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

Tim de Silva† Eugene Larsen-Hallock‡ Adam Rej‡ David Thesmar§

February 11, 2025

**Very preliminary. Please do not circulate without permission.**

**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 propose a framework that connects these facts, in which the underlying process is non-Gaussian and forecasters use a simplified linear model. After showing that this framework can quantitatively explain our facts, we show it provides a qualitative explanation for patterns in asset prices, such as the non-linearity of the returns to momentum and responses to earnings announcements. In ongoing work, we test additional predictions of our theory in an online forecasting experiment where the underlying process has fat tails.

*We are grateful to Xavier Gabaix for feedback and help with asymptotic results. We thank seminar participants at CFM, Columbia and Yale for their constructive feedback.

†Stanford University, Graduate School of Business and SIEPR, tdesilva@stanford.edu.

‡Capital Fund Management

§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 forecasts deviate from the full information rational expectations (FIRE) model that predominates in economic modeling because forecast errors are predictable using information within forecasters' information sets. Most theories that have been proposed to explain this predictability start by assuming that the data-generating process (DGP) of the variable that is being forecasted is simple, such as an AR1 process with normal shocks, and forecasters know this DGP. Despite knowing the DGP, forecasters in these theories 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, in which the underlying process is non-Gaussian, but forecasters use a simplified linear model. The key challenge in this framework is characterizing rational expectations, which we do analytically using results from empirical Bayes theory (Robbins, 1956) and numerically using a particle filter (Fernandez-Villaverde and Rubio-Ramirez, 2007). Finally, after showing that this model can quantitatively explain our facts, we show it provides a qualitative explanation for patterns in asset prices, such as the non-linearity of the returns to momentum and responses to earnings announcements. In ongoing work, we test additional predictions of our theory in an online forecasting experiment similar to Afrouzi et al. (2023), but where the underlying process has fat tails.

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 well-defined. 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, a feature that is not discussed in the existing literature. In some papers, revisions linearly and *positively* predict forecast errors, a feature commonly interpreted as evidence of under-reaction (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., 2018, 2019; Afrouzi et al., 2023). In our panel of forecasts and realizations of sales growth,

which is much larger than typical studies, we find evidence that both co-exist. For intermediate values of revisions, forecasters underreact to news (an increasing relation between revisions and errors). For large values of revisions, forecaster overreact (a decreasing relation between revisions and errors). Given the size of our sample, we can establish that this non-linearity is a robust feature of the data.

The second fact that we document is that the distribution of sales growth has fat tails. This fact has already been documented in the literature (Stanley et al., 1996; Axtell, 2001), and we revisit it in our sample. To detect the presence of fat tails, we examine the log-density plots, as in Guvenen et al. (2021) who study the distribution of income growth. In the bulk of the sales growth distribution, we find that a Gaussian density, which is quadratic in logs, fits well. However, in the top and bottom 10% tails of the distribution, sales growth in the data has much thicker and longer tails than a normal distribution that can be well approximated by a power law (i.e., Pareto) with a tail parameter between 2 and 3, imply slightly thinner tails than Zipf's law (Gabaix, 2009). Importantly, fat tails in the distribution of sales growth do not arise for heterogeneity in volatility, as in Wyart and Bouchaud (2003): when we rescale sales growth by estimates of firm-level standard deviations, the resulting distribution still has very fat tails, consistent with Moran et al. (2024).

Our third and final motivating fact is that the conditional expectation of future sales growth is non-linear in current growth. In particular, we find that the relationship between current and one-year-ahead sales growth is positive 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 we deviate from standard Gaussian models of the DGP in our model.

In the next part of the paper, we develop a forecasting model that is designed to connect the three facts that we documented in the data. In the model, the data-generating process (DGP) for the forecasting variable contains a permanent 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 that the transitory component is non-Gaussian and follows a power law distribution, a tractable way of characterizing processes with fat tails (Gabaix, 2009). For some results, we assume that the permanent component follows an AR1 process with normal innovations. This assumption is less crucial, but what is key is that the transitory component is non-Gaussian. Although the permanent component may also be non-Gaussian, a model with a Gaussian transitory component cannot match our third fact: the strong mean reversion in sales growth.

Given this DGP, the rational expectation of future growth conditional on current growth is non-

linear. In our model, characterizing rational expectations is not possible in closed-form due to the non-normality. However, under the assumption that the permanent component is an AR1 process with normal innovations, we characterize the expectation of future growth conditional on current growth as a function of the log marginal density of growth by applying a result from empirical Bayes theory known as Tweedie's formula (Robbins, 1956; Efron, 2012). The consequence of this result is that the rational expectation is linear when the distribution of sales growth locally resembles a Gaussian, as in the bulk of the distribution. However, when the distribution is non-Gaussian, as in the tails, the rational expectation of future growth conditional on current growth is less linear and potentially even decreasing in current growth. Intuitively, a forecaster who knows the DGP anticipates more mean reversion after extreme realizations because she infers that the large realizations were likely due to the transitory component and, hence, will not persist.

Having developed a model of the DGP that is consistent with our second and third facts, we show that our first fact on the non-linear relationship between forecast errors and revisions can be explained qualitatively using a single assumption: agents make linear forecasts. In particular, we assume that agents form forecasts according to a misspecified linear model for the relationship between current and future growth, where the model parameters are chosen via OLS so that the model has the best possible linear fit. We view the assumption that agents use a simple misspecified model as a natural form of bounded rationality (Gabaix, 2018), as in the natural expectations model of (Fuster et al., 2010). Another possible microfoundation that we discuss below is the model in Augenblick et al. (2024), in which forecasters incorrectly perceive signal quality and shrink it to a default. This is similar to using a linear model in our setting because extreme observations are more likely to come from the transitory shock. Therefore, these signals are less informative and should receive less weight, but our model assigns them equal weight. However, we do not take a specific stand on a micro-foundation for a linear forecasting rule and whether expectations are Bayesian but rather view it as a reduced form.

We show that the combination of a non-Gaussian DGP and linear forecasting is enough to generate overreaction to news in the tails of past growth and underreaction in the bulk. We show this formally, but the mechanism is intuitive. Large revisions are driven by large shocks to current growth, which are likely to come from the transitory shock. As a result, a rational forecaster would anticipate these shocks to mean-revert, while a linear forecaster does not and, therefore, overreacts to extreme (positive or negative) news.

Based on this prediction, we next quantitatively test our model by seeing whether it can jointly match our three facts. When taking the model to the data, we relax our assumption of a perfectly linear forecasting rule in two ways. First, allow forecasters to shrink their forecasts towards the

true rational expectation à la Gabaix (2018), which we compute using a particle filter (Fernandez-Villaverde and Rubio-Ramirez, 2007), and we estimate the extent of this shrinkage by matching our first fact. Second, we allow forecasters to observe private signals about future growth that are unobserved to the econometrician, as in de Silva and Thesmar (2024). We estimate the model using the simulated method of moments and find it can indeed generate all three of our facts quantitatively.

We conclude by documenting several additional results that qualitatively support our theory. First, we show that after a large shock to current growth, future growth mean-reverts faster, which is the key ingredient of the DGP that underlies expectation errors in our model. Second, we show that our three key facts are also true for earnings-to-price forecasts and realizations. Third, we examine medium-run momentum. Our model predicts that the relationship between past and future returns should be positive in the bulk of the distribution, where undereaction to news is dominant. In contrast, it predicts mean-reversion 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 (below median CRSP market cap). For these stocks, a momentum strategy that omits extreme winners and losers has better risk-adjusted returns. Fourth, we provide evidence that post-earnings announcement drift is more prominent for announcements with less extreme news, while the reverse is true for those with more extreme news. Of course, these latter two strategies may be too difficult to trade, but we view these findings as corroborating evidence for our theory of expectation formation.

Finally, in *ongoing* work, we test additional predictions of our theory in an online forecasting experiment similar to Afrouzi et al. (2023), but where the underlying process has fat tails. The key benefit of this survey experiment is that it allows us to directly test our theory of expectation formation by experimentally-varying the features of the data-generating process.

**Related literature.** This paper contributes to the recent empirical literature on expectations formation. Most papers in this space focus on linear and Gaussian data-generating processes. Forecasting rules may, or may not, be optimal, but are in general linear, so that the relationship between forecast errors and past revisions (or past errors) is also linear. Our paper emphasizes the non-linearity of such a relation, and as a result, suggests an different modelling approach for the data-generating process to account for this non-linearity. We emphasize non-Gaussian dynamics, similar to Kozlowski et al. (2020) and Farmer et al. (2021). However, unlike these papers, forecasters in our model use simplified forecasting rules rather than engaging in Bayesian learning.

In doing this, we connect the expectations formation literature with the empirical literature on firm dynamics, which emphasizes the omnipresence of power laws in the distribution of firms sizes (see Gabaix, 2009, for a survey of power laws in economics). In our paper, the key ingredient is

that sales *growth* (rather than level) has fat tails, a fact first documented by Stanley et al. (1996) and Bottazzi and Secchi (2006). We contribute to this descriptive literature by showing that the heavy tails of the growth distribution are robust to adjusting for firm-level volatility. This confirms that the distribution of growths cannot be explained by a mixture of Gaussian distribution with heterogeneous variances (Moran et al. (2024)). Also related is the literature on income dynamics, which emphasizes deviations from the canonical permanent-transitory income processes with Gaussian shocks (Guvenen et al., 2014, 2021). However, unlike these models, which emphasize the importance of non-normal permanent shocks, the key ingredient in our model is non-normal *transitory* shocks.[1]

Besides, our assumption that forecasters use a simple, linear, forecasting rule that is misspecified is inspired by the literature on bounded rationality, which assumes economic agents have a propensity to use oversimplified models to minimize computation costs (Fuster et al., 2010; Gabaix, 2018). Such models are correct on average (they are not *unconditionally biased*), since they are fitted on available data, but their misspecification gives rise to predictability in forecast errors (they are *conditionally biased*).

The most closely related papers are Augenblick et al. (2024), Graeber et al. (2024) and Kwon and Tang (2024). Augenblick et al. (2024) provide evidence that people tend to use an "average" reaction rule for both weak and strong signal. This leads to overreaction to weak signals and underreaction to strong ones. As we clarify in the paper, our mechanism could be interpreted this way: Past growth is a much stronger predictor of future growth in the bulk of the distribution than it is in the tails. Thus, our linear forecaster will be susceptible to the bias they document. Kwon and Tang (2024) focus on stock returns that news can arise from different potential DGPs, some with fat tails, others with thin tails. Large shocks are over-attributed to fat-tailed DGP (via the representativeness heuristic), generating overreaction (and thus mean-reversion in returns), while small shocks are over-attributed to thin tailed DGPs, generating under-reaction (and thus momentum). While they do not provide evidence on non-linearity of the forecast-revision relationship as we do, their framework predicts such a relation – even though it rests from very different foundations (we assume one single DGPs while they assume several ones). An important difference between their DGP and ours is that we make dynamic predictions of the sensitivity of future growth on past growth, and we find strong support for these dynamics in our data. Finally, Graeber et al. (2024) analyze a non-linear relationship between (earnings announcement) errors and stock returns – or, equivalently, expectations revisions. They show that revisions respond very strongly to the direction of the error but much less to its intensity. They interpret this fact as evidence that people's perception of intensity

[1] 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, such as job loss.

– and thus optimally underreact to this component of the news. What happens in our set-up is very different, as we find overreaction to large shocks.

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 the main fact: future errors are a S-shaped function of past revision. Section 3 lays out the simple framework that we build in order to explain this novel pattern. Section 4 tests additional predictions from this model and performs robustness checks.

# 1 Data

Analyst forecasts come from IBES Adjusted Summary Statistics files, which are available both in the U.S. and internationally. Summary statistics files contains "current" estimates as of the third Wednesday of each month. While EPS (earnings per share) forecasts have received greater attention in the literature, sales forecasts have the advantage of never being negative, so that we can easily calculate sales growth, and do not run into normalization issues. Indeed, a downside of EPS is that it is frequently negative or small rendering the calculation of EPS growth forecast impractical. Thus, the literature on EPS forecasts studies a variable that is, in essence, non-stationary (a problem typically solved by normalizing forecasts and future realizations by a common price level, as for instance in Bouchaud et al. (2019)). Our focus on log sales growth does not have these issues.

## 1.1 Definition of realizations

We start with the definition of sales growth, which is our realization in this paper. We denote sales of firm $i$ at date $t$ as $R_{it}$. These data from IBES actual files, which ensures comparability with analyst forecasts (more 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. 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.

For each firm $i$, $T_i$ is the number of years for which we have a growth observation. Denote $\mu_i =$

$\frac{1}{T_i}\sum_t G_{it}$ the empirical average of growth observations for firm $i$. Further, $\sigma_i = \frac{1}{T_i}\sum_t |G_{it} - \mu_i|$ is 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. This helps us define *normalized* growth as:

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

where we impose that $T_i \leq 10$. Otherwise, we define adjusted growth as missing. By making growth processes more comparable accross firms, it also makes it easier to fit a single data-generating process on the whole cross-section of firms.

## 1.2 Definitions of forecasts, 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 $R_{it+1}$. $F_t R_{it+1}$ is obtained from IBES summary files as the mean consensus forecast extracted *three months after the end of fiscal year* $t$ (i.e. nine months prior to the end of fiscal year $t+1$) to ensure that sales results for fiscal year $t$ are (mostly) available when the forecast for $t+1$ is formed. Similarly, the two-year ahead forecast $F_{t-1}R_{it+1}$ is measured three months after the end of fiscal year $t-1$.

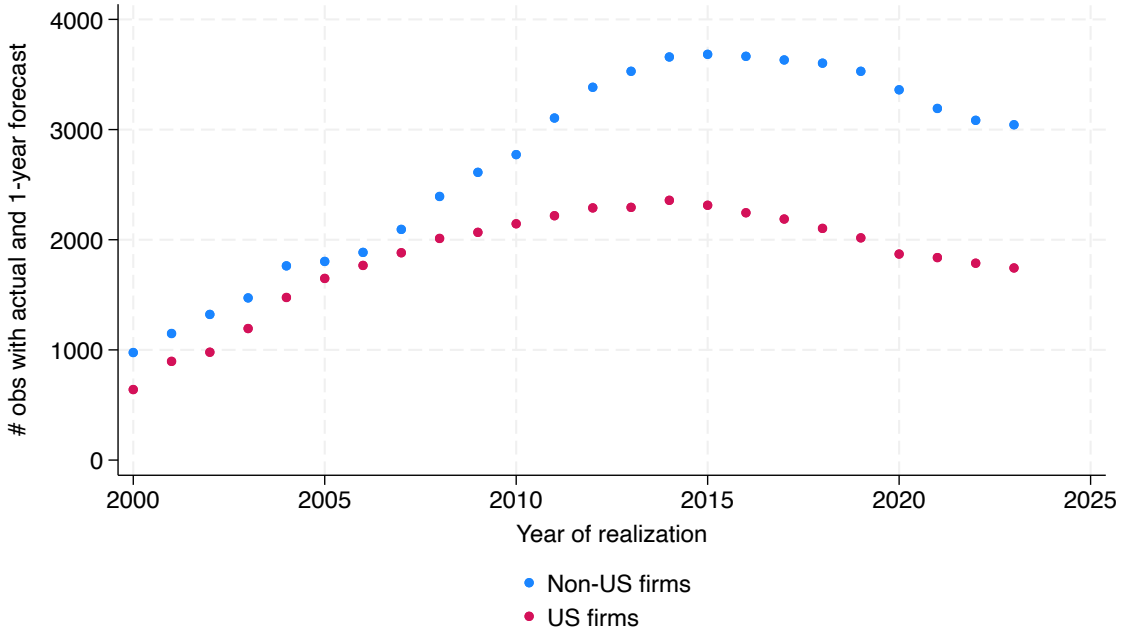
Our key variable of interest in this paper is the forecast of log sales growth, which has the advantage of being stationary. We construct the one-year forecast of *raw* sales growth as $F_t G_{it+1} = \log F_t R_{it+1} - \log R_{it}$. Most of our analysis will focus on "normalized" sales growth forecast, which we construct as $F_t g_{it+1} = \frac{1}{\sigma_i}(\log F_t G_{it+1} - \mu_i)$. Similarly, we also construct two-year ahead forecasts. At the two-year horizon, the raw forecast is $F_t G_{it+2} = \log F_t R_{it+2} - \log F_t R_{it+1}$, while the normalized one is given by: $F_t g_{it+2} = \frac{1}{\sigma_i}(\log F_t G_{it+2} - \mu_i)$.

Finally, in the spirit of the expectations formation literature (Coibion and Gorodnichenko, 2015; Bouchaud et al., 2019), construct use forecast errors and revisions. Raw forecast errors are defined as $ERR_t G_{it+1} = G_{it+1} - F_t G_{it+1}$, while the normalized version is $ERR_t g_{it+1} = g_{it+1} - F_t g_{it+1}$. Similarly, for "raw" growth, forecast revision is given by $R_t G_{it+1} = F_t G_{it+1} - F_{t-1} G_{it+1}$, while the normalized version is given by $R_t g_{it+1} = F_t g_{it+1} - F_{t-1} g_{it+1}$.

## 1.3 Descriptive statistics

Throughout this paper, the data consists of the entire international IBES summary file, to which we make a small number of restrictions. First, we focus on firms for which at least 10 realizations of sales growth are observed. Second, we focus on years for which sales forecasts are well populated enough, i.e. the period from 2000 to 2023. Finally, we require that forecasts can be observed 3 months after the end of the previous fiscal year end (see previous section). We end up with 122,395 observations. We then trim all of our variables at the $1^{st}$ and $99^{th}$ percentiles.

Figure 1: Main Sample: Number of Observations




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

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 as much as 3,729 internationally and 2,231 in the U.S. Overall, sales growth is well covered by IBES after 2000.

Table 1 shows summary statistics for the variables studied in this paper. The s.d. of raw log sales growth is 17%, and the mean is 6%. Normalized growth has a mean and median closer to zero, and a s.d. of 1.19 – it is not exactly equal to 1 because we normalize by the mean absolute deviation, not the firm-level standard deviation. Forecast errors and revisions have less variance than growth

Table 1: Descriptive statistics

| Variable | Mean | p25 | Median | p75 | s.d. | Obs. |
|---|---|---|---|---|---|---|
| *Panel A: Full sample* | | | | | | |
| Log sales growth (raw) | 0.06 | -0.02 | 0.06 | 0.14 | 0.17 | 108,694 |
| Growth forecast error (raw) | -0.01 | -0.05 | -0.00 | 0.05 | 0.13 | 108,682 |
| Growth forecast revision (raw) | -0.01 | -0.04 | -0.00 | 0.03 | 0.11 | 105,599 |
| Log sales growth (normalized) | 0.00 | -0.70 | 0.01 | 0.72 | 1.20 | 108,694 |
| Growth forecast error (normalized) | -0.06 | -0.52 | -0.02 | 0.45 | 0.96 | 108,682 |
| Growth forecast revision (normalized) | -0.04 | -0.39 | -0.03 | 0.32 | 0.80 | 105,598 |
| *Panel B: US Firms* | | | | | | |
| Log sales growth (raw) | 0.06 | -0.02 | 0.05 | 0.13 | 0.17 | 64,870 |
| Growth forecast error (raw) | -0.01 | -0.06 | -0.00 | 0.05 | 0.13 | 64,674 |
| Growth forecast revision (raw) | -0.00 | -0.04 | -0.00 | 0.03 | 0.10 | 63,880 |
| Log sales growth (normalized) | 0.01 | -0.70 | 0.02 | 0.73 | 1.20 | 64,717 |
| Growth forecast error (normalized) | -0.07 | -0.59 | -0.04 | 0.47 | 1.01 | 64,556 |
| Growth forecast revision (normalized) | -0.02 | -0.36 | -0.01 | 0.34 | 0.79 | 63,700 |
| *Panel C: Non US Firms* | | | | | | |
| Log sales growth (raw) | 0.07 | -0.01 | 0.06 | 0.15 | 0.18 | 43,824 |
| 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 | 41,719 |
| Log sales growth (normalized) | -0.00 | -0.70 | -0.01 | 0.72 | 1.20 | 43,977 |
| Growth forecast error (normalized) | -0.03 | -0.43 | -0.01 | 0.41 | 0.88 | 44,126 |
| Growth forecast revision (normalized) | -0.07 | -0.43 | -0.05 | 0.30 | 0.81 | 41,898 |

Note.

itself, consistent with the idea that forecasters cannot accurately predict future growth. We show in Appendix Figure B.1 the distribution of forecast error, which centered around zero, and non-normal, consistent with the non-normality of growth that we document below. Finally, variables on U.S. and international firms have similar distributions.

# 2 Motivating Facts

In this section, we document the three key facts that will be the building blocks of our model presented in Section 3.

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

Let us start with the key empirical fact of this paper. Following Coibion and Gorodnichenko (2015), expectations formation research typically estimates the following linear relationship between forecast errors and revision:

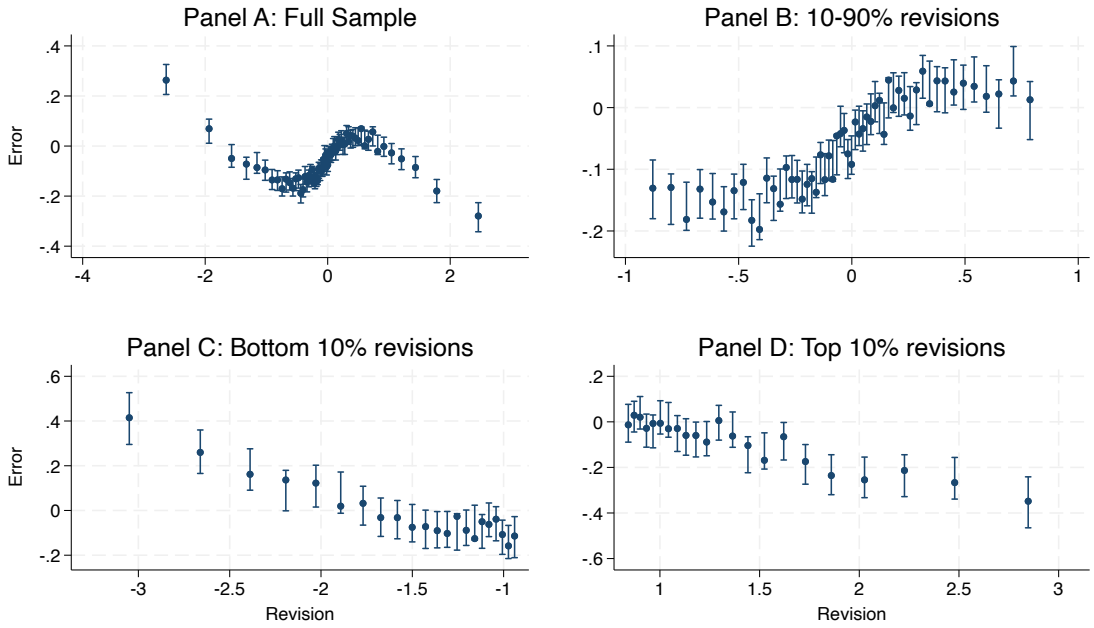
$$ERR_t Y_{it+1} = \alpha + \beta R_t Y_{it+1} + \epsilon_{it+1} \tag{1}$$

for any variable $Y$. Full information rationality predicts $\beta = 0$ for consensus forecasts (Coibion and Gorodnichenko, 2015). Plain rationality predicts $\beta = 0$ for individual forecasts. $\beta > 0$ is typically interpreted as evidence of information frictions (Coibion and Gorodnichenko, 2015), or, if run at the forecaster level, plain under-reaction (Bouchaud et al., 2019, Ma et al., 2020). In contrast, $\beta < 0$ is interpreted as evidence of overreaction (Bordalo et al., 2020). All these papers restrict their analyses to linear functional forms, as in equation (1).

Figure 2 provides preliminary evidence that, for revenue growth forecasts, this relationship is actually non-linear. This figure is a binned scatter plot of errors $ERR_t Y_{it+1}$ as a function of revisions $R_t Y_{it+1}$, using 100 bins. As in the rest of the paper, we focus on normalized sale growth $Y_{it+1} = g_{it+1}$ (as defined in Section 1). The chart is quite similar for raw (non-normalized) sales growth, as shown in Appendix figure B.2.

Both panels show strong evidence of non-linearity (Panels A1 and B1 show the non-linear relationship at all levels of revisions). For revisions in the bulk of the distribution, we find that errors are increasing in revisions (Panels A2 and B2, which focus on revisions between the $20^{th}$ and

Figure 2: The Error-revision relationship




We show here the binned scatter plot of the error-revision relationship for *normalized* 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 quintiles of revisions 2,3 and 4 (i.e. percentiles 20-80). Panel C focuses on the first quantile of revisions (i.e. below the $20^{th}$ percentile) and panel D on the top quantile of revisions (i.e. above the $80^{th}$ percentile).

$80^{th}$ percentils). Thus, forecasters are *under-reacting* in response to moderately-sized news shocks. This is consistent with evidence from Bouchaud et al. (2019) on EPS forecasts in United states. Ma et al. (2020) find similar evidence on log revenue forecasts from managers' expectations in the U.S. (using managerial guidance data from IBES) and Italy (using a survey from the Bank of Italy).

For large revisions, however, this relationship is reversed. In the face of extreme news, forecasters seem to be *over-reacting*: large revisions lead to surprises in the opposite direction (Panels A3,A4 and B3,B4, which focus on the top and bottom $20^{th}$ percentiles of the revision distributions). For instance, a positive news is followed by negative surprise, i.e. a realization below expectations. A similar non-linearity is visible, but not discussed, in U.S. earnings forecasts in Bouchaud et al. (2019).

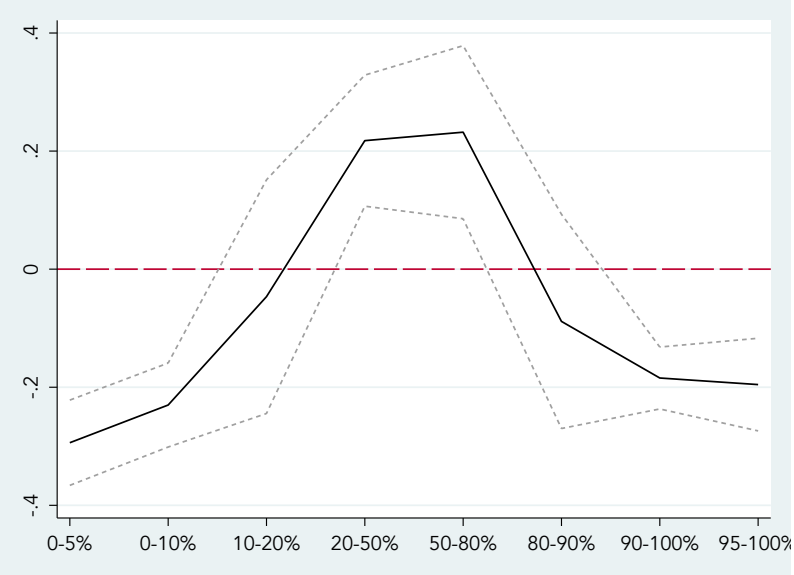
In Figure 3, we provide more rigorous statistical evidence of the S-shaped pattern shown in Figure 2. In this Figure, we report the results of 8 regressions. Each regression uses the specification in equation (1)), double-clustering error terms at the firm and the year level. We focus here on normalized growth – but results are similar with raw growth as shown in Appendix Figure B.3. Each linear regression is run on a different subsample, whose range is reported on the x-axis. On the y axis, we report the point estimate of $\beta$ (solid black line) and its 95% confidence interval (dashed gray lines).

The results in Figure 3 confirm what can be seen from Figure 2. Between centiles 20 and 80, errors are a significantly *positive* function of revisions. Outside of these bounds, in the tails of the distribution of revision, errors are significantly *negatively* correlated with revisions.

Most of the literature on analysts forecasts focuses on U.S. firms, for which data are believed to be of higher quality. We use international data in order to maximize the number of observations in the tails of the distribution – our empirical analysis and our theory relies on these tails. WE do, however, confirm that our results do not depend on the presence of international firms by splitting our sample into U.S. and international firms. We report the results in Figure 4. This Figure has exactly the same structure as 3, and shows that, indeed, underreaction in the bulk, and overreaction in the tail, is a feature present both in the U.S. and abroad.

Overall, the evidence on forecast errors and revisions points towards a different treatment of large v. 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.

Figure 3: CG regression coefficient by quantile

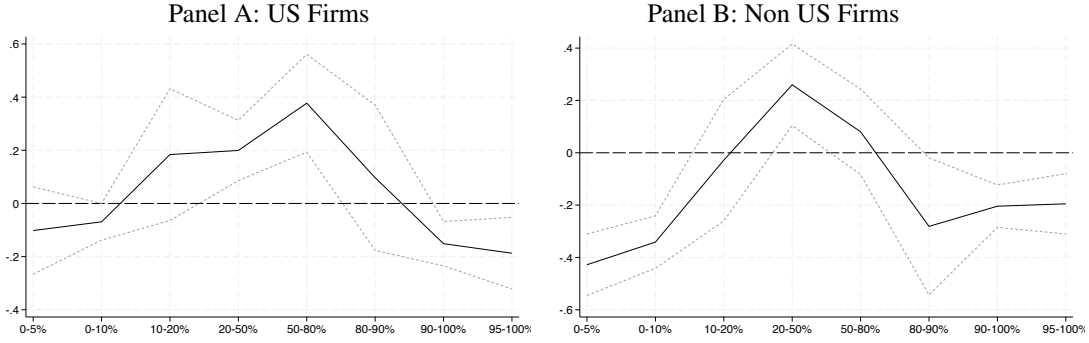



Note: In this Figure, we report the estimates of $\beta$ in the following regression:

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

where $g_{it}$ is the normalized log sales growth rate, 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 revisions. The point estimate of $\beta$ is the solid black line, while the dashed lines corresponds to the 95% confidence interval.

Figure 4: The Error-revision relationship: US vs non-US




Note: 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, 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 revisions. The point estimate of $\beta$ is the solid black line, while the dashed lines corresponds to the 95% confidence interval. Panel A runs this analysis on U.S. firms only, while Panel B runs it on international firms only.

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

We now document that, in our data, the distribution of sales growths has fat tails, one of the two building blocks of our model. Many variables relevant for finance and economics are not normally distributed Gabaix (2009). It is for instance well-known that the distribution of firm sizes follows a Zipf law (Axtell, 2001). For growth rates, Stanley et al. (1996) and Bottazzi and Secchi (2006) find that the distribution of COMPUSTAT firms follows a Laplace distribution (which has fatter tails than gaussian distributions). Here, we document the presence of fat tails in our sample.

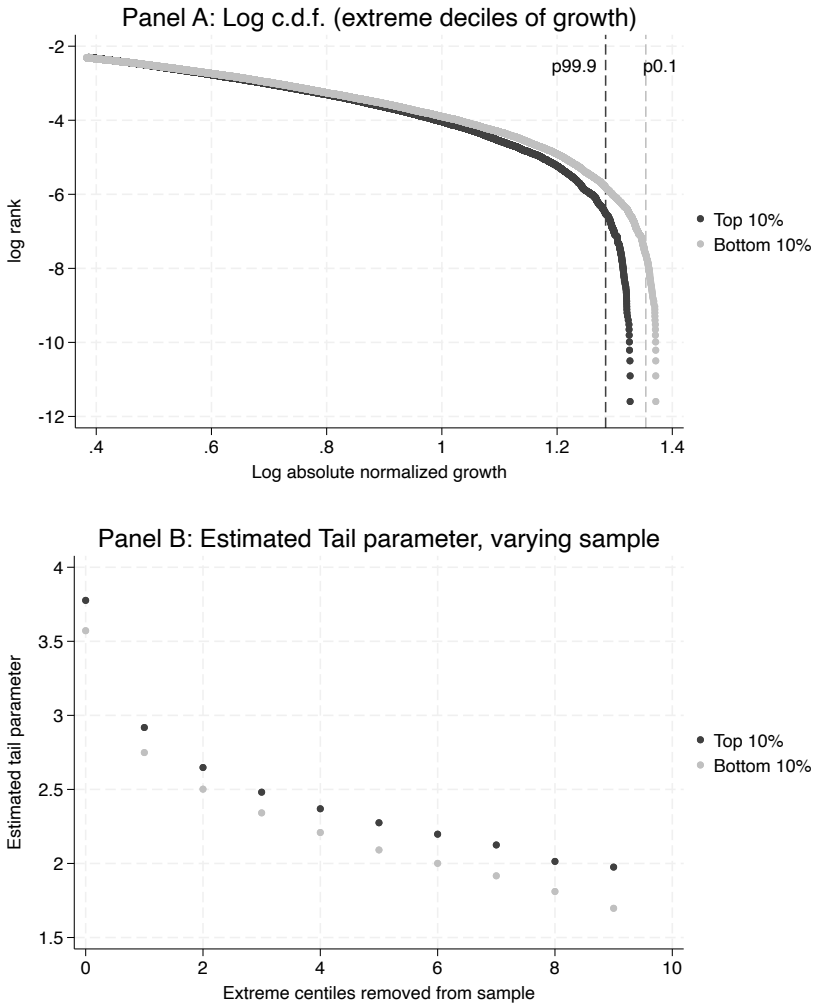
We look at the tail of the distribution in Figure 5. In this figure, we focus on the distribution of log normalized growth $g_{it}$. We focus on *normalized* growth here, so that the mean growth is zero, and our results do not depend on the presence of heterogeneous volatilies (Moran et al., 2024) (it turns out our results are not very sensitive to this adjustement). Further, we restrict our analysis of the tails to the top and bottom 10% of observations.

Figure 5, Panel A shows that the tails of the growth distribution are quite symmetric, and reasonably well fitted by power laws with low tail coefficients (somewhere between 2 and 3). Both tails are thinning out in the very end, but this only represents a small fraction of each decile. In Panel A, we classicaly plot, on the x-axis the log of growth ($\log g_{it}$ for the top decile, $-\log(-g_{it})$ for the bottom decile) and on the y-axis the log rank of the observations ($\log(1 - H(g_{it})$ for the top decile, $\log(1 - H(-g_{it})$ for the bottom decile, where $H$ is the empirical c.d.f. of growth). The power low approximation – the straight line – is a good approximation except into the top 0.1 percent of the distribution. In Panel B, the black dots show the coefficient $\beta$ of the following regression:

$$\log(1 - H(g_{it})) = \alpha + \beta \log g_{it} + \epsilon_{it}$$

run on various subsamples of the top decile, excluding the top centiles of the distribution (for instance, when the x coordinate is 1, we exclude from the regression the top centile of the top decile). The gray dot correspond to the symmetric analysis in the right tail. Overall, this second graph confirms what is obvious in Panel A: the power law approximation has much thicker tails if one removes extreme observations from the tails. Overall, it seems that a value of the tail parameter of about 2.5, on both sides, is a reasonable approximation. Thus, tails for growth rates are thinner than tails for firm size (for which the second moment does not even exist, as in Axtell (2001)), but still quite thick. For comparison, we also show, in Appendix Figure B.4 the exact same analysis on the universe of COMPUSTAT firms between 2000 and 2023 (therefore nearly 110k observations of firms listed in the U.S.). Reassuringly, the tail parameter is roughly identical as in our IBES actuals

Figure 5: Sales growth tail: Pareto approximation



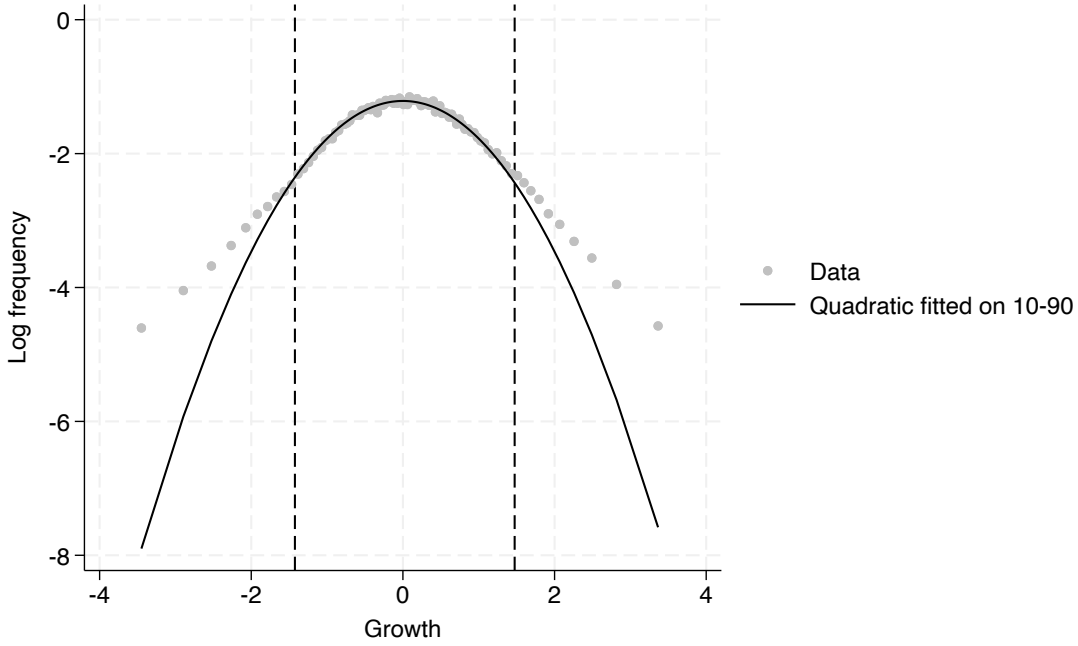

Note: We describe here the tails of the distribution of normalized log growth rates $\log g_{it}$. In Panel A, we show the scatter plot of the rank (one minus c.d.f.) of growth in the top and bottom deciles (c.d.f. being normalized to 1 at the extreme value of each decile to make things comparable). The x-axis corresponds to $\log g_{it}$ for the right tail, and $-\log(-g_{it})$ for the left tail. We restrict ourselves to the top and bottom 10% of growth values. The dashed lines highlight percentiles 0.1 and 99.9. Panel B shows how the tail parameter varies as a function of how much of the tail one excludes from the regression. Let $H$ be the c.d.f. of growth. In the top decile of the distribution of $g_{it}$, we run the regression:

$$\log(1 - H(g_{it})) = \alpha + \beta \log(g_{it}) + \epsilon_{it}$$

on ten different subsample, each one excluding the most extreme centiles of the distribution. When $x = 0$, the subsample does not exclude any observation. When $x = 1$, the subsample only excludes the top centile, etc. On the y-axis, for the black dots, we represent the slope estimate $\hat{\beta}$. The gray dot emerge from the same methodology, except focusing on the left tail (the lowest decile). Reading: If one runs the above regression in the top decile but after excluding the top 5 percentiles, one obtains a tail parameter of 2.3. If one does so for the bottom decile, while excluding the 2 lowest percentiles, one obtains 2.5.

sample.

Figure 6: Log density of normalized sales growth




Note: We show here the log density of the log of normalized log sales growth. We group observations by centiles of growth. For each centile, we compute the log density by taking the difference between the log fraction of observations ($-\log(100)$) and the log growth range of the centile. On the figure, each gray dot corresponds to one centile of growth, with the mean value of growth on the x-axis, and the log density on the y-axis. The two dashed vertical lines correspond to the cutoff values of the top and bottom decile of the distribution of log log growth. Finally, the black line corresponds to a quadratic fit estimated on all quantiles in the bulk of the growth distribution, i.e. between the $11^{th}$ and the $90^{th}$ one.

Last, we explore the distribution of growth rates in the bulk of the distribution by plotting the log p.d.f of sales growth, along with a simple quadratic fit. We do this in Figure 6. To compute this density, we first group observations by centile of sales growth. For each centile, we compute the average growth, which we use on the x-axis. On the y-axis, we show the log p.d.f. We 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).

We also fit a simple quadratic function on the bulk of the distribution, i.e. all centiles except the top and bottom decile. For centile $c \in [10; 90]$, we estimate the following relationship:

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

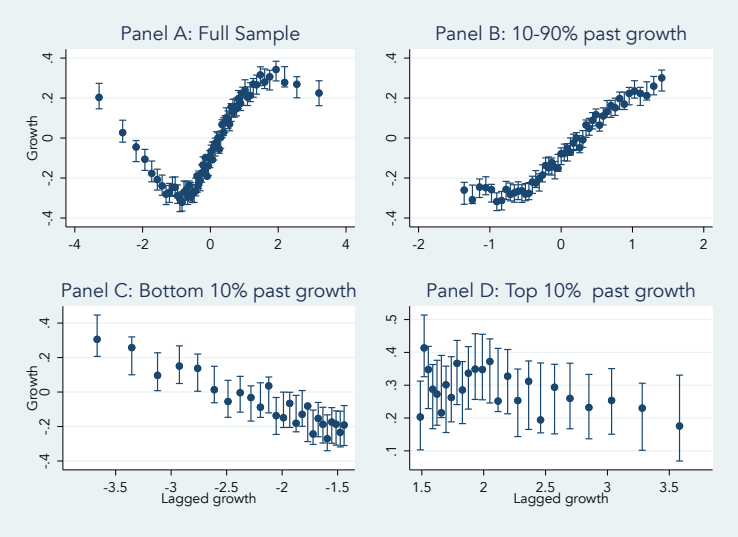
where $g_c$ is the mean growth of centile $c$ and $\log h(g_c)$ the corresponding log density. In the dataset made of these 80 centiles, the $R^2$ of this regression is 0.98, so the fit is quite good in the bulk of the distribution, and we obtain $\frac{1}{\Sigma^2} = 1.14$. Thus, in the bulk, the distribution is close to a normal with mean zero (by normalization of $g$) and standard deviation $\Sigma = \frac{1}{\sqrt{1.14}} \approx 0.93$. Note that the sample s.d. of growth is 1.19, so the gaussian approximation in the bulk has lower variance than the overall distribution. This is consistent with the fact that the distribution has heavier tails than normal.

Overall, the distribution of sales growth is reasonably well approximated by a Gaussian distribution in its bulk, and a power law in the tails. This observation will guide our theoretical modeling, to which we now turn.

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

Our final fact that will be important for the model of the DGP that we develop is that the conditional expectation of growth is non-linear. Figure 7 shows a binned scatter plot of $g_{it+1}$ against $g_{it}$ in Panel A. Panels B,C,D then zoom in on three segments of the data: the bulk of the distribution, and the bottom and top deciles. As is clear from this Figure, future growth is not a linear function of past growth in the data. We also show in Appendix Figure B.5 that our large sample of COMPUSTAT firms generates an almost identical shape.

Figure 7: Future growth against current growth in the data




Note: This figure reports 4 binned scatter plots of $g_{it+1}$ 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.

# 3 Model

In this section we develop a parsimonious model that features extreme shocks and linear expectations in order to capture the non-linear behavior of expectation errors of Figure 2. We start with the two assumptions on the DGP and expectations formation. We then derive formal results and provide numerical simulations showing that this model can reproduce our main empirical fact.

## 3.1 Data-Generating Process

The first piece of the model is the data-generating process. Overall, the unconditional mean of $g_t$ is assumed to be zero, because we are modelling normalized growth rates. We omit here the firm index $i$ for clarity's sake. We assume that adjusted growth follows the following data-generating process:

$$g_{t+1} = g^*_{t+1} + \epsilon_{t+1} \tag{2}$$

$$g^*_{t+1} = \rho g^*_t + u_{t+1} \tag{3}$$

where $u_{t+1}$ and $\epsilon_{t+1}$ are assumed uncorrelated and i.i.d. $g^*_t$ is the *unobservable* latent state that follows the classic linear AR1 dynamics described in Equation (3). The distribution of $u_t$ is assumed to have zero mean and variance $\sigma^2_u$ – we label $u$'s p.d.f. as $\phi$. Given the AR1 structure, the underlying state $g^*_t$ has mean zero and variance $\sigma^2_{g^*} = \frac{\sigma^2_u}{1-\rho^2}$ – the p.d.f. of $g^*$ is labelled $\phi^*$. For the underlying state, we will work with gaussian and non-gaussian process (assuming $u$ gaussian buys us a few results).

The key difference with the existing literature on expectations formation is that $\epsilon_{t+1}$ follows a probability distribution density that also has mean zero but heavy tails. We note its p.d.f. as $f$, its variance $\sigma^2_\epsilon$, and will assume:

$$f(\epsilon) \propto |\epsilon|^{-\nu} \text{ as } |\epsilon| \to +\infty \tag{4}$$

This assumption is in line with evidence from Section 2, in which we found that the tails of the growth distribution were well approximated by a power law (of tail parameter between 2.5 and 3.5). In our numerical calibrations and estimations, we will use for $f$ a Student's t distribution with $\nu$ degrees of freedom. In the tails, the Student distribution converges to a Pareto distribution of tail

parameter $\nu$, consistent with Equation (4), but the student distribution is defined for all real numbers and is symmetric around zero.

## 3.2 Subjective Expectations Formation

After the DGP, the second building block of the model is the formation of expectations. Our core assumption is that forecasters (wrongly) assume that the growth process is linear. This assumption is based on the idea that economic agents use simplified, "sparse", models of reality to formulate expectations (Fuster et al., 2012; Gabaix, 2018). We thus assume that agents assume $g_t$ follows a linear AR1 model, estimate it on data and use this model to form forecasts. One advantage of this representation is that the term structure of forecasts is naturally defined, as agents calculate mathematical expectations under this model.

Hence, we assume that the forecaster's subjective expectations are given by:

$$F_t g_{t+1} = EL(g_{t+1}|g_t) = \beta g_t \tag{5}$$

where we set the constant to zero because $g_t$ is normalized growth, so that $Eg_t = 0$. As a result, the OLS estimate of the constant, $\bar{g}_t - \hat{\beta}\bar{g_{t-1}}$, is centered around, and asymptotically converges to, zero.

Using the formula for the OLS estimate of $\beta$, along with the structure of our model, it is straightforward to see that:

$$F_t g_{t+1} = \rho \frac{\sigma^2_{g^*}}{\sigma^2_g} g_t \tag{6}$$

where $\sigma^2_g$ is the unconditional variance of $g_t$.

Thus, the linear forecaster uses a persistence coefficient that is lower than $\rho$. This is rational, as the underlying state $g^*$ is measured with error. Such a shrinkage of the persistence becomes bigger as the share of the variance of $g$ explained by the persistent process is smaller and smaller.

Finally, we similarly assume that two-period ahead forecast are also driven by a simple linear projection:

$$F_{t-1}g_{t+1} = EL(g_{t+1}|g_{t-1}) = \beta g_{t_1} = \rho^2 \frac{\sigma^2_{g^*}}{\sigma^2_g} g_{t-1} \tag{7}$$

## 3.3 Characterizing rational expectations

To understand the structure of the forecast error implied by equation (5), we need to understand the expectation of $g_{t+1}$ conditional on $g_t$. In this Section, we show that this conditional expectation is a non-linear and characterize its shape, under the assumption that the hidden state $g^*_t$ is gaussian. This characterization is achieved through a direct application of Tweedie's formula (Robbins, 1956; Efron, 2012), which requires $g^*_t$ to be gaussian. We summarize these characteristics in the following proposition:

**Proposition 1.** *Assume that the hidden state $g^*_t$ is gaussian. Assume $h$ is the marginal distribution of the growth rate, given by the formula:*

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

*where $\phi_{g^*}$ is the normal p.d.f. of $g^*$. Then,*

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

$$E(g_{t+1}|g_t) = -\rho \sigma^2_{g^*} \left( \frac{d}{dg} \log h(g) \right)(g_t) \tag{8}$$

- *The derivative of the conditional expectation w.r.t. $g_t$ is given by:*

$$\left( \frac{E(g_{t+1}|g)}{dg} \right)(g_t) = \rho \left( 1 - \frac{1}{\sigma^2_{g^*}} Var\left(g^*_t|g_t\right) \right)$$

- *As a result, the linear forecaster is less sensitive ("underreacts") to $g$ than the rational one iff:*

$$\frac{\sigma^2_\epsilon}{\sigma^2_g} > \frac{Var\left(g^*_t|g_t\right)}{\sigma^2_{g^*}}$$

*and overreacting if the inequality is reversed. Further, when $\epsilon$ is normal, this relationship holds with equality: the linear forecaster is rational.*

*Proof.* See Appendix A.1. □

We do not have a closed-form solution for $h$, but the beauty of Tweedie's formula is that it involves the marginal density of growth $h$, which is directly observable from the data – we return to this point Section 4. Further, proposition 1 provides a few interesting insights. First, as shown in Figure 6, in the bulk of the distribution, the growth distribution is well approximated by a normal distribution: $\log h(g) \approx -\frac{g^2}{2\Sigma^2} + \text{cst}$. Then, the rational expectation is given by:

$$E(g_{t+1}|g_t) \approx \rho \frac{\sigma^2_{g^*}}{\Sigma^2} g_t \tag{9}$$

which is linear. This is the well-known result, widely used in the literature, that conditional expectations with gaussian mixtures are linear. Here, Tweedie's formula allows us to extend this result to local approximation. If $h$ is locally normal, then the conditional expectation is locally linear. As we will see in the following section, the approximation (9) also holds asymptotically for any distribution of $u$.

Moreover, comparing equations (9) and (6), it is clear that the rational expectation is more sensitive to past growth than the linear forecast if and only if:

$$\Sigma > \sigma_g$$

i.e. when the s.d. of the bulk normal approximation is lower than the s.d. of the entire distribution. As mentionned in Section 2.2, this is the case, as the s.d. of the normal approximation is 0.93 while the overall s.d. is 1.19, due to fatter tails than normal. Thus, fat tails explain why the slope of the linear approximation is less than the rational expectation.

Second, the flip side of this reasoning is that, in our model, the conditional expectation are not linear everywhere. Since $g_t = g^*_t + \epsilon_t$, with $\epsilon$ heavy-tailed, the mixture distribution $h$ is not normal. Formula (1) allows us to be more precise. We saw in Figure 6 that, in the top and bottom deciles of growths, the distribution $h$ looks like a power law: $\log h(g) \approx -\nu \log(|g|)$. Formula allows us to write that, in this case:

$$E(g_{t+1}|g_t) \approx \rho \sigma^2_{g^*} \frac{\nu}{g_t} \tag{10}$$

so that, in the tails of growth distribution, the expectation is *decreasing* in past growth. The intuition is that very large (and, say, positive) values of $g_t$ are presumably 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 slow-moving state $g^*$. As a result, the rational forecaster reverts to the unconditional expectation (zero, here) as her best possible forecast. This is in stark constrast with the linear forecaster of equation (5) whose forecasts increases linearly with $g_t$. As we will see below, the above formula does not require to assume that $g^*$ is normal when we look asymptotically.

Third, the sensitivity of the rational forecast to $g_t$ is higher, the lower the variance of $g_t^*$ conditional on $g_t$. This the second bullet point of the proposition. In the limit case where $g_t$ is perfectly informative about the underlying state, the rational forecaster will be fully responsive to past realization of $g_t$ to forecast $g_{t+1}$ – i.e. will apply the full persistence coefficient $\rho$. On the contrary, if $g_t$ is uninformative about the hidden state, then the conditional expectation does not depend at all on $g_t$.

Fourth, the linear forecaster is more likely to *underreact* to past growth when the share of variance of $g$ that is accounted by the tail shock is high. This is because in this case, the linear forecaster will not use the persistence of growth very much. Second, underreaction is more likely when $g_t$ is more informative about future growth. In this case, the rational forecaster will use this informativeness more than the linear forecaster.

## 3.4 Asymptotic results

When the underlying state $g^*$ is non-gaussian, Tweedie's approximation no longer holds but we can still characterize the rational approximation for large absolute values of growth. This is what we do in this Section. We start with the case of our model, where :

**Proposition 2.** *If in the limit, the distribution of $\epsilon$ is Pareto:*

$$f(\epsilon) \propto |\epsilon|^{-\nu} \text{ as } |\epsilon| \to +\infty$$

*Then, for any distribution of $u$, the expectation of future growth conditional on past growth only is given by:*

$$E(g_{t+1}|g_t) \approx \rho\nu\frac{\sigma_g^2}{g_t} \text{ as } |g_t| \to +\infty$$

*where $\sigma_g^2$ is the unconditional variance of $g$.*

*Proof.* See Appendix A.2. □

Proposition 2 confirms that, when transitory shocks $\epsilon_t$ are non-Gaussian, the "rational" expectation $E(g_{t+1}|g_t)$ is not linear in $g_t$. When shocks are asymptotically Pareto, we can characterize this non-linearity for large $g_t$. In the limit where $g_t$ goes to infinity, the expected value is simply zero. The intuition here is that, if last year's growth was very high, the rational forecaster is almost certain this is due to the transitory shock. It is then impossible to know what the true $g_t^*$ was, so the best guess for $g_{t+1}$ is the unconditional mean, zero.

In turn, when $g_t$ is large but not infinite, there is still some indication that the hidden state $g_t^*$ was positive: This points towards a forecast of $g_{t+1}$ above the unconditional mean. The difference is smaller when $\epsilon$ is more fat-tailed ($\nu$ is closer to 1): In this case, past growth is more likely to come from a past large temporary shock. Also, the difference decreases as past growth increases, as it becomes more and more likely that it was generated by the fat-tailed temporary process.

To further build intuition on these mechanisms, the following proposition makes a different assumption but the shape of the distribution of $\epsilon$:

**Proposition 3.** *If in the limit, the distribution of $\epsilon$ has the form:*

$$f(\epsilon) \propto e^{-\frac{1}{\sigma_\epsilon^2}\epsilon^\theta} \text{ as } |\epsilon| \to +\infty$$

*Then, the following two properties hold:*

1. *If $\theta = 1$ (exponential distribution), the conditional expectation tends towards a constant for $g$ large:*
$$E(g_{t+1}|g_t) \approx L \text{ as } |g_t| \to +\infty$$
2. *If the distribution of $u$ is Gaussian, then:*
$$E(g_{t+1}|g_t) \approx \frac{\sigma_g^2}{\sigma_\epsilon^2}\theta g_t^{\theta-1} \text{ as } |g_t| \to +\infty$$

*Proof.* See Appendix A.3. □

This proposition has mostly illustrative purposes, as the distribution of shocks here is non-standard (in the rest of the paper, we assume that $f$ is asymptotically Pareto). The first result of the proposition holds for any distribution of the underlying growth rate $u$: If the temporary shock is exponential, i.e. thicker-tailed than normal but not as much as Pareto, then the rational expectation is no longer *decreasing* in $g_t$ but is flat. The second result only holds for Gaussian underlying processes $g_t^*$. First,

as soon as tails are thinner than exponential ($\theta > 1$), rational expectations become increasing in $g_t$ (instead of constant). But they are concave, not linear, as long as $\theta < 2$. When $\theta = 2$ (temporary shock is Gaussian), rational expectations are linear, as is the case in almost all of the literature on information frictions. Overall, as soon as tails are thicker than Gaussian, the expectation of $g_{t+1}$ becomes less than linear in $g_t$. When the shock is asymptotically exponential, it is constant, and when it is Pareto, it becomes decreasing.

Combining all the above results, it is then straightforward to show that overreaction predominates in the tails of the distribution, while underreaction needs to predominate the rest of the distribution. This is summarized in the following proposition:

**Proposition 4.** *If in the limit, the distribution of $\epsilon$ has the form:*

$$f(\epsilon) \propto \epsilon^{-\nu} \text{ as } |\epsilon| \to +\infty$$

*Then, there exists a growth threshold $\overline{g}$ s.t.:*

$$E\left(g_t(g_{t+1} - F_t g_{t+1}) || g_t| > \overline{g}\right) < 0$$
$$E\left(g_t(g_{t+1} - F_t g_{t+1}) || g_t| < \overline{g}\right) > 0$$

*i.e. there is underreaction for $|g| < \overline{g}$, and overreaction for $|g| > \overline{g}$*

The proof of this proposition rests on the definition of $F$, which ensures that $E\left(g_t(g_{t+1} - F_t g_{t+1})\right) = 0$. Then, the conditional expectation of this product is negative for large absolute growth, so it must be negative otherwise. This result of course also holds if $f$ is a Laplace distribution in the tails, given the results of Proposition 3.

The conclusion of this analysis is that, if the temporary shock $\epsilon$ has fat tails, the mean value of $g_{t+1}$ is less than linear in $g_t$ when past growth is large. In this case, the linear forecaster is mechanically overreacting to large news. This is consistence with the evidence presented in figure 2.

## 3.5 Numerical Results

In this section, we use simulations to show that our model predicts underreaction conditional on intermediate value of $g_t$, while confirming our formal results that the linear forecaster overreacts

to large news. Simulations are necessary for two reasons: (1) there are no formal results on the forecasting rule outside of $|g_t| \to +\infty$ (we cannot establish underreaction for intermediate values of past growth) and (2) rather than predicting the reaction of the forecast error to the lagged level of $g_t$, we are interested in the response to past revision, $F_t g_{t+1} - F_{t-1} g_{t+1}$. For this last relation, there is no formal result either.

To run our simulations and build intuition, we set the parameters of the DGP somewhat arbitrarily – we estimate them in the next section. For the underlying state $g^*$, we set the persistence parameter $\rho = .3$, while the innovation $u$ follows a normal distribution with unit variance. For the transitory shock $\epsilon$, we use a student distribution with degrees of freedom $\nu$. The Student distribution has the advantage of being defined for both negative and positive values of $\epsilon$, while the pd.f. converges to a power law – with tail parameter $\nu$ – in the tails. We vary $\nu$ from 2 to 10. We simulate 1 million firms over 10 years each – this gives a total of 10 milions observations for which we have realizations of the DGP.[2]

To generate forecasts, we run, on the entire simulated dataset, two linear regressions:

$$g_{it} = \alpha_1 + \beta_1 g_{it-1} + \omega_{1,it}$$
$$g_{it} = \alpha_2 + \beta_2 g_{it-2} + \omega_{2,it}$$

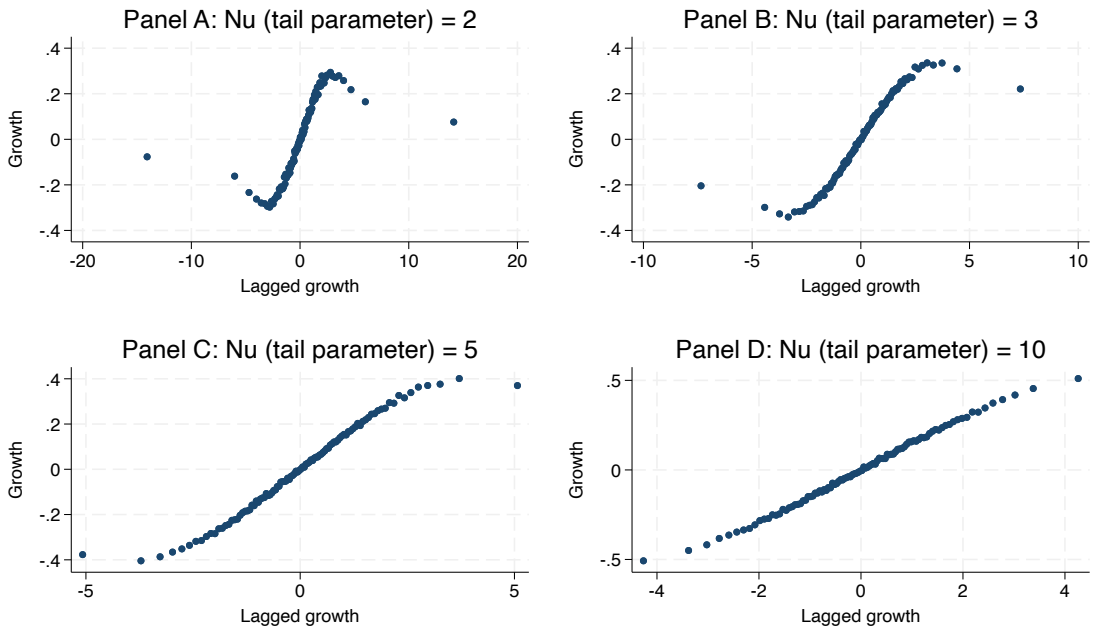
and use the estimates to define the forecasts. We define the one-period ahead forecast $F_{t-1} x_{it} = \widehat{\alpha_1} + \widehat{\beta}_1 g_{it-1}$. Simlarly, we define the two-period ahead forecast as: $F_{t-2} x_{it} = \widehat{\alpha_2} + \widehat{\beta}_2 g_{it-2}$. We then use these forecasts to generate the errors and revisions as in the data.

Before analyzing forecast errors, we first show the rational expectation $E(g_{it+1}|g_{it})$, i.e. the mean value of $g_{it+1}$ as a function of $g_{it}$, in Figure 8. In this figure, we show a binned scatter plot of $g_{it+1}$ against $g_{it}$ obtained from our simulated data. Consistent with our formal results, for low levels of $\nu$, the relationship is linear in the body of the distribution, but, as predicted by proposition 2 experiences "reversals" in the tails. As $\nu$ gets larger, the conditional expectation becomes closer to linear on the entire domain. This is consistent with the idea that, as the distribution becomes close to being Gaussian, the optimal expectations is closer to linear.

We are now ready to investigate the relationship between forecast errors and past revisions that is implied by our model. Figure 9 shows that, as long as the temporary shock has sufficiently fat tails, our model can generate the non-linear pattern we found in the data. The effect seems to disappear when $\nu$ is larger than 5 or 10. This is quite intuitive: As the process becomes closer and closer to

[2]We initialize the process with a draw of $g^*$ from its long-term distribution with mean 0 and variance $\frac{1}{1-\rho^2}$.

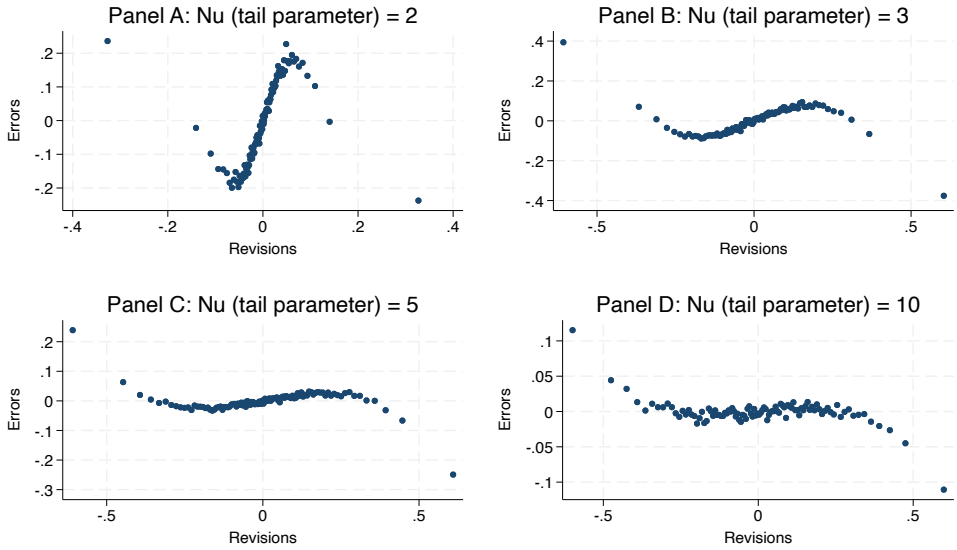
Figure 8: Actual vs. lagged actual in the simulated model: Varuing the tail parameter




Note: We show here the relationship between current and past growth in our model. The data-generating process for growth is described by equations (2)-(3). $u$ is a standardized normal, $\rho = .3$, and $\epsilon$ is a student distribution with varying degrees of freedom: 2 (panel A), 3 (Panel B), 5 (panel C) and 10 (Panel D). Our simulated data have 10 million observations and the plots are binned scatter plots with 100 bins.

Gaussian, the linear forecast is optimal and errors become harder and harder to predict.

Figure 9: Errors v. revisions in the simulated model: Varying the tail parameter




Note: We show here the relationship between forecast errors $g_{it+1} - F_t g_{it+1}$ and revisions $F_t g_{it+1} - F_{t-1} g_{it+1}$ in our simulated data. The data-generating process for growth is described by equations (2)-(3). $u$ is a standardized normal, $\rho = .3$, and $\epsilon$ is a student distribution with varying degrees of freedom: 2 (panel A), 3 (Panel B), 5 (panel C) and 10 (Panel D). Our simulated data have 10 million observations and the plots are binned scatter plots with 100 bins.

# 4 Quantitative Analysis

In this Section, our objective is to show that our simple model of forecast formation can quantitatively account for the main stylized fact described in Figure 3. To do this, we first need to obtain a quantitative model of sales growth that fits equations (2). This dynamical model is fitted on realized growth rates only. Then, we check that a linear forecast assumed in Equation (5), fitted on the entire data, can quantitatively fit the evidence. Finally, we discuss the implication of our findings for stock returns.

**Note: this section is a work-in-progress and is still incomplete**.

## 4.1 Testing the Non-Linear Conditional Expectation Quantitatively

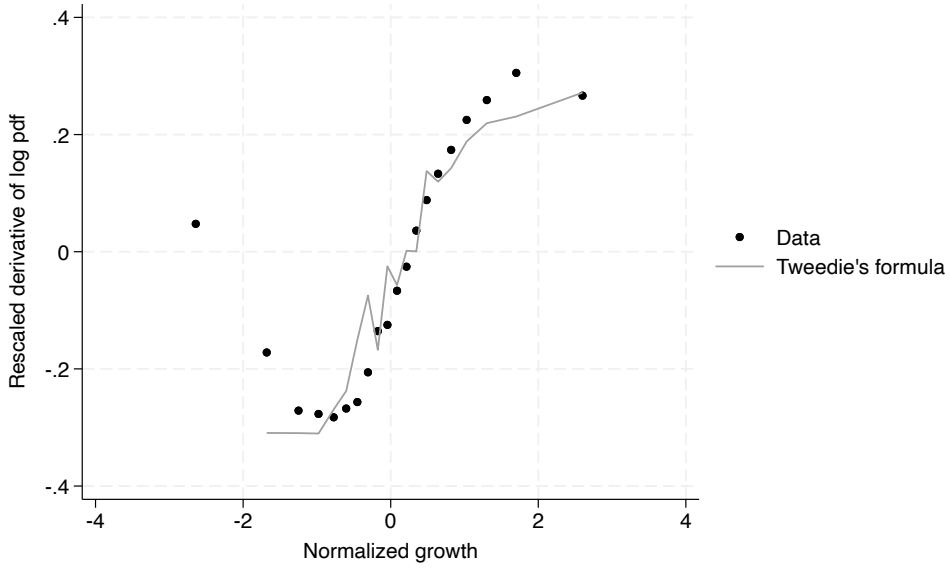
Proposition 1 allows us test our model more quantitatively. Indeed equation (8) states that:

$$E(g_{t+1}|g_t) = -\rho\sigma^2_{g^*}\left(\frac{d}{dg}\log h(g)\right)(g_t)$$

which is directly testable as the distribution $h$ (and therefore its log derivative) of normalized growth is observable in the data – as in Figure 6.

Based on this formula, we estimate the distribution $h$ in the data, then differentiate it, and check whether it is proportional to an empirical measure of $E(g_{t+1}|g_t)$. We report the result of this test in Figure 10. The black dots are a binned scatter plot of growth against past growth, exactly identical to the panel A of Figure 7. The grey line represent the fit of Tweedie's formula (which rests on the assumptions that (1) $g^*$ is a gaussian AR1 and (2) $g$ is the sum of $g^*$ and a temporary shock hidden in the distribution of $g$). We compute this fit the following way. Using the 20 vingitiles of $g_{it}$ from the data, we first estimate the p.d.f. $h$ by dividing the number of observations in each bin (equal $1/20$ of the total) by the size of each bin (which depends on the distance between the quantile breakpoints). We obtain 20 values of $\log h$. We then estimate the derivative by calculating *minus* the slope of $\log h$ between each quantile. We then regress the average future realization of $g_{it+1}$ (our estimate of the conditional expectation) on this slope, and use the predicted value.

Figure 10: Actual v lagged growth: Data v predicted from equation (8)




Note: The black dots correspond to the average realization of $g_{it+1}$ for 20 bins of $g_{it}$, the same analysis as in Figure 7, Panel A. The grey line represents the fit of our estimate of $\left(\frac{d}{dg}\log h(g)\right)(g_{it})$ (for all observations in the quantile) on the average future realization of $g_{it+1}$ in the same quantile of past growth.

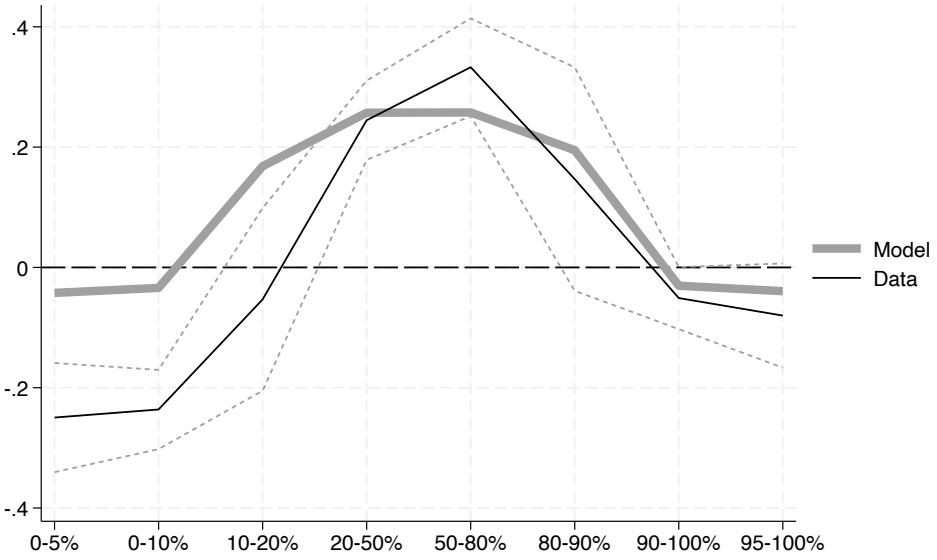
## 4.2 Estimating the data-generating process

**TBD: explanation of SMM estimation procedure**

We show in Figure 11 the fit of our model. We report the empirical coefficients of growth on lagged growth on the different subsamples, along with the confidence intervals, as in Figure 7, and overlay the coefficients obtained from our model. These coefficients are obtained using the parameters estimated above and 10 million observations.

Looking at Figure 11, we find that the model matches the data reasonably well, albeit not perfectly. The model strongly generate a stronger slope in the bulk of the distribution than in the tails. But in the tails, the slope is not negative as it is in the data – especially in the left tail. XXX

Figure 11: Model Fit




Note: We report here the empirical and predicted coefficients of growth on lagged growth, on various subsample. The empirical estimates, along with confidence intervals, are the same ones as in Figure 7. The dark line represents the coefficients obtained from simulating 1 million firms over 10 years, using the estimated parameters.

## 4.3 Calibrating the forecasting process

We now turn to the forecast. Our main model assumes forecasts are linear, but as it turns out, forecasts doTo match the data better, we allow actual forecasters to be partly rational:

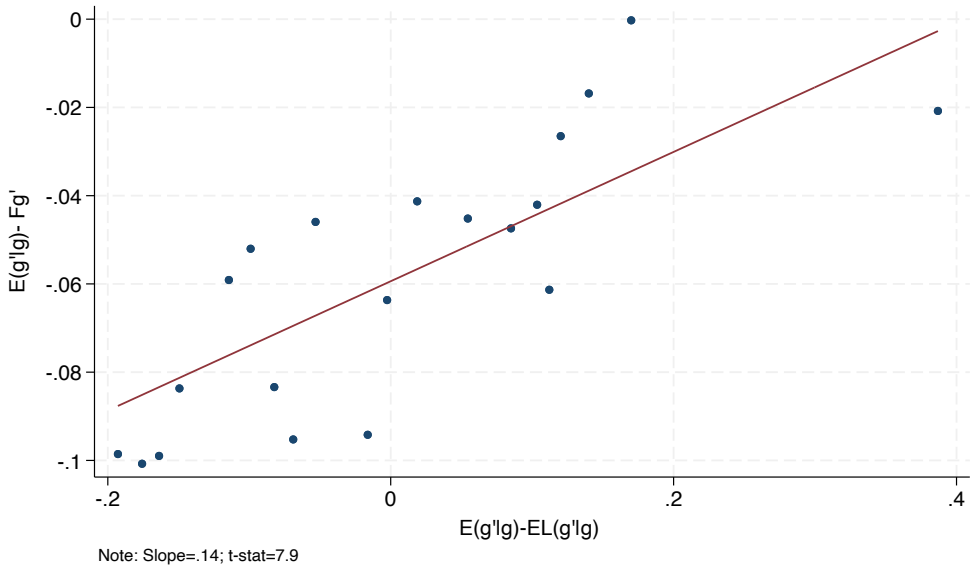
$$F_t g_{t+1} = \lambda EL(g_{t+1}|g_t) + (1-\lambda)E(g_{t+1}|g_t)$$

where $EL$ stands for the linear expectation. By definition of conditional expectations, the equivalent moment condition writes as:

$$E\left(\underbrace{(g_{t+1} - F_t g_{t+1})}_{\text{subjective error}}|g_t\right) = \lambda E\left(\underbrace{(g_{t+1} - EL(g_{t+1}|g_t))}_{\text{linear error}}|g_t\right) \tag{11}$$

so that the conditional subjective error should be linear with the conditional linear error. We test this relation in Figure 12. To compute the two conditional expectations, we split the sample into 20 quantiles of lagged growth, and calculate the averages of subjective and linear errors conditional on each one of these quantiles. We then produce in Figure 12 a binned scatter plot of conditional subjective on conditional linear error. Equation (11) predicts that the relation should be linear, which is confirmed the plot.

Figure 12: Subjective forecasts as functions of linear forecasts



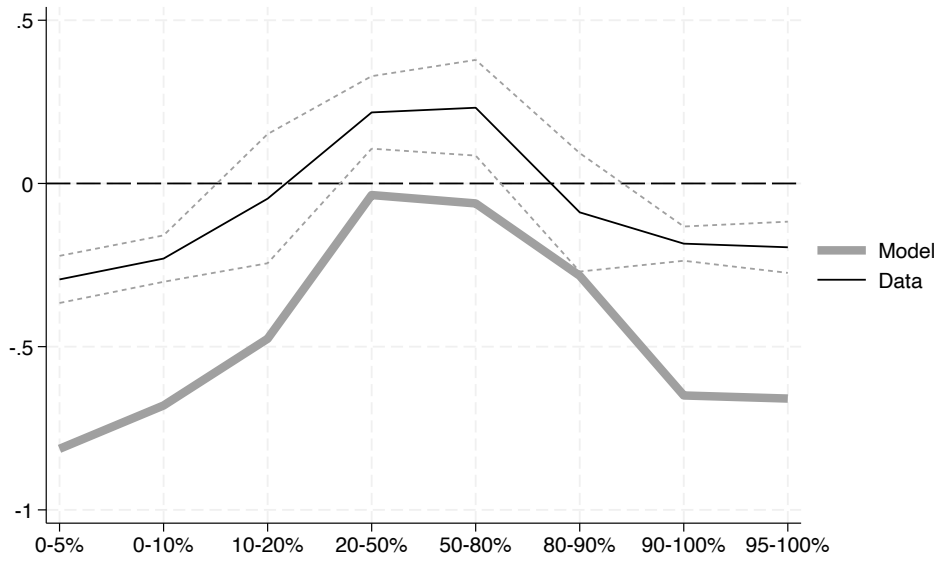

Note: In this graph

We then run the formal regression in equation (12) using our 20 quantiles. We report the result in Figure 12. We find forecasts to be 83% rational, and 17% linear. We now turn to the main prediction of our model.

## 4.4 Quantitative fit of the model

In Figure 13, we compare the sensitivity of errors to revisions in the model to the data. We plot the same sequence of coefficients, subsample by subsample, as in Figure 3.

Figure 13: Error on revisions: Model Fit




Note:

# 5 Application and Robustness

**Note: this section is a work-in-progress and is still incomplete**.

## 5.1 EPS forecasts

In this Section, we explore similar implication for EPS forecasts. EPS forecasts are much broadly studied in the literature than sales forecasts. Our focus on sales growth forecasts is relatively novel, because sales growth is a natural construct, it has a symmetric distribution and is stationary. EPS growth is harder to manipulate since many firms have negative earnings, and firms close to zero EPS will display very large growth rates. So instead, we follow here the literature and study EPS to price ratios (as, for example, in Bouchaud et al. (2019)).

More precisely, we extract EPS forecasts and their realizations from IBES, and run the following regression:

$$\underbrace{\frac{EPS_{it+1} - F_t EPS_{it+1}}{P_{it-2}}}_{\text{ERR EPS}_{it+1}} = \alpha + \beta \underbrace{\frac{F_t EPS_{it+1} - F_{t-1} EPS_{it+1}}{P_{it-2}}}_{\text{ERR EPS}_{it+1}} + \epsilon_{it+1}$$

on different subsamples of revision in order to analyze non-linearity. We start in 1990 instead 2000 since EPS forecasts are much better reported for earlier years. The common normalization by $P_{it-2}$ ensure that (1) the price is in the information set at the time of forecast and (2) the forecasting variable does not trend over the period (the level of EPS in dollars does trend). The price $P_{it}$ comes from the IBES "actual" files, where $t$ is the date at which the forecast is made. We further lag the price by one full year to be conservative, but this does not affect our results at all.

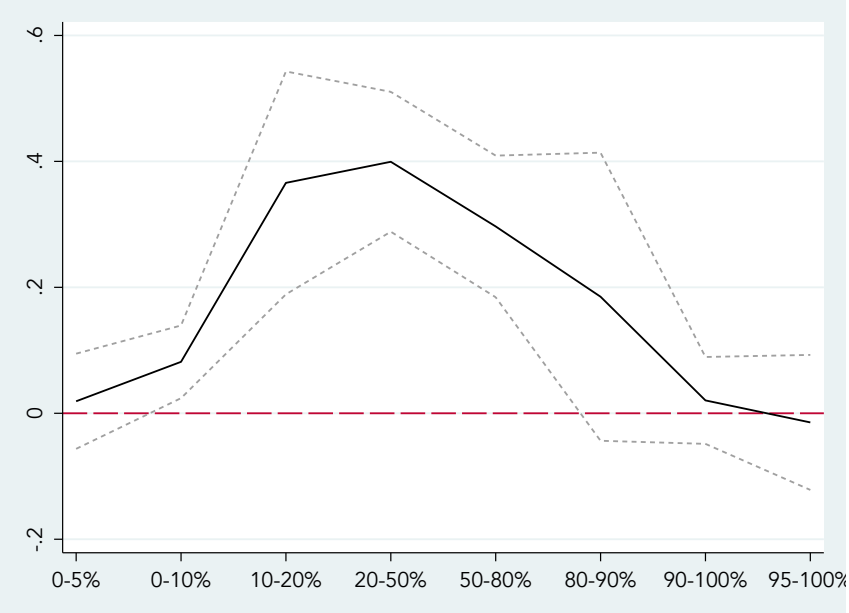
We show in Figure 14 the same analysis as the one we produce for growth forecasts in Figure 3. Quite clearly, the relationship between errors and past revisions is non-linear: significantly positive and linear for revisions between the second and the ninth decile, but flat in the two extreme deciles. This non linearity is already apparent in Bouchaud et al. (2019) but in our data it is more visible as the size of our sample is much bigger (they have approximately 54k observation, while ours is approximately twice as large).

In Appendix Figure B.6, we provide evidence that the distribution of EPS to price ratios can be reasonably well approximated by a power law of tail coefficients approximately 2 in the left tail, and 3 in the right tail. We show in Figures B.8 and B.7 that the relationship between EPS and lagged EPS is highliy non-linear, and similar to what we obtained for sales growth. This suggests a similar quantitative analysis, using the same DGP, could be implemented on EPS forecasts. The choice of normalizing EPS by a price level does, however, make the algeabra somewhat awkward and, at times, slightly inconsistent. Our focus on sales growth allows for a much more straightforward analysis.

## 5.2 Asset pricing implications

A natural implication of our result is that returns should also be a non-linear function of past returns. We start by a few equations to clarify what we mean, and then provide evidence that, indeed, on U.S. stock returns, future returns are a non-linear function of past returns.

Figure 14: CG regression coefficient by quantile




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

$$\frac{EPS_{it+1} - F_t EPS_{it+1}}{P_{it-1}} = \alpha + \beta \frac{F_{t-1} EPS_{it+1} - F_{t-1} EPS_{it+1}}{P_{it-1}} + \epsilon_{it+1}$$

where $EPS_{it}$ is the IBES Earning per share. 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 revisions. The point estimate of $\beta$ is the solid black line, while the dashed lines corresponds to the 95% confidence interval.

In the spirit of Bouchaud et al. (2019), assume stock prices are given by the formula:

$$P_t = \sum_{s \geq 1} \frac{F_t \pi_{t+s}}{(1+r)^s} \tag{12}$$

where $F_t pi_{t+s}$ is the forecasted cash-flows, and $r$ is a constant discount rate – we rule out time-varying expected returns here. With a slight abuse of notation, we label log profit growth $g_t$. Given our DGP and our forecasting rule, the Campbell-Shiller log-linearization of returns gives:

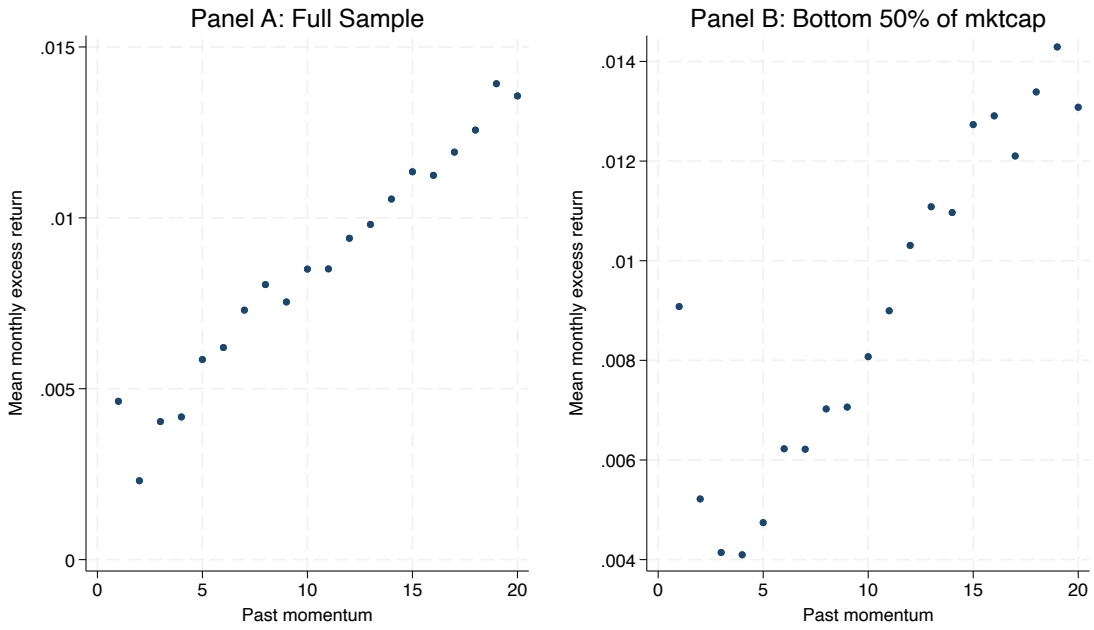
$$\begin{aligned} r_t &\approx r + \sum_{s=0}^{\infty} \rho^s \left( F_t g_{t+s} - F_{t-1} g_{t+s} \right) \\ &\approx r + \left( 1 + \frac{\sigma^2_{g^*}}{\sigma^2} \frac{\rho^2}{1-\rho^2} \right) g_t - \frac{\sigma^2_{g^*}}{\sigma^2} \frac{\rho}{1-\rho^2} g_{t-1} \end{aligned}$$

which shows that the dynamics of returns is dominated by the dynamics of $g_t$ (though $g_{t-1}$ also plays a more modest role). We thus expect $E(r_{t+1}|r_t)$ to be non-linear as $g_t$ is.

We now provide a small test of this. Our data is the CRSP universe of monthly returns, merged with factor returns from the Ken French library. We use the entire available period, from July 1926 to december 2023. As is common in empiricall asset pricing, we restrict ourselves to share codes 10 and 11, and to AMEX, NASDAQ and NYSE. For each stock $i$ and date $t$, we calculate the market capitalization as shares outstanding (shrout) times price (prc) at $t-12$. Our return variable corresponds to total returns (ret), and when firms delist, we apply the delisting return (DLRET). We then use, as the main sorting variable, the cumulative return between $t-12$ and $t-2$. We exclude the last month of return following the standard practice in the literature. Finally, we restrict ourselves to years after 1973, so that the number of firms considered is reasonably stable. None of these choices – which are standard – make a material difference to our results.

Figure 15 offers a visual test of our hypothesis, that returns should exhibit underreaction in the bulk of news, and overreaction in the tails. Every month, we sort stocks into 20 quantiles of momentum (our x-axis). We show the average return in the following month on the y-axis. The non-linearity is barely visible on the entire sample, but is stronger as we zoom on smaller firms, indicating that, for larger firms, arbitrage make our pricing equation (12) fail (marginal investors are more rational). There is also an asymmetry: past extreme losers win, but extreme winners do not lose.

Figure 15: Returns sorted on cross-sectional momentum




Note: In this figure, we sort stocks by the traditional momentum signal (past annual returns except the last month), and report, per vingitile of past return, the next month return. Portfolios are formed every month. Panel A uses the entire CRSP sample. Panel B is restricted to firms below the median CRSP capitalization (measured 12 months before portfolio formation) and Panel C to firms in the bottom 25% of market capitalization. Panel D further restrict the sample to years after 1990.

The analysis in Figure 15 suggests that the traditional momentum strategy would perform better if one removes extreme values of the signal. This is what we show in Appendix Figure B.9. There, we show the cumulative returns of a clipped, and an "unclipped", momentum strategy. Every month, we sort stocks by past momentum. We then compute a rank normalized to be between -0.5 and +0.5. The unclipped momentum portfolio uses these weights. The clipped portfolio also uses these weights, but removes the top and bottom deciles of momentum. The clipped portfolio does not outperform in absolute term, but exhibits significantly less volatility, so that its sharpe ratio is 1.5 (instead of 1.1 for the raw momentum strategy). This difference is robust to clipping choices, but becomes smaller as we include larger firms (on the entire sample, we get a sharpe of 1.8 for clipped vs 1.6 for unclipped).

## 5.3 Noise

In this Section, we show that noise cannot explain the error-revision pattern we have in our data. Assume here that forecasts are noisy, i.e. that they are given by the following relationship:

$$F_t g_{t+k} = \lambda EL(g_{t+k}|g_t) + (1-\lambda)E(g_{t+k}|g_t) + \eta_{t,k}$$

where $\eta_{t,k}$ is i.i.d. with variance $\sigma^2_\eta$, independent across horizons of forecasts, and independent of $g_t$.

The presence of noise can affect the CG coefficient since it creates a mechanical negative correlation between error (which contains a term $-\eta_{t,1}$) and the revision (which also contains the term $\eta_{t,1}$). This makes the coefficient negative, even when expectations formation have no other bias (Angeletos et al. (2020), de Silva and Thesmar (2024)).

To appraise the effect of noise, neutralize, in our model, the main effect of this paper by setting $\epsilon = 0$. In this case, the DGP is a simple AR1 and expectations become:

$$F_t g_{t+k} = \rho^k g_t + \eta_{t,k}$$

so that the error and the revision are:

$$ERR_t g_{t+1} = u_{t+1} - \eta_{t,1}$$

$$R_t g_{t+1} = \rho u_t + \eta_{t,1} - \eta_{t-1,2}$$

so that the CG regression coefficient simply writes as: $-\frac{\sigma_\eta^2}{\rho^2\sigma_u^2+2\sigma_\eta^2}$. Clearly, this regression coefficient is negative, as in our data for very large absolute revisions. But it does not have the ability to become positive in the bulk of the distribution of revisions. Alone, the presence of noise cannot explain our headline empirical result.

## 5.4 Adding a private signal

In this section, we show that the presence of a private signal about future growth does not change our structural interpretation. In the real world, forecasters do not just use past growth to predict future growth. Let us call the relevant information $s_t$, which we assume to be i.i.d. across periods. Now, the DGP writes:

$$g^*_{t+1} = \rho g^*_t + s_t + u_{t+1}$$
$$g_{t+1} = g^*_{t+1} + \epsilon_{t+1}$$

which is pretty generic. $s_t$ is a sufficient statistic for all the information available to the forecaster at date $t$, that comes on top of $g_t$. We further assume that $s_t$ is independent of $g^*_t$, $u_t$ and $\epsilon_t$. This model has the same properties as our main model of section 3. The rational and linear forecasts are given by:

$$E(g_{t+1}|g_t, s_t) = s_t + \rho E(g^*_t|g_t)$$
$$EL(g_{t+1}|g_t, s_t) = s_t + \rho\frac{\sigma^2_{g^*}}{\sigma^2_g} g_t$$

since $s_t$ is independent of $g_t$, the projection of these forecasts on $g_t$ is identical as in the rest of the paper. Further, this explains why subjective forecasts are not exactly linear, but subject to statistical noise (which correspond to actual information here).

# 6 Conclusion

In this paper we emphasize that boundedly rational agents, when faced with fat-tailed processes, will make predictable mistakes. In order to explore such processes, we need large samples. Our empirical research here leverages the international version of IBES which gives us a large panel of sales growth forecasts. Consistently with the firm demographics literature, we find that sales growth dynamics are well described by the sum of a short-run and and long-run processes. The long-run process is a simple Gaussian, AR1 process, but the short-run process has fat tails. As a result, a simple, linear filtering rule will not be optimal. This simple model of expectations formation matches a lot of the key features of the data.

# References


Afrouzi, Hassan, Spencer Yongwook Kwon, Augustin Landier, Yueran Ma, and David Thesmar, 2023, Overreaction and Working Memory, *Quarterly Journal of Economics* .

Angeletos, Marios, Zhen Huo, and Karthik Sastry, 2020, Imperfect macroeconomic expectations: Evidence and theory, *NBER Macro Annual* .

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, 2018, Over-reaction in Macroeconomic Expectations, Technical Report w24932, National Bureau of Economic Research, Cambridge, MA.

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.

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 Silva, Tim, and David Thesmar, 2024, Noise in expectations: Evidence from analyst forecasts, *Review of Financial Studies* .

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, *Revuew of Economic Studies* 74, 1059–1087.

Fuster, Andreas, Benjamin Hebert, and David Laibson, 2012, Natural Expectations, Macroeconomic Dynamics, and Asset Pricing, *NBER Macroeconomics Annual* 26, 1–48, Publisher: University of Chicago Press Chicago, IL.

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, 2018, 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.

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, 2024, Extreme categories and overreaction to news, Technical report.

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.

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* .

Wyart, Matthieu, and Jean-Philippe Bouchaud, 2003, Statistical models for company growth, *Physica A: Statistical Mechanics and its Applications* 326, 241–255.

# A Proofs

## A.1 Proof of proposition 1

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

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

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 $y$).

The second bullet point comes from differentiating the log marginal density two times:

$$\frac{E\left(g_{t+1}|g\right)}{dg} = -\rho\sigma_{g^*}^2 \left( \frac{h''(g)}{h(g)} - \left( \frac{h'(g)}{h(g)} \right)^2 \right)$$

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

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

which gives the formula in the proposition.

The condition is straightforward. Underreaction happens if:

$$\rho\left(1-\frac{\text{Var}\left(g_t^*|g\right)}{\sigma_{g^*}^2}\right) > \rho\frac{\sigma_{g^*}^2}{\sigma_g^2}$$
$$1-\frac{\sigma_{g^*}^2}{\sigma_g^2} > \frac{\text{Var}\left(g_t^*|g\right)}{\sigma_{g^*}^2}$$
$$\frac{\sigma_\epsilon^2}{\sigma_g^2} > \frac{\text{Var}\left(g_t^*|g\right)}{\sigma_{g^*}^2}$$

Now, assume that $\epsilon$ is Gaussian. Then, the projection of $g^*$ on $g$ is linear, so that: $g^* = \frac{\sigma_{g^*}^2}{\sigma_g^2}g + u$ where $u \perp g^*$. Then, Var $(g_t^*|g)$ = Var u $= \frac{\sigma_\epsilon^2\sigma_{g^*}^2}{\sigma_g^2}$. This shows that, with gaussian $\epsilon$, the equality holds. QED

## A.2 Proof of proposition 2

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 $g \to +\infty$

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

Then, the following Lemma holds:

**Lemma A.1.** *As $g \to +\infty$, then:*

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

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. QED

## A.3 Proof of proposition 3

Now, given the exponential assumption, and for given $x$, as $g \to +\infty$

$$f(g-x) \to e^{-\frac{1}{2\sigma_\epsilon^2}(g-x)^\theta} \approx e^{-\frac{1}{2\sigma_\epsilon^2}\left(g^\theta - \theta x g^{\theta-1}\right)}$$

Then, the following Lemma holds:

**Lemma A.2.** *As $g \to +\infty$, then:*

$$E(g^*|g) \to \frac{\int x e^{\frac{1}{2\sigma_\epsilon^2}\left(\theta x g^{\theta-1}\right)}\Phi(x)dx}{\int e^{\frac{1}{2\sigma_\epsilon^2}\left(\theta x g^{\theta-1}\right)}\Phi(x)dx}$$

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, first, it is easy to see that, if $\theta = 1$, the distribution is exponential, and $E(g^*|g)$ does not depend on $g$. It tends towards the constant:

$$E(g^*|g) \to \frac{\int x e^{\frac{1}{2\sigma_\epsilon^2}x} \Phi(x) dx}{\int e^{\frac{1}{2\sigma_\epsilon^2}x} \Phi(x) dx}$$

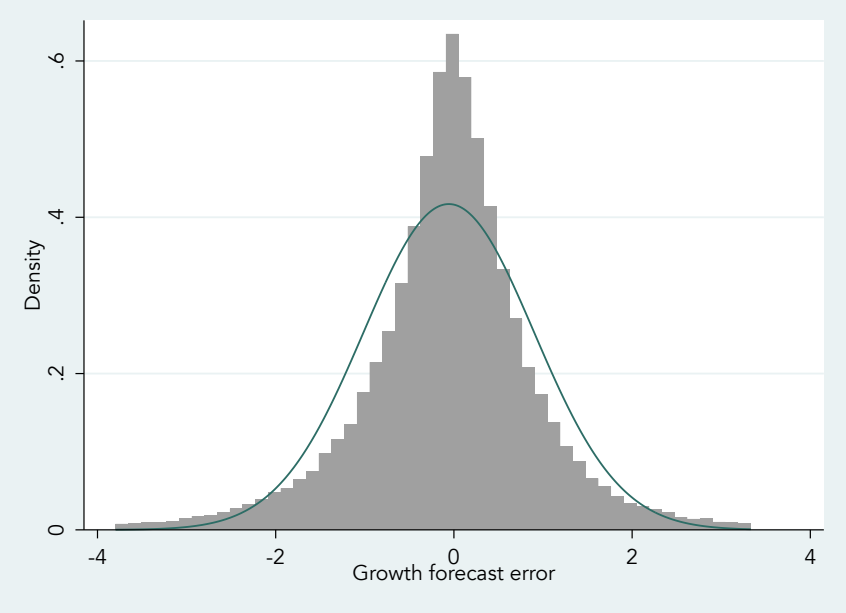
Now, if we assume that $\Phi$ is Gaussian, then the limit is easy to retrieve:

$$E(g^*|g) \to \frac{\sigma_g^2}{\sigma_\epsilon^2} \theta g^{\theta-1}$$

where we find again the familiar result that, if $\epsilon$ is Gaussian too ($\theta = 2$), then the expectation is linear. But as soon as tails are thicker than this, i.e. $\theta < 2$, $E(g^*|g)$ becomes concave, and even decreasing is $\theta < 1$. QED

# B Additional Figures

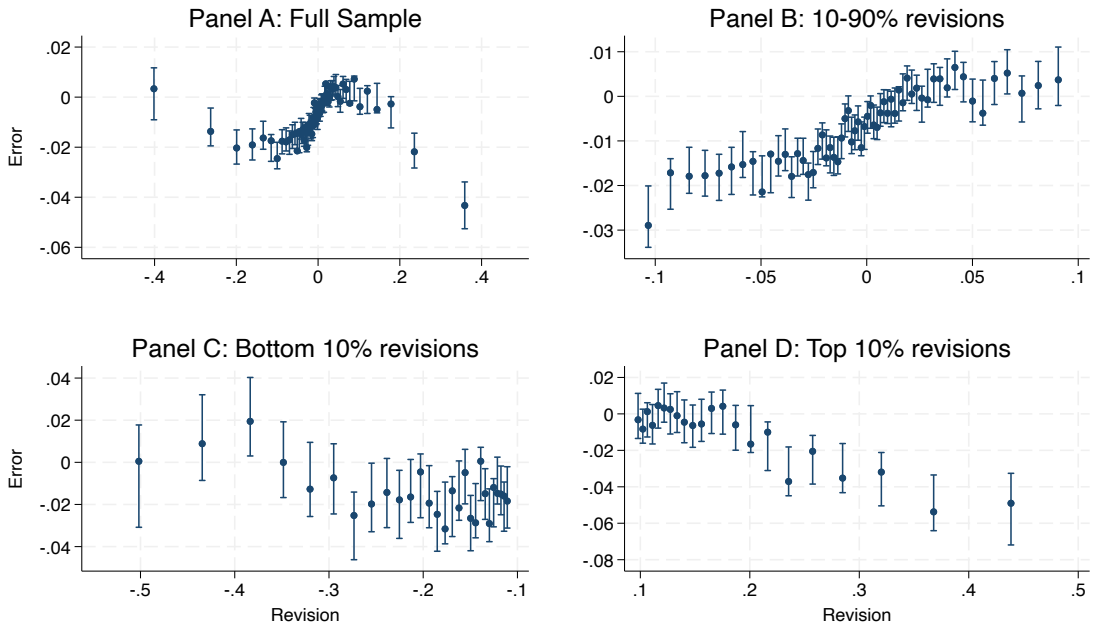
Figure B.1: Forecast error of log growth (normalized)




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

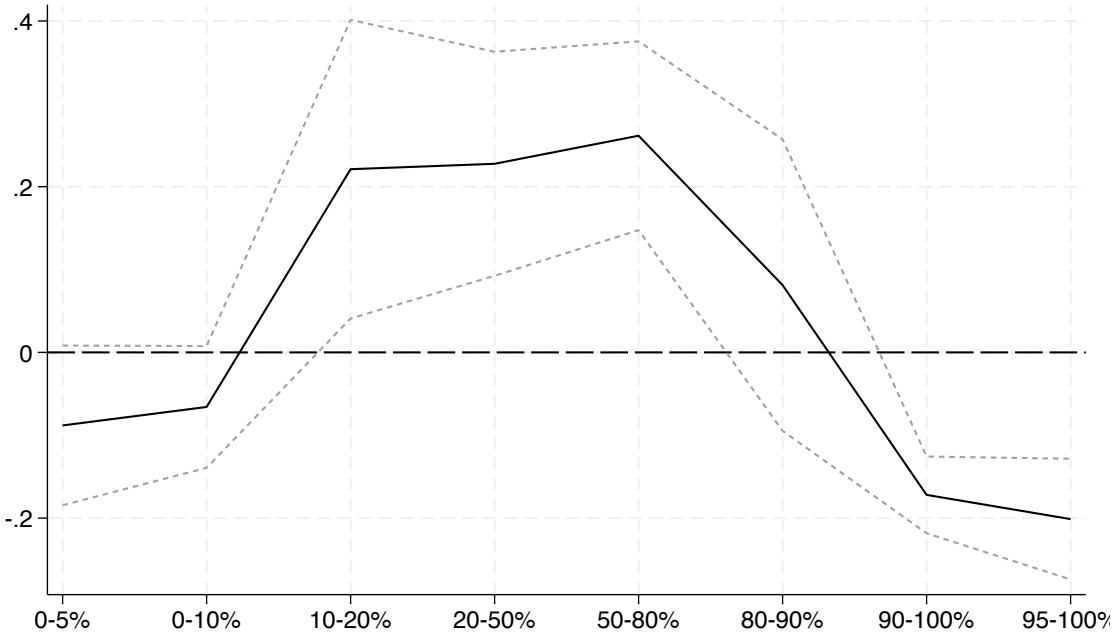
Figure B.2: The Error-revision relationship: Raw growth rate

Panel A: Unadjusted sales growth




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 quintiles of revisions 2,3 and 4 (i.e. percentiles 20-80). Panel C focuses on the first quantile of revisions (i.e. below the $20^{th}$ percentile) and panel D on the top quantile of revisions (i.e. above the $80^{th}$ percentile).

Figure B.3: CG regression coefficient by quantile: Raw growth rate

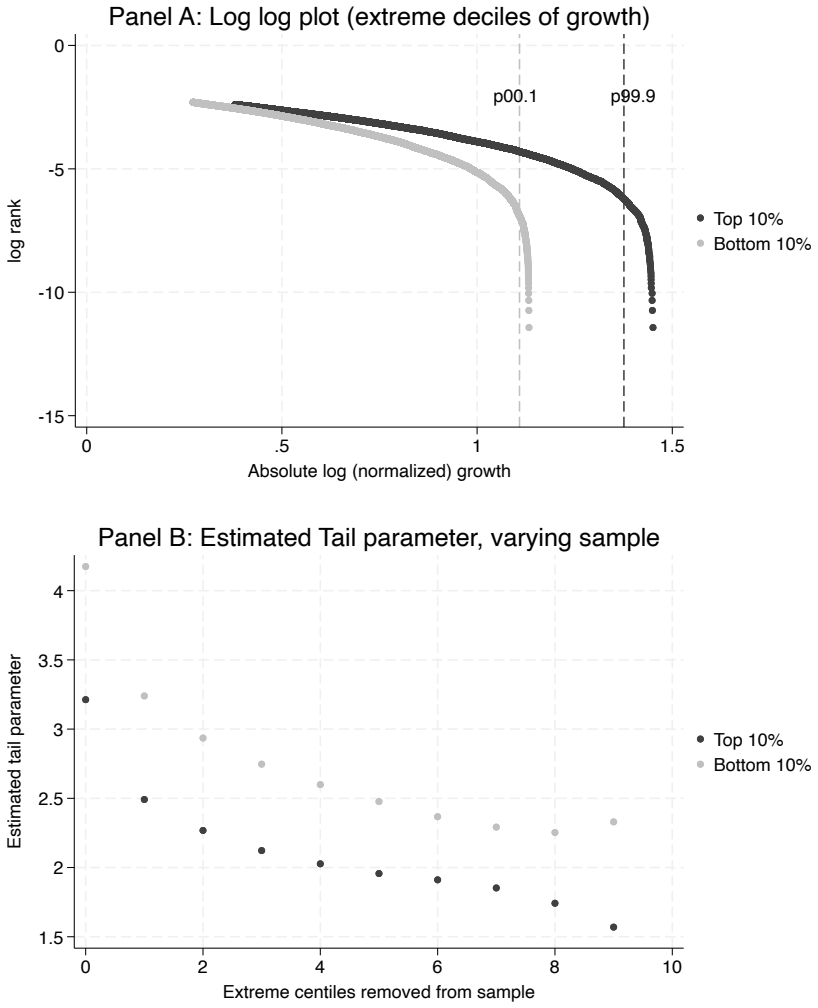



Note: In this Figure, we report the estimates of $\beta$ in the following regression:

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

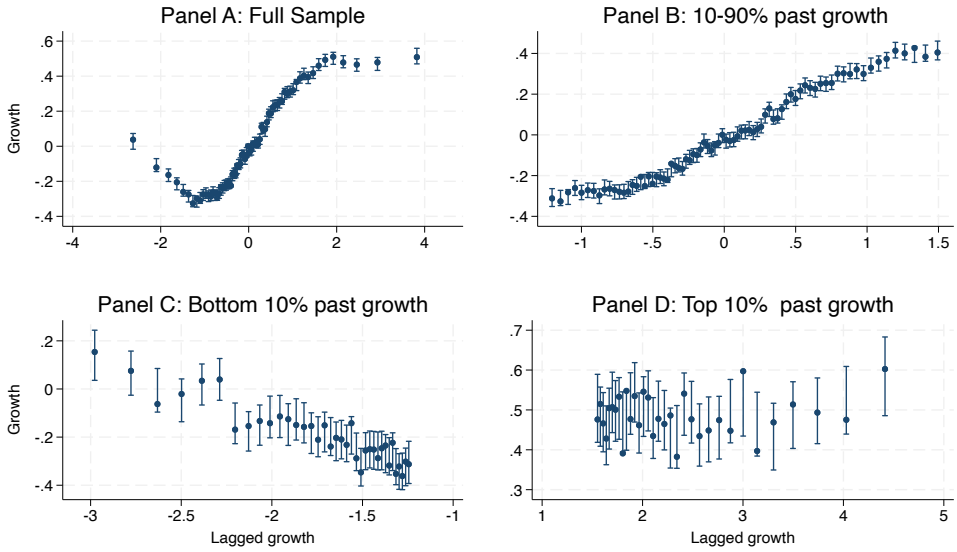
where $g_{it}$ is the normalized log sales growth rate, and error terms are assumed to be correlated within firm and within years. This regression is run on 8 different subsamples, whose ranges are described in the x-axis of this chart. These subsamples corresponds to the tails and the bulk of the distribution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

Figure B.4: Sales growth tail: COMPUSTAT sample



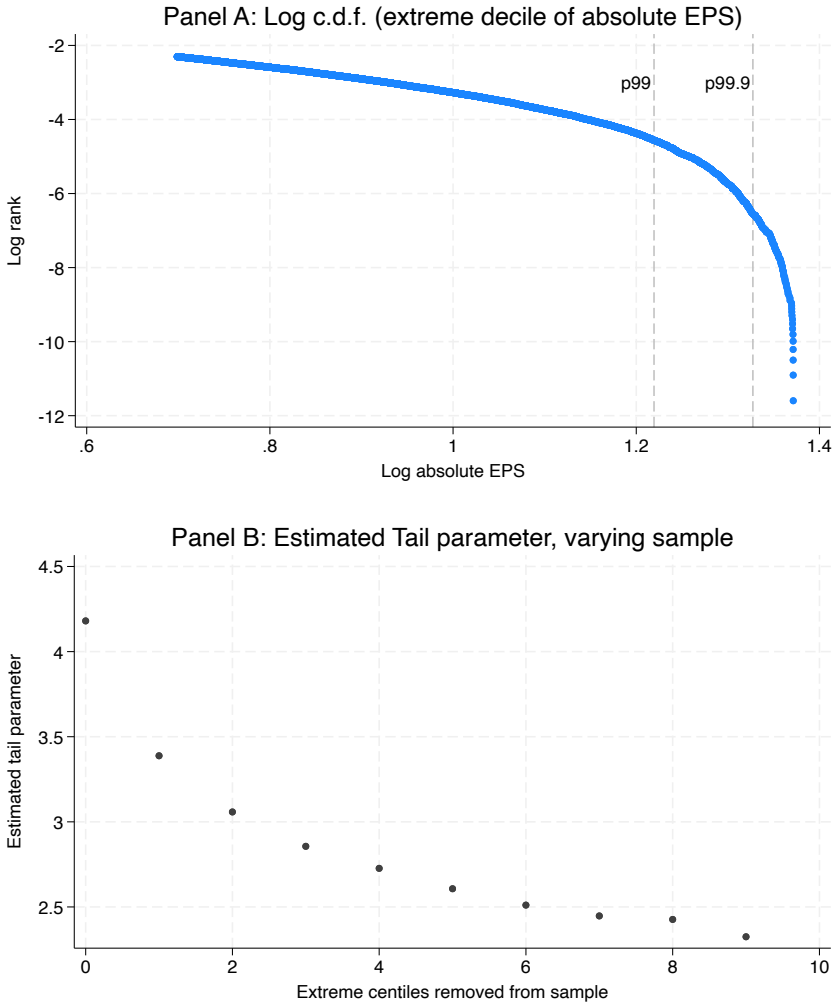

Note: We reproduce here exactly the same graph as Figure 5, except that we use a sample of COMPUSTAT firms. We restrict ourselves to all firms in COMPUSTAT (fundamentals annuals) between 2000 and 2023, and to all firms present at least 10 years over the period. Normalized log sales growth is computed exactly as in our IBES sample. In the end, our sample has nearly 110k observation, all U.S. firms.

Figure B.5: Future growth against current growth: COMPUSTAT sample




Note: We reproduce the exact same analysis as in Figure 7, except that we use a sample of COMPUSTAT firms. We restrict ourselves to all firms in COMPUSTAT (fundamentals annuals) between 2000 and 2023, and to all firms present at least 10 years over the period. Normalized log sales growth is computed exactly as in our IBES sample. In the end, our sample has nearly 110k observation, all U.S. firms.

Figure B.6: EPS/price tail: Pareto approximation

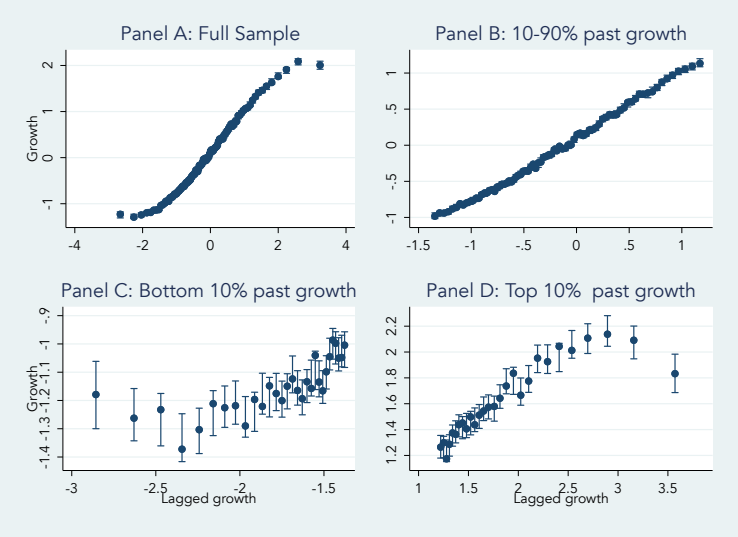



Note: We describe here the tails of the distribution of normalized EPS/price. In Panel A, we show the scatter plot of the rank (one minus c.d.f.) in the top and bottom deciles (c.d.f. being normalized to 1 at the extreme value of each decile to make things comparable). The x-axis corresponds to the log of absolute EPS to price for the right tail. We restrict ourselves to the top and bottom 10% of growth values. The dashed lines highlight percentiles 1 and 99. Panel B shows how the tail parameter varies as a function of how much of the tail one excludes from the regression. Let $H$ be the c.d.f. In the top decile, we run the regression:

$$\log(1 - H(\pi_{it})) = \alpha + \beta \log(\pi_{it}) + \epsilon_{it}$$

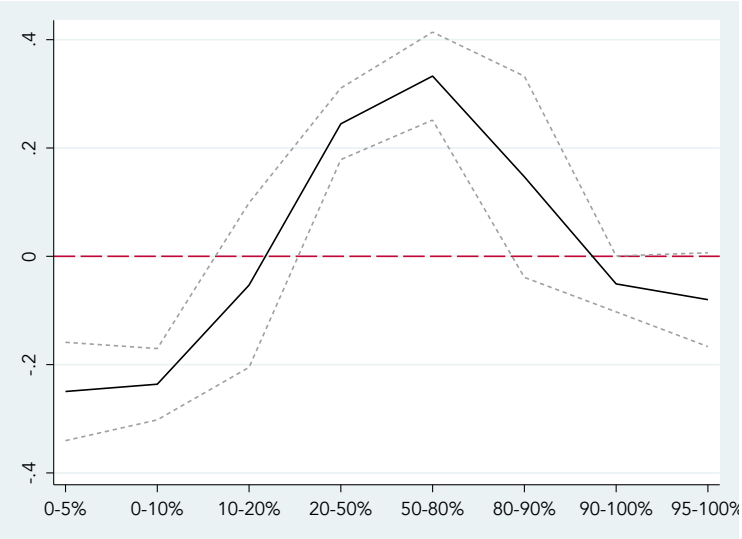
on ten different subsample, each one excluding the most extreme centiles of the distribution ($\pi_{it}$ denotes the EPS to price ratio). When $x = 0$, the subsample does not exclude any observation. When $x = 1$, the subsample only excludes the top centile, etc. On the y-axis, for the black dots, we represent the slope estimate $\hat{\beta}$. The gray dot emerge from the same methodology, except focusing on the left tail (the lowest decile). Reading: If one runs the above regression in the top decile but after excluding the top 5 percentiles, one obtains a tail parameter of 2. If one does so for the bottom decile, while excluding the 2 lowest percentiles, one obtains 3.3.

Figure B.7: Future EPS against current EPS in the data




Note: This figure reports 4 binned scatter plots of $\frac{EPS_{it+1}}{P_{it-2}}$ against $\frac{EPS_{it}}{P_{it-2}}$. Panel A shows the entire sample, Panel B conditions on the bulk of the distribution, while C and D focus on the tails. We normalize both values of EPS by the same price levels to filter out potential price dynamics.

Figure B.8: Regression of EPS on lagged EPS, various subsamples

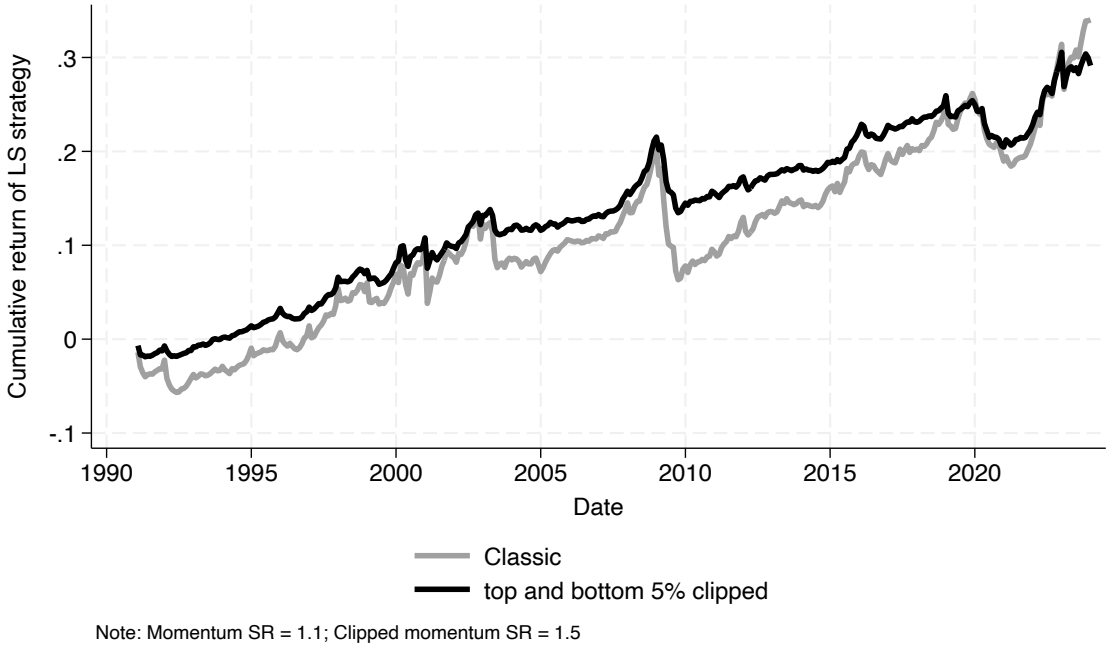



Note: We report here the empirical coefficients of growth on lagged growth, on various subsamples. We run the following specification:

$$\frac{EPS_{it+1}}{P_{it-2}} = \alpha + \beta \frac{EPS_{it}}{P_{it-2}} + \epsilon_{it+1}$$

We show the estimates of $\beta$ as well as confidence intervals. The subsamples are the same ones as in Figure 7.

Figure B.9: Momentum performance: Raw vs Clipped




Note: In this figure, we show the cumulative return of two version of the momentum strategy on all stocks whose market cap is below median. Each period, we rank stocks by past momentum, and transform this rank into a weight going from -0.5 to +0.5. The grey line shows the return of a long-short strategy using these weights every month. The black line shows the return of a long-short strategy using the very same weight, but excluding the top and bottom deciles of past momentum. We also report the sharpe ratios of these two strategies.