

# 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>

April 14, 2025

[Latest Version ↗](#)

## Abstract

This paper studies expectations formation when the underlying process has fat tails. Using a large sample of firm sales growth expectations, we document three facts: (i) the relationship between forecast revisions and future forecast errors is strongly non-linear, (ii) the distribution of sales growth has fat tails, and (iii) extreme values of sales growth tend to mean-revert. 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 fully. We estimate this model and show it quantitatively explains our three facts. Finally, we show the model is consistent with evidence from an online forecasting experiment where the underlying process is non-Gaussian and the non-linearity in the momentum of stock returns.

---

\*We are grateful to Xavier Gabaix for very helpful feedback, and thank Mauro Cazzaniga for his outstanding research assistance. We also thank Nick Barberis, Kelly Shue, and seminar participants at CFM, Columbia, MIT, Yale, and NYU for their thoughtful comments.

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

<sup>‡</sup>Capital Fund Management

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

Expectations formation is a core question in economics. In recent years, a strand of literature in macroeconomics and finance has been collecting empirical regularities using survey data on subjective forecasts. A common finding in this literature is that forecast errors are predictable using information within forecasters' information sets, which is inconsistent with rational expectations. Most theories proposed to explain this predictability assume that the data-generating process (DGP) being forecasted is simple, such as an AR1 process with normal shocks, and forecasters know this DGP. However, despite knowing the DGP, forecasters make predictable errors because of behavioral biases or cognitive limitations, such as sluggish updating (Bouchaud et al., 2019), representativeness (Bordalo et al., 2019, 2020), and imperfect memory (Afrouzi et al., 2023).

In this paper, we take a different approach and argue that recognizing the complexity of the underlying DGP is crucial for understanding the properties of subjective forecasts. We begin by documenting three facts using data on sales growth forecasts 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 distribution of the underlying process has fat tails; and (iii) the conditional expectation of future sales growth is non-linear in current growth, with 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. After showing formally that our model can explain the three facts we documented in the data, we estimate it and show that it does so quantitatively. Finally, we show that our framework is consistent with evidence from an online forecasting experiment where the underlying process is non-Gaussian and that it provides an explanation for non-linearity in the momentum of stock returns.

Our empirical analysis uses data on analyst forecasts of sales growth and their realizations from IBES. The advantage of focusing on sales growth rather than earnings-per-share (as is typically done) is that sales growth is stationary and provides a larger sample. Using these data, the first and most important fact that we document is that the relationship between 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, 2020). In our panel of forecasts and realizations of sales growth, we find evidence that both co-exist. For intermediate values of revisions, forecasters underreact to news (a positive relationship between revisions and errors). For large values of revisions, forecasters overreact (a negative relationship between revisions and errors). Given the size of our sample, we can establish that this non-linearity is a robust feature of the data and is present in a variety of robustness checks.

Our second empirical fact is that the distribution of sales growth has fat tails, as in [Stanley et al. \(1996\)](#). To detect the presence of fat tails, we examine the log-density plots, similar to the literature on income dynamics ([Guvenen et al., 2021](#)). In the bulk of the sales growth distribution, we find that a Gaussian density, which is quadratic in logs, provides a close approximation. However, in the top and bottom 10% of the distribution, sales growth has much thicker and longer tails than those of a normal distribution. Instead, we find that these tails are well-approximated by a power law with a tail parameter between 2 and 3, thinner than Zipf’s law ([Gabaix, 2009](#)). We show that these fat tails do not arise from heterogeneity in volatility across firms ([Wyart and Bouchaud, 2003](#)), consistent with [Moran et al. \(2024\)](#), and do not arise from time-varying aggregate volatility.

Our third and final motivating fact is that the conditional expectation of future sales growth on current growth is non-linear. In particular, we find that the relationship between current and one-year-ahead sales growth is increasing and linear in the bulk of the distribution. However, in the tails of the distribution, we find that this relationship flips sign, indicating that extreme values of sales growth 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 forecasting model that is designed to connect the above three facts. In the model, the DGP for the forecasting variable contains a persistent and transitory component, which is common in models of dividend growth ([Lettau and Wachter, 2007](#)) and income dynamics ([Guvenen et al., 2014](#)). Importantly, we assume the transitory component is non-Gaussian and follows a power law distribution, a tractable way of characterizing processes with fat tails ([Gabaix, 2009](#)). In contrast, we assume that the persistent component follows an AR1 process with normal innovations. This assumption is less crucial, but it is important that the transitory component has thicker tails than the persistent one. Having fat tails only in the persistent component would not match our third fact: the strong mean reversion in sales growth.

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 growth conditional on current growth—our third fact—we apply a result from empirical Bayes theory known as Tweedie’s formula ([Efron, 2012](#)). Although this expectation can not be characterized in closed-form due to the non-normality, this result characterizes it as a function of the (observable) density function of growth. We show that this result implies that the expectation of future growth conditional on current growth is locally linear when the density is locally Gaussian, as in the bulk of the distribution. However, when the density is locally non-Gaussian, as in the tails, the expectation of future growth conditional on current growth is no longer linear and is asymptotically decreasing in current growth. Intuitively, very large

realizations of growth are likely due to the transitory component and, hence, will not persist.

Given this model of the DGP that is consistent with our second and third facts, we show it 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 component followed a normal rather than power law distribution. We view the assumption that agents use a simple mis-specified model as a natural form of bounded rationality (as in [Fuster et al. 2010](#) and [Gabaix 2019](#)), but do not microfound it. We show 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 shocks to current growth, which 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 that 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.

Next, we assess our model’s ability to replicate our three facts quantitatively. We estimate the parameters governing the DGP using simulated minimum distance and show that they provide a close fit to our second and third facts that are specific to the DGP. However, when we turn to beliefs, we find that the estimated model generates too much underreaction in the bulk and overreaction in the tails relative to our first fact. This is not surprising, given our model has no additional free parameters that govern belief formation. Therefore, we enrich our model by assuming that expectations are a weighted average of the Kalman gain and the rational expectation (as in [Fuster et al. 2010](#) and [Gabaix 2019](#)), which we numerically compute using a particle filter ([Fernandez-Villaverde and Rubio-Ramirez, 2007](#)). We estimate the degree of shrinkage towards the rational expectation and find that the model can match our first fact with forecasters placing a 71% weight on the rational expectation and a 29% weight on the Kalman filter.

We conclude by providing two additional tests of our model. 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 expectation 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, like in our first fact. In contrast, when forecasters forecast a similar process with no fat tails, 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 our model makes predictions for return momentum (Jegadeesh and Titman, 2011) that are supported by the data. To translate sales growth expectations into returns, we apply the Campbell (1991) return decomposition with a constant subjective discount rate (as in Bouchaud et al. 2019 and Nagel and Xu 2019). We find that 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 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.** This paper contributes to the recent and growing literature on expectations formation. Most of this literature works with models in which data-generating processes are Gaussian, implying conditional expectations are linear. In these models, forecasters may or may not be rational, but they, in general, are linear, implying a linear relationship between errors and revisions. Our empirical analyses highlight how this relationship can be quite non-linear, and we provide a theory that links this non-linearity to features of the DGP. Our focus on a non-Gaussian DGP is similar to Kozlowski et al. (2020) and Farmer et al. (2021), but we focus on whether forecasters ignore these non-Gaussian dynamics rather than learning. 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 (Fuster et al., 2010; Gabaix, 2019).

Three closely related papers on belief formation are Kwon and Tang (2025), Augenblick et al. (2024), and Graeber et al. (2024). 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 to 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, our model could be interpreted in this way. Given our model of the DGP, past growth is a much stronger predictor of future growth in the bulk of the distribution rather than in the tails. However, our forecasters that ignore non-Gaussian dynamics do not realize this, causing them to underreact to strong signals and overreact to weak signals. Finally, Graeber et al. (2024) analyze the S-shaped relationship between returns and earnings surprises, which map

into revisions and forecast errors in our framework. 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 zeros comes from dampening within a category due to noisy perceptions. Our first empirical 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 the omnipresence of power laws (Gabaix, 2009). We show that sales *growths* rather than levels have fat tails, a fact first documented by Stanley et al. (1996). Our finding that the heavy tails of the growth distribution cannot be explained by a mixture of Gaussian distribution with heterogeneous variances is consistent with Moran et al. (2024). Also related is the literature on income dynamics, which emphasizes deviations from the canonical income processes with Gaussian shocks (Guvenen et al., 2014, 2021). Unlike these models, which emphasize the importance of non-normal persistent shocks, the key ingredient in our model is non-normal *transitory* shocks.<sup>1</sup>

**Outline.** Section 1 describes the publicly available data we use, on analyst forecasts (IBES), and the confidential data on international stock returns that we use at the end of the paper. Section 2 documents our three main facts. Section 3 lays out the simple framework we build to connect and explain these facts. Section 4 estimates our model and shows that it quantitatively explains these facts. Section 5 provides additional tests of our model from an online return experiment and data on stock returns.

# 1 Data and Variables

## 1.1 Data Source

Our analysis primarily relies on a large annual panel of analyst forecasts for yearly revenues at one and two-year horizons. We obtain these forecasts from IBES Adjusted Summary Statistics files, which provide data for both U.S. and international firms. The summary statistics files contain “current” estimates as of the third Wednesday of each month. We extract mean forecasts reported in the third month of each fiscal year  $t + 1$ , after fiscal year  $t$  earnings have been announced. The forecasts we use correspond to two horizons: fiscal years  $t + 1$  and  $t + 2$ . The resulting panel covers the period 2000-2023, and includes both U.S. and foreign firms.

---

<sup>1</sup>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.

We focus on revenue forecasts rather than EPS (earnings per share) for several methodological advantages. First, revenues are consistently positive, making past realized revenues a natural normalization base (some normalization is needed to ensure stationarity). Revenue growth forecasts exhibit a well-behaved distribution with minimal outliers—a crucial attribute given our emphasis on distribution tails. Second, revenues are not reported on a per-share basis, eliminating the confounding effects of unexpected stock splits that can generate substantial jumps in EPS or forecast errors unrelated to the focus of this paper. Finally, the distribution of EPS-to-price ratios—a common normalization for most studies—is notably non-normal with a characteristic bulge above zero, whereas the log sales growth distribution demonstrates smooth, well-behaved properties.

In our robustness checks, we also examine individual analyst forecasts and EPS forecasts (normalized by price). Similar to our revenue data, we extract these at annual horizons in the third month of the fiscal year.

## 1.2 Definition of Forecasting Variable: Sales Growth

We start with the definition of sales growth, which is our forecasting variable of interest. We denote sales of firm  $i$  at date  $t$  as  $R_{it}$ . These data are from the IBES actual files, which ensures comparability with analyst forecasts (described below). (Raw) log sales growth is given by  $G_{it} = \log R_{it} - \log R_{it-1}$ .

Most of our analysis actually focuses on *adjusted* sales growth. We do this for two reasons. First, this adjustment makes different firms comparable with one another, which makes it easier to fit a single data-generating process on the whole cross-section of firms (we do this in Section 4). Second, as discussed in [Wyart and Bouchaud \(2003\)](#), the thick tail of the growth distribution may mechanically emerge from the combination of normal processes interacted with heterogeneous variances. Our adjustment takes care of these two issues, but we will later explore robustness.

For each firm  $i$ , denote  $T_i$  as the number of years for which we have a growth observation and  $\mu_i = \frac{1}{T_i} \sum_t G_{it}$  as the empirical average of growth observations for firm  $i$ . Additionally, denote  $\sigma_i = \frac{1}{T_i} \sum_t |G_{it} - \mu_i|$  as the mean absolute deviation, an estimate of the standard deviation of growth at the firm level. The advantage of this measure is that it is less sensitive to outliers than variance, because it has no squared term. We then define *adjusted* growth as:

$$g_{it} = \frac{G_{it} - \mu_i}{\sigma_i}.$$

### 1.3 Definitions of Forecast Errors and Revisions

We now turn to the definition of forecast errors and revisions, the key objects of our analysis. For each firm  $i$  and each year  $t$ , we denote  $F_t R_{it+1}$  the forecast made in year  $t$  for the future realization of sales  $R_{it+1}$ . As previously argued,  $F_t R_{it+1}$  is obtained from IBES summary files as the mean consensus forecast extracted in the third month after the end of fiscal year  $t$ . Similarly, the two-year ahead forecast  $F_{t-1} R_{it+1}$  is measured three months after the end of fiscal year  $t - 1$ .

Our key variable of interest in this paper is the forecast of log sales growth. We construct the  $h$ -year ahead forecast of *raw* sales growth as

$$F_t G_{it+h} = \log F_t R_{it+h} - \log R_{it},$$

and the  $h$ -year forecast of *adjusted* sales growth as

$$F_t g_{it+h} = \frac{1}{\sigma_i} (\log F_t G_{it+h} - \mu_i).$$

We focus on  $h = 1$  and  $h = 2$  for one- and two-year ahead forecasts. Note that the way we translate from forecasts of raw to adjusted sales growth implicitly ignores a Jensen's inequality term because  $\log F_t G_{it+h} \neq F_t \log G_{it+h}$ . We ignore this adjustment for simplicity, but it does not materially affect our analysis. First, on a theoretical level, a constant Jensen's term would simply shift the unconditional level of forecasts, while our analysis focuses on the conditional properties of forecasts errors. Second, we show that our empirical results are robust to using percent (instead of log) growth.

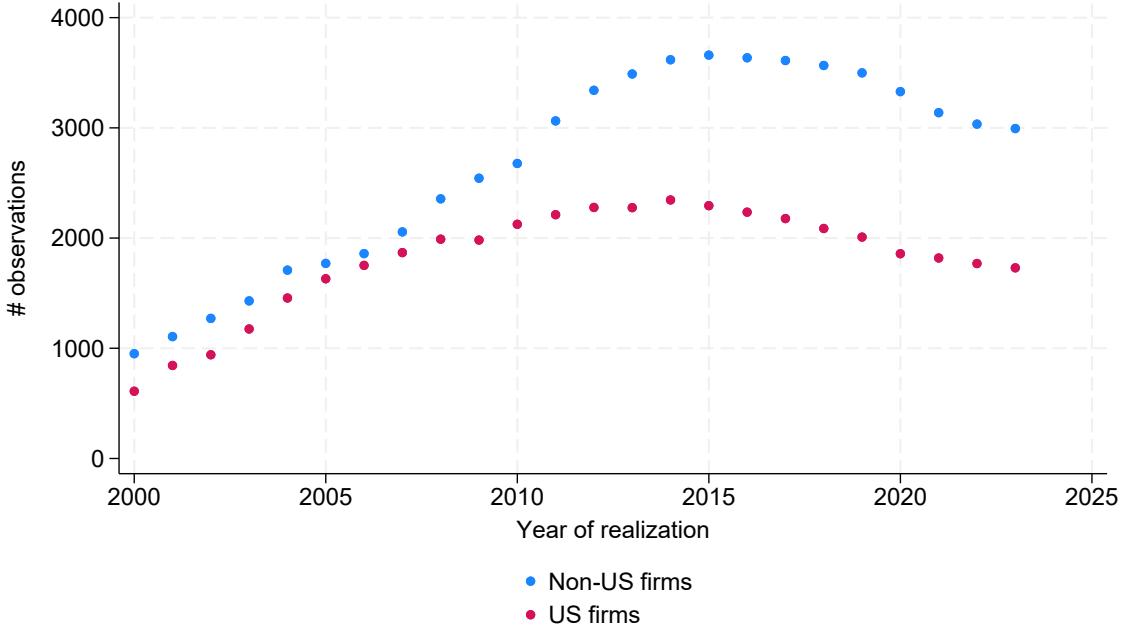
Using these definitions, we then construct forecasts errors and revisions in line with the literature on expectations formation ([Coibion and Gorodnichenko, 2015](#); [Bouchaud et al., 2019](#)). In particular, raw and adjusted forecast errors are defined as

$$ERR_t G_{it+1} = G_{it+1} - F_t G_{it+1}, \quad ERR_t g_{it+1} = g_{it+1} - F_t g_{it+1},$$

while raw and normalized forecast revisions are defined as

$$R_t G_{it+1} = F_t G_{it+1} - F_{t-1} G_{it+1}, \quad R_t g_{it+1} = F_t g_{it+1} - F_{t-1} g_{it+1}.$$

Figure 1: Number of observations in our sample



*Notes:* This figure shows the total number of observations in our data for which we have a non-missing (adjusted) growth and one-year forecast.

## 1.4 Summary Statistics

The sample in our analysis consists of the entire international IBES summary file, subject to several sample restrictions. First, we focus on firms for which at least 10 realizations of sales growth are observed (i.e.,  $T_i \geq 10$ ). Second, we focus on years for which sales forecasts are sufficiently well-populated, which is from 2000 to 2023. We end up with 122,395 observations. We then trim the top and bottom 1% of all of our variables. This trimming procedure does not materially affect our results, but discards extreme outliers which would compress our plot axes.

Figure 1 shows the total number of observations in our data for which we have a non-missing growth and one-year forecast. As early as 2000, there are already 1,700 such observations (1,000 international and 700 in the U.S.). Then, the number of firms grows to about 3,700 internationally and 2,200 in the U.S. Overall, sales growth is well covered by IBES after 2000.

Table 1 shows summary statistics for our main variables of interest. The standard deviation of raw log sales growth is 17 ppt, and the mean is 6%. Normalized growth has a mean and median closer to zero, and a standard deviation of 1.20 – it is not exactly equal to 1 because we normalize by the mean absolute deviation, not the firm-level standard deviation. Forecast errors have less

Table 1: Summary Statistics

Variable	Mean	P25	Median	P75	SD	# of Obs.
<i>Panel A: Full Sample</i>						
Log sales growth (raw)	0.06	-0.02	0.06	0.14	0.17	108,706
Growth forecast error (raw)	-0.01	-0.05	-0.00	0.05	0.13	108,694
Growth forecast revision (raw)	-0.01	-0.04	-0.00	0.03	0.11	105,609
Log sales growth (normalized)	0.00	-0.70	0.01	0.72	1.20	108,706
Growth forecast error (normalized)	-0.06	-0.52	-0.02	0.45	0.96	108,694
Growth forecast revision (normalized)	-0.04	-0.39	-0.03	0.32	0.80	105,609
<i>Panel B: US Firms</i>						
Log sales growth (raw)	0.06	-0.02	0.05	0.13	0.17	64,881
Growth forecast error (raw)	-0.01	-0.06	-0.00	0.05	0.13	64,686
Growth forecast revision (raw)	-0.00	-0.04	-0.00	0.03	0.10	63,892
Log sales growth (normalized)	0.01	-0.70	0.02	0.73	1.20	64,727
Growth forecast error (normalized)	-0.07	-0.59	-0.04	0.47	1.01	64,569
Growth forecast revision (normalized)	-0.02	-0.36	-0.01	0.34	0.79	63,716
<i>Panel C: Non-US Firms</i>						
Log sales growth (raw)	0.07	-0.01	0.06	0.15	0.18	43,825
Growth forecast error (raw)	-0.01	-0.05	-0.00	0.05	0.12	44,008
Growth forecast revision (raw)	-0.01	-0.05	-0.01	0.03	0.12	41,717
Log sales growth (normalized)	-0.00	-0.70	-0.01	0.72	1.20	43,979
Growth forecast error (normalized)	-0.03	-0.43	-0.01	0.41	0.88	44,125
Growth forecast revision (normalized)	-0.07	-0.43	-0.05	0.30	0.81	41,893

*Notes* Source: IBES summary files. Raw growth corresponds to unadjusted log growth ( $G_{it}$  in the main text). Adjusted growth subtracts the firm-level mean and divides by the mean absolute distance ( $g_{it}$  in the main text). All variables are trimmed at the bottom and top 1%.

variance than growth itself, consistent with the idea that forecasters can somewhat predict future growth. [Figure A.1](#) provides suggestive evidence that the distribution of forecast errors has thicker tails than a Gaussian fit.

## 2 Motivating Facts

In this section, we document the three key facts that motivate the model of expectations formation that we develop in Section 3. In the main body of the paper, we present these facts using normalized log sales growth,  $g_{it}$ , but discuss a series of robustness checks and report their results in Appendix.

## 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:

$$ERR_t Y_{it+1} = \alpha + \beta R_t Y_{it+1} + e_{it+1} \quad (1)$$

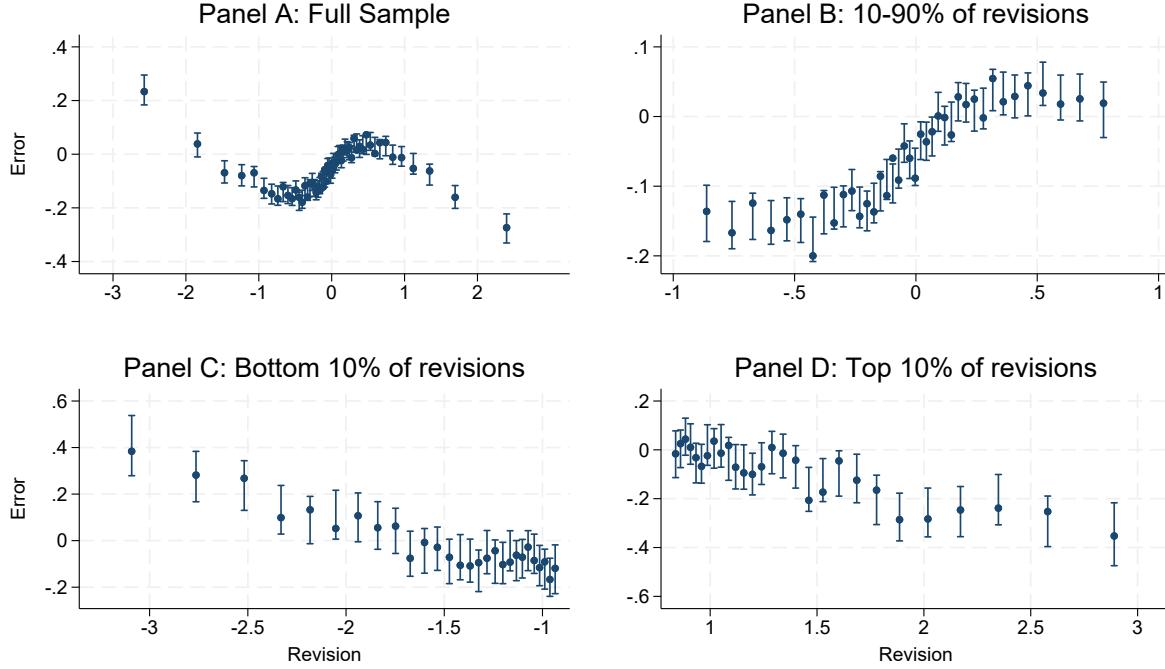
for any forecasting variable  $Y_{it}$ . This regression is useful because its slope coefficient can be used to distinguish between different models of expectations formation, requiring only panel data on expectations. Full-information rational expectations predicts  $\beta = 0$  for consensus forecasts, while limited-information rational expectations predicts  $\beta = 0$  for individual forecasts. 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](#)). However, 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., 2020](#)). 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, in this section we use our large sample of sales expectations to provide evidence that this relationship is non-linear.

[Figure 2](#) provides preliminary evidence that, for  $Y = g_{it}$ , the relationship between forecast errors and revisions is non-linear. Panel A shows a binned scatterplot of forecast errors,  $ERR_t g_{it+1}$ , as a function of revisions  $R_t g_{it+1}$ , using 100 bins. This plot provides evidence of significant non-linearity. Panels B, C, and D show the same plot separately for observations in the 10-90th, 0-10th, and 90-100th 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., 2020](#)).

Panels C and D of [Figure 2](#) shows 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 *over-reaction*. In other words, forecasters appear to overreact in response to news that generate large revisions, while underreacting to more moderate news.

[Figure 3](#) provides sharper statistical evidence of the non-linear relationship between forecast

Figure 2: Non-Linear Relationship Between Forecast Errors and Revisions

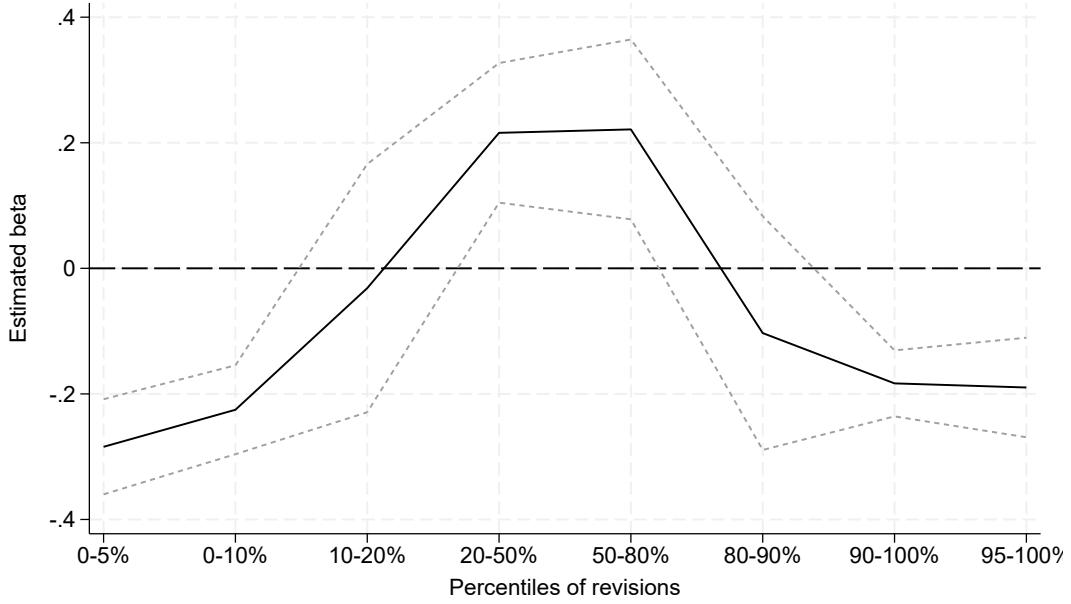


*Notes:* This figure shows binned scatterplots of forecast errors on forecast revisions for normalized log sales growth. See definitions of raw and normalized growth in Section 1. Panel A shows the results for the entire sample; Panel B restricts the sample to the 10-90th percentiles of revisions; Panel C restricts the sample to below the 10th percentile of revisions; Panel D restricts the sample to above the 90th 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”).

errors and revisions shown in Figure 2. In particular, we report the slope coefficient from estimating the CG error-revision regression on eight different subsamples based on different percentiles of revisions shown in the horizontal axis. In each one of these regressions, we double cluster error terms at the year and firm levels. The results in Figure 3 confirm the findings in Figure 2: between the 20th and 80th percentiles of revisions, errors are a significantly *increasing* function of revisions. In contrast, outside of these bounds in the tails of the distribution of revisions, errors are significantly *negatively* correlated with revisions.

**Robustness.** This non-linear relation is robust to a battery of robustness checks. First, while our adjustments from raw to normalized (log) growth are done to adjust for heterogeneous variances, it could be that this adjustment mechanically generates mean-reversion. We show in Figure A.2 that this is not the case: raw (log) growth displays the same non-linear pattern. Second, working with the natural logarithm of growth has the property of somewhat compressing the tails of the distribution. However, Figure A.3 shows that percent growth exhibits the same non-linear relationship between

Figure 3: Error-Revision Regression Coefficient by Percentiles of Revisions



Notes: In this Figure, we report the estimates of  $\beta$  in the following regression:

$$ERR_t g_{it+1} = \alpha + \beta R_t g_{it+1} + \epsilon_{it+1}$$

where  $g_{it}$  is the normalized log sales growth rate defined in Section 1. This regression is run on eight different subsamples, whose ranges are described in the x-axis of this chart. These subsamples corresponds to the tails and the bulk of the distribution of revisions. The point estimate of  $\beta$  is the solid black line, while the dashed lines corresponds to the 95% confidence interval based on standard errors that are double-clustered by firm and year.

errors and revisions. Third, we attempt to control for aggregate shocks by making a different adjustment to growth. Each year, we compute the cross-sectional mean absolute deviation of raw log growth (from that year's mean log growth) as a measure of aggregate dispersion. We then divide raw error and raw revision by this measure of dispersion. [Figure A.4](#) shows that, after such adjustment for time-varying aggregate volatility, the non-linear relationship between errors and revisions is still very strong. Fourth, because we seek to maximize the size of our sample, our data include non-US firms and therefore somewhat differs from most existing research. We show in [Figure A.5](#) that the non-linear pattern is clearly present both in US and international firms separately. Fifth, forecasting noise has been suggested as a potential source of downward bias in the CG coefficient ([de Silva and Thesmar, 2024](#)). To assess the importance of noise, [Figure A.6](#) reimplements the exercise on individual analyst forecasts, which likely contain more noise than the consensus. Consistent with the presence of noise, we find the intermediate positive slope, is smaller and the tail negative slopes are more negative. However, the non-linear relationship between errors and revisions is still very strong, suggesting that noise is unlikely to explain our main fact. Finally,

while the present paper focuses on sales growth because of its empirically convenient properties, most literature on analyst forecasts analyzes EPS forecasts. In [Figure A.7](#) we show that raw EPS forecast exhibit the same type of non-linearity, albeit slightly less pronounced (it is also present in [Bouchaud et al. 2019](#)).<sup>2</sup>

Overall, the evidence on forecast errors and revisions points towards a different treatment of large versus smaller shocks. Such evidence is hard to square with established models of expectations formations, which feature linear DGPs (typically, AR1 models) and linear expectations models. We will deviate from the existing literature in allowing for fat tails in the growth process. To guide our theory, we first document two additional facts on the fat tails of firm dynamics.

## 2.2 Fact #2: Fat Tails in Distribution of Sales Growth

Our second fact is that the distribution of sales growth has fat tails, which informs the specification of the data-generating process in our model. 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 Pareto 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](#)).

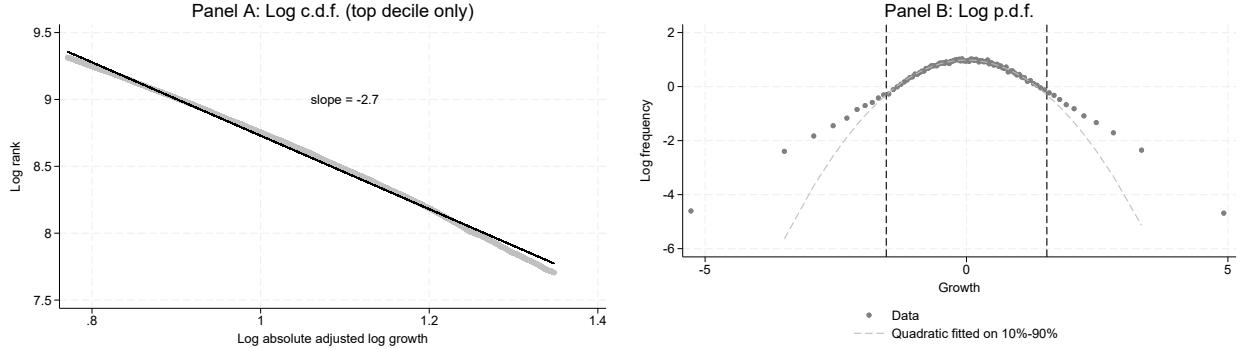
[Figure 4](#) examines the tails of log normalized growth,  $g_{it}$ , in our sample. Panel A shows a classic plot in the literature on power laws: a plot of log rank against  $\log |g_{it}|$  in the tail of the distribution, where each point corresponds to an observation ([Gabaix, 2009](#)). In this plot, we focus on the top 10% of observations of absolute growth  $|g_{it}|$ , and exclude the top 1%. The plot also shows the OLS regression line, which has a slope of -2.7. The fact that the relationship between these two variables is approximately linear shows that in the top decile of the distribution, the density function of  $g_{it}$  is well-approximated by a power law of tail coefficient 2.7 (it has a variance but no kurtosis).

Another way to illustrate the fat tails in the distribution of sales growth is shown in Panel B of [Figure 4](#). This plot shows the log of the probability density function of  $g_{it}$  for the entire distribution. We compute this density by first grouping observations into centiles of sales growth. For each centile, we then compute the average growth, which is shown on the x-axis. On the y-axis, we

---

<sup>2</sup>To normalize EPS, we use the standard practice of dividing by a common lagged value of stock price, as in [de Silva and Thesmar \(2024\)](#). Hence, the forecast error is defined as  $\frac{EPS_{it+1} - F_{jt}EPS_{it+1}}{P_{it-2}}$  for analyst  $j$ , firm  $i$  at date  $t$ . Similarly, revisions are defined as  $\frac{F_{jt}EPS_{it+1} - F_{jt-1}EPS_{it+1}}{P_{it-2}}$ .

Figure 4: CDF and PDF of Sales Growth Distribution



*Notes:* Panel A of this figure shows a scatterplot of log rank of  $|g_{it}|$  against log  $|g_{it}|$ . We restrict ourselves to the top decile of absolute growth and remove the top one percent. The panel reports the slope of the regression of log rank on log growth estimated by OLS. Panel B shows the log density of  $g_{it}$  computed as follows. For each centile, we estimate density as the log of the number of observations in the centile divided by its range. The dashed line is a quadratic fit on the centiles between the 10th and 90th centiles. The two dashed vertical lines correspond to the cutoff values of the top and bottom decile of the distribution.

calculate the density as the difference between the log frequency in the centile (equal to 1/100) and the log range of the centile, normalized the overall range of growth in the sample.

As a point of comparison, Panel B also shows the fit of a quadratic approximation between the 10th and 90th percentiles of the distribution. In particular, for centile  $c \in [10; 90]$ , we estimate the following relationship:

$$\log h(g_c) = \alpha - \frac{1}{\Sigma^2} \frac{g_c^2}{2} + \epsilon_c$$

where  $g_c$  is the mean growth of centile  $c$  and  $\log h(g_c)$  the corresponding log density. In the dataset made of these 80 centiles, the  $R^2$  of this regression is 0.98, indicating the fit is quite good in the bulk of the distribution. If the distribution of sales growth was Gaussian, its log-PDF would be well-approximated by a quadratic function for the entire distribution. However, in the top and bottom decile, the log density is much larger than predicted by the quadratic fit, illustrating the presence of non-Gaussian fat tails.

**Robustness.** In the Appendix, we examine the robustness of our second fact in several ways. First, in [Figure A.8](#), we check the fat tails are not driven by our use of normalized log growth. As expected, we find the opposite is true: in the absence of adjustment, the asymptotic tail coefficient is estimated to be 2, lower than the 2.7 obtained after adjustment. This is consistent with the idea that part of the tail thickness in raw log growth comes from heterogeneous variances. Second, in [Figure A.9](#), we look at percent growth instead of raw growth. We find that percent growth has thinner tails with an estimated tail coefficient of 3.5, but is still far from being well-approximated

by a Gaussian distribution. Finally, [Figure A.10](#) shows log growth rates adjusted for time variation in mean growth and mean absolute distance in the cross-section. This attempts to correct aggregate changes in mean and volatility of growth. We find that the tails of this distribution are still very thick, with a tail coefficient of 2.5.

### 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 current growth is non-linear in past growth. To illustrate this point, Panel A of [Figure 5](#) shows a binned scatter plot of  $g_{it}$  versus  $g_{it-1}$ . As is evident from the figure, the conditional expectation of  $g_{it}$  conditional on  $g_{it-1}$  exhibits significant non-linearity reminiscent of the non-linearity in the relationship between forecast errors and revisions in [Figure 2](#).

Panels B, C, and D of [Figure 5](#) zoom in on the three parts of the distribution of past growth. These panels illustrate how  $g_{it}$  is approximately linearly-increasing in  $g_{it-1}$  in the bulk of the distribution, while it is decreasing in  $g_{it-1}$  in the tails of the distribution.<sup>3</sup> In [Figure A.13](#), we provide a statistical evidence of non-linearity analogous to [Figure 3](#) by splitting the sample of past growth into 8 quantiles, and then, within each quantile, regressing  $g_{it}$  on  $g_{it-1}$  (as with revisions we double cluster at the firm and year level). We find the slope is significantly positive in the bulk and significantly negative in the tails of past growth. Put differently, growth appears to be persistent for intermediate levels of past growth, but mean-reverting in the tails.

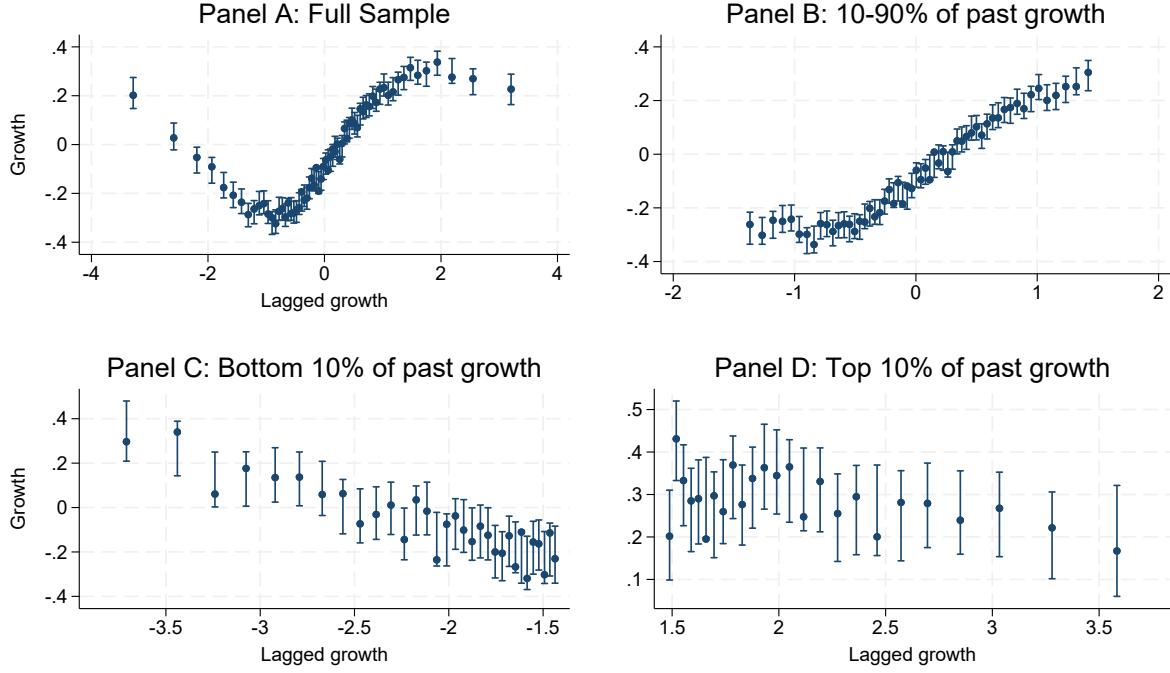
## 3 Model

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

---

<sup>3</sup>In [Figure A.11](#) and [Figure A.12](#), we show that this same non-linearity is present for raw log growth, as well as percent growth.

Figure 5: Non-Linear Relationship Between Current and Past Growth



*Notes:* This figure shows binned scatterplots of current growth on past growth for normalized log sales growth. Panel A shows the results for the entire sample; Panel B restricts the sample to the 10-90th percentiles of past growth; Panel C restricts the sample to below the 10th percentile of past growth; Panel D restricts the sample to above the 90th percentile of past growth.

### 3.1 First Block: Data-Generating Process

The first piece of the model is the data-generating process (DGP) for adjusted log growth,  $g_t$ , our forecasting variable of interest. Since this has already been adjusted, our model does not have any firm-specific heterogeneity, and we therefore omit the firm index  $i$  for brevity. 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

$$g_{t+1} = g_{t+1}^* + \sigma_\epsilon \epsilon_{t+1}, \quad \epsilon \sim f(\cdot), \quad (2)$$

$$g_{t+1}^* = \rho g_t^* + \sigma_u u_{t+1}, \quad u_t \sim N(0, \sigma_u^2), \quad (3)$$

where  $u_{t+1}$  and  $\epsilon_{t+1}$  are IID shocks with a unit variance. The DGP for  $g_t$  therefore consists of two components: (i) a persistent component,  $g_t^*$ , that follows an AR1 process with normal shocks, and (ii) a transitory shock,  $\epsilon_t$ , with PDF  $f(\cdot)$ . 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 (normal) PDF by  $\phi_{g^*}(\cdot)$ .

We also denote the marginal PDF of  $g_t$  by  $h(\cdot)$ , which is given by:

$$h(g) = \int_{-\infty}^{+\infty} \phi_{g^*}(g - \epsilon) f(\epsilon) d\epsilon.$$

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

We now introduce the key way in which our model of the DGP differs from existing literature on expectations formation. Unlike existing literature, we assume the PDF of  $\epsilon_t$ ,  $f(\cdot)$ , has heavy tails in the sense that it is well-approximated by a power-law for large values of  $\epsilon_t$ . Formally, we assume that:

$$f(|\epsilon|) \propto |\epsilon|^{-\nu} \text{ as } |\epsilon| \rightarrow \infty \quad (4)$$

Given that the sum of a Gaussian and a random variable that asymptotically follows a power law also exhibits a power law in its tail, our DGP for  $g_t$  will be well-approximated by a power law in the tails, consistent with our second empirical fact.

The specification of growth as the sum of a persistent and transitory component is consistent with [Lettau and Wachter \(2007\)](#), who use this same model with a normally-distributed  $\epsilon_t$  to model dividend growth. Additionally, this specification of the DGP is reminiscent of those used in the recent literature on income dynamics ([Guvenen et al., 2014, 2021](#)), which models income processes as the sum of persistent and transitory components. However, a key difference from these models is that fat tails typically come from persistent shocks that have non-normal distributions (often a mixture of normals), unlike our model in which transitory shocks have fat tails. The reason for this difference is that a model with fat-tailed persistent shocks, while being consistent with our second fact, is inconsistent with our third fact: in a model with fat-tailed persistent shocks, large values of  $g_{t-1}$  would be more positively related to  $g_t$ , the opposite of what we find in the data.

### 3.2 Second Block: Belief Formation

Given the DGP, the second building block of our model is 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 distribution of  $\epsilon_t$  to be Gaussian such that (4) does not hold. We do not micro-found 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 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  $\epsilon_t$  as being Gaussian, the solution to this filtering problem implies that their expectations will be characterized by the Kalman filter. In our theoretical analysis, 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  $K$ , agent’s expectations at horizon  $k$ ,  $F_t g_{t+k}$ , are characterized by:

$$F_t g_{t+k} = \rho^k K \sum_{s \geq 0} (\rho(1 - K))^k g_{t-s}. \quad (5)$$

### 3.3 A Simplified Model to Build Intuition

To build intuition about the bias of forecasting rule (5), we assume *in this section only* that forecasters have a “short-memory” and only use  $g_t$  to forecast  $g_{t+k}$ , in addition to misperceiving  $\epsilon_t$  as being Gaussian. This assumption buys us a few analytical results that provide intuition valid with the more general forecasting rule in (5). Formally, we assume that agents forecast future values of  $g_{t+k}$  using the population regression coefficient of  $g_{t+k}$  on  $g_t$ :

$$\hat{F}_t g_{t+k} = \rho^k \frac{\sigma_{g^*}^2}{\sigma_g^2} g_t. \quad (6)$$

We refer to this forecast as the “linear” forecast. The key advantage of working with the simplified forecasting rule is that it is easy to assess forecast rationality (holding fixed forecasters’ short memory) by simply comparing  $\hat{F}_t g_{t+k}$  to  $E(g_{t+k}|g_t)$ : linear forecasters overreact (underreact) when  $\rho \frac{\sigma_{g^*}^2}{\sigma_g^2}$ , which does not depend on  $g_t$ , is larger (smaller) than the sensitivity of rational expectations  $E(g_{t+1}|g_t)$  to  $g_t$ , which will depend on  $g_t$  given our third fact. In the following proposition, we use a result from statistical theory known as Tweedie’s formula (Efron, 2012) to characterize  $E(g_{t+1}|g_t)$  directly.

#### **Proposition 1. Characterizing short-memory rational expectations**

*Assume that the only information available to rational and subjective forecasters is the current realization,  $g_t$ . Then, the following results hold:*

- The rational expectation is a function of the (observable) distribution of  $g$ :

$$E(g_{t+1}|g_t) = -\rho\sigma_{g^*}^2 \frac{d \log h}{dg}(g_t). \quad (7)$$

- As  $|g_t| \rightarrow \infty$ ,

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

- The linear forecaster overreacts to  $g_t$  relative to the rational expectation if and only if:

$$\left( \frac{d^2}{dg^2} \log h(g) \right)(g) > -\frac{1}{\sigma_g^2} \quad (9)$$

and underreacts if the inequality is reversed. When  $\epsilon$  is normal, this relationship holds with equality, meaning the linear forecaster is rational.

- The conditional variance of  $g_{t+1}$  given  $g_t$  is given by:

$$\text{var}(g_{t+1}|g_t) = \sigma_g^2 + \rho^2\sigma_{g^*}^4 \left( \frac{d^2}{dg^2} \log h(g) \right)(g_t). \quad (10)$$

*Proof.* See Appendix B.1. □

Proposition 1 provides several insights, all drawn from Tweedie's formula, (7). This equation informs us about the shape of  $E(g_{t+1}|g_t)$ . While we do not have a closed-form for  $h(\cdot)$ , we can measure it directly in the data. In the bulk of the distribution, Panel B of Figure 4 shows that the growth distribution is approximately Gaussian:  $\log h(g) \approx -\frac{g^2}{2\Sigma^2} + \text{constant}$ . Given this approximation, the rational expectation of  $g_{t+1}$  given  $g_t$  in the bulk of the distribution is then:

$$E(g_{t+1}|g_t) \approx \rho \frac{\sigma_{g^*}^2}{\Sigma^2} g_t, \quad (11)$$

which is linear in  $g_t$ . This is the well-known result: the rational expectations given a signal of a state with additive Gaussian noise is linear. The benefit of (7) is that it allows us to extend this result to distributions that are locally Gaussian. In a nutshell, with short memory, rational expectations should be *linear* for moderate values of  $g_t$ , which is consistent with Figure 5.

Outside of the bulk of the distribution, Panel A of Figure 4 shows that the distribution  $h(\cdot)$  looks like a power law in the top decile of absolute growth. Denoting the tail parameter of this power law by  $\nu$  (in our model, the tail of  $g_t$  is dominated by  $\epsilon$ , whose tail parameter is  $\nu$ ), and assuming  $h(\cdot)$

is symmetric, we obtain  $\log h(g) \approx -\nu \log(|g|)$  for large  $|g|$ . Plugging this approximation into (7) allows us to characterize the rational expectation in the bulk of the distribution:

$$E(g_{t+1}|g_t) \approx \rho \sigma_{g^*}^2 \frac{\nu}{g_t}. \quad (12)$$

which shows that, in the tails of growth distribution, the rational expectation of  $g_{t+1}$  is *decreasing* in past growth, again consistent with Figure 5. Equation (8) shows that the above formula applies asymptotically.<sup>4</sup> The intuition is that very large values of  $g_t$  are likely coming from the fat-tailed part of the process, which is not persistent. In the limit where  $g_t$  is very large in absolute value, it is completely uninformative about the persistent latent state  $g_t^*$ . As a result, the rational forecaster reverts to the unconditional expectation as her best possible forecast. This is in stark contrast with the linear forecaster of equation whose prediction increases linearly with  $g_t$ .

The second insight of Proposition 1 is in the third bullet point, which follows directly from the first one. Inequality (9) comes from comparing the local sensitivity of the rational expectation to  $g_t$  and the slope of subjective expectations. It shows that the linear forecaster is too sensitive to  $g_t$  (compared to the rational one) when  $h$  is not log concave enough. This is reminiscent of O'Hagan (1979) theory of outlier rejection, where log convex densities are a necessary condition for outlier proneness.<sup>5</sup> If  $h(\cdot)$  is exactly Gaussian, then (9) holds with equality and expectations neither under- nor overreact. But because  $h(\cdot)$  is approximately a power law when  $|g|$  is large, it is log-convex, which in turn leads to overreaction. Conversely, for small values of past growth,  $h(\cdot)$  is locally Gaussian and underreaction holds if  $\Sigma > \sigma_g$ , i.e. when the standard deviation of the normal approximation in the bulk is lower than that of the entire distribution, as is the case in Figure 4.

The final bulletpoint in Proposition 1 describes the precision of the rational forecast, and allows to connect our theory with recent literature. In line with intuitions above,  $g_t$  is a less precise signal about  $g_{t+1}$  when  $h(\cdot)$  is more log-convex. As shown in Figure 4,  $\log h(\cdot)$  is concave in the bulk (Panel B), but convex in the tails (Panel A). Recast into equation (10), this means that  $g_t$  is a more precise signal of  $g_{t+1}$  for smaller absolute values of past growth, while it is less precise in the tails of the distribution. This aspect connects our theory with that of Augenblick et al. (2024), in which forecasters rely too much on weak signals and too little on strong ones. Proposition 1 shows that our linear forecaster does exactly that: she relies too much on  $g_t$  in the tails, where it is not very informative, and too little in the bulk, where it is more informative. Equation (10) ties this intuition with the log convexity of  $h$ .

---

<sup>4</sup>We are grateful to Xavier Gabaix for helping us with this asymptotic algebra.

<sup>5</sup>We thank Kelly Shue for pointing out this reference to us.

### 3.4 The Forecast Error-Revision Relationship in the Full Model

We now revert back to our core assumption that subjective forecast differ from rational only in that they ignore non-normality. Thus, forecasters now have long memory, and form forecasts according to the (steady-state) Kalman filter in (5). Defining errors and revisions as in our empirical section:

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

the following proposition characterizes the relationship between the two.

**Proposition 2. Relationship between errors and revisions under Kalman filter**

1. *Errors are asymptotically linear in revisions. In particular, as  $|REV_t| \rightarrow \infty$ ,*

$$E(ERR_{t+1}|REV_t) \rightarrow \underbrace{\frac{1}{K} \left[ \frac{1}{1 + \frac{K^2 \rho^2 + \sigma_u^2 / \sigma_\epsilon^2}{1 - (1-K)^2 \rho^2}} - \frac{1}{1 + \frac{K^\nu \rho^\nu}{1 - (1-K)^\nu \rho^\nu}} \right]}_{\equiv \mu} REV_t.$$

Therefore, if  $\nu \geq 2$ , then there is overreaction in the tails of revisions ( $\mu < 0$ ).

2. *There exists a threshold  $\bar{R}$  such that:*

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

Therefore, there is underreaction in the bulk of the distribution of revisions.

*Proof.* See Appendix B.2. □

This proposition shows that our model of the data-generating process, which can replicate our second and third facts by construction, can also replicate our first fact when forecasters form forecasts using the Kalman filter. The first part of the proposition is the most crucial. The challenging part of showing this result is that past revisions of the linear forecaster may be large and positive for two reasons: (1) the most recent realization of  $g_t$  is very positive due to a large positive realization of  $\epsilon_t$ , or (2) past realizations  $g_{t-k}$  are very negative. Both of these histories will push the linear forecast too high, and generate negative future errors, implying overreaction. The first point of the

proposition shows that this always happens, as long as  $\nu \geq 2$ .<sup>6</sup> The second point is straightforward. Since the linear forecaster uses the optimal Kalman gain, the unconditional covariance between errors and revisions is, by definition, zero. Since for large revisions, the conditional covariance is negative, it has to be positive, on average, for non-extreme revisions.

## 4 Quantitative Fit of the Model

This section assesses whether the model in Section 3 can quantitatively account for our three main empirical facts.

### 4.1 Simulation Details

The model that we take to the data is the same model described in Section 3 with two changes. First, 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 power law with tail parameter  $\nu$ , and has the nice property of converging to a normal distribution as  $\nu \rightarrow \infty$ . Second, 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 (2) and (3) under the (incorrect) assumption that  $\epsilon_t \sim N(0, 1)$ :

$$F_t g_{t+h} = \rho^h F_t g_t^* \quad (13)$$

$$\begin{aligned} F_t g_t^* &= (1 - K_t) F_{t-1} g_t^* + K_t g_t \\ K_t &= \frac{\Sigma_t}{\Sigma_t + \sigma_\epsilon^2} \\ \Sigma_{t+1} &= \rho^2 (1 - K_t) \Sigma_t + \sigma_u^2 \\ F_0 g_0^* &= g_0^*, \quad \Sigma_0 = 0 \end{aligned} \quad (14)$$

In our simulations, we sample time series of  $g_t$  according to equations (2) and (3) with 100,000 observations. We repeat this simulation 100 times, where  $g_0^*$  is drawn from its stationary distribution,

---

<sup>6</sup> $\nu \leq 2$  is a pathological case where no variance exists. In our data, we estimate a value of  $\nu$  above 2.

and 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>7</sup>

## 4.2 Estimating Parameters of the DGP

We start by estimating the parameters of the DGP  $g_t$  (equations (2)-(3)). Our model has four parameters that control the data-generating process:  $\rho$ , 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. We use the inverse-diagonal covariance as the weighting matrix.<sup>8</sup>

We use 9 moments to estimate the 4 parameters. First, we use 6 slope coefficients from regressing  $g_{t+1}$  on  $g_t$  for 6 different subsamples based on percentile breakpoints of  $g_t$ : [0; 10], [10; 20], [20; 50], [50; 80], [80; 90], [90; 100]. We also ask the estimated model to match four additional moments: (1) the Pareto tail coefficient from Panel A of Figure 4, (2) the 10-90 percentile difference of  $g_t$  (3) the 10-90 percentile difference of  $g_{t+1} - g_t$ , and (4) the 10-90 percentile difference of  $g_{t+3} - g_t$ .

The idea behind these moments is to try to get the model to match Facts #2 and #3 in the data. The first six regression slopes capture Fact #3: the non-linearity in the conditional expectation of  $g_{t+1}$  given  $g_t$  in Figure 5.<sup>9</sup> These regression coefficients jointly identify  $\rho$ ,  $\sigma_u$ , and  $\nu$ , but they do not separately identify each of these parameters. Therefore, we include a seventh statistic that captures Fact #2: the Pareto tail coefficient in Figure 4. This moment is useful because it is only affected by  $\nu$  and, hence, can be used to separately identify it. To summarize, the 6 slopes and the tail coefficient jointly help pin down the Gaussian and non-Gaussian part of the process, but not its scale.

The eighth statistic, the dispersion of  $g_t$ , serves this purpose. We choose to use the spread between the 10th and 90th percentiles to avoid sensitivity to outliers, which inevitably occur in a process with fat tails. This statistic primarily identifies the scale of the DGP, in particular  $\sigma_u$  and  $\sigma_\epsilon$ . The

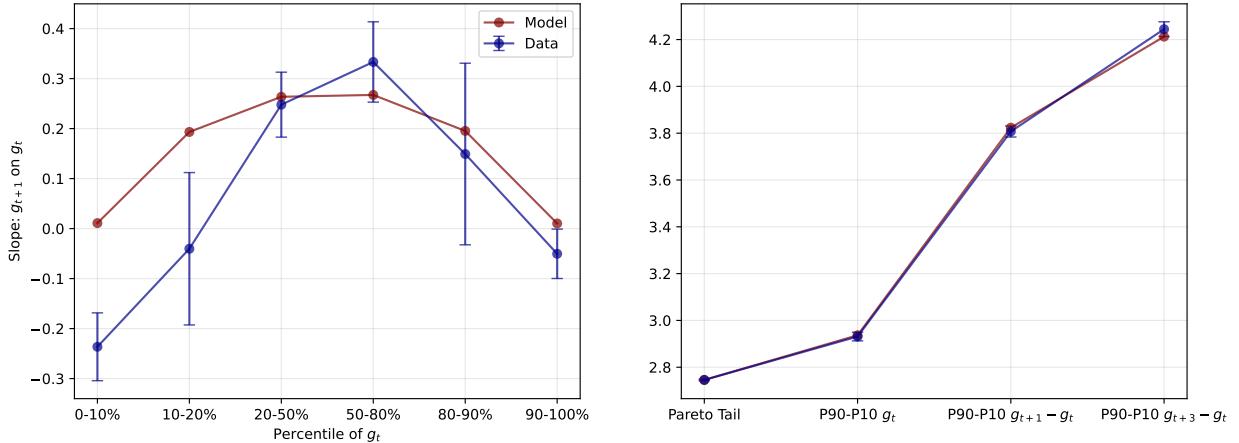
---

<sup>7</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 because when we compute the rational expectation using a particle filter described below.

<sup>8</sup>Because the SMD objective function depends on four parameters, we need to be careful that we reach a global rather than local minimum. We perform this optimization using a two-step procedure in which we first search of a quasi-random grid of 20,000 points, and then run local Nelder-Mead optimizations using the top 10 points as starting points. Our result is then the parameter vector that has the lowest objective function from any of these Nelder-Mead optimizations. We have verified that with our choice of statistics, these local optimizations all converge to similar points.

<sup>9</sup>Before running these regressions in both the model and the data, consistent with our empirical analysis, we trim observations at the 1st and 99th percentiles to avoid the influence of extreme outliers.

Figure 6: Fit of Estimated Model on Data-Generating Process



*Notes:* This figure shows the fit of the estimated model on the statistics used to estimate it. The values of the statistics in the data are shown with 95% confidence intervals. The values in the model are computed from simulations, as described in Section 4.1, at the set of parameters shown in (15).

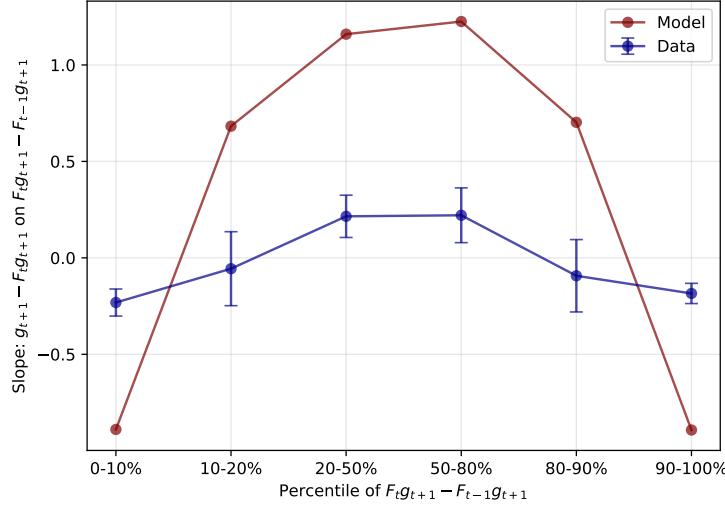
final two statistics that we use are the spread between the 10th and 90th percentiles of one- and three-year changes in  $g_t$ . These additional statistics are useful for separately identifying  $\rho$  and  $\sigma_u$ . Without these moments, the estimation has several local minima with different values of  $\rho$  and  $\sigma_u$ . Including these moments is useful because the extent to which the three-year changes tend to larger than one-year changes is directly affected by  $\rho$  and no 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 procedure delivers the following estimates:

$$\rho = 0.529, \sigma_u = 0.631, \sigma_\epsilon = 1.325, \nu = 2.533. \quad (15)$$

Figure 6 shows the fit of the model on the targeted statistics using the parameters in (15). The results show that the model is able to fit all the moments in the data relatively well. The main dimension on which the model misses is relationship between  $g_{t+1}$  and  $g_t$  in the left tail of the distribution of  $g_t$ . In the data, there is some mean reversion that manifests in a negative slope coefficient, which the model cannot generate.

Figure 7: Fit of Estimated Model on Fact #1: Non-Linearity in Error-Revision Relationship



*Notes:* This figure shows the fit of the model on the error-revision relationship in the data. The values of the statistics in the data are shown with 95% confidence intervals, which are repeated from Figure 3. The values in the model are computed from simulations, as described in Section 4.1, at the set of parameters shown in (15). The left panel shows results when agents form forecasts using a simplified short-memory linear rule based on the population regression coefficient of  $g_{t+1}$  on  $g_t$ . The right panel shows results when agents form forecasts according to the Kalman filter in (13). Before running these regressions in both the model and the data, we trim observations at the 1st and 99th percentiles to avoid the influence of extreme outliers.

### 4.3 Estimating the Expectations Formation Model

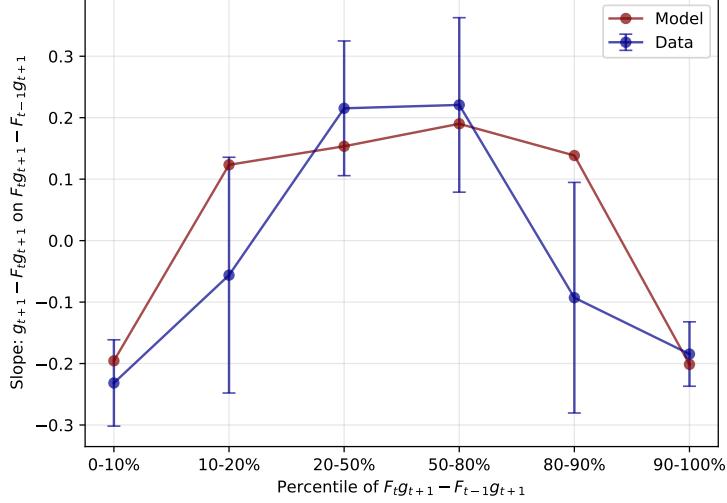
We now turn to the model fit of our first fact. First, we note that the expectations model (13), based on the Kalman filter, generates “too much” bias on our estimated DGP. We show this in Figure 7. Consistent with our theoretical results in Section 3.4, the Kalman filter generates a non-linear relationship between errors and revisions, with positive coefficients (i.e., underreaction) for intermediate revisions and negative coefficients (i.e., overreaction) for large revisions. However, relative to the data, the model generates too much non-linearity: the error-revision slope is too positive in the bulk of revisions, and too negative in the tails.

To attenuate the bias of our linear forecasting model, we allow forecasts to be anchored in rational expectation (as in Fuster et al. 2010 and Gabaix 2019). In particular, we assume that expectations are a weighted-average of the Kalman filter and the true (non-linear) rational expectation:

$$F_t^\lambda g_{t+h} = \lambda F_t g_{t+h} + (1 - \lambda) E_t g_{t+h} \quad (16)$$

where  $E_t$  is the rational expectation given all past history of  $g_t$  and  $\lambda$  is to the weight that is placed on Kalman filter forecast.  $\lambda = 1$  corresponds to the case in Figure 7, while  $\lambda = 0$  corresponds to

Figure 8: Fit of Estimated Model with  $\lambda = 0.290$  on Fact #1: Non-Linearity in Error-Revision Relationship



*Notes:* This figure shows the fit of the estimated model with the forecasting equation (16). The values of the statistics in the data are shown with 95% confidence intervals. The values in the model are computed from simulations, as described in Section 4.1, at the set of parameters shown in (15) and (17).

the case of rational expectations in which error-revision coefficients would always be zero.

We estimate  $\lambda$  to assess that the anchored model in (16) can match our first fact quantitatively. In order to do this, we first need to compute the rational expectation,  $E_t g_{t+1}$ , which we do using the particle filtering algorithm from [Fernandez-Villaverde and Rubio-Ramirez \(2007\)](#) (also known as sequential importance sampling). See Appendix C for a detailed description of this procedure. Having computed the rational expectation, we then estimate  $\lambda$  using SMD with six statistics and the inverse-diagonal covariance as the weighting matrix. The six statistics that we use are the error-revision regression coefficients in [Figure 7](#). This SMD procedure delivers the following estimate:

$$\lambda = 0.290. \quad (17)$$

Our estimated value of  $\lambda$  implies that forecasters place 29% of their weight on the Kalman filter forecast, and 71% weight on the true forecast. [Figure 8](#) shows the fit of the model on the targeted error-revision coefficients using the parameters in (15) and (17). The fit is much better than in [Figure 7](#), where the model predicts way too much non-linearity. Now almost all of the error-revision coefficients are within the 95% confidence intervals of the data. One area where the model misses is that it generates positive rather negative coefficients in the 10–20% and 80–90% bins, like in the case of the Kalman filter in [Figure 7](#).

## 4.4 Accuracy Loss Relative to Rational Expectation

Having computed the (limited-information) rational expectation, we next examine its accuracy relative to the Kalman filter and the forecasts formed in (16) with the estimated value of  $\lambda$ . Table 2 shows the percent loss in mean-squared error (MSE) of these two forecasts in separate panels relative to the rational expectation. The different rows correspond to different values of  $\rho$ , while the different columns correspond to different values of  $\nu$ . At the estimated values of  $\rho = 0.529$  and  $\nu = 2.533$ , the loss in MSE from using the Kalman filter is around 1.2%, which is non-trivial. This loss tends to increase as  $\rho$  increases and  $\nu$  increases, which is to be expected. At low values of  $\rho$ , forecasts of  $g_{t+1}$  are not very dependent on the solution to the filtering problem, since  $g_t^*$  is not persistent. At high values of  $\nu$ , the  $f(\cdot)$  becomes closer to a normal distribution, meaning the Kalman filter provides a closer approximation to the true solution. In contrast, the loss in MSE from agents forecasts with  $\lambda = 0.290$  is much smaller. At the estimated values of  $\rho$  and  $\nu$ , the loss is around 0.1%. Like in Panel A, this loss increases as  $\rho$  increases and  $\nu$  decreases. Overall, the cost of forecasting errors induces by the Kalman filter is modest, which supports the idea that it would be a “default” in a model of bounded rationality (Gabaix, 2019).

## 5 Additional Tests of Model Predictions

This section presents two additional tests of our model’s predictions. The first one comes from an online forecasting experiment, and the second comes from data on stock returns.

### 5.1 Forecasting Experiment

The design of our online forecasting experiment is taken from Afrouzi et al. (2023). Participants are asked to predict the outcome of an AR1 process with parameter values that vary across conditions, and their compensation depends on the accuracy of their forecasts. Participants are recruited on Amazon MTurk and 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, although robustness checks confirm that this is not important. The interface is graphical and user-friendly: participants click with their mouses 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

Table 2: Percent Loss in One-Period Ahead Mean-Squared Error Relative to Rational Expectation

<i>Panel A: Kalman Filter Forecast</i>								
$\rho$	$\nu$							
	2.1	2.5	2.533	3.0	3.5	4.0	4.5	5.0
0.1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
0.2	0.1%	0.2%	0.2%	0.1%	0.1%	0.1%	0.1%	0.0%
0.3	0.2%	0.4%	0.4%	0.3%	0.2%	0.2%	0.1%	0.1%
0.4	0.4%	0.7%	0.7%	0.6%	0.4%	0.3%	0.2%	0.2%
0.5	0.6%	1.1%	1.1%	0.9%	0.7%	0.5%	0.4%	0.3%
0.529	0.7%	1.2%	1.2%	1.0%	0.8%	0.6%	0.4%	0.3%
0.6	1.0%	1.6%	1.6%	1.3%	1.0%	0.7%	0.5%	0.4%
0.7	1.4%	2.2%	2.2%	1.8%	1.4%	1.0%	0.8%	0.6%
0.8	1.9%	3.0%	2.9%	2.4%	1.8%	1.4%	1.1%	0.8%
0.9	2.5%	3.6%	3.6%	3.0%	2.3%	1.8%	1.4%	1.1%

<i>Panel B: Forecast in (16) with Estimated <math>\lambda</math></i>								
$\rho$	$\nu$							
	2.1	2.5	2.533	3.0	3.5	4.0	4.5	5.0
0.1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
0.2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
0.3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
0.4	0.0%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
0.5	0.1%	0.1%	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%
0.529	0.1%	0.1%	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%
0.6	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.0%	0.0%
0.7	0.1%	0.2%	0.2%	0.2%	0.1%	0.1%	0.1%	0.0%
0.8	0.2%	0.3%	0.2%	0.2%	0.2%	0.1%	0.1%	0.1%
0.9	0.2%	0.3%	0.3%	0.2%	0.2%	0.1%	0.1%	0.1%

*Notes:* This table shows the percent loss in mean-squared error (MSE) in model simulation of different one-period ahead forecasts relative to the one-period ahead rational expectation computed using Algorithm 1. Panel A shows the results for the Kalman filter forecast, computed as in (13); Panel B shows the results for the forecast in (16) with the estimated value of  $\lambda$  in (17). The different rows correspond to different values of  $\rho$ , while the different columns correspond to different values of  $\nu$ . For each different combination of  $\rho$  and  $\nu$ ,  $\sigma_e$  and  $\sigma_u$  are adjusted so that the analytical variance of  $g_t$  and  $g_t^*$  remain the same as in the model with the estimated parameters in (15).

each period. We refer the reader to Afrouzi et al. (2023) for further details about this design; see Figure A.14 an example of the interface our participants see for these parameters.

Starting with this design, we modify the DGP in the experiment to the following:

$$g_{t+1} = g_{t+1}^* + 0.608\epsilon_{t+1} \quad (18)$$

$$g_{t+1}^* = 0.529g_t^* + 0.631u_{t+1} \quad (19)$$

where  $\epsilon$  is drawn from a student distribution with 2.533 degrees of freedom and  $u$  is Gaussian. These particular values comes from a simulated minimum distance estimation of our model's DGP that is described in Section 4. We then ran the experiment on 201 participants on March 17th, 2025. Given that each participant makes 40 forecasts, our data has 8,040 observations. Notice that this is significantly smaller than our empirical data, which has approximately 120,000 observations. As a result, we should expect as much precision in our estimates, especially in the tails of the distribution. This is apparent in [Figure A.15](#), which plots  $g_{it+1}$  against  $g_{it}$  where  $i$  is the participant and  $t$  the round of forecasting. The relationship is clearly non-linear, but the number of observations for which this non-linearity shows up is rather small.

To test for non-linearity in the relationship between errors and revisions, we estimate the following regression

$$ERR_{it+1} = \alpha + \beta REV_{it} + \gamma_L Q1\ REV_{it} + \gamma_H Q1\ REV_{it} + \beta_L ERR_{it} \times Q1\ REV_{it} + \beta_H ERR_{it} \times Q5\ REV_{it} + \nu_{it+1} \quad (20)$$

where  $Q1\ REV_{it}$  and  $Q5\ REV_{it}$  are indicator variables that equal one when revisions are in the top or bottom quintile. Our prediction is that the relationship between errors and revisions is more negative in the tails, so we expect  $\beta_H$  and  $\beta_L$  to be negative. Column (2) of [Table 3](#) shows the results from estimating (20). We find evidence of non-linearity in the bottom and top 20% of revisions like in our sample of growth expectations, but it is only significant in the bottom 20%. Column (3) shows the results from imposing  $\beta_L = \beta_H$ , which results in significant non-linearity at the 5% significant level. This estimate implies that the error-revision coefficient in the bulk of the distribution is -0.28, while it is -0.46 in the tails.

A key difference between these results and those in our model and data, however, is that the slope in the bulk is negative, not positive. This level-shift in the error-revision coefficients likely comes from the experimental set-up, which features expectations noise that biases downwards the relationship between errors and revisions ([de Silva and Thesmar, 2024](#)), and from massive overreaction when the process is weakly persistent ([Afrouzi et al., 2023](#)). To highlight this and the importance of fat tails, columns (4)-(6) of [Table 3](#) replicate the first three columns using data from [Afrouzi et al. \(2023\)](#) where the process is a Gaussian AR1 with  $\rho = 0.2$ .<sup>10</sup> Like in our sample, column (4) shows that the error-revision coefficient is of a similar magnitude and negative on average. However, columns (5) and (6) show that there is no significant non-linearity in this relationship, unlike in the case with non-Gaussian dynamics.

---

<sup>10</sup>We take 0.2 because it is closest experimental condition in [Afrouzi et al. \(2023\)](#) to the regression coefficient of  $g_{t+1}$  on  $g_t$  in our model estimated in Section 4.

Table 3: Non-Linear Relationship Between Forecast Errors and Revisions in Experimental Data

	Dependent Variable: $ERR_{it+1}$					
	Non-Gaussian DGP with Fat Tails			Gaussian AR1		
	(1)	(2)	(3)	(4)	(5)	(6)
Revision	-0.40*** (0.02)	-0.28*** (0.06)	-0.28*** (0.06)	-0.44*** (0.02)	-0.42*** (0.06)	-0.42*** (0.06)
Revision $\times$ Bottom 20 %		-0.27*** (0.09)			-0.12 (0.09)	
Revision $\times$ Top 20 %			-0.11 (0.08)		-0.07 (0.07)	
Revision $\times$ Top & Bottom 20 %				-0.18** (0.08)		-0.09 (0.07)
Top 20 % of revision			-1.46 (4.41)	4.19 (4.24)		-1.10 (2.52)
Bottom 20 % of revision			-11.04** (4.55)	-5.42 (3.76)		-10.17*** (3.38)
Constant	-10.95*** (2.08)	-10.74*** (2.41)	-10.74*** (2.41)	-5.03*** (1.57)	-3.30* (1.70)	-3.30* (1.70)
N	7839	7839	7839	5421	5421	5421

*Notes:* We run variants of regression (20). In columns (1)-(3), we use the experimental data described in the main text, where the DGP is fitted on our data – see equations (18) and (19). In columns (4)-(6), we use data from Afrouzi et al. (2023), where the DGP is a Gaussian AR1 with persistence 0.2. Columns (1) and (4) just regress errors on revision. Columns (2) and (5) estimate equation (20). Columns (3) and (6) impose the restriction that  $\beta_L = \beta_H$ . All standard errors are clustered at the participant level. These regressions have fewer than 8,040 observations because computing forecasts revisions loses one observation per subject.

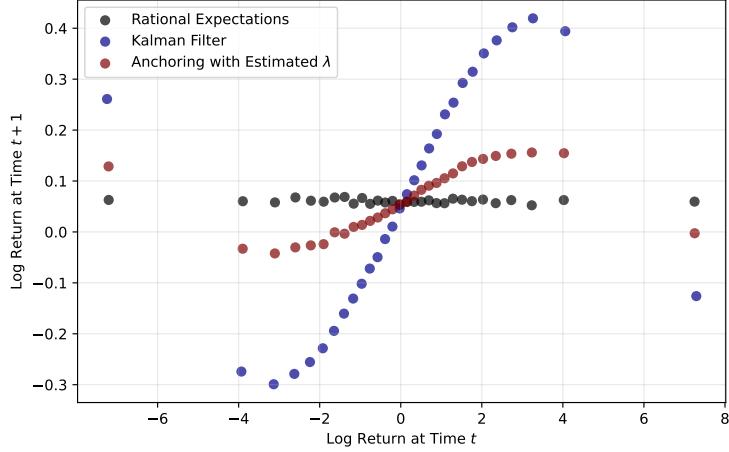
## 5.2 Implications for Return Predictability

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).

### 5.2.1 Model Prediction: Non-Linearity in Momentum

To derive predictions about returns, we two simplifying assumptions: (i) earnings are a constant fraction of sales, and (ii) the subjective discount rate is constant (as in Bouchaud et al. 2019 and Nagel and Xu 2019), which is consistent with evidence in De la O and Myers (2021). These assumptions allow us to simulate a dataset of realized returns consistent with realized and expected profits generated by our estimated DGP and forecasting model. As shown in Appendix B.3, the

Figure 9: 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 (21), as described in Appendix B.3. 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 (16) with the value of  $\lambda$  in (17). We conduct this simulation assuming  $r_f = 1.01$ ,  $\pi = 5.5\%$ ,  $c = 0.96$ , and  $\rho$  is set to the estimated value in (15).

Campbell (1991) decomposition coupled with these assumptions allows us to link returns with changes in expectations of sales growth  $g_t$  using the following expression:

$$r_t = \log(r_f + \pi) + (g_t - F_{t-1}g_t) + \sum_{k=1}^{\infty} c^k REV_t g_{t+k}, \quad (21)$$

where  $\pi$  is the (constant) ERP,  $r_f$  is the risk-free rate,  $c$  is a linearization constant, and  $REV_t g_{t+k}$  denotes investors' subjective revisions about future growth between  $t-1$  and  $t$ . This expression is intuitive: given that discount rates are fixed, returns are driven earnings surprises and revisions of future growth. As detailed in Appendix B.3, we use (21) to generate a panel of simulated returns based on panel of realized and expected sales growth. The parameters used are the ones estimated in the previous section.

Figure 9 shows the prediction that our model makes for momentum by showing a binned scatterplot of (log) returns  $r_t$  on 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, 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.

### 5.2.2 Momentum Non-Linearity in the Data

We now examine if momentum indeed exhibits mean-reversion in the tails in the data. We use CRSP monthly returns from 1927 to 2023, restrict ourselves to all firms listed on NYSE, AMEX and NASDAQ, and adjust returns for delisting. 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 also compute firm size as the market capitalization 12 months prior to  $t$ .

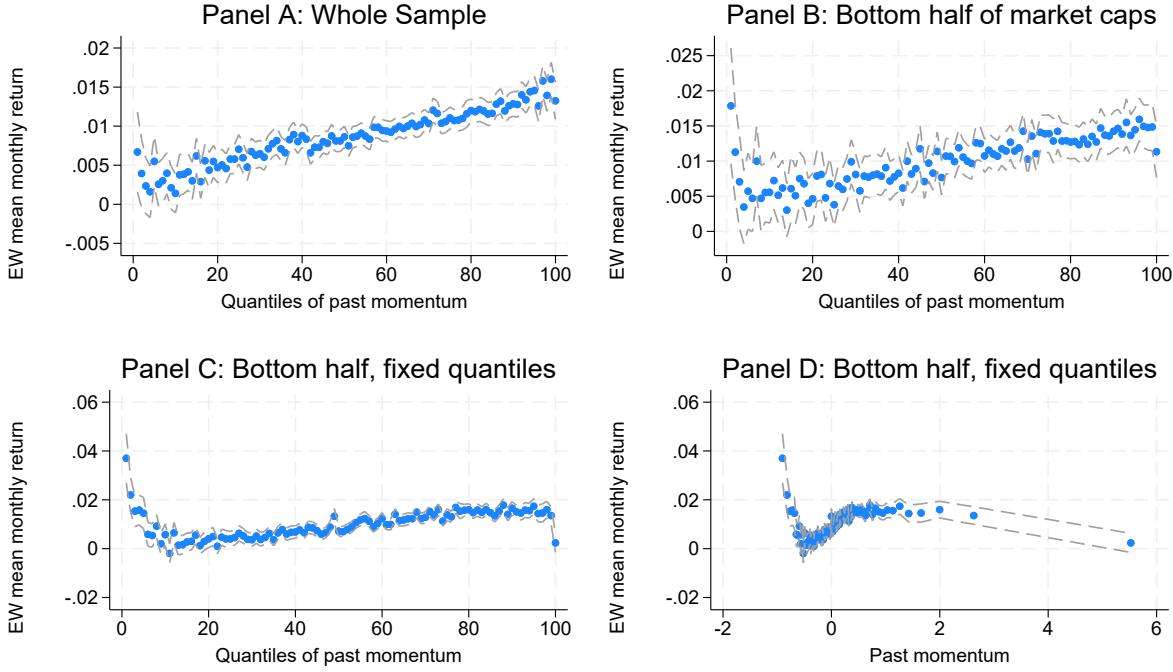
Motivated by [Figure 9](#), we first plot returns against centiles of momentum that are redefined each month, as in standard asset pricing tests. Panel A shows the results for the entire CRSP sample: the line is upward sloping, consistent with the presence of momentum. However, we also find that there is a bit of mean-reversion for “super losers”, but not for “extreme winners”. Panel B shows that the non-linearity is a bit more pronounced for smaller firms, which are defined as firms with below-median size in each month. This is to be expected as smaller stocks are more expensive to trade and therefore should display more predictability (Novy-Marx and Velikov, 2016).

In Panels C and D, we define the centiles on the entire sample, thereby accounting for times series variation in momentum returns. Panel C shows that, using fixed centiles and focusing on firms below the 50% size cutoff, the non-linear relationship starts appearing significantly. This suggests that the time series of momentum returns generate some mean-reversion in the tails: when volatility is high, losers are more likely to overperform, and winners more likely to underperform. This is consistent with the literature on momentum crashes (Clara and Barroso, 2015) and with our model, to the extent that tail risk emerges when aggregate volatility is high. Panel D 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 practice of forming portfolios, as is typically done in asset pricing tests, but is more consistent the x-axis scale in our [Figure 9](#). Again, there is significant non-linearity in the return-momentum relationship, in line with the prediction of our model.

## 6 Conclusion

In this paper, argue that recognizing the complexity of the underlying DGP, in particular it’s fat tails are non-Gaussian dynamics, are crucial for understanding the properties of subjective forecasts. We document three facts using data on sales growth forecasts by equity analysts: (i) the relationship between forecast revisions and future forecast errors—the variables used in [Coibion and](#)

Figure 10: Relationship Between Current and Past Returns in the Data



*Notes:* This figure shows binned scatterplots of monthly returns,  $r_t$ , as 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 months 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.

Gorodnichenko (2015) regressions—is strongly non-linear; (ii) the distribution of the underlying process has fat tails; and (iii) the conditional expectation of future sales growth is non-linear in current growth, with 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. After showing formally that our model can explain the three facts we documented in the data, we estimate it and show that it does so quantitatively. Finally, we show that our framework is consistent with evidence from an online forecasting experiment where the underlying process is non-Gaussian and that it provides an explanation for 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 micro-foundation for why forecasters ignore fat tails in data-generating processes, which would allow us to study how this bias would manifest for other data-generating processes. Second, it would be useful to try to estimate our shrinkage

parameter using other data on subjective forecasts. Variation in this shrinkage parameter across different data-generating processes would be useful for disciplining a more micro-founded model of belief formation.

## References

- Afrouzi, Hassan, Spencer Yongwook Kwon, Augustin Landier, Yueran Ma, and David Thesmar, 2023, Overreaction in expectations: Evidence and theory, *Quarterly Journal of Economics* .
- Augenblick, Ned, Eben Lazarus, and Michael Thaler, 2024, Overinference from weak signals and underinference from strong signals, *Quarterly Journal of Economics* .
- Axtell, Robert L., 2001, Zipf Distribution of U.S. Firm Sizes, *Science* 293, 1818–1820.
- Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer, 2020, 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.
- 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, Publisher: Wiley Online Library.
- Campbell, John Y., 1991, A Variance Decomposition for Stock Returns, *The Economic Journal* 101, 157, Publisher: Oxford University Press (OUP).
- 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.
- Clara, Pedro Santa, and Pedro Barroso, 2015, Momentum has its moments, *Journal of Financial Economics* 116, 111–120.
- 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* .

Denuit, Michel, Patricia Ortega-Jimenez, and Christian Robert, 2024, Conditional expectations given the sum of independent random variables with regularly varying densities, Technical report, LIDAM Discussion Paper ISBA.

Efron, Bradley, 2012, Tweedie's formula and selection bias, *Journal of the American Statistical Association* .

Farmer, Leland E., Emi Nakamura, and Jón Steinsson, 2021, Learning About the Long Run, *SSRN Electronic Journal* .

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* .

Graeber, Thomas, Christopher Roth, and Marco Sammon, 2024, Categorical processing in an imperfect world, Technical report.

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.

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.

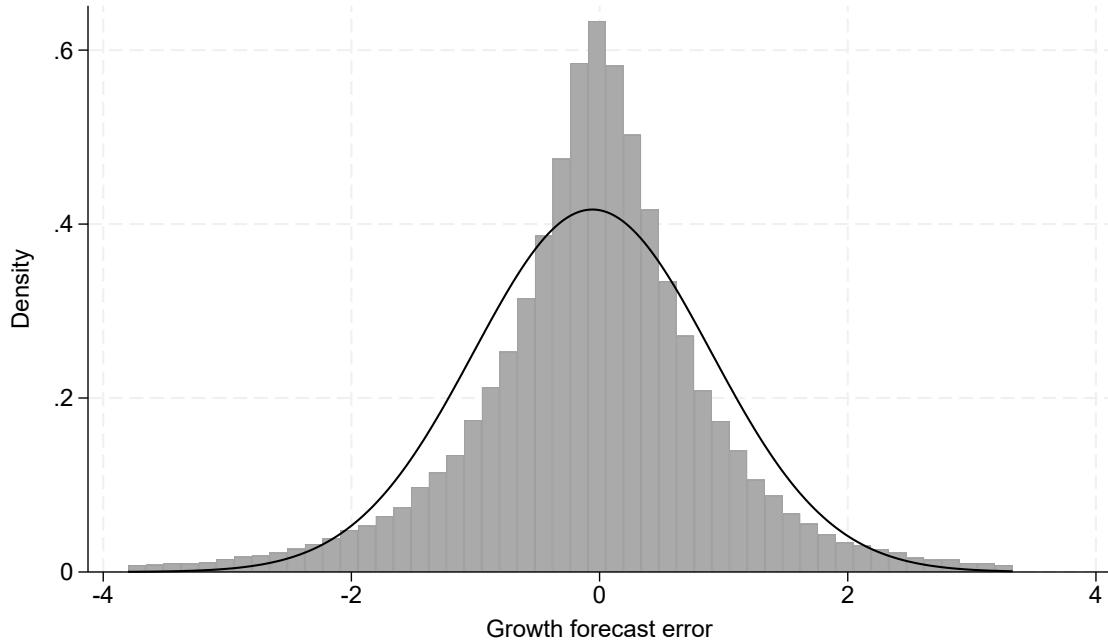
Kwon, Spencer, and Johnny Tang, 2025, Extreme categories and overreaction to news, *Review of Economic Studies, forthcoming* .

- 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, 2020, A Quantitative Analysis of Distortions in Managerial Forecasts, Technical Report w26830, National Bureau of Economic Research, Cambridge, MA.
- Moran, José, Angelo Secchi, and Jean-Philippe Bouchaud, 2024, Revisiting granular models of firm growth.
- Nagel, Stefan, and Zhengyang Xu, 2019, Asset Pricing With Fading Memory, *Review of Financial Studies* Forthcoming.
- Novy-Marx, Robert, and Mihail Velikov, 2016, A taxonomy of anomalies and their trading costs, *Review of Financial Studies* 29, 104–147.
- O'Hagan, A., 1979, On outlier rejection phenomena in bayesian inference, *Journal of the Royal Statistical Society. Series B (Methodological)* .
- Robbins, Herbert, 1956, An empirical bayes approach to statistics, *Berkeley Symposium on Mathematical Statistics and Probablility* 157—163.
- 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.
- Wyart, Matthieu, and Jean-Philippe Bouchaud, 2003, Statistical models for company growth, *Physica A: Statistical Mechanics and its Applications* 326, 241–255.

# ONLINE APPENDIX

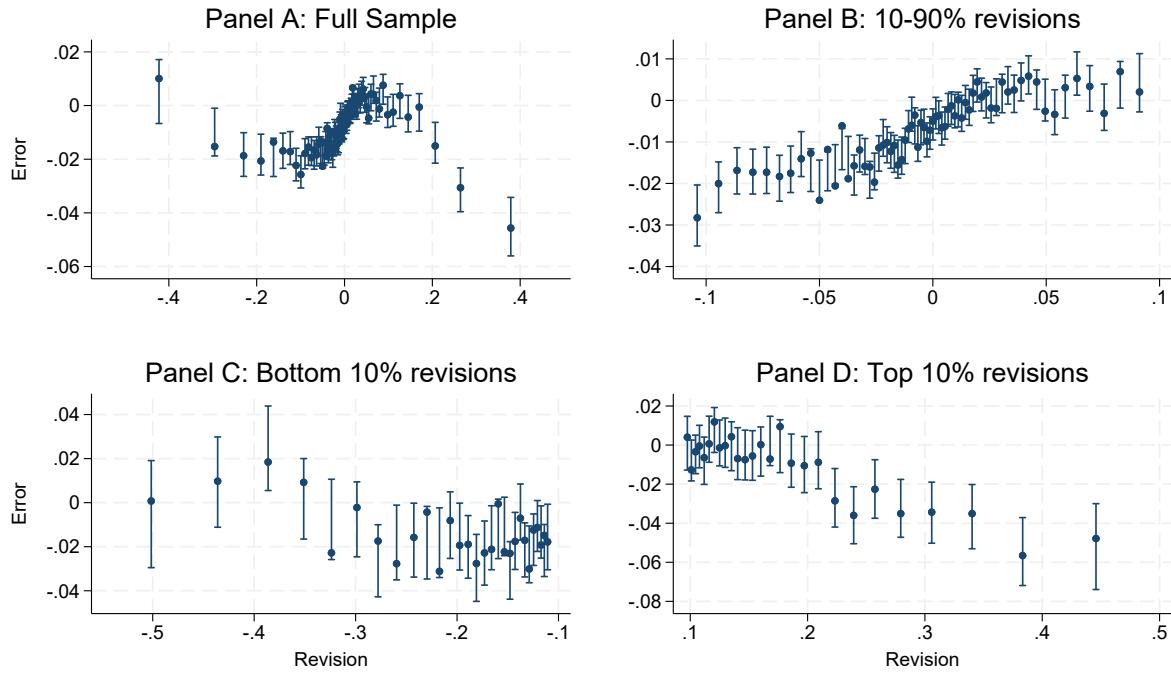
## A Additional Figures

Figure A.1: Forecast error of log growth (normalized)



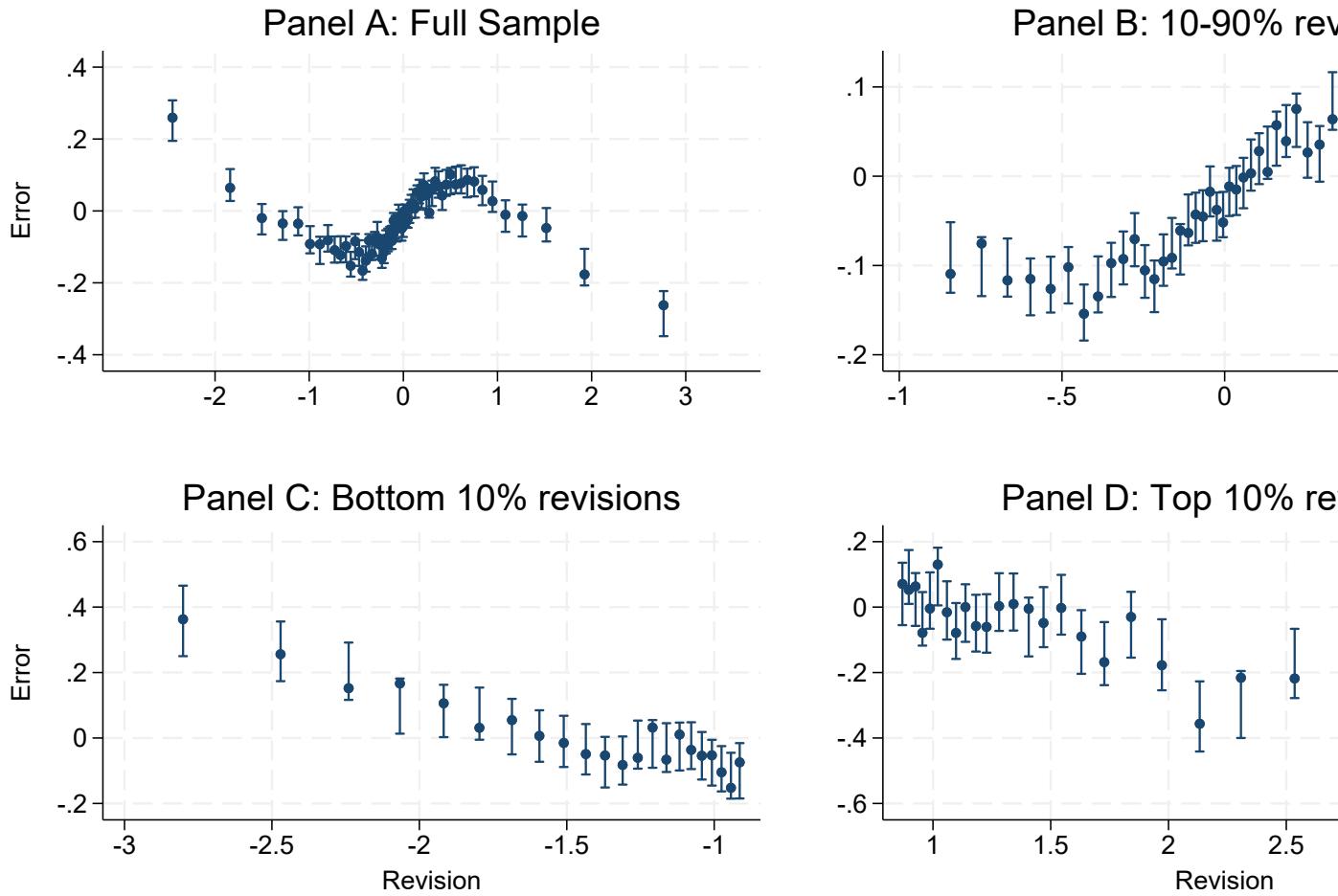
Note: We show here the distribution of normalized log growth  $g_{it}$ . The line corresponds to a Gaussian fit.

Figure A.2: The Error-revision relationship: Raw log growth rate



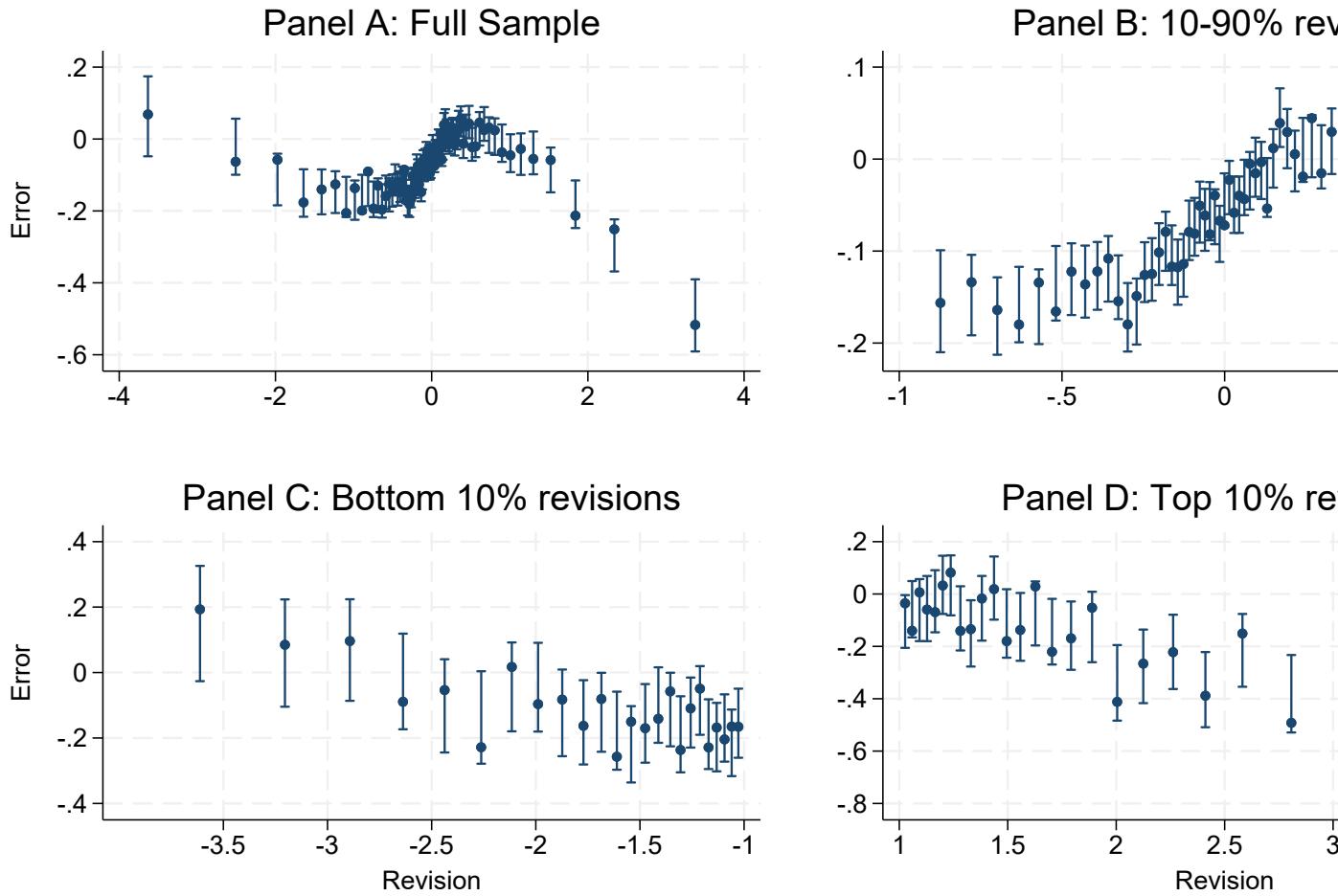
Note: We show here the binned scatter plot of the error-revision relationship for *raw* log sales growth. See definitions of raw and normalized growth in Section 1. Panel A is for the entire sample, panel B restricts the sample to deciles 2 to 9. Panel C focuses on the bottom decile and panel D on the top decile of revisions. Vertical bars represent 95% confidence intervals, assuming the relationship is piecewise linear and continuous (option “ci(1 1)” in stata command “binsreg”).

Figure A.3: The Error-revision relationship: Adjusted percent growth rate



Note: We show here the binned scatter plot of the error-revision relationship for adjusted *percent* sales growth. See definitions of raw and normalized growth in Section 1. Panel A is for the entire sample, panel B restricts the sample to deciles 2 to 9. Panel C focuses on the bottom decile and panel D on the top decile of revisions. Vertical bars represent 95% confidence intervals, assuming the relationship is piecewise linear and continuous (option “ci(1 1)” in stata command “binsreg”).

Figure A.4: The Error-revision relationship: Macro-adjusted log growth rate



Note: We use here an adjustment of errors and revisions at the *macro level*. Each year, we compute the mean absolute deviation of raw log growth, as a measured of cross-sectional volatility. Then, we take the error and revision of raw log growth and normalize them by this MAD. We show here the binned scatter plot of the error-revision relationship after such adjustment. Panel A is for the entire sample, panel B restricts the sample to deciles 2 to 9. Panel C focuses on the bottom decile and panel D on the top decile of revisions. Vertical bars represent 95% confidence intervals, assuming the relationship is piecewise linear and continuous (option “ci(1 1)” in stata command “binsreg”).

## B Proofs

### B.1 Proof of Proposition 1

*Proof.* Let us start with the first bullet point. Directly from the definition of the process, we obtain:

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

and the result obtains from a straightforward application of Tweedie's formula (Robbins, 1956; Efron, 2012). This formula states that, for  $y = x + \eta$ , where  $x$  is gaussian and  $x$  and  $\epsilon$  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  $x$  (the gaussian part of  $y$ ).  $\square$

*Proof.* The second bullet point relies on the property that the distribution of  $\epsilon$ ,  $f$ , is asymptotically power law. First, given the DGP, we know that:

$$E(g_{t+1}|g_t) = \rho E(g_t^*|g_t^* + \epsilon_t) \equiv \rho E(g^*| \underbrace{g^* + \epsilon}_{\equiv g})$$

where drop the  $t$  subscript for convenience.

Call  $\Phi$  the p.d.f. of  $g^*$ , then the conditional expectation writes as:

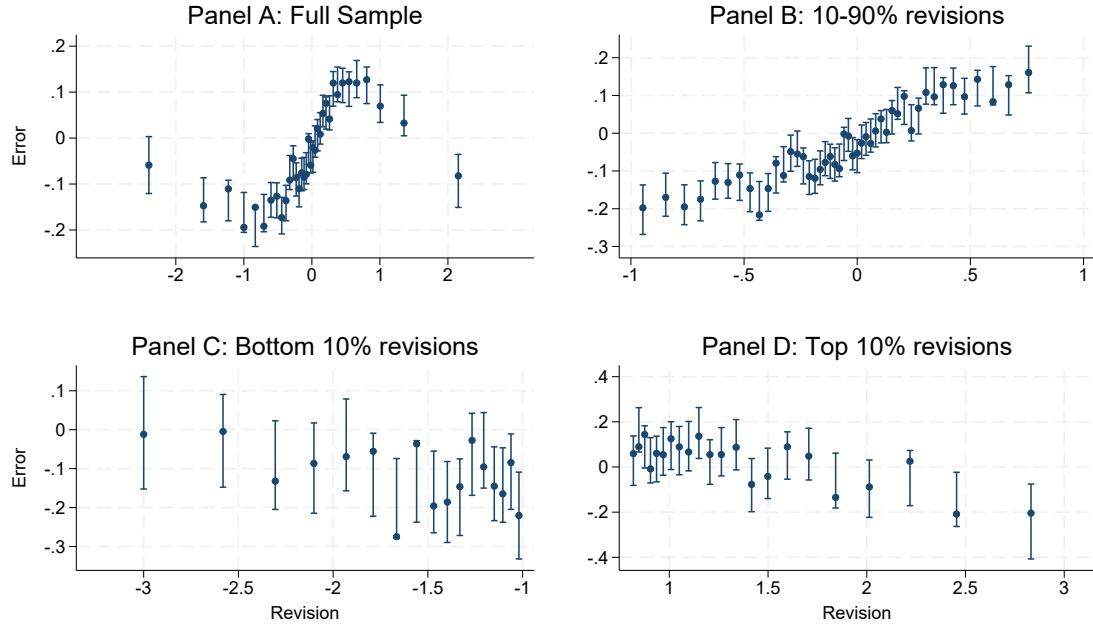
$$E(g^*|g) = \frac{\int x f(g-x) \Phi(x) dx}{\int f(g-x) \Phi(x) dx}$$

Now, given the Pareto assumption, and for given  $x$ , as  $g \rightarrow +\infty$

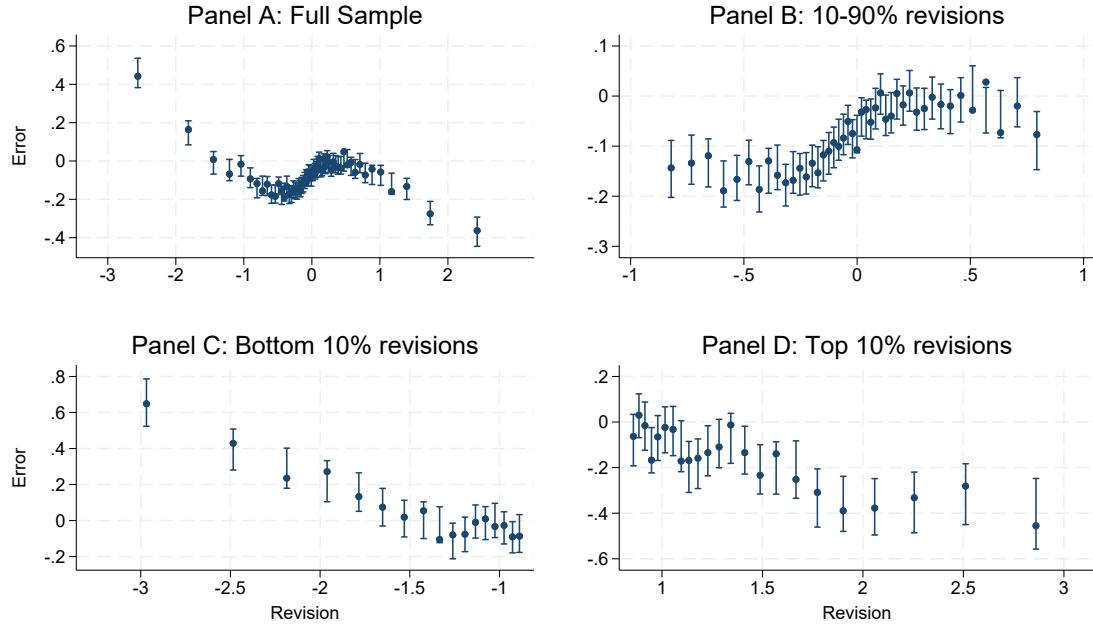
$$f(g-x) \rightarrow (g-x)^{-\nu} \approx g^{-\nu} \left(1 + \nu \frac{x}{g}\right)$$

Figure A.5: The Error-revision relationship: US v non US Firms

Panel 1: US Firms

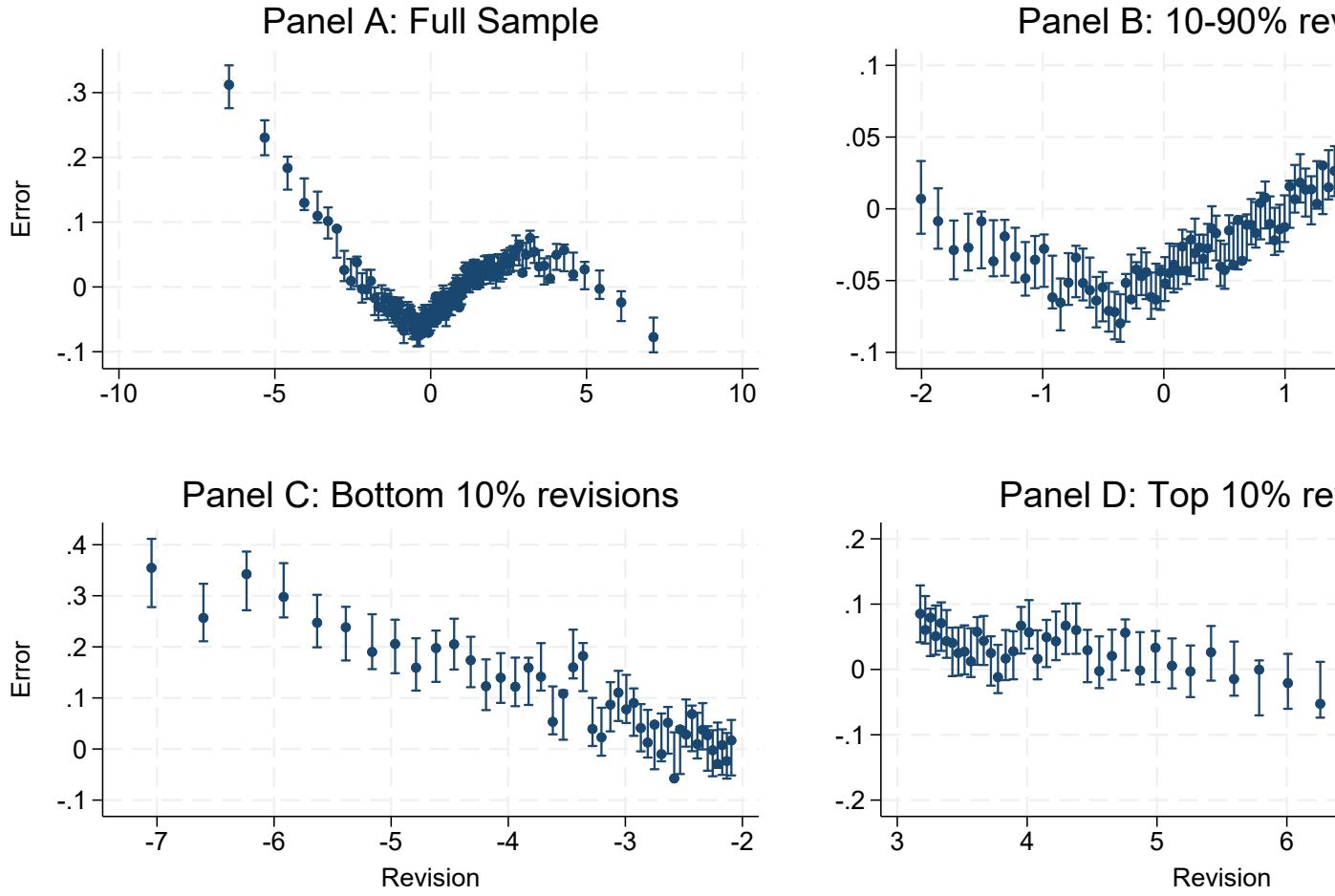


Panel 2: Non US Firms



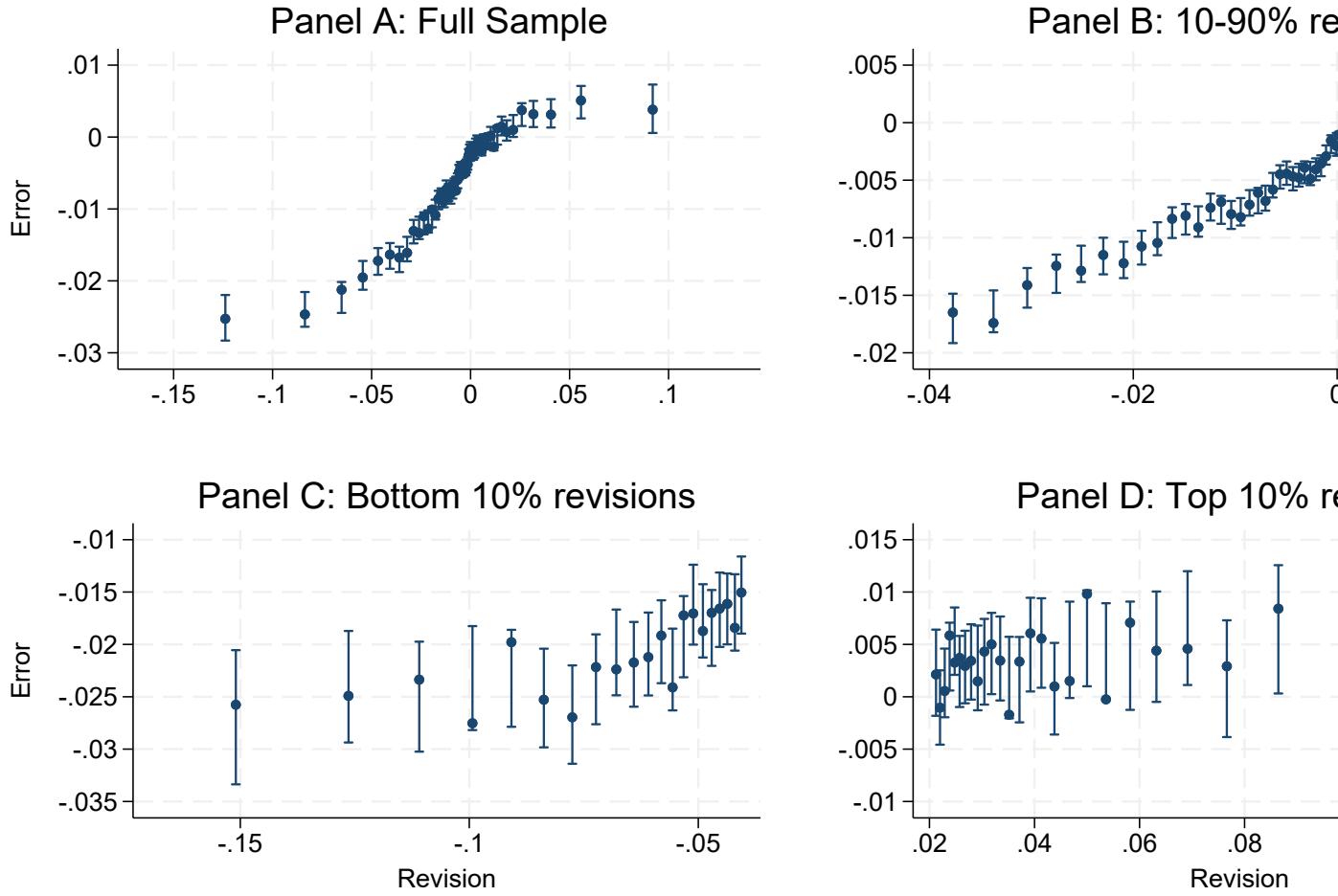
Note: We show here the binned scatter plot of the error-revision relationship for adjusted log sales growth. See definitions of raw and normalized growth in Section 1. Panel 1 is for US firms, and Panel 2 for international firms. Each one of these two panels has 4 subpanels: Panel A is for the entire sample, panel B restricts the sample to deciles 2 to 9. Panel C focuses on the bottom decile and panel D on the top decile of revisions. Vertical bars represent 95% confidence intervals, assuming the relationship is piecewise linear and continuous (option “ci(1)” in stata command “binsreg”).

Figure A.6: The Error-revision relationship: Individual Analyst forecasts



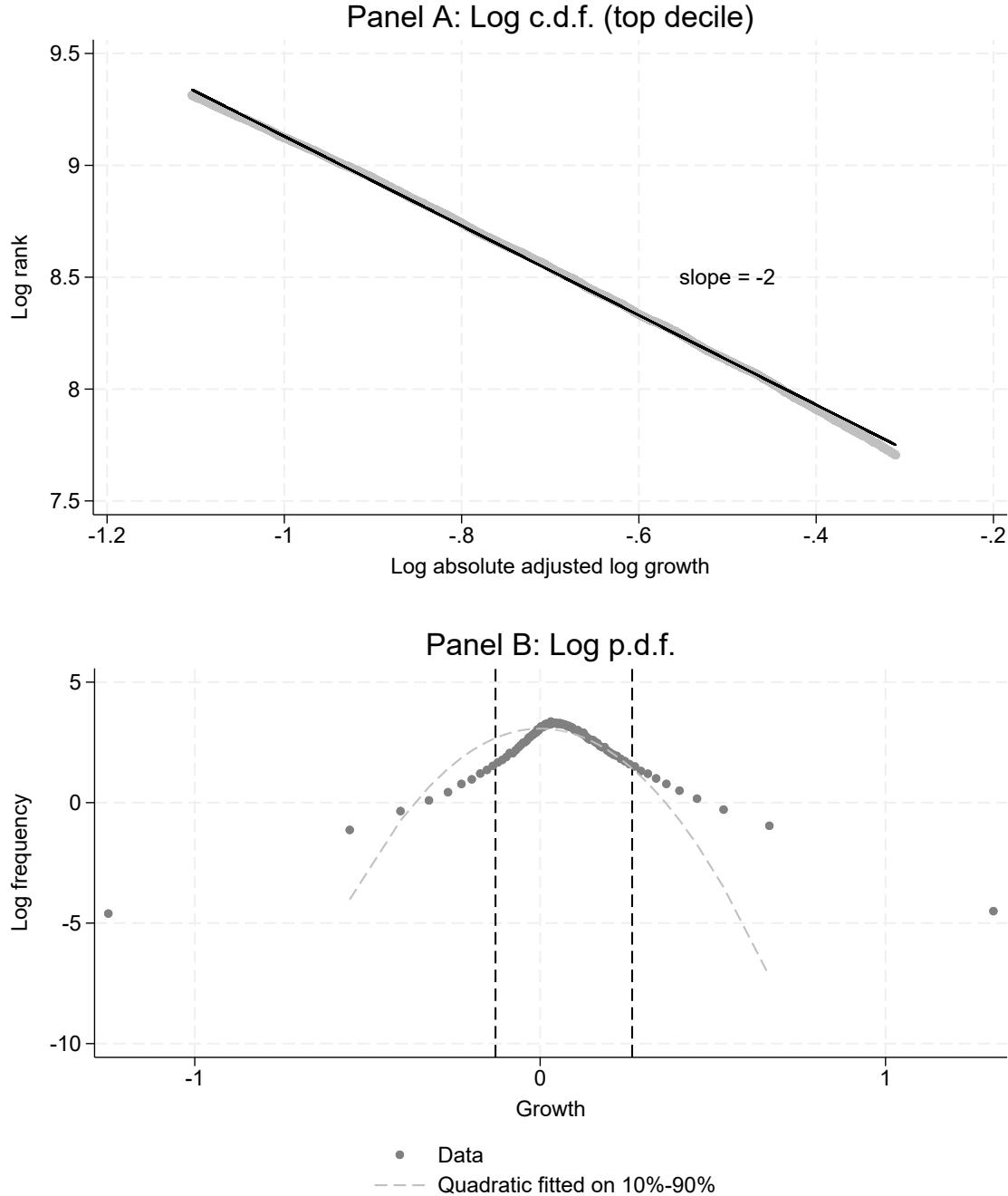
Note: We show the binned scatter plot of the error-revision relationship for adjusted log sales growth. Crucially, we use here *individual* (not consensus) analyst forecasts. The forecast error is defined as  $g_{it+1} - F_{jt}g_{it+1}$  for analyst  $j$ , firm  $i$  at date  $t$ . Similarly, revisions are defined as  $F_{jt}g_{it+1} - F_{jt-1}g_{it+1}$ . See definitions of raw and normalized growth in Section 1. Panel A is for the entire sample, panel B restricts the sample to deciles 2 to 9. Panel C focuses on the bottom decile and panel D on the top decile of revisions. Vertical bars represent 95% confidence intervals, assuming the relationship is piecewise linear and continuous (option “ci(1 1)” in stata command “binsreg”).

Figure A.7: The Error-revision relationship: Normalized EPS to Price



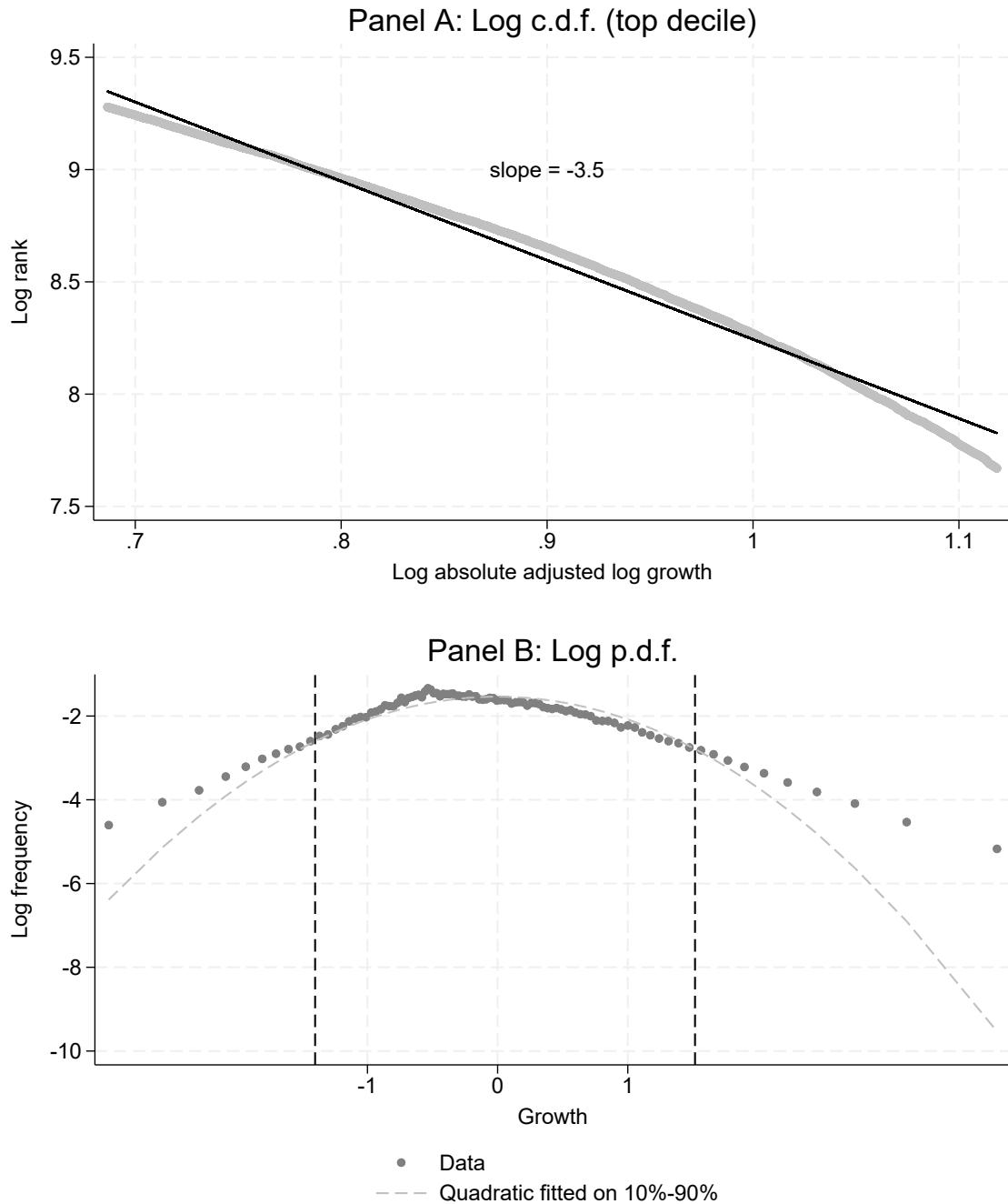
Note: We show the binned scatter plot of the error-revision relationship for raw *EPS to Price ratios*. The forecast error is defined as  $\frac{EPS_{it+1} - F_{jt}EPS_{it+1}}{P_{it-2}}$  for analyst  $j$ , firm  $i$  at date  $t$ . Similarly, revisions are defined as  $\frac{F_{jt}EPS_{it+1} - F_{jt-1}EPS_{it+1}}{P_{it-2}}$ . Panel A is for the entire sample, panel B restricts the sample to deciles 2 to 9. Panel C focuses on the bottom decile and panel D on the top decile of revisions. Vertical bars represent 95% confidence intervals, assuming the relationship is piecewise linear and continuous (option “ci(1 1)” in stata command “binsreg”).

Figure A.8: Fat tails of the sales growth distribution: Raw log growth



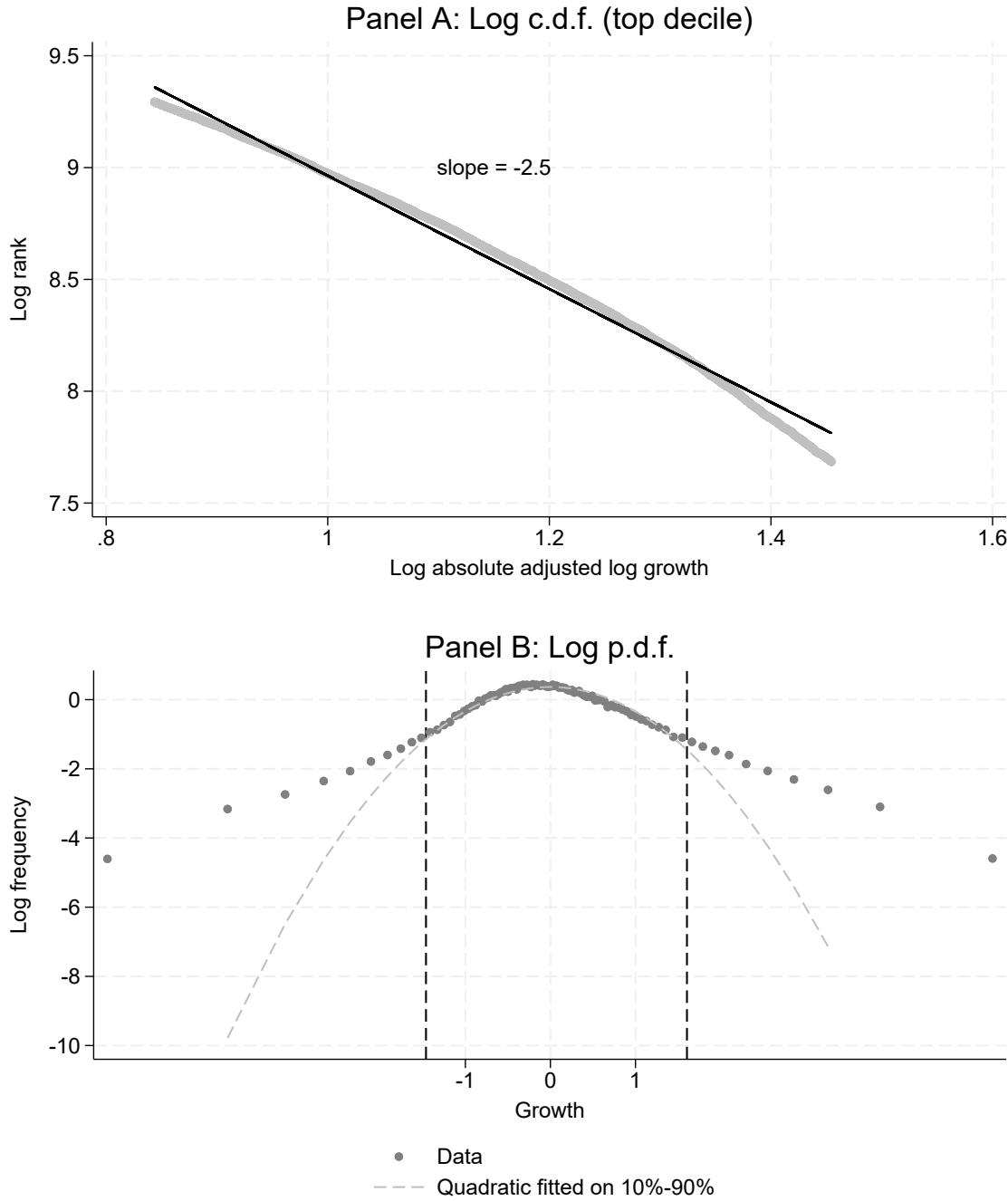
Note: We describe here the tail of the distribution of *raw* log growth rates  $G_{it}$ . In Panel A, we show the scatter plot of log rank of  $|G_{it}|$  against  $\log |G_{it}|$ . We restrict ourselves to the top decile of absolute growth and remove the top percent. We also report the slope of the regression of log rank on log growth, estimated by OLS (-2.7). In Panel B, we show the log density of  $G_{it}$ . For each centile, we estimate density as the log of the number of observations in the centile divided by its range. The dashed line is a quadratic fit on the centiles between the 11th and 89th centiles. The two dashed vertical lines correspond to the cutoff values of the top and bottom decile of the distribution of log log growth.

Figure A.9: Fat tails of the sales growth distribution: Adjusted percent growth



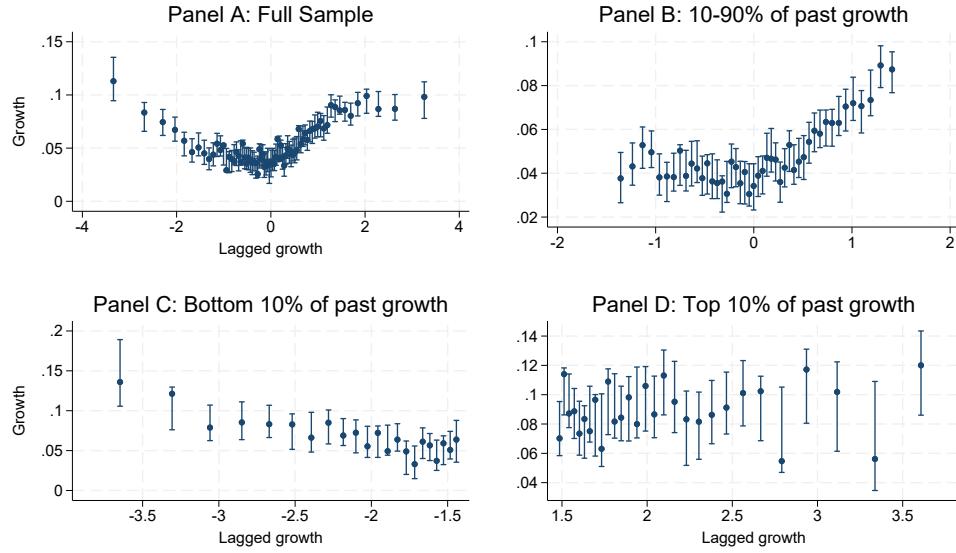
Note: We describe here the tail of the distribution of adjusted *percent* growth rates  $\hat{g}_{it}$ . In Panel A, we show the scatter plot of log rank of  $|\hat{g}_{it}|$  against  $\log |\hat{g}_{it}|$ . We restrict ourselves to the top decile of absolute growth and remove the top percent. We also report the slope of the regression of log rank on log growth, estimated by OLS (-2.7). In Panel B, we show the log density of  $\hat{g}_{it}$ . For each centile, we estimate density as the log of the number of observations in the centile divided by its range. The dashed line is a quadratic fit on the centiles between the 11th and 89th centiles. The two dashed vertical lines correspond to the cutoff values of the top and bottom decile of the distribution of log log growth.

Figure A.10: Fat tails of the sales growth distribution: Macro-adjusted log growth rate



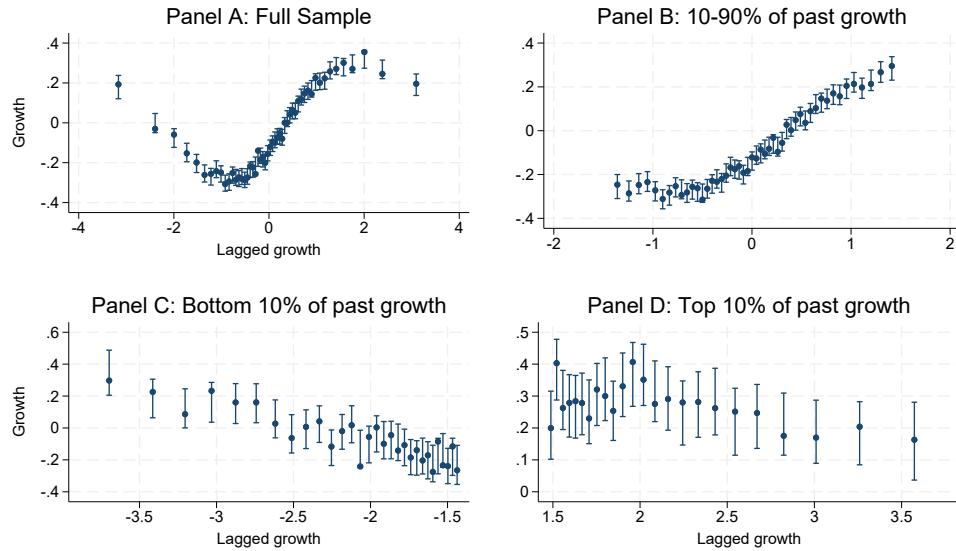
Note: We describe here the tail of the distribution of macro adjusted log growth rates  $\tilde{g}_{it}$ . Each year, we compute the deviation raw log growth rates from their mean, and divide this by their cross-sectional mean absolute distance from that mean. In Panel A, we show the scatter plot of log rank of  $|\tilde{g}_{it}|$  against  $\log |\tilde{g}_{it}|$ . We restrict ourselves to the top decile of absolute growth and remove the top percent. We also report the slope of the regression of log rank on log growth, estimated by OLS (-2.7). In Panel B, we show the log density of  $\tilde{g}_{it}$ . For each centile, we estimate density as the log of the number of observations in the centile divided by its range. The dashed line is a quadratic fit on the centiles between the 11th and 89th centiles. The two dashed vertical lines correspond to the cutoff values of the top and bottom decile of the distribution of log log growth.

Figure A.11:  $g_{it+1}$  as a function of  $g_{it}$ : unadjusted log growth



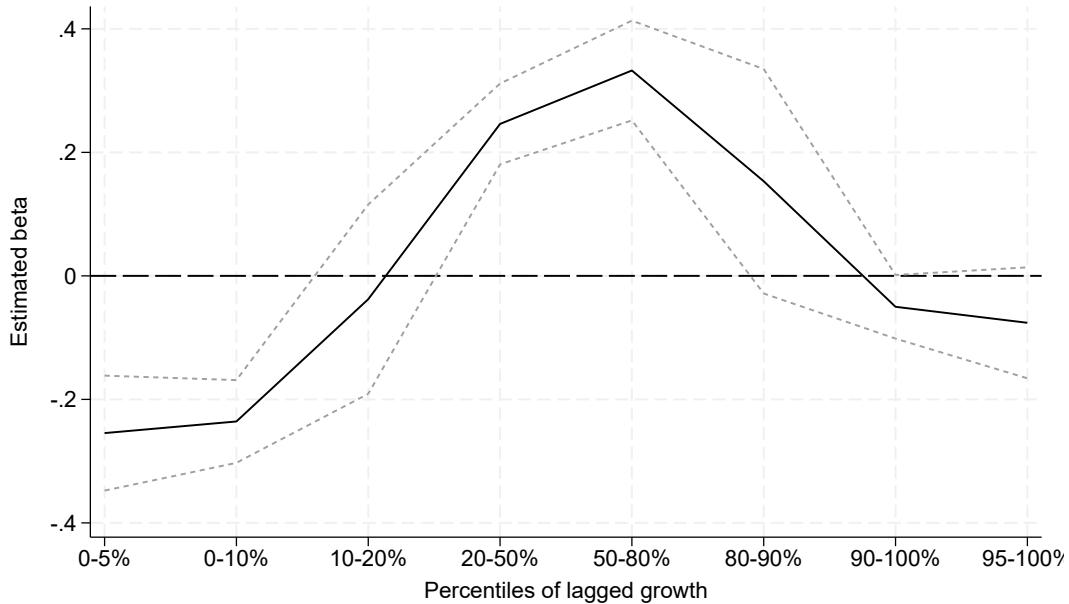
Note: This figure reports 4 binned scatter plots of unadjusted log growth  $G_{it+1}$  against  $G_{it}$ . Panel A shows the entire sample. Panel B restricts to all observations of  $g_{it}$  between the first and the last decile breakpoints. Panel C focuses on the bottom decile and panel D on the top decile. Vertical bars represent 95% confidence intervals, assuming the relationship is piecewise linear and continuous (option “ci(1 1)” in stata command “binsreg”).

Figure A.12:  $g_{it+1}$  as a function of  $g_{it}$ : adjusted percent growth



Note: This figure reports 4 binned scatter plots of adjusted percent growth  $\hat{g}_{it+1}$  against  $\hat{g}_{it}$ . Panel A shows the entire sample. Panel B restricts to all observations of  $g_{it}$  between the first and the last decile breakpoints. Panel C focuses on the bottom decile and panel D on the top decile. Vertical bars represent 95% confidence intervals, assuming the relationship is piecewise linear and continuous (option “ci(1 1)” in stata command “binsreg”).

Figure A.13: Regression coefficients of  $g_{it+1}$  on  $g_{it}$  by quantile

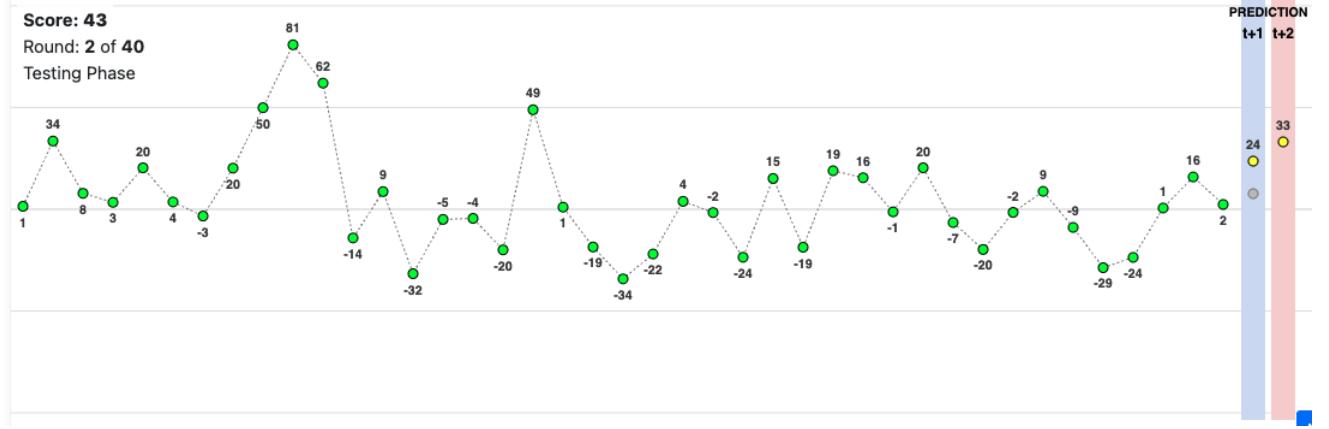


Note: In this Figure, we report the estimates of  $\beta$  in the following regression:

$$g_{it+1} = \alpha + \beta g_{it} + \epsilon_{it+1}$$

where  $g_{it}$  is the normalized log sales growth rate, and error terms are assumed to be correlated within firm and within years. This regression is run on 8 different subsamples, whose ranges are described in the x-axis of this chart. These subsamples corresponds to the tails and the bulk of the distribution of  $g_{it}$ . The point estimate of  $\beta$  is the solid black line, while the dashed lines corresponds to the 95% confidence interval.

Figure A.14: Experiment interface



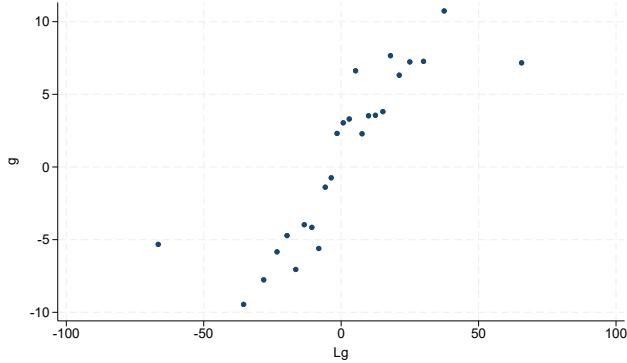
Note: This Figure is a screenshot of our experiment interface. The green dots are generated by the DGP:

$$g_{t+1} = g_{t+1}^* + 0.608\epsilon_{t+1}$$

$$g_{t+1}^* = 0.529g_t^* + 0.631u_{t+1}$$

which is obtained by matching several moments of the data (see “Structural estimation” section).  $\epsilon_t$  is drawn from a student distribution with 2.533 degrees of freedom. The two yellow dots corresponds to forecasts produced by the participant. The operation is repeated 40 times.

Figure A.15: Conditional expectation  $E(g_{it+1}|g_{it})$  in the experiment



Note: This Figure is a binned scatter plot of  $g_{it+1}$  against  $g_{it}$ , where  $t = 1, \dots, 40$  is a round and  $i$  a participant. The data is coming from our experimental data (8,040 observations). The DGP in our experiment is given by:

$$g_{t+1} = g_{t+1}^* + 0.608\epsilon_{t+1}$$

$$g_{t+1}^* = 0.529g_t^* + 0.631u_{t+1}$$

$\epsilon_t$  is drawn from a student distribution with 2.533 degrees of freedom.

Then, the following Lemma holds:

**Lemma B.1.** *As  $g \rightarrow +\infty$ , then:*

$$E(g^*|g) \rightarrow \frac{\int x \left(1 + \nu \frac{x}{g}\right) \Phi(x) dx}{\int \left(1 + \nu \frac{x}{g}\right) \Phi(x) dx}$$

The proof of this Lemma comes from a Tauberian logic. The conditional expectation is a weighted average. As  $g$  becomes very large, the weights converge to a function. So the weighted average converges to an average using the limiting function.

Now, it is easy to see that:

$$\int x \left(1 + \nu \frac{x}{g}\right) \Phi(x) dx = Eg + \nu \frac{Eg^2}{g} = \nu \frac{\sigma_g^2}{g}$$

and:

$$\int \left(1 + \nu \frac{x}{g}\right) \Phi(x) dx = 1$$

which leads to the result.  $\square$

*Proof.* The third bullet point is a straight application of the first one, combined with the slope of subjective expectations (6). The forecaster overreacts the her sensitivity to  $g_t$  is larger than the sensitivity of short-term forecasters:

$$-\rho \sigma_{g^*}^2 \left( \frac{d^2}{dg^2} \log h(g) \right) (g_t) > \rho \frac{\sigma_{g^*}^2}{\sigma_g^2}$$

This gives the third result.  $\square$

*Proof.* To prove the fourth bullet point see that:

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

Differentiate the log density two times:

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

As we have just seen,  $\frac{h'(g)}{h(g)} = -\frac{E(g_t^*|g)}{\sigma_{g^*}^2}$ . Simple algebra shows that:

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

so combining the two gives:

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

which then allows us to inject  $\text{var}(g_t^*|g_t)$  into the formula for  $\text{var}(g_{t+1}|g_t)$ .  $\square$

## B.2 Proof of Proposition 2

*Proof.* We start with the first item. First, note that, by definition of subjective expectations:

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

Thus, the conditional expectation writes:

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

We thus need to focus on the second term on the right hand side. Given the definition of forecasts in equation (5), simple algebra leads to the following decomposition for revisions:

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

where, by definition,  $\epsilon_t$ ,  $E_t$  and  $U_t$  are independent.

Clearly,  $U_t$  is gaussian (it is a linear combination of gaussians).  $E_t$  is asymptotically Pareto. Assume  $E \rightarrow +\infty$ , since  $\epsilon$  is asymptotically Pareto, we have that:

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

where the second line is a property of regularly varying functions (of which Pareto is a subcategory). See for instance [Jessen and Mikosch \(2006\)](#), Lemma 3.1. This proves that  $E_t$  is also asymptotically Pareto with tail  $\nu$  but a different scale given by the formula.

Now, we need to compute, for large revisions:

$$\begin{aligned} E(\epsilon_t | REV_t) &= E(\epsilon_t | K\rho\sigma_\epsilon(\epsilon_t + E_t + U_t)) \\ &= E(\epsilon_t | \epsilon_t + E_t + U_t) \\ &\approx E(\epsilon_t | \epsilon_t + E_t) \end{aligned}$$

since large revisions are asymptotically driven by the fat-tailed processes only.

To compute the expectations of one pareto variable conditional on the sum of two, one needs to know the relative scale of these two variables. The above algebra shows that:

$$f(\epsilon) \approx \frac{1 - \rho^\nu(1 - K)^\nu}{K^\nu \rho^\nu} g(E)$$

where we label  $g$  the p.d.f. of  $E_t$ . Property 4.9 of Denuit et al. (2024) establishes that, in this case:

$$\begin{aligned} E(\epsilon_t | \epsilon_t + E_t) &\approx \frac{\frac{1 - \rho^\nu(1 - K)^\nu}{K^\nu \rho^\nu}}{1 + \frac{1 - \rho^\nu(1 - K)^\nu}{K^\nu \rho^\nu}} \cdot (\epsilon_t + E_t) \\ &\approx \frac{1}{1 + \frac{K^\nu \rho^\nu}{1 - \rho^\nu(1 - K)^\nu}} \cdot \frac{REV_t}{K \rho} \end{aligned}$$

We plug this result into equation (B.1) and obtain that:

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

Now, to obtain the formula in the first point, we need to use the expression of the Kalman gain  $K$ :

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

whose combination shows that:

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

This completes the proof of the first point. □

*Proof.* We first need to find the sign of  $\mu$ . Assume  $\nu > 2$ . From its expression, and given that  $\rho$ ,  $K$  and  $1 - K$  are all positive but smaller than 1, we have that:

$$\frac{K^2 \rho^2}{1 - (1 - K)^2 \rho^2} > \frac{K^\nu \rho^\nu}{1 - (1 - K)^\nu \rho^\nu}$$

as long as  $\nu > 2$ . Combined with the fact that  $\sigma_u^2 / \sigma_\epsilon^2 > 0$ , this ensures that  $\mu < 0$ .

The second result is straightforward. By definition of the Kalman gain, we have that:

$$E(ERR_{t+1} REV_t) = 0$$

Now, for large enough revisions,  $E(ERR_{t+1}|REV_t) \approx \mu REV_t$ , so that  $E(ERR_{t+1} REV_t | REV_t) \approx \mu REV_t^2 < 0$ . This gives the result in the proposition.  $\square$

### B.3 Additional Details on Asset Pricing Model in Section 5.2

We first derive (21). Define the return on a stock as  $R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t}$  that has sales  $S_t$ , earnings  $\mathcal{E}_t$ , and a payout ratio  $DE_t = \frac{D_t}{\mathcal{E}_t}$ . Assume that earnings are a constant fraction of sales:  $\mathcal{E}_t = \gamma S_t$ . As in prior sections, denote the (log) growth rate of sales as  $g_t = \log S_t - \log S_{t-1}$ . Denoting lower-case letters as logs and following [Campbell and Shiller \(1988\)](#), we can approximate the price-earnings ratio using to first order around the mean price-dividend ratio as:

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

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

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

Imposing the usual transversality condition, the previous equation becomes:

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

Now, rearranging the first equation above, we have

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

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

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

Rewriting, we obtain:

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

To derive predictions based on our model of subjective expectations, we follow [Bouchaud et al. \(2019\)](#) and [Nagel and Xu \(2019\)](#) 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 (21):

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

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:

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

## C Computing Rational Expectations Using the Particle Filter

Constructing the forecasts in (16) requires computing the rational expectation,  $E(g_{t+1} | g_0, \dots, g_t)$ . In the case where  $\epsilon_t$  is normally-distributed, this corresponds to the Kalman filter, and takes the simple closed form in (13). 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  $p(g_t^* | g_1, \dots, g_t)$  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 for  $u_t$ . The new particles are then evaluated against the observed data using a likelihood function  $f(\cdot)$ , the density function for 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  $f(\cdot)$  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  $f(\cdot)$ . Intuitively, this weighting step incorporates the non-normality of  $f(\cdot)$  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 lengths 100,000 with  $P$  set to 10,000.<sup>11</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>11</sup>We choose this particular value of  $P$  because it is the largest value of  $P$  such that all the particles fit into 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 such that the particle filter recovers the true conditional expectations.

---

**Algorithm 1** Particle Filter

---

- 1: **Fix** a number of particles,  $P$ , and 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:

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

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

$$q_t^p = \frac{f(g_t - \xi_t^p)}{\sum_{p=1}^P f(g_t - \xi_t^p)},$$

- 7: **Resample**  $\{\xi_t^p\}_{p=1}^P$  from  $\{\xi_t^p\}_{p=1}^P$  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 2.
  - 12: **end if**
-