is line 9 of NIPA Table 1.1.9 (FRED ID: A008RD3A086NBEA). Since physical capital is recorded at the end of the year, we use the computed physical capital for year $t-1$ as the year t physical capital input in the estimation. As an example, in our notation, $k_{i,2000}$ refers to the end of 1999 end of year value of physical capital for firm i.

The firm's intangible capital stock is intialized following a similar procedure used by Eisfeldt and Papanikolaou (2013) and Peters and Taylor (2017). We define gross intangible investment as

$$\text{IntInv}_{i,t} = \text{XRD}_{i,t} + \text{XAD}_{i,t} + 0.3 \times (\text{XSGA}_{i,t} - \text{XRD}_{i,t} - \text{XAD}_{i,t}), \tag{A1}$$

where $\text{XRD}_{i,t}$ and $\text{XAD}_{i,t}$ are the research and development and advertising expenses from Compustat, respectively. We compute the sample median growth rate of intangible investment and initialize each firm's intangible capital stock as the maximum of 0 and the first year's intangible divided by the sum of the median growth rate plus depreciation. We set annual depreciation rate to be 30% following Eisfeldt and Papanikolaou (2013). We deflate these values using the GDP deflator which is from line 1 of NIPA Table 1.1.9 (FRED ID: A191RD3A086NBEA). As with physical capital, intangible capital follows the same timing convention.

Finally, recall that we define materials as $\text{COGS}_{i,t}$ plus $\text{XSGA}_{i,t}$ less the wage bill, rent (XRENT in Compustat), and the non-labor portion of the intangible investment. Since intangible investment is constructed using line items that are included in COGS or SGA, we must adjust for this in the definition of materials. Also, since we already remove the total wage bill, we only need to remove the non-labor portion of the intangible investment. Lehr (2023) estimates the labor share of research and development to be 79%. We use this figure and extend it for all intangible investment to compute the non-labor share of intangible investment.

Value-Added, Labor Productivity, Labor Share, and Profit. The firm's value-added is computed following Donangelo et al. (2019) and Seegmiller (2023), and is given below

$$\text{VA}_{i,t} = \text{OIBDP}_{i,t} + \Delta\text{INVFG}_{i,t} + \text{WB}_{i,t}, \tag{A2}$$

where $\text{OIBDP}_{i,t}$ is operating income before depreciation, $\Delta\text{INVFG}_{i,t}$ is the first difference in inventories, and $\text{WB}_{i,t}$ is the wage bill. The changes in inventories are set to 0 when missing. Given Equation (A2) we can compute log labor productivity and the labor share of value-added, which are given by Equations (A3) and (A4), respectively,

$$\ln(\text{Labor Productivity}_{i,t}) = \ln\left(\frac{\text{VA}_{i,t}}{\text{EMPLBD}_{i,t}}\right), \tag{A3}$$

$$\text{LSVA}_{i,t} = \frac{\text{WB}_{i,t}}{\text{VA}_{i,t}}, \tag{A4}$$

where $\text{EMPLBD}_{i,t}$ is the LBD-based employment measure.

The firm's profit in the data is defined as

$$\Pi_{i,t} = \text{OIBDP}_{i,t} - (r_{f,t} + \delta + \text{RP})K_{i,t}, \tag{A5}$$

where $r_{f,t}$ is the real risk-free rate (1-year Treasury rate), $\delta = 0.1$ is the annual depreciation rate of physical capital, RP $= 0.02$ a risk premium, $K_{i,t}$ is the physical capital stock in real terms. Equation (A5) follows the definition and imputations of De Loecker, Eeckhout and Unger (2020) for comparability. The firm's profit share is defined as the ratio of profit to sales. For the rest of the paper and in particular Sections 5 and 6 we use a slightly different definition.

Firm-Level Output Elasticities and TFPR. We estimate translog production functions. Therefore, the output elasticity of input j is given by

$$\theta^j_{i,t} = \beta_j + 2\beta_{j,j} x^j_{i,t} + \sum_{j' \in \mathcal{J} \setminus \{j\}} \beta_{j,j'} x^{j'}_{i,t}. \tag{A6}$$

We assume that $\beta_{j,j'} = \beta_{j',j}$ for all $j, j' \in \mathcal{J}$. With Equation (A6) we can compute the ratio estimators in Equations (3) and (4). Given the estimates of $\boldsymbol{\beta}$, firm-level log inputs $\mathbf{x}_{i,t}$, log output $y_{i,t}$, and first-stage estimate of $\varepsilon_{i,t}$, we can also recover log TFPR $\omega_{i,t}$ from re-arranging Equation (5) evaluated at the estimated parameters to isolate for $\omega_{i,t}$.

Aggregate Price Markup and Wage Markdown Indices. First we compute the NAICS2 industry-level aggregate price markups and wage markdowns following the given weighting methodology. Then we normalize each NAICS2 aggregate series to its 1977 value to 1. From here we aggregate across these by taking a weighted average using the corresponding industry weights (i.e. the sales-weighted index uses the total NAICS2 industry-level sales; the unweighted or simple index uses the observation count as the weight). This indexation approach creates a weighted-average growth rate and bypasses the common multiplicative bias at the NAICS2 level since production functions are estimated at the NAICS2 level. In practice we find that this method produces very similar results to simply normalizing the final aggregate series' 1977 value to 1.

B.2 Proof of Proposition 1

Proof of Proposition 1. Consider the profit maximization problem of a firm that uses $J > 1$ inputs to produce one output, which is given by

$$\begin{gathered} \max_{\mathbf{X}_{i,t} \in \mathbb{R}^J_{++}} P_{i,t}(Y_{i,t}) Y_{i,t} - \sum_{j=1}^{J} W^j_{i,t}(X^j_{i,t}) X^j_{i,t} \\ \text{subject to} \\ Y_{i,t} \leq F(\mathbf{X}_{i,t}; \omega_{i,t}), \end{gathered} \tag{A7}$$

where $\mathbf{X}_{i,t}$ is the vector of inputs, $P_{i,t}(\cdot)$ is the inverse demand function, $Y_{i,t}$ is output, $W^j_{i,t}(X^j_{i,t})$ is the inverse supply function of input j, $F(\cdot;\cdot)$ is the production function that satisfies the usual assumptions, and $\omega_{i,t}$ is the firm's Hicks-neutral log productivity. Finally, let $\mathcal{J}$ denote the set of inputs. Note that we assume that the inverse demand function of a given input j only depends on the quantity of that input.

Then we can take the first-order condition of the problem (A7) with respect to a flexible input. This is given by

$$\begin{aligned}\frac{\partial P_{i,t}(Y_{i,t})Y_{i,t}}{\partial X^f_{i,t}} - \frac{\partial W^f_{i,t}(X^f_{i,t})X^f_{i,t}}{\partial X^f_{i,t}} &= \frac{\partial P_{i,t}(Y_{i,t})}{\partial Y_{i,t}}\frac{\partial Y_{i,t}}{\partial X^f_{i,t}}Y_{i,t} + P_{i,t}(Y_{i,t})\frac{\partial Y_{i,t}}{\partial X^f_{i,t}} - \bar{W}^f_{i,t} \\ &= 0\end{aligned}$$

We can rearrange this expression as follows

$$\left[\frac{\partial P_{i,t}(Y_{i,t})}{\partial Y_{i,t}}\frac{Y_{i,t}}{P_{i,t}(Y_{i,t})} + 1\right]^{-1} = \frac{\partial Y_{i,t}}{\partial X^f_{i,t}}\frac{P_{i,t}(Y_{i,t})}{W^f_{i,t}},$$

Notice that the first two terms of the product on the right-hand side is the markup that follows the definition from (1). The expression for marginal costs follows from the dual problem (cost minimization). The right-hand side can be further rearranged as follows

$$\begin{aligned}\frac{\partial Y_{i,t}}{\partial X^f_{i,t}}\frac{P_{i,t}(Y_{i,t})}{W^f_{i,t}} &= \frac{\partial Y_{i,t}}{\partial X^f_{i,t}}\frac{X^f_{i,t}}{Y_{i,t}}\frac{P_{i,t}(Y_{i,t})Y_{i,t}}{W^f_{i,t}X^f_{i,t}} \\ &= \frac{\theta^f_{i,t}}{\alpha^f_{i,t}} = \mu_{i,t},\end{aligned}$$

which is the relationship in (3).

With the price markup and the corresponding ratio estimator recovered we can proceed to the input markdown and its ratio estimator. The first-order condition with respect to a monopsonistic input is given by

$$\begin{aligned}\frac{\partial P_{i,t}(Y_{i,t})Y_{i,t}}{\partial X^j_{i,t}} - \frac{\partial W^j_{i,t}(X^j_{i,t})X^j_{i,t}}{\partial X^j_{i,t}} &= \frac{\partial P_{i,t}(Y_{i,t})}{\partial Y_{i,t}}\frac{\partial Y_{i,t}}{\partial X^j_{i,t}}Y_{i,t} + P_{i,t}(Y_{i,t})\frac{\partial Y_{i,t}}{\partial X^j_{i,t}} - \frac{\partial W^j_{i,t}(X^j_{i,t})}{\partial X^j_{i,t}}X^j_{i,t} - W^j_{i,t}(X^j_{i,t}) \\ &= 0.\end{aligned}$$

This expression can be rearranged as follows

$$\begin{aligned}\left[\frac{\partial P_{i,t}(Y_{i,t})}{\partial Y_{i,t}}\frac{Y_{i,t}}{P_{i,t}(Y_{i,t})} + 1\right]P_{i,t}(Y_{i,t})\frac{\partial Y_{i,t}}{\partial X^j_{i,t}} &= \frac{\partial W^j_{i,t}(X^j_{i,t})}{\partial X^j_{i,t}}X^j_{i,t} + W^j_{i,t}(X^j_{i,t}) \\ \frac{\mu^{-1}_{i,t}}{W^j_{i,t}(X^j_{i,t})}P_{i,t}(Y_{i,t})\frac{\partial Y_{i,t}}{\partial X^j_{i,t}} &= \left[\frac{\partial W^j_{i,t}(X^j_{i,t})}{\partial X^j_{i,t}}\frac{X^j_{i,t}}{W^j_{i,t}(X^j_{i,t})} + 1\right].\end{aligned}$$

The left-hand side of the first line is the marginal revenue product with respect to $X^j_{i,t}$; note that with product market power the MRPL accounts for the change in price when selling one more unit. Then on the second line, the left-hand side is the marginal revenue product of input j to the price of input j, which is the definition of a markdown following (2). This expression can be further changed to yield

$$\begin{aligned}\frac{\mu_{i,t}^{-1}}{W^j_{i,t}(X^j_{i,t})}P_{i,t}(Y_{i,t})\frac{\partial Y_{i,t}}{\partial X^j_{i,t}} &= \frac{P_{i,t}(Y_{i,t})Y_{i,t}}{W^j_{i,t}(X^j_{i,t})X^j_{i,t}}\frac{\partial Y_{i,t}}{\partial X^j_{i,t}}\frac{X^j_{i,t}}{Y_{i,t}}\mu_{i,t}^{-1} \\ &= \frac{\theta^j_{i,t}}{\alpha^j_{i,t}}\mu_{i,t}^{-1} = \nu^j_{i,t},\end{aligned}$$

thus we recover the relation in (4). Thus, we have proved Proposition 1. □

B.3 Detailed Steps on Production Function Estimation Procedure

This section discusses the details on the implementation of the GMM procedure to estimate the second stage. First, we discuss the moment conditions and specifically how we incorporate the constant returns to scale (CRS) restriction. Then we discuss the procedure we utilize to address a common numerical optimization issue.

We implement the CRS restriction similar to the method described by Yeh, Macaluso and Hershbein (2022) and Ren and Zhang (2025). However, first we discuss how the other moment conditions are constructed. Following the notation from Section 2.2 (in logs), we have materials $m_{i,t}$, physical capital $k_{i,t}$, labor $l_{i,t}$, and intangible capital $n_{i,t}$. We write the vector of the production function's parameters as

$$\boldsymbol{\beta} = \left(\beta_l, \beta_m, \beta_k, \beta_n, \beta_{l,l}, \beta_{l,m}, \beta_{l,k}, \beta_{l,n}, \beta_{m,m}, \beta_{m,k}, \beta_{m,n}, \beta_{k,k}, \beta_{k,n}, \beta_{n,n}\right)^\intercal. \tag{A8}$$

Next, we define the vector of instruments and then we construct the first set of moment conditions. The vector of instruments is given by

$$\tilde{\mathbf{z}}_{i,t} = \left(\tilde{\mathbf{z}}^\intercal_{1,i,t}, \tilde{\mathbf{z}}^\intercal_{2,i,t}\right)^\intercal, \tag{A9}$$

where

$$\begin{aligned}\tilde{\mathbf{z}}_{1,i,t} &= \left(l_{i,t-1}, m_{i,t-1}, k_{i,t}, n_{i,t}\right)^\intercal, &(\text{A10})\\ \tilde{\mathbf{z}}_{2,i,t} &= \Big(l^2_{i,t-1}, l_{i,t-1}m_{i,t-1}, l_{i,t-1}k_{i,t}, l_{i,t-1}n_{i,t}, \\ &\qquad m^2_{i,t-1}, m_{i,t-1}k_{i,t}, m_{i,t-1}n_{i,t}, k^2_{i,t}, k_{i,t}n_{i,t}, n^2_{i,t}\Big)^\intercal. &(\text{A11})\end{aligned}$$

Notice that in Equations (A9) to (A11) we impose a timing assumption for identification. Following the standard assumptions, we assume that the firm observes their idiosyncratic productivity shock $\xi_{i,t}$ and then make their input decisions. However, since the current stock of physical and intangible capital are assumed to be chosen in the period before, their current value is orthogonal to $\xi_{i,t}$. Similarly, since materials and labor are chosen contemporaneously, we use their lagged values as instruments as those are uncorrelated with $\xi_{i,t}$. Now we move on to the CRS restriction's moment condition.

Let $\Sigma_{i,t}(\cdot)$ denote a firm's returns to scale that takes the production function parameters as an input. The returns to scale are given by

$$\Sigma_{i,t}(\boldsymbol{\beta}) = \sum_{j \in \mathcal{J}} \frac{\partial f(\mathbf{x}_{i,t}; \boldsymbol{\beta})}{\partial x_{i,t}^{j}}. \tag{A12}$$

CRS implies that Equation (A12) is set to 1. Therefore, the moment conditions that includes the CRS restriction are given by

$$\mathbb{E}\begin{bmatrix} \xi_{i,t}(\hat{\boldsymbol{\beta}}) \cdot \tilde{\mathbf{z}}_{i,t} \\ \Sigma_{i,t}(\hat{\boldsymbol{\beta}}) - 1 \end{bmatrix} = \mathbf{0}. \tag{A13}$$

Since we are using a translog production function, we can write the CRS restriction as a linear operator. Let $\tilde{\boldsymbol{\beta}} = (1, \boldsymbol{\beta}^{\intercal})^{\intercal}$ and $\tilde{\mathbf{x}}_{i,t} = \left(1, \mathbf{x}_{i,t}^{\intercal}\right)^{\intercal}$, then we have

$$\Sigma_{i,t}(\boldsymbol{\beta}) - 1 = (R\tilde{\boldsymbol{\beta}})^{\intercal}\tilde{\mathbf{x}}_{i,t}, \tag{A14}$$

where

$$R = \begin{pmatrix} -1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 2 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 2 \end{pmatrix}.$$

When we compute the norm of the moment conditions, we use the identity matrix as the weight matrix for our baseline specification with the CRS restriction. To estimate a version without this restriction, we replace the last diagonal entry of the weight matrix from 1 to 0, which corresponds to Equation (A14).

Once we construct these objects, we simply need to run the optimization procedure to recover the estimates. However, optimization routines do not necessarily find the global minimum for a given starting point and boundary restriction. Indeed, if we utilize different starting points we generally obtain different solutions. Therefore, we run the estimation 1,000 times with different (randomly selected) starting points within the boundaries. We collect the estimates and the value of the objective function for each iteration across all industries. We select the estimates associated with the minimum evaluated objective function value across all iterations for each industry to be the "global" solution. While this approach still does not guarantee a global solution, we find that the approach is robust to picking 100, 200, and 500 iterations. The parameter bounds are set to be $[0,1]$ for the first-order terms and $[-0.05, 0.05]$ for the second-order and interaction terms. The estimated parameters for all industries have all their respective estimates within these bounds.

C Proofs and Derivations for Section 3

C.1 Proof of Proposition 2

Proof of Proposition 2. We start with Equation (15) and substitute in $\gamma = \rho_\zeta / \rho_g$, which yields

$$\tilde{\mathbb{E}}_t \left[g_{i,t+s}\right] = \mathbb{E}_t \left[g_{i,t+s}\right] + \theta \sum_{n\geq 1} \left(\frac{\rho_\zeta}{\rho_g}\right)^{n-1} \left(\mathbb{E}_{t+1-n}[g_{i,t+s}] - \mathbb{E}_{t-n}[g_{i,t+s}]\right).$$

From Equation (11), we have that $\mathbb{E}_{t-k}\left[g_{i,t+s}\right] = \rho_g^{s-1}\,\mathbb{E}_{t-k}\left[g_{i,t+1}\right]$, which we substitute into the expression above to obtain

$$\tilde{\mathbb{E}}_t \left[g_{i,t+s}\right] = \mathbb{E}_t \left[g_{i,t+s}\right] + \theta \rho_g^{s-1} \sum_{n\geq 1} \left(\frac{\rho_\zeta}{\rho_g}\right)^{n-1} \left(\mathbb{E}_{t+1-n}[g_{i,t+1}] - \mathbb{E}_{t-n}[g_{i,t+1}]\right).$$

Using Equation (11) again, we have that $\mathbb{E}_{t+1-n}[g_{i,t+1}] = \rho_g^{n-1}\,\mathbb{E}_{t+1-n}\left[g_{i,t+2-n}\right]$, which we substitute in to get

$$\tilde{\mathbb{E}}_t \left[g_{i,t+s}\right] = \mathbb{E}_t \left[g_{i,t+s}\right] + \theta \rho_g^{s-1} \sum_{n\geq 1} \rho_\zeta^{n-1} \left(\mathbb{E}_{t+1-n}[g_{i,t+2-n}] - \mathbb{E}_{t-n}[g_{i,t+2-n}]\right).$$

The rational revision is given by

$$\begin{aligned}
\mathbb{E}_{t+1-n}[g_{i,t+2-n}] - \mathbb{E}_{t-n}[g_{i,t+2-n}] &= \rho_g g_{i,t+1-n} + \kappa_{i,t+1-n} - \rho_g^2 g_{i,t-n} - \rho_g \kappa_{i,t-n}, \\
&= \rho_g^2 g_{i,t-n} + \rho_g \tau_{i,t+1-n} + \rho_g \kappa_{i,t-n} + \kappa_{i,t+1-n} - \rho_g^2 g_{i,t-n} - \rho_g \kappa_{i,t-n}, \\
&= \rho_g \tau_{i,t+1-n} + \kappa_{i,t+1-n}.
\end{aligned}$$

Thus, we obtain

$$\tilde{\mathbb{E}}_t \left[g_{i,t+s}\right] = \mathbb{E}_t \left[g_{i,t+s}\right] + \theta \rho_g^{s-1} \sum_{n\geq 1} \rho_\zeta^{n-1} \left(\rho_g \tau_{i,t+1-n} + \kappa_{i,t+1-n}\right).$$

Then using the definition of $u_{i,t+1-n}$ in Equation (18), we obtain

$$\tilde{\mathbb{E}}_t \left[g_{i,t+s}\right] = \mathbb{E}_t \left[g_{i,t+s}\right] + \rho_g^{s-1} \sum_{n\geq 1} \rho_\zeta^{n-1} u_{i,t+1-n}.$$

Iterating backward Equation (17), we get that

$$\zeta_{i,t} = \sum_{n\geq 1} \rho_\zeta^{n-1} u_{i,t+1-n},$$

thus we obtain

$$\tilde{\mathbb{E}}_t \left[g_{i,t+s}\right] = \mathbb{E}_t \left[g_{i,t+s}\right] + \rho_g^{s-1}\zeta_{i,t},$$

which is Equation (16). Thus, we have proved Proposition 2. □

C.2 Closed-Form Expressions of Regression Coefficients

Here we provide the closed-form expressions of the regression coefficients of interest in Equations (19) and (20) as well as discuss how these form the moments for the estimation procedure in Section 4.2. See Bordalo et al. (2024*a*) for a more detailed discussion. To derive these expressions, however, it is helpful to derive the expressions for the forecast error in earnings growth and the revision to earnings growth expectations. The expected forecast error follows Equation (16)

$$\mathbb{E}_t \left[g_{i,t+s}\right] - \tilde{\mathbb{E}}_t \left[g_{i,t+s}\right] = -\rho_g^{s-1}\zeta_{i,t}.$$

The revision of earnings growth expectations is given by

$$\begin{aligned}
\tilde{\mathbb{E}}_t \left[g_{i,t+s}\right] - \tilde{\mathbb{E}}_{t-1} \left[g_{i,t+s}\right] &= \mathbb{E}_t \left[g_{i,t+s}\right] - \mathbb{E}_{t-1} \left[g_{i,t+s}\right] + \rho_g^{s-1}\zeta_{i,t} - \rho_g^{s}\zeta_{i,t-1}, \\
&= \rho_g^{s-1}(\rho_g g_{i,t} + \kappa_{i,t}) - \rho_g^{s}(\rho_g g_{i,t-1} + \kappa_{i,t-1}) + \rho_g^{s-1}\zeta_{i,t} - \rho_g^{s}\zeta_{i,t-1}, \\
&= \rho_g^{s-1}\left(\rho_g \tau_{i,t} + \kappa_{i,t}\right) + \rho_g^{s-1}\zeta_{i,t} - \rho_g^{s}\zeta_{i,t-1}, \\
&= \rho_g^{s-1}\left(\rho_g \tau_{i,t} + \kappa_{i,t}\right) + \rho_g^{s-1}\left(\rho_\zeta \zeta_{i,t-1} + u_{i,t} - \rho_g \zeta_{i,t-1}\right), \\
&= \rho_g^{s-1}\left[(1+\theta)\left(\rho_g \tau_{i,t} + \kappa_{i,t}\right) + (\rho_\zeta - \rho_g)\zeta_{i,t-1}\right].
\end{aligned}$$

The derivation for this expression follows from the definitions from Equations (11), (17), and (18).

These expressions are then used to compute the following

$$\gamma_{11} \equiv \text{Var}\left[\tilde{\mathbb{E}}_t\left[g_{i,t+s}\right] - \tilde{\mathbb{E}}_{t-1}\left[g_{i,t+s}\right]\right] = \rho_g^{2(s-1)}\sigma_u^2\left[(1+\theta)^2 + (\rho_g - \rho_\zeta)\frac{\theta^2}{1-\rho_\zeta^2}\right],$$

$$\gamma_{12} \equiv \text{Cov}\left[\tilde{\mathbb{E}}_t\left[g_{t+s}\right] - \tilde{\mathbb{E}}_{t-1}\left[g_{t+s}\right], \tilde{\mathbb{E}}_{t-1}\left[g_{t+s}\right]\right] = -\rho_g^{2s-1}(\rho_g - \rho_\zeta)\theta\sigma_u^2\left[\frac{1}{1-\rho_g\rho_\zeta} + \frac{\theta}{1-\rho_\zeta^2}\right],$$

$$\gamma_{22} \equiv \text{Var}\left[\tilde{\mathbb{E}}_{t-1}\left[g_{t+s}\right]\right] = \rho_g^{2s}\sigma_u^2\left[\frac{1}{1-\rho_g^2} + \frac{2\theta}{1-\rho_g\rho_\zeta} + \frac{\theta^2}{1-\rho_\zeta^2}\right],$$

$$\gamma_{1Y} \equiv \text{Cov}\left[\tilde{\mathbb{E}}_t\left[g_{t+s}\right] - \tilde{\mathbb{E}}_{t-1}\left[g_{t+s}\right], g_{t+s} - \tilde{\mathbb{E}}_t\left[g_{t+s}\right]\right] = -\rho_g^{2(s-1)}\theta\sigma_u^2\left[1 + \theta\left(1 - \rho_\zeta\frac{\rho_g - \rho_\zeta}{1-\rho_\zeta^2}\right)\right],$$

$$\gamma_{2Y} \equiv \text{Cov}\left[\tilde{\mathbb{E}}_{t-1}\left[g_{t+s}\right], g_{t+s} - \tilde{\mathbb{E}}_t\left[g_{t+s}\right]\right] = -\rho_g^{2s-1}\rho_\zeta\theta\sigma_u^2\left[1 + \frac{\theta}{1-\rho_\zeta^2}\right],$$

where $\sigma_u^2 \equiv \rho_g^2\sigma_\tau^2 + \sigma_\kappa^2$. With these expressions we can yield expressions for the regression coefficients of interest in Equations (19) and (20). The regression coefficients are given by

$$\beta_1^{\text{FE}} = \frac{\gamma_{22}\gamma_{1Y} - \gamma_{12}\gamma_{2Y}}{\gamma_{11}\gamma_{22} - \gamma_{12}^2},$$
$$\beta_2^{\text{FE}} = \frac{-\gamma_{12}\gamma_{1Y} + \gamma_{11}\gamma_{2Y}}{\gamma_{11}\gamma_{22} - \gamma_{12}^2},$$
$$\beta_1^{\text{Rev}} = \frac{\gamma_{12}}{\gamma_{22}}.$$

We can also recover an expression for the total variance of LTG revisions with γ_{11}, which we use to estimate σ_u in Section 4.2.

D Implementation of Diagnostic Expectations Parameter Estimation

This section provides additional detail on the parameter estimation procedure described in Section 4.2. Since the model moments are available in closed form, the implementation of the minimization problem in (26) is relatively straightforward once the model moments, empirical moments, and the weighting matrix W have been computed.

Because the objective function is highly nonlinear and the solution can depend on the initial conditions, we solve the problem 250 times using randomly drawn starting values from the interior of the parameter bounds. For all firm types, we impose the following parameter bounds:

$$\begin{aligned}
\theta &\in [0, 10], \\
\rho_\zeta &\in [0, \hat{\rho}_g + \varepsilon], \\
\rho_g &\in [0, 1], \\
\sigma_u &\in [0, 0.05],
\end{aligned}$$

where $\hat{\rho}_g$ is the estimated AR(1) persistence parameter of the realized LTG, and $\varepsilon = 10^{-6}$ is a small positive constant used to ensure smooth behavior of the optimizer near the boundary. We retain the solution that yields the smallest value of the objective function across the 250 initializations. In addition, we scale the objective function by a factor of 10^6 to improve numerical precision and obtain a more accurate solution.

Once we obtain the estimates $\hat{\Theta}$, we compute standard errors using the Delta method, following Hansen (1982), Hansen and Singleton (1982), and Newey and McFadden (1994). The estimated variance-covariance matrix of the parameters is given by

$$\widehat{\mathrm{Var}}[\hat{\Theta}] = \left(D(\hat{\Theta})^\intercal \hat{W}^{-1} D(\hat{\Theta}) \right)^{-1}, \tag{A15}$$

where $\hat{W}$ is the estimated weighting matrix, and $D(\hat{\Theta})$ denotes the Jacobian of the model moments with respect to the parameters, evaluated at $\hat{\Theta}$. We compute this Jacobian using automatic differentiation. The standard errors are given by the square roots of the diagonal elements of $\widehat{\mathrm{Var}}[\hat{\Theta}]$.

To estimate W, we implement a clustered bootstrap procedure as in Horowitz (2001) and Cameron and Miller (2015). We define firm as the clustering unit. Given N unique firms, we draw N clusters with replacement for a total of $B = 1{,}000$ iterations. From the resulting B empirical moment vectors, we compute the sample variance-covariance matrix and use it as our estimate of W.

E Computational Method

The model presented in Section 5 is solved using standard numerical techniques, with policy function iteration employed to efficiently recover both the capital policy function and the value function. The problem features two exogenous state variables, productivity $(\omega_{i,t})$ and belief distortions $(\zeta_{i,t})$, which evolve according to the stochastic processes described in Equations (17) and (28), as well as one endogenous state variable, capital $(k_{i,t})$. We discretize the exogenous state spaces using the method of Rouwenhorst (1995), generating a 35-point grid for productivity and a 7-point grid for belief distortions. Capital follows a 150-point grid whose limits depend on the model being solved.

The first-order conditions of the problem (36) are given by

$$[l_{i,t}]: \quad p_{i,t}\frac{\partial y_{i,t}}{\partial l_{i,t}} + \frac{\partial p_{i,t}}{\partial y_{i,t}}\frac{\partial y_{i,t}}{\partial l_{i,t}} y_{i,t} - w_{i,t} - \frac{\partial w_{i,t}}{\partial l_{i,t}} l_{i,t} = 0 \tag{A16}$$

$$[k_{i,t+1}]: \quad -1 - \psi\left(\frac{k_{i,t+1}}{k_{i,t}} - 1\right) + \beta_i \tilde{\mathbb{E}}_t\left[\frac{\partial V_{i,t+1}}{\partial k_{i,t+1}}\right] = 0, \tag{A17}$$

and the envelope condition is given by

$$\frac{\partial V_{i,t}}{\partial k_{i,t}} = (1-\delta) - \frac{\psi}{2}\left[1 - \left(\frac{k_{i,t+1}}{k_{i,t}}\right)^2\right] + \frac{p_{i,t}}{\mu_{j(i)}}\frac{\partial y_{i,t}}{\partial k_{i,t}}. \tag{A18}$$

Since the labor policy function is static, given the state variables we can immediately recover the current labor choice. Equation (A16) yields the pricing rule

$$p_{i,t} = \mu_{j(i)} \nu_{j(i)} w_{i,t} \left(\frac{\partial y_{i,t}}{\partial l_{i,t}}\right)^{-1}. \tag{A19}$$

We can substitute in the production function, residual product demand, and residual product supply (Equations (27), (29), and (30)) into Equation (A19) to obtain the labor policy function

$$l_{i,t}(\omega_{i,t}, \zeta_{i,t}, k_{i,t}) = \left[\mu_{j(i)} \nu_{j(i)} \frac{\omega_{i,t}^{\frac{1}{\varepsilon_i}} k_{i,t}^{\frac{\alpha}{\varepsilon_{j(i)}} - \alpha}}{(1-\alpha)\omega_{i,t}}\right]^{-\frac{\varepsilon_{j(i)}\eta_{j(i)}}{\alpha\varepsilon_{j(i)}\eta_{j(i)} - \alpha\eta_{j(i)} + \varepsilon_{j(i)} + \eta_{j(i)}}}. \tag{A20}$$

We can use Equations (A17) and (A18) as well as substitute in Equations (33) and (34) to recover the future capital policy function. This is given by

$$k_{i,t+1} = k_{i,t} + k_{i,t}\psi^{-1}\left[-1 + \beta\tilde{\mathbb{E}}_t\left[(1-\delta) - \frac{\psi}{2}\left[1 - \left(\frac{k_{i,t+2}}{k_{i,t+1}}\right)^2\right] + \frac{p_{i,t+1}}{\mu_{j(i)}}\frac{\partial y_{i,t+1}}{\partial k_{i,t+1}}\right]\right]. \tag{A21}$$

Then we substitute Equation (A20) into Equation (A21) during the iteration process to numerically recover Equation (A21).