

# Expectations Formation with Fat-Tailed Processes: Evidence and Theory\*

Tim de Silva<sup>†</sup> Eugene Larsen-Hallock<sup>‡</sup> Adam Rej<sup>‡</sup> David Thesmar<sup>§</sup>

February 6, 2026

[Latest Version](#) 

## Abstract

Using a large sample of forecasts of 22 firm-level variables, we document three facts: (i) the relationship between forecast revisions and future forecast errors is strongly non-linear, (ii) the distributions of the underlying processes have fat tails, and (iii) extreme realizations tend to mean-revert. Next, we formally show that these three facts are consistent with a model in which the underlying process is non-Gaussian, but forecasters fail to recognize this. Finally, we provide additional evidence in support of our theory by showing that it can explain evidence from an online forecasting experiment, that the non-linear relationship between errors and revisions is also present in macroeconomic forecasts, and that it provides an explanation for the presence of non-linearity in the momentum of stock returns.

---

\*We are grateful to Xavier Gabaix for very helpful feedback and Mauro Cazzaniga for outstanding research assistance. We also thank Nick Barberis, Eben Lazarus, Stefan Nagel, Larry Schmidt, Kelly Shue, and seminar participants at CFM, Columbia, MIT, Yale SOM, NYU Stern, University of Toronto, Stanford GSB, Wharton, Chicago Booth, NBER Behavioral Finance, and NBER Forecasting & Empirical Methods for helpful comments and suggestions.

<sup>†</sup>Stanford University, Graduate School of Business, tdesilva@stanford.edu.

<sup>‡</sup>Capital Fund Management

<sup>§</sup>MIT Sloan School of Management, CEPR, and NBER, thesmar@mit.edu.

The recent literature on expectations formation has accumulated evidence that subjective forecast errors are predictable. To explain this predictability, theories of expectations formation have been proposed with a variety of ingredients such as information frictions (Coibion and Gorodnichenko, 2015), sluggish updating (Bouchaud et al., 2019; Ma et al., 2024), representativeness bias (Bordalo et al., 2019, 2020a), and imperfect memory (Afrouzi et al., 2023). A common feature of these theories is that they assume that the underlying data-generating process (DGP) is Gaussian. However, in many economic applications, tails are “fat”, meaning that extreme outcomes occur much more frequently than in a Gaussian distribution. For example, fat tails are present in the cross-sectional distributions of many economic variables, including population size, income, wealth, and firm size (Gabaix, 2009). More importantly, fat tails are also common in dynamic processes about which agents need to continuously update their expectations, such as stock returns (Fama, 1965), firm growth (Stanley et al., 1996), and income growth (Guvenen et al., 2014). Therefore, understanding how expectations are formed in the presence of fat tails is important for studying many dynamic decisions, such as corporate investment, household consumption-savings, or portfolio choice.

In this paper, we study expectations formation in the presence of fat tails, both empirically and theoretically. We begin with an empirical analysis of a large dataset of subjective forecasts made by professional analysts about over twenty firm outcomes, including sales growth, earnings per share, dividend payouts, and profitability ratios. The large size of the dataset is useful because it allows us to study the tails of the distribution. Using these data, we document three robust empirical facts: (i) the relationship between forecast revisions and future forecast errors—the variables used in Coibion and Gorodnichenko (2015) regressions—is strongly non-linear for all of the firm-level outcomes; (ii) the distributions of the underlying processes have fat tails; and (iii) the conditional expectation of future realizations is non-linear in current realizations, with persistence in the bulk and mean reversion in the tails. Next, we build an expectations formation model that connects these facts. The key ingredients in our model are that the underlying process is non-Gaussian, but forecasters fail to recognize this and incorrectly fit a linear model to the data. Finally, we provide additional evidence in support of our theory by showing that it can explain expectations formation in an online forecasting experiment, that the non-linear relationship between errors and revisions is also present in macroeconomic forecasts, and that it provides an explanation for the presence of non-linearity in the momentum of stock returns. Overall, our analysis suggests that forecasters partially ignore the presence of fat tails and rely too much on linear models, which is an insight that could be leveraged in models of household and firm behavior with non-Gaussian DGPs (Guvenen et al., 2021; Boar et al., 2025).

The data that we use in our empirical analysis are a large panel of analyst forecasts and realizations of 22 firm-level variables from IBES that cover the most important financial statement items, such

as revenue, earnings, cash flow, and debt levels. The main benefit of these data is that the sample is large, with over 3 million observations covering more than 20 years. Using these data, the first and most important fact that we document is that the relationship between current forecast revisions and future forecast errors is strongly non-linear. In some papers, revisions linearly and *positively* predict forecast errors, a feature commonly interpreted as evidence of underreaction (Coibion and Gorodnichenko, 2015; Bouchaud et al., 2019). In others, revisions linearly and *negatively* predict forecast errors, which is evidence of overreaction (Bordalo et al., 2019, 2020a). In our panel, we show that both coexist. For large revisions, there is a *negative* relationship between revisions and future errors, indicating overreaction to news. For more moderate revisions, the relationship between revisions and future errors becomes more positive, most of the time indicating underreaction to news (or a much milder amount of overreaction). We show that this non-linear relation between errors and revisions is present for all 22 firm-level variables in our sample.

This non-linear relationship between errors and revisions is robust to a battery of robustness checks that we conduct. For example, we show that this non-linearity is present in both individual *and* consensus forecasts and, therefore, is not driven by forecast aggregation (Bordalo et al., 2020a). Additionally, we show that the non-linearity is driven by cross-sectional rather than aggregate variation and does not vary with analysts' forecasting experience. These findings suggest that it is unlikely to be driven by a slow convergence of learning that can occur with one short time series (Bianchi et al., 2022; Farmer et al., 2024).

Our second empirical fact is that the realizations of all 22 firm-level variables in our sample have fat tails. To detect the presence of fat tails, we measure the kurtosis of each variable and the extent to which it is well-approximated by a power law in the tails of the distribution (Gabaix, 2009). We find that most variables in our sample have kurtosis above six and power law tail parameters below three (with  $R^2$ 's of log-log regressions above 0.99), both of which indicate the presence of fat tails relative to a normal distribution. These fat tails do not arise from heterogeneity in volatility across firms (Wyart and Bouchaud, 2003) and do not arise from time-varying aggregate volatility.

Our third and final fact is that the conditional expectation of future realizations given the current realization is non-linear in a way that is reminiscent of the non-linearity in our first fact. In particular, we find that the relationship between current and one-year-ahead realizations is increasing and linear in the bulk of the distribution for most of the variables in our sample. However, in the tails of the distribution, this persistence is less strong for all variables and often changes sign, indicating that extreme values tend to mean-revert. As described below, this non-linearity helps inform the exact way the DGP in our model deviates from that of standard Gaussian models.

In the next part of the paper, we develop a model of expectations formation that is designed to connect the above three facts, while also being simple and portable. In the model, the DGP for the forecasting variable contains a persistent and transitory component, which is common in models of dividend growth (Bansal and Yaron, 2004; Lettau and Wachter, 2007) and income dynamics (Guvenen et al., 2014). Our two key assumptions are that the transitory component follows a power law distribution, a tractable way of characterizing processes with fat tails (Gabaix, 2009), and that the persistent component has a thinner tail than the transitory component.<sup>1</sup>

We show that this model of the DGP is consistent with our second and third facts. The model generates fat tails by assumption, consistent with our second fact. To show that it generates the non-linear conditional expectation of future realizations conditional on current realizations—our third fact—we leverage a result from empirical Bayes theory known as Tweedie’s formula (Efron, 2012). Although this expectation cannot be expressed in closed-form due to the non-normality, this result allows us to characterize it as a function of the observable density of realizations. We show that this result implies that the conditional expectation is locally linear when the density is locally Gaussian, as in the bulk of the distribution of most variables. However, when the density is locally non-Gaussian, as in the tails, the conditional expectation is no longer linear and is asymptotically decreasing in the current realization. Intuitively, very large realizations are likely due to the transitory component and, hence, are not likely to persist.

Given this DGP, consistent with our second and third facts, we can also explain the non-linear relationship between forecast errors and revisions—our first fact—with a single assumption: agents construct their forecasts ignoring fat tails. Formally, we assume that agents form forecasts according to the Kalman filter, which would be the rational expectation in our model if the transitory and persistent components were both normally distributed, and is equivalent to assuming that forecasters use the linear forecast with the lowest possible mean-squared error. Using results from the theory of regularly varying functions (Bingham et al., 1987; Geluk and de Vries, 2006), we prove that this assumption is enough to generate overreaction in the tails and underreaction in the bulk. This is because large revisions are driven by large realizations that asymptotically come from the transitory component of the DGP. While a rational forecaster would recognize that these extreme shocks are unlikely to persist, our agents who ignore fat tails do not and, therefore, overreact. However, because our agents are unbiased unconditionally, overreaction in the tails has to be compensated by underreaction in the bulk, consistent with our first fact.

While we do not provide a deep psychological microfoundation for why agents ignore fat tails,

---

<sup>1</sup>The latter assumption is consistent with our third fact: if the persistent component had the thicker tail, we would not observe mean reversion in the tails of the distribution. Additionally, our results generalize to the case in which the persistent component is regularly varying (Bingham et al., 1987).

we view it as a form of bounded rationality in which forecasters use a misspecified model, as in [Fuster et al. \(2010\)](#) and [Gabaix \(2019\)](#). To illustrate this point, we show that our assumption that forecasters use the Kalman filter is analogous to assuming that forecasters are insensitive to the strength of the signal that they observe. The latter assumption is consistent with laboratory and field evidence in [Augenblick et al. \(2024\)](#) and [Ba et al. \(2024\)](#), and suggests that insensitivity to signal strength in our time series setting is a possible reason why forecasters ignore the presence of fat tails.

We conclude by providing three pieces of evidence from different domains that support our theory that forecasters do not fully recognize the presence of fat tails. First, we run an online forecasting experiment similar to [Afrouzi et al. \(2023\)](#), but where the underlying process has fat tails. The key benefit of this experiment is that it allows us to test our theory of expectations formation directly by experimentally varying the features of the data-generating process. When we run the experiment using our estimated DGP, we find that the relationship between errors and revisions is non-linear, consistent with our first fact. In contrast, when forecasters forecast a similar process with no fat tails (as in [Afrouzi et al. 2023](#)), we find no evidence of a non-linear relationship between errors and revisions. These findings provide direct evidence that the non-linear relationship between errors and revisions in the data is driven by the fat tails of the DGP, which is the key result of our theory.

Second, we show that the non-linear relationship between forecast errors and revisions—our first fact—is also present in forecasts of macroeconomic outcomes from the [Survey of Professional Forecasters \(1968–2025\)](#). Of the 14 variables that are available, we find that the relationship between individual-level forecast errors and revisions is non-linear, with a more negative slope in the tails of the distribution for all but one variable. For most variables, the slope is positive in the bulk of the distribution and negative in the tails of the distribution, similar to our findings from firm-level forecasts. These findings suggest that the prevalence of more overreaction in the tails of the distribution is more general than our sample of firm-level forecasts.

Finally, we show that our model makes predictions for return momentum that are supported by the data. We translate expectations of sales growth, one of the variables in our firm-level sample, into returns by applying the [Campbell \(1991\)](#) return decomposition with a constant subjective discount rate (as in [Bouchaud et al. 2019](#) and [Nagel and Xu 2022](#)). Our model predicts that the relationship between past and future returns should be positive in the bulk of the distribution, where underreaction to news is dominant. In contrast, it predicts mean-reversion of returns in the tails, where agents fail to recognize that the extreme shocks are not persistent. We find some support for this prediction in the universe of smaller stocks: for these stocks, momentum tends to mean-revert for extreme losers and winners.

**Related literature.** Economists have long recognized the benefits of using survey data on expectations to study economic behavior (Manski, 2004), but recent literature has shifted the emphasis to the study of forecast error predictability (e.g., Coibion and Gorodnichenko 2015). This paper contributes to this branch of literature in two ways. First, we contribute to the empirical literature that has documented under- and overreaction across many forecasting variables and horizons. Broadly speaking, this literature tends to find evidence of underreaction when looking at shorter-term or consensus forecasts (Coibion and Gorodnichenko, 2015; Bouchaud et al., 2019), and overreaction when looking at longer-term or individual forecasts (Bordalo et al., 2019, 2020a; Wang, 2021). Relative to this empirical literature, our contributions are to provide field evidence of both under- and overreaction within the *same* forecasting variable and horizon, and lab evidence that the degree of overreaction can vary within a sample depending on the tail parameter in the DGP.

Second, we contribute to the literature that proposes models of belief formation that can generate under- and overreaction, including constant-gain learning (Nagel and Xu, 2022), selective recall (Bordalo et al., 2020b), and biased perceptions of autocorrelation (Wang, 2021). Most of this literature works with models in which data-generating processes are Gaussian, implying that conditional expectations are linear. In contrast, our empirical analyses highlight how this relationship can be quite non-linear, and we provide a theory that links this non-linearity to the presence of under- and overreaction within a given process. Our focus on non-Gaussian dynamics is similar to Kozlowski et al. (2020), but we focus on forecasters ignoring these dynamics rather than learning.<sup>2</sup> With forecasts of a single time series, rational learning can converge relatively slowly (Farmer et al., 2024), generating an in-sample relationship between errors and revisions even under full rationality (Singleton, 2021; Bianchi et al., 2022). In contrast, one benefit of having the large cross-section in our data is that, under fairly mild assumptions, it averages out the portion of the error-revision predictability that is driven by a short time series. Our assumption that forecasters (partially) ignore non-Gaussian dynamics is inspired by the literature on bounded rationality, which argues that economic agents use simplified models to minimize computation costs or because of cognitive constraints (Fuster et al., 2010; Gabaix, 2019; Molavi et al., 2024). However, it is possible that this assumption could be microfounded via Bayesian learning about the tail parameter of the process.

Three closely related papers are Kwon and Tang (2025), Augenblick et al. (2024), and Graeber et al. (2025). Kwon and Tang (2025) also provide a model of belief formation with non-Gaussian dynamics. In their model, news events belong to categories with different power law distributions, and forecasters have diagnostic expectations, causing them to overreact to news from categories with fatter tails and underreact to news from categories with thinner tails. This prediction is similar

---

<sup>2</sup>See Dew-Becker et al. (2024) for a characterization of Bayesian learning with arbitrary non-Gaussian dynamics.

to that of our model. An important difference is that this provides a theory of why over- and underreaction would vary depending on the category from which a realization is drawn, while our model provides a theory of why over- and underreaction would vary even *within* a category. [Augenblick et al. \(2024\)](#) propose a model in which forecasters incorrectly perceive signal quality and shrink it to a default, which leads to overreaction to weak signals and underreaction to strong ones. As we discuss in the paper, insensitivity to signal strength is a possible microfoundation for ignoring fat tails. Finally, [Graeber et al. \(2025\)](#) analyze the S-shaped relationship between returns and earnings surprises, which loosely map into revisions and forecast errors. In their theory, the strong sensitivity between returns and surprises around zero comes from overreaction that occurs at a category boundary, while the lower sensitivity away from zero comes from dampening within a category due to noisy perceptions. Our main fact—underreaction in the bulk and overreaction in the tails—points to a different mechanism in our setting.

Through our model of the DGP with fat tails, we connect the expectations formation literature with the literature on power laws. The omnipresence of power laws ([Gabaix, 2009](#)) suggests that the misperception of fat tails that we document is likely important for understanding subjective forecasts in other settings. The facts that we document about the data-generating process of firm-level realizations are consistent with the literature on firm dynamics. For example, we find that sales growth (rather than its level) has fat tails, which is a violation of Gibrat’s law that was first documented by [Stanley et al. \(1996\)](#) and recently emphasized by [Boar et al. \(2025\)](#). Our third fact is consistent with [Jaimovich et al. \(2025\)](#), who find that revenue is more persistent in the bulk of the distribution than in the tails and exhibits fat tails.

**Outline.** Section 1 describes our data on forecasts of 22 firm-level variables. Section 2 documents our three main facts. Section 3 describes the model that we build to connect and explain these facts. Section 4 provides additional evidence in support of our theory from an online forecasting experiment, macroeconomic forecasts, and data on stock returns.

# 1 Firm-Level Forecast Data

## 1.1 Data Source and Sample

Our analysis primarily relies on an annual panel between 2000 and 2023 of forecasts issued by over 38,000 security analysts who follow around 36,000 U.S. and international firms. These forecasts cover 22 different annual firm-level variables at different horizons. We obtain these analyst-level forecasts from LSEG IBES Unadjusted Detail files (henceforth “Detail”). In robustness analysis, we

also study consensus analyst forecasts, which come from the LSEG IBES Summary Statistics files (henceforth “Summary”). Both of these files contain current estimates as of the third Wednesday of each month. We extract the forecasts issued in the third month of each fiscal year  $t + 1$ , which ensures that the results for fiscal year  $t$  have been released. We focus on forecasts for two different horizons: fiscal years  $t + 1$  and  $t + 2$ . Along with one and two-year forecasts, LSEG IBES also provides data on the realizations of the forecasted variables, which the provider ensures are measured consistently with the forecasts made by analysts.

## 1.2 Definition of Forecasting Variables

We study forecasts and realizations of all the variables available in the Summary and Detail files. The list of these variables is provided in [Table 1](#). These variables fall into one of three categories: accounting items in local currency (e.g., Revenue), items in local currency on a “per-share” basis (e.g., Earnings Per Share), and accounting ratios in percentage points (e.g., Return on Assets). To avoid imposing unnecessary sample filters, we choose to include all 22 variables regardless of their data coverage, which we discuss below.

The only transformation we apply to the variables of interest is done to ensure they are stationary. Out of the 22 variables, three are accounting ratios that are already in percentage points, which are presumably already stationary and therefore not transformed. The remaining 19 variables are accounting items in local currency or on a per-share basis. To make them stationary, we convert them from levels to growth rates, as in [Bordalo et al. \(2020a\)](#). More specifically, let  $X_{it+h}$  denote the realization of one of these variables for firm  $i$  at  $t + h$ . We transform this variable into a log growth rate according to  $g_{it+h} = \log X_{it+h} - \log X_{it+h-1}$ . To simplify exposition, for the remaining three accounting ratios, we define  $g_{it+h}$  as the untransformed realization for firm  $i$  at  $t + h$ . Moving forward, we refer directly to  $g_{it+h}$  as our realization of interest.

## 1.3 Definitions of Forecast Errors and Revisions

For each firm  $i$  and each year  $t$ , we denote  $F_t^j X_{it+h}$  as the forecast of  $X_{it+h}$  made in year  $t$  by analyst  $j$  that comes from the Detail files. We denote the consensus forecast by  $F_t X_{it+h}$ , which is defined as the mean forecast across analysts from the Summary files. For both the analyst-level and consensus, we use forecasts extracted three months after the end of fiscal year  $t$ , which ensures that the results for fiscal year  $t$  have been released by the time these forecasts are made. Similarly, we denote  $F_{t-1}^j X_{it+h}$  and  $F_{t-1} X_{it+h}$  as the analyst-level and consensus forecasts extracted three months after the end of fiscal year  $t - 1$ .

For forecasts of the three untransformed accounting ratios, we can work with our raw forecasts directly, so  $F_t^j g_{it+h} = F_t^j X_{it+h}$ . For the remaining 19 variables, we define forecasts of growth rates as  $F_t^j g_{it+h} = \log F_t^j X_{it+h} - \log F_t^j X_{it+h-1}$ , with the convention that  $F_t^j X_{it} = X_{it}$ . For consensus forecasts, variables are constructed analogously without the  $j$  superscript. Note that the way we translate from forecasts of levels to growth rates implicitly ignores a Jensen’s inequality term because  $\log F_t^j X_{it+h} \neq F_t^j \log X_{it+h}$ . We ignore this adjustment for simplicity, but we do not think it materially affects our analysis for two reasons. First, a constant Jensen’s term would simply shift the unconditional level of forecasts, while our analysis focuses on the conditional properties of forecast errors. Second, we show that our results are robust to using percent (instead of log) growth.

Having defined  $F_t^j g_{it+h}$  for each of the 22 variables, we then construct forecast errors and revisions following the literature on expectations formation (Coibion and Gorodnichenko, 2015; Bouchaud et al., 2019):

$$ERR_{it+1}^j = g_{it+1} - F_t^j g_{it+1}, \quad REV_{it}^j = F_t^j g_{it+1} - F_{t-1}^j g_{it+1}.$$

For consensus forecasts, we again use the analogous definitions without the  $j$  superscript.

## 1.4 Summary Statistics

**Table 1** shows summary statistics on the realizations, analyst-level forecast errors, and analyst-level forecast revisions for each of our 22 variables. **Table IA.1** shows the same table for consensus-level forecasts. For the sample of each variable in the table, we remove outliers by dropping the top and bottom one percent of each distribution. The number of observations reported in the table is the number of firm-year observations for which we can compute all of the variables of interest. The size of our sample varies a lot across variables, but is in general quite large. Earnings Per Share (EPS), which is studied by most of the literature on analyst forecasts, is one of the most covered variables. However, many other variables have good coverage, such as Net Sales, which is even more covered than EPS with over 407k observations. ROE and Net Income are nearly as covered as EPS. These four variables, as well as EBITDA, EBIT, and Pre-Tax Profit, have more than 200k observations. 13 variables have more than 100k observations, and 19 more than 50k. Two variables, EPS Before Goodwill and Funds From Operations Per Share, have fewer than 10k observations. In the following section, we report analysis results for all variables, ranking them by their number of observations.

In the main text of this paper, we focus on analyst-level forecasts to avoid the possibility that our results are affected by aggregating across forecasters (Bordalo et al., 2020a), but also to be

Table 1: Summary Statistics: Analyst-Level Forecasts

Variable	Transformation (1)	Realizations		Errors		Revisions		# of Analysts (8)	# of Obs. (9)
		Mean (2)	SD (3)	Mean (4)	SD (5)	Mean (6)	SD (7)		
Net Sales	Growth	0.06	0.16	-0.00	0.10	-0.02	0.13	6.33	399809
Earnings Per Share	Growth	0.06	0.44	-0.04	0.31	-0.07	0.34	6.35	360011
Return On Equity	Level	0.13	0.13	-0.01	0.07	-0.01	0.07	5.17	345082
Net Income	Growth	0.06	0.43	-0.03	0.31	-0.06	0.34	6.24	325230
EBITDA	Growth	0.06	0.29	-0.02	0.20	-0.04	0.23	5.82	280111
Pre-tax Profit	Growth	0.06	0.41	-0.04	0.30	-0.06	0.32	5.71	269883
EBIT	Growth	0.06	0.37	-0.02	0.27	-0.07	0.31	6.32	231455
GAAP EPS	Growth	0.06	0.55	-0.08	0.40	-0.03	0.41	4.85	171523
Book Value Per Share	Growth	0.05	0.17	-0.01	0.16	-0.02	0.13	3.82	170561
Return On Assets	Level	0.05	0.06	-0.01	0.03	-0.01	0.03	3.47	157063
Gross Margin	Level	0.41	0.21	-0.00	0.04	-0.01	0.03	4.56	150220
Net Asset Value	Growth	0.06	0.18	-0.00	0.17	-0.02	0.14	3.12	119885
Cash Flow Per Share	Growth	0.04	0.48	-0.03	0.41	-0.07	0.48	3.78	112323
Net Debt	Growth	0.04	0.41	0.07	0.41	0.10	0.45	3.69	90126
Enterprise Value	Growth	0.04	0.32	0.04	0.29	0.02	0.26	2.98	88791
Dividend Per Share	Growth	0.06	0.37	-0.02	0.31	-0.04	0.32	2.82	63840
Capital Expenditures	Growth	0.04	0.41	-0.06	0.45	0.10	0.52	2.89	62099
Operating Profit	Growth	0.06	0.40	-0.05	0.28	-0.05	0.30	3.57	57756
EBITDA Per Share	Growth	0.05	0.31	-0.02	0.21	-0.05	0.25	1.44	16824
Cash Earnings Per Share	Growth	0.05	0.31	-0.02	0.27	-0.06	0.28	1.85	11578
Funds From Operations Per Share	Growth	0.03	0.14	-0.00	0.08	-0.04	0.11	4.74	9279
EPS Before Goodwill	Growth	0.09	0.36	-0.01	0.27	-0.06	0.30	2.19	2558

*Notes:* This table reports the mean, standard deviation, and number of observations for each one of the 22 variables forecasted by analysts in our sample. The sample includes all observations for which an analyst's forecast error and revision are available for a given variable. Variables are ranked by decreasing number of observations. 19 variables are transformed into log growth as explained in the text; 3 are left untransformed. Column (8) reports the average number of analysts forecasts available per firm-year, and column (9) reports the total number of observations.

consistent with our experimental evidence in Section 4.1. However, after presenting our main results in Section 2.1, we discuss the results from using consensus forecasts, which are presented in the Appendix. Additionally, we also conduct a series of robustness checks that are also presented in the Appendix, including showing that our results do not depend on our particular log growth transformation.

## 2 Three Facts

In this section, we document the three key facts that motivate the model of expectations formation that we develop in Section 3.

## 2.1 Fact #1: Non-Linear Relationship Between Forecast Errors and Revisions

Our first and central empirical fact is based on a regression of forecast errors on forecast revisions. This regression was introduced by Coibion and Gorodnichenko (2015) (CG) and takes the following form, for forecaster  $j$ , firm  $i$  and year  $t$ :

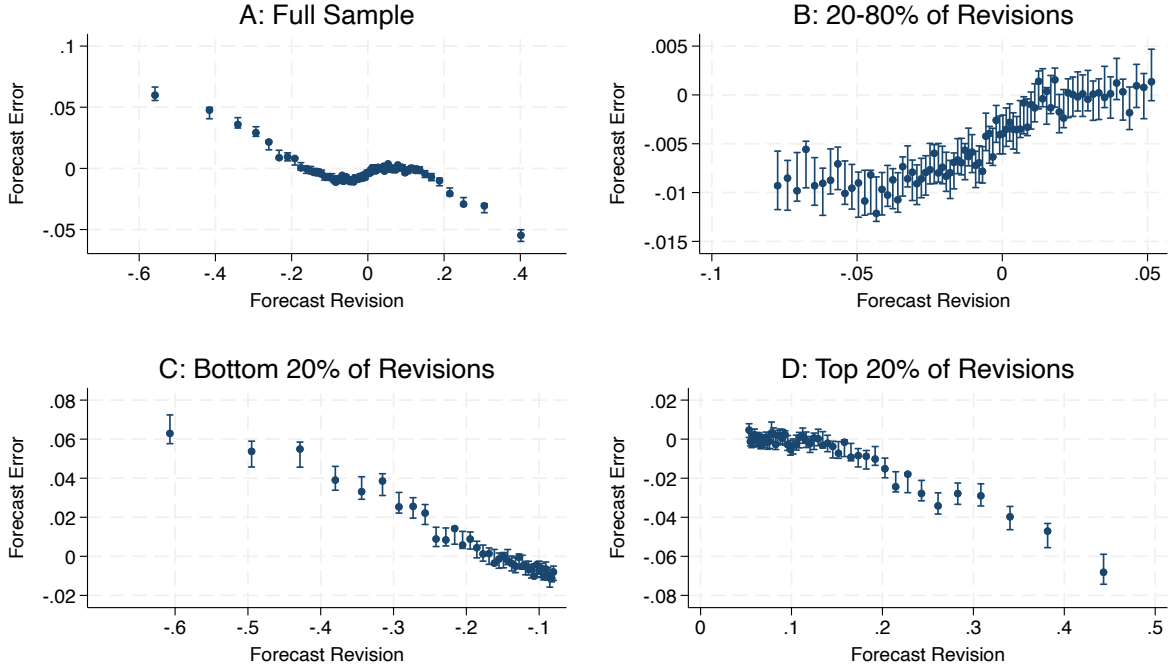
$$ERR_{it+1}^j = \alpha + \beta REV_{it}^j + e_{it+1}^j \quad (1)$$

where the forecast error  $ERR_{it+1}^j$  and revision  $REV_{it}^j$  for each variable are defined in Section 1. The slope coefficient is typically used to distinguish between different models of expectations formation, requiring only panel data on expectations and realizations. The rational expectations hypothesis predicts  $\beta = 0$ , since forecast errors cannot be predicted using data in the forecaster’s information set. When this regression is run using consensus forecasts,  $\beta > 0$  is typically interpreted as evidence of information frictions, as in models of sticky or noisy information (Coibion and Gorodnichenko, 2015). With individual forecasts,  $\beta > 0$  is interpreted as non-Bayesian underreaction (Bouchaud et al., 2019), while  $\beta < 0$  is interpreted as overreaction (Bordalo et al., 2020a). A key feature of prior literature is that it restricts analysis to linear functional forms, as in equation (1). While this is a natural starting point, especially in settings with small sample sizes, this section uses our large sample of firm-level expectations to provide evidence that this relationship is non-linear.

**Graphical evidence.** Figure 1 provides a non-parametric look at the relationship between forecast errors and revisions for Net Sales, the best-populated variable with more than 400k observations. The figure shows that the relationship between forecast errors  $ERR_{it+1}^j$  and forecast revisions  $REV_{it}^j$  is highly non-linear. Panel A shows a binned scatterplot on the full sample with 100 bins; Panels B, C, and D show the same plot separately for observations in the 20-80th, 0-20th, and 80-100th percentiles of revisions. For revisions in the bulk of the distribution, Panel B shows that errors are increasing in revisions, consistent with forecasters underreacting to news that causes moderately-sized revisions in forecasts. This finding is not novel to our paper: it is consistent with the underreaction in analysts’ EPS forecasts for US firms (Bouchaud et al., 2019), as well as managers’ revenue forecasts in the US and Italy (Ma et al., 2024).

Panels C and D of Figure 1 show that for large (positive or negative) revisions, the positive relationship between forecast errors and revisions reverses and becomes negative: large positive (negative) forecast revisions are predictive of negative (positive) future forecast errors. Unlike the relationship in the bulk of the distribution, which is consistent with underreaction, this finding is consistent with *overreaction* to news. In other words, forecasters appear to overreact in response to

Figure 1: Non-Linear Error-Revision Relationship: Analyst-Level Net Sales Forecasts



*Notes:* This figure shows binned scatterplots of analyst-level forecast errors on forecast revisions for Net Sales. Panel A shows the results for the entire sample; Panel B restricts the sample to the 20-80th percentiles of revisions; Panel C restricts the sample to below the 20th percentile of revisions; Panel D restricts the sample to above the 80th percentile of revisions. Vertical bars represent 95% confidence intervals, assuming the relationship is piecewise linear and continuous (option “ci(1 1)” in Stata command “binsreg”).

news that generates large revisions, while underreacting to more moderate news.

**Regression evidence.** While the results in Figure 1 are specific to Net Sales, we now show that this same non-linearity is present for almost all of the other variables in Table 1. To do so, we estimate the following variant of the regression in (1) separately for each variable with OLS:

$$ERR_{it+1}^j = \alpha + \beta REV_{it}^j + \gamma REV_{it}^j \times (P_{ijt}^+ + P_{ijt}^-) + \delta^+ P_{ijt}^+ + \delta^- P_{ijt}^- + e_{it+1}^j \quad (2)$$

where  $P_{ijt}^+$  (resp.  $P_{ijt}^-$ ) is equal to 1 if  $REV_{it}^j$  belongs to the top (resp. bottom) 20% of all revisions for this variable. In this specification,  $\beta$  is the error-revision coefficient for revisions between the 20th and 80th percentiles of the distribution of revisions, while  $\gamma$  captures the change in this slope in the tails of the distribution (i.e., below the 20th and above the 80th percentiles). Equation (2) crucially allows intercepts to differ in the two tails, as is the case in Figure 1, but estimates a pooled slope in the tails of the distribution. We choose to estimate a pooled slope to improve

Table 2: Non-Linear Error-Revision Relationship: Analyst-Level Forecasts

Error <sub>it+1</sub>	Revision <sub>it</sub>		Revision <sub>it</sub> × Top/Bottom 20% of Revision <sub>it</sub>		Top 20% of Revision <sub>it</sub>		Bottom 20% of Revision <sub>it</sub>		# of Obs.
Variable	$\hat{\beta}$ (1)	$t$ -stat (2)	$\hat{\gamma}$ (3)	$t$ -stat (4)	$\hat{\delta}^+$ (5)	$t$ -stat (6)	$\hat{\delta}^-$ (7)	$t$ -stat (8)	
Net Sales	0.11	6.31	-0.26	-14.99	0.02	5.86	-0.02	-5.09	399809
Earnings Per Share	0.08	2.97	-0.26	-9.44	0.02	1.48	-0.07	-5.09	360011
Return On Equity	0.23	9.75	-0.27	-11.70	0.00	2.56	-0.02	-5.99	345082
Net Income	0.04	1.53	-0.25	-11.08	0.02	1.63	-0.06	-4.91	325230
EBITDA	-0.04	-1.51	-0.15	-5.28	0.01	1.72	-0.02	-2.00	280111
Pre-tax Profit	0.01	0.47	-0.22	-9.05	0.03	2.38	-0.05	-3.07	269883
EBIT	-0.05	-1.60	-0.22	-8.12	0.03	2.94	-0.03	-2.06	231455
GAAP EPS	-0.02	-0.73	-0.14	-4.99	0.01	0.31	-0.04	-1.91	171523
Book Value Per Share	-0.06	-1.52	-0.42	-8.53	0.01	2.44	-0.04	-6.90	170561
Return On Assets	0.41	10.35	-0.45	-10.15	-0.01	-5.32	-0.01	-8.49	157063
Gross Margin	0.09	3.90	-0.15	-6.23	0.00	1.70	-0.01	-6.62	150220
Net Asset Value	0.04	0.99	-0.56	-11.26	0.03	7.75	-0.03	-4.61	119885
Cash Flow Per Share	-0.28	-15.86	-0.06	-3.27	0.05	4.05	-0.00	-0.36	112323
Net Debt	-0.17	-8.31	-0.12	-5.46	0.12	10.06	0.12	9.14	90126
Enterprise Value	-0.20	-5.82	-0.26	-8.25	0.11	11.91	-0.01	-0.82	88791
Dividend Per Share	-0.12	-4.54	-0.17	-6.15	-0.01	-1.13	-0.03	-3.08	63840
Capital Expenditures	-0.36	-21.93	-0.16	-8.53	0.06	5.20	-0.05	-4.32	62099
Operating Profit	0.12	3.07	-0.35	-7.83	0.04	3.40	-0.06	-4.54	57756
EBITDA Per Share	-0.09	-2.84	-0.13	-3.14	0.02	2.24	-0.03	-1.85	16824
Cash Earnings Per Share	-0.19	-4.95	-0.20	-3.83	0.03	1.98	-0.02	-1.01	11578
Funds From Operations Per Share	-0.01	-0.25	-0.09	-1.72	0.01	1.61	-0.02	-2.00	9279
EPS Before Goodwill	-0.13	-2.29	-0.11	-1.36	0.04	2.82	-0.09	-3.21	2558

Notes: This table reports the OLS estimates of equation (2), for each one of the 22 variables. We report the estimates of  $\beta$  (column 1),  $\gamma$  (column 3),  $\delta^+$  (column 5), and  $\delta^-$  (column 7), along with the corresponding  $t$ -statistics with standard errors clustered by firm and time. Variables are ranked by number of observations.

power and because our focus is on the non-linearity of the error-revision relationship rather than any asymmetries.<sup>3</sup>

Table 2 shows the estimated coefficients in (2) for all of the 22 variables in our sample. Starting with Net Sales in the first line, we find  $\hat{\beta} = 0.1$  ( $t = 6.3$ ) and  $\hat{\gamma} = -0.26$  ( $t = -15$ ). This is consistent with Figure 1: in the bulk of the distribution, errors are positively related to revisions with slope  $\hat{\beta} = 0.1$ , consistent with underreaction, while in the tails of the distribution, errors are negatively related to revisions with slope  $\hat{\beta} + \hat{\gamma} = 0.1 - 0.26 = -0.16$ , consistent with overreaction.

<sup>3</sup>Table IA.2 shows the results of a specification that allows for asymmetric slopes. While the strong non-linearity shown here is present for all variables, some variables exhibit asymmetry.

The remaining rows in [Table 2](#) show that the non-linear relationship between forecast errors and revisions that we find for sales is present for all 22 variables in our data. For 20 out of 22 variables, we find that  $\hat{\gamma}$  is negative and significant at the 1% level.  $\hat{\beta}$  varies more across variables, with some variables having underreaction in the bulk of the distribution (e.g., Net Sales, Earnings Per Share, Return On Equity), while others have overreaction (e.g., Enterprise Value, Dividend Per Share). Overall, across all variables available in our data, these results show that the error-revision relationship is highly non-linear.

**Robustness.** We now discuss a battery of robustness checks that we conduct to ensure that the non-linear relationship between forecast errors and revisions is robust. We discuss each one of these robustness checks in a different paragraph below. For each robustness check, we produce an analogous version of [Table 2](#) that is referenced and shown in the Appendix.

First, we check that the relationship continues to hold for consensus forecasts. A potential issue with analyst-level forecasts is that they are sensitive to analyst-level noise, which introduces a negative bias in the estimate of  $\beta$  in (1) ([de Silva and Thesmar, 2024](#)). Therefore, our results could be driven by the fact that expectation noise is larger for bigger revisions, inducing a stronger bias in the tails. However, [Table IA.3](#) shows that this is not the case. In particular, when working with consensus forecasts, we find that  $\hat{\gamma}$  is statistically significantly negative for all but 5 variables, with magnitudes that are similar to the analyst-level results.

Second, we show that our results are not sensitive to our log growth transformation. One concern with this transformation is that we ignore a Jensen’s inequality term when constructing  $F_t^j g_{it+h}$  (the log of a forecast is not necessarily the forecast of the log). To address this concern, we redefine growth as  $g_{it+1} = \frac{X_{it+1} - X_{it}}{X_{it}}$  and forecasted growth at horizon  $h$  as  $F_t^j g_{it+h} = \frac{F_t^j X_{it+h} - F_t^j X_{it+h-1}}{F_t^j X_{it+h-1}}$  (for the 19 variables that are transformed). The benefit of this approach is that our construction of  $F_t^j g_{it+1}$  is no longer affected by Jensen’s inequality.<sup>4</sup> [Table IA.4](#) shows that, when using percent growth,  $\hat{\gamma}$  is negative for all 19 transformed variables, and significantly negative for 16 variables with similar magnitudes.

Third, we argue later in the paper that the driver of the non-linear error-revision relationship is the presence of fat tails. One concern with this argument is that fat tails may instead reflect a mixture of thin-tailed processes with heterogeneous variances ([Wyart and Bouchaud, 2003](#)). To address this concern, we restrict our sample to firms for which we have at least 10 observations and compute the

---

<sup>4</sup>We still need to ignore a Jensen’s inequality term when computing  $F_{t-1}^j g_{it+1}$ , which is needed to compute forecast revisions. However, the fact that our results are quantitatively unchanged when using percent growth to remove this effect on  $F_t^j g_{it+1}$  makes us less concerned that this is an important issue for our results.

standard deviation of the realization of each variable. We then normalize realizations and forecasts by this firm-specific standard deviation. [Table IA.5](#) shows that, even after adjusting for firm-specific volatility, our results are quantitatively similar with  $\hat{\gamma}$  estimated to be significantly negative for all but one variable. In [Table IA.6](#), we perform a similar analysis but instead normalize by the annual cross-sectional standard deviation of each realization. Again, we find quantitatively similar results, suggesting our findings are not driven by time-varying aggregate volatility.

Finally, part of the literature on expectation formation has emphasized that rational learning can occur slowly ([Farmer et al., 2024](#)), which can generate a correlation between errors and revisions in short time series that may not be present out of sample ([Bianchi et al., 2022](#)). Our use of cross-sectional variation alleviates this concern to the extent that our results are driven by firm-specific variation, but not if they are driven by aggregate variation. [Table IA.7](#) shows that our results are indeed driven by the former: they are unchanged after absorbing aggregate shocks by removing the cross-sectional mean of errors and revisions in each year. Additionally, we directly test whether learning drives the non-linear error revision relationship by splitting our sample into four quartiles of analyst experience, measured by the number of firms followed by each analyst up to the current year. [Table IA.8](#) shows that the non-linear pattern is quantitatively similar in all quartiles. This finding suggests that learning is not a first-order driver of our results, which are likely driven by our use of extensive cross-sectional variation.

## 2.2 Fact #2: Fat Tails in Distribution of Realizations

The strong evidence in [Section 2.1](#) that the relationship between forecast errors and revisions is non-linear points towards a different treatment of large versus small shocks. This evidence is hard to square with established models of expectation formation, which feature linear DGPs (e.g., AR1s) and linear expectations models. Next, we document two additional facts about the realizations of the variables being forecasted that guide our theory by informing our model of the DGP.

Our second fact is that the distributions of realizations have thick tails. It is well known that many financial and economic variables have tails that are fatter than a normal distribution ([Gabaix, 2009](#)). For example, the distribution of firm sizes follows a power law distribution with a tail coefficient of one ([Axtell, 2001](#)), which is known as Zipf’s law, and the distribution of growth rates in COMPUSTAT follows a Laplace distribution, which has fatter tails than a normal distribution ([Stanley et al., 1996](#); [Bottazzi and Secchi, 2006](#)). To measure the extent to which our variables are Gaussian, we use two metrics on the set of realizations at the firm-year level. The first is kurtosis, which would be three if the variables are normal. The second is the slope coefficient in a log rank-log size regression, which is a standard way of estimating the tail parameter of a power law

distribution that would be infinity for a normal distribution (Gabaix, 2009). Specifically, for each variable  $g_{it}$ , we regress the log rank of  $|g_{it}|$  on  $\log |g_{it}|$  using observations between the 90th and 99th percentiles of  $|g_{it}|$ . We then report the negative of the slope coefficient, which corresponds to our estimate of the tail parameters.

Table 3 shows that the realizations of all 22 forecasting variables in our sample have tails that are much fatter than those of a Gaussian distribution. Apart from Return on Assets and Gross Margin, all variables have a tail coefficient that is below three. These imply very fat tails: a tail parameter of two or lower implies that the distribution has an infinite variance. Apart from Capital Expenditures, Enterprise Value, and Gross Margin, all variables have a kurtosis of five or more. These estimates are very far from normal, and formal tests based on kurtosis reject the null of a normal distribution at conventional significance levels.

**Robustness.** In the Appendix, we show that the presence of fat tails documented in Table 3 is robust to the alternative variable definitions discussed in the robustness checks in Section 2.1. First, Table IA.9 shows that fat tails are still present after dividing by firm-specific estimates of volatility, which suggests that they do not arise from a mixture of thin-tailed processes with heterogeneous variances (Wyart and Bouchaud, 2003). Second, Table IA.10 shows that adjusting for time-varying aggregate volatility does not change the results. Finally, Table IA.11 shows that tail coefficients remain low when using percent instead of log growth.

### 2.3 Fact #3: Non-Linear Conditional Expectation

The third and final fact that informs the specification of the data-generating process in our model is that the conditional expectation of the current realization is a non-linear function of the lagged realization. Figure 2 illustrates this point graphically for Net Sales, the best-populated variable. In Panel A, we plot the relationship between Net Sales at  $t + 1$  and Net Sales at  $t$  using a binned scatter plot. As is evident from the figure, the conditional expectation of Net Sales conditional on its lag exhibits significant non-linearity reminiscent of the non-linearity in the relationship between forecast errors and revisions in Figure 1. Panels B, C, and D of Figure 2 zoom in on the three parts of the distribution of past growth. These panels show that Net Sales is approximately increasing and linear in the bulk of the distribution, while the slope is severely attenuated, or even reversed, in the tails of the distribution. This evidence is consistent with Jaimovich et al. (2025), who find that sales growth is more persistent in the bulk of the distribution than in the tails and exhibits fat tails.

While the evidence in Figure 2 is specific to Net Sales, we now show that this non-linear

Table 3: Fat Tails in the Distribution of Realizations

Variable	Transformation	Kurtosis	Log-Log Regression		# of Obs.
			Tail Parameter	$R^2$	
	(1)	(2)	(3)	(4)	(5)
Net Sales	Growth	5.86	2.67	0.99	140482
Earnings Per Share	Growth	6.41	2.39	0.99	129735
Return On Equity	Level	9.49	2.61	1.00	127268
Net Income	Growth	6.38	2.40	0.99	113862
Pre-tax Profit	Growth	6.25	2.45	0.99	102386
EBITDA	Growth	6.16	2.52	0.99	99731
Book Value Per Share	Growth	8.60	2.09	0.99	89884
EBIT	Growth	6.44	2.42	0.99	82100
Return On Assets	Level	6.46	3.40	0.99	79479
GAAP EPS	Growth	5.91	2.56	0.99	74210
Net Asset Value	Growth	7.92	2.16	0.99	71494
Gross Margin	Level	2.44	8.00	0.98	64483
Cash Flow Per Share	Growth	5.56	2.63	0.99	62446
Dividend Per Share	Growth	7.07	2.45	0.99	49402
Net Debt	Growth	6.57	2.33	0.99	48338
Enterprise Value	Growth	4.58	3.00	0.99	48140
Capital Expenditures	Growth	4.98	2.82	0.99	44623
Operating Profit	Growth	6.23	2.54	0.99	33033
EBITDA Per Share	Growth	6.26	2.41	0.99	15343
Cash Earnings Per Share	Growth	6.04	2.40	0.99	10013
Funds From Operations Per Share	Growth	8.39	2.38	0.99	3399
EPS Before Goodwill	Growth	6.25	2.87	0.99	2413
Normal Distribution	.	3	$\infty$	.	.

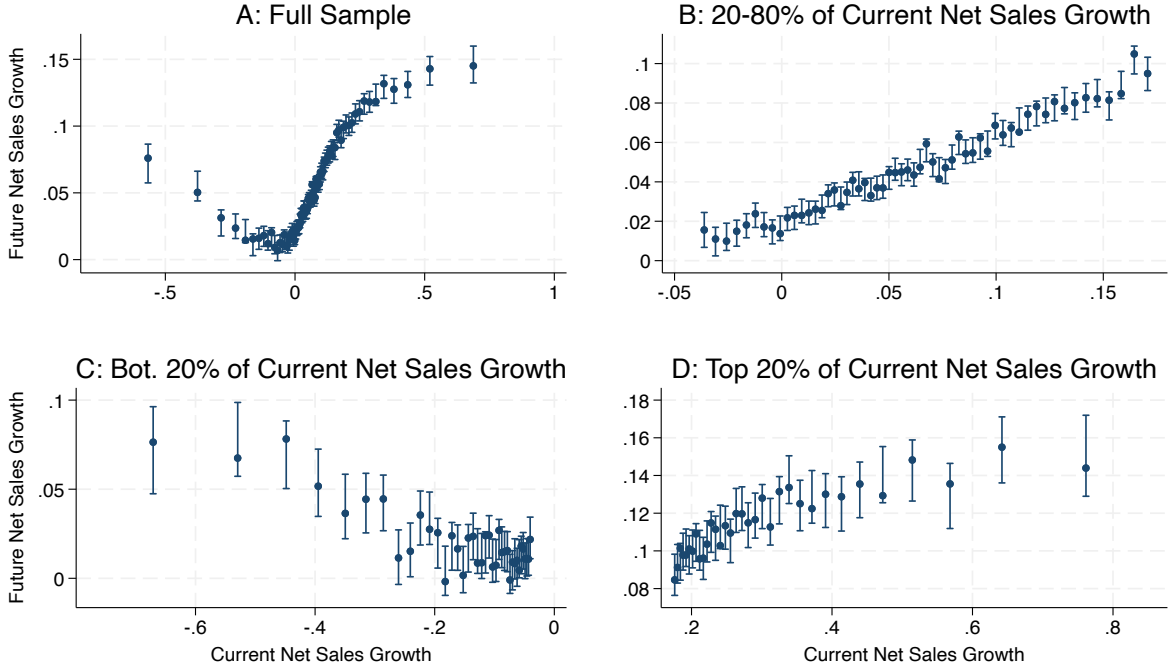
*Notes:* This table reports the kurtosis and tail parameters of the 22 variables in our data. Tail parameters, reported in column (3), are obtained through the following procedure. For each variable  $g_{it}$ , we regress the log rank of  $|g_{it}|$  on  $\log |g_{it}|$ , restricting ourselves to the centiles of  $|g_{it}|$  between 90 and 99. We report the  $R^2$  of these regressions in column (4). We restrict ourselves to realizations available in LSEG IBES. Variables are ranked by number of observations in the final column.

relationship holds for all other 22 variables. For each variable,  $g_{it}$ , we estimate the following regression with OLS:

$$g_{it+1} = \alpha_g + \beta_g g_{it} + \gamma_g g_{it} \times (M_{it}^+ + M_{it}^-) + \delta_g^+ M_{it}^+ + \delta_g^- M_{it}^- + e_{it+1} \quad (3)$$

where  $M_{it}^+$  (resp.  $M_{it}^-$ ) is equal to 1 if  $g_{it}$  belongs to the top (resp. bottom) 20% of its distribution. This specification is analogous to equation (2), but for realizations instead of forecasts:  $\beta_g$  is the coefficient for the relationship between  $g_{it+1}$  and  $g_{it}$  in the bulk of the distribution, while  $\gamma_g$  captures the change in this slope in the tails. As in (2), we allow the intercepts in the two tails to differ, but impose that the slopes are equal.

Figure 2: Non-Linear Conditional Expectation of Realizations: Net Sales



*Notes:* This figure shows binned scatter plots of current Net Sales on past Net Sales. Panel A shows the results for the entire sample; Panel B restricts the sample to the 20-80th percentiles of past Net Sales; Panel C restricts the sample to below the 20th percentile of past Net Sales; Panel D restricts the sample to above the 80th percentile of past Net Sales.

Table 4 shows that  $\hat{\beta}_g > 0$  in 20 out of 22 cases, and is statistically significant in 18 cases. Additionally, we find that  $\hat{\gamma}_g < 0$  in all cases, and it is statistically significant in all but one case. Overall, this evidence shows that the relationship between current and lagged values for our variables is highly non-linear, with persistence in the bulk and more mean-reversion in the tails of the distribution.

**Robustness.** In the Appendix, we show that this non-linear conditional expectation is robust to the alternative variable definitions discussed in the robustness checks in Section 2.1: Table IA.12 adjusts by firm-level volatility; Table IA.13 adjusts by aggregate volatility; and Table IA.14 uses percent instead of log growth. In all cases, the non-linearity is strong and pervasive.

### 3 Model

In this section, we develop a parsimonious model that ties the non-linear relationship between forecast errors and revisions, our Fact #1, to Facts #2 and #3, which are about the data-generating

Table 4: Non-Linear Conditional Expectation of Realizations

$g_{it+1}$	$g_{it}$		$g_{it} \times$ Top/Bottom 20% of $g_{it}$		Top 20% of $g_{it}$		Bottom 20% of $g_{it}$		# of Obs.
Variable	$\hat{\beta}_g$ (1)	$t$ -stat (2)	$\hat{\gamma}_g$ (3)	$t$ -stat (4)	$\hat{\delta}_g^+$ (5)	$t$ -stat (6)	$\hat{\delta}_g^-$ (7)	$t$ -stat (8)	
Net Sales	0.43	12.40	-0.44	-8.12	0.10	6.28	0.00	0.23	140482
Earnings Per Share	0.20	6.18	-0.49	-13.11	0.23	9.63	-0.03	-1.19	129735
Return On Equity	0.86	68.21	-0.28	-16.55	0.08	14.23	0.01	3.44	127268
Net Income	0.18	5.60	-0.47	-11.48	0.24	9.39	-0.02	-0.66	113862
Pre-tax Profit	0.22	5.92	-0.48	-11.61	0.23	8.96	0.00	0.03	102386
EBITDA	0.23	7.03	-0.40	-10.66	0.15	8.65	-0.00	-0.15	99731
Book Value Per Share	0.47	13.87	-0.53	-13.11	0.08	10.37	-0.03	-6.04	89884
EBIT	0.25	6.48	-0.46	-10.67	0.18	6.85	-0.01	-0.55	82100
Return On Assets	0.90	89.07	-0.28	-18.41	0.04	13.39	0.00	6.29	79479
GAAP EPS	0.01	0.28	-0.36	-7.56	0.22	8.84	0.01	0.44	74210
Net Asset Value	0.46	12.83	-0.44	-11.91	0.08	13.88	-0.03	-3.76	71494
Gross Margin	0.96	209.91	-0.05	-4.53	0.03	4.77	0.01	4.58	64483
Cash Flow Per Share	-0.18	-5.43	-0.14	-4.80	0.14	9.39	0.05	2.84	62446
Dividend Per Share	0.44	7.82	-0.66	-9.91	0.17	8.37	-0.03	-1.67	49402
Net Debt	0.09	4.35	-0.27	-8.79	0.16	9.73	0.00	0.13	48338
Enterprise Value	-0.05	-0.98	-0.10	-1.69	0.06	2.34	-0.01	-0.52	48140
Capital Expenditures	0.05	1.49	-0.22	-5.51	0.07	4.21	-0.06	-3.45	44623
Operating Profit	0.24	4.56	-0.51	-8.79	0.25	9.60	-0.01	-0.69	33033
EBITDA Per Share	0.13	3.28	-0.27	-4.76	0.08	3.12	0.00	0.05	15343
Cash Earnings Per Share	0.12	2.83	-0.36	-5.06	0.10	3.64	-0.00	-0.19	10013
Funds From Operations Per Share	0.37	3.48	-0.35	-2.78	0.04	2.46	-0.02	-1.54	3399
EPS Before Goodwill	0.13	2.02	-0.34	-7.87	0.13	4.13	-0.03	-1.50	2413

Notes: This table reports the OLS estimates of equation (3), for each one of the 22 variables. We report the estimates of  $\beta_g$  (column 1),  $\gamma_g$  (column 3),  $\delta_g^+$  (column 5), and  $\delta_g^-$  (column 7), along with the corresponding  $t$ -statistics with standard errors clustered by firm and time. Variables are ranked by number of observations. The sample includes one observation per firm-year.

process. We start by describing a model of the DGP that is consistent with our latter two facts, and then turn to a model of belief formation that, given this model of the DGP, generates the first fact. We then show the model can quantitatively replicate our facts, and a potential microfoundation based on [Augenblick et al. \(2024\)](#).

### 3.1 Data-Generating Process

The first piece of the model is the data-generating process (DGP) for  $g_t$ , our forecasting variable of interest. Our model does not have any firm-specific heterogeneity, and we therefore omit the firm

index  $i$  for brevity.<sup>5</sup> Without loss of generality, we normalize the unconditional mean of  $g_t$  to zero. We then assume that the DGP for  $g_t$  takes the following form

$$g_{t+1} = g_{t+1}^* + \sigma_\epsilon \epsilon_{t+1}, \quad \epsilon_t \sim f_\epsilon(\cdot), \quad (4)$$

$$g_{t+1}^* = \rho g_t^* + \sigma_u u_{t+1}, \quad u_t \sim f_u(\cdot), \quad (5)$$

where  $\epsilon_t$  and  $u_t$  are IID shocks with a zero mean and unit variance. The DGP for  $g_t$  therefore consists of two components: (i) a persistent component,  $g_t^*$ , that follows an AR1 process with shocks,  $u_t$  that have PDF  $f_u(\cdot)$ , and (ii) a transitory shock,  $\epsilon_t$ , with PDF  $f_\epsilon(\cdot)$ . Throughout, we assume that  $g_t^*$  is an *unobservable* latent state. We denote its unconditional variance by  $\sigma_{g^*}^2 = \frac{\sigma_u^2}{1-\rho^2}$ , and denote its PDF by  $\Phi(\cdot)$ . We also denote the marginal PDF of  $g_t$  by  $h(\cdot)$ , which is given by:

$$h(g) = \int_{-\infty}^{+\infty} \Phi(g - \sigma_\epsilon \epsilon) f_\epsilon(\epsilon) d\epsilon.$$

and denote  $\sigma_g^2 = \sigma_\epsilon^2 + \sigma_{g^*}^2$  its unconditional variance.

At this point, if we assumed that  $\epsilon_t$  and  $u_t$  were Gaussian, our model of the DGP would be identical to the models of dividend growth in [Bansal and Yaron \(2004\)](#) and [Lettau and Wachter \(2007\)](#). However, unlike these models and existing literature on belief formation, we instead assume the PDF of  $\epsilon_t$ ,  $f_\epsilon(\cdot)$ , has heavy tails in the sense that it is well approximated by a power law with tail parameter  $\nu$  for large values of  $\epsilon_t$ . Formally, we make the following assumption, where we use the notation  $a(x) \sim b(x)$  to denote asymptotic equivalence in the sense that  $\lim_{x \rightarrow \infty} \frac{a(x)}{b(x)} = 1$ .<sup>6</sup>

**Assumption 1.** *The distribution of  $\epsilon_t$  is symmetric and follows a power law:*

$$P(\epsilon_t > x) = \int_x^\infty f_\epsilon(\epsilon) d\epsilon \sim \Theta x^{-\nu} \text{ as } x \rightarrow \infty$$

where  $\nu > 2$  is the tail index and  $\Theta$  is a constant of proportionality.

An example of a case in which Assumption 1 holds would be if  $f_\epsilon(\cdot)$  is the PDF of a  $t$ -distribution, where  $\nu$  would be the degrees of freedom of the distribution. Finally, we need to make an assumption about the distribution of  $u_t$ . In all of our theoretical results, we assume that  $u_t$  has “thinner” tails than  $\epsilon_t$ , which is formalized in the following assumption.

<sup>5</sup>The lack of firm-specific heterogeneity in the model is consistent with the fact that our results are unchanged when standardizing at the firm-level.

<sup>6</sup>We prove all of our results in Appendix B under a weaker assumption that  $\epsilon_t$  is a regularly varying random variable such that  $P(\epsilon_t > x) \sim x^{-\nu} L(x)$ , where  $L(x)$  is a slowly varying function.

**Assumption 2.** *The distribution of  $u_t$  is symmetric and exhibits tail behavior that is asymptotically negligible compared to  $\epsilon_t$ :*

$$P(u_t > x) = \int_x^\infty f_u(u) du = o(x^{-\nu}).$$

For our theoretical results that characterize the properties of beliefs, Assumptions 1 and 2 are sufficient. However, to characterize the properties of the data-generating process for  $g_t$  in closed form, we will make a stronger assumption than Assumption 2 that  $u_t$  is normally distributed.

### 3.2 Replicating Facts #2 and #3: Fat Tails and Non-Linear $E(g_{t+1}|g_t)$

Before turning to our model of beliefs, we show that our model of the DGP can generate our second and third facts. Our second fact makes clear that we need a model of the DGP with fat tails. The fact that our model generates this follows from the properties of power laws: the tail parameter of a sum of independent random variables is the minimum of the tail parameters (Jessen and Mikosch, 2006). Given that  $g_t^*$  is the sum of random variables that have thinner tails than  $\epsilon_t$  and  $g_t$  is the sum of  $g_t^*$  and  $\epsilon_t$ , the tail parameter of  $g_t$  will be the same as that of  $\epsilon_t$ , which we denoted by  $\nu$ .

Showing our model replicates our third fact—the non-linearity in  $E(g_{t+1}|g_t)$ —is more difficult because there is no closed-form expression for  $E(g_{t+1}|g_t)$ . Nevertheless, the following proposition leverages a result from empirical Bayes theory known as Tweedie’s formula (Efron, 2012) to characterize it.

**Proposition 1** (Tweedie’s formula). *If  $u_t$  is normally distributed, the expectation of future growth conditional on current growth takes the following form:*

$$E(g_{t+1}|g_t) = -\rho\sigma_{g^*}^2 \frac{d}{dg} \log h(g_t). \quad (6)$$

*In particular, it is a function of the **observable** distribution of growth,  $g_t$ .*

*Proof.* See Appendix B.1. □

Proposition 1 shows that  $E(g_{t+1}|g_t)$  is a function of  $h(\cdot)$ , the marginal PDF of  $g_t$ . This result is useful because  $h(\cdot)$  is observable, which means we can characterize the shape of  $E(g_{t+1}|g_t)$  even without a closed form. For example, suppose that this distribution is approximately Gaussian

in the bulk of the distribution:  $\log h(g) \approx -\frac{g^2}{2\Sigma_g} + \text{constant}$ , which [Table IA.15](#) shows is a good approximation, especially for Net Sales. Given this approximation, [Proposition 1](#) implies

$$E(g_{t+1}|g_t) \approx \rho \frac{\sigma_{g^*}^2}{\Sigma_g} g_t, \quad (7)$$

which is linear in  $g_t$ . This is consistent with our third fact: in the bulk of the distribution, the expectation of  $g_{t+1}$  is linearly increasing in  $g_t$ . Intuitively, in the bulk of the distribution,  $g_t$  is dominated by  $g_t^*$  and  $g_t^*$  is persistent, implying that  $g_t$  is a good approximation of  $g_{t+1}$ . Note that if  $g_t$  were globally Gaussian, then this equation would hold with equality, and we would recover the well-known result that the rational expectation given a signal of a state with additive Gaussian noise is linear. The benefit of [Proposition 1](#) is that it allows us to extend this result to distributions that are locally Gaussian.

In the tails of the distribution, [Table 3](#) shows that the growth distribution is well approximated by a power law, consistent with [Assumption 1](#). Under [Assumption 1](#), we can use the properties of regularly varying functions to derive the following result that characterizes  $E(g_{t+1}|g_t)$ .

**Proposition 2.** *If  $u_t$  is normally distributed, [Assumption 1](#) implies that:*

$$E(g_{t+1}|g_t) \sim \frac{\rho\sigma_{g^*}^2(\nu + 1)}{g_t} \text{ as } |g_t| \rightarrow \infty. \quad (8)$$

*Proof.* See [Appendix B.2](#). □

[Proposition 2](#) shows that, in the tails of the growth distribution, the expectation of  $g_{t+1}$  is *decreasing* in  $g_t$ , unlike in the bulk of the distribution. The intuition is that extreme values of  $g_t$  are likely driven by  $\epsilon_t$ . However, because  $\epsilon_t$  is not persistent, this value of  $g_t$  is unlikely to persist at  $t + 1$ . In the limit of  $|g_t| \rightarrow \infty$ ,  $g_t$  reflects  $\epsilon_t$  with probability one, and  $E(g_{t+1}|g_t)$  converges to the unconditional expectation of zero.<sup>7</sup> While it is clear that our model of the DGP needs to have fat tails, [Proposition 2](#) also explains why an alternative model in which the distribution of permanent shocks,  $u_t$ , has fatter tails than that of transitory shocks,  $\epsilon_t$ , would be inconsistent with our third fact.<sup>8</sup> With fatter-tailed permanent shocks, large shocks would be more likely to persist, and the conditional expectation of  $g_{t+1}$  given  $g_t$  would be *increasing* in  $g_t$  in the tails of the distribution.

<sup>7</sup>A related result is developed in [Chambers and Healy \(2011\)](#), who show that when a distribution has sufficiently fat tails such that MLRP does not hold, a positive signal may generate a negative update about the underlying parameter.

<sup>8</sup>Our specification of the DGP is reminiscent of the recent literature on income dynamics ([Guvenen et al., 2014, 2021](#)), which models income processes as the sum of persistent and transitory components. However, this literature emphasizes the importance of non-normal persistent shocks, while the key ingredient in our model is non-normal *transitory* shocks. Another difference is that our focus is primarily on generating the kurtosis in the data rather than skewness. The latter is a pervasive feature of income data due to extreme negative events like job loss.

### 3.3 Expectations Formation

Given our model of the DGP, we now describe our model of subjective expectation formation. We assume that expectations do not have “full information,” in the sense that forecasters only observe realizations of  $g_t$  but not  $g_t^*$ . However, this assumption alone cannot explain our first fact, given it would imply that forecast errors should not be predictable by revisions, which are in forecasters’ information set. Therefore, we also assume that forecasts are not rational given this information set. In particular, our core assumption is that forecasters incorrectly perceive the distributions of  $\epsilon_t$  and  $u_t$  to be Gaussian, inconsistent with Assumptions 1 and 2. We do not microfound this misperception, but we view it as consistent with the idea that economic agents use simplified, or “sparse”, models of reality to form their beliefs (Fuster et al., 2010; Gabaix, 2019). Because of this model misspecification, forecast errors will be (conditionally) predictable.

Since only past values of  $g_t$  are observable, agents have to solve a filtering problem to compute their expectations about  $g_t^*$  given  $g_0, \dots, g_t$ , which they in turn use to forecast future growth. Under the assumption that agents perceive  $\epsilon_t$  and  $u_t$  as Gaussian, the solution to this filtering problem implies that their expectations will be characterized by the Kalman filter. Given the DGP, our assumption on belief formation is equivalent to assuming that forecasters use the *linear* forecast with the lowest possible mean squared error. For our theoretical results, we assume that agents are in a steady state in the sense that the posterior variance of the Kalman filter and, hence, the Kalman gain is constant. Denoting this steady state Kalman gain as  $K$ , agents’ expectations at horizon  $k$ ,  $F_t g_{t+k}$ , are characterized by:

$$F_t g_{t+k} = \rho^k K \sum_{s \geq 0} \rho^s (1 - K)^s g_{t-s}. \quad (9)$$

Our assumption that forecasters use exactly the Kalman filter is not crucial for our results. For example, the results in the next section about the non-linear error-revision relationship would be present if forecasters linearly extrapolated current observations. What is important is that forecasters use a forecasting rule that does not fully capture the presence of fat tails. Aside from its tractability, we choose to use the Kalman filter specifically because it would correspond to rational expectations under the assumption that  $\epsilon_t$  and  $u_t$  were normal. This clarifies that our analytical results are entirely about ignoring the presence of fat tails, rather than an arbitrary assumption about the history or functional form that forecasts take.

### 3.4 Replicating Fact #1: Non-Linear Error-Revision Relationship

We now show that our model of belief formation, combined with our model of the DGP that replicates Facts #2 and #3, also generates Fact #1.

**Proposition 3.** *Define forecast errors and forecast revisions as follows:*

$$\begin{aligned} ERR_{t+1} &= g_{t+1} - F_t g_{t+1}, \\ REV_t &= F_t g_{t+1} - F_{t-1} g_{t+1}. \end{aligned}$$

*Then, under Assumptions 1 and 2, forecast errors are asymptotically linear in forecast revisions:*

$$E(ERR_{t+1} \mid REV_t) \sim C \times REV_t \quad \text{as } |REV_t| \rightarrow \infty, \quad (10)$$

where  $C < 0$ .

*Proof.* See Appendix B.3. □

Proposition 3 shows that in the tails of the distribution of revisions, the relationship between forecast errors and revisions is linear and negative. This is consistent with our first fact in Figure 1: there is overreaction in the tails of the distribution of revisions. The proof of this result relies on showing that revisions are driven by changes in  $g_t$ , which are asymptotically large (in absolute value) for one of two reasons: (i) a large realization of  $\epsilon_t$  or (ii) a large realization of past  $\epsilon_{t-h}$ . The intuition for why these large revisions reflect overreaction is that the forecaster does not realize that they are less likely to be persistent because they are driven by the fat-tailed component of the process,  $\epsilon_t$ , which is transitory. If  $\epsilon_t$  and  $u_t$  were Gaussian, there would still be transitory shocks, but the change in  $g_t$  would not be informative about the relative size of transitory and persistent shocks. In contrast, when  $\epsilon_t$  has fatter tails than  $u_t$ , large values of  $|g_t|$  asymptotically only reflect  $\epsilon_t$ .

The definition of the Kalman filter implies that it is the minimum linear mean squared error predictor. Combining this with Proposition 3 implies the following result.

**Corollary 1.** *Under Assumptions 1 and 2, there exists an  $\bar{R} > 0$  such that:*

$$\begin{aligned} E\left( ERR_{t+1} \times REV_t \mid |REV_t| > \bar{R} \right) &< 0 \\ E\left( ERR_{t+1} \times REV_t \mid |REV_t| < \bar{R} \right) &> 0 \end{aligned}$$

*Proof.* See Appendix B.4. □

Corollary 1 shows that our model can generate the non-linear error-revision relationship in Figure 1. While Proposition 3 shows there is overreaction in the tails, Corollary 1 shows that errors and revisions are positively correlated *on average* in the bulk of the distribution of revisions. The intuition for why forecasters underreact in the bulk is similar to the intuition for overreaction in the tails: the forecasters do not realize that intermediate values of  $g_t$ , which generate smaller revisions, are more likely to reflect  $u_t$  than  $\epsilon_t$ , and hence are more likely to be persistent. Note, however, that while Corollary 1 shows that errors and revisions are positively correlated *on average* in the bulk where  $|REV_t| < \bar{R}$ , we cannot show that this is true for all  $|REV_t| < \bar{R}$  without further assumptions on  $f_\epsilon(\cdot)$  and  $f_u(\cdot)$ . Additionally, we cannot characterize the exact value of  $\bar{R}$  at which the switch between underreaction and overreaction occurs.

### 3.5 Quantitative Fit

We now assess whether our model can quantitatively account for our three main empirical facts. Because of computational constraints, we focus on the one variable that has the largest sample, Net Sales.

**Simulation procedure.** To simulate from our model, we make three additional assumptions. First, we make the stronger version of Assumption 2 that  $u_t$  is normally distributed with unit variance, which is motivated by the evidence in Table IA.15. Second, we assume that  $\epsilon_t$  is distributed according to a  $t$ -distribution with  $\nu > 2$  degrees of freedom normalized to have a unit variance. The  $t$ -distribution is asymptotically a power law with tail parameter  $\nu$ , and has the nice property of converging to a normal distribution as  $\nu \rightarrow \infty$ . Finally, we relax the assumption that the Kalman filter updating equations are applied using a constant Kalman gain, which would only apply in a steady state. Instead, we use the following updating equations to compute subjective forecasts, which follow from applying standard Kalman filter results to equations (4) and (5) under the (incorrect) assumption that  $\epsilon_t \sim N(0, 1)$ :

$$\begin{aligned} F_t g_{t+h} &= \rho^h F_t g_t^*, \\ F_t g_t^* &= (1 - K_t) F_{t-1} g_t^* + K_t g_t, \quad F_0 g_0^* = g_0^*, \\ K_t &= \frac{\Sigma_t}{\Sigma_t + \sigma_\epsilon^2}, \\ \Sigma_{t+1} &= \rho^2 (1 - K_t) \Sigma_t + \sigma_u^2, \quad \Sigma_0 = 0. \end{aligned} \tag{11}$$

In our simulations, we sample time series of  $g_t$  according to equations (4) and (5) with 100,000 observations. We repeat this simulation 100 times, where  $g_0^*$  is drawn from its stationary distribution;

the length of the simulation burn-in period is 50 observations for each series. This gives us a total of 10 million simulated observations.<sup>9</sup>

**Fit on Facts #2 and #3.** We start by estimating the four parameters of the DGP for  $g_t$  in (4) and (5):  $\rho$ , the persistence of  $g_t^*$ ,  $\sigma_\epsilon$ , the scale parameter for  $\epsilon_t$ ,  $\nu$ , the tail parameter of  $\epsilon$ , and the innovation volatility,  $\sigma_u$ . We estimate these parameters using Simulated Minimum Distance (SMD), minimizing the difference between a set of statistics computed from simulated data in the model and the corresponding value of those statistics in the data for Net Sales with the inverse-diagonal of the sample covariance matrix as the weighting matrix.<sup>10</sup> We use the following six statistics: our estimates of  $\beta_g$  and  $\gamma_g$  from Table 4; the tail parameter from Table 3; the 10-90 percentile difference of  $g_t$ ,  $g_{t+1} - g_t$ , and  $g_{t+3} - g_t$ . The first two statistics,  $\beta_g$  and  $\gamma_g$ , capture Fact #3: the non-linearity in the conditional expectation of  $g_{t+1}$  given  $g_t$ . These regression coefficients jointly identify  $\rho$ ,  $\sigma_u$ , and  $\nu$ , but they do not separately identify each of these parameters. The tail parameter that captures Fact #2 identifies  $\nu$  separately. The dispersion of  $g_t$  helps identify the scale of the DGP, in particular  $\sigma_u$  and  $\sigma_\epsilon$ . Finally, the dispersion of one- and three-year changes in  $g_t$  is useful for separately identifying  $\rho$  and  $\sigma_u$  because the extent to which the three-year changes tend to be larger than one-year changes is affected by  $\rho$  much more than other parameters. Our choice of one- and three-year changes, specifically, follows the literature on income dynamics, which uses the same moments to identify persistence and volatility in similar data-generating processes (e.g., Guvenen et al., 2014, 2021).

Our SMD estimates for the DGP parameters are shown in the first four columns of Table 5. Consistent with Fact #2, we estimate  $\nu = 2.411$ . Combined with our estimate of a relatively persistent process for  $g_t^*$ ,  $\rho = 0.784$ , this generates Fact #3. The first six rows of Table 6 show the fit of the model on the targeted statistics at these estimated parameters. The results show that the model is able to fit the data very well, replicating all of the targeted statistics almost exactly. In Section 4.3, we will examine the out-of-sample prediction that this estimated model makes for return predictability.

**Fit on Fact #1.** We now turn to the model’s fit of Fact #1. Assuming expectations are formed according to the Kalman filter in (11), the final column panel of Table 6 shows the model’s prediction

<sup>9</sup>We choose this simulation size to be as large as possible without exceeding the RAM of our GPU. The simulation itself is not memory or computationally intensive—these constraints only become binding when we compute the rational expectation using a particle filter described below.

<sup>10</sup>We perform this optimization using a two-step procedure in which we first search on a quasi-random grid of 15,000 points, and then run local Nelder-Mead optimizations using the top 5 points as starting points. Our result is then the parameter vector that has the lowest objective function from any of these Nelder-Mead optimizations.

Table 5: Simulated Minimum Distance Parameter Estimates

	$\rho$	$\sigma_u$	$\sigma_\epsilon$	$\nu$	$\lambda$
Estimate	0.784	0.336	1.058	2.411	0.197
Std. Error	0.020	0.018	0.046	0.037	0.024

*Notes:* This table shows the parameter estimates and standard errors from the two-step Simulated Minimum Distance estimation described in the main text. Standard errors are computed using the covariance matrix of the estimation moments and the gradient of the moment vector computed using central differencing with a step size equal to 1% of the estimated parameter value. We adjust our standard error for  $\lambda$  to account for the first step of the estimation using the procedure in [Murphy and Topel \(1985\)](#).

Table 6: Fit of Model on Estimation Targets

Moment	Data	Model	Model ( $\lambda = 1$ )
$\hat{\beta}_g$	0.429	0.425	.
$\hat{\gamma}_g$	-0.438	-0.437	.
Tail Parameter	2.670	2.670	.
P90-P10 $g_t$	2.146	2.147	.
P90-P10 $g_{t+1} - g_t$	2.595	2.594	.
P90-P10 $g_{t+3} - g_t$	2.844	2.842	.
$\hat{\beta}$	0.106	0.144	1.816
$\hat{\gamma}$	-0.265	-0.255	-2.861

*Notes:* This table shows the fit of the estimated model on the statistics used to estimate it, where the top half are the statistics used to estimate the four DGP parameters and the bottom half are the two statistics used to estimate  $\lambda$ . The first column shows their values in the data. The second column shows their value in the estimated model computed from simulations at the estimated parameter values shown in [Table 5](#). The final column shows values at the same model parameters, changing only  $\lambda$  to be set to 1. The set of statistics that correspond to the data-generating process are not reported because they do not depend on  $\lambda$ .

$\beta$  and  $\gamma$  from [Table 2](#). Comparing this with the data in the first column, we see that the model qualitatively generates the non-linear error-revision relationship in the data, with underreaction in the bulk of the distribution and overreaction in the tails, consistent with our theoretical results. However, the degree of this non-linearity is an order of magnitude too large. The fact that our model cannot fit the data is not surprising given that our model of beliefs has no additional degrees of freedom, given our estimates of the DGP parameters.

To attenuate the bias of our linear forecasting model, we follow [Fuster et al. \(2010\)](#) and [Gabaix \(2019\)](#) and assume that expectations are formed by shrinking away from the rational expectation towards a default model, which we consider as the Kalman filter:

$$F_t^\lambda g_{t+h} = \lambda F_t g_{t+h} + (1 - \lambda) E_t g_{t+h} \quad (12)$$

where  $E_t$  is the rational expectation given the full history of  $g_t$  and  $\lambda$  is the weight that is placed on the Kalman filter forecast.  $\lambda = 1$  corresponds to the case in the final column of [Table 6](#), while

$\lambda = 0$  corresponds to the case of rational expectations in which error-revision coefficients would always be zero. We then estimate  $\lambda$  to assess whether the anchored model in (12) can match our first fact quantitatively. Doing so requires computing the rational expectation,  $E_t g_{t+1}$ , which we do using the particle filter from [Fernandez-Villaverde and Rubio-Ramirez \(2007\)](#) as described in Appendix C. Having computed the rational expectation, we estimate  $\lambda$  using SMD with  $\beta$  and  $\gamma$  as statistics and the inverse-diagonal of the covariance matrix as a weighting matrix.<sup>11</sup>

Our estimated value of  $\lambda$ , shown in the last column of [Table 5](#), implies that forecasters place 20% of their weight on the Kalman filter forecast, and 80% of their weight on the true forecast. The second column of [Table 6](#) shows that with  $\lambda = 0.197$ , the model is able to quantitatively replicate the non-linear error-revision relationship in the data.

### 3.6 A Microfoundation Based on Insensitivity to Signal Strength

In this section, we show that our assumption that forecasters ignore fat tails and use the Kalman filter is analogous to assuming that forecasters are insensitive to the strength of the signals that they observe each period, as in the theory of [Augenblick et al. \(2024\)](#) (ALT).

In order to make this point, we first note that rational expectations, which cannot be computed in closed form, are well-approximated by the following state-dependent Kalman filter:<sup>12</sup>

$$\begin{aligned} F_t^{SD} g_{t+h} &= \rho^h F_t^{SD} g_t^*, \\ F_t^{SD} g_t^* &= F_{t-1} g_t + \tilde{K}(v_t) v_t, \quad v_t = g_t - F_{t-1}^{SD} g_t \\ \tilde{K}(v_t) &= K \frac{\gamma_0}{1 + \exp(|v_t| - \gamma_1)}, \end{aligned} \tag{13}$$

where  $K$  is the steady state Kalman gain and  $(\gamma_0, \gamma_1)$  are parameters. (13) corresponds to a state-dependent version of the steady state Kalman filter, where  $\tilde{K}(\cdot)$  tilts away from the steady state Kalman gain based on the time  $t$  forecast error,  $v_t$ .<sup>13</sup> We estimate  $\gamma_0 = 4.55$  and  $\gamma_1 = 0.89$  by minimizing the mean squared difference between the one-period-ahead forecasts in (13) and the rational expectation in our full sample of simulated data. We find that this state-dependent filter provides a very close approximation to the particle filter with an  $R^2$  of 99.4% and a slope coefficient that is nearly one.

<sup>11</sup>In principle, we could have estimated  $\lambda$  jointly with the DGP parameters. However, this is too computationally-intensive because it requires rerunning the particle filter for each different set of DGP parameters. Nevertheless, we adjust our standard error for  $\lambda$  to account for the two-step estimation procedure following [Murphy and Topel \(1985\)](#).

<sup>12</sup>We thank Stefan Nagel for suggesting the idea of using a state-dependent Kalman gain.

<sup>13</sup>We tilt away from the steady state Kalman gain because convergence occurs after only three periods.

Given the quasi-rational expectation in (13), the theory in ALT provides a microfoundation for forecasters ignoring fat tails. To see this, observe that our assumption that forecasters use a constant Kalman gain implies that forecasters place too much weight on  $v_t$  when it is large and too little weight when it is small. For extremely large forecast errors, the quasi-rational weight on recent information should be  $\lim_{v \rightarrow \infty} \tilde{K}(v) = 0$ , while the constant Kalman gain uses a positive weight. For small errors, the quasi-rational weight on  $v_t$  is  $\tilde{K}(0) = K \frac{4.55}{1 + \exp(-0.89)}$ , which is larger than the constant Kalman gain. Therefore, using a constant Kalman gain, as we assume in our model, is analogous to not recognizing that different signals have different strengths and giving too much weight to large uninformative shocks but too little to small informative ones. As a result, forecasters overreact to imprecise signals (larger errors) and underreact to precise ones (smaller errors), just like in ALT.

An alternative way to see the connection between underreaction versus overreaction and ALT is the following result.

**Proposition 4.** *If  $u_t$  is normally distributed, Assumption 1 implies that:*

$$\text{var}(g_{t+1}|g_t) = \sigma_g^2 + \rho^2 \sigma_{g^*}^4 \frac{d^2}{dg^2} \log h(g_t).$$

*Proof.* See Appendix B.5. □

Proposition 4 shows that the conditional variance of  $g_{t+1}$  given  $g_t$  depends on the second derivative of the log-density of  $g_t$ . Given our model of the DGP,  $\log h(\cdot)$  is concave in the bulk and convex in the tails. Using Proposition 4, these patterns imply that  $g_t$  is a more precise signal of  $g_{t+1}$  for smaller values of past growth, while it is less precise in the tails of the distribution. Therefore, our model delivers a similar result to ALT: forecasters are underreacting in the parts of the distribution where signals are strong, and overreacting when signals are weak.<sup>14</sup>

## 4 Additional Evidence from an Experiment, Macro Forecasts, and Stock Returns

This section presents additional evidence in support of our model from three different domains. The first comes from an online forecasting experiment, the second comes from macro forecasts, and

---

<sup>14</sup>This result is also consistent with the model in Ba et al. (2024) (BBI), which (under parameter restrictions) makes the same prediction for the relationship between underreaction and overreaction and signal strength as ALT. While both models in ALT and BBI make this same prediction, the mechanisms are quite different. This leads them to make other different predictions, such as how overreaction responds to complexity. See Appendix C of BBI for additional details.

the third comes from data on stock returns.

## 4.1 Evidence from a Forecasting Experiment

**Experimental design.** The design of our online forecasting experiment is taken from Afrouzi et al. (2023) (AKLMT). Participants are asked to predict the outcome of a process, and their compensation depends on the accuracy of their forecasts. We recruit participants on Amazon MTurk, and they are reasonably representative of the general population. The experiment does not require participants to have any prior knowledge of statistics, and participants do not know the DGP (AKLMT show that this is not important). The interface is graphical and user-friendly: participants click with their mice to provide their forecasts at one- and two-period-ahead horizons. Prior to making their first forecast, they see 40 prior realizations of the process, and then sequentially provide both forecasts in 40 periods, seeing the realization of the process between each period. We refer the reader to AKLMT for further details about this design; see Figure IA.1 for an example of the interface our participants see for these parameters.

Starting with this design, we set the DGP of the process being forecasted to be:

$$g_{t+1} = g_{t+1}^* + 12.16\epsilon_{t+1}, \quad \epsilon_t \sim t(2.533) \quad (14)$$

$$g_{t+1}^* = 0.529g_t^* + 12.62u_{t+1}, \quad u_t \sim N(0, 1) \quad (15)$$

This is the DGP in our theoretical model (equations (4)-(5)) with parameters selected to approximate our data on Net Sales.<sup>15</sup> We set  $\sigma_u$  and  $\sigma_\epsilon$ , which simply scale the DGP, to be larger numbers so that values are easier for participants to interpret. As shown in Figure IA.1, the tail of our DGP is thick enough that participants typically observe a few large movements in the DGP in the past 40 realizations they are initially given. We are not in a situation where large shocks of  $\epsilon$  rarely happen.

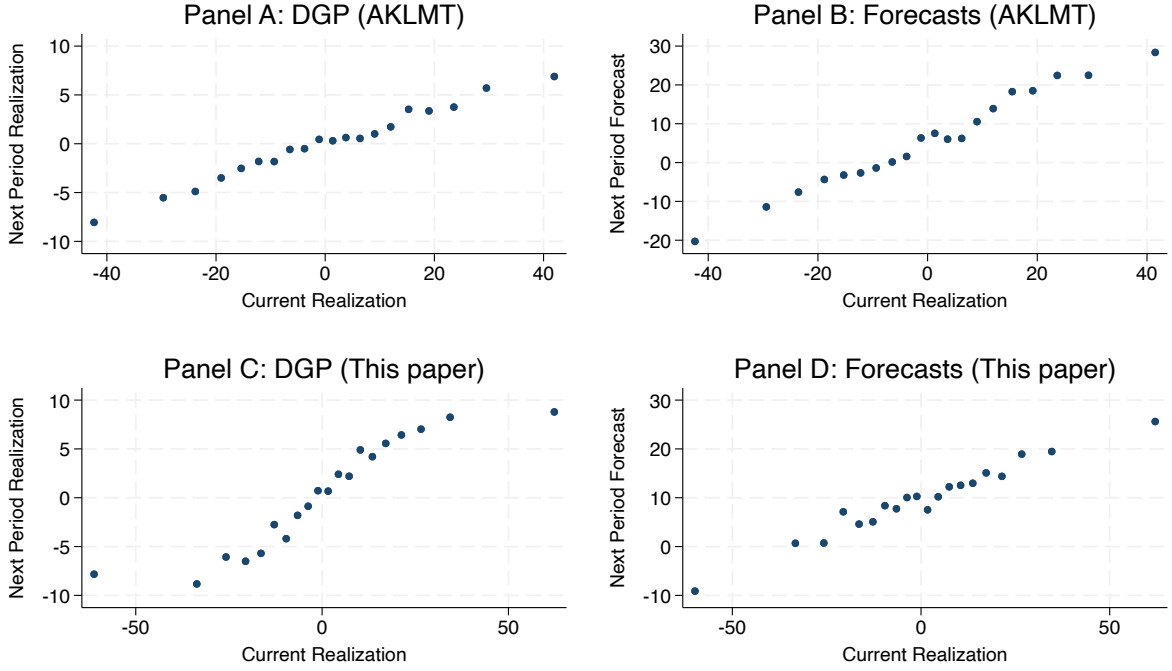
We ran the experiment in two waves with 201 participants on March 17, 2025 and 202 participants on April 29, 2025. Given that each participant makes 40 forecasts, our data have 16,120 observations. We compare our results using the DGP in (14) and (15) with those from a condition in AKLMT in which the process is an AR1 with Gaussian shocks and persistence parameter of 0.2.<sup>16</sup>

Before describing the results, we briefly discuss the relative benefits and costs of this experiment relative to our evidence in Section 2. The main benefit of running an experiment is that it allows us to directly control the parameters of the DGP. However, this benefit comes with several downsides.

<sup>15</sup>These parameters are from an estimation in a prior draft that used a smaller sample of data on Net Sales.

<sup>16</sup>We use 0.2 because it is the closest experimental condition in AKLMT to the regression coefficient of  $g_{it+1}$  and  $g_{it}$  when simulating from (14) and (15).

Figure 3: Data-Generating Process and Forecasts in Experimental Data



*Notes:* Panels A and C show binned scatterplots of  $g_{it+1}$  versus  $g_{it}$  in experimental data, where  $i$  is the participant and  $t$  the round of forecasting. Panels B and D show binned scatterplots of  $F_{it}g_{it+1}$  against  $g_{it}$ . Panels A and B use data from [Afrouzi et al. \(2023\)](#), where the DGP is a Gaussian AR1 process with a persistence parameter of 0.2. Panels C and D use the experimental data described in the main text, where the DGP is fitted to our data – see equations (14) and (15).

First, participants in the experiment are fundamentally different from the forecasters in our data, who are professional equity analysts with incentives aside from accuracy. Second, the professional forecasters in our data are likely to have a better understanding of the DGP, while we cannot control the priors of the participants in our experiment. Finally, the environment of forecasting a variable in our experiment is obviously quite different from forecasting in the real world. Despite these differences, we still view this experiment as useful because, by comparing our results with those in AKLMT, we can perform a direct test of our theory’s key prediction that transitory fat-tailed shocks create a non-linear relationship between errors and revisions.

**Results.** Panel A of [Figure 3](#) shows the relationship between  $g_{it+1}$  and  $g_{it}$  in the data from AKLMT, where  $i$  is the participant and  $t$  is the round of forecasting. As expected, this relationship is linear. In contrast, Panel C shows the same relationship in our experimental data, where the DGP is in (14) and (15). Although our experimental data have one order of magnitude fewer observations than our firm-level data, these data still replicate the non-linear relationship in Fact #3.

Panels B and D of [Figure 3](#) provide evidence that, despite the differences in the DGP, participants' one-period ahead forecasts,  $F_{it}g_{it+1}$ , are still linear in  $g_{it}$ , unlike the true DGP. This is consistent with our theory that forecasters ignore the fact that large shocks are likely to be transitory. Comparing Panels A and C with B and D also illustrates that forecasters substantially overestimate the persistence of the process:  $E(F_{it}g_{it+1} | g_{it})$  is steeper than  $E(g_{it+1} | g_{it})$  at all points. The combination of this extrapolation (a core result of AKLMT) and the likely presence of expectation noise ([de Silva and Thesmar, 2024](#)) leads to a strong negative correlation between errors and revisions on average. Columns (1) and (3) of [Table 7](#) show that the relationship between forecast errors and revisions is significantly negative in our experiment and in AKLMT, with a similar magnitude.

Despite the presence of overreaction on average, the fact that  $E(F_{it}g_{it+1} | g_{it})$  is linear in Panel D while  $E(g_{it+1} | g_{it})$  is non-linear in Panel C provides preliminary evidence that forecasters are likely to overreact more in the tails of the distribution, consistent with our theory. We test this statistically by estimating (2), which is the same regression that we estimated on analyst forecast data in Section 2.1. For the data from AKLMT, whose DGP is fully linear and Gaussian, column (2) shows no difference in the amount of overreaction that is present in the bulk of the distribution and the tails: the extra slope for the top and bottom quintile of past revisions is small and not statistically different from zero. In contrast, column (4) shows that, in our experimental data, this extra slope is three times larger and has a  $t$ -stat of 3.3, which is lower than in the analyst data but consistent with the fact that the sample is an order of magnitude smaller. Collectively, these findings provide direct evidence for our theory's key prediction that transitory fat-tailed shocks create a non-linear relationship between  $g_{t+1}$  and  $g_t$  and, as a result, between errors and revisions.

## 4.2 Evidence from SPF Macro Forecasts

We next study forecasts of macroeconomic outcomes from the [Survey of Professional Forecasters \(1968–2025\)](#) (SPF), which have been studied extensively in the literature on belief formation ([Bordalo et al., 2020a](#)). While we find qualitatively similar results to our findings at the firm-level, we interpret these results as suggestive given the much smaller sample size. In constructing our sample, we attempt to follow the exact choices made in the replication package of [Bordalo et al. \(2020a\)](#) (BGMS) as closely as possible, and refer the reader to BGMS for additional details.

**Forecasting variables.** The SPF is conducted around the end of the second month in each quarter and collects forecasts from individual forecasters that are anonymized and identified by forecaster IDs. We study the same set of outcome variables in the SPF as in BGMS, extending their sample

Table 7: Non-Linear Forecast Error-Revision Relationship: Experimental Data

	Dependent Variable: $\text{Error}_{it+1}$			
	Afrouzi et al. (2023)		New Experiment	
	(1)	(2)	(3)	(4)
Revision <sub>it</sub>	-0.42 (-21.56)	-0.40 (-10.24)	-0.35 (-20.51)	-0.27 (-6.73)
Revision <sub>it</sub> × Top/Bottom 20% of Revision <sub>it</sub>		-0.06 (-1.17)		-0.17 (-3.28)
Top 20% of Revision <sub>it</sub>		-2.37 (-0.98)		7.64 (3.55)
Bottom 20% of Revision <sub>it</sub>		-7.57 (-2.94)		-3.96 (-1.53)
Constant	-6.98 (-5.33)	-5.22 (-4.03)	-8.40 (-7.86)	-9.14 (-7.40)
# of Obs.	6859	6859	15003	15003

*Notes:* This table shows results from regressions of forecast errors onto revisions. In columns (1) and (2), we use data from Afrouzi et al. (2023), where the DGP is a Gaussian AR1 process. In columns (3) and (4), we use the experimental data described in the main text, where the DGP is fitted to our data – see equations (14) and (15). Columns (1) and (3) regress errors on revisions. Columns (2) and (4) estimate equation (2). All standard errors are clustered at the participant level. These regressions have fewer than 16,120 observations because computing forecast revisions loses one observation per subject and because we trim the top and bottom 1% of errors and revisions (as in Table 2).

forward to the second quarter in 2025.<sup>17</sup> Following BGMS, for variables that are not stationary in levels (e.g., GDP or CPI), we look at annual growth rates from the most recent realized quarter to three quarters in the future. For the remaining variables, we look at levels three quarters in the future. While macroeconomic variables are released quarterly, they are often subsequently revised. Therefore, we follow BGMS and focus on initial releases from the [Real-Time Data Set for Macroeconomists \(1968–2025\)](#). As an example, actual GDP growth from quarter  $t - 1$  to  $t + 3$  is defined as the initial release at  $t + 4$  of the GDP level in  $t + 3$  divided by the contemporaneous release of the GDP level at  $t - 1$ . For financial variables, the actual outcomes are not revised, and we use historical data from the Federal Reserve Bank of St. Louis.

**Forecast errors and revisions.** The SPF contains forecasts for the levels of variables, and we focus on an annual forecast horizon. Forecasters report forecasts of outcomes in the current and next four quarters. Given the timing of this survey, forecasts made in quarter  $t$  are made with knowledge of the realized values of variables with quarterly releases in quarter  $t - 1$  and the realized values of variables with monthly releases up to the previous month. For the variables that we study in growth

<sup>17</sup>The only variable that we do not include is housing starts, which is only available in a monthly vintage and has many missing observations.

rates, we transform forecasts of the levels of these variables into growth rates by dividing the level forecasts by the most recent realized level from the previous quarter. For the remaining variables, we use the level forecasts directly. Denoting  $y_t$  as a realized outcome in quarter  $t$  and  $F_{jt}y_{t+h}$  as the forecast made in quarter  $t$  for quarter  $t+h$  by forecaster  $j$ , we construct the forecast errors as  $ERR_t y_{t+4} = y_{t+4} - F_{jt}y_{t+4}$ , and the quarterly forecast revision as  $REV_t y_{t+4} = F_{jt}y_{t+4} - F_{j,t-1}y_{t+4}$ . We winsorize outliers by removing, for each forecast horizon in a given quarter, forecasts that are more than 10 interquartile ranges away from the median. We also require that the forecasters in our sample make at least five forecasts.

**Results.** Table 8 performs the same analysis as Table 2 by estimating equation (2) at the forecaster-level for each of the variables in our SPF sample. As shown in the table, the size of our sample is much smaller, which reflects the fact that we only have one time-series for each variable. Nevertheless, we find patterns that are similar to our first fact at the firm-level: the error-revision relationship is non-linear with more overreaction in the tails of the distribution. Specifically, we find that  $\hat{\beta} > 0$  for most variables, but  $\hat{\gamma} < 0$  for all but one variable. These findings, coupled with our experimental evidence in Section 4.1, suggest that the non-linear error-revision relationship is more general than our sample of firm-level forecasts.

### 4.3 Evidence from Stock Return Predictability

This section examines and tests the predictions that our model of belief formation makes for return predictability. We focus on one type of predictability: momentum, the fact that past returns predict future returns (Jegadeesh and Titman, 2011).

**Model prediction: non-linearity in momentum.** To derive predictions about returns, we make two simplifying assumptions: (i) earnings growth is a constant fraction of sales growth, and (ii) the subjective discount rate is constant (as in Bouchaud et al. 2019 and Nagel and Xu 2022), which is consistent with evidence in De la O and Myers (2021). As shown in Appendix B.6, these assumptions coupled with the Campbell (1991) decomposition allow us to link returns with changes in expectations of our measure of Net Sales,  $g_t$ , using the following expression:

$$\log(1 + R_t) \approx \log(1 + R_f + \pi) + (g_t - F_{t-1}g_t) + \sum_{k=1}^{\infty} c^k REV_t g_{t+k}, \quad (16)$$

where  $R_t$  is the net return,  $\pi$  is the (constant) expected return,  $R_f$  is the gross risk-free rate,  $c$  is a linearization constant, and  $REV_t g_{t+k}$  denotes investors' subjective revisions about future growth

Table 8: Non-Linear Error-Revision Relationship: SPF Macro Forecasts

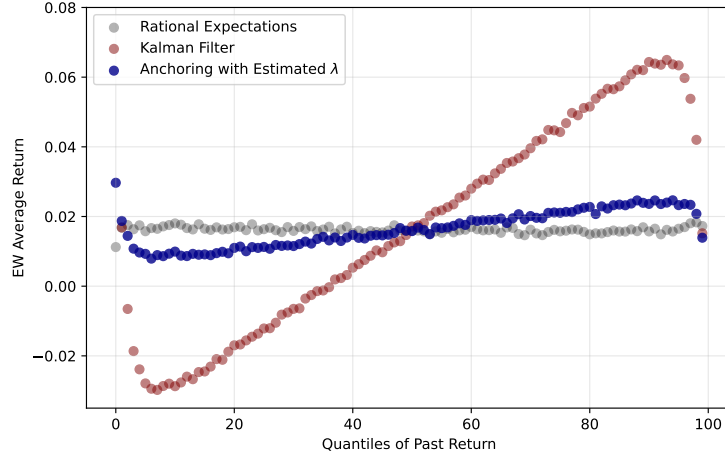
Error $_{it+1}$	Revision $_{it}$		Revision $_{it} \times$ Top/Bottom 20% of Revision $_{it}$		Top 20% of Revision $_{it}$		Bottom 20% of Revision $_{it}$		# of Obs.
	$\hat{\beta}$ (1)	$t$ -stat (2)	$\hat{\gamma}$ (3)	$t$ -stat (4)	$\hat{\delta}^+$ (5)	$t$ -stat (6)	$\hat{\delta}^-$ (7)	$t$ -stat (8)	
Nominal GDP	0.29	2.01	-0.81	-3.90	0.01	3.50	-0.00	-1.39	6026
Real GDP	0.43	2.49	-0.96	-4.32	0.01	2.72	-0.01	-1.75	6045
Real Consumption	0.07	0.42	-0.75	-2.66	0.01	3.62	-0.00	-0.60	4425
Unemployment	0.48	2.55	-0.98	-3.46	0.46	2.35	-0.52	-4.14	6161
GDP Price Index	0.42	2.02	-0.70	-2.93	0.01	4.03	-0.00	-0.82	5995
CPI	0.32	1.56	-0.65	-3.04	0.01	2.38	-0.00	-1.94	4525
Industrial Production	0.16	0.76	-0.66	-2.35	0.01	2.07	-0.01	-1.87	5567
Real Non-Residential Investment	0.47	2.23	-1.02	-3.34	0.02	2.55	-0.02	-1.98	4308
Real Residential Inv.	0.32	1.61	-0.92	-3.38	0.03	2.50	-0.05	-3.30	4302
Real Federal Govt. Consump.	-0.62	-3.51	0.05	0.21	-0.01	-1.63	-0.01	-2.30	4158
Real State/Local Govt. Consump.	-0.15	-0.83	-0.43	-2.23	0.00	1.41	-0.00	-1.92	4158
3-Month Treasury Rate	0.80	4.68	-0.76	-3.34	0.28	1.95	-0.56	-3.07	4491
10-Year Treasury Rate	0.14	0.67	-0.46	-1.81	0.28	2.78	-0.12	-0.91	3773
AAA Corporate Bond Rate	-0.15	-1.15	-0.13	-0.78	0.08	0.91	-0.09	-0.77	3655

*Notes:* This table reports the OLS estimates of equation (2), for each one of the variables in our SPF sample that is constructed following [Bordalo et al. \(2020a\)](#). We report the estimates of  $\beta$  (column 1),  $\gamma$  (column 3),  $\delta^+$  (column 5), and  $\delta^-$  (column 7), along with the corresponding  $t$ -statistics with standard errors clustered by forecaster and time. Variables are ordered following [Bordalo et al. \(2020a\)](#).

between  $t - 1$  and  $t$ . This expression is intuitive: given that discount rates are fixed, returns are driven by earnings surprises and revisions of future growth. As detailed in [Appendix B.6](#), we use (16) to generate a panel of simulated returns based on the simulated panel of realized and expected sales growth from our estimated model.

[Figure 4](#) shows the prediction that our model makes for momentum by showing a binned scatterplot of current returns  $r_t$  against past returns  $r_{t-1}$ . The model predicts momentum in the bulk of past returns, but mean-reversion in the tails. This is consistent with our expectations formation model, which predicts overreaction to large news, and underreaction to intermediate ones. As the figure also shows, the predictability is stronger for the more “biased” expectation: Kalman filter expectations (blue dots) generate the strongest predictability, while forecasts anchored to rational expectations (red dots) generate weaker predictability. Predictability disappears completely when forecasts are rational (gray dots) or when  $\epsilon$  is Gaussian, in which case the Kalman filter is rational.

Figure 4: Model-Implied Relationship Between Current and Future Returns



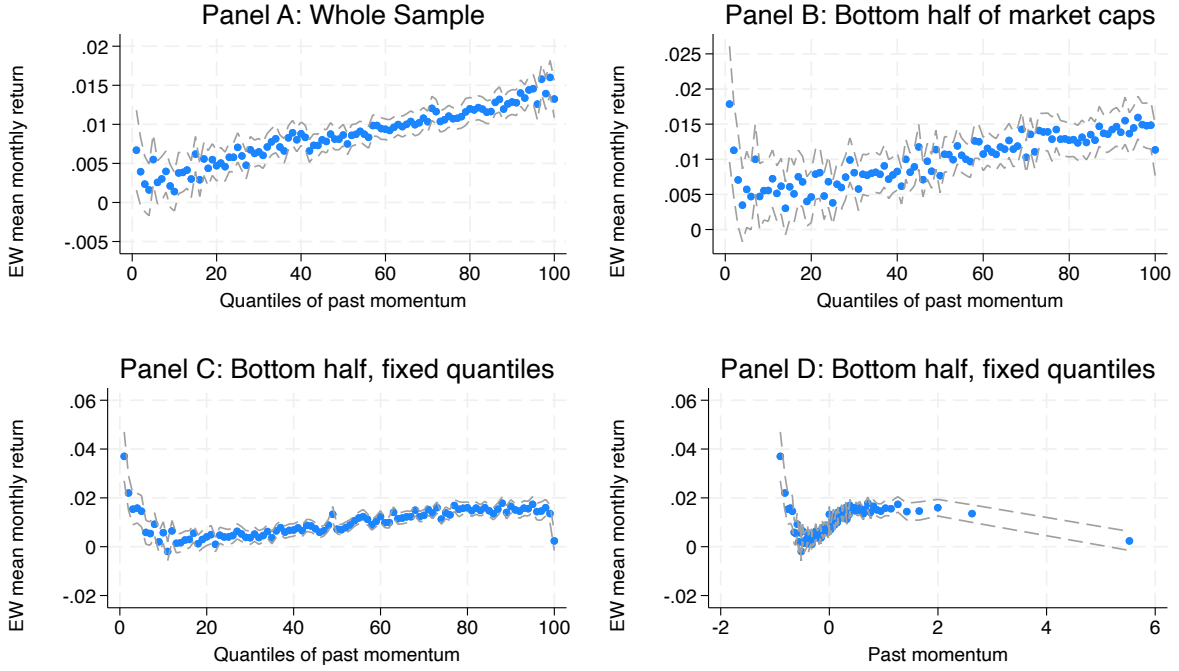
*Notes:* This figure shows a binned scatterplot of future returns against current returns in our simulated model. Given a set of earnings growth expectations, we compute returns using (16), as described in Appendix B.6. The three sets of points on the graph correspond to three cases in which beliefs are set equal to (i) the rational expectation computed with the particle filter, (ii) the Kalman filter, and (iii) the combination of the former two as in (12) with the value of  $\lambda$  in Table 5. We conduct this simulation assuming  $R_f = 1.01$ ,  $\pi = 5.5\%$ ,  $c = 0.96$ , and  $\rho$  is set to the estimated value in Table 5. We set the constant of proportionality between sales and earnings growth such that the standard deviation of returns is 15%.

**Non-linearity in momentum in the data.** We now examine whether momentum indeed exhibits mean-reversion in the tails in the data. We use CRSP monthly returns from 1927 to 2023 adjusted for delisting, and restrict to the sample of firms listed on NYSE, AMEX, and NASDAQ. Our measure of momentum follows the literature (Jegadeesh and Titman, 2011): every month  $t$ , momentum is the cumulative return between months  $t - 12$  and  $t - 1$  (thereby excluding the last month of past returns, which exhibit reversals). We define firm size as the market capitalization at  $t - 12$ .

We begin by plotting returns against centiles of momentum that are redefined each month, as in standard asset pricing tests. Panel A of Figure 5 shows the results for the entire CRSP sample: the line is upward sloping, consistent with the presence of momentum. However, we also find that there is a bit of mean-reversion for “super losers”, but not for “extreme winners”. In Panel B, we show that the non-linearity is more pronounced for smaller firms, which we define as those with below-median size. This is to be expected, as smaller stocks are more expensive to trade and therefore display more predictability (Novy-Marx and Velikov, 2016).

In Panel C of Figure 5, we show results when the centiles of past returns are defined on the entire sample rather than within each month, as we do within our model in Figure 4. While this is less standard in asset pricing tests because it implies the portfolios are not tradeable, it allows for the possibility that some of the non-linearity in momentum returns may come from time-series variation, which is more aligned with our model that does not distinguish between time-series and

Figure 5: Relationship Between Current and Past Returns in the Data



*Notes:* This figure shows binned scatterplots of monthly returns,  $r_t$ , as a function of past annual returns excluding month  $t - 1$ ,  $r_{t-12,t-1}$ . There are 100 bins in each panel. Panels A-C use the standard convention of using bins of past returns  $r_{t-12,t-1}$  as the x-axis. Panel D uses past returns,  $r_{t-12,t-1}$ , directly as the x-axis. Panel A uses the entire CRSP sample of stocks traded on AMEX, NASDAQ and NYSE. Panel B uses only the bottom half of stocks ranked by 12-month lagged market cap. In contrast to Panels A and B, quantiles of returns are defined on the entire sample in Panels C and D, rather than separately for each month. Dashed lines are 95% confidence bands based on standard errors assuming that returns are independent.

cross-sectional variation. Panel C shows that, using fixed centiles and focusing on firms below the 50% size cutoff, the non-linear relationship starts to appear significantly with a shape that is remarkably similar to that of our estimated model in Figure 4. The differences relative to Panel B suggest that the time series of momentum returns generates some mean-reversion in the tails, consistent with the literature on momentum crashes (Santa-Clara and Barroso, 2015): when volatility is high, losers are more likely to overperform, and winners more likely to underperform. Panel D of Figure 5 reproduces the analysis of Panel C, except that the x-axis is now rescaled to the average values of past returns. This last graph is less consistent with the practice of forming portfolios, as is typically done in asset pricing tests. Nevertheless, there is still significant non-linearity in the return-momentum relationship, in line with the prediction of our model.

## 5 Conclusion

In this paper, we argue that recognizing the complexity of the underlying data generating process, in particular its fat tails and non-Gaussian dynamics, is crucial for understanding the properties of subjective forecasts. We document three facts using data on forecasts of 22 firm-level variables by equity analysts: (i) the relationship between forecast revisions and future forecast errors—the variables used in [Coibion and Gorodnichenko \(2015\)](#) regressions—is strongly non-linear; (ii) the distributions of the underlying processes have fat tails; and (iii) the conditional expectation of future realizations is non-linear in current realizations, with persistence in the bulk and mean reversion in the tails. Next, we build a forecasting model that connects these facts. The key ingredients in our model are that the underlying process is non-Gaussian, but forecasters fail to recognize this and incorrectly fit a linear model to the data. Finally, we provide additional evidence in support of our theory by showing that it can explain evidence from an online forecasting experiment, that the non-linear relationship between errors and revisions is also present in macroeconomic forecasts, and that it provides an explanation for the presence of non-linearity in the momentum of stock returns.

Our paper raises several questions for further work. First, our model of belief formation is reduced-form. It would be fruitful to try to provide a microfoundation for why forecasters ignore fat tails in data-generating processes, which would allow us to study how this bias would manifest in other settings. One possible microfoundation that we emphasize is that forecasters may be insensitive to signal strength, as in [Augenblick et al. \(2024\)](#) and [Ba et al. \(2024\)](#). Second, our paper raises the question of how learning occurs in an environment with fat-tailed processes, which is likely to occur much more slowly and be more difficult. Finally, many of the driving processes in dynamic models of firm and household behavior have fat tails, such as revenue growth ([Stanley et al., 1996](#); [Boar et al., 2025](#); [Jaimovich et al., 2025](#)) and income growth ([Guisen et al., 2014](#)). Our findings suggest that allowing agents to, at least partially, ignore the presence of the fat tails might improve the ability of these models to explain firm and household behavior.

## References

- Afrouzi, Hassan, Spencer Y. Kwon, Augustin Landier, Yueran Ma, and David Thesmar, 2023, Overreaction in expectations: Evidence and theory, *The Quarterly Journal of Economics* 138, 1713–1764.
- Augenblick, Ned, Eben Lazarus, and Michael Thaler, 2024, Overinference from weak signals and underinference from strong signals, *Quarterly Journal of Economics* 140, 335–401.
- Axtell, Robert L., 2001, Zipf Distribution of U.S. Firm Sizes, *Science* 293, 1818–1820.
- Ba, Cuimin, J. Aislinn Bohren, and Alex Imas, 2024, Over- and underreaction to information, Working Paper, University of Pennsylvania and University of Chicago, 2024.
- Bansal, Ravi, and Amir Yaron, 2004, Risks for the long run: A potential resolution of asset pricing puzzles, *The Journal of Finance* 59, 1481–1509.
- Bianchi, Francesco, Sydney C. Ludvigson, and Sai Ma, 2022, Belief distortions and macroeconomic fluctuations, *American Economic Review* 112, 2269–2315.
- Bingham, N.H., C.M. Goldie, and L. Teugels J., 1987, *Regular Variation*, volume 27 of *Encyclopaedia of Mathematics and its Applications* (Cambridge University Press, Cambridge and New York).
- Boar, Corina, Denis Gorea, and Virgiliu Midrigan, 2025, Why are returns to private business wealth so dispersed?, Working Paper, 2025.
- Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer, 2020a, Overreaction in Macroeconomic Expectations, *American Economic Review* 110, 2748–82.
- Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer, 2019, Diagnostic Expectations and Stock Returns, *The Journal of Finance* 74, 2839–2874.
- Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, 2020b, Memory, Attention, and Choice, *The Quarterly Journal of Economics* 135, 1399–1442.
- Bottazzi, Giulio, and Angelo Secchi, 2006, Explaining the distribution of firm growth rates, *The RAND Journal of Economics* 37, 235–256.
- Bouchaud, Jean-Philippe, Philipp Krueger, Augustin Landier, and David Thesmar, 2019, Sticky Expectations and the Profitability Anomaly, *Journal of Finance* 74, 639–674.

- Campbell, John Y., 1991, A Variance Decomposition for Stock Returns, *The Economic Journal* 101, 157–179.
- 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, 195–228.
- Chambers, Christopher P., and Paul J. Healy, 2011, Reversals of signal-posterior monotonicity for any bounded prior, *Mathematical Social Sciences* 61, 178–180.
- Coibion, Olivier, and Yuriy Gorodnichenko, 2015, Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts, *American Economic Review* 105, 2644–2678.
- De la O, Ricardo, and Sean Myers, 2021, Subjective Cash Flow and Discount Rate Expectations, *The Journal of Finance* 76, 1339–1387.
- de Silva, Tim, and David Thesmar, 2024, Noise in expectations: Evidence from analyst forecasts, *Review of Financial Studies* 37, 1494–1537.
- Denuit, Michel, Patricia Ortega-Jimenez, and Christian Robert, 2025, Conditional expectations given the sum of independent random variables with regularly varying densities, *Astin Bulletin* 55, 449–485.
- Dew-Becker, Ian, Stefano Giglio, and Pooya Molavi, 2024, The dependence of belief dynamics on beliefs: Implications for stock returns, Working Paper, Northwestern University, Yale University, and University of Illinois, 2024.
- Efron, Bradley, 2012, Tweedie’s formula and selection bias, *Journal of the American Statistical Association* 106, 1602–1614.
- Fama, Eugene F., 1965, The behavior of stock-market prices, *The Journal of Business* 38, 34–105.
- Farmer, Leland E., Emi Nakamura, and Jón Steinsson, 2024, Learning About the Long Run, *Journal of Political Economy* 132, 3334–3378.
- Fernandez-Villaverde, Jesus, and Juan F. Rubio-Ramirez, 2007, Estimating macroeconomic models: A likelihood approach, *Review of Economic Studies* 74, 1059–1087.
- Fuster, Andreas, David Laibson, and Brock Mendel, 2010, Natural Expectations and Macroeconomic Fluctuations, *Journal of Economic Perspectives* 24, 67–84, Publisher: American Economic Association.

- Gabaix, Xavier, 2009, Power Laws in Economics and Finance, *Annual Review of Economics* 1, 255–294.
- Gabaix, Xavier, 2019, Behavioral Inattention, *Handbook of Behavioral Economics* .
- Geluk, J. L., and C. G. de Vries, 2006, Weighted sums of subexponential random variables and applications to reinsurance, *Insurance: Mathematics and Economics* 38, 230–246.
- Graeber, Thomas, Christopher Roth, and Marco Sammon, 2025, Categorical processing in an imperfect world, *Working Paper* .
- Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song, 2021, What Do Data on Millions of U.S. Workers Reveal About Lifecycle Earnings Dynamics?, *Econometrica* 89, 2303–2339.
- Guvenen, Fatih, Serdar Ozkan, and Jae Song, 2014, The nature of countercyclical income risk, *Journal of Political Economy* 122, 621–660.
- Jaimovich, Nir, Stephen J. Terry, and Nicholas Vincent, 2025, The empirical distribution of firm dynamics and its macro implications, *Working Paper*.
- Jegadeesh, Narasimhan, and Sheridan Titman, 2011, Momentum, *Annual Review of Financial Economics* 3, 493–509.
- Jessen, Anders Hedegaard, and Thomas Mikosch, 2006, Regularly varying functions, *Publications de l'institut mathématique* 79, 2303–2339.
- Kozlowski, Julian, Laura Veldkamp, and Venky Venkateswaran, 2020, The Tail that Wags the Economy: Belief-Driven Business Cycles and Persistent Stagnation, *Journal of Political Economy* 128, 2839–2880.
- Kwon, Spencer Y., and Johnny Tang, 2025, Extreme categories and overreaction to news, *The Review of Economic Studies* Forthcoming.
- Lettau, Martin, and Jessica A. Wachter, 2007, Why Is Long-Horizon Equity Less Risky? A Duration-Based Explanation of the Value Premium, *The Journal of Finance* 62, 55–92.
- Ma, Yueran, Tiziano Ropele, David Sraer, and David Thesmar, 2024, A quantitative analysis of distortions in managerial forecasts.
- Manski, Charles F., 2004, Measuring expectations, *Econometrica* 72, 1329–1376.
- Molavi, Pooya, Alireza Tahbaz-Salehi, and Andrea Vedolin, 2024, Model complexity, expectations, and asset prices, *Review of Economic Studies* 91, 2462–2507.

- Murphy, Kevin M., and Robert H. Topel, 1985, Estimation and inference in two-step econometric models, *Journal of Business & Economic Statistics* 3, 370–379.
- Nagel, Stefan, and Zhengyang Xu, 2022, Asset pricing with fading memory, *The Review of Financial Studies* 35, 2190–2245.
- Novy-Marx, Robert, and Mihail Velikov, 2016, A taxonomy of anomalies and their trading costs, *Review of Financial Studies* 29, 104–147.
- Real-Time Data Set for Macroeconomists, 1968–2025, Historical data files for the real-time data set for macroeconomists, Federal Reserve Bank of Philadelphia, Accessed September 2025.
- Robbins, Herbert, 1956, An empirical bayes approach to statistics, *Berkeley Symposium on Mathematical Statistics and Probability* 157—163.
- Santa-Clara, Pedro, and Pedro Barroso, 2015, Momentum has its moments, *Journal of Financial Economics* 116, 111–120.
- Singleton, Kenneth J., 2021, Presidential address: How much “rationality” is there in bond-market risk premiums?, *The Journal of Finance* 76, 1611–1654.
- Stanley, Michael, Luis Amaral, Sergey Bouldyrev, Shlomo Havlin, Heiko Leschhorn, Philipp Maass, Michael Salinger, and Eugene Stanley, 1996, Scaling behavior in the growth of companies, *Nature* 379, 804–806.
- Survey of Professional Forecasters, 1968–2025, Historical data files for the survey of professional forecasters, Federal Reserve Bank of Philadelphia, Accessed September 2025.
- Wang, Chen, 2021, Under- and overreaction in yield curve expectations, Working Paper, Mendoza College of Business, University of Notre Dame, October 2021.
- Wyart, Matthieu, and Jean-Philippe Bouchaud, 2003, Statistical models for company growth, *Physica A: Statistical Mechanics and its Applications* 326, 241–255.

# Online Appendix to “Expectations Formation with Fat-Tailed Processes: Evidence and Theory”

## Contents

<b>A</b>	<b>Additional Tables and Figures</b>	<b>43</b>
<b>B</b>	<b>Proofs and Additional Derivations</b>	<b>59</b>
B.1	Proof of Proposition 1 . . . . .	59
B.2	Proof of Proposition 2 . . . . .	59
B.3	Proof of Proposition 3 . . . . .	63
B.4	Proof of Corollary 1 . . . . .	65
B.5	Proof of Proposition 4 . . . . .	65
B.6	Additional Details on Asset Pricing Model in Section 4.3 . . . . .	66
<b>C</b>	<b>Computing Rational Expectations Using the Particle Filter</b>	<b>68</b>

## A Additional Tables and Figures

Table IA.1: Summary Statistics: Consensus-Level Forecasts

Variable	Transformation (1)	Realizations		Errors		Revisions		# of Obs. (8)
		Mean (2)	SD (3)	Mean (4)	SD (5)	Mean (6)	SD (7)	
Net Sales	Growth	0.06	0.18	-0.02	0.14	-0.00	0.12	144567
Return On Equity	Level	0.12	0.15	-0.01	0.09	-0.01	0.06	128459
Earnings Per Share	Growth	0.05	0.47	-0.09	0.39	-0.02	0.30	126341
Net Income	Growth	0.06	0.47	-0.08	0.39	-0.01	0.31	112501
Pre-tax Profit	Growth	0.05	0.45	-0.09	0.38	-0.01	0.28	106043
Book Value Per Share	Growth	0.05	0.19	-0.03	0.19	-0.01	0.13	99542
EBITDA	Growth	0.06	0.31	-0.05	0.25	-0.00	0.20	98311
Return On Assets	Level	0.05	0.06	-0.01	0.04	-0.01	0.03	89610
Net Asset Value	Growth	0.06	0.18	-0.00	0.19	-0.02	0.13	89572
Dividend Per Share	Growth	0.06	0.31	-0.01	0.28	-0.01	0.22	89522
EBIT	Growth	0.05	0.40	-0.07	0.34	-0.01	0.26	83365
GAAP EPS	Growth	0.05	0.57	-0.11	0.47	0.01	0.39	82230
Capital Expenditures	Growth	0.04	0.54	-0.12	0.61	0.19	0.61	78556
Gross Margin	Level	0.40	0.22	-0.01	0.07	-0.01	0.04	77380
Cash Flow Per Share	Growth	0.05	0.52	-0.05	0.45	-0.02	0.45	70384
Net Debt	Growth	0.04	0.42	0.09	0.43	0.07	0.38	44476
Enterprise Value	Growth	0.05	0.33	0.06	0.31	0.02	0.21	43661
Operating Profit	Growth	0.05	0.40	-0.08	0.34	-0.00	0.27	42824
Cash Earnings Per Share	Growth	0.04	0.32	-0.04	0.30	-0.03	0.26	22621
EBITDA Per Share	Growth	0.04	0.31	-0.04	0.26	-0.02	0.22	22411
Funds From Operations Per Share	Growth	0.01	0.16	-0.02	0.12	-0.03	0.11	3547
EPS Before Goodwill	Growth	0.08	0.35	-0.01	0.29	-0.04	0.25	2616

*Notes:* This table reports the mean, standard deviation, and number of observations for each one of the 22 firm variables forecasted by analysts in our sample. The sample includes all observations for which the consensus forecast error and revision are available for a given variable. Variables are ranked by decreasing number of observations. 19 variables are transformed into log growth as explained in the text; 3 are left untransformed. [Click here to go back to main text.](#)

Table IA.2: Non-Linear Error-Revision Relationship: Allowing for Asymmetric Slopes

Variable	Revision <sub>it</sub>		Revision <sub>it</sub> × Top 20% of Revision <sub>it</sub>		Revision <sub>it</sub> × Bottom 20% of Revision <sub>it</sub>		Top 20% of Revision <sub>it</sub>		Bottom 20% of Revision <sub>it</sub>		# of Obs.
	$\hat{\beta}$ (1)	$t$ -stat (2)	$\hat{\gamma}^+$ (3)	$t$ -stat (4)	$\hat{\gamma}^-$ (5)	$t$ -stat (6)	$\hat{\delta}^+$ (7)	$t$ -stat (8)	$\hat{\delta}^-$ (9)	$t$ -stat (10)	
Net Sales	0.11	6.31	-0.26	-11.33	-0.27	-12.74	0.02	12.25	-0.02	-5.72	399809
Earnings Per Share	0.08	2.97	-0.24	-6.65	-0.27	-9.64	0.01	1.48	-0.08	-7.58	360011
Return On Equity	0.23	9.75	-0.37	-8.59	-0.22	-11.03	0.01	11.30	-0.01	-6.94	345082
Net Income	0.04	1.53	-0.21	-6.16	-0.27	-12.16	0.01	1.00	-0.07	-8.05	325230
EBITDA	-0.04	-1.51	-0.11	-2.29	-0.17	-6.55	0.00	1.43	-0.03	-3.58	280111
Pre-tax Profit	0.01	0.47	-0.16	-4.03	-0.27	-11.90	0.00	1.01	-0.07	-6.67	269883
EBIT	-0.05	-1.60	-0.14	-3.09	-0.26	-11.11	0.00	0.56	-0.05	-4.61	231455
GAAP EPS	-0.02	-0.73	-0.10	-2.84	-0.18	-7.87	-0.02	-1.01	-0.07	-3.97	171523
Book Value Per Share	-0.06	-1.52	-0.49	-8.97	-0.37	-7.56	0.02	5.10	-0.03	-5.08	170561
Return On Assets	0.41	10.35	-0.70	-11.25	-0.35	-9.63	-0.00	-0.77	-0.01	-6.80	157063
Gross Margin	0.09	3.90	-0.23	-6.44	-0.12	-4.28	0.00	5.18	-0.00	-3.61	150220
Net Asset Value	0.04	0.99	-0.41	-6.85	-0.68	-13.71	0.01	3.11	-0.05	-9.43	119885
Cash Flow Per Share	-0.28	-15.86	0.04	1.47	-0.15	-6.36	-0.01	-0.94	-0.07	-5.97	112323
Net Debt	-0.17	-8.31	0.13	5.69	-0.54	-21.33	-0.06	-5.67	-0.05	-4.02	90126
Enterprise Value	-0.20	-5.82	-0.01	-0.12	-0.44	-11.09	0.02	1.80	-0.07	-7.76	88791
Dividend Per Share	-0.12	-4.54	-0.18	-4.51	-0.16	-5.66	-0.01	-0.61	-0.03	-2.89	63840
Capital Expenditures	-0.36	-21.93	-0.14	-5.98	-0.19	-6.85	0.05	3.53	-0.07	-5.62	62099
Operating Profit	0.12	3.07	-0.31	-5.97	-0.37	-7.95	0.02	3.59	-0.07	-6.42	57756
EBITDA Per Share	-0.09	-2.84	-0.15	-2.37	-0.12	-2.72	0.03	3.45	-0.02	-1.67	16824
Cash Earnings Per Share	-0.19	-4.95	-0.15	-2.12	-0.24	-4.17	0.01	0.79	-0.04	-2.11	11578
Funds From Operations Per Share	-0.01	-0.25	-0.03	-0.46	-0.11	-1.90	0.00	1.01	-0.02	-2.09	9279
EPS Before Goodwill	-0.13	-2.29	-0.18	-1.46	-0.06	-0.71	0.06	2.61	-0.06	-3.33	2558

Notes: This table reproduces Table 2 allowing for asymmetric slopes. Instead of a single interaction term  $\gamma$  for both top and bottom 20% of revisions, we estimate separate interaction coefficients  $\gamma^+$  for the top 20% and  $\gamma^-$  for the bottom 20%. We report the estimates of  $\beta$  (column 1),  $\gamma^+$  (column 3),  $\gamma^-$  (column 5),  $\delta^+$  (column 7), and  $\delta^-$  (column 9), along with the corresponding  $t$ -statistics with standard errors clustered by firm and time. Variables are ranked by number of observations. [Click here to go back to main text.](#)

Table IA.3: Non-Linear Error-Relationship: Consensus Forecasts

Error <sub><i>it</i>+1</sub>	Revision <sub><i>it</i></sub>		Revision <sub><i>it</i></sub> × Top/Bottom 20% of Revision <sub><i>it</i></sub>		Top 20% of Revision <sub><i>it</i></sub>		Bottom 20% of Revision <sub><i>it</i></sub>		# of Obs.
	$\hat{\beta}$	<i>t</i> -stat	$\hat{\gamma}$	<i>t</i> -stat	$\hat{\delta}^+$	<i>t</i> -stat	$\hat{\delta}^-$	<i>t</i> -stat	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
Net Sales	0.23	5.34	-0.40	-8.31	0.01	0.92	-0.04	-11.76	145225
Return On Equity	0.40	7.63	-0.44	-6.86	0.00	2.01	-0.03	-10.09	128761
Earnings Per Share	0.02	0.49	-0.22	-4.82	-0.03	-2.58	-0.07	-6.14	126893
Net Income	-0.08	-1.79	-0.12	-2.77	-0.03	-1.76	-0.05	-4.69	112990
Pre-tax Profit	-0.07	-1.38	-0.13	-2.69	-0.02	-1.63	-0.04	-3.18	106455
Book Value Per Share	-0.21	-7.02	-0.28	-8.62	-0.01	-3.15	-0.04	-10.06	99986
EBITDA	-0.03	-1.29	-0.18	-5.02	-0.01	-0.92	-0.05	-3.94	98819
Net Asset Value	-0.31	-9.97	-0.23	-6.44	0.01	2.89	-0.01	-1.11	89954
Dividend Per Share	-0.00	-0.02	-0.30	-4.20	-0.00	-0.22	-0.04	-4.75	89927
Return On Assets	0.46	11.50	-0.51	-11.36	-0.01	-9.72	-0.02	-14.44	89807
EBIT	-0.16	-4.82	-0.08	-2.13	-0.02	-1.83	-0.03	-2.57	83770
GAAP EPS	-0.19	-4.19	0.01	0.23	-0.02	-1.12	-0.00	-0.30	82578
Capital Expenditures	-0.40	-25.03	-0.23	-9.78	0.10	6.04	-0.09	-5.38	78871
Gross Margin	0.08	2.39	-0.27	-6.58	0.00	3.16	-0.01	-7.67	77518
Cash Flow Per Share	-0.33	-23.51	0.03	1.50	0.01	1.05	0.02	2.94	70787
Net Debt	-0.23	-7.75	-0.00	-0.13	0.11	10.77	0.14	13.20	44693
Enterprise Value	-0.10	-1.50	-0.36	-5.79	0.09	7.82	-0.01	-1.13	43855
Operating Profit	-0.01	-0.15	-0.32	-4.52	-0.00	-0.36	-0.07	-3.93	43025
Cash Earnings Per Share	-0.39	-12.04	0.01	0.12	-0.02	-1.99	-0.02	-0.90	22777
EBITDA Per Share	-0.09	-1.91	-0.16	-2.56	-0.00	-0.26	-0.03	-2.78	22528
Funds From Operations Per Share	-0.00	-0.04	-0.23	-2.29	-0.00	-0.13	-0.04	-4.77	3548
EPS Before Goodwill	-0.29	-1.90	0.08	0.48	0.00	0.30	-0.02	-0.71	2646

Notes: This table reproduces Table 2 using consensus-level forecasts. See Table 2 for additional notes. [Click here to go back to main text.](#)

Table IA.4: Non-Linear Error-Relationship: Percent Growth

Error <sub>it+1</sub>	Revision <sub>it</sub>		Revision <sub>it</sub> × Top/Bottom 20% of Revision <sub>it</sub>		Top 20% of Revision <sub>it</sub>		Bottom 20% of Revision <sub>it</sub>		# of Obs.
	$\hat{\beta}$ (1)	t-stat (2)	$\hat{\gamma}$ (3)	t-stat (4)	$\hat{\delta}^+$ (5)	t-stat (6)	$\hat{\delta}^-$ (7)	t-stat (8)	
Net Sales	0.10	5.67	-0.27	-13.42	0.03	7.86	-0.02	-3.65	399832
Earnings Per Share	0.00	0.05	-0.23	-7.17	0.06	4.11	-0.05	-2.06	364593
Net Income	-0.04	-1.67	-0.22	-7.25	0.06	4.14	-0.04	-1.89	329340
EBITDA	-0.08	-3.25	-0.13	-3.94	0.02	2.65	-0.01	-0.73	281746
Pre-tax Profit	-0.07	-3.18	-0.18	-6.45	0.05	3.26	-0.01	-0.52	273285
EBIT	-0.12	-4.30	-0.14	-3.88	0.04	3.11	0.00	0.24	234200
GAAP EPS	-0.11	-5.92	-0.15	-4.91	0.07	2.89	-0.02	-0.74	173552
Book Value Per Share	-0.11	-2.77	-0.40	-8.73	0.02	3.55	-0.04	-7.12	170925
Net Asset Value	-0.02	-0.60	-0.45	-8.97	0.04	6.39	-0.02	-2.17	120417
Cash Flow Per Share	-0.30	-17.77	-0.02	-1.12	0.00	0.37	0.00	0.07	113717
Net Debt	-0.09	-4.25	-0.04	-1.78	0.09	6.20	0.25	16.00	96338
Enterprise Value	-0.17	-4.60	-0.04	-0.89	0.05	5.12	0.06	4.64	88700
Dividend Per Share	-0.19	-6.65	-0.11	-3.04	-0.00	-0.27	-0.01	-0.61	64154
Capital Expenditures	-0.38	-25.04	-0.46	-24.24	0.29	15.58	-0.21	-15.90	62076
Operating Profit	0.02	0.79	-0.28	-6.18	0.06	3.32	-0.03	-1.76	58369
EBITDA Per Share	-0.15	-5.08	-0.12	-2.86	0.04	3.14	-0.02	-1.33	16905
Cash Earnings Per Share	-0.22	-7.03	-0.15	-2.73	0.01	0.95	-0.01	-0.27	11651
Funds From Operations Per Share	0.01	0.18	-0.14	-3.17	0.01	2.32	-0.02	-2.23	9324
EPS Before Goodwill	-0.18	-2.33	-0.18	-2.28	0.10	4.41	-0.11	-1.75	2560

Notes: This table reproduces [Table 2](#) where we change our transformation of the variables that are converted into growth rates from log growth rates to percent growth rates. We compute percent growth rates as  $g_{it} = \frac{X_{it} - X_{it-1}}{X_{it-1}}$ , and redefine forecasts for these variables as  $F_t^j g_{it+h} = \frac{F_t^j X_{it+h} - F_t^j X_{it+h-1}}{F_t^j X_{it+h-1}}$ , with the convention that  $F_t^j X_{it} = X_{it}$ . See [Table 2](#) for additional notes. [Click here to go back to main text.](#)

Table IA.5: Non-Linear Error-Relationship: Adjusting for Firm-Level Volatility

Error <sub><i>it</i>+1</sub>	Revision <sub><i>it</i></sub>		Revision <sub><i>it</i></sub> × Top/Bottom 20% of Revision <sub><i>it</i></sub>		Top 20% of Revision <sub><i>it</i></sub>		Bottom 20% of Revision <sub><i>it</i></sub>		# of Obs.
	$\hat{\beta}$	<i>t</i> -stat	$\hat{\gamma}$	<i>t</i> -stat	$\hat{\delta}^+$	<i>t</i> -stat	$\hat{\delta}^-$	<i>t</i> -stat	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
Net Sales	0.11	6.81	-0.33	-16.89	0.20	9.95	-0.16	-9.31	303623
Earnings Per Share	0.03	1.49	-0.26	-7.67	0.15	7.12	-0.12	-3.85	272283
Return On Equity	0.18	11.53	-0.31	-15.35	0.13	6.05	-0.17	-8.09	248053
Net Income	-0.02	-1.11	-0.24	-7.07	0.17	7.12	-0.10	-2.90	242091
EBITDA	-0.03	-1.67	-0.29	-11.04	0.19	7.35	-0.13	-4.85	205224
Pre-tax Profit	-0.00	-0.18	-0.32	-11.37	0.20	8.94	-0.13	-3.43	196802
EBIT	-0.06	-2.21	-0.41	-10.22	0.30	12.92	-0.21	-4.35	159223
Book Value Per Share	-0.12	-2.92	-0.40	-8.27	0.16	5.64	-0.17	-5.36	119693
GAAP EPS	-0.03	-1.06	-0.20	-5.21	0.15	4.30	-0.06	-1.97	110490
Return On Assets	0.23	8.95	-0.39	-11.66	0.00	0.07	-0.38	-9.18	95610
Gross Margin	0.11	5.23	-0.25	-11.33	0.12	5.29	-0.17	-9.62	85574
Cash Flow Per Share	-0.27	-16.27	-0.21	-9.07	0.27	11.26	-0.11	-2.68	74669
Net Asset Value	-0.07	-2.03	-0.62	-10.91	0.41	16.04	-0.24	-4.08	70815
Net Debt	-0.15	-5.53	-0.22	-5.64	0.38	8.35	0.25	6.64	58422
Enterprise Value	-0.15	-3.65	-0.46	-11.48	0.55	16.25	-0.11	-2.99	43748
Dividend Per Share	-0.21	-8.59	-0.15	-4.54	0.12	4.02	-0.07	-2.22	38038
Capital Expenditures	-0.33	-16.59	-0.13	-5.00	0.10	1.66	-0.08	-2.57	33460
Operating Profit	0.15	2.85	-0.46	-7.14	0.26	6.52	-0.12	-2.47	22761
Funds From Operations Per Share	0.01	0.18	-0.11	-1.86	0.11	3.03	-0.11	-2.27	6941
Cash Earnings Per Share	-0.25	-5.81	-0.26	-4.16	0.23	3.90	-0.28	-3.31	4483
EBITDA Per Share	-0.04	-1.08	-0.36	-6.07	0.25	4.43	-0.24	-5.92	4068

Notes: This table reproduces [Table 2](#) after normalizing forecasts and realizations by estimates of the firm-level standard deviation of the realizations. The sample here consists of firms for which we have at least 10 observations. See [Table 2](#) for additional notes. This table does not include EPS Before Goodwill because we do not have enough observations. [Click here to go back to main text.](#)

Table IA.6: Non-Linear Error-Relationship: Adjusting for Aggregate Volatility

Error <sub><i>it</i>+1</sub>	Revision <sub><i>it</i></sub>		Revision <sub><i>it</i></sub> × Top/Bottom 20% of Revision <sub><i>it</i></sub>		Top 20% of Revision <sub><i>it</i></sub>		Bottom 20% of Revision <sub><i>it</i></sub>		# of Obs.
	$\hat{\beta}$	<i>t</i> -stat	$\hat{\gamma}$	<i>t</i> -stat	$\hat{\delta}^+$	<i>t</i> -stat	$\hat{\delta}^-$	<i>t</i> -stat	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
Net Sales	0.11	5.46	-0.27	-12.06	0.06	4.90	-0.09	-7.87	399555
Earnings Per Share	0.08	2.53	-0.25	-10.00	0.02	1.26	-0.11	-5.65	359817
Return On Equity	0.21	8.94	-0.22	-8.29	0.00	1.05	-0.03	-5.32	344222
Net Income	0.03	1.42	-0.24	-13.95	0.02	1.25	-0.10	-5.90	325134
EBITDA	-0.04	-1.34	-0.14	-5.99	0.02	1.18	-0.05	-2.64	279943
Pre-tax Profit	0.00	0.09	-0.21	-10.90	0.03	1.79	-0.08	-3.97	269692
EBIT	-0.04	-1.37	-0.22	-9.49	0.04	2.50	-0.06	-2.93	231398
GAAP EPS	-0.02	-0.91	-0.13	-4.67	-0.00	-0.03	-0.06	-2.21	171482
Book Value Per Share	-0.07	-1.71	-0.41	-7.92	0.03	2.49	-0.12	-6.63	170367
Return On Assets	0.50	8.27	-0.52	-8.59	-0.08	-4.80	-0.14	-8.02	156935
Gross Margin	0.10	4.23	-0.15	-5.62	0.00	1.33	-0.02	-5.75	150124
Net Asset Value	0.02	0.62	-0.55	-11.89	0.11	8.67	-0.11	-5.41	119865
Cash Flow Per Share	-0.28	-15.30	-0.08	-4.21	0.08	4.39	-0.01	-0.52	112303
Net Debt	-0.17	-8.09	-0.12	-5.44	0.20	10.25	0.20	8.31	90095
Enterprise Value	-0.19	-5.20	-0.28	-8.53	0.26	12.27	-0.03	-1.03	88775
Dividend Per Share	-0.14	-5.46	-0.19	-7.40	0.00	0.01	-0.08	-4.87	63788
Capital Expenditures	-0.37	-23.67	-0.16	-6.75	0.12	4.93	-0.10	-4.19	62079
Operating Profit	0.12	3.29	-0.35	-8.27	0.07	3.57	-0.11	-4.80	57718
EBITDA Per Share	-0.09	-3.02	-0.13	-3.47	0.04	1.88	-0.06	-1.98	16816
Cash Earnings Per Share	-0.18	-4.54	-0.19	-3.41	0.04	1.20	-0.04	-0.90	11563
Funds From Operations Per Share	0.05	1.62	-0.15	-2.72	0.03	1.19	-0.06	-1.67	9231
EPS Before Goodwill	-0.11	-2.23	-0.10	-1.19	0.06	1.26	-0.13	-2.30	2552

Notes: This table reproduces [Table 2](#) after normalizing forecasts and realizations by estimates of the cross-sectional standard deviation of the realizations in each year. See [Table 2](#) for additional notes. [Click here to go back to main text.](#)

Table IA.7: Non-Linear Error-Relationship: Removing Time Fixed Effects

Error <sub><i>it</i>+1</sub>	Revision <sub><i>it</i></sub>		Revision <sub><i>it</i></sub> × Top/Bottom 20% of Revision <sub><i>it</i></sub>		Top 20% of Revision <sub><i>it</i></sub>		Bottom 20% of Revision <sub><i>it</i></sub>		# of Obs.
	$\hat{\beta}$ (1)	<i>t</i> -stat (2)	$\hat{\gamma}$ (3)	<i>t</i> -stat (4)	$\hat{\delta}^+$ (5)	<i>t</i> -stat (6)	$\hat{\delta}^-$ (7)	<i>t</i> -stat (8)	
Net Sales	0.08	3.35	-0.25	-11.65	0.02	6.84	-0.02	-6.34	399799
Earnings Per Share	0.07	2.31	-0.24	-8.60	0.01	0.84	-0.06	-6.39	359967
Return On Equity	-0.00	-0.00	-0.24	-0.66	0.10	1.04	-1.31	-1.42	340601
Net Income	0.02	0.87	-0.22	-10.67	0.01	0.90	-0.05	-5.34	325194
EBITDA	-0.03	-1.23	-0.15	-6.43	0.01	1.25	-0.02	-3.11	280094
Pre-tax Profit	-0.01	-0.30	-0.19	-7.03	0.02	1.87	-0.04	-3.76	269840
EBIT	-0.04	-1.39	-0.23	-9.58	0.03	3.25	-0.02	-2.40	231446
GAAP EPS	-0.05	-0.98	-0.10	-2.08	-0.01	-0.94	-0.04	-2.81	171544
Book Value Per Share	-0.10	-2.11	-0.39	-7.23	0.02	5.64	-0.03	-6.55	170543
Return On Assets	0.21	3.08	-0.24	-3.20	-0.00	-1.61	-0.01	-4.42	157075
Gross Margin	0.09	4.65	-0.14	-6.09	0.00	2.75	-0.00	-6.62	150215
Net Asset Value	-0.03	-0.65	-0.49	-8.02	0.04	11.91	-0.02	-3.21	119887
Cash Flow Per Share	-0.25	-16.94	-0.09	-5.15	0.05	4.70	-0.00	-0.30	112347
Net Debt	-0.15	-6.46	-0.14	-6.44	0.11	10.47	0.11	11.70	90108
Enterprise Value	-0.20	-5.78	-0.26	-7.32	0.09	10.94	-0.02	-2.10	88772
Dividend Per Share	-0.16	-4.84	-0.15	-4.92	-0.00	-0.59	-0.03	-3.87	63847
Capital Expenditures	-0.37	-21.85	-0.16	-7.53	0.04	4.03	-0.06	-6.36	62094
Operating Profit	0.04	0.98	-0.26	-6.43	0.04	3.97	-0.04	-5.15	57758
EBITDA Per Share	-0.04	-0.91	-0.16	-3.31	0.01	0.80	-0.02	-1.36	16820
Cash Earnings Per Share	-0.21	-5.63	-0.18	-3.36	0.04	2.56	-0.02	-1.01	11578
Funds From Operations Per Share	-0.01	-0.15	-0.10	-1.74	0.01	1.45	-0.01	-1.32	9280
EPS Before Goodwill	-0.13	-2.12	-0.07	-0.92	0.01	0.65	-0.06	-2.40	2558

Notes: This table reproduces Table 2 after removing the cross-sectional mean from forecast errors and forecast revisions, separately, for each year. See Table 2 for additional notes. [Click here to go back to main text.](#)

Table IA.8: Non-Linear Error-Relationship: Split by Analyst Experience

Quartile of Past Experience	1		2		3		4	
Variable	$\hat{\gamma}$ (1)	$t$ -stat (2)	$\hat{\gamma}$ (3)	$t$ -stat (4)	$\hat{\gamma}$ (5)	$t$ -stat (6)	$\hat{\gamma}$ (7)	$t$ -stat (8)
Net Sales	-0.28	-16.28	-0.28	-14.70	-0.25	-11.09	-0.25	-7.54
Earnings Per Share	-0.27	-8.91	-0.25	-6.12	-0.27	-7.83	-0.26	-6.78
Return On Equity	-0.27	-9.27	-0.30	-12.03	-0.27	-9.72	-0.22	-4.86
Net Income	-0.24	-8.63	-0.24	-6.00	-0.26	-10.26	-0.24	-8.32
EBITDA	-0.16	-5.06	-0.15	-4.81	-0.13	-3.53	-0.13	-4.18
Pre-tax Profit	-0.26	-7.28	-0.20	-5.70	-0.25	-8.57	-0.19	-6.45
EBIT	-0.23	-7.12	-0.19	-6.82	-0.21	-4.97	-0.24	-6.00
GAAP EPS	-0.13	-3.66	-0.12	-3.58	-0.14	-4.52	-0.15	-3.32
Book Value Per Share	-0.40	-8.37	-0.40	-6.14	-0.42	-7.13	-0.46	-5.52
Return On Assets	-0.48	-9.55	-0.43	-6.74	-0.43	-7.70	-0.37	-7.01
Gross Margin	-0.16	-4.52	-0.15	-2.76	-0.14	-2.91	-0.17	-3.91
Net Asset Value	-0.59	-14.20	-0.57	-8.99	-0.49	-8.11	-0.59	-6.65
Cash Flow Per Share	-0.03	-0.71	-0.05	-1.96	-0.08	-3.10	-0.10	-2.18
Net Debt	-0.13	-3.98	-0.08	-2.41	-0.11	-2.15	-0.18	-5.92
Enterprise Value	-0.23	-4.56	-0.26	-5.86	-0.30	-8.44	-0.26	-5.61
Dividend Per Share	-0.22	-4.31	-0.15	-3.49	-0.13	-2.44	-0.16	-2.90
Capital Expenditures	-0.15	-6.30	-0.13	-3.79	-0.10	-2.82	-0.26	-6.44
Operating Profit	-0.33	-7.17	-0.32	-4.20	-0.36	-5.86	-0.39	-5.49
EBITDA Per Share	-0.22	-2.89	-0.02	-0.29	-0.14	-1.97	-0.11	-1.70
Cash Earnings Per Share	-0.23	-3.43	-0.12	-0.86	-0.24	-1.94	-0.15	-3.34
Funds From Operations Per Share	0.03	0.33	-0.14	-1.80	-0.09	-1.08	-0.16	-1.34
EPS Before Goodwill	-0.02	-0.22	-0.03	-0.17	-0.06	-0.22	-0.30	-2.31

Notes: Each pair of columns in this table reports the OLS estimates of equation (2), for each one of the 22 variables on a different subsample based on analyst experience. For each variable, the experience of analyst  $j$  is defined at date  $t$  as the number of firm-years for which analyst  $j$  has issued a forecast for this variable prior to  $t$ . The table reports estimates of  $\gamma$ , along with the corresponding  $t$ -statistics with standard errors clustered by firm and time. Variables are ranked by number of observations. [Click here to go back to main text.](#)

Table IA.9: Fat Tails in Realizations: Adjusting for Firm-Level Volatility

Variable	Transformation	Kurtosis	Log-Log regression		# of Obs.
			Tail Parameter	$R^2$	
	(1)	(2)	(3)	(4)	(5)
Net Sales	Growth	3.40	4.14	0.99	87500
Earnings Per Share	Growth	3.70	4.17	0.98	80346
Return On Equity	Level	3.76	4.15	1.00	73100
Net Income	Growth	3.60	4.25	0.98	69126
Pre-tax Profit	Growth	3.66	4.13	0.98	59902
EBITDA	Growth	3.40	4.40	0.98	59770
Book Value Per Share	Growth	3.64	3.94	0.99	50930
EBIT	Growth	3.49	4.46	0.98	44622
Return On Assets	Level	4.01	3.92	0.99	38226
GAAP EPS	Growth	3.47	4.59	0.98	36745
Cash Flow Per Share	Growth	3.23	4.58	0.98	33233
Net Asset Value	Growth	3.32	3.86	0.99	33099
Gross Margin	Level	5.38	3.17	0.99	27309
Net Debt	Growth	3.80	3.76	0.98	24395
Dividend Per Share	Growth	4.19	3.93	0.99	23469
Enterprise Value	Growth	2.80	5.07	0.98	18837
Capital Expenditures	Growth	2.80	5.26	0.98	17345
Operating Profit	Growth	3.60	4.24	0.98	10142
EBITDA Per Share	Growth	3.19	4.92	0.99	3089
Cash Earnings Per Share	Growth	3.10	5.08	0.97	2975
Funds From Operations Per Share	Growth	4.02	3.88	0.97	2250
Normal Distribution	.	3	$\infty$	.	.

Notes: This table reproduces [Table 3](#) after normalizing forecasts and realizations by estimates of the firm-level standard deviation of the realizations. The sample here consists of firms for which we have at least 10 observations. See [Table 3](#) for additional notes. [Click here to go back to main text.](#)

Table IA.10: Fat Tails in Realizations: Adjusting for Aggregate Volatility

Variable	Transformation	Kurtosis	Log-Log regression		# of Obs.
			Tail Parameter	$R^2$	
	(1)	(2)	(3)	(4)	(5)
Net Sales	Growth	5.84	2.70	0.99	140481
Earnings Per Share	Growth	6.34	2.42	0.99	129735
Return On Equity	Level	7.62	2.76	1.00	127268
Net Income	Growth	6.28	2.42	0.99	113861
Pre-tax Profit	Growth	6.18	2.47	0.99	102386
EBITDA	Growth	6.07	2.53	0.99	99731
Book Value Per Share	Growth	8.38	2.13	1.00	89882
EBIT	Growth	6.36	2.48	0.99	82099
Return On Assets	Level	6.02	3.16	0.99	79477
GAAP EPS	Growth	5.88	2.57	0.99	74212
Net Asset Value	Growth	7.84	2.17	0.99	71494
Gross Margin	Level	2.47	7.21	0.98	64481
Cash Flow Per Share	Growth	5.52	2.64	0.99	62446
Dividend Per Share	Growth	6.87	2.50	0.99	49401
Net Debt	Growth	6.57	2.33	0.99	48338
Enterprise Value	Growth	4.44	3.12	0.99	48140
Capital Expenditures	Growth	4.96	2.82	0.99	44623
Operating Profit	Growth	6.08	2.54	0.99	33033
EBITDA Per Share	Growth	6.19	2.38	0.99	15344
Cash Earnings Per Share	Growth	6.01	2.44	0.99	10012
Funds From Operations Per Share	Growth	7.11	2.33	1.00	3399
EPS Before Goodwill	Growth	5.70	3.07	0.98	2413
Normal Distribution	.	3	$\infty$	.	.

Notes: This table reproduces [Table 3](#) after normalizing realizations by the yearly cross-sectional standard deviation of the realizations. See [Table 2](#) for additional notes. [Click here to go back to main text.](#)

Table IA.11: Fat Tails in Realizations: Percent Growth

Variable	Transformation	Kurtosis	Log-Log regression		# of Obs.
			Tail Parameter	$R^2$	
	(1)	(2)	(3)	(4)	(5)
Net Sales	Growth	8.27	2.35	1.00	140482
Earnings Per Share	Growth	25.66	1.29	1.00	129736
Net Income	Growth	27.02	1.29	1.00	113862
Pre-tax Profit	Growth	23.34	1.36	1.00	102386
EBITDA	Growth	14.73	1.89	1.00	99730
Book Value Per Share	Growth	9.80	2.26	0.99	89884
EBIT	Growth	19.63	1.52	1.00	82100
GAAP EPS	Growth	28.02	1.16	1.00	74221
Net Asset Value	Growth	11.92	2.08	1.00	71494
Cash Flow Per Share	Growth	23.00	1.31	1.00	62446
Dividend Per Share	Growth	15.61	1.66	0.99	49401
Net Debt	Growth	28.51	1.17	1.00	48339
Enterprise Value	Growth	9.01	2.06	1.00	48140
Capital Expenditures	Growth	16.21	1.60	1.00	44624
Operating Profit	Growth	19.50	1.55	1.00	33033
EBITDA Per Share	Growth	13.53	2.01	0.99	15344
Cash Earnings Per Share	Growth	11.62	2.05	0.99	10013
Funds From Operations Per Share	Growth	7.36	2.29	1.00	3399
EPS Before Goodwill	Growth	23.33	1.63	0.99	2414
Normal Distribution	.	3	$\infty$	.	.

Notes: This table reproduces [Table 3](#) where we change our transformation of the variables that are converted into growth rates from log growth rates to percent growth rates. We compute percent growth rates as  $g_{it} = \frac{X_{it} - X_{it-1}}{X_{it-1}}$ . We do not report results for the three “level variables” which are not transformed anyway. See [Table 3](#) for additional notes. [Click here to go back to main text.](#)

Table IA.12: Non-Linear Conditional Expectation of Realizations: Adjusting for Firm-Level Volatility

$g_{it+1}$	$g_{it}$		$g_{it} \times$ Top/Bottom 20% of $g_{it}$		Top 20% of $g_{it}$		Bottom 20% of $g_{it}$		# of Obs.
Variable	$\hat{\beta}_g$ (1)	$t$ -stat (2)	$\hat{\gamma}_g$ (3)	$t$ -stat (4)	$\hat{\delta}_g^+$ (5)	$t$ -stat (6)	$\hat{\delta}_g^-$ (7)	$t$ -stat (8)	
Net Sales	0.43	11.87	-0.36	-6.89	0.63	6.90	0.10	3.80	87500
Earnings Per Share	0.30	9.77	-0.53	-12.58	0.77	12.90	0.10	2.42	80346
Return On Equity	0.89	74.58	-0.03	-2.00	0.22	2.73	0.13	6.02	73100
Net Income	0.30	10.46	-0.51	-11.63	0.75	10.10	0.12	2.72	69126
Pre-tax Profit	0.30	9.28	-0.54	-13.47	0.79	12.29	0.09	2.39	59902
EBITDA	0.28	9.90	-0.40	-10.64	0.64	9.13	0.10	3.04	59770
Book Value Per Share	0.51	17.75	-0.45	-13.20	0.76	14.56	0.16	6.28	50930
EBIT	0.33	10.13	-0.53	-10.57	0.81	9.82	0.08	2.00	44622
Return On Assets	0.87	51.63	0.06	3.60	-0.20	-2.41	0.06	2.29	38226
GAAP EPS	0.12	3.74	-0.46	-11.98	0.72	11.63	0.11	2.33	36745
Cash Flow Per Share	-0.13	-3.89	-0.16	-4.44	0.36	9.43	0.14	5.13	33233
Net Asset Value	0.49	15.54	-0.40	-9.22	0.74	10.05	0.06	1.71	33099
Gross Margin	1.00	293.41	-0.01	-2.07	0.07	2.78	0.21	6.66	27309
Net Debt	0.12	4.18	-0.23	-6.82	0.34	9.23	0.09	2.23	24395
Dividend Per Share	0.44	10.73	-0.51	-9.41	0.85	9.68	0.15	4.85	23469
Enterprise Value	-0.03	-0.66	-0.16	-2.64	0.28	3.48	0.02	0.31	18837
Capital Expenditures	0.03	0.87	-0.11	-2.27	0.15	2.10	-0.07	-1.05	17345
Operating Profit	0.22	4.03	-0.48	-6.28	0.77	7.66	0.06	1.03	10142
EBITDA Per Share	0.30	4.89	-0.29	-2.81	0.55	2.64	0.22	1.88	3089
Cash Earnings Per Share	0.12	1.97	-0.29	-2.53	0.52	2.88	0.19	1.81	2975
Funds From Operations Per Share	0.39	5.53	-0.39	-3.50	0.51	3.84	-0.10	-1.01	2250

Notes: This table reproduces Table 4 after normalizing forecasts and realizations by estimates of the firm-level standard deviation of the realizations. The sample here consists of firms for which we have at least 10 observations. This explains why “EPS Before Goodwill” is not reported here (there are not enough observations). See Table 4 for additional notes. [Click here to go back to main text.](#)

Table IA.13: Non-Linear Conditional Expectation of Realizations: Adjusting for Aggregate Volatility

$g_{it+1}$	$g_{it}$		$g_{it} \times$ Top/Bottom 20% of $g_{it}$		Top 20% of $g_{it}$		Bottom 20% of $g_{it}$		# of Obs.
Variable	$\hat{\beta}_g$ (1)	$t$ -stat (2)	$\hat{\gamma}_g$ (3)	$t$ -stat (4)	$\hat{\delta}_g^+$ (5)	$t$ -stat (6)	$\hat{\delta}_g^-$ (7)	$t$ -stat (8)	
Net Sales	0.42	12.19	-0.41	-8.60	0.32	6.58	0.01	0.56	140481
Earnings Per Share	0.20	5.75	-0.47	-15.52	0.32	10.57	-0.03	-1.04	129735
Return On Equity	0.44	2.30	0.08	0.41	0.02	0.31	-0.03	-0.68	127268
Net Income	0.19	5.96	-0.45	-13.33	0.33	9.35	-0.02	-0.63	113861
Pre-tax Profit	0.19	5.25	-0.43	-10.98	0.34	10.26	-0.01	-0.28	102386
EBITDA	0.22	7.38	-0.39	-11.49	0.31	8.89	-0.01	-0.38	99731
Book Value Per Share	0.47	14.84	-0.51	-13.69	0.23	10.52	-0.08	-5.45	89882
EBIT	0.24	5.58	-0.44	-10.61	0.30	8.28	-0.01	-0.36	82099
Return On Assets	0.64	4.03	-0.04	-0.20	0.10	0.72	0.04	0.77	79477
GAAP EPS	0.02	0.34	-0.37	-8.35	0.28	10.25	0.00	0.01	74212
Net Asset Value	0.47	13.30	-0.45	-12.15	0.24	12.54	-0.08	-4.83	71494
Gross Margin	0.90	16.42	-0.23	-1.65	0.69	1.73	0.53	1.80	64481
Cash Flow Per Share	-0.18	-6.21	-0.13	-4.34	0.18	7.33	0.07	3.56	62446
Dividend Per Share	0.43	7.51	-0.65	-11.34	0.35	9.69	-0.06	-2.07	49401
Net Debt	0.08	4.18	-0.27	-8.77	0.24	10.35	0.00	0.18	48338
Enterprise Value	-0.06	-1.19	-0.09	-1.43	0.13	2.35	-0.03	-0.98	48140
Capital Expenditures	0.06	1.67	-0.22	-5.86	0.11	4.71	-0.10	-3.42	44623
Operating Profit	0.24	4.77	-0.49	-10.77	0.43	10.51	-0.02	-0.46	33033
EBITDA Per Share	0.12	2.67	-0.24	-3.11	0.13	2.12	0.01	0.09	15344
Cash Earnings Per Share	0.08	2.12	-0.34	-5.68	0.21	4.36	-0.03	-0.86	10012
Funds From Operations Per Share	0.46	4.25	-0.46	-3.72	0.22	2.09	-0.09	-1.08	3399
EPS Before Goodwill	0.21	2.09	-0.54	-3.97	0.40	3.45	-0.08	-1.00	2413

Notes: This table reproduces Table 4 after normalizing realizations by yearly cross-sectional dispersion of realizations. See Table 4 for additional notes. [Click here to go back to main text.](#)

Table IA.14: Non-Linear Conditional Expectation of Realizations: Percent Growth

$g_{it+1}$	$g_{it}$		$g_{it} \times$ Top/Bottom 20% of $g_{it}$		Top 20% of $g_{it}$		Bottom 20% of $g_{it}$		# of Obs.
Variable	$\hat{\beta}_g$ (1)	$t$ -stat (2)	$\hat{\gamma}_g$ (3)	$t$ -stat (4)	$\hat{\delta}_g^+$ (5)	$t$ -stat (6)	$\hat{\delta}_g^-$ (7)	$t$ -stat (8)	
Net Sales	0.44	12.74	-0.39	-8.69	0.11	7.31	0.02	2.17	140482
Earnings Per Share	0.14	4.50	-0.17	-5.25	0.14	7.03	0.35	8.50	129736
Net Income	0.14	5.26	-0.17	-6.14	0.15	7.09	0.37	8.16	113862
Pre-tax Profit	0.19	6.40	-0.20	-7.08	0.15	7.04	0.33	7.33	102386
EBITDA	0.21	6.68	-0.24	-6.75	0.12	7.32	0.13	5.53	99730
Book Value Per Share	0.45	14.10	-0.52	-15.55	0.10	14.80	-0.01	-2.31	89884
EBIT	0.23	6.87	-0.25	-6.98	0.13	5.89	0.23	5.69	82100
GAAP EPS	-0.05	-1.54	0.01	0.40	0.08	3.16	0.60	9.23	74221
Net Asset Value	0.46	14.08	-0.45	-13.88	0.10	18.03	-0.01	-1.19	71494
Cash Flow Per Share	-0.20	-6.26	0.16	5.62	0.01	0.79	0.51	15.81	62446
Dividend Per Share	0.36	7.92	-0.45	-8.44	0.17	9.34	0.13	4.57	49401
Net Debt	0.07	2.47	-0.10	-2.97	0.11	6.62	0.30	11.81	48339
Enterprise Value	-0.06	-1.19	0.00	0.03	0.05	1.99	0.09	6.29	48140
Capital Expenditures	0.04	1.37	-0.10	-3.37	0.09	8.06	0.11	4.57	44624
Operating Profit	0.19	4.49	-0.22	-5.24	0.17	7.36	0.25	7.34	33033
EBITDA Per Share	0.11	3.09	-0.13	-2.72	0.07	2.51	0.13	2.90	15344
Cash Earnings Per Share	0.08	2.01	-0.20	-3.90	0.10	4.60	0.14	5.47	10013
Funds From Operations Per Share	0.34	3.53	-0.32	-2.54	0.05	2.45	-0.01	-0.39	3399
EPS Before Goodwill	0.13	2.37	-0.16	-2.37	0.09	1.43	0.17	3.53	2414

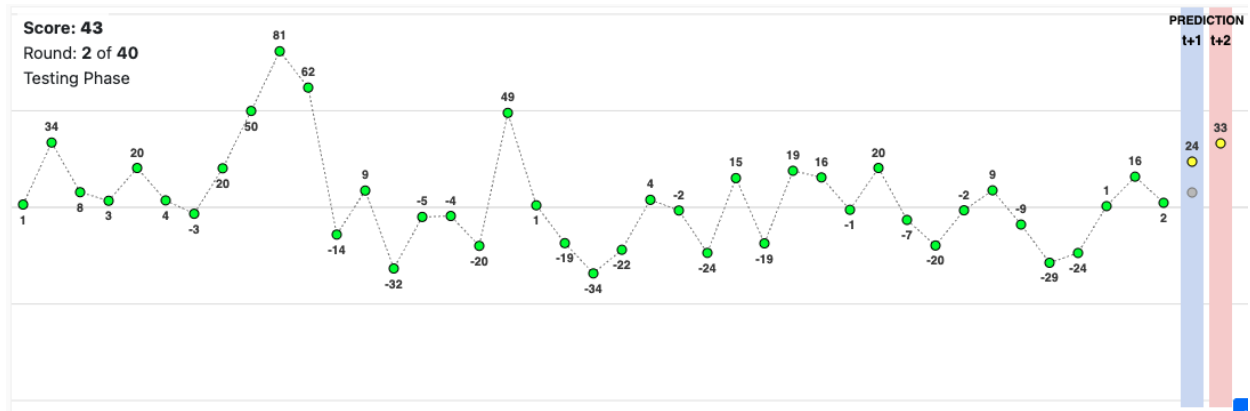
Notes: This table reproduces [Table 4](#) where we change our transformation of the variables that are converted into growth rates from log growth rates to percent growth rates. We compute percent growth rates as  $g_{it} = \frac{X_{it} - X_{it-1}}{X_{it-1}}$ . We do not report results for the three “level variables” which are not transformed anyway. See [Table 4](#) for additional notes. [Click here to go back to main text.](#)

Table IA.15: Fit of a Gaussian Approximation in the Bulk of the Distribution

Variable	$R^2$ (1)	# of Obs. (2)
Net Sales	0.954	87500
Earnings Per Share	0.939	80345
Return On Equity	0.831	73100
Net Income	0.949	69126
Pre-tax Profit	0.940	59903
EBITDA	0.951	59771
Book Value Per Share	0.950	50929
EBIT	0.941	44620
Return On Assets	0.717	38226
GAAP EPS	0.908	36745
Cash Flow Per Share	0.929	33233
Net Asset Value	0.902	33099
Gross Margin	0.508	27309
Net Debt	0.894	24396
Dividend Per Share	0.893	23469
Enterprise Value	0.776	18838
Capital Expenditures	0.826	17344
Operating Profit	0.771	10142
EBITDA Per Share	0.631	3089
Cash Earnings Per Share	0.600	2976
Funds From Operations Per Share	0.583	2249

*Notes:* This table tests whether the central portion of the distribution of realizations is approximately Gaussian. For each variable,  $x$ , we keep one observation per firm-year. We restrict the sample to observations between the 20th and 80th percentiles. We then compute the empirical log-PDF for each centile bin  $c$  as  $p_c = \log \left( \frac{1}{100} \frac{1}{x_c^+ - x_c^-} \right)$ , where  $x_c^+$  is the upper bound of  $c$  and  $x_c^-$  is the lower bound of  $c$ . We compute  $\hat{x}_c$  as the mean value of  $x$  in centile  $c$ . We then regress  $p_c$  on  $\hat{x}_c$  and  $\hat{x}_c^2$ . For a Gaussian distribution, the  $R^2$  of this regression should equal to 1. Variables are ranked by number of observations. [Click here to go back to main text.](#)

Figure IA.1: Screenshot of the Experiment Interface



Notes: This figure shows a screenshot of our experiment. [Click here to go back to main text.](#)

## B Proofs and Additional Derivations

In this appendix, we prove all of the results in the main text. We prove our results under a weaker version of Assumption 1, stated below.

**Assumption IA.1.** *The distribution of  $\epsilon_t$  is symmetric and regularly varying:*

$$P(\epsilon_t > x) = \int_x^\infty f_\epsilon(\epsilon) d\epsilon \sim x^{-\nu} L(x) \text{ as } x \rightarrow \infty,$$

where  $\nu > 2$  is the index and  $L(x)$  is slowly varying in the sense that, for all  $c$ ,  $\frac{L(cx)}{L(x)} \xrightarrow{x \rightarrow \infty} 1$ .

### B.1 Proof of Proposition 1

*Proof.* From the definition of the process, we obtain:

$$E(g_{t+1}|g_t) = \rho E(g_t^*|g_t)$$

We now apply Tweedie's formula (Robbins, 1956; Efron, 2012), which states that, for  $y = x + \eta$ , where  $x$  is Gaussian and  $x$  and  $\eta$  are independent, then:

$$E(x|y) = -\sigma_x^2 \frac{d \log h(y)}{dy}$$

where  $h$  is the marginal distribution of  $y$  and  $\sigma_x^2$  the variance of  $x$ . Combining the last two equations with  $x = g_t^*$  and  $\eta = \epsilon_t$  delivers the result.  $\square$

### B.2 Proof of Proposition 2

*Proof.* By the definition of a conditional expectation, we have:

$$E(g^*|g) = \frac{\int x f_\epsilon\left(\frac{g-x}{\sigma_\epsilon}\right) \Phi(x) dx}{\int f_\epsilon\left(\frac{g-x}{\sigma_\epsilon}\right) \Phi(x) dx} = \frac{\int x \frac{f_\epsilon\left(\frac{g-x}{\sigma_\epsilon}\right)}{f_\epsilon\left(\frac{g}{\sigma_\epsilon}\right)} \Phi(x) dx}{\int \frac{f_\epsilon\left(\frac{g-x}{\sigma_\epsilon}\right)}{f_\epsilon\left(\frac{g}{\sigma_\epsilon}\right)} \Phi(x) dx}. \quad (\text{IA.1})$$

Given Assumption IA.1, we have

$$\frac{f_\epsilon\left(\frac{g-x}{\sigma_\epsilon}\right)}{f_\epsilon\left(\frac{g}{\sigma_\epsilon}\right)} = \frac{(g-x)^{-\nu-1} L_1\left(\frac{g-x}{\sigma_\epsilon}\right)}{g^{-\nu-1} L_1\left(\frac{g}{\sigma_\epsilon}\right)} \quad (\text{IA.2})$$

$$= \left(1 - \frac{x}{g}\right)^{-\nu-1} \frac{L_1\left(\frac{g-x}{\sigma_\epsilon}\right)}{L_1\left(\frac{g}{\sigma_\epsilon}\right)}. \quad (\text{IA.3})$$

where the first equality follows from the assumption that  $\epsilon$  is regularly varying for some slowly varying  $L_1(\cdot)$ . Now we will derive some bounds on each component of the previous expression, with the goal of deriving bounds on  $\frac{f_\epsilon\left(\frac{g-x}{\sigma_\epsilon}\right)}{f_\epsilon\left(\frac{g}{\sigma_\epsilon}\right)}$ . By the binomial expansion, we have:

$$\left(1 - \frac{x}{g}\right)^{-\nu-1} = 1 + (\nu+1)\frac{x}{g} + O\left(\frac{x^2}{g^2}\right).$$

When  $|x| \leq |g|^\beta$  and  $\beta \in (0, 1/2)$ , this remainder term is  $O(g^{2\beta-2}) = o\left(\frac{1}{g}\right)$ . Next, using the Uniform Convergence Theorem for slowly varying functions (Bingham et al., 1987, Theorem 1.2.1), we have

$$\lim_{|g| \rightarrow \infty} \sup_{|x| \leq |g|^\beta} \left| \frac{L_1\left(\frac{g-x}{\sigma_\epsilon}\right)}{L_1\left(\frac{g}{\sigma_\epsilon}\right)} \right| = 1.$$

Multiplying the previous two results, we have, for  $|x| \leq |g|^\beta$ ,

$$\frac{f_\epsilon\left(\frac{g-x}{\sigma_\epsilon}\right)}{f_\epsilon\left(\frac{g}{\sigma_\epsilon}\right)} = 1 + (\nu+1)\frac{x}{g} + o\left(\frac{1}{g}\right).$$

Next, consider the region  $|g/2| > |x| > |g|^\beta$ . Taylor's theorem implies that, for some constant  $C$ ,

$$\left| \left(1 - \frac{x}{g}\right)^{-\nu-1} - 1 - (\nu+1)\frac{x}{g} \right| \leq C \left(\frac{x}{g}\right)^2$$

Since  $L_1$  is slowly varying, the Uniform Convergence Theorem implies that

$$\sup_{|x| \leq |g|/2} \left| \frac{L_1\left(\frac{g-x}{\sigma_\epsilon}\right)}{L_1\left(\frac{g}{\sigma_\epsilon}\right)} - 1 \right| \rightarrow 0 \quad \text{as } |g| \rightarrow \infty.$$

Hence, uniformly for  $|x| \leq |g|/2$ ,

$$\frac{L_1\left(\frac{g-x}{\sigma_\epsilon}\right)}{L_1\left(\frac{g}{\sigma_\epsilon}\right)} = 1 + o(1).$$

Combining these results, we obtain

$$\frac{f_\epsilon\left(\frac{g-x}{\sigma_\epsilon}\right)}{f_\epsilon\left(\frac{g}{\sigma_\epsilon}\right)} = \left(1 - \frac{x}{g}\right)^{-\nu-1} \frac{L_1\left(\frac{g-x}{\sigma_\epsilon}\right)}{L_1\left(\frac{g}{\sigma_\epsilon}\right)}$$

$$= 1 + (\nu + 1) \frac{x}{g} + O\left(\frac{x^2}{g^2}\right) + o(1),$$

where the  $o(1)$  term is uniform over  $|x| \leq |g|/2$ . Therefore, for  $|g|^\beta < |x| < |g|/2$ ,

$$\left| \frac{f_\epsilon\left(\frac{g-x}{\sigma_\epsilon}\right)}{f_\epsilon\left(\frac{g}{\sigma_\epsilon}\right)} - 1 - (\nu + 1) \frac{x}{g} \right| \leq C' \frac{x^2}{g^2} + o(1),$$

for some constant  $C' > 0$ .

Finally, we consider the region  $|x| \geq |g|/2$ . In this region, Taylor's theorem does not give a uniform bound because the remainder term approaches  $\infty$  as  $|x| \rightarrow |g|$ . However, since  $\epsilon$  is regularly varying and has a bounded density, we have

$$\frac{f_\epsilon\left(\frac{g-x}{\sigma_\epsilon}\right)}{f_\epsilon\left(\frac{g}{\sigma_\epsilon}\right)} \leq B|g|^{\nu+1},$$

for some  $B > 0$ .

Next, return to the integrals in (IA.1). Focusing on the denominator, we have

$$\begin{aligned} \int \frac{f_\epsilon\left(\frac{g-x}{\sigma_\epsilon}\right)}{f_\epsilon\left(\frac{g}{\sigma_\epsilon}\right)} \Phi(x) dx &= \int \left(1 + (\nu + 1) \frac{x}{g}\right) \Phi(x) dx + \int \left(\frac{f_\epsilon\left(\frac{g-x}{\sigma_\epsilon}\right)}{f_\epsilon\left(\frac{g}{\sigma_\epsilon}\right)} - 1 - (\nu + 1) \frac{x}{g}\right) \Phi(x) dx \\ &= \int \left(1 + (\nu + 1) \frac{x}{g}\right) \Phi(x) dx + \int_{|x| \leq g^\beta} \left(\frac{f_\epsilon\left(\frac{g-x}{\sigma_\epsilon}\right)}{f_\epsilon\left(\frac{g}{\sigma_\epsilon}\right)} - 1 - (\nu + 1) \frac{x}{g}\right) \Phi(x) dx \\ &\quad + \int_{|x| > g^\beta} \left(\frac{f_\epsilon\left(\frac{g-x}{\sigma_\epsilon}\right)}{f_\epsilon\left(\frac{g}{\sigma_\epsilon}\right)} - 1 - (\nu + 1) \frac{x}{g}\right) \Phi(x) dx \end{aligned}$$

where the final equality just breaks up the integral. Focusing on the first integral, we have

$$\begin{aligned} \left| \int_{|x| \leq g^\beta} \left(\frac{f_\epsilon\left(\frac{g-x}{\sigma_\epsilon}\right)}{f_\epsilon\left(\frac{g}{\sigma_\epsilon}\right)} - 1 - (\nu + 1) \frac{x}{g}\right) \Phi(x) dx \right| &\leq \sup_{|x| \leq g^\beta} \left(\frac{f_\epsilon\left(\frac{g-x}{\sigma_\epsilon}\right)}{f_\epsilon\left(\frac{g}{\sigma_\epsilon}\right)} - 1 - (\nu + 1) \frac{x}{g}\right) \int_{|x| \leq g^\beta} \Phi(x) dx \\ &= o\left(\frac{1}{g}\right) \end{aligned}$$

where the second equality follows from the bounds above. Turning to the second integral, we split it into two regions based on the bounds above. The first of these two regions is  $|x| \in (g^\beta, |g|/2) \equiv \mathcal{R}$ . In this region,

we have

$$\begin{aligned}
\left| \int_{|x| \in \mathcal{R}} \left( \frac{f_\epsilon \left( \frac{g-x}{\sigma_\epsilon} \right)}{f_\epsilon \left( \frac{g}{\sigma_\epsilon} \right)} - 1 - (\nu+1) \frac{x}{g} \right) \Phi(x) dx \right| &\leq \int_{|x| \in \mathcal{R}} \left| \frac{f_\epsilon \left( \frac{g-x}{\sigma_\epsilon} \right)}{f_\epsilon \left( \frac{g}{\sigma_\epsilon} \right)} - 1 - (\nu+1) \frac{x}{g} \right| \Phi(x) dx \\
&\leq \frac{C'}{g^2} \int_{|x| \in \mathcal{R}} x^2 \Phi(x) dx + o(1) \int_{|x| \in \mathcal{R}} \Phi(x) dx \\
&= o\left(\frac{1}{g}\right)
\end{aligned}$$

where the first line follows from the triangle inequality, the second inequality follows from the bound established above, and the final line follows because a Gaussian PDF implies  $\int_{|x| > g^\beta} x^k \Phi(x) dx \rightarrow 0$  for any finite  $k$  as  $|g| \rightarrow \infty$ .

For the second region,  $|x| \geq |g|/2$ , we have

$$\begin{aligned}
\left| \int_{|x| \geq |g|/2} \left( \frac{f_\epsilon \left( \frac{g-x}{\sigma_\epsilon} \right)}{f_\epsilon \left( \frac{g}{\sigma_\epsilon} \right)} - 1 - (\nu+1) \frac{x}{g} \right) \Phi(x) dx \right| &\leq \int_{|x| \geq |g|/2} \left( B|g|^{\nu+1} + 1 + (\nu+1) \frac{|x|}{|g|} \right) \Phi(x) dx \\
&= O\left(|g|^{\nu+1} e^{-g^2/(8\sigma_{g^*}^2)}\right) \\
&= o\left(\frac{1}{g}\right)
\end{aligned}$$

where the first line follows from the bounds above, the second line follows from Gaussian tail bounds, and the third line follows from the fact that exponential decay dominates polynomial growth.

Putting together the expressions for these two integrals, we have

$$\int \frac{f_\epsilon \left( \frac{g-x}{\sigma_\epsilon} \right)}{f_\epsilon \left( \frac{g}{\sigma_\epsilon} \right)} \Phi(x) dx = \int \left( 1 + (\nu+1) \frac{x}{g} \right) \Phi(x) dx + o\left(\frac{1}{g}\right).$$

We can then use the same arguments to show that

$$\int x \frac{f_\epsilon \left( \frac{g-x}{\sigma_\epsilon} \right)}{f_\epsilon \left( \frac{g}{\sigma_\epsilon} \right)} \Phi(x) dx = \int x \left( 1 + (\nu+1) \frac{x}{g} \right) \Phi(x) dx + o\left(\frac{1}{g}\right).$$

Combining the previous two results and using the fact that  $\int \Phi(x) dx = 1$ ,  $\int x \Phi(x) dx = 0$ , and  $\int x^2 \Phi(x) dx = \sigma_{g^*}^2$ , we have, as  $|g| \rightarrow \infty$ ,

$$E(g^* | g) \sim \frac{(\nu+1)\sigma_{g^*}^2}{g}.$$

Applying the data-generating process delivers the desired result.  $\square$

### B.3 Proof of Proposition 3

*Proof.* By definition of subjective expectations:

$$ERR_{t+1} = \frac{1-K}{K}REV_t + \underbrace{(g_{t+1} - \rho g_t)}_{\sigma_\epsilon \epsilon_{t+1} + \sigma_u u_{t+1} - \rho \sigma_\epsilon \epsilon_t} \quad (\text{IA.4})$$

This implies:

$$E(ERR_{t+1}|REV_t) = \frac{1-K}{K}REV_t - \rho \sigma_\epsilon E(\epsilon_t|REV_t) \quad (\text{IA.5})$$

Now, we focus on the second term on the right-hand side. Given the definition of forecasts in equation (9), we can express forecast revisions using the following orthogonal decomposition:

$$REV_t = K\rho\sigma_\epsilon \left[ \underbrace{\epsilon_t - K\rho \sum_{s \geq 0} ((1-K)\rho)^s \epsilon_{t-s-1}}_{\equiv E_t} + \underbrace{\frac{\sigma_u}{\sigma_\epsilon} \sum_{s \geq 0} ((1-K)\rho)^s u_{t-s}}_{\equiv U_t} \right]$$

Under Assumption IA.1, we can make use of the following result for the asymptotic behavior of regularly varying random variables to characterize the tail behavior of  $E_t$ :

**Lemma IA.1** (Geluk and de Vries 2006, Theorem 1). *Suppose  $X_0, X_1, \dots$  are IID regularly varying random variables and  $c_0, c_1, \dots$  are weights that represent a convergent geometric series. Then*

$$P\left(\sum_{s \geq 0} c_s X_s > x\right) \sim \sum_{s \geq 0} P(c_s X_s > x) \text{ as } x \rightarrow \infty$$

Applying Lemma IA.1 to the definition of  $E_t$ , we have, as  $E \rightarrow +\infty$ ,

$$\begin{aligned} P(E_t > E) &= P\left(K\rho \sum_{s \geq 0} ((1-K)\rho)^s \epsilon_{t-s-1} > E\right) \\ &\sim \sum_{s \geq 0} P(K\rho ((1-K)\rho)^s \epsilon_{t-s-1} > E) \\ &\sim \sum_{s \geq 0} L(E) \left(\frac{K\rho ((1-K)\rho)^s}{E}\right)^\nu \\ &= L(E) \frac{1}{E^\nu} \times \frac{K^\nu \rho^\nu}{1 - \rho^\nu (1-K)^\nu} \\ &\sim P(\epsilon_t > E) \times \frac{K^\nu \rho^\nu}{1 - \rho^\nu (1-K)^\nu} \end{aligned}$$

The final expression shows that  $E_t$  is also regularly varying with index  $\nu$  but a different scale given by the

formula. Given this result, we can characterize the  $E(\epsilon_t | REV_t)$  for large revisions as follows:

$$\begin{aligned} E(\epsilon_t | REV_t = r) &= E(\epsilon_t | K\rho\sigma_\epsilon(\epsilon_t - E_t + U_t) = r) \\ &= E\left(\epsilon_t \mid \epsilon_t - E_t + U_t = \frac{r}{K\rho\sigma_\epsilon}\right) \\ &\sim E\left(\epsilon_t \mid \epsilon_t - E_t = \frac{r}{K\rho\sigma_\epsilon}\right) \text{ as } r \rightarrow +\infty \end{aligned}$$

where the last line follows from the fact that  $E_t$  has the same tail parameter as  $\epsilon_t$ , which is thicker than that of  $U_t$  by Assumption 2. To compute the expectations of one regularly varying variable conditional on the sum of two, one needs to know the relative scale of these two variables. As shown above,  $E_t$  has the same index as  $\epsilon_t$ , but its counter-CDF differs by a scale of  $\frac{K^\nu\rho^\nu}{1-\rho^\nu(1-K)^\nu}$  in the tails. Applying Proposition 4.9 from [Denuit et al. \(2025\)](#) along with the fact that  $\epsilon_t$  (and hence  $E_t$ ) is symmetric, we have, as  $r \rightarrow +\infty$ :

$$\begin{aligned} E\left(\epsilon_t \mid \epsilon_t - E_t = \frac{r}{K\rho\sigma_\epsilon}\right) &\sim \frac{L(\frac{r}{K\rho\sigma_\epsilon})}{L(\frac{r}{K\rho\sigma_\epsilon}) + L(\frac{r}{K\rho\sigma_\epsilon})\frac{K^\nu\rho^\nu}{1-\rho^\nu(1-K)^\nu}} \frac{r}{K\rho\sigma_\epsilon} \\ &= \frac{1}{1 + \frac{K^\nu\rho^\nu}{1-\rho^\nu(1-K)^\nu}} \frac{r}{K\rho\sigma_\epsilon} \end{aligned}$$

Now, we plug this result into equation (IA.5) and obtain that:

$$E(ERR_{t+1} | REV_t) \sim \underbrace{\left[ \frac{1-K}{K} - \frac{1}{K} \frac{1}{1 + \frac{K^\nu\rho^\nu}{1-\rho^\nu(1-K)^\nu}} \right]}_{\equiv C} REV_t.$$

Repeating the analogous argument for the case in which  $REV_t \rightarrow -\infty$  and using the fact that  $\epsilon_t$  and  $u_t$  are symmetric delivers the first desired result.

To show that  $C < 0$ , we use the definition of the steady state Kalman gain,  $K$ , and prior variance,  $P$ :

$$\begin{aligned} K &= \frac{P}{P + \sigma_\epsilon^2} \\ P(1 - \rho^2(1 - K)) &= \sigma_u^2 \end{aligned}$$

whose combination shows that

$$1 - K = \frac{K}{\frac{\sigma_u^2}{\sigma_\epsilon^2} + K\rho^2}.$$

For  $C < 0$ , we need

$$\frac{1-K}{K} < \frac{1}{K} \cdot \frac{1}{1 + \frac{K^\nu\rho^\nu}{1-\rho^\nu(1-K)^\nu}}.$$

Multiplying both sides by  $K$  and cross-multiplying yields

$$(1 - K)K^\nu \rho^\nu + K(1 - K)^\nu \rho^\nu < K.$$

Dividing by  $K$  and factoring

$$(1 - K)\rho^\nu [K^{\nu-1} + (1 - K)^{\nu-1}] < 1.$$

Since  $\nu > 2$ , we have  $\nu - 1 > 1$ . For any  $x \in (0, 1)$ , this implies  $x^{\nu-1} < x$ . Applying this to both  $K$  and  $1 - K$ :

$$K^{\nu-1} + (1 - K)^{\nu-1} < K + (1 - K) = 1.$$

Therefore:

$$(1 - K)\rho^\nu [K^{\nu-1} + (1 - K)^{\nu-1}] < (1 - K)\rho^\nu < 1$$

where the final inequality comes from the fact that  $\rho, K \in (0, 1)^2$ . This is the desired inequality two lines above, which concludes the proof.  $\square$

## B.4 Proof of Corollary 1

*Proof.* The definition of the Kalman filter implies that it is the minimum linear mean squared error predictor. Combined with the law of iterated expectations, this implies that:

$$\begin{aligned} E(ERR_{t+1}REV_t) &= E\left[E(ERR_{t+1}REV_t || REV_t \geq r) P(|REV_t| \geq r) \right. \\ &\quad \left. + E(ERR_{t+1}REV_t || REV_t < r) P(|REV_t| < r)\right] = 0. \end{aligned}$$

By Proposition 3, there exists an  $\bar{R}$  such that  $E(ERR_{t+1}REV_t || REV_t \geq \bar{R}) < 0$ . Therefore, the fact that  $E(ERR_{t+1}REV_t) = 0$  implies that  $E(ERR_{t+1}REV_t || REV_t < \bar{R}) > 0$ , which is the desired result.  $\square$

## B.5 Proof of Proposition 4

*Proof.* By the definition of the process, we have:

$$\text{var}(g_{t+1}|g_t) = \rho^2 \text{var}(g_t^*|g_t) + \sigma_u^2 + \sigma_\epsilon^2$$

Differentiating the log density twice gives:

$$\frac{d^2}{dg^2} \log h(g) = \frac{h''(g)}{h(g)} - \left( \frac{h'(g)}{h(g)} \right)^2.$$

Proposition 1 implies that  $\frac{h'(g)}{h(g)} = -\frac{E(g_t^*|g)}{\sigma_{g^*}^2}$ , which in turn gives:

$$\frac{h''(g)}{h(g)} = -\frac{1}{\sigma_{g^*}^2} + \frac{E((g_t^*)^2|g)}{\sigma_{g^*}^4}.$$

Combining the previous two expressions gives

$$\begin{aligned} \frac{d^2}{dg^2} \log h(g) &= -\frac{1}{\sigma_{g^*}^2} + \frac{E((g_t^*)^2|g)}{\sigma_{g^*}^4} - \left( \frac{E(g_t^*|g)}{\sigma_{g^*}^2} \right)^2 \\ &= \frac{1}{\sigma_{g^*}^4} (-\sigma_{g^*}^2 + \text{var}(g_t^*|g_t)). \end{aligned}$$

Using the first formula above for  $\text{var}(g_t^*|g_t)$  gives the desired result.

□

## B.6 Additional Details on Asset Pricing Model in Section 4.3

We first derive (16). Define the return on a stock as  $R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t} - 1$ , which has sales  $S_t$ , earnings  $\mathcal{E}_t$ , and a payout ratio  $DE_t = \frac{D_t}{\mathcal{E}_t}$ . Consistent with our variable definitions and transformations in Section 1, denote the (log) growth rate of sales as  $g_t = \log S_t - \log S_{t-1}$ . Assume that earnings growth is a constant fraction of sales growth:  $\log\left(\frac{\mathcal{E}_t}{\mathcal{E}_{t-1}}\right) = \gamma g_t$ . Denoting lower-case letters as logs and following Campbell and Shiller (1988), we can approximate the price-earnings ratio to first order around the mean price-dividend ratio as:

$$p_t - e_t = \kappa + \gamma g_{t+1} - r_{t+1} + (1 - c)(d_{t+1} - e_{t+1}) + c(p_{t+1} - e_{t+1}),$$

where  $c = \frac{e^{\overline{pd}}}{1 + e^{\overline{pd}}}$ ,  $\overline{pd}$  is the mean of the log price-dividend ratio, and  $\kappa$  is a constant. Assuming a constant log payout ratio of  $\overline{de}$ , we can iterate the first equation forward to obtain:

$$p_t - e_t = \frac{\kappa}{1 - c} + \overline{de} + \sum_{k=1}^{\infty} c^{k-1} (\gamma g_{it+k} - r_{it+k}) + \lim_{k \rightarrow \infty} c^k (p_{it+k} - e_{it+k}).$$

Imposing the usual transversality condition, the previous equation becomes:

$$p_t - e_t = \frac{\kappa}{1 - c} + \overline{de} + \sum_{k=1}^{\infty} c^{k-1} (\gamma g_{it+k} - r_{it+k}).$$

Now, rearranging the first equation above, we have

$$r_{t+1} = \kappa + \gamma g_{t+1} + (1 - c)\overline{de} + c(p_{t+1} - e_{t+1}) - (p_t - e_t).$$

Letting  $F_t$  denote investors' subjective beliefs, the previous two equations imply:

$$F_{t+1}r_{t+1} - F_t r_{t+1} = \gamma F_{t+1}g_{t+1} - \gamma F_t g_{t+1} + c[(p_{t+1} - e_{t+1}) - F_t(p_{t+1} - e_{t+1})].$$

Rewriting, we obtain:

$$r_{t+1} - F_t r_{t+1} = \gamma (F_{t+1} - F_t) \sum_{k=0}^{\infty} c^k g_{t+1+k} - (F_{t+1} - F_t) \sum_{k=1}^{\infty} c^k r_{t+1+k}.$$

To derive predictions based on our model of subjective expectations, we follow [Bouchaud et al. \(2019\)](#) and [Nagel and Xu \(2022\)](#) and assume that subjective risk premia are constant and equal to the risk-free rate,  $r_f$ , plus a constant risk premium,  $\pi$ . Under this assumption, the final term in the previous equation is zero and we obtain (16):

$$r_{t+1} = \log(r_f + \pi) + \underbrace{\gamma (F_{t+1} - F_t) \sum_{k=0}^{\infty} c^k g_{t+1+k}}_{=\sum_{k=1}^{\infty} c^k REV_t g_{t+k}}.$$

Given a panel of simulated earnings growth expectations, we can compute the final two terms in the previous equation by recognizing that our model of belief formation implies:

$$F_t \sum_{k=0}^{\infty} c^k g_{t+1+k} = F_t g_{t+1} \sum_{k=0}^{\infty} c^k \rho^k = \frac{F_t g_{t+1}}{1 - c\rho}.$$

Using this relationship, we can now simulate a path of return realizations using the following relationship:

$$r_{t+1} = \log(1 + R_f + \pi) + \frac{\gamma}{1 - c\rho} (g_{t+1} - F_t g_{t+1}).$$

We then set  $\gamma$  to generate a volatility of  $R_t$  equal to 15%, after winsorizing all returns that are above 100%.

## C Computing Rational Expectations Using the Particle Filter

Constructing the forecasts in (12) requires computing the rational expectation,  $E(g_{t+1} \mid g_0, \dots, g_t)$ . In the case where  $\epsilon_t$  is normally distributed, this corresponds to the Kalman filter and takes the simple closed form in (11). However, outside of this special case, this expectation cannot be computed in closed form and instead must be computed using sequential Monte Carlo methods. We choose to compute this expectation using the particle filtering algorithm from [Fernandez-Villaverde and Rubio-Ramirez \(2007\)](#) (also known as sequential importance sampling), which is described in Algorithm 1.

The goal of this algorithm is to estimate the posterior distribution of a latent state process,  $\{g_t^*\}$ , given a sequence of observed data,  $\{g_t\}_{t=1}^T$ . The algorithm proceeds by approximating the filtering distribution  $p(g_t^* \mid g_1, \dots, g_t)$  using a collection of  $P$  particles, each representing a possible realization of the latent state. At each time step  $t$ , these particles are propagated forward via the state transition equation for  $g_t^*$ , incorporating stochastic innovations sampled from the distribution of  $u_t$ . The new particles are then evaluated against the observed data using a likelihood function  $f_\epsilon(\cdot)$ , the density function of the distribution of  $\epsilon_t$ . Particles are assigned weights according to this likelihood, and a resampling step is used to adjust the distribution of the particles based on their posterior probabilities. Under mild conditions, [Fernandez-Villaverde and Rubio-Ramirez \(2007\)](#) show that the expectations computed in Step 5, which depend on  $P$ , converge to their population counterparts as  $P \rightarrow \infty$ . This is the sense in which the algorithm “works.”

In the case where  $f_\epsilon(\cdot)$  is the density function of a normal distribution, the expectations computed in Step 5 converge to those of the Kalman filter as  $P \rightarrow \infty$ . However, in the presence of non-Gaussian shocks, as in our case, these two solutions will differ. The key step in this algorithm is Step 6, where each particle’s weight is computed using  $f_\epsilon(\cdot)$ . Intuitively, this weighting step incorporates the non-normality of  $f_\epsilon(\cdot)$  by “tilting” the contribution of each particle to the posterior distribution of the latent state according to how well it explains the data under the true distribution. As a result, particles that better align with the observed data under the correct distribution are favored during the resampling step (Step 7), allowing the particle approximation to capture features—like skewness or fat tails—that the Kalman filter necessarily misses.

We run the particle filter on each of our 100 simulations of length 100,000 with  $P$  set to 10,000.<sup>18</sup> The most computationally intensive part of this Algorithm 1 is Step 7, which requires resampling a large number of particles from a non-uniform distribution.

---

<sup>18</sup>We choose this particular value of  $P$  because it is the largest value of  $P$  such that all the particles fit in the memory of our GPU. We have found very similar results with  $P$  set to 5,000, which makes us confident that 10,000 is sufficiently large that the particle filter recovers the true conditional expectations.

---

**Algorithm 1** Particle Filter

---

- 1: **Fix** a number of particles,  $P$ , and a time series of length  $T$ ,  $\{g_t\}_{t=1}^T$ .
- 2: **Initialize** particles,  $\{\xi_0^p\}_{p=1}^P$ , at  $\xi_0^p = g_0^*$  for all  $p$ , and set  $t = 1$ .
- 3: **Sample**  $\{u_t^p\}_{p=1}^P$  from  $N(0, \sigma_u^2)$ .
- 4: **Update** particles according to:

$$\{\xi_t^p\}_{p=1}^P = \{\rho \xi_{t-1}^p + u_t^p\}_{p=1}^P.$$

- 5: **Compute** and store:

$$E(g_t \mid g_1, \dots, g_{t-1}) = E(g_t^* \mid g_1, \dots, g_{t-1}) = \frac{1}{P} \sum_{p=1}^P \xi_t^p$$

- 6: **Define**  $f_\epsilon(\cdot)$  as the PDF of  $\epsilon_t$  and compute:

$$q_t^p = \frac{f_\epsilon\left(\frac{g_t - \xi_t^p}{\sigma_\epsilon}\right)}{\sum_{p=1}^P f_\epsilon\left(\frac{g_t - \xi_t^p}{\sigma_\epsilon}\right)},$$

- 7: **Resample**  $\{\xi_t^p\}_{p=1}^P$  from  $\{\xi_{t-1}^p\}_{p=1}^P$  with replacement and sampling weights  $\{q_t^p\}$ .
  - 8: **if**  $t = T$  **then**
  - 9:     **Stop**.
  - 10: **else**
  - 11:     **Set**  $t \leftarrow t + 1$  and go to Step 3.
  - 12: **end if**
-