Abstract

We empirically analyze a large sample of firm sales growth expectations. We find that the relationship between forecast errors and lagged revision is non-linear. Forecasters underreact to typical (positive or negative) news about future sales, but overreact to very significant news. To account for this non-linearity, we propose a simple framework, where (1) sales growth dynamics have a fat-tailed high frequency component and (2) forecasters use a simple linear rule. This framework qualitatively fits several additional features of data on sales growth dynamics, forecast errors, and stock returns.
1 Introduction

Expectations formation is a core question in economics. In recent years, a strain of literature in macroeconomics and finance has been collecting empirical regularities using survey data on subjective forecasts. It finds that forecasts largely deviate from the full information model that predominates in economic modelling: forecast errors are biased and predictable using past errors and past revisions. Two types of explanations for this have been put forward. The first one is that the data-generating process (DGP) is simple and known to forecasters, but forecasting rules are irrational but linear, featuring for instance under-reaction (Bouchaud et al., 2019) or overreaction (Bordalo et al., 2019, 2020; Afrouzi et al., 2020). The second approach to explaining observed biases is the tenet that the data-generating process is too complex to be known by forecasters. Thus, they use a misspecified model calibrated on the data they observe. This may come from the fact that the DGP is hard to learn (for recent contributions along these lines see Kozlowski et al., 2020; Farmer et al., 2022), or alternatively from bounded rationality of the forecasters. They can only use simple forecasting rules (Fuster et al., 2010; Gabaix, 2018). In any case, forecast errors are predictable because forecasters use an imperfect model. In this paper, we find evidence consistent with the second view, i.e. that, facing complex (non-Gaussian) processes, forecasters use simple rules.

We use data on some 63,601 analyst forecasts of corporate revenue growth and their realizations. An advantage of focusing on revenue growth (instead of EPS as the literature typically does) is that revenue is always positive so that growth rate is always well defined. We first show that the relationship between forecast revisions and future forecast error is non-linear, a feature that is not reported in the existing literature. In some settings, revisions linearly and positively predict forecast errors, a feature commonly interpreted as evidence of under-reaction (Coibion and Gorodnichenko, 2015a). In others, revisions linearly and negatively predict forecast errors, which is considered as evidence of overreaction (Bordalo et al., 2019, 2020). In our sample, which is much larger than the existing studies, and which focuses on a rather new object, sales growth, we find evidence of both. 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). This non-linearity is robust. It holds in U.S. data and international data. It holds across most industry groups.

The remainder of the paper is dedicated to explaining this fact. Our framework is based on the simple assumption that forecasters use a linear rule to forecast sales growth, but that this rule is misspecified because the true DGP is more complex. Taking inspiration from the literature on firm size distribution (in particular, [Axtell], [2001]; [Bottazzi and Secchi], [2006]), we posit that sales growth distribution may be modelled by the sum of a low-frequency and high-frequency shock. The low frequency shock is Gaussian, while the high-frequency shock is non-Gaussian. It may have a very large (positive or negative) realizations. With such a model, the optimal forecast of future growth, conditional on current growth, is non-linear. A perfectly rational forecaster anticipates more reversion to the mean when realizations are extreme and more persistence when realizations are intermediate. We assume, however, that agents stick to a linear rule to make their forecasts. The fact that agents use a misspecified model may be grounded in bounded rationality (i.e., agents use a simple rule even if the process is complex, as in [Fuster et al.], [2010]) or the difficulty of learning about complex processes (shocks with multiple frequencies are hard to learn [Farmer et al.], [2022]; shocks with fat tails also [Kozlowski et al.], [2020]).

Combined, these two assumptions (linear forecasting rule but short-term non-Gaussian shocks) are enough to generate the non-linear relation between forecast errors and past revisions that observe empirically. The mechanism is intuitive. When revisions are large, the rational forecaster should anticipate mean reversion, but the linear forecaster won’t. She overreacts to big positive (or negative) news. When fitting her forecasting rule to the data, she does, however, take this overreaction into account, and optimally attenuates the sensitivity of her forecast to recent observations in the bulk of the distribution. As a result, she underreacts to news of lesser significance.

We then qualitatively test four additional predictions of the model. We start with two natural predictions of the data-generating process. The first such prediction is that the distribution of sales growths has fat tails, a fact that holds strongly in
the data (and previously shown by Bottazzi and Secchi, 2006). In particular, we check that this fact is not driven by an alternative model of firm dynamics, where firms have heterogeneous volatility, but Gaussian dynamics. In such a setting, large growth shocks could be generated by the subset of firms who are more volatile than average (Wyart and Bouchaud, 2003). We thus rescale sales growths by estimates of firm-level standard deviation and find that the resulting distribution still has very fat tails, suggesting that growth shocks occur within firms, not across firms.

The second prediction from our DGP is that, conditional on past growth, future growth should follow a S-shaped pattern as discussed above. We show that this holds in the data, whether we normalize sales growth by firm-level standard deviation or not.

The third prediction is on forecast errors. A natural prediction of our forecasting model is that the autocorrelation of forecast errors should have the same non-linear relation as the relation between errors and lagged revision. In our model, where the forecasting rule is linear, they are the same. Large past errors are equivalent to big shocks and therefore transient ones: This leads to overreaction, as in the error-revision relation. We find that this pattern holds in the data: forecast error are positively correlated for intermediate values and negatively for large absolute values.

Our fourth and last prediction is on stock returns. Assuming risk-neutral pricing and that equity cash-flows follow a dynamic similar to revenues, it is easy to show that our model predicts that the autocorrelation of returns should have a shape similar to the autocorrelation of forecast errors. For intermediate values of past returns, momentum should dominate, but for extreme values of returns, stock returns should mean revert. We find this pattern to hold in the data. Our findings line up with recent research from Schmidhuber (2021), who also finds evidence of momentum for “normal past returns” and mean-reversion for extreme values of returns. We conclude from this analysis that the risk-adjusted performance of momentum strategies would be considerably improved by excluding stocks whose past returns have been large in absolute value.

This paper contributes to the recent empirical literature on expectations for-
mation. 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 in firms’ growth (as Kozlowski et al., 2020 have done in a different setting and in their case with a focus on Bayesian learning).

In doing this we also connect the expectations formation literature with the empirical literature on firm dynamics, which has since Axtell (2001) emphasized the omnipresence of power laws in the distribution of firms sizes (see Gabaix 2009, for a survey of power laws in economics). That sales growths (rather than log sales) have fat tails is a less well-known fact, although it was first uncovered by Bottazzi and Secchi (2006).

Last, 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, 2012, Gabaix, 2018). Such models are correct on average, they are fitted on available data, but their misspecification gives rise to predictability in forecast errors.

Section 2 describes the data we use: publicly available data on analyst forecasts (IBES) and confidential data on international stock returns from CFM. Section 3 documents the main fact: future errors are a S-shaped function of past revision. Section 4.1 lays out the simple framework that we build in order to explain this novel pattern. Section 5 tests four additional predictions from this model. Section 6 concludes.
2 Data

2.1 Analyst Forecast data

This paper focuses on firm revenue (sales) forecasts made by analysts. 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 Earnings per Share forecasts have received greater attention in the literature, sales forecasts are, in fact, better populated in the data than EPS forecasts in recent years. Another advantage of revenue forecasts is that they are never negative, so that we can easily calculate sales growth. 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 (typically, future EPS normalized by current stock price).

For each firm $i$ and each year $t$, we denote sales by $R_{it}$, and $F_t R_{it+1}$ the forecast made in year $t$ for the future realization of $R_{it+1}$. We compute $F_t R_{it+1}$ as the consensus 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 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 $t - 1$. Finally, we retrieve the realization of $R_{it+1}$ from the IBES actual files, which is designed to recover the realization of the quantity actually forecast by analysts.

In this paper, we focus on log sales growth and forecast of log sales growth. We define $g_{it+1} = \log R_{it+1} - \log R_{it}$ the log-growth of this quantity. The growth forecast is defined as $F_t g_{it+1} = \log F_t R_{it+1} - \log R_{it}$ for the one-year ahead growth forecast, and $F_{t-1} g_{it+1} = \log F_{t-1} R_{it+1} - \log F_{t-1} R_{it}$ for the two-year ahead forecast of annual growth.

Finally, in the spirit of the expectations formation literature (Coibion and Gorodnichenko, 2015a; Bouchaud et al., 2019), we construct two empirical variables: the forecast error $ERR_{it+1} = g_{it+1} - F_t g_{it+1}$ and the forecast revision $R_t g_{it+1} = F_t g_{it+1} - F_{t-1} g_{it+1}$. These two variables will be the main focus of our analysis.
To ensure forecast quality and improve sample consistency when we examine returns, restrict our analysis of forecasts to firms that belong to one of the major global stock indexes.\footnote{The list of stock markets used consists of: AEX, AS5, CAC, DAX, HSC, HSI, IBE, IND, KOS, MID, NDX, NIF, NKY, OMX, SMI, SPT, RAY, SX5, TOP, TPX, TWY, UKX} Further, we restrict ourselves to firm-year observations for which both the forecast error $ERR_{it+1}$ and the revision $R_{tg_{it+1}}$ are available. We give more details about the number of observations and the start date in Table \[1\]

2.2 International Data on Stock Returns

In examining returns we restrict our sample to equities included in a major national index. We rely on proprietary return data purchased and maintained by CFM. The start of data availability differs by index and is shown in Table \[2\]. For all indexes data has been obtained through January, 2022. Each observation is a ticker-month, and returns are log returns.

3 Motivating Facts

In this section we describe new evidence on expectations formation and document a strong non-linearity in the link between forecast error and revisions.

Since (Coibion and Gorodnichenko, 2015b) many papers in the expectations formation literature estimate the following linear relationship between forecast errors and revision:

$$ERR_{it+1} = \alpha + \beta R_{tg_{it+1}} + \epsilon_{it+1}$$ (1)

which is intuitive to interpret. Full information rationality predicts $\beta = 0$ for consensus forecasts (Coibion and Gorodnichenko, 2015b). Plain rationality predicts $\beta = 0$ for individual forecasts. $\beta > 0$ is typically interpreted as evidence of information frictions (Coibion and Gorodnichenko, 2015a), or, if run at the forecaster level, plain under-reaction (Bouchaud et al., 2019) study EPS forecasts; Ma et al., 2020 study the revenue forecasts of managers). In contrast, $\beta < 0$ is interpreted as
| Index | Total | 2000 | 2005 | 2010 | 2015 | 2020 |
|-------|-------|------|------|------|------|------|
| AEX   | 533   | 0    | 32   | 19   | 30   | 28   |
| AS5   | 3228  | 48   | 122  | 167  | 196  | 161  |
| CAC   | 921   | 0    | 40   | 45   | 49   | 48   |
| DAX   | 680   | 0    | 29   | 38   | 37   | 35   |
| HSC   | 972   | 7    | 24   | 41   | 75   | 74   |
| HSI   | 572   | 15   | 24   | 28   | 29   | 29   |
| IBE   | 715   | 0    | 35   | 37   | 40   | 35   |
| IND   | 746   | 1    | 38   | 38   | 41   | 39   |
| KOS   | 1540  | 34   | 29   | 30   | 101  | 124  |
| MID   | 13016 | 10   | 586  | 818  | 782  | 646  |
| NDX   | 1174  | 1    | 47   | 67   | 72   | 61   |
| NIF   | 1037  | 13   | 24   | 47   | 66   | 64   |
| NKY   | 4959  | 207  | 206  | 226  | 224  | 233  |
| OMX   | 605   | 19   | 26   | 31   | 32   | 31   |
| RAY   | 15923 | 4    | 525  | 995  | 1057 | 881  |
| SMI   | 479   | 8    | 21   | 21   | 27   | 23   |
| SPT   | 998   | 0    | 34   | 56   | 67   | 63   |
| SX5   | 215   | 0    | 10   | 11   | 13   | 11   |
| TOP   | 493   | 0    | 18   | 14   | 40   | 32   |
| TPX   | 10836 | 372  | 486  | 531  | 504  | 574  |
| TWY   | 1314  | 13   | 40   | 71   | 80   | 74   |
| UKX   | 2645  | 82   | 110  | 131  | 142  | 119  |
evidence of overreaction (Bordalo et al., 2019; study long-term EPS growth forecasts; Bordalo et al., 2018 focus on macroeconomic expectations). All these papers restrict their analyses to linear functional forms, as in equation (1).

Figure 1: Revenue Forecast Error as a Function of Past Revision

Note: In this figure we use international sample of firm revenue expectations to report the binned scatter plot of future log forecast errors $g_{t+1} - F_t g_{t+1}$ as a function of past revision $F_t g_{t+1} - F_{t-1} g_{t+1}$. The blue line is a local polynomial approximation, centered in the middle of each centile.

In this section we show that for revenue growth forecasts this relationship is actually non-linear. In Figure 1 we represent the relationship in a non-parametric way through a binned scatter plot where the x-axis are the revisions $R_t g_{t+1}$ and the y-axis is the average forecast error $FE_{it+1}$. Each black dot corresponds to a centile of the distribution of revisions, with the x coordinate being the average revision in this centile and the y coordinate being the average forecast error. The grey shaded area shows a bootstrapped 95% confidence interval. The blue line shows the predicted error from a local polynomial regression (or LOESS) model estimated at the center of each percentile of lagged revision. The kernel for this local regression model is Gaussian with the bandwidth set equal to the average of the distances between the centers of the 1st and 2nd, and the 99th and 100th, percentiles.
For revisions of relatively small to moderate magnitude we find that errors are increasing in revision. Thus, forecasters are under-reacting in response to moderately-sized news shocks. This consistent with evidence from [Bouchaud et al. (2019)] on EPS forecasts in United states. [Ma et al. (2020)] find similar evidence on revenue forecasts from managers’ expectations in the U.S. (using guidance data) and Italy (using a survey from the Bank of Italy). Their samples are, however, much smaller than ours (a few 10,000 observations at most), which precludes observing the tails of the distribution of revisions.

The key difference is in the tails of the distribution of revisions, for which this relationship is reversed. In the face of exceptionally bad news, forecasters are over-reacting: a larger, positive revision leads to more negative surprises. A similar non-linearity is marginally observable in U.S. EPS forecasts in [Bouchaud et al. (2019)], but the S shape in not complete there.

We then explore the robustness of this relationship across sub-categories in Figure 2. This figure has two panels: one that splits between U.S. and non-U.S. firms (Panel A) and one that splits the sample into industries (Panel B). In both cases we only show the prediction from flexible polynomial approximation. In both subcategories the S-shaped function emerges. In particular, it is visible both in US and non-US firms, although more pronounced among U.S. firms.

Overall, the evidence on log 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. In what follows, we set up a simple model that features extreme (i.e. non-Gaussian) shocks and linear expectations formation in order to captures the above non-linearity.

4 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 1.
Figure 2: The Error-Revision relationship: Sample Splits

Panel A: US v RoW  Panel B: By Industry

Note: In this figure we use our international sample of firm revenue expectations to report the binned scatter plot of future log forecast errors $g_{t+1} - F_t g_{t+1}$ as a function of past revision $F_t g_{t+1} - F_{t-1} g_{t+1}$. The blue line is a local polynomial approximation, centered in the middle of each centile. In Panel A, we split the sample between U.S. and non U.S. observations. In Panel B, we split the sample into 1 digit GICS industries.

4.1 Modeling Sales Growth

The first piece of the model is the data-generating process. We will omit the firm index $i$ for clarity’s sake and assume that log sales growth, $g_{t+1}$, evolves according to:

\begin{align}
  g_{t+1} &= g_{t+1} + \epsilon_{t+1} \\
  g_{t+1} &= \bar{g} + \phi(g_t - \bar{g}) + u_{t+1}
\end{align}

where $g_{t+1}$ is the unobservable latent state that follows the classic linear-Gaussian AR1 dynamics. The key difference with most existing models of expectations formation is that $\epsilon_{t+1}$ follows a probability distribution density that has heavy tails. Because it fits the data quite well (as we document below), we assume that $\epsilon_{t+1}$ follows a Student’s t distribution with $\nu$ degrees of freedom. Thus:

\begin{align}
  \epsilon_{t+1} &\sim \text{Student-t}(0, 1, \nu) \\
  u_{t+1} &\sim \text{Normal}(0, 1)
\end{align}
Note that, although we analyze the cross-section of firms, we assume a single process for all firms. In this paper, we do not explore the consequences of firm heterogeneity for forecasting biases. For instance, such biases could arise from forecasters using one single forecasting model for firms following different processes. We believe such an avenue is interesting, but beyond the scope of this paper, which focuses on one single deviation from the classical model, i.e. that temporary shocks have fat tails. In order to bring the data closer to the model however, we will conduct all of our analysis with “normalized growth” data, thereby ensuring that all firms have the same growth volatility. We discuss this adjustment extensively in Section 5.

In our simple model the conditional expectation \( E(g_{it+1}|g_{it}) \) is non-linear. We show this numerically in Figure 3. For different values of \( \nu \), we numerically simulate the process and compute the conditional expectation \( E(g_{it+1}|g_{it}) \) on simulated data. As shown in Figure 3 this relationship is indeed quite linear in the body of the distribution, but experiences “reversals” in the tails. While not visible in Figure 3, all finite values of \( \nu \) leads to such reversals in the tail, but as \( \nu \) gets larger (and \( \epsilon \) is closer to being Gaussian) they get pushed out farther into the tails and are very sharp and localized.

The economic intuition is simple. Since the underlying state variable is Gaussian, extreme negative or positive realizations are more likely to come from the transitory process \( \epsilon \) than the persistent one \( g \), since it features more extreme shocks. As a result, a large sales growth realization today is unlikely to translate into future large sales growth tomorrow: This suggests the presence of “reversals” in the tails, as we see in Figure 3.

The above process has several predictions about the distribution of growth rates, one of them being that the cross-sectional distribution of growth rates should have fat tails. We will explore these predictions in Section 5.

4.2 Expectations Formation

The second building block of the model is the formation of expectations. Our core assumption is that forecasters fail to perceive the non-linearity of true expectations
Figure 3: Actual vs. lagged actual in model simulation.

Note: We simulate the model over \( T \) periods with \( u \) following a normal distribution \( N(0, 1) \), and \( \epsilon \) following Student-\( t(0, 1, \nu) \). We then show local polynomial regressions of \( g_t \) on \( g_{t-1} \) estimated at the center of each percentile of lagged realization \( g_{t-1} \). We explore values of \( \nu \) from 1.6 (fat tailed) to \( \infty \) (Gaussian).
\[ E(g_{t+1}|g_t) \] and use a linear rule. This assumption is based on the idea that economic agents use simplified, “sparse”, model of reality to formulate expectations (Fuster et al., 2012, Gabaix, 2018). Agents assume \( g_{it} \) follows a linear AR(p) 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 the AR(p) model. Hence, we assume that the forecaster believes growth follows the following AR(p) model:

\[
g_{t+1} = g + \sum_{s=0}^{p-1} \beta_k (g_{t-k} - \bar{g}) + u_{t+1}
\]

We denote the subjective expectation operator by \( F_t g_{t+k} \equiv EL(g_{t+k}|g_t) \).

We assume that this prior is dogmatic. The forecaster is willing to re-estimate the model’s parameters as new data comes in, but does not explore models outside of the AR(p) set-up. As a result, the agent does not really formulate rational expectations since she does not estimate the right DGP, as in Fuster et al. (2010). One foundation for such dogmatism is that learning is extremely slow in non-Gaussian, non-linear environments, so that it takes many periods to modify the prior about the model (in recent literature, see Kozlowski et al. (2020) and Farmer et al. (2022)).

Thus the agent estimates the parameters of the misspecified model using OLS on expanding windows – using all information until date \( t \). Let \( \hat{g} \) and \( \hat{\beta}_k \) be these estimates. The one-period ahead forecast and the revision are given by:

\[
F_t g_{t+1} = \hat{g} + \sum_{s=0}^{p-1} \hat{\beta}_k (g_{t-k} - \hat{g})
\]

### 4.3 Predictions of the Model: Errors on Revisions

We now check that our model indeed generates the non-linear relation between revenue forecast errors and revenue forecast revisions shown in Figure 1.

In Figure 4 we report results from simulations, assuming that forecasts are based on a fitted AR(2) model. We vary the thickness of the tail of the temporary shock \( \epsilon \), which is governed by \( \nu \). \( \nu = +\infty \) corresponds to a normal distribution,
while $\nu = 1.6$ is the thickest tail possible.

Figure 4: Error as a function of lagged revision

Note: We simulate the model over $T$ periods with $u$ following a normal distribution $N(0, 1)$, and $\epsilon$ following Student-t$(0, 1, \nu)$. We then show local polynomial regressions of error $ERR_{t+1} = g_{t+1} - F_t g_{t+1}$ on revision $R_t g_{t+1} = F_t g_{t+1} - F_{t-1} g_{t+1}$ estimated at the center of each percentile of revision. We explore values of $\nu$ from 1.6 (fat tailed) to $\infty$ (Gaussian). Forecasters are assumed to employ an AR(3) model when predicting dividend growth rates.

Figure 4 shows that, as long as the temporary shock has sufficiently fat tails, the linear expectations model generates predictable forecast error that display a non-linear pattern similar to Figure 1. This is quite intuitive. As shown previously, the true conditional expectation is non-linear (see Figure 3). When the past realization of revenue growth is large, it is likely that it was driven by the temporary fat-tailed process. As a result, the rational forecaster would expect some mean-reversion, but the linear forecaster does not. This creates overreaction to large shocks. In contrast, when past realizations are moderate, there is underreaction. This comes from the fact that the linear forecaster is on average rational: She fits a linear relation on the S-shaped data of Figure 3. The slope of forecasts for smaller realizations incorporates some of the overreaction in the tails.

To gain further insights in Figure 3 we vary the fatness of the tail $\nu$. The less
thick-tailed the innovation process, the less predictable errors are. When \( \nu = +\infty \),
the temporary shock \( \epsilon \) is Gaussian and forecast errors are very close to zero for all
lagged realizations (the black dots line up on the x-axis). This is because in this case
the linear AR2 forecasting rule is nearly rational. Indeed, in this case, the rational
expectation is a Kalman filter:

\[
K_t g_{t+h} = g + \phi^h G \sum_{s=0}^{+\infty} (1 - G)^s \left( g_{t-s} - g \right)
\]

where \( G \) is the Kalman “gain”. The AR2 process is close enough to the above
equation that forecast errors are nearly zero in our simulations.

The bottom line of this analysis is that the non-linear structure of expectations
errors can easily arise when forecasters use linear models when the data generating
process has temporary shocks that have fat tails. Indeed in this case, the optimal
forecasting rule is non-linear, even though the process is itself linear.

4.4 An Additional Prediction: Error on Lagged Error

The empirical expectations literature also investigates a different moment: The au-
tocorrelation of expectation errors (for instance, Ma et al. (2020) and Farmer et al.
(2022) among many others).

In our model the autocorrelation of errors is equivalent to the error-revision co-
efficient. This happens because revisions are directly proportional to current fore-
cast errors:

\[
F_t g_{t+1} - F_{t-1} g_{t+1} \equiv R_t g_{t+1} = \hat{\beta}_0 \cdot (g_t - F_{t-1} g_t)
\]

which means that a positive surprise translates into a positive revision about future
growth. The fact that the prior is linear makes this relationship linear, whatever the
number of lags \( p \).

As a result, we expect the non-linear relation between errors and lagged revision
of Figure 1 to also hold between error and lagged errors. We test this additional
prediction in Section 5.
4.5 Model Prediction on Returns: Building Intuition

We also derive predictions on stock returns. Our simple model, as we will see, predicts that momentum occurs for intermediate returns and mean-reversion occurs for extreme returns.

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

\[
P_t = \sum_{s \geq 1} \frac{F_t D_{t+s}}{(1+r)^s}
\]

(5)

where \( F_t D_{t+s} \) is based on the forecasting rule described above in Section 4.2. Hence, the stock is priced by investors who form expectations based on a linear AR2 fitted on past realizations, while we assume dividends to follow the process described in Section 4.1. We also assume for simplicity that investors are risk-neutral, so that the discount rate is fixed at \( r \).

In this very simple asset pricing model we expect returns to be a non-linear function of past returns, similar to what we documented for the error-revision relation in Figure 1. Before we discuss simulation results and economic intuition, it is worth showing the algebra. The standard first order Campbell-Shiller approximation writes as:

\[
r_{t+1} - r_t r_{t+1} \approx (F_{t+1} - F_t) \sum_{s=0}^{\infty} \rho^s g_{t+1+s} - (F_{t+1} - F_t) \sum_{s=1}^{\infty} \rho^s r_{t+1+s}
\]

where we denote log dividend growth as \( g \) with a slight abuse of notation (\( g \) stands for log revenue growth in the rest of the paper). Equation (5) assumes constant expected returns \( F_t r_{t+k} = r \) (investors may be biased but are risk neutral), so that the CS decomposition simplifies into:
\[ r_{t+1} - r = (F_{t+1} - F_t) \sum_{s=0}^{\infty} \rho^s g_{t+1+s} \]

\[ = g_{t+1} - F_t g_{t+1} + \rho (F_{t+1} - F_t) \sum_{s=0}^{\infty} \rho^s g_{t+2+s} \]

It then remains to describe the infinite sum of discounted dividend growth. In this paper, we assume that forecasters (mistakenly) estimate dividend growth as an AR(p) process:

\[ g_t - \underline{g} = \sum_{s=1}^{p} \beta_s (g_{t-s} - g) + \epsilon_t \]

We can then stack the estimated AR(p) coefficients \( \beta_1, \ldots, \beta_p \) into “companion” form:

\[
\begin{bmatrix}
  g_t - \underline{g} \\
  g_{t-1} - \underline{g} \\
  \vdots \\
  g_{t-p} - \underline{g}
\end{bmatrix}
= \begin{bmatrix}
  \beta_1 & \cdots & \beta_{p-1} & \beta_p \\
  1 & \cdots & 0 & 0 \\
  \vdots & \ddots & \vdots & \vdots \\
  0 & \cdots & 1 & 0
\end{bmatrix}
\begin{bmatrix}
  g_{t-1} - \underline{g} \\
  g_{t-2} - \underline{g} \\
  \vdots \\
  g_{t-p-1} - \underline{g}
\end{bmatrix}
+ \begin{bmatrix}
  \epsilon_t \\
  0 \\
  \vdots \\
  0
\end{bmatrix}
\]

or more compactly:

\[ G_t = BG_{t-1} + \epsilon_t \]

Time \( t \) forecasts for \( g_{t+s} - \underline{g} \) are then given by:

\[ F_t (g_{t+s} - \underline{g}) = e_1' B^s G_t \]

where \( e_1 \) is a “selector” vector picking out the first element of the following vector.

The infinite sum of discounted forecast dividend growth is given by:
\[ F_t \sum_{s=0}^{\infty} \rho^s (g_{t+1+s} - g) = \sum_{s=0}^{\infty} \rho^s e'_1 B^{s+1} G_t = e'_1 B (I - B)^{-1} G_t \]

We then plug this formula into the CS decomposition and obtain:

\[ r_{t+1} - r = e'_1 (G_{t+1} - B G_t) + \rho e'_1 B (I - B)^{-1} (G_{t+1} - B G_t) \]
\[ = e'_1 (I + \rho B (I - B)^{-1}) (G_{t+1} - B G_t) \]
\[ = ERR_{t+1} G_{t+1} \]

The above expression shows that under the AR(p) assumption, returns are a linear function of past forecast errors (\( ERR_{t+1} G_{t+1} \) is the vector of past \( p \) forecast error). In this simple asset-pricing model, returns are only predictable if dividend growth forecast errors are predictable. Under rational expectations (i.e. if the true DGP for dividends is an AR(p)), they are not. But if dividends are driven by a thick-tailed state variable, the true DGP is far from AR(p) as we have documented. Thus, expected forecast errors are non-linear functions of past errors, and the same should hold for returns and past returns. So our model predicts that returns should be a non-linear function of past returns, in other words, momentum should only be present for intermediate values of past returns.

4.6 Model Prediction on Returns: Simulations

In order to check that this prediction also holds without CS approximation, we proceed to simulate our model. On simulated data, we build returns as \( R_{t+1} = (P_{t+1} + D_t - P_t)/P_t \). We then plot average future returns by bins of past returns in Figure 5. In this very simple asset pricing model, returns are predictable as soon as the dividend process.

The intuition is the same one as before. Very high past returns likely emerge from surprises due to a large thick-tailed temporary shock. Linear forecasters over-
Figure 5: Momentum for Intermediate Past Returns; Reversal in the Tails

Note: We simulate the model over many periods with \( u \) following a normal distribution \( N(0, 1) \), and \( \epsilon \) following Student-t\((0, 1, \nu)\). We then show local polynomial regressions of \( R_{t+1} \) on \( R_t \) estimated at the center of each percentile of lagged realization \( R_{tg_{t+1}} \). We explore values of \( \nu \) from 1.6 (fat tailed) to \( \infty \) (Gaussian). Forecasters are assumed to employee an AR(3) model when predicting dividend growth rates.
react: large past return events are likely to be situations where dividend realization a one-off boon. As a result, linear forecasters overestimate the level of future dividends: The stock price rises too much and future returns are lower. Intermediate past returns, however, are likely generated by standard shocks. There, the linear forecaster under-reacts to small dividend news and the price does not respond enough to these news. Future returns are thus positively correlated with past returns for “smaller” absolute values of returns. Note that when $\nu = +\infty$, temporary shocks are Gaussian and, as discussed previously, linear expectations are quasi optimal – a true Kalman filter would be perfectly rational – and past returns do not predict future returns.

5 Testing the Model’s Predictions

5.1 Predictions of Growth Rate Dynamics

In this section we discuss two key predictions of our data-generating process (2)-(3). The first one is that the distribution of firms revenue growths should have fat tails. The second one is that the conditional expectation of $g_{t+1}$ on $g_t$ should be non-linear as in Figure 3.

First, a key prediction of our DGP (2)-(3) is that the distribution of firm size growth has fat tails. 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)]. Less well-known is also the fact that the distribution of firm growth rates has fat tails. [Bottazzi and Secchi (2006)] show that the distribution of Compustat firms follows a Laplace distribution. Here, we provide similar evidence from our sample.

In Figure 6 we show the QQ plot of the sales growth distribution in our sample, along some other textbook distributions. This QQ plot focuses on observations above the 90th percentile distribution of $|g_{it}|$, the absolute value of sales growth (so both negative and positive shocks). Each point of this chart corresponds to one quantile of the data distribution. For a given quantile $q$ and a c.d.f. $F$, the y coordinate of the point is the average value of absolute growth at quantile $q$ of
the data distribution. Since we focus on the top 10% of absolute sales growth, this
number is positive (so the y axis does not start at zero). The x coordinate of that
point is the average value of the same quantile of the chosen distribution $F$, or
$F^{-1}(q)$. The closer is $F$ to the data distribution, the more the chart will look like a
45 degree line (the black line on the Figure).

Figure 6: Tail of log sales growth distribution: Fit of various distributions

Note: This is a Q-Q plot of log sales growth vs. some textbook distributions (Laplace, Normal and
Student). This plot shows the tail above the 90th-percentile of the distribution of $|g_{it}|$. On the x-axis,
we report the value of the quantile ($F^{-1}(q)$) of the comparison distribution. On the y-axis, we report
the value of the same quantile of the data distribution ($F^{-1}_{\text{data}}(q)$). The data distribution is normalized
so that its variance is one. The comparison distribution also have unit variance. By design, the
“data” line is the 45 degree line.

Looking at Figure 6 it is clear that the distribution of growth rates is very dif-
fert from normal in the tail. The green line increases faster than the 45 degree
line meaning that the sales growth distribution has much heavier tails than normal
distribution. The best fit is obtained by fitting a Student distribution.

Our model crucially assumes that this distribution comes from temporary shocks
occurring within firms. Wyart and Bouchaud (2003) suggest an alternative explana-
tion for such thick tails: sales growth has a normal distribution at the firm-level, but
that the standard deviation of this process varies across firms. In this case, extreme growth rates would typically occur among firms that have very volatile growth rates (for instance, smaller firms). This alternative interpretation does not explain our findings, but it is worthwhile to analyze its validity.

In order to do that we normalize growth rates by a measure of firm-level “volatility”. To do this we compute the mean absolute deviation of log sales for each firm. This measure of volatility has the advantage of being more immune to fat tails in the growth distribution (since variance may not exist in such cases). For firm $i$, we thus compute:

$$MAD_i = \frac{1}{T_i} \sum_{t=0}^{T_i} |g_{it} - \bar{X}_i|$$

where $T_i$ is the number of observations for the firm, and $\bar{X}_i$ is the average sales growth at the firm level.

In Figure 7 we show the QQ plot of the distribution of normalized $\frac{g_{it}}{MAD_i}$. If heavy tails are driven by firms with larger growth variance, this adjustment should significantly reduce the fat tails of the data. The QQ plot shows that the distribution is still strongly non-normal, though now the fit of the Laplace distribution is much better (consistent with Bottazzi and Secchi, 2006) and the one of the student distribution is nearly perfect.

From this analysis we draw the conclusion that sales growth distribution has heavy tails and that the conditional expectation $E_t(g_{t+1}|g_{it})$ is non-linear. We will postulate below a model of growth dynamics that fits these two facts, and show how it can explain the non-linear relation between revisions and errors.

From now on we report results using the above normalization by the mean absolute distance. This allows to account for fat-tails effects stemming from heterogeneous growth variance. Also quite importantly our model (2)-(3) assumes homoskedasticity, so it is important to rescale the data so that they have the same property.

A second key prediction of our Data is that the conditional expectation $E_t(g_{t+1}|g_{it})$
Figure 7: Tail of log sales growth distribution (Normalized)

Note: This is a Q-Q plot of normalized log sales growth vs. some textbook distributions (Laplace, Normal and Student). This plot shows the tail above the 90th-percentile of the distribution of $|g_{it} - \bar{g}_i|/MAD_i$. On the x-axis, we report the value of the quantile ($F^{-1}(q)$) of the comparison distribution. On the y-axis, we report the value of the same quantile of the data distribution ($F_{data}^{-1}(q)$). The data distribution is normalized so that its variance is one. The comparison distribution also have unit variance. By design, the “data” line is the 45 degree line.
should be non-linear, as shown in Figure 3. As mentioned previously the intuition is that large revision presumably come from large shocks to firm sales growth. Since in our model large shock are transitory, the rational forecaster should not expect that large shocks should persist going forward. Smaller shocks are, however, much more likely to stem from the permanent component of revenue, and therefore the rational forecaster should expect them to persist. We now check whether this relationship holds in the data.

Figure 8: Future Growth as a Function of Past Growth (Normalized)

Note: This Figure reports the binned scatter plot of future growth \( |g_{it+1} - \bar{g}_t|/\text{MAD}_t \) by value of past growth \( |g_{it} - \bar{g}_t|/\text{MAD}_t \). Each bin corresponds to a centile of the past growth distribution. The blue line is a local polynomial approximation centered around each one of these centiles.

In Figure 8 we construct a binned scatter plot of sales growth against lagged sales growth. To make sure all firms have the same growth volatility (as in the model), we normalize growth by our estimate of the firm level standard deviation \( \text{MAD}_t \). Each black dot on this figure represents a centile of the distribution of lagged log growth of sales. The x axis shows the average lagged growth and the y axis measures the average current growth. This chart shows that the relationship between current and lagged growth is far from being linear and looks like the S curve shown in Figure 1. For intermediate levels of growth (between 0 and 1 stan-
standard deviation), past growth translates into higher future growth, with a coefficient of about 0.3. The relationship does, however, become much flatter for high growth (with a slightly negative slope in the tail). For negative growth the slope becomes strongly negative. The lower the past growth, the higher the future growth will be, which is consistent with the idea of a rebound. Conditional on survival, very poor past performance predicts strong future growth, as in our model.

5.2 Predicting Forecast Errors

The model was designed to predict that forecast errors be a non-linear function of past revisions. Given that the model assumes that all firms have the same variance of shocks, it is natural to check that our main empirical results holds after rescaling by firm-level variance. Another reason why it is important to perform such a robustness check is discussed in the previous Section. Assume, following Wyart and Bouchaud (2003), that large news purely come from a separate group of firms (those with more volatile, but still Gaussian, shocks). Then, if forecasters use a firm-level linear forecasting rule, then their forecast errors should be close to unpredictable (to the extent that the AR(p) model they use mimics the optimal Kalman filter)\footnotemark

Figure \ref{fig:normalized} shows that normalized error and normalized revisions follow the same relationship as in our headline Figure \ref{fig:headline}. In this Figure, we simply show the binned scatter plot of future log forecast errors $\frac{g_{t+1} - F_{t+1}}{MAD_i}$ as a function of past revision $\frac{F_{t} - F_{t-1}}{MAD_i}$. This suggests that the non-linear relationship does not stem from firm volatility heterogeneity.

Another natural prediction of our forecasting model is that current and past forecast errors should follow a similar relationship. This comes from the fact that in our linear forecasting model, errors and revisions are proportional (equation \ref{eq:4}). Thus, it mechanically follows that if error and lag revisions are linked by the S-shaped curve of Figure \ref{fig:headline} then error and lagged error should follow the same relationship.

\footnotetext[2]{If, however, forecasters were to use a global forecasting rule (a single rule estimated on all firms), the non-linear shape may be predicted. Indeed, assume all shocks are Gaussian, but firms differ in the volatility of their temporary shock $\epsilon$. In this case, a unique forecasting rule would overestimate the persistence of large shocks. We do not explore this lead in this paper since we have documented in the previous Section that normalized growth is far from being normally distributed.}
Figure 9: Revenue Forecast Error as a Function of Past Revision (Normalized)

Note: In this figure we use our international sample of firm revenue expectations to report the binned scatter plot of future log forecast errors $\frac{g_{t+1} - F_{t+1}}{MAD_t}$ as a function of past revision $\frac{F_{t} - g_{t+1}}{MAD_t}$. The blue line is a local polynomial approximation, centered in the middle of each centile.
We look at the relation between error and lagged error in Figure 10. It shows the binned scatter plot of future log forecast errors \( g_{t+1} - F_t g_{t+1} \) as a function of past error \( g_t - F_t^{-1} g_t \), both of them normalized by firm-level volatility. As can be seen from this figure, the forecast errors follow a linear relationship for intermediate values (until about 1 unit of volatility), but the relationship reverses for larger past errors.

![Figure 10: Revenue Forecast Error as a Function of Past Error (Normalized)](image)

Note: In this figure we use our international sample of firm revenue expectations to report the binned scatter plot of future log forecast errors \( g_{t+1} - F_t g_{t+1} \) as a function of past error \( g_t - F_t^{-1} g_t \). The blue line is a local polynomial approximation, centered in the middle of each centile.

### 5.3 Evidence from Returns

Our model in Section 4.5 predicts that past returns should predict future returns in a non-linear way. We now provide evidence on returns based on CFM’s international monthly stock returns data described in the Data section of this paper.

In Figure 11 we first show a smoothened binscatter plot of future returns on past returns. Future returns are monthly and past returns are calculated over the past 12 months excluding the last month of returns, as is common in the literature on stock momentum. The only difference here with the standard literature is that we take
Table 2: Sample size by exchange (returns)

| Index | Total | 2000 | 2005 | 2010 | 2015 | 2020 |
|-------|-------|------|------|------|------|------|
| AEX   | 5892  | 0    | 290  | 300  | 278  | 283  |
| AS5   | 47252 | 0    | 2127 | 2340 | 2164 | 2361 |
| CAC   | 9946  | 0    | 475  | 474  | 474  | 480  |
| DAX   | 6531  | 0    | 360  | 360  | 360  | 351  |
| HSC   | 5946  | 0    | 0    | 281  | 435  | 585  |
| HSI   | 8949  | 0    | 0    | 522  | 582  | 597  |
| IBE   | 8663  | 0    | 398  | 419  | 410  | 419  |
| IND   | 5377  | 0    | 0    | 360  | 360  | 348  |
| KOS   | 36595 | 0    | 0    | 2381 | 2345 | 2369 |
| MID   | 120017| 4585 | 4773 | 4770 | 4583 | 4750 |
| NDX   | 21624 | 0    | 1171 | 1200 | 1234 | 1197 |
| NIF   | 6794  | 0    | 0    | 52   | 599  | 600  |
| NKY   | 60783 | 2642 | 2674 | 2698 | 2676 | 2698 |
| OMX   | 7565  | 0    | 348  | 360  | 360  | 360  |
| RAY   | 830673| 29792| 32936| 33833| 31915| 33247|
| SMI   | 3715  | 0    | 0    | 233  | 240  | 228  |
| SPT   | 10200 | 0    | 0    | 707  | 708  | 720  |
| SX5   | 9744  | 0    | 0    | 600  | 600  | 587  |
| TOP   | 5229  | 0    | 0    | 5    | 504  | 452  |
| TPX   | 518734| 16490| 19074| 19783| 21722| 25588|
| TWY   | 13663 | 0    | 0    | 355  | 1168 | 1028 |
| UKX   | 29775 | 1086 | 1180 | 1196 | 1180 | 1187 |
the log of returns (this is done because this analysis tends to focus on extreme past returns).

Figure 11: Binscatter Plot of Returns by Past Returns

Panel A: Entire Sample
Panel B: US vs International
Panel C: By holding Period
Panel D: By S index

Note: These 4 panels represent smoothed binned scatter plots of future log returns as a function of log past cumulative returns of the past 12 months excluding the last month. Panel A is the entire sample, Panel B splits the sample into US and International stocks, Panel C computes future returns using different holding periods, and Panel D by size quintile.

Figure 11 shows the binned scatter plot for different splits of the data. Panel A looks at the entire dataset and provides a picture consistent with our prediction: There is momentum for most levels of past returns, but for extreme values it is mean-reversion that prevails. Panel B shows that this pattern holds both on US data and non-US returns. Panel C investigates the role of various holding periods, i.e. looking for future returns over the following 1, 3, 6, 9 and 12 months. We find that the S-shaped curve emerges as soon as this holding period is longer than one month. In panel D, we sort stocks into market cap quintiles at the index-month
level. Even when examining different sizes of stocks, the S-shaped pattern is to be seen everywhere.

This finding suggests that the performance of traditional momentum strategies could be “boosted” by allowing for a region of reversal in both tails. To test this, we consider a self-financing strategy that goes long on momentum for moderate values of the momentum signal, and goes short on a momentum (i.e. long on reversal) for more extreme values of the signal. Let $s_{t-1}$ be a momentum signal calculated from past returns. Specifically, we calculate the momentum signal as the cross-sectional rank transform of cumulative returns over 11 months from $t-12$ to $t-1$, normalized such that $s_{i,t-1} = 0.5$ for firm with the greatest past returns, and $s_{i,t-1} = -0.5$ for the firm with the least. A portfolio with weights $w(s_i)$ is then formed at time $t$ as follows:

$$
w_{it} = \begin{cases} 
0.5 - \frac{s_{i,t-1}}{a} & \text{if } s_{i,t-1} \leq a \\
\frac{s_{i,t-1} - b}{b-a} - 0.5 & \text{if } a < s_{i,t-1} \leq b \\
0.5 - \frac{s_{i,t-1} - b}{1-b} & \text{if } b \leq s_{i,t-1}
\end{cases}
$$

where $a$ and $b$ are constant “inflection” points at which our strategy flips from reversal to momentum, and then from momentum to reversal.

Figure 12: The Error-Revision relationship: Sample Splits

Panel A: Portfolio weights
Panel B: Sharpe ratio by inflection point

Note: Portfolio formed across all firms in our equity sample, both U.S. and international. See text for details of momentum signal construction.

Our findings are depicted in figure 12. The inflection points of our strategy,
\(a\) and \(b\) are chosen to optimize the Sharpe ratio of our strategy over the sample period. Panel A shows the Sharpe ratio maximizing weighting function, while panel B shows how the Sharpe ratio varies for different upper and lower inflection points. Note that a strategy with a low inflection point of 0 and upper inflection point of 100 corresponds to a traditional momentum strategy, with no tail reversal. Clearly, our approach contains some look ahead bias, as the coefficients \(a\) and \(b\) are estimated on the entire sample. A more systematic investigation of these returns is beyond the scope of this paper.

In line with our earlier empirical results, we find that the Sharpe ratio of our strategy is maximized when the lower inflection point is set at the 13th cross-sectional percentile of the normalized momentum signal, and the upper inflection point is set at the 86th percentile. The maximum Sharpe ratio to this strategy (0.61) is 1.27x the Sharpe ratio we observe for a pure momentum strategy (0.48).

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.

Natural extensions of our work consists in exploring alternative forecasts data. Macro forecasts are unlikely to provide us with non-Gaussian processes and are in general too sparse to measure the tails of the DGP with enough accuracy. Within IBES studying EPS forecasts is another natural research direction, although it presents a scaling challenge. Growth cannot be computed for a large number of firms. Internal sales forecast from large companies could be another path.
References

AFROUZI, H., S. Y. KWON, A. LANDIER, Y. MA, AND D. THESMAR (2020): “Overreaction and Working Memory,” Tech. Rep. w27947, National Bureau of Economic Research, Cambridge, MA.

AXTELL, R. (2001): “Zipf Distribution of U.S. Firm Sizes,” Science, 293.

BORDALO, P., N. GENNAIOLI, Y. MA, AND A. SHLEIFER (2020): “Overreaction in Macroeconomic Expectations,” American Economic Review.

BORDALO, P., N. GENNAIOLI, R. L. PORTA, AND A. SHLEIFER (2019): “Diagnostic Expectations and Stock Returns,” The Journal of Finance, 74, 2839–2874.

BORDALO, P., N. GENNAIOLI, AND A. SHLEIFER (2018): “Diagnostic Expectations and Credit Cycles,” Journal of Finance, 73, 199–227, publisher: Wiley Online Library.

BOTTAZZI, G. AND A. SECCHI (2006): “Explaining the distribution of firm growth rates,” The RAND Journal of Economics, 37, 235–256.

BOUCHAUD, J., P. KRÜGER, A. LANDIER, AND D. THESMAR (2019): “Sticky Expectations and the Profitability Anomaly,” The Journal of Finance, 74, 639–674.

COIBION, O. AND Y. GORODNICHENKO (2015a): “Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts,” American Economic Review, 105, 2644–2678.

——— (2015b): “Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts,” American Economic Review, 105, 2644–78.

FARMER, L., E. NAKAMURA, AND J. STEINSSON (2022): “Learning About the Long Run.”

FUSTER, A., B. HEBERT, AND D. LAIBSON (2012): “Natural Expectations, Macroeconomic Dynamics, and Asset Pricing,” NBER Macroeconomics Annual, 26, 1–48, publisher: University of Chicago Press Chicago, IL.
Fuster, A., D. Laibson, and B. Mendel (2010): “Natural Expectations and Macroeconomic Fluctuations,” *Journal of Economic Perspectives*, 24, 67–84, publisher: American Economic Association.

Gabaix, X. (2009): “Power Laws in Economics and Finance,” *Annual Review of Economics*, 1, 255–294.

——— (2018): “Behavioral Inattention,” *Handbook of Behavioral Economics*.

Kozlowski, J., L. Veldkamp, and V. Venkateswaran (2020): “The Tail that Wags the Economy: Belief-Driven Business Cycles and Persistent Stagnation,” *Journal of Political Economy*.

Ma, Y., T. Ropele, D. Sraer, and D. Thesmar (2020): “A Quantitative Analysis of Distortions in Managerial Forecasts,” Working Paper.

Schmidhuber, C. (2021): “Trends, reversion, and critical phenomena in financial markets,” *Physica A: Statistical Mechanics and its Applications*.

Wyart, M. and J.-P. Bouchaud (2003): “Statistical models for company growth,” *Physica A: Statistical Mechanics and its Applications*, 326, 241–255.