

DIAGNOSTIC EXPECTATIONS AND STOCK RETURNS

Pedro Bordalo, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer¹

First Draft, November 2016, This Draft December 15 2018

Abstract

We revisit LaPorta's (1996) finding that returns on stocks with the most optimistic analyst long-term earnings growth forecasts are lower than those for stocks with the most pessimistic forecasts. We document the joint dynamics of fundamentals, expectations, and returns of these portfolios, and explain the facts using a model of belief formation based on the representativeness heuristic. Analysts forecast fundamentals from observed earnings growth, but overreact to news by exaggerating the probability of states that have become objectively more likely. We find supportive evidence for the model's distinctive predictions. An estimation of the model quantitatively accounts for the key patterns.

¹ Saïd Business School, University of Oxford, Bocconi University and IGER, Brown University, and Harvard University, respectively. Gennaioli thanks the European Research Council (GA 647782) and Shleifer thanks the Pershing Square Venture Fund for Research on the Foundations of Human Behavior for financial support of this research. We are grateful to seminar participants at Brown University and Sloan School, and especially to Josh Schwartzstein, Jesse Shapiro, Pietro Veronesi, Yang You, and the referees for helpful comments. We also thank V. V. Chari, who encouraged us to confront our model of diagnostic expectations with the Kalman filter. The authors have no relevant or material financial interests that relate to the research described in this paper.

I. Introduction

La Porta (1996) shows that expectations of stock market analysts about long-term earnings growth of the companies they cover have strong predictive power for these companies' future stock returns. Companies whose earnings growth analysts are most optimistic about earn poor returns relative to companies whose earnings growth analysts are most pessimistic about. Betting against extreme analyst optimism has been on average a good idea. La Porta (1996) interprets this finding as evidence that analysts, as well as investors who follow them or think like them, are too optimistic about stocks with rapidly growing earnings, and too pessimistic about stocks with deteriorating earnings. As a result, the former stocks are overvalued, the latter are undervalued, and the predictability of returns is caused by the correction of expectations.

We revisit this puzzle using a model of Kahneman and Tversky's representativeness heuristic developed by Gennaioli and Shleifer (GS 2010) and Bordalo, Coffman, Gennaioli and Shleifer (BCGS 2016). Relative to existing work, we make three innovations. First, we rely on a portable and psychologically founded model that has been used to describe the dynamics of beliefs in disparate settings, not only in the one at hand. Second, we assess both the qualitative and the quantitative performance of this model in explaining not only a cross-section of stock returns, but also the path of fundamentals leading to overvaluation, and crucially the path of analyst expectations. Third, we test new predictions that distinguish our model from mechanical models of extrapolation such as adaptive expectations. These new predictions follow from the fact that expectations in our model contain a "kernel of truth": they exaggerate true features of reality. Beliefs move in the correct direction, but by more than the right amount.

As a first step, we look at the data (described in Section II). Section III confirms with 20 additional years of data La Porta's finding that stocks with low long term earnings growth

forecasts, LLTG stocks, outperform stocks with high long term earnings growth forecasts, HLTG stocks. We present three additional facts. First, HLTG stocks exhibit fast past earnings growth, which slows down going forward. Second, forecasts of future earnings growth of HLTG stocks are too optimistic, and are systematically revised downward later. Third, HLTG stocks exhibit good past returns but their returns going forward are low. The opposite dynamics obtain for LLTG stocks, but in a less extreme form, an asymmetry we do not account for in our model.

The data suggests that analysts use a firm's past performance to infer its future performance, but overreact. This updating mechanism arises naturally from overweighting of representative types. GS (2010) and BCGS (2016) model a type t to be representative of a group G when it occurs more frequently in that group than in a reference group $-G$. For instance, after a positive medical test, the representative patient is $t = \text{"sick"}$, because sick people are truly more prevalent among those who tested positive than in the overall population. After such a positive test, the representative sick type quickly comes to mind and the doctor inflates its probability too much, which may still be objectively low if the disease is rare (Casscells et al. 1978). There is a kernel of truth in departures from rationality: the doctor overreacts to the objectively useful information from the test.

Applied to an analyst learning about a firm's unobserved fundamentals in light of a noisy signal such as current earnings, representativeness yields a distorted Kalman filter – which we call the “Diagnostic Kalman Filter” – that overinflates the probability of future earnings growth realizations whose likelihood has objectively increased the most in light of recent news. After strong earnings growth, the probability that the firm is the next “Google” goes up. This type becomes representative and analysts inflate its probability excessively, even though Googles remain rare in absolute terms. As good news stop arriving, over-optimism cools off. Strong

earnings growth thus causes overvaluation and subsequently disappointment, leading to reversals of optimism and abnormally low returns.

Section V shows that this model makes predictions consistent with the evidence in Section III. Section VI performs a first pass quantitative assessment of the model. We first show that expectations data point to overreaction to news, in the sense that revisions of the forecast of long term growth negatively predict errors in that forecast. We then use the Simulated Method of Moments (SMM) to estimate the parameter controlling the strength of representativeness by matching observed overreaction and the empirical moments of the earnings process. We estimate that forecasters react about twice as much to information as is objectively warranted. This estimate lies in the ballpark of those we obtained using the expectations of professional forecasters about credit spreads (Bordalo, Gennaioli, and Shleifer, BGS 2018) and many macroeconomic variables (Bordalo, Gennaioli, Ma, Shleifer 2018). We also find that stocks overreact to news over a time scale of about 3 years. Notably, while the estimation uses expectations and fundamentals alone, it predicts reasonably well the observed return spread between LLTG and HLTG stocks, as well as several other features of the data.

In Section VII we test three novel predictions of the model, all following from the “kernel of truth” property as applied to the present dynamic setting. First, we show that the HLTG group contains a fat right tail of future exceptional performers, and that analysts attach an excessively high probability that HLTG firms are in that tail. There are very few Googles, but they are concentrated in the HLTG group, and analysts exaggerate their frequency. This is precisely what the kernel of truth predicts.

Second, we show that the return spread between LLTG and HLTG stocks widens among firms having more volatile or persistent fundamentals. This is also in line with the kernel of truth:

in both cases, good news are even more informative about strong future performance, which render Googles even more representative. Third, we show that expectations about HLTG (LLTG) stocks revert downward (upward) even in the absence of bad (good) news. This is also in line with the kernel of truth: analyst forecasts reflect the true mean reversion in earnings growth.

Mechanical model of beliefs, such as adaptive expectations, cannot yield these facts, which are due to the forward looking but not fully rational learning we see in the data.

Our paper follows extensive research on over-reaction and volatility, which begins with Shiller (1981), DeBondt and Thaler (1985, 1987), Cutler, Poterba, and Summers (1990, 1991), and DeLong et al. (1990a,b). This work often uses mechanical rules for belief updating such as adaptive expectations or adaptive learning (e.g., Barsky and DeLong 1993, Barberis et al. 2015, Adam, Marcet, and Beutel 2017).² Barberis, Shleifer, and Vishny (BSV, 1998) is the paper closest in spirit, though not in formulation, to our current work, since it is also motivated by representativeness. BSV present a model of Bayesian learning in which the decision maker is trying to distinguish models that are all incorrect. BSV do not model representativeness explicitly. As such, the BSV specification does not allow for a tight link between measurable reality and measurable beliefs that is central to our theory and evidence. Daniel, Hirshleifer, and Subramanyam (1998) and Odean (1998) model investor overconfidence, the tendency of decision-makers to exaggerate the precision of private information, which causes inaccurate beliefs and excess trading. Unlike overconfidence, our model yields overreaction to not just private but also public information, such as earnings news.

² Other papers include Barberis and Shleifer (2003), Glaeser and Nathanson (2015), Hong and Stein (1999), Marcet and Sargent (1989), and Adam, Marcet, and Nicolini (2016). In Adam, Marcet, and Beutel (2017), agents learn the mapping between fundamentals and prices, but they are rational about fundamentals. Pastor and Veronesi (2003, 2005, 2009) present rational learning models in which uncertainty about the fundamentals of some firms can yield predictability in aggregate stock returns. This approach does not analyze expectations data or cross sectional differences in returns.

Our model is based on a psychologically founded distortion applicable whenever agents make probability judgments, and is not specific to asset pricing. It allows to unify several biases hitherto viewed as separate, such as extrapolation, overreaction to information, and neglect of tail risk. The model is portable to different domains (Rabin 2013), and has been used to shed light on fallacies in probabilistic judgments (GS 2010), social stereotypes (BCGS 2016), beliefs about gender (BCGS 2017), credit cycles (BGS 2017), and macroeconomic forecasts (BGMS 2018). Unification of biases, quantification, and cross-context applicability discipline the model, helping to identify realistic and robust alternatives to rational expectations.

We use expectations data not only to predict returns, but also as a central element of the model and its empirical estimation.³ Such data are becoming increasingly common in recent work (Ben David, Graham and Harvey 2013, Greenwood and Shleifer 2014, Gennaioli, Ma, and Shleifer 2015). Bouchaud, Landier, Krueger and Thesmar (2018) use analyst expectations data to study the profitability anomaly, and offer a model in which expectations underreact to news, in contrast with our focus on overreaction. As we show in Section VI.A, in our data there is also some short-term underreaction, but at the long horizons of LTG forecasts overreaction prevails. Daniel, Klos, and Rottke (2017) show that stocks featuring high dispersion in analyst expectations and high illiquidity earn high returns, but do not offer a theory of expectations and their dispersion.

II. Data and Summary Statistics

II.A. Data

³ Early uses of beliefs data include Dominguez (1986) and Frankel and Froot (1987, 1988). A large literature on analyst expectations shows that they are on average too optimistic (Easterwood and Nutt 1999, Michaely and Womack 1999, Dechow, Hutton, and Sloan 2000).

We gather data on analysts' expectations from IBES, stock prices and returns from CRSP, and accounting information from CRSP/COMPUSTAT. Below we describe the measures used in the paper and, in parentheses, provide their mnemonics in the primary datasets.

From the IBES Unadjusted US Summary Statistics file we obtain mean analysts' forecasts for earnings per share and their expected long-run growth rate (*meanest*, henceforth "LTG") for the period December 1981, when LTG becomes available, through December 2016. IBES defines LTG as the "expected annual increase in operating earnings over the company's next full business cycle", a period ranging from three to five years. From the IBES Detail History Tape file we get analyst-level data on earnings forecasts. We use CRSP daily data on stock splits (*cfacshr*) to adjust IBES earnings per share figures. On December of each year between 1981 and 2015, we form LTG decile portfolios based on stocks that report earnings in US dollars.⁴

The CRSP sample includes all domestic common stocks listed on a major US stock exchange (i.e. NYSE, AMEX, and NASDAQ) except for closed-end funds and REITs. Our sample starts in 1978 and ends in 2016. We present results for both buy-and-hold annual returns and daily cumulative-abnormal returns for various earnings' announcement windows. We compute annual returns by compounding monthly equally-weighted returns for LTG portfolios. If a stock is delisted, whenever a post-delisting price exists in CRSP, we use it in the computations. When CRSP is unable to determine the value of a stock after delisting, we assume that the investor was able to trade at the last quoted price. Given that IBES surveys analysts around the middle of the month (on Thursday of the third week of the month), LTG is in the information set when we form portfolios. Daily cumulative abnormal returns are defined relative to CRSP's equally-weighted index. We also gather data on market capitalization in December of year t as

⁴ We form portfolios in December of each year because IBES data on LTG starts in December of 1981. Unlike Fama and French (1993), we know exactly when the information required for an investable strategy is public.

well as the pre-formation 3-year return ending on December of year t . Finally, we rank stocks into deciles based on market capitalization using breakpoints for NYSE stocks.

We get from the CRSP/COMPUSTAT merged file data on assets (at), sales ($sale$), net income (ni), book equity, common shares used to calculate earnings per share ($cshpri$), adjustment factor for stock splits ($adjex_f$), and *Wall Street Journal* dates for quarterly earnings' releases (rdq). Our CRSP/COMPUSTAT data covers 1978-2016. We use annual and quarterly accounting data. We define book equity as stockholders' equity (depending on data availability seq , $ceq+pfd$, or $at-lt$) plus deferred taxes (depending on data availability $txditc$ or $txdb+itcb$) minus preferred equity (depending on data availability $pstkr$, $pstkl$, or $pstk$). We define operating margin as the difference between sales and cost of goods sold ($cogs$) divided by assets, and return on equity as net income divided by book equity. We compute the annual growth rate in sales per share in the most recent 3 fiscal years. When merging IBES with CRSP/COMPUSTAT, we follow the literature and assume that data for fiscal periods ending after June becomes available during the next calendar year.

II.B. Summary Statistics

Table 1 reports the statistics for the LTG decile portfolios. The number of stocks with CRSP data on stock returns and IBES data on LTG varies by year, ranging from 1,310 in 1981 to 3,849 in 1997. On average, each LTG portfolio contains 236 stocks. The forecasted growth rate in earnings per share ranges from 4% for the lowest LTG decile (LLTG) to 38% for the highest decile (HLTG), an enormous difference. LLTG stocks are larger than HLTG stocks in terms of both total assets (7,123 MM vs. 1,027 MM) and market capitalization (3,862 MM vs. 1,641 MM). However, differences in size are not extreme: the average size decile is 5.0 for LLTG and 3.4 for

HLTG. Finally, LLTG firms are older than HLTG firms (27.5 vs. 6.8 years) and more likely to remain publicly-traded during the five years following the formation period (70% vs. 60%).

LLTG stocks have lower operating margins to asset ratios than HLTG stocks but higher return on equity (5% vs -7%). In fact, 31% of HLTG firms have negative eps while the same is true for only 12% of LLTG stocks. The high incidence of negative eps companies in the HLTG portfolio underscores the importance of the definition of LTG in terms of annual earnings growth over a full business cycle. Current negative earnings do not hinder these firms' future prospects.

Table 1 – Descriptive Statistics for Portfolios Formed on *LTG*.

We form decile portfolios based on analysts' expected growth in earnings per share (*LTG*) in December of each year between 1981 and 2015. The table reports time-series means of the variables described below for equally-weighted *LTG* portfolios. Unless otherwise noted, accounting variables pertain to the most recently available fiscal year, where we follow the standard assumption that data for fiscal periods ending after June become available during the next calendar year. *Assets* is book value of total assets (in millions). *Market capitalization* is the value of common stock on the last trading day of year *t* (in millions). *Size decile* refers to deciles of market capitalization with breakpoints computed using only NYSE stocks. *Years publicly traded* is the number of years since the first available stock price on CRSP. *%Listed after 5 years* is the percentage of firms that remain publicly traded 5 years after the formation period. *Operating margin to assets* is the difference between sales and cost of goods sold divided by *assets*. *Return on equity* is net income divided by book equity. *Percent eps positive* is the fraction of firms with positive earnings. *Observations* is the number of observations in a year. All variables are capped at the 1% and 99% levels.

	<i>LTG decile</i>									
	1	2	3	4	5	6	7	8	9	10
Expected growth eps (<i>LTG</i>)	4%	9%	11%	12%	14%	16%	18%	20%	25%	38%
Assets (MM)	7,123	9,387	9,398	7,508	5,008	3,591	2,340	1,974	1,373	1,027
Market capitalization (MM)	3,862	4,692	5,316	4,552	3,967	3,270	2,749	2,431	1,679	1,641
Size decile	5.0	5.0	5.4	5.1	5.0	4.5	4.4	3.8	3.5	3.4
Years publicly traded	27.5	24.4	23.0	21.3	18.9	15.9	14.0	11.0	8.5	6.8
%Listed after 5 years	70%	66%	68%	69%	68%	67%	67%	65%	62%	60%
Operating margin to assets	22%	27%	32%	35%	39%	41%	44%	44%	43%	39%
Return on equity	5%	7%	9%	10%	10%	8%	10%	7%	2%	-7%
Percent eps positive	88%	90%	93%	94%	93%	91%	92%	88%	82%	69%
Observations per year	246	246	230	244	237	249	214	236	235	222

III. A New Look at the Data.

To revisit the La Porta (1996) finding, we sort stocks by analysts' forecast of their long-term growth in earnings per share (LTG). The LLTG portfolio is the 10% of stocks with most pessimistic forecasts, the HLTG portfolio is the 10% of stocks with most optimistic forecasts. Figure 1 reports geometric averages of one-year returns on equally weighted portfolios.

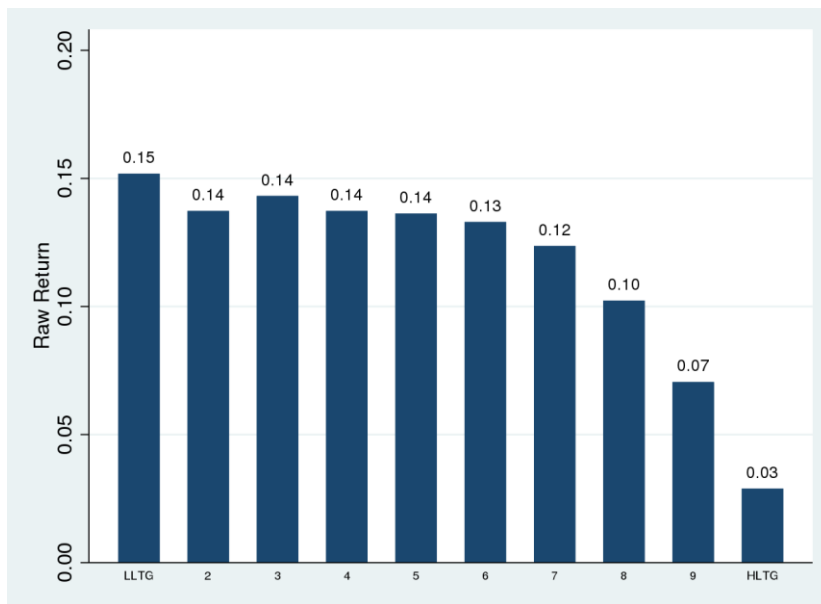


Figure 1. Annual Returns for Portfolios Formed on LTG. In December of each year between 1981 and 2015, we form decile portfolios based on ranked analysts' expected growth in earnings per share and report the geometric average one-year return over the subsequent calendar year for equally-weighted portfolios with monthly rebalancing. A portfolio that is long on LLTG stocks and short on HLTG earns an average yearly (log) return of 13.6%, with a t-statistic of 2.22.

Consistent with La Porta (1996), the LLTG portfolio earns a compounded average return of 15% in the year after formation, the HLTG portfolio earns only 3%.⁵ Adjusting for systematic risk only deepens the puzzle: the HLTG portfolio has higher market beta than the LLTG portfolio, and performs worse in market downturns (see Appendix B, Table B.1). A conventional factor analysis reveals that the excess returns of LLTG stocks survives controlling for three Fama-French factors, although it is positively correlated with the value premium, and becomes

⁵ The spread in Figure 1 is in line with, although smaller than, previous findings. La Porta (1996) finds an average yearly spread of 20% but employed a shorter sample (1982 to 1991). Dechow and Sloan (1997) use a similar sample to La Porta (1996) and find a 15% spread. Figure B.1 in Appendix B shows that the spread also holds in sample subperiods.

insignificant after one additionally controls for profitability or for betting against beta (Table B.3). This is unsurprising: LLTG stocks have low market to book ratios, low beta, and good profitability despite their low earnings growth.⁶ Kozak, Nagel, and Santosh (2018) show that it is generally incorrect to interpret these factor controls as adjustments for fundamental risk.

So why do LLTG stocks appear undervalued and HLTG overvalued? To assess the expectations-based hypothesis that analysts and the stock market may be too bullish on firms they are optimistic about and too bearish on firms they are pessimistic about, we document some basic facts connecting firms' performance, expectations and returns.

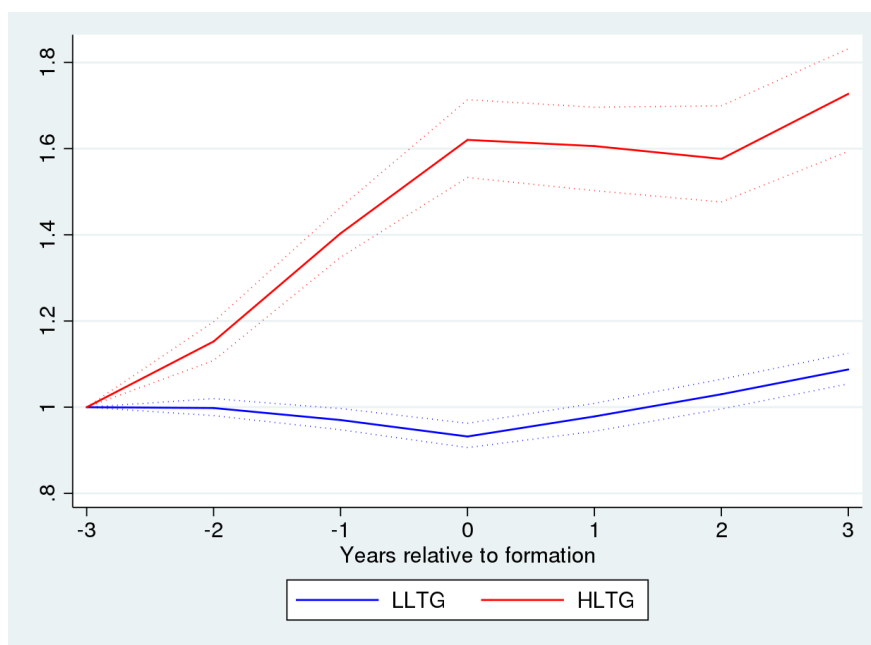


Figure 2. Evolution of EPS. In December of years (t) 1981, 1984, ..., 2011, and 2013, we form decile portfolios based on ranked analysts' expected growth in earnings per share (LTG). We report the (bootstrapped) mean value of earnings per share for the highest (HLTG) and lowest (LLTG) LTG deciles for each year between $t-3$ and $t+3$. We exclude firms with negative earnings in $t-3$ and normalize to 1 the value of earnings per share in $t-3$. To do so, we restrict the computation to firms that have positive earnings at $t-3$. The dotted lines indicate 5th and 95th confidence levels determined via nonparametric bootstrapping using 1,000 samples (the results are robust to changing the size of sample draws).

⁶ The LLTG-HLTG spread stays roughly constant across momentum terciles and size terciles, see Appendix B, Tables B.2 and B.3 (because the underperforming HLTG stocks tend to be small, the usual rationale for value-weighting does not apply).

Figure 2 reports average earnings per share of HLTG and LLTG portfolios in years $t - 3$ to $t + 3$ where $t = 0$ corresponds to portfolio formation. We normalize year $t - 3$ earnings per share of both portfolios to \$1. Earnings per share for HLTG stocks exhibit explosive growth during the pre-formation period, rising from \$1 in year -3 to \$1.56 in year 0. Earnings of LLTG firms decline to \$0.92 during the corresponding period. But the past does not repeat itself after portfolio formation: earnings growth of HLTG firms slows down, earnings of LLTG firms recover during the post-formation period. HLTG firms remain more profitable on average than LLTG firms 3 years after formation, but actual growth rates are much closer than the LTG differences in Table 1.

Figure 3 shows the average LTG for the HLTG and LLTG portfolios over the same time window. Prior to portfolio formation, expectations of long-term growth for HLTG firms rise dramatically in response to strong earnings growth (compare with Figure 2), while expectations for LLTG drop. After formation, expectations for HLTG firms are revised downwards, particularly during the first year, whereas those for LLTG firms are revised up. Three years after portfolio formation, earnings of HLTG firms are still expected to grow faster than those of LLTG firms, but the spread in expected growth rates of earnings has narrowed considerably. Data attrition is not responsible for this and other findings in this Section: the level of attrition is similar across HLTG and LLTG portfolios (40% for HLTG and 30% for LLTG five years post formation). The findings of the Figures also hold when we restrict attention to firms that survive for five years after formation.

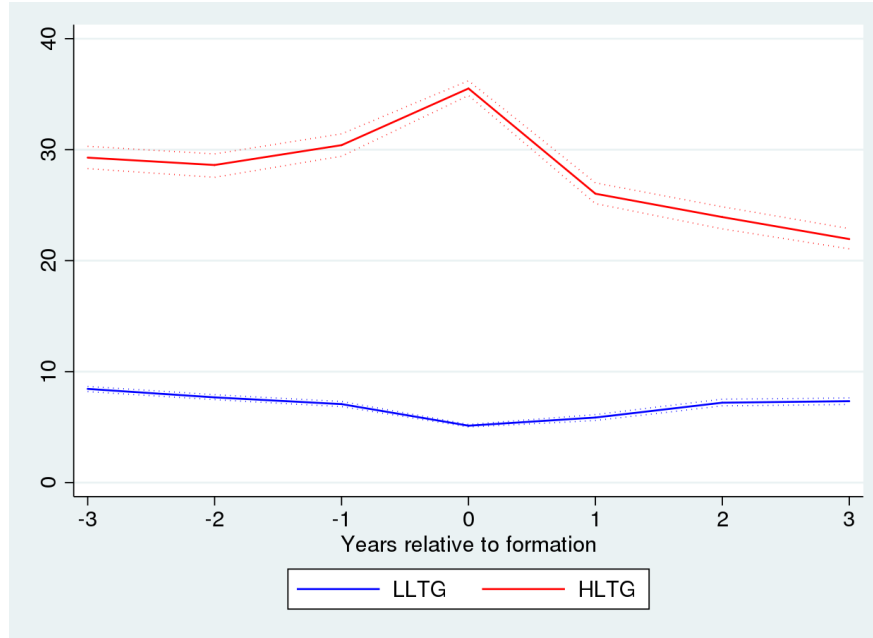


Figure 3. Evolution of LTG. In December of each year t between 1984 and 2013, we form decile portfolios based on ranked analysts' expected growth in earnings per share (LTG) and report the (bootstrapped) mean value of LTG on December of years $t-3$ to $t+3$ for the highest (HLTG) and lowest (LLTG) LTG deciles. We include in the sample stocks with LTG forecasts in year $t-3$. Values for $t+1$, $t+2$, and $t+3$ are based on stocks with IBES coverage for those periods. The dotted lines indicate 5th and 95th confidence levels determined via bootstrapping

Expectations of long-term growth follow the pattern of actual earnings per share of HLTG and LLTG portfolios displayed in Figure 2. Analysts seem to be learning about firms' earnings growth from their past performance. Fast pre-formation growth leads analysts to place a firm in the HLTG category. Post-formation growth slowdown triggers a downward revision of forecasts.

Mean reversion in forecasts may be caused by mean reversion in fundamentals, which is evident in Figure 2, or by the correction of analysts' expectations errors at formation. To see the role of expectations errors, Figure 4 reports the difference between realized earnings growth and analysts' LTG expectations in each portfolio, from formation to year $t + 3$. There is strong over-optimism, i.e. very negative forecast errors, for HLTG firms. There is also over-optimism for LLTG firms, consistent with many previous studies and usually explained by distorted analyst incentives (e.g., Dechow et al. 2000, Easterwood and Nutt 1999, Michaely and Womack 1999).

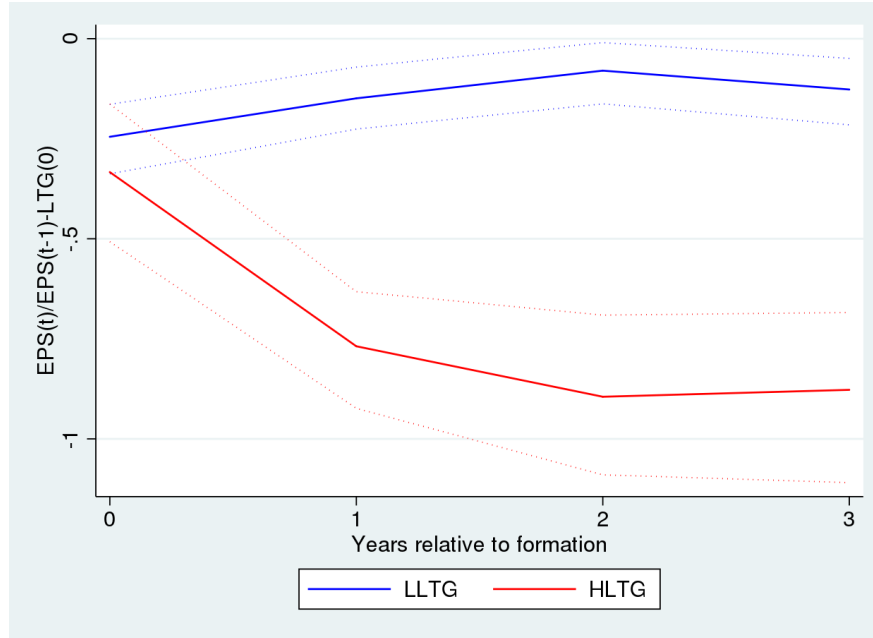


Figure 4. Realized earnings per share vs. Expected growth in long-term earnings. In December of each year t between 1981 and 2013 we form decile portfolios based on ranked analysts' expected growth in eps. We plot the (bootstrapped) mean difference between the annual growth rate in earnings per share in each year between t and $t+3$ and the forecast for long-term growth in earnings made in year t . To compute earnings growth, here we restrict to firms with positive earnings at $t = -1$. The dotted lines indicate 5th and 95th confidence levels determined via bootstrapping.

The overestimation of earnings growth for HLTG firms is economically large. By year 3, actual earnings are a small fraction of what analysts forecast: earnings per share grow from 0.16 at formation to 0.21 in $t + 3$, compared to the prediction of 0.70 based on LTG at formation.

There are two concerns with this evidence, both of which deal with the interpretation and reliability of our expectations data. The first is whether Figure 4 reflects genuine errors in analyst beliefs or alternatively their distorted incentives. For instance, analysts may blindly follow the market, reporting high LTG for firms that investors are excited about. This possibility does not undermine our analysis. To the extent that analysts' forecasts reflect investor beliefs, they are informative about the expectations shaping market prices. A more radical concern is that analysts' beliefs are unrelated to investor beliefs. To assess this possibility Figure 5 shows stock returns around earnings announcements, both pre- and post-formation. For every stock in the HLTG and

LLTG portfolios, we compute the 12-day cumulative return during the four quarterly earnings announcement days, in years $t - 3$ to $t + 4$, following the methodology of La Porta et al. (1997).

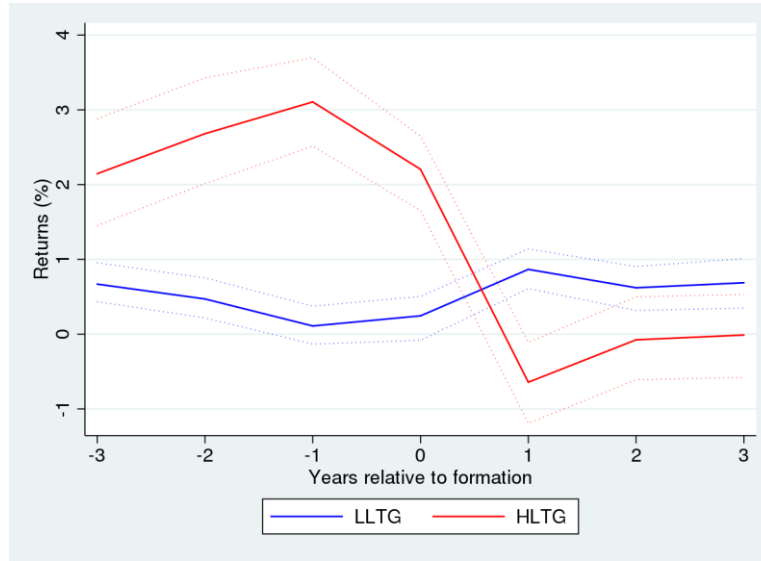


Figure 5. Twelve-day Returns on Earnings Announcements for LTG Portfolios. In December of each year t between 1981 and 2013, we form decile portfolios based on ranked analysts' expected growth in earnings per share. Next, for each stock, we compute the 3-day market-adjusted return centered on earnings announcements in years $t - 3, \dots, t + 3$. Next, we compute the annual return that accrues over earnings announcements by compounding all 3-day stock returns in each year. We report the equally-weighted (bootstrapped) average annual return during earnings announcements for the highest (HLTG) and lowest (LLTG) LTG deciles. Excess returns are defined relative to the equally-weighted CRSP market portfolio. The dotted lines indicate 5th and 95th confidence levels determined via bootstrapping.

HLTG stocks positively surprise investors with their earnings announcements in the years prior to portfolio formation, and their LTG is revised up.⁷ Returns are low afterwards, especially in year 1, consistent with the sharp decline in LTG in this period. Analysts' over-optimism thus seems to be shared by investors, so that HLTG stocks consistently disappoint in the post formation period. The converse holds for LLTG stocks, but in a milder form. Analysts' expectations are not noise: they correlate both with actual earnings and stock returns.

⁷ The persistence of positive surprises and high returns exhibited by the HLTG portfolio pre-formation in Figure 5 should not be confused with time-series momentum. On average, firms require a sequence of positive shocks to be classified as HLTG.

In summary, the data point to three key findings. First, optimism about HLTG firms follows the observation of fast earnings growth, and is reversed when growth slows down. Analysts appear to be learning the fundamentals of firms based on past performance. Second, one should be skeptical of analyst rationality, as shown by the evidence on systematic forecast errors in the HLTG group (which includes more than 200 firms per year). Third, the dynamics of returns match the dynamics of expectations and their errors, both pre and post formation.

We next present a model derived from psychological first principles that sheds light on this evidence. It assumes that the consensus analyst forecasts are the relevant expectations that shape prices and seeks to jointly explain these forecasts, their errors, and return predictability.⁸

IV. A model of learning with representativeness

IV.A. The Setup

At time t , the natural logarithm of the earnings per share (eps) $x_{i,t}$ of firm i is given by:

$$x_{i,t} = bx_{i,t-1} + f_{i,t} + \epsilon_{i,t}, \quad (1)$$

where $b \in [0,1]$ captures mean-reversion in eps, and $\epsilon_{i,t}$ denotes a transitory i.i.d. normally distributed shock to eps, $\epsilon_{i,t} \sim \mathcal{N}(0, \sigma_\epsilon^2)$. The term $f_{i,t}$, which we call the firm's "fundamental", captures the firm's persistent earnings capacity. It follows the law of motion:

$$f_{i,t} = a \cdot f_{i,t-1} + \eta_{i,t}. \quad (2)$$

⁸ As suggested by Pietro Veronesi, predictable returns (but not forecast errors) can arise under full rationality if analysts jointly learn a firm-specific required return and earnings. We return to this point in Section VI.

where $a \in [0,1]$ is persistence and $\eta_{i,t} \sim \mathcal{N}(0, \sigma_\eta^2)$ is an i.i.d. normally distributed shock. We can think of firms with exceptionally high $f_{i,t}$ as “Googles” that will produce very high earnings in the future, and firms with low $f_{i,t}$ as “lemons” that will produce low earnings in the future. We assume stationarity of earnings by imposing $b \leq a$.

The analyst observes eps $x_{i,t}$ but not the fundamental $f_{i,t}$. The Kalman filter characterizes the forecasted distribution of $f_{i,t}$ at any time t conditional on the firm’s past and current earnings $(x_{i,u})_{u \leq t}$. Given the mean forecasted fundamental $\hat{f}_{i,t-1}$ for firm i at $t - 1$ and its current earnings $x_{i,t}$, the firm’s current forecasted fundamental is normally distributed with variance σ_f^2 and mean:⁹

$$\hat{f}_{i,t} = a\hat{f}_{i,t-1} + K(x_{i,t} - bx_{i,t-1} - a\hat{f}_{i,t-1}), \quad (3)$$

where $K \equiv \frac{a^2\sigma_f^2 + \sigma_\eta^2}{a^2\sigma_f^2 + \sigma_\eta^2 + \sigma_\epsilon^2}$ is the signal to noise ratio.

The new forecast of fundamentals starts from the history-based value $a\hat{f}_{i,t-1}$ but adjusts it in the direction of the current surprise $x_{i,t} - bx_{i,t-1} - a\hat{f}_{i,t-1}$. The extent of adjustment increases in K . Absent transitory shocks ($\sigma_\epsilon^2 = 0$), earnings are so informative about fundamentals that the adjustment is full (i.e., $\hat{f}_{i,t} = x_{i,t} - bx_{i,t-1}$). As transitory shocks get more frequent, earnings become a noisier signal, which gets discounted in assessing fundamentals ($K < 1$). The signal to noise ratio thus separates the transitory from the persistent earnings shock (i.e., $\eta_{i,t}$ from $\epsilon_{i,t}$).

We next describe how the representativeness heuristic distorts this learning process.

IV.B. Representativeness and the Diagnostic Kalman filter

⁹ In the presence of fundamental shocks $\eta_{i,t}$, a firm’s fundamental is never learned with certainty. Equation (3) obtains when the variance of fundamentals converges to its steady state σ_f^2 , which is defined as the solution to:

$$a^2\sigma_f^4 + \sigma_f^2[\sigma_\eta^2 + (1 - a^2)\sigma_\epsilon^2] - \sigma_\eta^2\sigma_\epsilon^2 = 0$$

Kahneman and Tversky (KT 1972) argue that the automatic use of the representativeness heuristic causes individuals to estimate a type as likely in a group when it is merely representative. KT define representativeness as follows: “an attribute is representative of a class if it is very diagnostic; that is, the relative frequency of this attribute is much higher in that class than in a relevant reference class (TK 1983).” Starting with KT (1972), experimental evidence has found ample support for the role of representativeness.

Consider a decision maker assessing the distribution $h(T = \tau|G)$ of a variable T in a group G . Gennaioli and Shleifer (2010) define the representativeness of the specific type τ for G as:

$$R(\tau, G) \equiv \frac{h(T = \tau|G)}{h(T = \tau|-G)}. \quad (4)$$

As in KT, a type is more representative if it is *relatively* more frequent in G than in the comparison group $-G$. To capture overestimation of representative types, BCGS (2016) assume that probability judgments are formed using the representativeness-distorted density:

$$h^\theta(T = \tau|G) = h(T = \tau|G) \left[\frac{h(T = \tau|G)}{h(T = \tau|-G)} \right]^\theta Z, \quad (5)$$

where $\theta \geq 0$ and Z is a constant ensuring that the distorted density $h^\theta(T = \tau|G)$ integrates to 1. The extent of probability distortions increases in θ , with $\theta = 0$ capturing the rational benchmark.

In Kahneman and Tversky’s quote, as well as in Equation (4), the representativeness of a type depends on its *true* relative frequency in group G . GS (2010) interpret this feature on the basis of limited and selective memory. True information $h(T = \tau|G)$ and $h(T = \tau|-G)$ about a group is stored in a decision maker’s long term memory. Representative types, being distinctive of the group under consideration, are more readily recalled than other types and thus overweighted.

This setup can be applied to prediction and inference problems (as in BCGS 2016 and BGS 2018). Consider for example the doctor assessing the health status of a patient, $T = \{healthy, sick\}$ in light of a positive medical test, $G = positive$. The positive test is assessed in the context of untested patients ($-G = untested$). Applying the previous definition, being sick is representative of patients who tested positive if and only if:

$$\frac{\Pr(T = sick|G = positive)}{\Pr(T = sick|-G = untested)} > \frac{\Pr(T = healthy|G = positive)}{\Pr(T = healthy|-G = untested)},$$

namely when $\Pr(G = positive|T = sick) > \Pr(G = positive|T = healthy)$. The condition holds if the test is even minimally informative of the health status. A positive test brings “sick” to mind because the true probability of this type has increased the most after the positive test is revealed. The doctor may then deem the sick state likely even if the disease is rare (Casscells et al. 1978), committing a form of base rate neglect described in TK’s (1974).

We apply this logic to the problem of forecasting a firm’s earnings. The analyst must infer the firm’s type $f_{i,t}$ after observing the current earnings surprise $x_{i,t} - bx_{i,t-1} - af_{i,t-1}$. Such surprise is akin to seeing the medical test. As we saw previously, the *true* conditional distribution of firm fundamentals $f_{i,t}$ is normal, with variance σ_f^2 and the mean given by Equation (3). This is our target distribution $h(T = \tau|G)$. As in the medical example, the information content of the earnings $x_{i,t}$ for fundamentals $f_{i,t}$ is assessed relative to the background information set in which no news is received, namely if the earnings surprise is zero $x_{i,t} - bx_{i,t-1} = af_{i,t-1}$. The comparison distribution $h(T = \tau|-G)$ is then also normal with mean $af_{i,t-1}$ and variance σ_f^2 .

The assumption of normality implies the monotone likelihood ratio property of posteriors relative to priors, which makes representativeness easy to characterize. After good news, the most representative firm types are on the right tail. These firms are outweighed in beliefs while firms

in the left tail are neglected. After bad news the reverse is true. Appendix A proves that these distortions generate diagnostic beliefs as follows:

Proposition 1 (*Diagnostic Kalman filter*) *In the long run, upon seeing $x_{i,t} - bx_{i,t-1}$, the analyst's posterior about the firm's fundamentals is normally distributed with variance σ_f^2 and mean:*

$$\hat{f}_{i,t}^\theta = a\hat{f}_{i,t-1} + K(1 + \theta)(x_{i,t} - bx_{i,t-1} - a\hat{f}_{i,t-1}). \quad (6)$$

When analysts overweight representative types, their beliefs exaggerate the signal to noise ratio relative to the standard Kalman filter, inflating the fundamentals of firms receiving good news and deflating those of firms receiving bad news. Exaggeration of the signal to noise ratio is reminiscent of overconfidence, but here over-reaction occurs with respect to public as well as private news.¹⁰ The psychology is in fact very different from overconfidence: in our model, as in the medical test example, overreaction is caused by the neglect of base rates. After good news, the most representative firms are Googles, those with high earnings capacity $f_{i,t}$. This firm type readily comes to mind and its probability is inflated, despite the fact that Googles are rare. After bad news, the most representative firms are lemons. The analyst exaggerates the probability of this type, despite the fact that lemons are also quite rare. Overreaction to news increases in θ . At $\theta = 0$ the model reduces to rational learning.

The key property of diagnostic expectations is “the kernel of truth”: distortions in beliefs exaggerate true patterns in the data. In Equation (6) departures from rationality are a function of the true persistence a of fundamentals and of the true signal to noise ratio K , which depends on the volatility of fundamentals. The kernel of truth distinguishes our approach from alternative

¹⁰ In fact, overconfidence predicts under-reaction to public news such as earnings releases (see Daniel et al. 1998).

theories of extrapolation such as adaptive expectations or BSV (1998). In Section VII we present some empirical tests comparing alternative models.

To streamline the analysis, we defined representativeness with respect to the recent news $x_{i,t} - bx_{i,t-1} - a\hat{f}_{i,t-1}$. As we discuss in BGS (2018), two alternative specifications may be relevant. First, memory can cause the reference group $-G$ to move slowly, leading the news to be compared to conditions further in the past. Formally, representativeness at t can be defined with respect to the information available at $t - s$. In this case, the comparison distribution $h(t| - G)$ is the true distribution of fundamentals conditional on the state $a^s \hat{f}_{i,t-s}$ that would arise if no news are attended to for the last s periods. Slow moving $-G$ causes more persistent departures from rationality, which last for s periods in our model. In Section VI we show that this is an important dimension for calibrating the model, although it does not affect its qualitative properties.

Second, the reference group may alternatively be formed by the lagged *distorted* expectations, as opposed to the unbiased beliefs based on past information. This may occur because news brings to mind the recently held expectations (rather than the full distribution stored in memory), so that the comparison distribution $h(t| - G)$ is the diagnostic forecast at $t - s$ of fundamentals at t , which is normal with mean $a^s \hat{f}_{i,t-s}^\theta$. This specification loses the key predictive feature of the model, namely the kernel of truth -- the hypothesis that beliefs overreact to objective information. It also is not portable across different domains (e.g. it cannot be applied to static problems such as group stereotypes).¹¹

V. The Model and the Facts

¹¹ In BGS (2018) we work out this specification and show that, over long time series, it also exhibits overreaction to objective news but at each time t the distortion of expectations propagates arbitrarily far into the future. In Section VI we show that, at the relevant time scales, the two specifications perform similarly.

To link our model to the data we shift attention from the level of earnings to the growth rate of earnings, which directly maps into LTG. Denote by h the horizon over which the growth forecast applies, which is about 4 years for LTG. The LTG of firm i at time t as the firm's expected earnings growth over this horizon, $LTG_{i,t} = \mathbb{E}_{i,t}^\theta(x_{t+h} - x_t)$. By Equations (1) and (6), this boils down to:

$$LTG_{i,t} = -(1 - b^h)x_t + a^h \frac{1 - (b/a)^h}{1 - (b/a)} \hat{f}_{i,t}^\theta.$$

Expectations of long-term growth are shaped by mean reversion in eps and by fundamentals. LTG is high when firms have experienced positive news, so $\hat{f}_{i,t}^\theta$ is high, and/or when current earnings x_t are low, which also raises future growth. Both conditions line up with the evidence, which shows HLTG firms have experienced fast growth (Figure 2), and have low eps (Table 1).

We next show that the model accounts for the previously documented facts regarding the long run distribution of fundamentals $f_{i,t}$ (which has zero mean and variance $\frac{\sigma_\eta^2}{1-\alpha^2}$) and analysts' mean beliefs $\hat{f}_{i,t}^\theta$ (which has zero mean and variance $\sigma_{f^\theta}^2$).¹² At time t we identify the high LTG group HLTG_t as the 10% of firms with the highest assessed future earnings growth, and the low LTG group LLTG_t as the 10% of firms with the lowest assessed future earnings growth.

V.A. Representativeness and the Features of Expectations

We first consider the patterns of fundamentals and expectations documented in Figures 2, 3 and 4. In Section V.B, we review the patterns of returns documented in Figures 1 and 5.

¹² There is no distortion in the average diagnostic expectation across firms because in steady state the average earning surprise is zero. Thus, the average diagnostic expectation coincides with the average rational expectation. However, diagnostic beliefs are fatter-tailed than rational ones, because they exaggerate the frequency of Googles and Lemons.

We start from Figure 2, which says that HLTG firms experience a period of pronounced growth before portfolio formation, while LLTG firms experience a period of decline.

Proposition 2. *Provided a, b, K, θ satisfy*

$$b^h + a^h \frac{1 - (b/a)^h}{1 - (b/a)} \left[\frac{K(1 + \theta) - a}{1 - a} \right] > 1, \quad (7)$$

the average HLTG_t (LLTG_t) firm experiences positive (negative) earnings growth pre-formation.

Positive earnings news have two conflicting effects on LTG. On the one hand, they raise estimated fundamentals $\hat{f}_{i,t}^\theta$, which raise future growth forecasts. On the other hand, they lower future growth via mean reversion. Condition (7) ensures that the former effect dominates, so that HLTG firms are selected from the recent good performers, while LLTG firms are selected from recent bad performers. Condition (7) is more likely to hold the less severe is mean reversion (i.e., when b is close to 1) and the higher is the signal to noise ratio K , as well as for high θ .

Combined with mean reversion of earnings, Proposition 2 accounts for Figure 2, in which HLTG firms experience positive growth pre-formation, which subsequently cools off, while LLTG firms go through the opposite pattern. But the model can also account for the fact documented in Figure 4 that expectations for the long-term growth of HLTG firms are excessively optimistic.

Proposition 3. *If analysts are rational, they make no systematic error in predicting the log growth of earnings of HLTG_t and LLTG_t portfolios:*

$$\mathbb{E}(x_{t+h} - x_t - LTG_t^\theta | \text{HLTG}_t) = \mathbb{E}(x_{t+h} - x_t - LTG_t^\theta | \text{LLTG}_t) = 0 \quad \text{for } \theta = 0.$$

Under diagnostic expectations, in contrast, analysts systematically over-estimate growth of HLTG_t firms and under-estimate growth of LLTG_t firms:

$$\mathbb{E}(x_{t+h} - x_t - LTG_t^\theta | \text{HLTG}_t) < 0 < \mathbb{E}(x_{t+h} - x_t - LTG_t^\theta | \text{LLTG}_t) \text{ for } \theta > 0.$$

Under rational expectations, no systematic forecast error can be detected by an econometrician looking at the data (expectations are computed using the true steady state probability measure). Indeed, when $\theta = 0$, the average forecast within the many firms of the HLTG_t and LLTG_t portfolios is well calibrated to the respective means. Diagnostic expectations, in contrast, cause systematic errors. Firms in the HLTG_t group are systematically over-valued: analysts over-react to their pre-formation positive surprises, and form excessively optimistic forecasts of fundamentals.¹³ As a consequence, the realized earnings growth is on average below the forecast. Firms in the LLTG_t group are systematically under-valued. As a consequence, their realized earnings growth is on average above the forecast. As noted in Section III, systematic pessimism about LLTG firms is not borne out in the data.

Finally, our model yields the boom-bust LTG pattern in the HLTG group (and a reverse pattern in the LLTG group) documented in Figure 3. By Proposition 2, the improving pre-formation forecasts of HLTG firms are due to positive earnings surprises, while the deteriorating pre formation forecasts of LLTG firms are due to negative ones. But the model also predicts post formation reversals in LTG for both groups of firms. To see this, we compare $LTG_{i,t}$ forecasts made at t , with forecasts made for the same firm at $t + s$, namely $LTG_{i,t+s}$.

Proposition 4 *Under rational expectations ($\theta = 0$) we have that:*

¹³ In general, forecasters err not only by overestimating growth for HLTG firms but also by misclassifying firms as HLTG. Analysts miss out on firms that have high growth potential but whose recent performance is poor.

$$\mathbb{E}(\text{LTG}_{t+s}^\theta | \text{HLTG}_t) - \mathbb{E}(\text{LTG}_t^\theta | \text{HLTG}_t) < 0$$

Under diagnostic expectations ($\theta > 0$) we have that:

$$\mathbb{E}(\text{LTG}_{t+s}^\theta | \text{HLTG}_t) - \mathbb{E}(\text{LTG}_t^\theta | \text{HLTG}_t) = \mathbb{E}(\text{LTG}_{t+s}^{\theta=0} | \text{HLTG}_t) - \mathbb{E}(\text{LTG}_t^{\theta=0} | \text{HLTG}_t) - \theta\Psi$$

for some $\Psi > 0$. The opposite pattern, with reversed inequality and $\Psi < 0$, occurs for LLTG.

Mean reversion in LTG obtains under rational expectations, due to mean reversion in fundamentals. Under diagnostic expectations, however, mean reversion is amplified by the correction of initial forecast errors. Post-formation, the excess optimism of HLTG firms on average dissipates, causing a cooling off in expectations $\theta\Psi$ that is more abrupt than what would be implied by mean reversion alone. The cooling off of excess optimism arises because there are no news on average in the HLTG portfolio, which on average causes no overreaction. Likewise, the excess pessimism of LLTG firms dissipates, strengthening the reversal in that portfolio.

The assumption that the comparison group $-G$ immediately adapts to the recent state implies that forecast errors are on average corrected in one period. If $-G$ is slow moving, it takes longer for forecast errors to be corrected. As we show in Section VI, the horizon at which systematic forecast errors are corrected allows us to estimate the speed of adjustment of $-G$, which has important implications for the quantitative fit of the model, including return dynamics.

V.B. The diagnostic Kalman filter and returns

To study the pattern of returns, we take the required return $R > 1$ as given. The pricing condition for a firm i at date t is then given by:

$$\mathbb{E}_t^\theta \left(\frac{P_{i,t+1} + D_{i,t+1}}{P_{i,t}} \right) = R, \quad (8)$$

so that the stock price of firm i is the discounted stream of expected future dividends as of t . We assume that R is high enough that the discounted sum converges. By Equation (8), the equilibrium price at t is $P_{i,t} = \mathbb{E}_t^\theta(P_{i,t+1} + D_{i,t+1})/R$. Using this formula and Proposition 2, we find:

Proposition 5. *Denote by $R_{t,P}$ the realized return of portfolio $P = \text{HLTG}, \text{LLTG}$ at formation date t . Then, under the condition of Proposition 2, we have that:*

$$R_{t,\text{HLTG}} > R > R_{t,\text{LLTG}}$$

Because firms in the HLTG portfolio receive positive news before formation (Proposition 2), they earn higher returns than required. Our model thus yields positive abnormal pre-formation returns for HLTG stocks, as well as the low pre-formation returns of LLTG stocks, as in Figure 5. This effect arises also under rationality, $\theta = 0$. The key implication of representativeness is predictability of post-formation returns (Figures 1 and 5).

To see this, note that the average realized return going forward (according to the true probability measure) for a given firm i is given by:

$$\mathbb{E}_t \left(\frac{P_{i,t+1} + D_{i,t+1}}{P_{i,t}} \right) = \frac{\mathbb{E}_t(P_{i,t+1} + D_{i,t+1})}{\mathbb{E}_t^\theta(P_{i,t+1} + D_{i,t+1})} R, \quad (9)$$

which is below the required return R when at t investors over-value the future expected price and dividend of firm i . Conversely, the average realized return at $t + 1$ is higher than the required return R when at t investors under-value the firm's future expected price and dividend.

Diagnostic expectations thus yield the return predictability patterns of Figure 5.

Proposition 6. (Predictable Returns) Denote by $\mathbb{E}_t(R_{t+1}^\theta|P)$ the average future return of portfolio $P = HLTG, LLTG$ at $t + 1$. Under rationality, excess returns are not predictable:

$$\mathbb{E}_t(R_{t+1}^\theta|HLTG) = R = \mathbb{E}_t(R_{t+1}^\theta|LLTG) \quad \text{for } \theta = 0$$

Diagnostic expectations generate predictable excess returns:

$$\mathbb{E}_t(R_{t+1}^\theta|HLTG) < R < \mathbb{E}_t(R_{t+1}^\theta|LLTG) \quad \text{for } \theta > 0$$

Under rational expectations (i.e., for $\theta = 0$) realized returns may differ from the required return R for particular firms. However, the rational model cannot account for systematic return predictability in a large portfolio of firms sharing a certain forecast. Conditional on current information, rational forecasts are on average (across firms) correct and returns are unpredictable.

Under the diagnostic filter ($\theta > 0$), in contrast, the HLTG portfolio exhibits abnormally high returns up to portfolio formation and abnormally low returns after formation. The converse holds for the LLTG portfolio, just as we saw in Table 1 and Figure 1. Indeed, post-formation expectations systematically revert to fundamentals, and in particular investors are systematically disappointed in HLTG firms and their returns are abnormally low. The speed at which post-formation returns go back to R depends on the sluggishness of the comparison group – G .

VI. Overreaction to News and Quantification of Diagnostic Expectations

Diagnostic expectations yield a coherent account of the dynamics of news, analyst expectations, and returns documented in Section III. The predictability of returns, and especially of expectations errors, cannot be accounted for by Bayesian learning. Departures from rationality

are summarized by the parameter θ , which controls the extent of overreaction to information. This raises two questions. First, can overreaction be detected in expectations data? Second, can we quantitatively assess the explanatory power of our model? We next address these questions.

VI.A Overreaction to News

To assess overreaction, we need a measure of news at each t . Coibion and Gorodnichenko (2015) propose to use analysts' forecast revision at t , which summarizes all information received by the forecaster in this period. Over- or underreaction to information can then be assessed by correlating their forecast revision with the subsequent forecast error. When expectations over-react, a positive forecast revision indicates excessive upward adjustment. As a consequence, it should predict negative forecast errors (i.e. realizations below the forecast). Likewise, when expectations over-react, the correlation between revisions and forecast errors should be positive. Bouchaud et al. (2018) use this method to diagnose under-reaction to news about firms' profitability.

Proposition 7. *Assume the condition (7) of Proposition 2. Consider the firm level regression*

$$x_{i,t+h} - x_{i,t} - LTG_{i,t} = \alpha + \gamma(LTG_{i,t} - LTG_{i,t-k}) + v_{i,t+h}. \quad (10)$$

Under rationality, $\theta = 0$, the estimated γ is zero, while it is negative for $\theta > 0$.

Table 2 reports the estimates from the univariate regression of forecast error, defined as the difference between average growth over $h = 3, 4, 5$ years and current LTG, and the revision of

LTG over the past $k = 1, 2, 3$ years. To estimate γ we use consensus forecasts rather than individual analyst estimates because many analysts drop out of the sample.¹⁴

Table 2: Coibion-Gorodnichenko regressions for EPS

Each entry in the table corresponds to the estimated coefficient of the forecast errors $(\text{eps}_{t+n} / \text{eps}_t)^{1/n} \text{-LTG}_t$ for $n=3, 4$, and 5 on the variables listed in the first column of the table and year fixed-effects (not shown).

	Dependent Variable		
	$(\text{eps}_{t+3} / \text{eps}_t)^{1/3} \text{-LTG}_t$	$(\text{eps}_{t+4} / \text{eps}_t)^{1/4} \text{-LTG}_t$	$(\text{eps}_{t+5} / \text{eps}_t)^{1/5} \text{-LTG}_t$
LTG _t -LTG _{t-1}	-0.0351 (0.0734)	-0.1253 ^a (0.0642)	-0.1974 ^a (0.0516)
LTG _t -LTG _{t-2}	-0.2335 ^a (0.0625)	-0.2687 ^a (0.0602)	-0.2930 ^a (0.0452)
LTG _t -LTG _{t-3}	-0.2897 ^a (0.0580)	-0.2757 ^a (0.0565)	-0.3127 ^a (0.0437)

Consistent with diagnostic expectations, an upward LTG revision predicts excess optimism, pointing to overreaction to news. The estimated γ tends to become more negative and more statistically significant at longer forecast horizons $h = 3, 4, 5$ (moving from left to right in Table 2), perhaps reflecting the difficulty of projecting growth into the future.¹⁵

The estimated γ also gets higher in magnitude and more statistically significant as we lengthen the revision period $k = 1, 2, 3$ (moving from top to bottom in Table 2). These patterns are informative about $-G$. Under Diagnostic Expectations, the correlation between forecast revisions and forecast errors is maximized when the lag k at which the revision is computed is

¹⁴ Estimating (10) on the consensus LTG may misleadingly indicate under-reaction if individual analysts observe noisy signals, so that there is dispersion in their forecasts (see Coibion and Gorodnichenko 2015).

¹⁵ Bouchaud et al. (2018) find under-reaction of forecasts for the level of eps over 1 or 2 year horizons. In Appendix D (see Table D.2) we show that at these horizons there is some evidence of under-reaction in our data also. Over- and under-reaction may be reconciled by combining diagnostic expectation with some short run rigidity in analyst forecasts.

equal to the lag of $-G$.¹⁶ Table 2 shows that the correlation is highest at $k = 3$, which roughly indicates that $-G$ should be defined by a three-year lag.

VI.B Model Estimation with Simulated Method of Moments

To quantify the model's explanatory power, we use the magnitude of overreaction in Table 2 to estimate the diagnostic parameter θ using the simulated method of moments (SMM), and then assess how well the estimated model can account for the observed spread in average returns.

The dynamics of earnings and expectations of a representative firm¹⁷ depend on six parameters: the persistence and conditional variance of log earnings per share (b and σ_ϵ from Equation 1), those of fundamentals f (a and σ_η from Equation 2), the strength of representativeness θ , and the sluggishness s of the lagged expectations $-G$. Physical parameters $a, b, \sigma_\eta, \sigma_\epsilon$ are defined at the quarterly timescale, the natural scale of the data generating process.

We set the six parameters $(a, b, \sigma_\eta, \sigma_\epsilon, \theta, s)$ to match: the autocorrelation of log earnings per share ρ_l at lags $l = 1, 2, 3$, and 4 years, and the two coefficients γ_k estimated in Table 2 linking forecast error $\left(\frac{eps_{t+4}}{eps_t}\right)^{1/4} - LTG_t$ to forecast revision $LTG_t - LTG_{t-k}$ for $k = 1, 3$ years.

For each parameter combination $(a, b, \sigma_\eta, \sigma_\epsilon, \theta, s)$, we first simulate a time series of log earnings x_t and compute the associated diagnostic expectations \hat{f}_t^θ about fundamentals.¹⁸ From

¹⁶ At sluggishness s , the diagnostic expectation is $\mathbb{E}_t^\theta(x_{t+h}) = \mathbb{E}_t(x_{t+h}) + \theta[\mathbb{E}_t(x_{t+h}) - \mathbb{E}_{t-s}(x_{t+h})]$. It can be shown that the correlation between $x_{t+h} - \mathbb{E}_t^\theta(x_{t+h})$ and $\mathbb{E}_t^\theta(x_{t+h}) - \mathbb{E}_{t-k}^\theta(x_{t+h})$ reaches its maximum at $k = s$.

¹⁷ In Section VI.D we look at the cross section of firms to assess the comparative statics predicted by the kernel of truth. We leave it to future work to perform a fuller estimation with heterogeneous firms. One problem is that for many firms time series of annual data is short, and can have negative earnings (which are not considered in the model).

¹⁸ We use one time series because the empirical estimates ρ_l and γ_k were both computed using the pooled data.

the earnings time series we then compute the autocorrelations $\hat{\rho}_l = \frac{cov(x_t, x_{t-4l})}{var(x_t)}$ at lags $l = 1$ through 4 years. To compute the model implied coefficients $\hat{\gamma}_1, \hat{\gamma}_3$ we first generate the expectations for long term growth LTG_t at a horizon of $h = 16$ quarters, and then regress the forecast error $x_{t+16} - x_t - LTG_t$ on the forecast revision over one year, $LTG_t - LTG_{t-4}$, or over three years $LTG_t - LTG_{t-12}$ (Equation 10). This yields the vector:

$$v(a, b, \sigma_\eta, \sigma_\epsilon, \theta) = (\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3, \hat{\rho}_4, \hat{\gamma}_1, \hat{\gamma}_3).$$

We repeat this exercise for each parameter combination on a grid defined by $a, b \in [0,1]$, $\sigma_\eta, \sigma_\epsilon \in [0,0.5]$, $\theta \in [0,3]$ and $s \in \{1,2, \dots, 20\}$ quarters.¹⁹ We estimate the parameters by picking the combination that minimizes the Euclidean distance loss function

$$\ell(v) = \|v - \bar{v}\|$$

where \bar{v} is the vector of target moments estimated from the pooled data of all firms, given by:

$$\bar{v} = (0.82, 0.75, 0.70, 0.65, -0.276, -0.126).$$

The table below reports the average and standard deviation of the parameter combinations that yield the lowest value of the Euclidean loss function, across 30 independent runs.

Table 3: Estimates of model parameters

a	b	σ_η	σ_ϵ	θ	s
0.97	0.56	0.14	0.08	.90	11
(0.00)	(0.01)	(0.03)	(0.03)	(0.07)	(1.9)

¹⁹ The grid was defined in steps of 0.025 to 0.1 (depending on the parameter and the parameter region) for all parameters other than the timescale.

Fundamentals (and, to a lesser extent, log earnings) are estimated to be very persistent. The variance of fundamentals σ_η is estimated to be about 60% higher than that of transitory earnings σ_ϵ , leading to a Kalman gain of about 0.76.

Turning to the diagnostic parameter, our estimation yields a significantly positive θ . The value of .90 coincides with the estimate for the same parameter obtained by BGS (2018) in the context of credit spreads ($\theta = 0.91$), and is close to the average of 0.6 found in the context of forecasts of macroeconomic variables (BGMS 2018). A θ of 1 intuitively implies that the magnitude of forecast errors is comparable to the magnitude of news. In the current context, it implies a doubling of the signal to noise ratio, to $(1 + \theta)K \sim 1.44$. A diagnostic Kalman gain above 1 captures the key finding of overreaction, and systematic reversal, of the consensus LTG forecasts to both private and public news.²⁰

Finally, consistent with the intuition from Table 2, we estimate the timescale of diagnostic distortions s to be about 3 years. As we show below, this is consistent with the boom-bust pattern in expectations, and in returns, to occur over a time scale of three years. At this time scale the data does not strongly differentiate between the specification where news are assessed relative to past conditions or relative to lagged distorted beliefs.²¹

²⁰ To test the robustness of our estimate of θ , we evaluate the matching between the empirical moments \bar{v} and the predicted counterpart when θ varies in the interval $[0, 2]$ (keeping the other parameters constant). Figure E.1 in Appendix E shows that the loss function is minimized for $\theta = 0.9$ and that it increases fast as θ departs from this value. As expected, this is driven entirely by an increasingly worse match of the expectations moments $\hat{\gamma}_1, \hat{\gamma}_3$.

²¹ Our benchmark specification achieves a loss function minimum that is about 2% lower than that of the alternative specification. To see why the specifications yield similar results, examine the role of the time scale s on the distortions, which under the benchmark are proportional to $\theta(x_{i,t} - bx_{i,t-1} - a^s \hat{f}_{i,t-s})$. The longer the time scale, the smaller the difference in lagged forecasts, $a^s \theta(\hat{f}_{i,t-s} - \hat{f}_{i,t-s}^\theta)$ and the greater the variance in the actual shocks $x_{i,t} - bx_{i,t-1} - a^s \hat{f}_{i,t-s}$. Thus, for long time scales, the distortions entailed by the two specifications are similar. To check this intuition, in our simulation we regress diagnostic expectations relative to distorted beliefs on those under our benchmark specification, setting all parameters but the time scale to the values in Table 3. The explanatory power (R^2) of diagnostic expectations under the benchmark specification increases monotonically from 0.87 to 0.97 as the time scale increases from 1 to 11 quarters.

Using these estimated parameter values, we generate and report in Figure 6 the simulated versions of Figures 1 through 5 (panels 1 through 5). To do so, we simulate earnings paths $x_{i,t}$ and expectations of fundamentals $\hat{f}_{i,t}^\theta$ for 4000 firms (indexed by i) described by the parameters in Table 3, over a period of 35 years. Each year, we sort firms into deciles of long term growth expectations, $LTG_{i,t}$. We then compute, at the portfolio level, one year ahead returns as well as the dynamics of earnings, LTG, forecast errors and earnings announcement returns in the years around portfolio formation. To compute returns, we set the required rate of return R such that the average return matches the historical value-weighted market return of 9.7%.²²

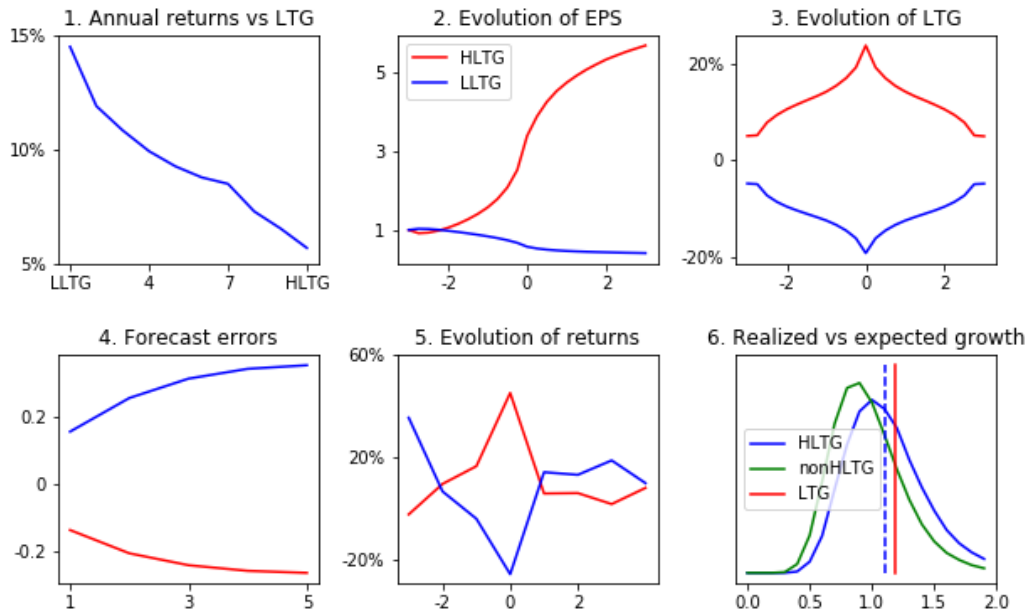


Figure 6. Simulation of the estimated model. Using the parameters in Table 3 we simulate 4000 firms over 140 time periods (quarters), generating time series of fundamentals, earnings, and growth expectations. The required rate of return R was set such that the average return matches the historical value-weighted market return of 9.7%. Every year (four periods), we sort firms on LTG forecasts. Panel 1 shows the average return spread 1 year post-formation across LTG deciles (compounding over the four quarters, then averaging over the portfolio). Finally we compute the geometric average of portfolio returns over time. Panels 2, 3, and 5 show the average EPS, LTG, and returns of the HLTG and LLTG portfolios from year $t-3$ to year $t+3$. Panel 4 shows the forecast error on HLTG and LLTG

²² In the simulation, average returns are higher than the required level of return, reflecting the fact that expectations are too volatile relative to fundamentals.

portfolios in the 5 years post-formation. Panel 6 shows the distribution of realized earnings growth after 5 years (annualized) for HLTG and for non-HLTG firms, together with the average long-term growth for HLTG firms (dotted blue line) and the LTG forecast for HLTG firms (solid red line, see Figure 9).

The model reproduces the main qualitative features of the data. In Panel 1, it reproduces the return spread between HLTG and LLTG stocks. In Panel 2, it reproduces the pattern of Figure 2 that pre-formation HLTG firms have fast growth, which then declines post formation. In Panel 3 the model reproduces the boom bust dynamics of LTG, with analysts' expectations becoming more optimistic pre-formation, and then reverting post-formation. Panel 4 reproduces the finding of Figure 4 of large forecast errors (excess optimism) for HLTG stocks. Panel 5 reproduces the boom-bust pattern in returns around portfolio formation. Returns for HLTG stocks are very high pre-formation, but then collapse below the required return R in the immediate post formation period, and eventually reverting back to their unconditional, long term value. The opposite happens with the return of LLTG stocks.

Panel 6 illustrates a central implication of the kernel of truth property, namely that HLTG firms have a fatter right tail of growth (and higher growth on average) than non-HLTG firms, but diagnostic expectations overestimate the prevalence of high growth firms in the HLTG portfolio. We examine this prediction empirically in Section VII.A (Figure 7).

The model fails to capture some qualitative features of the data, such as the negative forecast error for LLTG and the flat returns for firms with lower LTG expectations. We return to these issues when we summarize the results of the estimation.

We next assess the model's quantitative performance. For the cross sectional predictability of returns, we obtain an average LLTG-HLTG yearly return spread of 9.1% in year $t + 1$ (see Panel 1), with the LLTG and HLTG portfolios yielding respectively 14.7% and 5.6% average

returns. The empirical counterparts to these numbers are 15% and 3%, for a spread of 12% (Figure 1). This is a supportive and non-trivial finding: an estimation strategy using only earnings and expectations data provides a good fit for the evidence on spreads in returns.

The annualized levels of earnings growth (and expectations thereof) over a four-year horizon from the estimated model also provide a reasonable match to the data. The average yearly earnings growth for HLTG firms post formation is 11% (see Figure 7) in the data while in our model it is 9.8%. In turn, LTG for HLTG firms at formation is 39% while in the estimated model it is 22.5%.²³

We can also assess model performance with respect to the dynamics of earnings, expectations and returns around portfolio formation. As in Figure 3, growth expectations increase for three years pre-formation and decrease three years post formation, with most of the action happening in years $t - 1$ to $t + 1$. Crucially, event returns also exhibit a boom-bust pattern with an appropriate time scale: HLTG has strong returns for two years pre formation, while LLTG has low returns during this period. Post formation, HLTG returns are lower than LLTG returns for up to three years (though the trough happens in year $t + 1$ in the data and year $t + 3$ in the simulation). The fact that the time scale at which forecast errors are most predictable matches the time scale of abnormal returns suggests that predictability of returns is driven by expectation distortions.

Our overall assessment is that the model is able, both qualitatively and quantitatively, to account for several key features of the data, including the predictable return spread between HLTG and LLTG portfolios and the dynamics of expectations relative to the earnings process. At

²³ Expectations in our model are less exaggerated than equity analysts'. This is consistent with the fact that analysts expectations are on average too optimistic across the board, while in our model the average expectational error is zero. As noted before, the model predicts a positive forecast error for LLTG portfolio, which is not true in the data.

the same time, the model is very stylized, abstracting away from both firm and investor heterogeneity. These assumptions can be relaxed without compromising the tractability of the Diagnostic Kalman filter. An appropriate treatment of firm heterogeneity and variation in beliefs would likely help account for the features of the data which our model does not capture.

For example, the model exhibits too strong growth of EPS relative to Figure 2. Allowing for firm heterogeneity would improve the fit, because HLTG firms are disproportionately younger and smaller. Allowing for variation in size would lead to more reasonable estimates of growth, and would also capture asymmetries in performance between HLTG and LLTG firms. In turn, accounting for heterogeneity in investors' beliefs would capture the dispersion in LTG forecasts, which we abstract from here.

VII. Additional Predictions from Kernel of Truth

We next test some new cross-sectional predictions of the “kernel of truth” property of diagnostic expectations, the idea that expectations exaggerate true patterns in the data. This property allows to further distinguish diagnostic expectations from mechanical models of extrapolation such as adaptive expectations, in which individuals follow the ‘constant gain’ adaptive rule:

$$x_{t+h|t}^{ad} = x_{t+h|t-1}^{ad} + \mu(x_t - x_{t|t-1}^{ad}), \quad (11)$$

where $x_{t+h|t}^{ad}$ is the expectation held at t about the level of eps at $t + h$, x_t is the current realized level of eps, and $\mu \in [0,1]$ is a fixed coefficient.

Under adaptive expectations, beliefs depend only on surprising earnings ($x_t - x_{t|t-1}^{ad}$), but not on the forward looking data such as news, persistence, or volatility of the data generating process. We test the implications of the kernel of truth along three directions: i) the representativeness of future Googles in the HLTG group, ii) how the persistence and volatility of earnings modulate the LLTG-HLTG spread across different firms, and iii) the mechanics of expectations revisions.

VII.A Fat Tails and Representative Googles

Diagnostic Expectations account for excess optimism about HLTG firms as overreaction to the incidence of exceptional future performers, which are rare but representative. To assess this forward looking mechanism, Figure 7 plots the true distribution of future eps growth for the HLTG portfolio (blue curve) and the distribution of future eps growth of all other firms (orange curve).²⁴

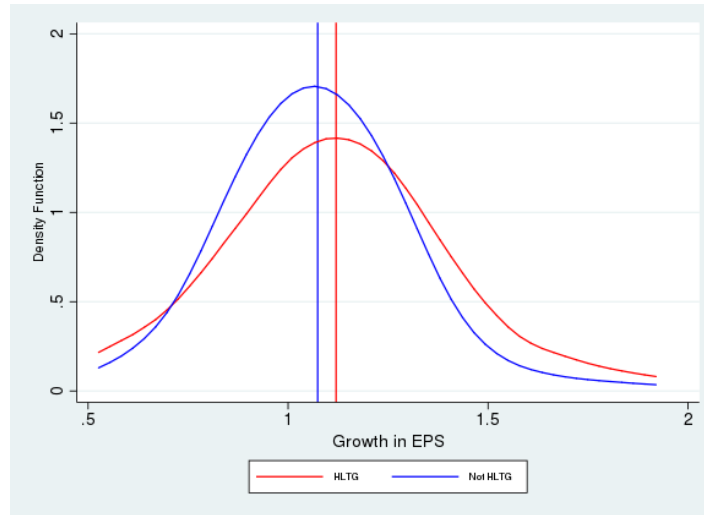


Figure 7. Kernel density estimates of growth in earnings per share for LTG Portfolios. In December of years (t) 1981, 1986, ..., and 2011, we form decile portfolios based on ranked analysts' expected growth in long term earnings

²⁴ The logic behind this cross-sectional test is that growth pre-formation positively predicts growth post formation, so non HLTG firms proxy for the performance of the HLTG portfolio under the counterfactual scenario of lower growth.

per share (*LTG*). For each stock, we compute the gross annual growth rate of earnings per share between t and $t+5$. We exclude stocks with negative earnings in year t and we estimate the kernel densities for stocks in the highest (*HLTG*) decile and for all other firms with *LTG* data. The graph shows the estimated density kernels of growth in earnings per share for stocks in the *HLTG* (blue line) and all other firms (orange line). The vertical lines indicate the means of each distribution (1.11 vs. 1.08, respectively). To compute earnings growth between t and $t+5$, we restrict the sample to firms having positive eps at these two dates.

Two findings stand out. First, *HLTG* firms have a higher average future eps growth than all other firms, as we saw in a somewhat different format in Figure 2. Second, and critically, *HLTG* firms display a fatter right tail of exceptional performers. Googles are representative for *HLTG* in the sense of definition (4). In fact, based on the densities in Figure 7, the most representative future growth realizations for *HLTG* firms are 40% to 60% annual growth (see Figure C.1 in Appendix C).²⁵

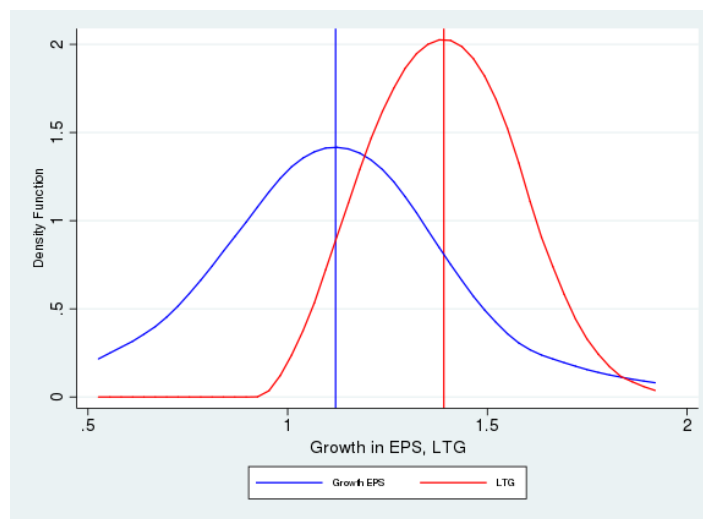


Figure 8. Realized vs. Expected Growth in eps. In December of years (t) 1981, 1986, ..., and 2011, we form decile portfolios based on ranked analysts' expected growth in long term earnings per share (*LTG*). We plot two series. First, we plot the kernel distribution of the gross annual growth rate in earnings per share between t and $t+5$. Second, we plot the kernel distribution of the expected growth in long term earnings at time t . The graph shows results for stocks in the highest decile of expected growth in long term earnings at time t . The vertical lines indicate the means of each distribution (1.11 vs. 1.39, respectively).

²⁵ Although *HLTG* firms tend to also have a slightly higher share of low performers, it is true that, as in our model, higher growth rates are more representative for *HLTG* firms. To compute earnings growth, Figure 7 is constructed using the subset of firms that have positive earnings at formation and five years after. The results are very robust to alternative assumptions, see Appendix C.

Diagnostic Expectations then predict that analysts should over-estimate the number of right-tail performers in the HLTG group. Figure 8 compares the distribution of future performance of HLTG firms (blue line) with the predicted performance for the same firms (red line).²⁶ It is indeed the case that analysts vastly exaggerate the share of exceptional performers, which are most representative of the HLTG group according to the true distribution of future eps growth. The kernel of truth can also shed light on the asymmetry between HLTG and LLTG firms. In Appendix C we show that future performance of LLTG firms tends to be concentrated in the middle, with a most representative growth rate of 0%. It is thus constant, rather than bad, performance that is representative of LLTG firms.²⁷ As a robustness check, we reproduce in Appendix C Figures 7 and 8 using two alternative measures of fundamentals (the change in earnings between t and $t + 5$ normalizing by the initial level of sales per share, and revenues minus cost of goods sold, which may be less noisy than eps). The evidence once again supports the kernel of truth hypothesis.

VII.B Volatility and Persistence of Earnings

Another prediction of kernel of truth is that the predictability of expectations errors and hence of returns should depend on true features of the data. In particular, it should depend on the volatility and the persistence of fundamentals. More volatile and more persistent fundamentals imply a higher signal to noise ratio of earnings, causing a larger overreaction to news. The LLTG-

²⁶ In making this comparison, bear in mind that analysts report point estimates of a firm's future earnings growth and not its full distribution (in our model, they report only the mean $f_{i,t}^\theta$ and not the variance σ_f^2). Thus, under rationality the LTG distribution would have the same mean but lower variance than realized eps growth.

²⁷ This can explain why expectations about LLTG firms and their market values are not overly depressed. One could capture this difference between HLTG firms (representative high growth) and LLTG firms (representative 0% growth) by relaxing the assumption of normality, or alternatively by allowing lower volatility for firms in the LLTG group.

HLTG spread should then be higher for firms with more volatile or persistent fundamentals. In Appendix E, Figure E.1 we illustrate these comparative statics in the context of a simulation.

We cannot estimate volatility and persistence firm by firm, because some firms are too short lived to allow a reliable computation. This is particularly true for HLTG firms, which tend to be young. To circumvent this problem, for each year's HLTG and LLTG portfolios, we match each firm to the group of firms in the same industry that are of comparable age. We require that each such group has at least 5 firms with at least 20 quarters of earnings data, which allows to estimate an average AR(1) coefficient for the earnings processes.²⁸ We then impute to each firm in the HLTG and LLTG portfolios the average estimated persistence of the matched group of firms and the average volatility of residuals.²⁹ We sort HLTG and LLTG firms into portfolios with firms in the top or bottom 30% of estimated persistence or of volatility estimated in this way. We compute the post formation returns for each portfolio, as well as the returns of the portfolio that is long on LLTG and short on HLTG in each of the persistence or volatility buckets. Table 4 reports the results.

Table 4: Return Spreads for Portfolios with Different Earnings Fundamentals

In December of each year between 1981 and 2015, we form decile portfolios based on ranked analysts' expected long-term growth in earnings per share and compute equally-weighted monthly return for decile portfolios. The table below reports annual average log returns for two-way sorts of extreme LTG portfolios (i.e. HLTG and LLTG), the portfolio that is long HLTG and short LLTG, and various portfolios formed on the basis of: (1) the average persistence of growth in EPS for firms in the same industry and year, and (2) the average variance of the prediction errors from regressing eps_t on its first lag for firms in the same industry and year. We measure persistence based on the AR(1) coefficient of eps_t on its lagged value in rolling regressions using 20 quarterly observations for years $t+5$ through t . Prediction errors are based on that regression. Finally, we form persistence and volatility portfolios based on the average persistence and volatility for stocks in the same industry and year. The "bottom 30%" is the portfolio of stocks in the bottom three deciles of persistence or volatility, as the case may be. Similarly, the "top 30%" is the portfolio of stocks in the top three deciles of persistence or volatility, as the case may be. The table also reports t-stats for the portfolio that is long stocks in the top 30% of persistence and short stocks in the bottom 30% of persistence.

Persistence of growth in EPS	Vol of prediction error of growth in EPS
------------------------------	--

²⁸ For some groups this condition is not satisfied, in which case we can either group based on year and firm age alone, or drop the requirement of 5 observations. Here we use the first method, but both give similar results.

²⁹ Data limitations do not allow us separate fundamentals and noise for individual firms in a Kalman filter estimation.

	Bottom 30%	Top 30%	t-stat Top- Bottom	Bottom 30%	Top 30%	t-stat Top- Bottom
LLTG	15.0%	12.8%	-0.73	14.9%	17.3%	0.80
HLTG	7.8%	0.9%	-2.46	4.9%	0.6%	-1.82
HLTG-LLTG	-7.2%	-11.5%	-2.05	-9.8%	-16.6%	-2.42

The average spread between LLTG and HLTG portfolios exhibits a significant increase from 7.2% (that is, 15%-7.8%) to 11.9% (that is, 12.8%-0.9%) as firms move from the bottom 30% to the top 30% of estimated persistence of earnings growth. The returns of a portfolio that is long on LLTG and short on HLTG stocks exhibits a similar increase, from 7.2% to 11.5%. A very similar pattern (10% vs 16.6%) is present when sorting on volatility of earnings growth. Consistent with the kernel of truth, return spreads are predictable in terms of the firms' underlying fundamental earnings process.

In sum, the evidence is in line with the key prediction that patterns in returns are predictable from fundamentals. Expectations link fundamentals to predictable returns. Together with our quantification, these results provide further support to the claim that predictability of returns is driven by forecast errors that exhibit overreaction.

VII.C Revisions of Expectations

The last kernel of truth prediction that we test concerns revisions of expectations. Under diagnostic expectations, believed fundamentals are revised from one period to the next by:

$$\begin{aligned} \hat{f}_{i,t+1}^\theta - \hat{f}_{i,t}^\theta &= K(1 + \theta)(x_{i,t+1} - bx_{i,t} - a\hat{f}_{i,t}^\theta) \\ &\quad - (1 - a)\hat{f}_{i,t} - K\theta(1 - aK(1 + \theta))(x_{i,t} - bx_{i,t-1} - a\hat{f}_{i,t-1}^\theta). \end{aligned}$$

Leaving aside current news (the term in the first line), beliefs about fundamentals are updated due to mean reversion of fundamentals (i.e. $-(1 - a)\hat{f}_{i,t}$), but also to the waning of the

overreaction to previous shocks (i.e., $x_{i,t} - bx_{i,t-1} - af_{i,t-1}$). For HLTG stocks, both forces point to a downward revision of beliefs regardless of the current news received, while for LLTG stocks the opposite holds, leading to systematic mean reversion of LTG forecasts.

Under adaptive expectations, in contrast, analysts revise growth expectations downward if and only if bad news arrive, that is, if $(x_t - x_{t|t-1}^{ad}) < 0$. This mechanical rule of expectations formation does not take mean reversion of fundamentals into account.

To see which theory better describes the data, we assess how LTG changes around earnings announcement dates, as a function of the surprise in earnings. Take a firm with Wall Street Journal announcement dates wsj_{t-1} and wsj_t for earnings in fiscal years $t-1$ and t , respectively. We compute earnings surprise at wsj_t as the difference between the realized earnings per share in (fiscal) year t announced at wsj_t and the eps forecast for time t made soon after the previous announcement, namely on day $wsj_{t-1}+45$. We then compute the change in LTG between $wsj_{t-1}+45$ and wsj_t+45 , which already incorporates the revision in response to the announcement at wsj_t .

To better visualize, we normalize earnings surprise by the stock price when that forecast was made, and then rank observations into surprise deciles. The results are reported in Figure 9 below.

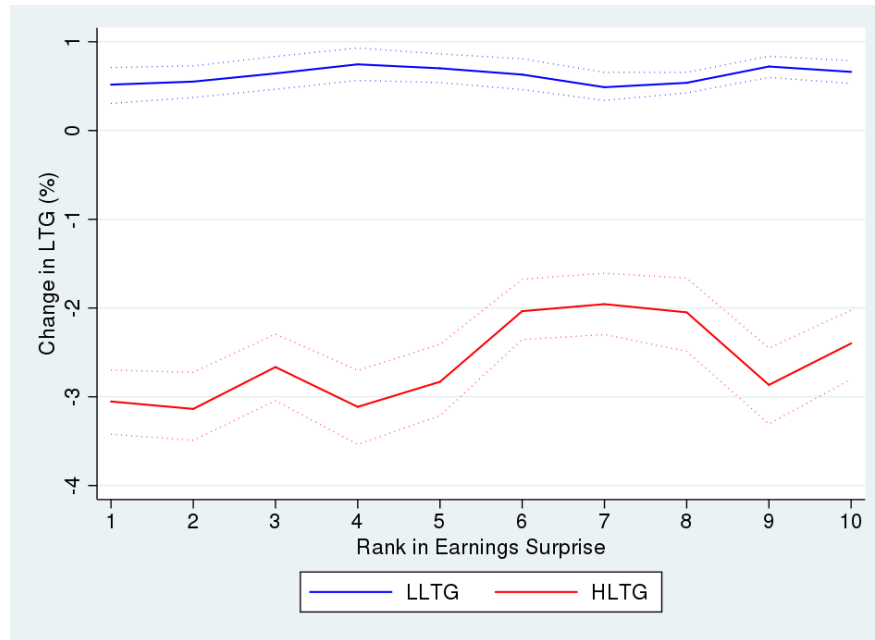


Figure 9. Evolution of Analysts' Beliefs in Response to Earnings' Announcements. For each analyst j , firm i , and fiscal year t , we compute the difference between the LTG forecasts made 45 following the announcement of earnings in : (a) fiscal year t (wsj_{t+45}) and (b) fiscal year $t-1$ (wsj_{t-1+45}). We measure forecasting errors for earnings based on forecasts 45 days following the announcement for earnings in fiscal year $t-1$ (wsj_{t-1+45}). We rank observations into deciles based on the ratio of the forecasting error for earnings per share in fiscal year t to the stock price when that forecast was made (wsj_{t-1+45}). The Figure reports the bootstrapped mean LTG change between wsj_{t-1+45} and wsj_{t+45} for the HLTG and LLTG portfolios. The dotted lines indicate 5th and 95th confidence levels determined via bootstrapping.

The data show strong evidence of systematic mean reversion. Regardless of the actual earnings surprise, long term growth expectations about HLTG firms (in red) deteriorate on average by 2.6% (Rafael check) while those about LLTG ones (in blue) improve on average by 0.6%. The reversal in Figure 9 is not simply due to the fact that HLTG firms on average receive bad surprises and LLTG firms on average receive good surprises. Rather, even HLTG firms that experience positive earnings surprises are downgraded, and even LLTG firms that experience negative earnings surprises are upgraded. These findings are puzzling from the perspective of

adaptive expectations, but are consistent with the forward-looking nature of diagnostic expectations.³⁰ In Appendix E, we show that the same pattern emerges in our estimated model.³¹

Overall, the data offer support for the kernel of truth prediction of diagnostic expectations, but are difficult to reconcile with a mechanical model of adaptive expectations. Another significant alternative to diagnostic expectations is the BSV (1998) model of investor sentiment. BSV is motivated by representativeness but does not model it explicitly. Rather, it assumes investors hold dogmatic, yet incorrect, priors in a way that is designed to produce excess optimism after a string of good news: while the true process driving a firm’s earnings is a random walk, analysts perform Bayesian updating across two incorrect models, one where earnings are believed to trend and one where they mean revert. Overreaction occurs because fast earnings growth leads the analyst to attach too high a probability that the firm is of the “trending type”, even though no firm is actually trending.

Diagnostic expectations capture this key intuition from BSV: after good performance analysts place disproportionate weight on strong fundamentals, and the reverse after bad performance. In our view, the new model has two main advantages relative to its antecedent. First, it seeks to provide a more explicit model of representativeness. The kernel of truth property, namely the hypothesis that beliefs overreact to objective information, is key for the model’s

³⁰ In Appendix D, we show that adaptive expectations predict no over-reaction to news after the persistence of the earnings process is accounted for. After controlling for current levels x_t , the adaptive forecast revision ($x_{t+1}^a - x_t^a$) should positively predict forecast errors as in the under-reaction models. In contrast, diagnostic expectations over-react to news regardless of the persistence of the data generating process.

³¹ Expectations may be formed, at least in part, by rational learning from prices. Suppose that investors’ required return is unobservable and follows a mean-reverting process. Price increases then signal an improvement in fundamentals but also a decrease in investors’ required return, so that expectations of earnings growth and expectations of returns should be negatively correlated. To test this prediction, we construct a measure of analysts’ expectations of returns by gathering IBES data on the projected price level forecasted by analysts within a 12-month horizon. We define target returns as the ratio of the mean target price across all analysts following the stock to the current stock price. Using this measure, the correlation between analysts’ expectations of long-term growth (LTG) and their expected returns is 0.23 (significant at the 1% level), in contrast to the prediction above. Our model does not provide a meaningful counterpart to this finding, as (diagnostic) expectations of returns are constant and equal to R .

generality and tractability, and generates provides novel predictions for the cross section of stock returns which is supported in the data. Second, and relatedly, diagnostic expectations are portable across different domains: unlike BSV they are not designed for a specific finance setting such as learning about a firm's earnings growth, and so can be easily applied to macroeconomic expectations, laboratory experiments, or social stereotyping. This allows us to assess and quantify the kernel of truth distortions in a variety of contexts, and to identify robust departures from rational expectations.

VIII. Conclusion.

This paper revisits what since Shiller (1981) has been perhaps the most basic challenge to rational asset pricing: over-reaction to news and the resulting excess volatility and mean reversion. We investigate this phenomenon in the context of individual stocks, for which we have extensive evidence on security prices, fundamentals, but also -- crucially -- expectations of future fundamentals. La Porta (1996) has shown empirically that securities whose long-term earnings growth analysts are most optimistic about earn low returns going forward. Here we propose a theory of belief formation that delivers this finding, but also provides a characterization of joint evolution of fundamentals, expectations, and returns that can be taken to the data.

A central feature of our theory is that investors are forward looking, in the sense that they react to news. However, their reaction is distorted by representativeness, the fundamental psychological principle that people put too much probability weight on states of the world that the news they receive is most favorable to. In psychology, this is known as the kernel of truth hypothesis: people react to information in the right direction, but too strongly. We call such belief formation diagnostic expectations, and show that a theory of security prices based on this model

of beliefs can explain not just previously documented return anomalies, but also the joint evolution of fundamentals, expectations, and returns.

The theory is portable in the sense that the same model of belief distortions has been shown to work in several other contexts. At the same time, the model can be analyzed using a variation of Kalman Filter techniques used in models of rational learning. The theory yields a number of strong empirical predictions, which have not been considered before, but which we have brought to the data. These predictions in particular distinguish the theory from adaptive expectations, and show that investors and analysts are forward looking in forming their beliefs. The model can also be estimated using the method of moments, and does a reasonable job in fitting several moments of the data. Significantly for our purposes, the critical representativeness parameter we estimate – our measure of overreaction – is comparable to estimates we obtained with very different data sets.

Of course, this is just a start. Our approach to expectation formation can be taken to other contexts, most notably aggregate stock prices but also macroeconomic time series. We have focused on distortions of beliefs about the means of future fundamentals, but the kernel of truth idea could be applied to thinking about other moments as well, such as variance or skewness. We hope to pursue these ideas in future work, but stress what we see as the central point: the theory of asset pricing can incorporate fundamental psychological insights while retaining the rigor and the predictive discipline of rational expectations models. And it can explain the data not just on the joint evolution of fundamentals and security prices, but also on expectations, in a unified dynamic framework. Relaxing the rational expectations assumption does not entail a loss of rigor; to the contrary it allows for a disciplined account of additional features of the data. An estimation

exercise suggests, moreover, that the model can replicate several quantitative, and not just qualitative, features of the data.

References

- Adam, Klaus, and Albert Marcet, and Juan Pablo Nicolini. 2016. "Stock market volatility and learning." *Journal of Finance* 71(1), 33 – 82.
- Adam, Klaus, Albert Marcet, and Johannes Beutel. 2017. "Stock price booms and expected capital gains." *American Economic Review*, 107 (8), 2352-2408.
- Barberis, Nicholas, and Andrei Shleifer. 2003. "Style investing." *Journal of Financial Economics* 68, 161-199.
- Barberis, Nicholas, Andrei Shleifer, and Robert Vishny. 1998. "A Model of investor sentiment." *Journal of Financial Economics* 49 (3), 307–343.
- Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer. 2015. "X-CAPM: an extrapolative capital asset pricing model." *Journal of Financial Economics* 115 (1), 1-24.
- Barsky, Robert and J. Bradford DeLong. 1993. "Why does the stock market fluctuate?" *Quarterly Journal of Economics* 108(2), 291-311.
- Ben David, Yitzhak, John Graham, and Campbell Harvey. 2013. "Managerial miscalibration." *Quarterly Journal of Economics* 128 (1), 547–84.
- Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. 2016. "Stereotypes." *Quarterly Journal of Economics* 131 (4), 1753-1793.
- Bordalo, Pedro, Nicola Gennaioli and Andrei Shleifer. 2018. "Diagnostic expectations and credit cycles." *Journal of Finance* 73 (1), 199-227.
- Bordalo, Pedro, Nicola Gennaioli, Yueran Ma and Andrei Shleifer. 2018. "Overreaction in macroeconomic expectations." Harvard working paper.
- Bouchaud, Jean-Philippe, Philipp Krueger, Augustin Landier, and David Thesmar. 2018. "Sticky expectations and the profitability anomaly." *Journal of Finance*, forthcoming.
- Casscells, Ward, Arno Schoenberger, and Thomas Graboys. 1978. "Interpretation by physicians of clinical laboratory results." *New England Journal of Medicine* 299 (18), 999–1001.
- Coibion, Olivier and Yuriy Gorodnichenko 2015. "Information rigidity and the expectations formation process: a simple framework and new facts" *American Economic Review* 105(8), 2644–2678.
- Cutler, David, James Poterba, and Lawrence Summers. 1990. "Speculative dynamics and the role of feedback traders." *American Economic Review* 80 (2), 63–68.
- . 1991. "Speculative dynamics." *Review of Economic Studies* 58 (3), 529–46.

- Daniel, Kent, David Hirshleifer, and Avanidhar Subramanyam. 1998. "Investor psychology and security market under- and over-reaction." *Journal of Finance* 53(6), 1839 – 1885.
- Daniel, Kent, Alexander Klos, and Simon Rottke. 2017. "Overconfidence, information diffusion, and mispricing persistence." Working paper.
- De Bondt, Werner and Richard Thaler. 1985. "Does the stock market overreact?" *The Journal of Finance* 40(3), 793-805.
- De Bondt, Werner and Richard Thaler. 1987, "Further evidence on investor overreaction and stock market seasonality." *The Journal of Finance* 42(3), 557-81.
- Dechow, Patricia, and Richard Sloan, 1997, "Returns to contrarian investment strategies: tests of naive expectations hypotheses." *Journal of Financial Economics* 43 (1), 3-27
- Dechow, Patricia, Amy Hutton, and Richard Sloan, 2000, "The relation between analysts' forecasts of long-term earnings growth and stock price performance following equity offerings." *Contemporary Accounting Research* 17(1), 1-32.
- DeLong, J. Bradford, Andrei Shleifer, Lawrence Summers, and Robert Waldmann 1990a. "Positive feedback investment strategies and destabilizing rational speculation." *Journal of Finance* 45(2), 379–95.
- DeLong, J. Bradford, Andrei Shleifer, Lawrence Summers, and Robert Waldmann 1990b. "Noise trader risk in financial markets." *Journal of Political Economy* 98(4), 703–738.
- Dominguez, Kathryn. 1986. "Are foreign exchange forecasts rational?: New evidence from survey data." *Economics Letters* 21 (3), 277–81.
- Easterwood, John, and Stacey Nutt, 1999, "Inefficiency in analysts' earnings forecasts: systematic misreaction or systematic optimism?" *Journal of Finance* 54 (5), 1777-1797.
- Fama, Eugene and Kenneth French. 1993. "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics* 33(1), 3 – 56.
- Frankel, Jeffrey, and Kenneth Froot. 1987. "Using survey data to test standard propositions regarding exchange rate expectations." *American Economic Review* 77 (1), 133–53.
- . 1988. "Explaining the demand for dollars: international rates of return, and the expectations of chartists and fundamentalists." In *Macroeconomics, agriculture, and the exchange rate*. Eds. R. G. Chambers and P. L. Paarlberg. Boulder, CO: Westview Press.
- Fuster, Andreas, David Laibson, and Brock Mendel. 2010. "Natural expectations and Macroeconomic Fluctuations." *Journal of Economic Perspectives* 24 (4), 67-84.
- Gennaioli, Nicola, Yueran Ma, and Andrei Shleifer. 2016. "Expectations and investment." *NBER Macroeconomics Annual* 30, 379-442.

- Gennaioli, Nicola, and Andrei Shleifer. 2010. "What Comes to mind." *Quarterly Journal of Economics* 125 (4), 1399–1433.
- Glaeser, Edward, and Charles Nathanson. 2015. "Housing bubbles." *Handbook of Regional and Urban Economics* 5, 701–751.
- Greenwood, Robin, and Andrei Shleifer. 2014. "Expectations of returns and expected returns." *Review of Financial Studies* 27 (3), 714–746.
- Hong, Harrison and Jeremy Stein. 1999. "A unified theory of underreaction, momentum trading, and overreaction in asset markets." *Journal of Finance* 56 (6), 2143 – 2184.
- Kahneman, Daniel, and Amos Tversky. 1972. "Subjective probability: a judgment of representativeness." *Cognitive Psychology* 3 (3), 430–454.
- Kahneman, Daniel, and Amos Tversky. 1973. "On the psychology of prediction." *Psychological Review* 80 (4), 237–251.
- Kozak, Serhiy, Stefan Nagel and Shrihari Santosh. 2018. "Interpreting factor models." *Journal of Finance* 73 (3): 1183 - 1223.
- Lakonishok, Joseph, Andrei Shleifer, and Robert Vishny. 1994. "Contrarian investment, extrapolation, and risk." *Journal of Finance* 49 (5), 1541–78.
- La Porta, Rafael. 1996. "Expectations and the cross-section of stock returns." *Journal of Finance* 51 (5), 1715-1742.
- La Porta, Rafael, Josef Lakonishok, Andrei Shleifer, and Robert Vishny. 1997. "Good news for value stocks: further evidence on market efficiency." *Journal of Finance* 52(2), 859–874.
- Marcet, Albert and Thomas Sargent. 1989. "Convergence of least squares learning mechanisms in self-referential linear stochastic models." *Journal of Economic Theory* 48 (2), 337 – 368.
- Marcet, Albert and Thomas Sargent. 1989. "Convergence of least-squares learning in environments with hidden state variables and private information." *Journal of Political Economy* 97(6), 1306-1322.
- Michaely, Roni, and Kent Womack. 1999. "Conflict of interest and the credibility of underwriter analyst recommendations." *Review of Financial Studies* 12 (4), 653-686.
- Odean, Terrance. 1998. "Volume, volatility, price, and profit when all traders are above average." *Journal of Finance* 53 (6), 1887-1934.
- Pastor, Lubos and Pietro Veronesi. 2003. "Stock valuation and learning about profitability." *The Journal of Finance* 58 (5), 1749-1789.
- Pastor, Lubos and Pietro Veronesi. 2005. "Rational IPO waves." *The Journal of Finance* 60 (4): 1713-1757.
- Pastor, Lubos and Pietro Veronesi. 2009. "Learning in financial markets." *Annual Review of Financial Economics* 1(1): 361-381.

- Rabin, Matthew. 2013. "An approach to incorporating psychology into economics." *American Economic Review* 103 (3): 617-22.
- Shiller, Robert. 1981. "The use of volatility measures in assessing market efficiency." *Journal of Finance* 36(2): 291 – 304.
- Sims, Christopher. 2003. "Implications of rational inattention." *Journal of Monetary Economics* 50 (3): 665–690.
- Skinner, Douglas and Richard Sloan. 2002. "Earnings surprises, growth expectations, and stock returns or don't let an earnings torpedo sink your portfolio." *Review of Accounting Studies* 7 (2), 289-312.
- Tversky, Amos, and Daniel Kahneman. 1974. "Judgment under uncertainty: heuristics and biases," *Science* 185 (4157), 1124-1131.
- Tversky, Amos, and Daniel Kahneman. 1983. "Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment." *Psychological Review* 90 (4): 293-315.

Internet Appendix.

A. Proofs

Proposition 1. Upon observing $g_{i,t} \equiv x_{i,t} - bx_{i,t-1}$, the analyst's believed distribution of firm fundamentals is given by:

$$h^\theta(f, g_{i,t}) = h(f, g_{i,t}) \cdot [R(f, g_{i,t})]^\theta \cdot Z$$

where $Z^{-1} = \int h(f, g_{i,t}) \cdot [R(f, g_{i,t})]^\theta \cdot df$ and

$$R(f, g_{i,t}) = \exp \left\{ \frac{(\hat{f}_{i,t} - a\hat{f}_{i,t-1})(2f - a\hat{f}_{i,t-1} - \hat{f}_{i,t})}{2\sigma_f^2} \right\}.$$

We expand the above expression using the assumption that $h(f, g_{i,t})$ is normally distributed with variance σ_f^2 and mean:

$$\hat{f}_{i,t} = a\hat{f}_{i,t-1} + K(g_{i,t} - a\hat{f}_{i,t-1})$$

We find:

$$h^\theta(f, g_{i,t}) = Z \cdot \exp \frac{1}{2\sigma_f^2} \left\{ -(f - \hat{f}_{i,t})^2 + \theta(\hat{f}_{i,t} - a\hat{f}_{i,t-1})(2f - a\hat{f}_{i,t-1} - \hat{f}_{i,t}) \right\}$$

The exponent then reads:

$$\begin{aligned} & -(f - \hat{f}_{i,t})^2 + \theta(\hat{f}_{i,t} - a\hat{f}_{i,t-1})(2f - a\hat{f}_{i,t-1} - \hat{f}_{i,t}) \\ &= -\left(f - \left(\hat{f}_{i,t} + \theta(\hat{f}_{i,t} - a\hat{f}_{i,t-1})\right)\right)^2 + c(\hat{f}_{i,t}, \hat{f}_{i,t-1}) \end{aligned}$$

where $c(\hat{f}_{i,t}, \hat{f}_{i,t-1})$ is a constant (does not depend on f). Taking normalization into account, we find

$$h^\theta(f, g_{i,t}) = \frac{1}{\sqrt{2\pi\sigma_f^2}} e^{-\frac{(f - (\hat{f}_{i,t} + \theta(\hat{f}_{i,t} - a\hat{f}_{i,t-1})))^2}{2\sigma_f^2}}$$

Using Equation (3) for the Bayesian expectation $\hat{f}_{i,t}$, the mean of this distribution can be written:

$$\hat{f}_{i,t}^\theta = \hat{f}_{i,t} + \theta(\hat{f}_{i,t} - a\hat{f}_{i,t-1}) = a\hat{f}_{i,t-1} + K(1 + \theta)(g_{i,t} - a\hat{f}_{i,t-1})$$

■

Proposition 2. Denote by $\lambda_H > 0$ the threshold in expected growth rate above which a firm is classified as HLTG (i.e., it is in the top decile). From the definition of LTG in Section V, firm i is classified as HLTG at time t provided:

$$LTG_{i,t} = -\varphi_h x_{i,t} + \vartheta_h \hat{f}_{i,t}^\theta \geq \lambda_H$$

where we have defined $\varphi_h \equiv (1 - b^h)$ and $\vartheta_h \equiv a^h \frac{1-(b/a)^h}{1-(b/a)}$. This can be written as:

$$-\varphi_h x_{i,t} + \vartheta_h a(1 - K')\hat{f}_{i,t-1} + \vartheta_h K'(1 - b)x_{i,t-1} + \vartheta_h K'(x_{i,t} - x_{i,t-1}) \geq \lambda_H,$$

where $K' \equiv K(1 + \theta)$. The left hand side of the above condition is a linear combination of mean zero normally distributed random variables. Denote it by $LHS_{i,t}$. By linear regression, the average growth rate $x_{i,t} - x_{i,t-1}$ experienced by firms whose $LHS_{i,t}$ is equal to λ is given by:

$$\mathbb{E}[x_{i,t} - x_{i,t-1} | LHS_{i,t} = \lambda] = \frac{cov(x_{i,t} - x_{i,t-1}, LHS_{i,t})}{var(LHS_{i,t})} \lambda.$$

Because for HLTG firms $\lambda \geq \lambda_H > 0$, their pre-formation growth is positive, $\mathbb{E}[x_{i,t} - x_{i,t-1} | LHS_{i,t} = \lambda] > 0$, provided $cov(x_{i,t} - x_{i,t-1}, LHS_{i,t}) > 0$. This occurs when the expression:

$$[K'\vartheta_h(1+b) - \varphi_h] \left(\text{var}(x_{i,t}) - \text{cov}(x_{i,t}, x_{i,t-1}) \right) + \vartheta_h a(1-K') \text{cov}(x_{i,t} - x_{i,t-1}, \hat{f}_{i,t-1})$$

is positive. For convenience, rewrite this as:

$$A[\text{var}(x_{i,t}) - \text{cov}(x_{i,t}, x_{i,t-1})] + B \text{cov}(x_{i,t} - x_{i,t-1}, \hat{f}_{i,t-1})$$

It is useful to rewrite the first term as:

$$A[(1-b)\text{var}(x_{i,t}) - a \cdot \text{cov}(f_{i,t-1}, x_{i,t-1})] = A \left[\frac{\text{var}(\epsilon)}{1+b} + \frac{\text{var}(f_{i,t})}{1-ab} \frac{1-a}{1+b} \right]$$

where we used $\text{var}(x_{i,t}) = \frac{1}{1-b^2} \left[\text{var}(\epsilon) + \frac{1+ab}{1-ab} \text{var}(f_{i,t}) \right]$. The second term reads:

$$\begin{aligned} & \text{cov}(x_{i,t} - x_{i,t-1}, a(1-K)\hat{f}_{i,t-2} + K(f_{i,t-1} + \epsilon_{i,t-1})) \\ &= a(1-K)\text{cov}(x_{i,t} - x_{i,t-1}, \hat{f}_{i,t-2}) - (1-b)K\text{cov}(x_{i,t-1}, f_{i,t-1}) \\ & \quad - (1-b)K\text{var}(\epsilon) + aK\text{var}(f_{i,t}) \end{aligned}$$

We can show that:

$$\text{cov}(x_{i,t} - x_{i,t-1}, \hat{f}_{i,t-2}) > a^2 \text{cov}(f_{i,t}, \hat{f}_{i,t-2}) - b(1-b)\text{cov}(x_{i,t-2}, f_{i,t-2}) - a(1-b)\text{var}(f_{i,t})$$

where we used $\text{cov}(x_{i,t-1}, f_{i,t-2}) > \text{cov}(x_{i,t-1}, \hat{f}_{i,t-2})$. Thus:

$$\begin{aligned} & \text{cov}(x_{i,t} - x_{i,t-1}, a(1-K)\hat{f}_{i,t-2} + K(f_{i,t-1} + \epsilon_{i,t-1})) \\ & > \text{var}(f_{i,t}) \left[\frac{aK}{1-a^2(1-K)} - (1-b) \left(K + a^2(1-K) + \frac{ab}{1-ab} \right) \right] \\ & \quad - (1-b)K\text{var}(\epsilon) \end{aligned}$$

So putting the two terms together we find

$$\begin{aligned}
& var(\epsilon) \left[\frac{A}{1+b} - (1-b)KB \right] \\
& + var(f_{i,t}) \left[\frac{A}{1+b} \frac{1-a}{1-ab} \right. \\
& \left. + B \left[\frac{aK}{1-a^2(1-K)} - (1-b) \left(K + a^2(1-K) + \frac{ab}{1-ab} \right) \right] \right]
\end{aligned}$$

A sufficient condition that makes both terms positive is:

$$b^h + \vartheta_h \left[\frac{K' - a}{1-a} \right] > 1$$

This condition is easier to satisfy for low mean reversion (large b), large signal to noise ratio (large K) and for strong overreaction (large θ). It is trivially satisfied when $b = 1$ provided $K(1 + \theta) \geq a$ (which holds in our estimation). ■

Proposition 3. From the law of motion of earnings we have that:

$$\mathbb{E}(x_{i,t+h} - x_{i,t} | HLTG_t) = \mathbb{E}(-\varphi_h x_{i,t} + \vartheta_h f_{i,t} | HLTG_t)$$

Because rational estimation errors $u_{i,t} \equiv f_{i,t} - \hat{f}_{i,t}$ are on average zero, we also have that:

$$\mathbb{E}(x_{i,t+h} - x_{i,t} | HLTG_t) = \mathbb{E}(-\varphi_h x_{i,t} + \vartheta_h \hat{f}_{i,t} | HLTG_t).$$

This implies that the average forecast error entailed in LTG is equal to:

$$\begin{aligned}
& \mathbb{E}(x_{i,t+h} - x_{i,t} - LTG_{i,t} | HLTG_t) = \vartheta_h \mathbb{E}(\hat{f}_{i,t} - \hat{f}_{i,t}^\theta | HLTG_t) = \\
& -\vartheta_h K\theta \mathbb{E}(x_{i,t} - bx_{i,t-1} - a\hat{f}_{i,t-1} | HLTG_t) = -\vartheta_h K\theta \mathbb{E}(\eta_{i,t} + \epsilon_{i,t} | HLTG_t).
\end{aligned}$$

The expectation $\mathbb{E}(\eta_{i,t} + \epsilon_{i,t} | HLTG_t)$ is positive because HLTG firms have positive recent performance (see Lemma 1). Under rationality, $\theta = 0$, forecast errors are unpredictable. Under diagnostic expectations, $\theta > 0$, forecast errors are predictably negative for the HLTG group. Conversely, the same argument shows that they are predictably positive for the LLTG group. ■

Proposition 4. The average LTG at future date $t + s, s \geq 1$, in the HLTG group is equal to:

$$\begin{aligned} \mathbb{E}(LTG_{i,t+s} | HLTG_t) &= \mathbb{E}(-\varphi_h x_{i,t+s} + \vartheta_h \hat{f}_{i,t+s}^\theta | HLTG_t) = \\ \mathbb{E}(-\varphi_h x_{i,t+s} + \vartheta_h \hat{f}_{i,t+s} + \vartheta_h (\hat{f}_{i,t+s}^\theta - \hat{f}_{i,t+s}) | HLTG_t) &= \\ \mathbb{E}(-\varphi_h x_{i,t+s} + \vartheta_h \hat{f}_{i,t+s} + \vartheta_h K\theta (g_{i,t+s} - a\hat{f}_{i,t+s-1}) | HLTG_t) &= \\ -\varphi_h b^s x_{i,t} + \vartheta_h a^s \hat{f}_{i,t}, \end{aligned}$$

where the last equality follows from the fact that within the HLTG group of stocks, $g_{i,t+s} - a\hat{f}_{i,t+s-1}$ is on average zero. This implies that within HLTG stocks future $LTG_{i,t+s}$ on average mean reverts, as implied by the power terms b^s and a^s . This occurs regardless of whether expectations are rational or diagnostic because $-\varphi_h b^s x_{i,t} + \vartheta_h a^s \hat{f}_{i,t}$ does not depend on θ . Between the formation date t and $t + 1$, however, mean reversion is stronger under diagnostic expectations. In fact, the condition

$$\mathbb{E}(g_{i,t+s} - a\hat{f}_{i,t+s-1} | HLTG_t) > 0$$

holds if and only if

$$b^h + \vartheta_h K' > 1$$

This condition is implied by the assumption of Proposition 2 (Equation 7), provided $K' > 1$, which holds in the estimation. We then find:

$$\mathbb{E}(LTG_{i,t+1}|HLTG_t) - \mathbb{E}(LTG_{i,t}|HLTG_t) = \mathbb{E}(LTG_{i,t+1}^{\theta=0}|HLTG_t) - \mathbb{E}(LTG_{i,t}^{\theta=0}|HLTG_t) - \Psi\theta$$

where $\Psi = \vartheta_h K \mathbb{E}(g_{i,t+s} - a\hat{f}_{i,t+s-1}) > 0$ (and $\theta > 0$). This is because under diagnostic expectation the average $LTG_{i,t}$ in the HLTG group is inflated relative to the rational benchmark.

The converse holds for stocks in the LLTG group at time t . ■

Proposition 5. The realized return at time t on HLTG stocks is equal to the average:

$$\mathbb{E}\left(\frac{P_{i,t} + D_{it}}{P_{i,t-1}}|HLTG_t\right) = R + \mathbb{E}\left(\frac{P_{i,t} - \mathbb{E}_{i,t-1}^{\theta}(P_{it}) + D_{i,t} - \mathbb{E}_{t-1}^{\theta}(D_{i,t})}{P_{i,t-1}}|HLTG_t\right).$$

An individual stock i in the HLTG portfolio therefore experiences positive abnormal returns pre formation provided:

$$P_{i,t} - \mathbb{E}_{t-1}^{\theta}(P_{i,t}) + D_{i,t} - \mathbb{E}_{t-1}^{\theta}(D_{i,t}) > 0$$

Consider the first term $P_{i,t} - \mathbb{E}_{t-1}^{\theta}(P_{i,t})$. Because prices are equal to discounted future dividends:

$$P_{i,t} - \mathbb{E}_{t-1}^{\theta}(P_{i,t}) = \sum_{s \geq 1} \frac{\mathbb{E}_t^{\theta}(D_{i,t+s}) - \mathbb{E}_{t-1}^{\theta}(D_{i,t+s})}{R^s},$$

which implies that abnormal returns are induced by an upward revision $\mathbb{E}_t^{\theta}(D_{i,t+s}) - \mathbb{E}_{t-1}^{\theta}(D_{i,t+s}) > 0$ of investors' beliefs of future dividends. Using previous notation, we have that:

$$\mathbb{E}_t^{\theta}(D_{i,t+s}) = \mathbb{E}_t^{\theta}(e^{x_{i,t+s}}) = e^{b^s x_{i,t} + \vartheta_s \hat{f}_{i,t}^{\theta} + \frac{1}{2} \text{var}_t(x_{t+s})}.$$

As a result, we have that $\mathbb{E}_t^\theta(D_{i,t+s}) - \mathbb{E}_{t-1}^\theta(D_{i,t+s}) > 0$ provided that, on average in HLTG:

$$b^s x_{i,t} + \vartheta_s \hat{f}_{i,t}^\theta > b^s \mathbb{E}_{t-1}^\theta(x_{i,t}) + \vartheta_s \mathbb{E}_{t-1}^\theta(f_{i,t})$$

We have:

$$\begin{aligned} & b^s (x_{i,t} - \mathbb{E}_{t-1}^\theta(x_{i,t})) + \vartheta_s (\hat{f}_{i,t}^\theta - \mathbb{E}_{t-1}^\theta(\hat{f}_{i,t}^\theta)) \\ &= b^s (S_{i,t} - aK\theta S_{i,t-1}) + \vartheta_s (K(1 + \theta)S_{i,t} - aK\theta S_{i,t-1}) \\ &= S_{i,t}(b^s + \vartheta_s K) + \theta K(\vartheta_s S_{i,t} - \vartheta_{s+1} S_{i,t-1}) \end{aligned}$$

where $S_t = x_{i,t} - bx_{i,t-1} - a\hat{f}_{i,t-1}$ is the news, or surprise, at time t . Intuitively, this suggests that high returns at t are associated with surprises $S_{i,t}$ that are not only positive, but also (for $\theta > 0$) sufficiently large compared to surprises in the previous period. Rewriting the above as $AS_{i,t} - BAS_{i,t-1}$ with $A = b^s + \vartheta_s K(1 + \theta)$ and $B = \theta K\vartheta_{s+1}$, we then have:

$$\mathbb{E}[AS_{i,t} - BAS_{i,t-1} | LTG_{i,t} = \lambda] = \frac{\text{cov}(AS_{i,t} - BAS_{i,t-1}, LTG_{i,t})}{\text{var}(LTG_{i,t})} \lambda.$$

We can write $LTG_{i,t}$ as:

$$LTG_{i,t} = [\vartheta_s K(1 + \theta) - \varphi_s] S_{i,t} - [\varphi_s(b + aK) - a\vartheta_s K] S_{i,t-1} + \text{terms at } t - 2$$

Because surprises in period t are uncorrelated with information at different periods, the numerator of the expectation above then reads:

$$\begin{aligned} & \text{cov}(AS_{i,t} - BAS_{i,t-1}, LTG_{i,t}) \\ &= A[\vartheta_s K(1 + \theta) - \varphi_s] \text{var}(S_{i,t}) - B[\varphi_s(b + aK) - a\vartheta_s K] \text{var}(S_{i,t-1}) \\ &= \text{var}(S_{i,t}) [A[\vartheta_s K(1 + \theta) - \varphi_s] - B[\varphi_s(b + aK) - a\vartheta_s K]] \end{aligned}$$

This is positive provided

$$\vartheta_s K(1 + \theta) + b^s > 1 + 2\theta K \vartheta_{s+1} [\varphi_s(b + aK) - a\vartheta_s K]$$

(which holds in our estimation). The assumption of Proposition 2 guarantees $\vartheta_s K(1 + \theta) + b^s > 1$. Thus, under rational expectations, $\theta = 0$, the condition holds trivially. For $\theta > 0$, a sufficient condition for the above to hold is that $\vartheta_s K > \varphi_s \left(\frac{b}{a} + K \right)$. Under this condition, $\mathbb{E}[x_{i,t} + LTG_{i,t} - x_{i,t-1} - LTG_{i,t-1} | LTG_{i,t} \geq \lambda_h]$ is positive, and the result follows. Note that this is implied by the assumption of Proposition 2 provided $\frac{b}{a} + K < \frac{1}{1+\theta}$. ■

Proposition 6. As shown in Equation (9), a stock's average return going forward into the next period is equal to:

$$\frac{\mathbb{E}_t(P_{i,t+1} + D_{i,t+1})}{\mathbb{E}_t^\theta(P_{i,t+1} + D_{i,t+1})} R.$$

Note that $D_{i,t+s} = e^{x_{i,t+s}}$ and that, given that price is the discounted sum of future dividends:

$$P_{i,t+1} + D_{i,t+1} = \sum_{s \geq 0} \frac{D_{i,t+1+s}}{R^s} = \sum_{s \geq 0} e^{x_{i,t+1+s} - s \ln R},$$

where, as usual, we assume that $\ln R$ is large enough that the sum converges. Given lognormality, we have that:

$$\mathbb{E}_t^\theta(P_{i,t+1} + D_{i,t+1}) = \sum_{s \geq 0} e^{\mathbb{E}_t^\theta(x_{i,t+1+s}) - s \ln R + \frac{1}{2} \text{Var}_t(x_{i,t+1+s})},$$

where rational expectations correspond to the special case of $\theta = 0$. For $\theta = 0$, then, the numerator and the denominator of Equation (9) are equal, so that the average realized return is

equal to the realized return R for all firms. As a result, the average realized post-formation return of the HLTG and LLTG portfolios should be equal to the required return R .

To see the role of diagnostic expectations, note that θ only influences the expected log dividend $\mathbb{E}_t^\theta(x_{i,t+1+s})$, but not the perceived variance $Var_t(x_{i,t+1+s})$. In particular:

$$\mathbb{E}_t^\theta(x_{i,t+s+1}) = b^{s+1}x_t + a^{s+1} \frac{1 - (b/a)^{s+1}}{1 - (b/a)} [\hat{f}_{i,t} + K\theta(g_{i,t} - a\hat{f}_{i,t-1})].$$

This implies that:

$$\begin{aligned} & \frac{\partial \mathbb{E}_t^\theta(P_{i,t+1} + D_{i,t+1})}{\partial \theta} \\ &= K\theta(g_{i,t} - a\hat{f}_{i,t-1}) \sum_{s \geq 0} a^{s+1} \frac{1 - (b/a)^{s+1}}{1 - (b/a)} e^{\mathbb{E}_t^\theta(x_{i,t+1+s}) - s \ln R + \frac{1}{2} Var_t(x_{i,t+1+s})} \end{aligned}$$

Under diagnostic expectations, pre-formation news drive mispricing. HLTG stocks experience positive surprises before formation, namely $(g_{i,t} - a\hat{f}_{i,t-1}) > 0$. As a result, the diagnostic expectation $\mathbb{E}_t^\theta(P_{i,t+1} + D_{i,t+1})$ is above the rational counterpart, so that realized post formation returns are on average below the required return R . For LLTG the opposite is true. ■

Proposition 7. Regressing $x_{i,t+h} - x_{i,t} - LTG_{i,t}$ on $LTG_{i,t} - LTG_{i,t-k}$ yields a coefficient

$$\beta = \frac{cov(x_{i,t+h} - x_{i,t} - LTG_{i,t}, LTG_{i,t} - LTG_{i,t-k})}{var(LTG_{i,t} - LTG_{i,t-k})}$$

The forecast error in the denominator reads

$$x_{i,t+h} - x_{i,t} - \mathbb{E}_t(x_{i,t+h} - x_{i,t}) - \theta \vartheta_h K(x_{i,t} - bx_{i,t-1} - a\hat{f}_{i,t-1})$$

The first two terms include only shocks after t and do not co-vary with any quantity at t . The last term arises only for $\theta > 0$, and captures the overreaction to news at t embedded in $LTG_{i,t}$. We thus have

$$\beta \propto -\theta \cdot cov(x_{i,t} - bx_{i,t-1} - a\hat{f}_{i,t-1}, LTG_{i,t} - LTG_{i,t-k})$$

Intuitively, a positive covariance means that positive surprises at t tend to be associated with upward revisions in LTG. The second argument reads:

$$\begin{aligned} & -\varphi_h(x_{i,t} - x_{i,t-1}) + \vartheta_h a(1 - K')(\hat{f}_{i,t-1} - \hat{f}_{i,t-2}) + \vartheta_h K'(x_{i,t} - bx_{i,t-1} - a\hat{f}_{i,t-1}) \\ & - \vartheta_h K'(x_{i,t-1} - bx_{i,t-2} - a\hat{f}_{i,t-2}) \end{aligned}$$

The surprise at t does not covary with either the update in beliefs at $t - 1$ (second term), nor with the surprise at $t - 1$ (last term), so these drop out. Write the first term as

$$-\varphi_h(x_{i,t} - x_{i,t-1}) = -\varphi_h(x_{i,t} - bx_{i,t-1} - a\hat{f}_{i,t-1}) - \varphi_h((1 - b)x_{i,t-1} - a\hat{f}_{i,t-1})$$

Again, because surprises at t are not predictable from information at $t - 1$, the second term drops out. We therefore get $(\vartheta_h K' - \varphi_h)var(x_{i,t} - bx_{i,t-1} - a\hat{f}_{i,t-1})$ which is positive if and only if

$$b^h + \vartheta_h K(1 + \theta) > 1$$

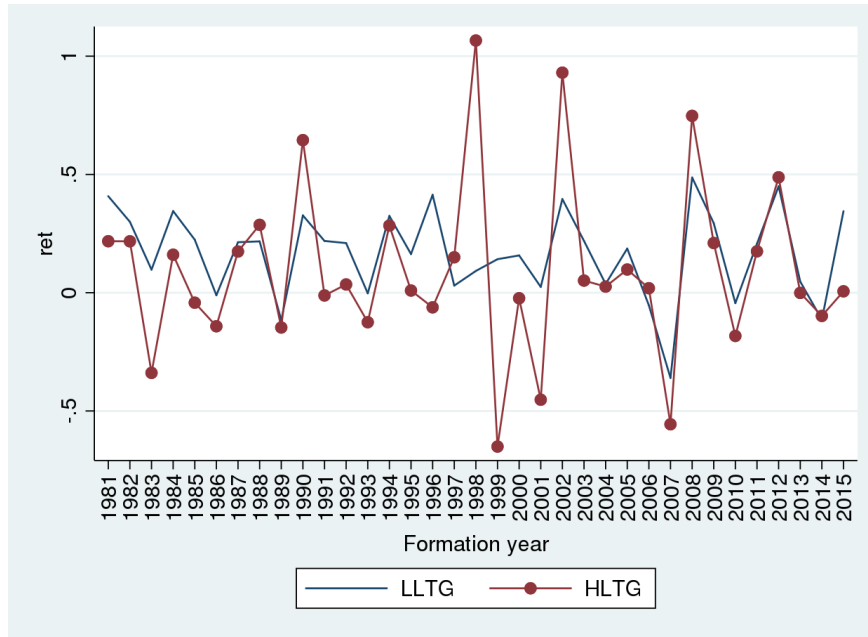
This condition is weaker than that of Proposition 2 provided $K(1 + \theta) > 1$, which holds in our estimation. ■

B. Robustness of LLTG-HLTG Return Differential

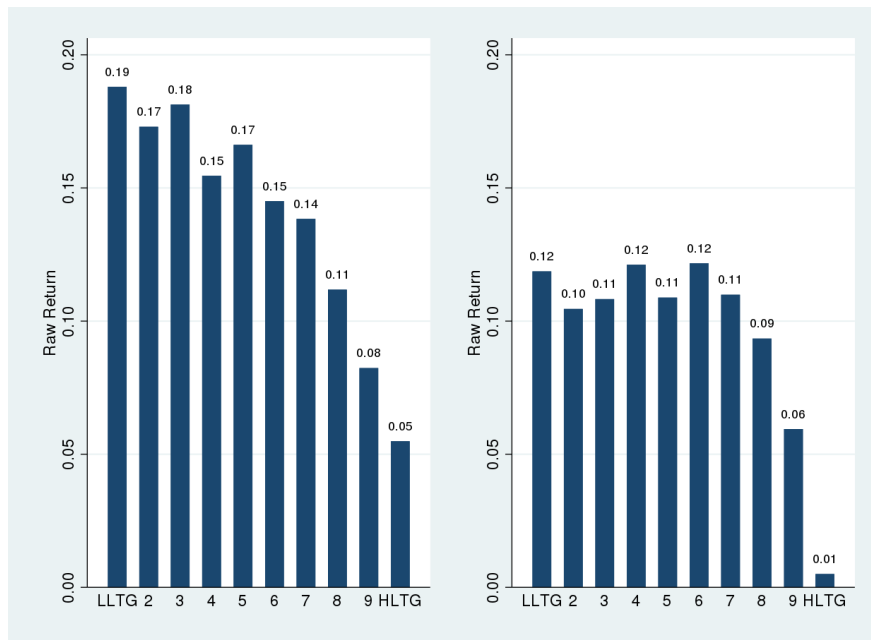
This section examines the robustness of the LLTG-HLTG return differential from a number of different perspectives. We first examine the performance of LTG portfolios across

different sub-samples. Next we relate the LLTG-HLTG spread in returns to the return factors commonly used in the finance literature. We then present two-way sorts of LTG and, alternatively, momentum and size. We conclude by presenting value-weighted returns for LTG portfolios.

We begin by analysing the consistency of the returns of HLTG and LLTG portfolios during subsamples. Panel A in Figure B1 illustrates the performance of the LLTG and HLTG portfolios over time. It shows that the HLTG portfolio exhibits extreme volatility, particularly during the 1998-2001 period. Panel B in Figure B1 splits the sample period roughly in half and shows the performance of LTG portfolios during the periods 1981-1997 (left panel) and 1998-2015. The LLTG-HLTG spread is roughly 14 percentage points during the first half of the sample and 12 percentage points during the second half. These calculations are sensitive to whether the 1998 formation period is included during the first or the second half of the sample. Specifically, if we do the former, the LLTG-HLTG spread is roughly 9 percentage points during the 1981-1998 period and 16 percentage points during the 1999-2015 period. Either way, the LLTG-HLTG spread is statistically undistinguishable between sub-samples.



Panel A



Panel B

Figure B.1. Annual Returns for Portfolios Formed on LTG. In December of each year between 1981 and 2015, we form decile portfolios based on ranked analysts' expected growth in earnings per share. Panel A shows the time-series of HLTG and LLTG returns during the full sample. Panel B shows returns for LTG portfolios formed during the periods 1981-1997 (on the left) and 1998-2015 (on the right). The returns in Panel B are geometric averages of one-year equally-weighted return over the relevant sample periods.

Next, we examine the relationship between the performance of a portfolio that is long LLTG stocks and short HLTG stocks and some of the return factors commonly used in the finance literature (see Fama and French 2015 and the references therein).

The factors that we consider are the five Fama-French factors, momentum (UMD), and betting against beta (BAB). The Fama-French factors are: (1) the excess return on the market relative to the one-month T-Bill (Mkt-RF), (2) the difference in the average return of a high book-to-market portfolio and the average return on a low book-to market portfolio (HML), (3) the difference in the average return on a portfolio of small firms and the average return on a portfolio of big stocks (SML), (4) the difference in the average return of a high operating profitability portfolio and a low operating profitability portfolio (RMW), and (5) the difference in the average return of a portfolio of low investment and the average return of a portfolio of high investment (CMA).

Table B.1 shows betas for LTG portfolios. As described in the text, beta increases monotonically with LTG and nearly doubles from the LLTG to HLTG portfolio (0.79 vs. 1.51).

Table B.1 – Average Beta is increasing across LTG Portfolios

We estimate slope coefficients from the following OLS regression

$$ret_{i,t} - rf_t = \alpha_i + \beta_i * (rm_t - rf_{i,t}) + \varepsilon_{i,t}$$

where $ret_{i,t}$ is the monthly return for firm i , rf_t is the risk-free rate (from Ken French's website) in period t , rm_t is the return on the equally-weighted index in period t (also from Ken French's website). We estimate the regression using a rolling window of 60 months. The table below reports average β s for LTG portfolios.

	LTG decile									
	1	2	3	4	5	6	7	8	9	10
β	0.79	0.93	1.00	1.06	1.12	1.18	1.28	1.37	1.46	1.51
										-0.72

Turning to the factor analysis, we begin by examining simple correlations of monthly equally-weighted portfolio returns in Table B.2, Panel A. The long-short LLTG-HLTG portfolio is negatively correlated with the market factor, consistent with Table B.1. It negatively correlated with the size factor and strongly positively correlated with the book-to-market factor as well as with the investment factor (CMA). As noted in the text, this is unsurprising since –relative to HLTG stocks -- LLTG stocks have high book-to-market ratios, low beta, and high profitability (Table 1). In contrast, the LTG portfolio only displays a small and non-significant correlation with momentum (UMD). This is confirmed in Panel B, which shows that the LLTG-HLTG spread is approximately constant across momentum buckets (i.e., bottom 30%, middle 40%, and top 30%). Finally, Table B.2 Panel C shows the results of a conventional factor analysis. The monthly excess return of the long-short portfolio is 1.16% when we only control for market risk and falls roughly in half when controlling for the standard three Fama-French factors (i.e. MktRF, HML, and SMB). The monthly excess return of the long-short LTG portfolio loses significance if we, alternatively, control for: (1) momentum, (2) betting-against-beta, and (3) profitability (RMW) plus investment (CMA).

Table B.2.

In December of each year between 1981 and 2015, we form decile portfolios based on ranked analysts' expected long-term growth in earnings per share and compute equally-weighted monthly return for decile portfolios. Panel A reports pair-wise correlations between the results of regressions of the return for the portfolio that is long LLTG and short HLTG (LTGLMH) on: (1) the excess return on the market relative to the one-month T-Bill (Mkt-RF), (2) the difference in the average return of a high book-to-market portfolio and the average return on a low book-to market portfolio (HML), (3) the difference in the average return on a portfolio of small firms and the average return on a portfolio of big stocks (SML), (4) the difference in the average return on a portfolio with high prior ($t-12$, $t-2$) returns and the average return on a portfolio with low previous returns (UMD), (5) the difference in the average return of a portfolio of low-beta stocks and the average return of a portfolio of high-beta stocks (BAB), (6) the difference in the average return of a high operating profitability portfolio and a low operating profitability portfolio (RMW), and (7) the difference in the average return of a portfolio of low investment and the average return of a portfolio of high investment (CMA). Data on BAB is from the AQR website. All other data on factor returns is from Ken French's website. In Panel B, we independently form ten portfolios based on ranked analysts' expected growth in earnings per share and three portfolios (i.e. bottom 30%, middle 40%, and top 30%) based on six-month momentum (i.e. July-December of year t). The table reports the average one-year return over the subsequent calendar year for equally-

weighted portfolios. Panel C reports the results of OLS regressions of the return for the portfolio that is long LLTG and short HLTG on Mkt-RF, HML, SMB, UMD, BAB, RMW, and CMA.

Panel A. Correlates of LLTG-HLTG. ^a denotes significance at the 1% level.

LLTG-HLTG	
MktRF	-54% ^a
HML	74% ^a
SMB	-43% ^a
RMW	60% ^a
CMA	63% ^a
BAB	48% ^a
UMD	11%

Panel B. Annual Returns for Portfolios Formed on LTG and six-month momentum

Returns in the previous 6 months				
LTG	Bottom 30%	Middle	Top 30%	Top-Bottom
LLTG	9.6%	15.9%	17.3%	7.7%
2	9.7%	15.2%	13.9%	4.2%
3	10.6%	15.2%	14.8%	4.2%
4	10.7%	14.4%	14.2%	3.5%
5	10.5%	15.0%	14.5%	3.9%
6	10.3%	14.4%	14.0%	3.6%
7	7.5%	12.3%	15.7%	8.2%
8	6.1%	11.3%	12.4%	6.3%
9	5.7%	7.1%	7.9%	2.2%
HLTG	-1.9%	4.5%	8.0%	9.8%
HLTG-LLTG	-11.5%	-11.4%	-9.3%	2.1%

Panel C. Factor Regressions

Mkt-RF	-0.8761 ^a (0.0815)	-0.5327 ^a (0.0465)	-0.4581 ^a (0.0436)	-0.4953 ^a (0.0383)	-0.3512 ^a (0.0423)
HML		1.5141 ^a (0.0813)	1.6182 ^a (0.0854)	1.3270 ^a (0.0732)	1.0829 ^a (0.0901)
SMB		-0.6685 ^a (0.0744)	-0.6861 ^a (0.0668)	-0.6823 ^a (0.0692)	-0.4604 ^a (0.0717)
UMD			0.2936 ^a (0.0759)		
BAB				0.4377 ^a (0.0798)	
RMW					0.7918 ^a (0.1162)

CMA					0.7052 ^a (0.1768)
Constant	1.1608 ^a (0.3090)	0.5526 ^a (0.1791)	0.2916 (0.1956)	0.1831 (0.1905)	0.0298 (0.1804)
Observations	408	408	408	408	408
Adjusted R-squared	0.293	0.761	0.792	0.802	0.825

We conclude by taking a closer look at the role of size in the performance of LTG portfolios and, related, value-weighted returns. Table B3 shows average annual compounded returns of portfolios formed independently based on LTG and size. The LLTG-HLTG spread holds within size buckets and ranges from 13.5 percentage points for small stocks to 9.5 percentage points. In contrast, size plays a muted role within LTG buckets, except for the LLTG portfolio where small stocks earn 3.9% higher returns than big stocks. The final column presents value-weighted returns. The LLTG-HLTG spread in (log) returns drops from 13.6%, (t-statistic of 2.22) when returns are equally-weighted to 6.4% when returns are value-weighted (t-statistic of 1.2). While the last result lacks significance, this should be interpreted with caution: first, the Table shows that equally-weighted returns of HLTG portfolios are roughly 3% across all size buckets. Second, value-weighted LTG returns are dominated by a handful of large cap stocks which had high returns, particularly during the early 2000s.

Table B.3. LTG and Value weighted returns

In December of each year between 1981 and 2015, we form decile portfolios based on ranked analysts' expected long-term growth in earnings per share and form decile portfolios. The first three columns of the table below report average compounded annual returns for independent two-way sorts of LTG and various size portfolios based for market capitalization in December of year t (using NYSE break-points). *Small* is the portfolio of stocks in the bottom three deciles of the size distribution. *Middle* is the portfolio of stocks in deciles 4 through 7 of the size distribution. *Big* is the portfolio of stocks in the top three deciles of the size distribution. SMB is the difference in the average return of the portfolio that is long *Small* and short *Big*. The last column shows value-weighted returns (WV Ret).

LTG	Size			SMB	VW Ret
	Small	Middle	Big		

1	16.5%	15.0%	12.6%	3.9%	13.5%
2	13.6%	13.0%	13.9%	-0.3%	13.1%
3	12.9%	14.9%	13.8%	-0.9%	12.1%
4	12.9%	14.2%	13.0%	-0.1%	12.1%
5	13.2%	14.6%	12.8%	0.4%	10.4%
6	13.8%	13.2%	13.0%	0.8%	12.8%
7	11.1%	13.2%	11.8%	-0.7%	11.5%
8	10.3%	10.9%	9.9%	0.4%	10.6%
9	8.4%	6.2%	6.4%	2.1%	7.3%
10	3.0%	2.1%	3.1%	-0.1%	7.1%
LLTG-HLTG	13.5%	12.9%	9.5%	4.0%	6.4%

C. Kernel of Truth

We begin by comparing the post-formation growth in earnings of HLTG and LLTG firms. For convenience, Figure C.1. reproduces Figure 6 (distribution of eps growth, top left panel) and Figure 7 (distribution of LTG expectations, bottom left panel) for the HLTG portfolio. The middle panel of Figure C plots the representativeness (i.e. the ratio of the densities in the top left panel) of HLTG vs. all other firms. As noted in the text, the most representative future growth realizations for HLTG firms are in the 50%-90% range of annual growth. Stocks in the LLTG portfolio (see right panel) stand in sharp contrast to those in the HLTG portfolio. The top right panel plots the distribution of earnings growth for all LLTG stocks relative to all other stocks. The right tail of the distribution of growth in earnings for stocks in the LLTG stocks is noticeably thinner than for stocks in all other portfolios. Critically, it is the left tail that is most representative for LLTG stocks (see middle panel). Despite the prominence of the left tail, analysts are not overly pessimistic about the performance of LLTG firms, which narrowly exceed growth expectations (i.e. the post-formation average growth in earnings averages 5% per year rather than the expected 3%).

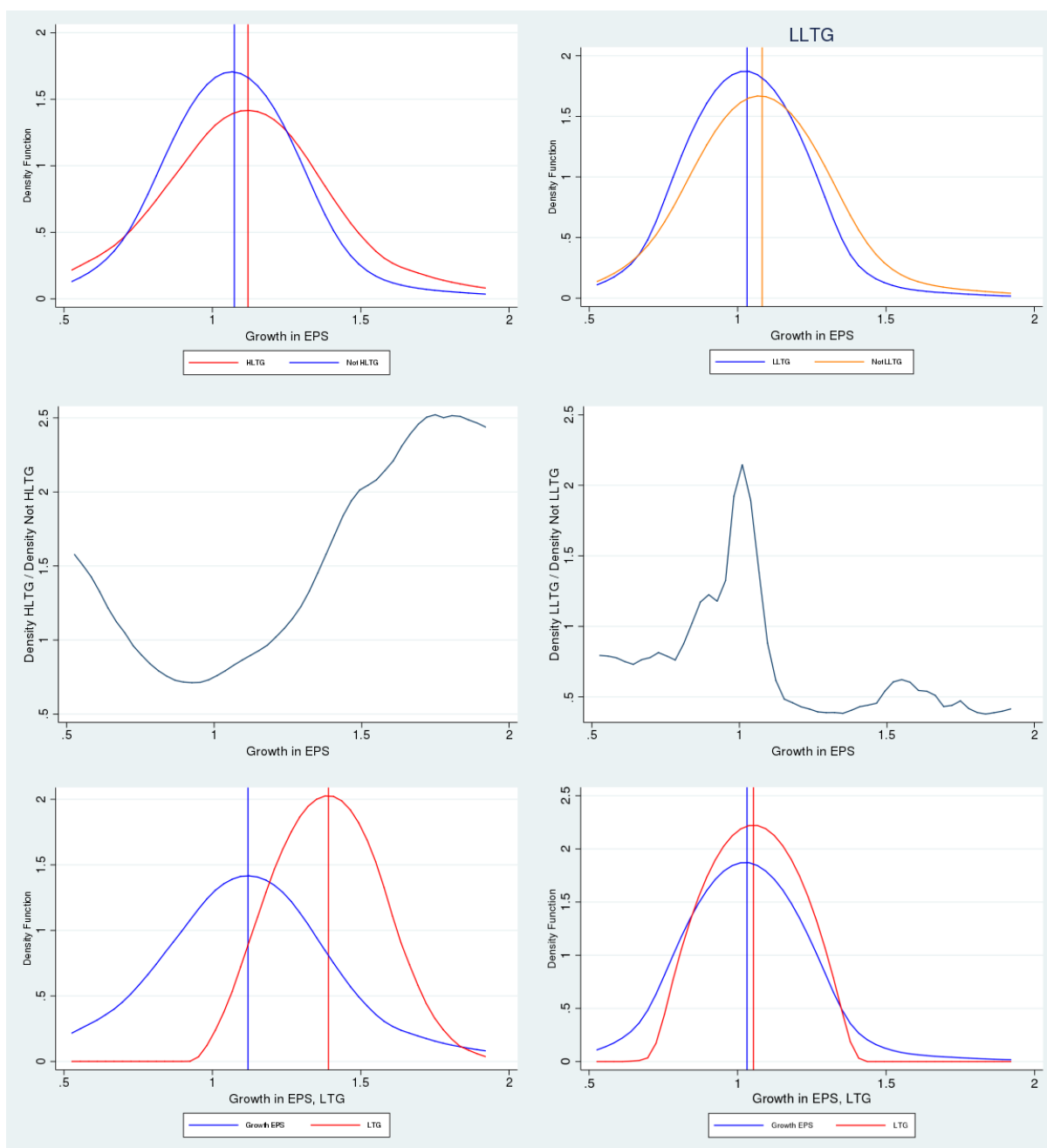


Figure C.1. EPS and expectations for the HLTG (left panels) and LLTG (right panels) portfolios. In December of years (t) 1981, 1986, ..., and 2011, we form decile portfolios based on ranked analysts' expected growth in long term earnings per share (LTG). For each stock, we compute the gross annual growth rate of earnings per share between t and $t+5$, excluding stocks with negative earnings in year t and $t+5$. We estimate the kernel densities for stocks in: (a) the highest ($HLTG$) decile vs. all other stocks, and (b) the lowest ($LLTG$) decile vs. all other stocks. The top panel shows the estimated density kernels of growth in earnings per share for stocks in the $HLTG$ (blue line) and all other firms (orange line) on the left and the estimated density kernels of growth in earnings per share for stocks in the $LLTG$ (blue line) and all other firms (orange line) on the right. The vertical lines indicate the means of each distribution (1.12 vs. 1.07 on the left panel and 1.03 vs. 1.08 on the right panel). The left middle panel shows the ratio of the densities of growth in earnings per share for stocks in the $HLTG$ portfolio versus stocks in the non- $LLTG$ portfolio. The right middle panel repeats the exercise for stocks in the $LLTG$ portfolio versus stocks in the non- $LLTG$ portfolio. The left bottom panel plots two series: the kernel distribution of the gross annual growth rate in earnings per share

between t and $t+5$ for stocks in the HLTG portfolio, and the kernel distribution of the expected growth in long term earnings at time t for stocks in the HLTG portfolio. The right bottom panel repeats the analysis for stocks in the LLTG portfolio vs. stocks in the non-LLTG portfolio. The vertical lines indicate the means of each distribution (1.12 vs. 1.39 on the left panel and 1.03 vs. 1.05 on the right panel).

The analysis so far has focused on firms with positive earnings. Next, we consider two metrics that make it possible to examine the performance of firms with negative earnings at the cost of shifting the focus away from the variable that theory predicts should matter for pricing stocks. The first such metric is growth in revenue minus operating, general, and sales costs (RMC), which takes advantage of the fact that most (97%) firms with negative earnings have positive values of RMC while staying close to the idea that what matters for expectations is growth. The second metric is defined as the ratio of changes in earnings between between t and $t+5$ scaled to the initial level of sales per share. The latter measure allows us to include all firm in the analysis but it is much less directly related to the model's predictions.

Figure C.2 presents results for growth in RMC in the left panel and changes in earnings in the right panel. Both panels confirm that HLTG firms have a slightly higher mean and a much fatter right tail of exceptional performers. Finally, the evidence in the bottom panel – a particularly with regards to growth in RMC – is consistent with the prediction of our model that the representativeness of Googles in the HLTG portfolio leads analysts to overestimate their future performance.

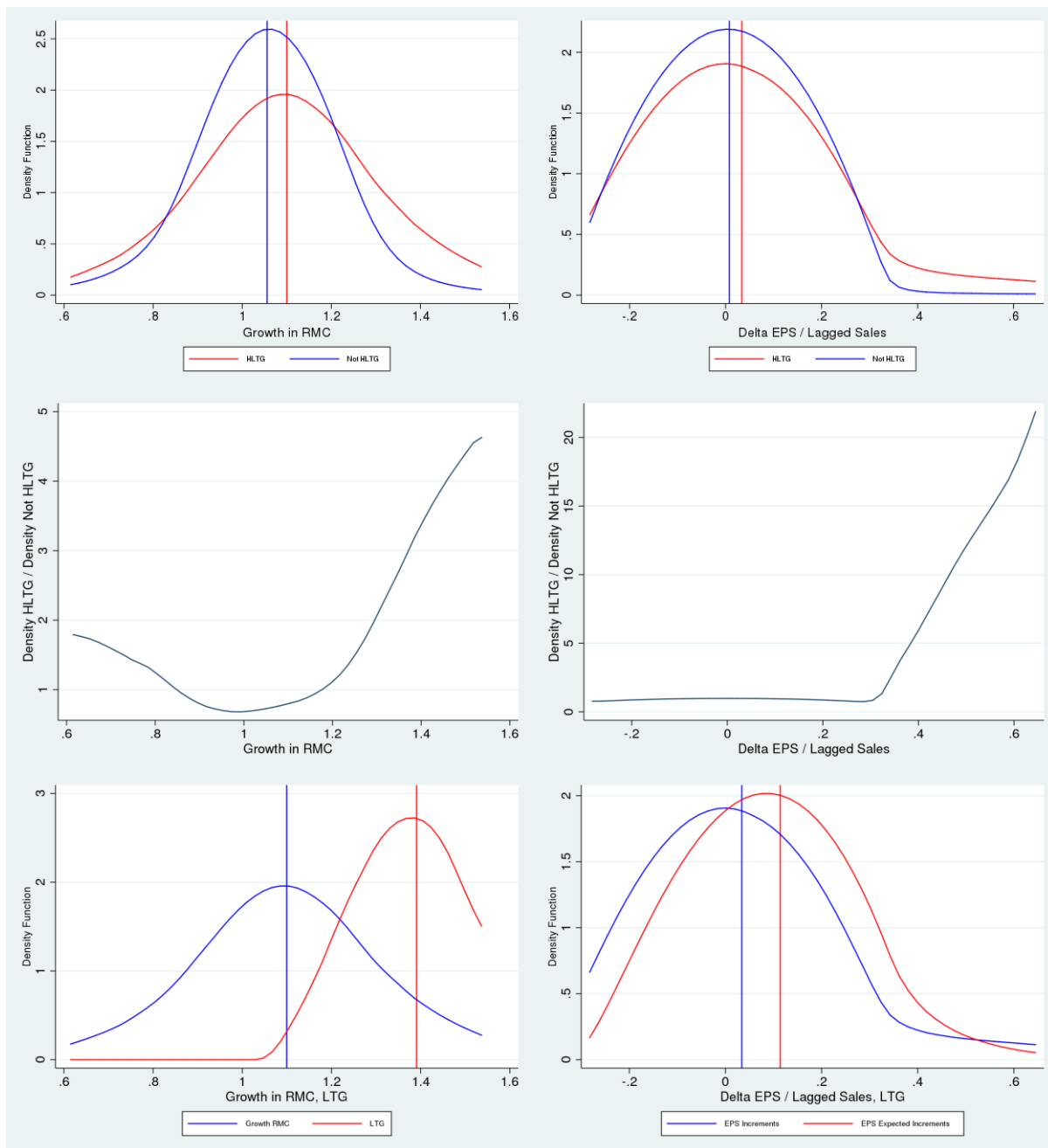


Figure C.2. Revenues minus cost of goods sold (RMC) (left panels) or change in EPS normalized by lagged sales (right panels) for the HLTG portfolios. In December of years (t) 1981, 1986, ..., and 2011, we form decile portfolios based on ranked analysts' expected growth in long term earnings per share (LTG). For each stock, we compute the gross annual growth rate of operating margin (i.e. revenue minus operating, general, and sales costs) per share between t and $t+5$ as well as the change in earnings per share between t and $t+5$ normalized by lagged sales per share. We exclude firms with negative margins in years t and $t+5$ for the computation of the growth rate of operating margin. The top panel shows the estimated density kernels of growth in operating margins per share for stocks in the *HLTG* (blue line) and all other firms (orange line) on the left and the estimated density kernels of the change in earnings for the *HLTG* (blue line) and all other firms (orange line) on the right. The vertical lines indicate the means of each distribution (1.10 vs. 1.06 on the left panel and 0.045 vs. 0.01 on the right panel). The left middle panel shows the ratio of the densities of growth in margins for stocks in the *HLTG* portfolio versus stocks in the non-*LTG* portfolio on the left and the analogous graph for changes in eps on the right. The left bottom panel plots two series:

the kernel distribution of the gross annual growth rate in operating margins per share between t and $t+5$ for stocks in the HLTG portfolio, and the kernel distribution of the expected growth in long term earnings at time t for stocks in the HLTG portfolio. The right bottom panel repeats the analysis for changes in eps of HLTG stocks. Expected earnings in year $t+5$ are computed as $\text{eps}_t (1+\text{LTG})^5$. The vertical lines indicate the means of each distribution (1.10 vs. 1.39 on the left panel and 0.045 vs. 0.11 on the right panel).

D. Coibion and Gorodnichenko Analysis

D.1 Overreaction to News vs Adaptive Expectations

Adaptive expectations (Equation 11) predict no over-reaction to news after the persistence of the earnings process is accounted for. From (11), the forecast error on an AR(1) process with persistence ρ is $x_{t+1} - x_{t+1}^a = (\rho - 1)x_t + \left(\frac{1-\mu}{\mu}\right)(x_{t+1}^a - x_t^a)$. Controlling for x_t fully accounts for mechanical over-reaction in processes with low persistence. The adaptive forecast revision $(x_{t+1}^a - x_t^a)$ should positively predict forecast errors as in the under-reaction models. This prediction is not shared by our model because diagnostic expectations over-react to news regardless of the persistence of the data generating process. Table C.1 reports the results. The coefficients on forecast revision become larger than those estimated in Table 2, but they remain mostly negative and statistically significant.

Table D.1: Forecast Errors

Each entry in the table corresponds to the estimated coefficient of the forecast errors $(\text{eps}_{t+n} / \text{eps}_t)^{1/n} - \text{LTG}_t$ for $n=3, 4$, and 5 on the variables listed in the first column of the table as well as $(\log) \text{eps}_t$ and year fixed-effects (not shown).

	Dependent Variable		
	$(\text{eps}_{t+3} / \text{eps}_t)^{1/3} - \text{LTG}_t$	$(\text{eps}_{t+4} / \text{eps}_t)^{1/4} - \text{LTG}_t$	$(\text{eps}_{t+5} / \text{eps}_t)^{1/5} - \text{LTG}_t$
$\text{LTG}_t - \text{LTG}_{t-1}$	0.0332 (0.0725)	-0.0733 (0.0660)	-0.1372 ^b (0.0589)
$\text{LTG}_t - \text{LTG}_{t-2}$	-0.0875 (0.0641)	-0.1435 ^b (0.0691)	-0.1842 ^a (0.0545)

LTG _t -LTG _{t-3}	-0.0956 (0.0578)	-0.1184 ^c (0.0627)	-0.1701 ^a (0.0517)
--------------------------------------	---------------------	----------------------------------	----------------------------------

D.2 Underreaction for forecasts at short time horizons

Table D.2 tests the predictability of forecast errors in the forecast of earnings levels, as opposed to the predictability of errors in LTG forecasts analysed in Table 2. The results suggest underreaction for forecasts at short time horizons (i.e. one year ahead), compatible with Bouchaud et al. (2018). As in Table 2, as the forecasting horizon increases to three years ahead, the coefficient becomes less positive and even negative in some specifications.

Table D.2: EPS Forecast Errors at short time horizons

Each entry in the table corresponds to the estimated coefficient of the forecast errors for t+1, t+3, and t+5 on the variables listed in the first column of the table as well as year fixed-effects (not shown). All forecast errors are scaled by lagged sales per share.

	$(\text{eps}_{t+1} - E_t \text{eps}_{t+1}) / \text{sps}_{t-1}$	$(\text{eps}_{t+3} - E_t \text{eps}_{t+3}) / \text{sps}_{t-1}$	$(\text{eps}_{t+5} - E_t \text{eps}_{t+2} * (1 + \text{LTG}_t)^3) / \text{sps}_{t-1}$
LTG _t -LTG _{t-1}	0.0839 (0.0554)	-0.3226 ^b (0.1312)	-0.5918 ^a (0.1435)
LTG _t -LTG _{t-2}	0.1629 ^a (0.0275)	0.1262 ^c (0.0677)	-0.0227 (0.0772)
LTG _t -LTG _{t-3}	0.0825 ^a (0.0195)	-0.0664 (0.0532)	-0.2145 ^b (0.0919)

D.3 Overreaction and predictable returns

We next try to tie over-reaction to news to the return spread between HLTG and LLTG portfolios. We estimate Equation (10) by pooling firms at the industry level, using the Fama and French classification. To capture industry and firm specific factors, we allow for industry \times year fixed effects. This yields an industry level estimate γ_s , where s indexes the industry, which we can correlate with the industry level LLTG-HLTG spread. These results should be taken with caution, due to the small number of industries.

In Figure D.1, we compare the post formation return spread across different terciles of the distribution of industry γ_s . Consistent with our prediction, the extra return obtained by betting against HLTG firms is highest in sectors that feature most over-reaction, namely those in the bottom tercile of γ_s . The return differential is sizable, though given the small sample size it is not statistically significant. Thus, the pattern of LLTG-HLTG return spreads across industries is consistent with a link from overreaction to news to overvaluation of HLTG stocks and thus to abnormally low returns of the HLTG portfolio.

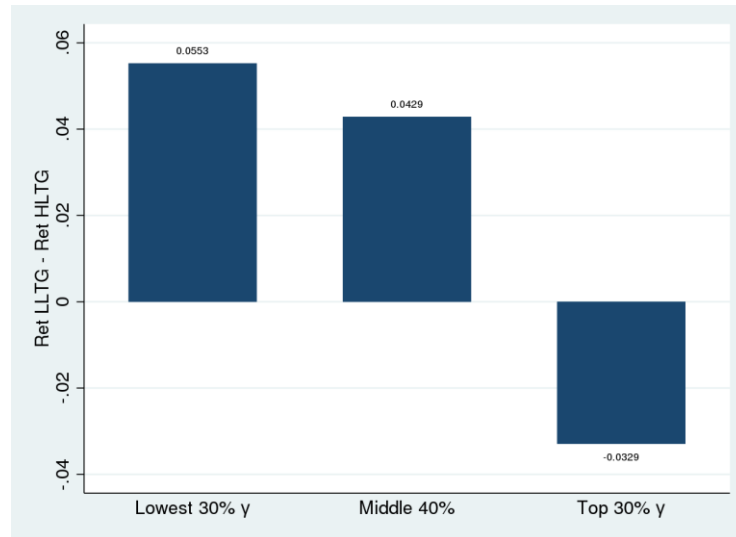


Figure D.1. Overreaction and return spread across industries. For each of the 48 Fama-French industries, we estimate the regression: $(EPS_{t+4}/EPS_t)^{1/4} - (1 + LTG_t) = \alpha_i + \mu_{i,t} + \gamma_i(LTG_t - LTG_{t-3}) + \varepsilon_{i,t}$ where μ_t are year fixed effects, EPS is earnings per share, and LTG is the forecast long-term growth in earnings. We rank

industries according to γ_{iS} and form the following three groups: (1) 14 industries with the lowest γ_{iS} , (2) 24 industries with intermediate values of γ_{iS} , and (3) 14 industries with the highest γ_{iS} . Finally, for each year and each group, we compute the difference in return for the LLTG (i.e. bottom 30% of LTG) and HLTG (i.e. highest 30% of LTG) portfolios. The graph shows the arithmetic mean of the LLTG-HLTG spread for grouping industries based on γ_{iS} from the regression.

E. Model Estimation

In this section, we extend our estimates with several exercises. We first test the robustness of our estimate $\theta = 0.9$. We then provide the simulation counterparts to the comparative static of overreaction relative to persistence and volatility of the process for growth fundamentals f_t (Table 4), and to the revision of expectations as a function of news (Figure 9).

To test the robustness of our estimate for θ , we evaluate the matching between the target vector of empirical moments $\bar{v} = (0.82, 0.75, 0.70, 0.65, -0.276, -0.126)$ and the predicted counterpart for $\theta \in [0, 2]$ (keeping the other parameters constant). Figure E.1 below shows that the match is indeed optimized for $\theta = 0.9$ and that it drops fast as θ departs from this value.

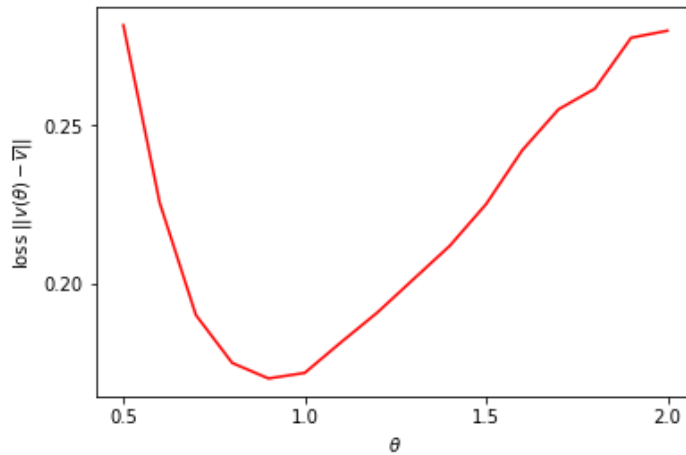


Figure E.1. For $(a, b, \sigma_\eta, \sigma_\epsilon, s) = (0.97, 0.56, 0.138, 0.082, 11)$, as given by our estimation, and each $\theta \in [0, 2]$ (in steps of 0.1) we compute the target vector of moments v . The figure plots the loss $\ell(v) = \|v - \bar{v}\|$.

Next, we return to the kernel of truth prediction that overreaction increases with the drivers of the signal to noise ratio. To illustrate this prediction we compute the returns of LLTG and HLTG portfolios under our estimates for $(b, \sigma_\epsilon, \theta, s)$ while allowing the persistence a and volatility σ_η of the growth fundamentals f_t to vary over parameter specifications comparable to those in Table 4. Figure E.2 presents the results. Consistent with the empirical evidence of Table 4, the LLTG-HLTG return spread increases with a and with σ_η .

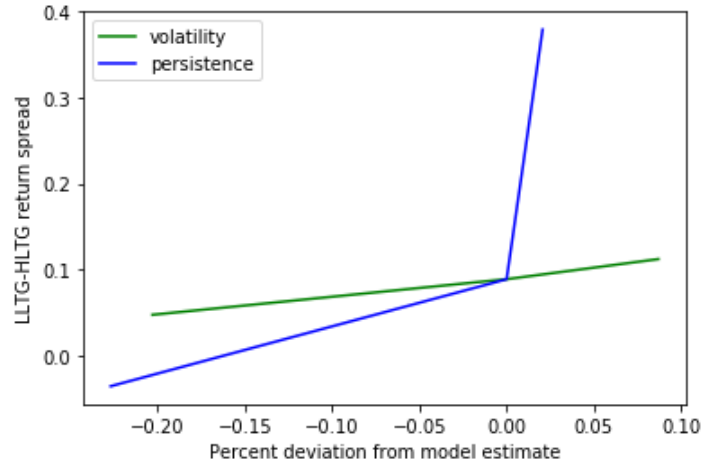


Figure E.2. For $(b, \sigma_\epsilon, \theta, s) = (0.56, 0.08, 0.90, 11)$, as given by our estimation, we vary the persistence of fundamentals $a \in \{0.75, 0.97, 0.99\}$ (keeping $\sigma_\eta = 0.138$) or the volatility of fundamentals $\sigma_\eta \in \{0.11, 0.138, 0.17\}$ (keeping $a = 0.97$). For each parameter combination we compute the LLTG-HLTG return spread.

Finally, we confirm numerically the pattern of Figure 9, namely that high LTG forecasts are on average revised downwards, even conditional on positive news, that is, earnings that exceeded (distorted) expectations. Figure E.3 below plots the equivalent of Figure 9 for our estimation.

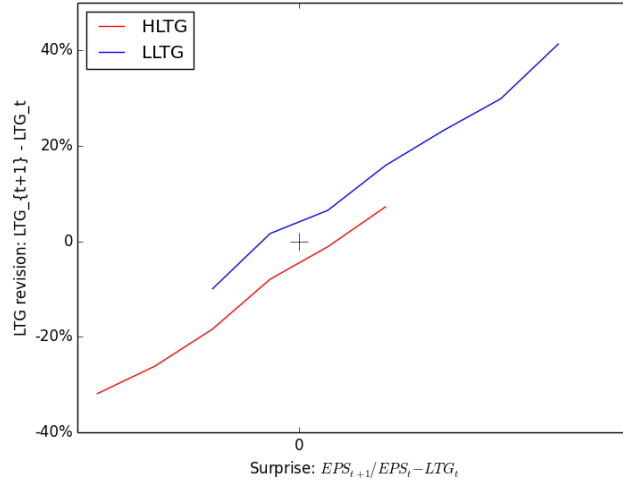


Figure E.3. For each period t and each firm in the $HLTG_t$ and $LLTG_t$ portfolios, we compute the surprise $\ln EPS_{t+1} - \ln EPS_t - LTG_t$. Pooling the data for all t and both portfolios, we rank the surprises in deciles and plot the average revision of LTG for each portfolio in each decile.

In the model, news and forecast revisions are positively correlated. Crucially, however, they may go in opposite directions. For a range of positive surprises – namely, realized growth above the diagnostic forecasts – forecasts about HLTG firms are still revised downwards. Naturally, for sufficiently large positive surprises (larger than $1.5\sigma_\epsilon$, in our estimation) forecast revisions are positive. The converse holds for LLTG.