DIAGNOSTIC EXPECTATIONS AND STOCK RETURNS

Pedro Bordalo, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer[1]

First Draft, November 2016, This Draft, July 2017

Abstract

We revisit La Porta's (1996) finding that returns on portfolios of stocks with the most optimistic analyst long term earnings growth forecasts are substantially lower than those for stocks with the most pessimistic forecasts. We document that this finding still holds, and present several further facts about the joint dynamics of fundamentals, expectations, and returns for these portfolios. We then propose a new approach to modeling belief formation and over-reaction to news that explains these facts, based on a portable psychological model of judgment by representativeness. This entails a learning model in which analysts forecast future fundamentals based on the history of earnings growth, but update excessively in the direction of states of the world whose objective likelihood rises the most in light of the news. Intuitively, fast earnings growth news predicts future Googles but not as many as analysts believe. The model delivers the empirical findings we initially document, and yields additional empirical predictions that distinguish it from both Bayesian learning and adaptive expectations. We test these predictions and find supportive evidence.

[1] Saïd Business School, University of Oxford, Universita Bocconi and IGIER, Brown University, and Harvard University, respectively. Gennaioli thanks the European Research Council 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 and Jesse Shapiro, for helpful comments.

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.

Figure 1 offers an update of this phenomenon. Stocks are sorted by analyst long-term earnings per share growth forecasts (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. The figure reports geometric averages of one year returns on equally weighted portfolios.

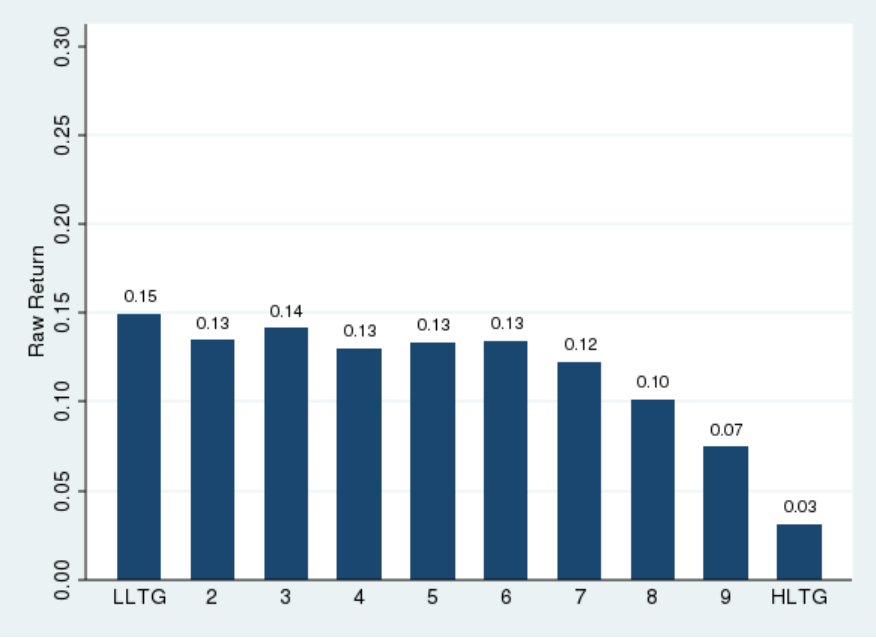


Figure 1. Annual Returns for Portfolios Formed on LTG (January 1981-December 2015). In December of each year between 1981 and 2014, we form decile portfolios based on ranked analysts' expected growth in earnings per share and report the average one-year return over the subsequent calendar year for equally-weighted portfolios.

Consistent with La Porta (1996), the LLTG portfolio earns an average return of 15% in the year after formation, while the HLTG portfolio earns only 3%.[2] Adjusting for systematic risk

[2] The spread in Figure 1 is in line with, although smaller than, previous findings. LaPorta (1996) finds an average spread of 20% between the one year returns of HLTG and LLTG portfolios, but employed a shorter sample (1982 to

deepens the puzzle: the HLTG portfolio has higher market beta than the LLTG portfolio, and performs much worse in market downturns.[3] Over the past 35 years, 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, extrapolate and make systematic errors of excessive optimism for stocks with rapidly growing earnings, and conversely for stocks with deteriorating earnings.

Our goal is to describe the dynamics of expectation formation and to offer a psychologically founded theory that accounts simultaneously for the empirical patterns of fundamentals, expectations, and returns. We propose a new learning model in which beliefs are forward looking just as with rational expectations, but distorted by representativeness, which biases the interpretation of the news. Specifically, analysts update excessively in the direction of states of the world whose objective likelihood rises the most in light of the news. The model delivers extrapolation and over-reaction to news. It also makes sharp predictions that distinguish it from both Bayesian learning and mechanical adaptive expectations. We test, and confirm, several of these new predictions.

After describing the data in Section 2, in Section 3 we document three facts. First, HLTG stocks exhibit fast past earnings growth, which slows down going forward. Second, forecasts of future earnings growth of HLTG stocks are excessively optimistic, and are systematically revised downward later. Third, HLTG stocks exhibit good past returns but their average returns going forward are low. The opposite, but much less extreme, dynamics obtain for LLTG stocks, an asymmetry we do not account for in our model.

1991). Dechow and Sloan (1997) use a similar sample to La Porta (1996), though eliminating NASDAQ firms, and find a 15% spread. Due to subsequent back-filling by IBES, our sample in this subperiod is 25% larger than that in the earlier papers. We also check that this spread holds in sample subperiods (see Appendix A).

[3] We find $\beta_{HLTG} = 1.51$, and $\beta_{LLTG} = 0.78$ (Appendix A). The HLTG-LLTG spread holds within size buckets and it is strongest for intermediate B/M levels.

Section 4 offers a psychological model of learning about earnings based on Gennaioli and Shleifer's (GS, 2010) model of Kahneman and Tversky's (1972) representativeness heuristic. As in GS (2010) and Bordalo, Coffman, Gennaioli and Shleifer (BCGS, 2016), a trait t is representative of a group G when it occurs more frequently in that group than in a reference group $-G$. The representative trait t is quickly recalled and its frequency in group G is exaggerated. To illustrate, consider a doctor assessing the health status of a patient after a positive test. The representative patient is t = "sick", because sick people are more common among patients who tested positive than in the overall population. The sick patient type quickly comes to mind and the doctor inflates its probability, which in reality may be low if the disease is rare.

In the present setting, we think of analysts learning about firms' unobserved fundamentals on the basis of a noisy signal (e.g., current earnings). The rational benchmark here is the Kalman Filter. Relative to the Kalman Filter, representativeness causes analysts to inflate the probability of firm types whose likelihood has increased the most in light of recent earnings. By analogy with the medical example, past earnings (the test) render certain fundamentals representative (the health status), inducing overreaction to news. After exceptionally high earnings growth, analysts think a stock is a Google, but they imagine too many Googles relative to reality. Following our work on credit cycles (BGS 2017), we say that such distorted inference follows a "Diagnostic Kalman Filter" to emphasize that it overweighs information diagnostic of certain firm types.

Section 5 maps the model to the data. It starts by considering a key implication of the model, namely that expectations are forward looking, in the sense that they exaggerate the incidence of Googles in the HLTG group because these firms are *relatively* more likely there. In line with this prediction, we find that the HLTG group has a fatter right tail of strong future performers than do non-HLTG firms. These exceptional performers are thus representative of the

HLTG group, even though they are unlikely in absolute terms. As the model predicts, we also show that analysts vastly exaggerate the true share of firms with exceptional earnings growth in the HLTG group. The model also accounts for the dynamics of fundamentals, expectations, and returns documented in Section 3. Because strong past eps growth is diagnostic of future strong growth, analysts become excessively optimistic about HLTG firms, driving up prices and generating negative forecast errors. Returns are low post formation as analysts correct their inflated forecasts.

We develop and test additional new predictions of the model in Section 6. We first assess whether the inflated expectations about HLTG stocks are due to overreaction to news. To do so, we proxy analysts' news by their forecast revisions, and show that upward revisions in LTG forecasts are associated with excessive optimism, as if analysts over-react to good news (Coibion and Gorodnichenko 2015). Second, we present evidence that over-reaction influences returns: industries with stronger measured overreaction exhibit a larger return differential between LLTG and HLTG firms. Finally, we show that the dynamics of expectations are hard to explain using mechanical extrapolation: inflated expectations mean revert even without bad news, suggesting that analysts are forward looking in incorporating fundamental mean reversion into their forecasts.

Our paper is related to several strands of research in finance. Empirical research on cross-sectional stock return predictability is framed in terms of concepts such as extrapolation (e.g., DeBondt and Thaler 1985, 1987, Cutler, Poterba and Summers 1991, Lakonishok, Shleifer, and Vishny 1994, Dechow and Sloan 1997), but most studies in this area do not use expectations data. Some older studies in finance that do use expectations 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). More recently, the use of survey expectations data not just by analysts but also by investors has been making a comeback (e.g., Ben David, Graham and Harvey 2013, Greenwood and Shleifer 2014, Gennaioli, Ma, and Shleifer 2015).

Bouchaud, Landier, Krueger and Thesmar (2016) use analyst expectations data to study the profitability anomaly, and offer a model in which expectations under-react to news, in contrast with our focus on over-reaction. As we show in Section 6, our data also shows some under-reaction at short horizons, but at the long horizons of LTG forecasts over-reaction 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.

Our paper is also related to research on over-reaction and volatility, which begins with Shiller (1981), DeBondt and Thaler (1985, 1987), Cutler, Poterba, and Summers (1990), and DeLong et al. (1990a). This work assumes mechanical, backward looking rules for belief updating, based either on adaptive expectations (e.g., DeLong et al 1990b, Barsky and DeLong 1993, Barberis and Shleifer 2003, Barberis et al. 2015, Glaeser and Nathanson 2015), adaptive learning (Marcet and Sargent 1989, Adam, Marcet, and Nicolini 2016), or rules of thumb (Hong and Stein, 1999). Pastor and Veronesi (2003, 2005) present rational learning models in which uncertainty about the fundamentals of some firms boosts the volatility of their returns and their market to book ratios. Under some conditions, learning dynamics also explain predictability in aggregate stock returns (Pastor and Veronesi 2006). This approach does not analyze expectations data or cross sectional differences in returns. Barberis, Shleifer, and Vishny (BSV, 1998) and Daniel, Hirshleifer, and Subramanyam (DHS, 1998) offer models grounded in psychology. The

BSV model is motivated by representativeness, the DHS model by investor overconfidence. We return to these in Section 6.

One advantage of our approach worth stressing is that our model is not designed for a specific finance setting but is more portable (Rabin 2013). Our formalization of representativeness was developed to account for biases in general probability assessments such as base rate neglect, conjunction and disjunction fallacies in a laboratory context (Gennaioli and Shleifer 2010). We have previously applied it to modeling social stereotypes (Bordalo, Coffman, Gennaioli and Shleifer 2016, 2017) and more closely credit cycles (Bordalo, Gennaioli and Shleifer 2017), where the patterns of over-reaction to news, systematic forecast errors, expectations revisions, and predictable returns are similar to those discussed here.

II. Data and Summary Statistics

II.A. Data

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 2015. 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 2014, 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 2015. We present results for both buy-and-hold annual returns and daily cumulative-abnormal returns for various earnings' announcement windows. We compute annual stock returns by compounding monthly returns. We focus on equally-weighted returns for LTG portfolios. If a stock is delisted, CRSP tries to establish its price after delisting. Whenever a post-delisting price exists, we use it in the computations for returns. 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. After a stock disappears from the sample, we replace its return until the end of the calendar year with the return of the equally-weighted market portfolio. 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 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 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 the period 1978-2015. We use annual and quarterly accounting

[4] We form portfolios in December of each year, because that is when IBES data on analyst expectations is released. Unlike in Fama and French (1993) we know exactly when the information required for an investable strategy is public.

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*) 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 means of some of the variables for LTG decile portfolios. The number of stocks with IBES data on LTG varies by year, ranging from 1,222 in 1981 to 3,845 in 1997. On average, each LTG portfolio contains 246 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 (7,982 MM vs. 1,061 MM) and market capitalization (3,699 MM vs. 1,617 MM). However, differences in size are not extreme: the average size decile is 5.1 for LLTG and 3.6 for HLTG.

LLTG stocks have lower operating margins to asset ratios than HLTG stocks but higher return on equity (4% vs -6%). In fact, 36% of HLTG firms have negative eps while the same is true for only 16% 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 2014. 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. *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.

	LTG decile									
	1	2	3	4	5	6	7	8	9	10
Expected growth in eps (*LTG*)	4%	9%	11%	12%	14%	15%	17%	20%	25%	38%
Assets (MM)	7,993	10,828	9,617	8,020	5,214	3,658	2,472	1,938	1,356	1,082
Market capitalization (MM)	3,699	4,620	5,048	4,531	3,819	3,196	2,669	2,285	1,641	1,617
Size decile	5.1	5.2	5.4	5.4	5.2	4.7	4.6	4.0	3.8	3.6
Operating margin to assets	19%	24%	29%	33%	37%	40%	41%	42%	42%	37%
Return on equity	5%	8%	10%	10%	10%	9%	9%	7%	3%	-6%
Percent eps positive	84%	87%	89%	90%	89%	87%	87%	84%	77%	64%
Observations	257	255	248	246	247	258	223	253	241	231

III. A New Look at the Data.

Figure 1 above suggests 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. To assess this possibility, we document some basic facts connecting firms' performance, expectations and returns.

Figure 2 reports average earnings per share of HLTG and LLTG portfolios in years $t-3$ to $t+4$ 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 \$1.71 in year 0. Earnings of LLTG firms decline to \$0.88 during the corresponding period. But the past does not repeat itself after

portfolio formation, because of mean reversion in earnings growth. Earnings growth of HLTG firms slows down, while earnings of LLTG firms recover during the post-formation period. HLTG firms remain more profitable on average than LLTG firms 4 years after portfolio formation, but the difference in actual growth rates is nowhere near the difference in LTG exhibited in Table 1.

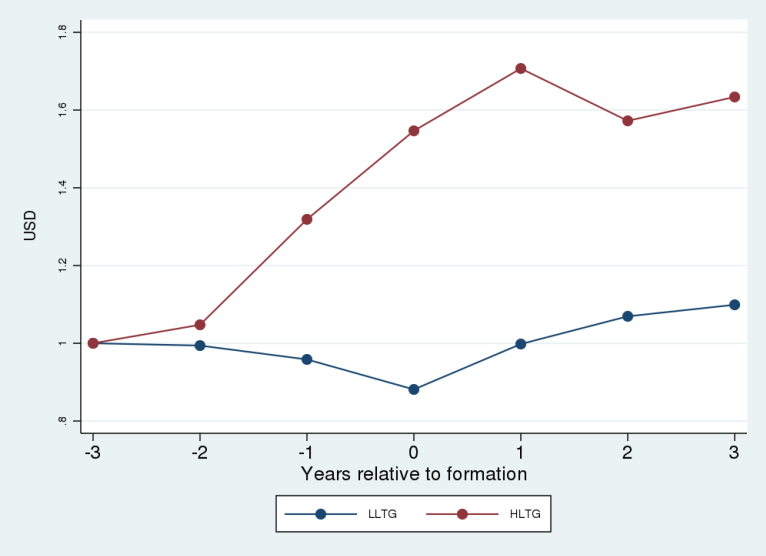


Figure 2. Evolution of EPS. In December of years (*t*) 1981, 1984., …, 2008, and 2011, we form decile portfolios based on ranked analysts' expected growth in earnings per share (LTG). We report the mean value of earnings per share for the highest (Hi) and lowest (Lo) LTG deciles for each year between 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.

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 sharply downwards, particularly during the first year, whereas LTG of LLTG firms is revised moderately 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. One potential concern about Figure 2 and in other Figures in this Section is the role

of attrition in the sample. Reassuringly, the level of attrition is similar across HLTG and LLTG portfolios (roughly 30% five years post formation), and the key features of the Figures continue to hold when we restrict the sample to firms that survive for five years post formation.

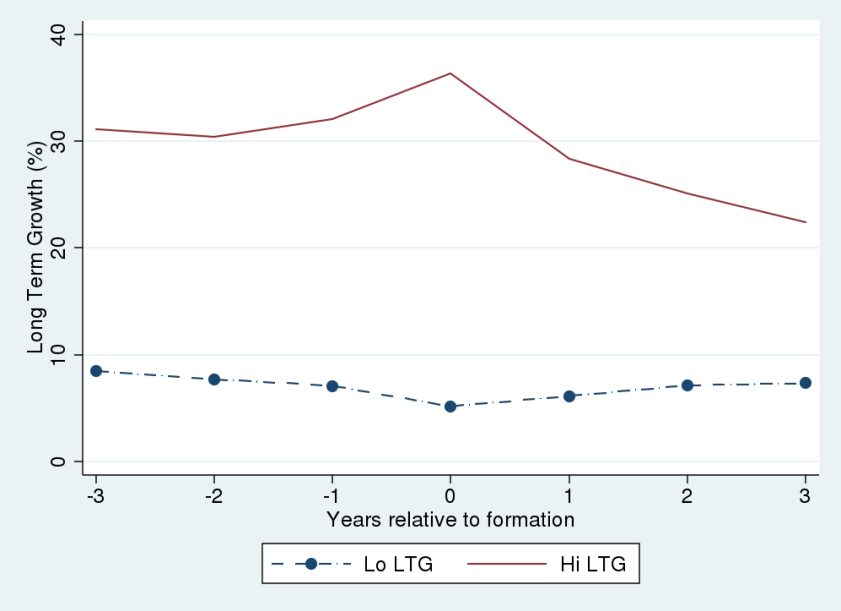


Figure 3. Evolution of LTG. In December of each year t between 1984 and 2011, we form decile portfolios based on ranked analysts' expected growth in earnings per share (LTG) and report the mean value of LTG on December of years t-3 to t+4 for the highest (Hi) and lowest (Lo) LTG deciles. We include in the sample stocks with LTG forecasts in year t-3. Values for t+1, t+2, t+3, and t+4 are based on stocks with IBES coverage for those periods.

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' capacity for 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 look at expectations errors, Figure 4 reports the difference between realized earnings growth and analysts' LTG expectations at formation in each portfolio, from formation to year $t + 4$. The data show dramatic over-optimism, i.e. very negative forecast errors, for HLTG firms over this period. It also shows modest 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).

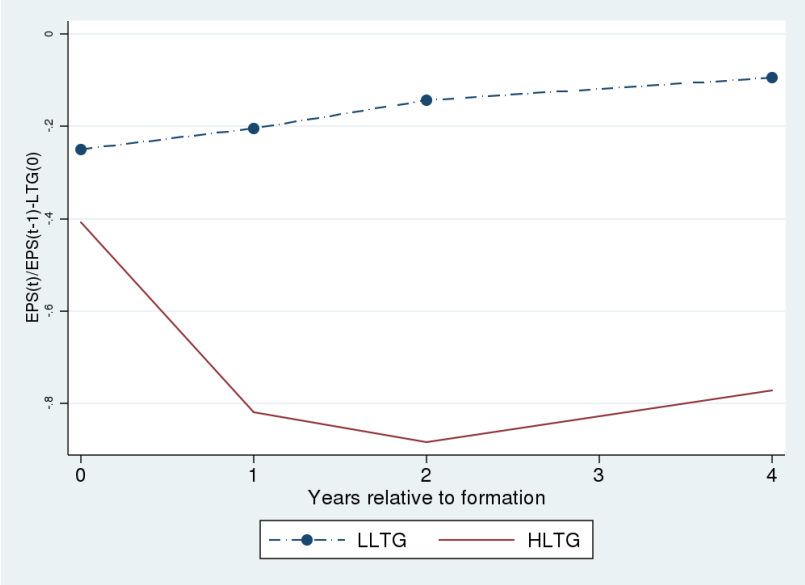


Figure 4. Realized earnings per share vs. Expected growth in long-term earnings. In December of each year *t* between 1981 and 2009 we form decile portfolios based on ranked analysts' expected growth in eps. We plot the difference between the annual growth rate in earnings per share in each year between t and t+4 and the forecast for long-term growth in earnings made in year *t*.

The overestimation of earnings growth for HLTG firms is economically large. By year 4, actual earnings are a small fraction of what analysts forecast: earnings per share grow from 0.16 upon formation to 0.21 in $t + 4$, compared to the prediction of 0.70 based on LTG at formation.

There are two potential concerns with Figure 4. First, analyst errors might be due to distorted incentives rather than beliefs. Second, analyst beliefs may not be shared by investors. To address these concerns we look at the dynamics of stock returns pre and post formation.

Figure 5 shows stock returns around earnings announcements. 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).

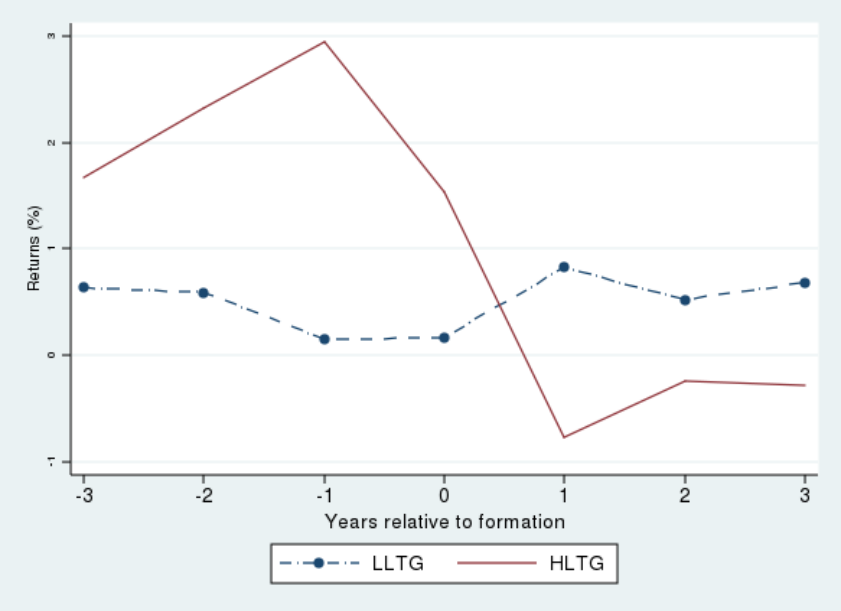


Figure 5. Twelve-day Returns on Earnings Announcements for LTG Portfolios. In December of each year t between 1981 and 2011, we form decile portfolios based on ranked analysts' expected growth in earnings per share. Next, for each stock, we compute the 3-day market-adjusted return centered on earnings announcements in years $t-3, \ldots, t+4$. 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 average annual return during earnings announcements for the highest (Hi) and lowest (Lo) LTG deciles. Excess returns are defined relative to the equally-weighted CRSP market portfolio.

HLTG stocks appear to positively surprise investors with their earnings announcements in the years prior to portfolio formation, when there is upward revision of LTG.[5] Returns are low afterwards, especially in year 1, consistent with the sharp decline in LTG in this period. This suggests that the over-optimism reported by analysts is shared by investors, so HLTG stocks

[5] The persistence of positive surprises and high returns exhibited by the HLTG portfolio in Figure 5 should not be confused with time-series momentum. Instead, it arises from selection: on average, firms require a sequence of positive shocks to be classified as HLTG. The basic HLTG-LLTG spread is orthogonal to short-term return momentum in the 6 months prior to portfolio formation, see Appendix A.

consistently disappoint in the post formation period. The converse holds for LLTG stocks, but in a much milder form.

Overall, this evidence provides three insights into the analyst expectations puzzle that play a key role in our theoretical analysis. First, analysts react to news. Optimism about HLTG firms follows the observation of fast earnings growth, and is reversed after earnings' growth slows down. It seems that analysts are trying to learn the earnings generating capacity of firms on the basis of the firm's observed performance. Second, there are reasons to be skeptical of analyst rationality, as shown by the evidence on systematic forecast errors in the HLTG group (which comprises more than 200 firms). Third, the dynamics of returns match the dynamics of expectations and their errors, both pre and post formation.

In the next section, we develop a model based on psychological first principles that sheds light on this evidence. This model takes the perspective that the consensus analyst forecast are the relevant expectations that shape prices and whose dynamics drive returns.[6]

IV. A model of learning with representativeness

IV.A. The Setup

There is a measure 1 of firms, $i \in [0,1]$. At time t, the natural logarithm of a firm's earnings per share (eps) $x_{i,t}$ is given by:

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

[6] We do not draw a distinction between analyst forecasts and investor expectations, and in particular neglect potential conflicts of interests between analysts and investors.

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

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

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 that may for instance result from a corporate reorganization or a change in market competition. 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 the additional condition $b \leq a$.

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 $\left(x_{i,u}\right)_{u \leq t}$. 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:[7]

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

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

The idea is intuitive: 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 earnings surprise $x_{i,t} - bx_{i,t-1} - a\hat{f}_{i,t-1}$. The extent of this adjustment increases in K. Absent transitory shocks ($\sigma_\varepsilon^2 = 0$), earnings are

[7] Equation (3) arises in the long run, when the variance of fundamentals has converged to its steady state σ_f^2. Given the presence of fundamental shocks $\eta_{i,t}$, a firm's fundamental is never learned with certainty. The passage of time exerts two opposite effects on the variance of fundamentals. On the one hand, it smooths transitory shocks, causing variance to decline. On the other hand, it accumulates fundamental shocks, causing the variance to increase. In the long run, these two forces balance out and variance stays constant at σ_f^2, which is defined as the solution to:

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

perfectly informative about fundamentals and the adjustment is full (i.e., $\hat{f}_{i,t} = x_{i,t} - bx_{i,t-1}$). As the importance of transitory shocks rises, earnings become a noisier signal and estimated fundamentals change less with earnings, so $K < 1$.

The signal to noise ratio solves the key inference problem here: to separate the extent to which current earnings $x_{i,t}$ are due to persistent or transitory shocks (i.e., $\eta_{i,t}$ versus $\varepsilon_{i,t}$). Among episodes of exceptionally high growth, the analyst must try to tell apart those due to luck and those due to the fact that the firm is the next Google.

Equation (3) yields not only the true conditional distribution of fundamentals, but also the assessment of fundamentals performed by a Bayesian agent seeking to forecast future earnings. We next describe how the representativeness heuristic distorts this learning process.

IV.B. Representativeness and the Diagnostic Kalman filter

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 of that group. 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.

Gennaioli and Shleifer (2010) propose a model of this phenomenon in which a decision maker assesses the distribution $h(T = \tau|G)$ of a variable T in a group G. The representativeness of the specific type τ for G is:

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

As in KT, a type is more representative if it is *relatively* more frequent in G than in the comparison group $-G$. The probability of representative types is overestimated relative to the truth. BCGS (2016) offer a convenient formalization of this process by assuming that probability judgments are formed using the distorted density:

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

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 the group G. The distorted probability in (5) thus depends on the true probability, which renders the model empirically testable. 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. This leads to belief distortions: representative types play an outsized role in judgments, while other types are relatively neglected.

This setup can be applied to prediction and inference problems (as in BCGS 2016 and BGS 2016). Consider the example from the introduction of a 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 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. Thus, the doctor may 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 $x_{i,t}$, or more precisely the current earnings surprise $x_{i,t} - bx_{i,t-1} - a\hat{f}_{i,t-1}$. This 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 $x_{i,t} - bx_{i,t-1} = a\hat{f}_{i,t-1}$. Formally, the comparison distribution of fundamentals is normal with mean $a\hat{f}_{i,t-1}$ and variance σ_f^2. This is the true distribution at t with no earnings surprise, and corresponds to $h(T = \tau| - G)$.

For firm i at time t, the prior believed fundamental is $a\hat{f}_{i,t-1}$. Suppose that after earnings surprise $x_{i,t} - bx_{i,t-1} - a\hat{f}_{i,t-1}$, the posterior is $\hat{f}_{i,t}$. Given normality, then, the representativeness of fundamental f for firm i at date t is:

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

If news are good, in the sense that they rationally imply better fundamentals, $\hat{f}_{i,t} > a\hat{f}_{i,t-1}$, representativeness is higher for higher types f. After bad news, implying $\hat{f}_{i,t} < a\hat{f}_{i,t-1}$,

representativeness is higher for lower types f. In the first case, high types are overweighed while low types are underweighted in judgments. In the latter case, the reverse is true.

The Appendix shows that these distortions generate diagnostic beliefs as follows:

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

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

When analysts overweight representative types, their beliefs resemble the optimal Kalman filter, but with a key difference: they exaggerate the signal to noise ratio, 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.[8] The psychology is in fact very different from overconfidence: in our model, as in the medical test example, overreaction is caused by neglect of base rates. After good news, the most representative firms are Googles. This firm type readily comes to mind and the analyst exaggerates its probability, 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. Exaggeration in the reaction to news increases in θ. At $\theta = 0$ the model collapses to rational learning.[9]

The key property of diagnostic expectations is "the kernel of truth": distortions in beliefs exaggerate true patterns in the data. The kernel of truth distinguishes our approach from alternative theories of extrapolation such as adaptive expectations or BSV (1998). As we map the

[8] In fact, overconfidence predicts under-reaction to public news such as earnings releases (see Daniel et al. 1998).
[9] The symmetric nature of Equation (6) for good and bad news depends on the assumption of normality. Relaxing normality may improve the fit with the data.

model to the facts of Sections 1 and 3, we first show that the kernel of truth is consistent with the data: Googles are overweighed in the HLTG portfolio because they occur much more often there than elsewhere.

V. The Model and the Facts

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

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

Expectations of long term growth are shaped by mean reversion in eps and fundamentals. LTG is high when firms have experienced positive news, so $\hat{f}^{\theta}_{i,t}$ 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 begin by testing for the kernel of truth. To this end, we first report in Figure 6 the true distribution of future eps growth for the HLTG portfolio (blue curve) against the distribution of future eps growth of all the other firms (orange curve).

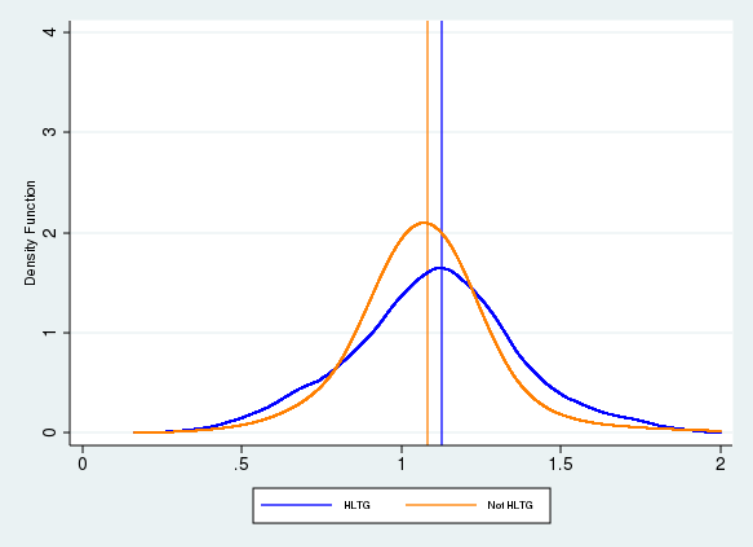


Figure 6 Kernel density estimates of growth in earnings per share for LTG Portfolios. In December of years (*t*) 1981, 1986, …, 2001, and 2006, we form decile portfolios based on ranked analysts' expected growth in long term earnings per share (*LTG*). For each stock, we compute the annual growth rate of operating margin per share between *t* and *t*+5, where operating margin is defined as revenues minus cost of goods sold. We exclude stocks with negative operating margin in year *t* and we estimate the kernel densities for stocks in the highest (*Hi*) decile and for all other firms. The graph shows the estimated density kernels of growth in operating margin per share for stocks in the *Hi* (blue line) and other (orange line) LTG portfolios. The vertical lines indicate the means of each distribution.

Two findings stand out. First, HLTG firms indeed have a higher average future eps growth than all other firms, as we saw in a somewhat different format in Figure 2. But second, and critically, the HLTG group displays a significantly fatter right tail of exceptional performers. Googles are thus representative for HLTG in the sense of definition (4). In fact, based on the densities in Figure 6, the most representative future growth realizations for HLTG firms are in the range of 40% to 60% annual growth.[10]

In light of these data, our model predicts that analysts should over-estimate the number of right-tail performers in the HLTG group. Figure 7 compares the distribution of future performance

[10] Although HLTG firms tend to have also a slightly higher share of low performers, it is true that, as in our model, higher growth rates are more representative for HLTG firms. See Figure B.1 in Appendix B.

of HLTG firms (blue line) with the predicted performance for the same firms (red line).[11] Consistent with diagnostic expectations, 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.[12] As a robustness check, we reproduce in Appendix B Figures 6 and 7 using as a measure of fundamentals revenues minus cost of goods sold (which may be less noisy that eps). With this metric as well, the evidence supports the kernel of truth hypothesis.

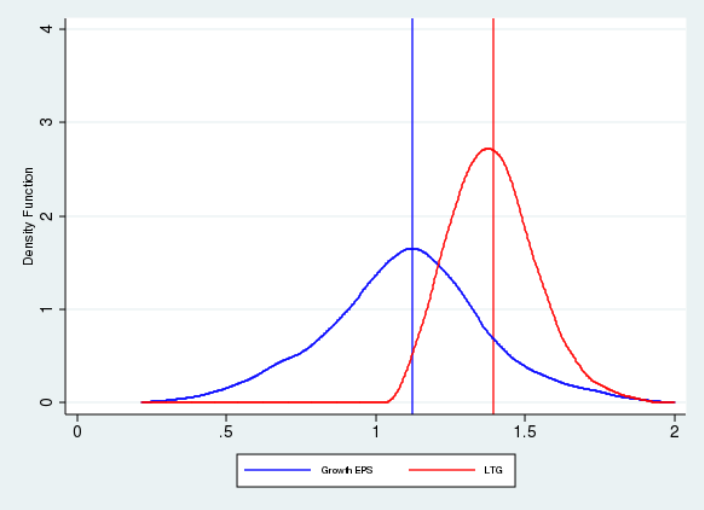


Figure 7. Realized vs. Expected Growth in eps. In December of years (*t*) 1981, 1986,…, 2001, and 2006, we form decile portfolios based on ranked analysts' expected growth in long term earnings per share (LTG). We plot two series. First, we plot the kernel distribution of the annual growth rate in operating margin per share between *t* and *t*+5, where operating margin is defined as revenues minus cost of goods sold. 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*.

[11] 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 the context of 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.

[12] The kernel of truth can also shed light on the asymmetry between HLTG and LLTG firms. In Appendix B we show that future performance of LLTG firms tends to be concentrated in the middle, with a most representative growth rate of 0%. Thus constant, rather than bad, performance is representative of LLTG firms, which can help explain why expectations about these firms and their market valuation are not overly depressed. This empirical asymmetry between HLTG firms (representative right tail) and LLTG firms (representative middle of the road performance) may be captured by relaxing the assumption of normality, or alternatively by allowing lower volatility for firms in the LLTG group.

We next show that the model accounts for the previously documented facts. To explore the dynamics of LTG in our model, we focus on the long run distribution of fundamentals $f_{i,t}$ (which has zero mean and variance $\frac{\sigma_\eta^2}{1-\alpha^2}$) and of analysts' mean beliefs $\hat{f}_{i,t}^{\theta}$ (which has zero mean and variance $\sigma_{f^\theta}^2$).[13] In line with our empirical analysis, at time t we identify the high LTG group $\mathrm{HLTG_t}$ as the 10% of firms with highest believed fundamentals, and hence with highest assessed future earnings growth, and the low LTG group $\mathrm{LLTG_t}$ as the 10% of firms with lowest believed fundamentals and hence lowest assessed future earnings growth.

V.A. Representativeness and the Features of Expectations

We first review 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.

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*

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

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

In our model, positive earnings surprises have two conflicting effects on long term growth prospects and thus on LTG. On the one hand, they raise estimated fundamentals $\hat{f}_{i,t}^{\theta}$, which

[13] There is no distortion in the average diagnostic expectation across firms because in steady state there are no systematic earnings surprises: the average earnings news in the population of firms is zero. As a consequence, 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.

enhance future growth. On the other hand, they lower future growth via mean reversion. Condition (7) ensures that the former effect dominates, so that firms with rosy future prospects (HLTG) are selected from those that have experience good recent performance, while firms with bad prospects (LLTG) are selected from those that have experienced bad recent performance.

The parametric restriction of Condition (7) is more likely to hold the less severe is mean reversion (i.e., when b is close to 1). It is also more likely to hold the higher is θ. Intuitively, this enhances the update of $\hat{f}^{\theta}$ in light of past growth, strengthening the first effect above.[14] The condition is more likely to hold the larger is the signal to noise ratio K and the larger is a, provided b is large ($b > a/(1-a)$). Note that condition (7) depends on the parameters of the earnings process because analysts do not mechanically extrapolate past performance.

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.

We next show that the model can account for the fact documented in Figure 4, namely that expectations for the long term growth of HLTG firms are excessively optimistic.

Proposition 3. *If analysts are rational, $\theta = 0$, they make no systematic error in predicting the log growth of earnings of* HLTG_t *and* LLTG_t *portfolios:*

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

Under diagnostic expectations $\theta > 0$, in contrast, analysts systematically over-estimate growth in HLTG_t *and under-estimate growth in* LLTG_t*:*

[14] The restriction $K(1+\theta) < 1$ only ensures positive autocorrelation of diagnostic beliefs at t and $t-1$.

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

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. 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 3, the prediction of systematic pessimism about LLTG firms is not borne out in the data.

Finally, our model also 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:*

$$\mathbb{E}\left(\mathrm{LTG}_{t+s}^{\theta=0} | HLTG_t\right) - \mathbb{E}\left(\mathrm{LTG}_t^{\theta=0} | HLTG_t\right) < 0$$

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

$$\mathbb{E}\left(\mathrm{LTG}_{t+s}^{\theta\geq 0} | HLTG_t\right) - \mathbb{E}\left(\mathrm{LTG}_t^{\theta>0} | HLTG_t\right) = \mathbb{E}\left(\mathrm{LTG}_{t+s}^{\theta=0} | HLTG_t\right) - \mathbb{E}\left(\mathrm{LTG}_t^{\theta=0} | HLTG_t\right) - \theta\Psi$$

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

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. The same force causes the excess pessimism of LLTG firms to dissipate, strengthening reversal in this portfolio.[15]

V.B. The diagnostic Kalman filter and returns

Consider now the return patterns documented in Figures 1 and 5. To explore the implications of diagnostic expectations for portfolio returns, we take the required return $R > 1$ as given. The pricing condition for a firm i at date t is then given by:

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

so that the stock price of firm i is the discounted stream of all 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:

[15] Going forward, firms in the HLTG portfolio receive neither positive nor negative true surprises on average, namely $\mathbb{E}_t(x_{i,t+1} - bx_{i,t-1} - a\hat{f}_{i,t}) = 0$. As a result, by Equation (6), their assessed fundamentals on average coincides with the rational value $a\hat{f}_{i,t}$. The same logic triggers mean reversion in the LLTG portfolio.

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

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

Because firms in the HLTG portfolio on average receive positive news before formation (Proposition 2), they earn returns higher than the required return R. This is also true under rational expectations. However, diagnostic expectations at t are too high for HLTG and too low for LLTG (Proposition 3). Propositions 2 and 5 thus predict positive abnormal pre-formation returns for HLTG stocks, as well as the low pre-formation returns of LLTG stocks, as in Figure 5.

Yet the post-formation returns in Figure 5, also illustrated in Figure 1, still need to be explained. To do so, we compute the average realized return going forward (according to the true probability measure) for a given firm i:

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

The average realized return at $t+1$ is below the required return R when at t investors over-value the future expected price and dividend of firm i, namely when the denominator is larger than the numerator in Equation (9). 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 $\mathbb{E}_t\left(R_{t+1}^{\theta}|P\right)$ the average future return of portfolio $P = \text{HLTG},\ \text{LLTG}$ at $t+1$. Under rationality, $\theta = 0$, excess returns are not predictable:*

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

Diagnostic expectations generate predictable excess returns:

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

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 the finding a large portfolio of firms sharing a certain forecast are on average overvalued. In that model, conditional on current information, forecasts are on average (across firms) correct and returns are unpredictable.

Under the diagnostic Kalman filter ($\theta > 0$), in contrast, the HLTG portfolio exhibits abnormally high returns up to portfolio formation and abnormally low returns after formation, and conversely for the LLTG portfolio, just as we saw in Table 1 and Figure 1. After formation, however, expectations in the portfolio revert back on average. The inflated fundamentals give way to a less extreme assessment. On average, then, HLTG investors are disappointed and returns are abnormally low, which pins down the inequality among the post formation returns.

To summarize, the return predictability documented in Figure 1 can be accounted by representativeness but not by rational learning. Rational learning can account for the pre-formation return patterns (Proposition 5), but it does not yield post formation reversals (Proposition 6) whereby HLTG stocks underperform LLTG stocks.

VI. Additional Predictions of the Model

Diagnostic expectations yield a coherent account of the dynamics of news, analyst expectations, and returns documented in Section 3. Two issues remain open. The first is whether

analyst expectations and returns indeed over-react to news. The challenge of testing this hypothesis is that news are hard to measure, as they include but are not restricted to observed earnings. The second issue is whether our model improves upon existing theories of extrapolation such as Barberis, Shleifer and Vishny's (BSV 1998) model of investor sentiment, or adaptive expectations. This section addresses these issues by highlighting additional implications of our model.

VI.A Overreaction to News and Returns

Coibion and Gorodnichenko (2015) propose measuring the information received by the forecaster at time t using their forecast revision at t. The forecast revision can in fact be interpreted as a summary index for all information received by the forecaster in the recent past. Forecasters' over or under reaction to information can be assessed by correlating their current forecast revision with the subsequent forecast error.

Coibion and Gorodnichenko show that, in sticky information or rational inattention models (e.g., Sims 2003, Mankiw and Reis 2002), consensus forecast revisions should positively correlate with subsequent consensus forecast errors. Intuitively, when expectations under-react, a positive forecast revision indicates insufficient upward adjustment. As a result, it should predict positive errors (i.e. realizations systematically above the forecast). Bouchaud et al. (2016) use this method to diagnose under-reaction to news about firms' profitability.

The same approach turns out to be useful for our analysis as well.

Proposition 7. *Consider the following firm level regression:*

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

The estimated coefficient β is negative under diagnostic expectations, $\theta > 0$, and zero under rational expectations, $\theta = 0$.

A negative coefficient β in (10) corresponds to a positive coefficient θ in the diagnostic expectation formula of Equation (6). If analysts overreact to news, as predicted by diagnostic expectations, an upward revision in LTG is symptomatic of excessive adjustment, which in turn predicts a negative forecast error (LTG above realized growth). Under rational expectations, forecast errors are unpredictable at the individual level.[16]

Table 2 below 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. We allow horizon h to vary because LTG refers to growth over a period between 3 to 5 years. To estimate β we use consensus forecasts rather than individual analyst estimates because many analysts drop out of the sample.[17]

Table 2: Coibion-Gorodnichenko regressions for EPS

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

	Dependent Variable		
	$(eps_{t+3}/\ eps_t)^{1/3}$-$LTG_t$	$(eps_{t+4}/\ eps_t)^{1/4}$-$LTG_t$	$(eps_{t+5}/\ eps_t)^{1/5}$-$LTG_t$
LTG_t-LTG_{t-1}	-0.0351	-0.1253[c]	-0.1974[a]
	(0.0734)	(0.0642)	(0.0516)

[16] A positive estimate of β would suggest, in our model, that analysts discount highly representative types. This is equivalent to having a negative θ in Equation (5).

[17] Estimating (10) on the consensus LTG may give the misleading impression of under-reaction if individual analysts observe different noisy signals, so that there is dispersion in their forecasts. For instance, Coibion and Gorodnichenko show that when different analysts observe different realizations of a noisy signal, the coefficient in Equation (10) is positive even if each analyst rationally revises his forecast. In this respect, finding negative β in a consensus regression is even stronger evidence of over-reaction to information.

LTG_t-LTG_{t-2}	-0.2335[a]	-0.2687[a]	-0.2930[a]
	(0.0625)	(0.0602)	(0.0452)
LTG_t-LTG_{t-3}	-0.2897[a]	-0.2757[a]	-0.3127[a]
	(0.0580)	(0.0565)	(0.0437)

Consistent with diagnostic expectations, upward LTG revision predicts excess optimism, pointing to over-reaction to news. This holds regardless of the forecast horizon h, so the pattern is robust to alternative interpretations of LTG. The estimated β tends to become larger in magnitude and more statistically significant at longer forecast horizons $h = 3,4,5$ (i.e. as we move from left to right in Table 2), perhaps reflecting the difficulty of projecting growth into the future.[18]

Interestingly, the estimated β also gets higher in magnitude and more statistically significant as we lengthen the revision period $k = 1,2,3$ (i.e. moving from top to bottom in Table 2). We view this evidence as being consistent with the kernel of truth. From Equation (6), over-reaction to information of the diagnostic filter $\hat{f}_{i,t}^{\theta}$ compared to the rational filter $\hat{f}_{i,t}$ is given by:

$$\hat{f}_{i,t}^{\theta} - \hat{f}_{i,t} = K\theta\left(x_{i,t} - bx_{i,t-1} - a\hat{f}_{i,t-1}\right),$$

where K is the true signal to noise ratio of information accruing during the revision period. It is plausible that persistent signals over 2 or 3 years have are objectively more informative about future earnings, i.e. have a higher K, than occasional signals accruing over one year. By the kernel of truth, then, such signals should induce more over-reaction, consistent with the data.

[18] This feature helps reconcile our evidence with the sluggishness documented by Bouchaud et al. (2016). They consider forecasts for the level of eps over short horizons such as 1 or 2 years. In Appendix C (see Table C.2) we show that at these horizons there is some evidence of under-reaction also in our data (with the usual caveat of the under-reaction bias entailed in estimating consensus regressions). The seemingly contradictory findings of over and under-reaction can be reconciled by combining diagnostic expectation with some rigidity in analyst forecasts, stemming for instance from sporadic revision times. In this case, a piece of news would initially trigger few adjustments, generating short term aggregate under-reaction, but will lead to overreaction as all analysts update.

We next try to tie over-reaction to news to the return spread between HLTG and LLTG portfolios. To the extent that over-reaction drives excess optimism about HLTG, and thus lower post formation returns, one may conjecture that stronger over-reaction should be associated with larger return spreads between HLTG and LLTG portfolios. To test this idea, 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×year fixed effects. This yields an industry level estimate β_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 8, we compare the post formation return spread across different terciles of the distribution of industry β_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 β_s. The return differential is sizable, though given the small sample size it is not statistically significant.

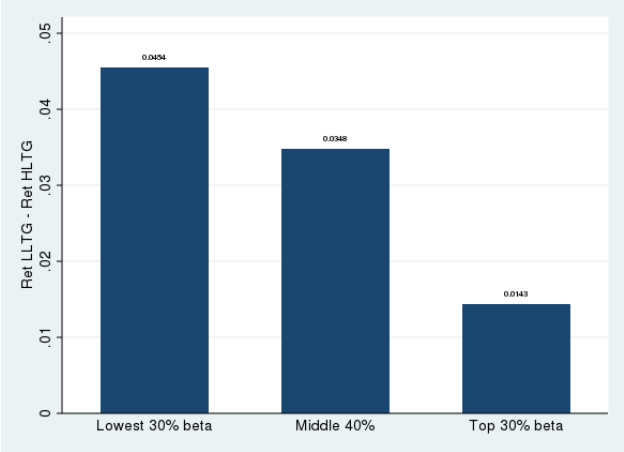


Figure 8. Overreaction and return spread across industries. For each of the 48 Fama-French industries, we estimate the regression: $(EPS_{t+4}/EPS_t)^{(1/4)} - (1+LTG_t) = \alpha_i + \mu_{i,t} + \beta_i * (LTG_t - LTG_{t-3}) + \varepsilon_{i,t}$ where μ_t are year fixed effects, EPS is earnings per share, and LTG is the forecast long-term growth in earnings. We rank industries according to β_is and form the following three groups: (1) 14 industries with the lowest β_is, (2) 24 industries with

intermediate values of β_is, and (3) 14 industries with the highest β_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 β_is from the regression.

Overall, this subsection offers evidence in support of two ideas. First, expectation formation about LTG features over-reaction to news, consistent with diagnostic expectations, but inconsistent with rational inattention or other theories of under-reaction.[19] Second, 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.

VI.B Alternative Mechanisms for Overreaction

We conclude by comparing diagnostic expectations with alternative models of over-reaction to news, such as the BSV model of investor sentiment and mechanical extrapolation. BSV is an early attempt to formalize the psychology of representativeness. It assumes that the true process driving a firm's earnings is a random walk, but analysts perform Bayesian updating across two incorrect models, one where earnings are believed to trend and one where they mean revert. Over-reaction occurs because periods of fast earnings growth induce the analyst to attach a high probability that the firm is of the "trending type", even though no firm is actually trending.

Our model captures the key intuition of BSV: after good performance analysts place disproportionate weight on strong fundamentals, and the reverse after bad performance. It has, however, two main advantages relative to its antecedent. First, in the BSV model extrapolation follows from belief in models. In contrast, our model yields the kernel of truth: the HLTG group

[19] The evidence of the over-reaction to news of consensus LTG forecasts is also inconsistent with theories of over-reaction based on analyst overconfidence such as Daniel et al. (1998). In these models, analysts over-react to their private information but under-react to common information, generating under-reaction of consensus forecasts.

features a relatively higher share of Googles, and analysts exaggerate this share in their assessment. This means, in contrast to the above, that belief distortions can be predicted from the data. Figures 6 and 7 are indeed consistent with this prediction. The second advantage of our model is portability: it is not designed for a specific finance setting, and so it can be easily applied to probability judgments, learning contexts or stereotyping.

The other conventional approach to over-reaction, mechanical extrapolation, implies that LTG is formed as a distributed lag of past earnings growth rates, following the adaptive rule:

$$x_{t+1}^{ad} = x_t^{ad} + \mu\left(x_t - x_t^{ad}\right), \tag{11}$$

where x_t^{ad} is the expectation held at t about the level or growth of eps at a certain period, x_t is the current realized level of growth of earnings, and $\mu \in [0,1]$ is a fixed coefficient. If μ is low, expectations under-react to news. But if μ is large relative to the persistence of the earnings process, expectations can over-react to news.

The difference between our model and mechanical extrapolation is that diagnostic expectations are forward looking. Under the mechanical rule of Equation (11), analysts revise growth expectations downward if and only if bad news arrive, namely if $\left(x_t - x_t^{ad}\right) < 0$. In contrast, under diagnostic expectations decision makers are influenced by the features of the data generating process such as the true share of Googles and the mean reversion of earnings. For instance, when considering firms that have grown fast in the past, such as HLTG ones, growth forecasts will cool of over time *even if no news is received*.

In fact, the revision of believed fundamentals from one period to the next is given by:

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

$$-(1-a)\hat{f}_{i,t} - K\theta\left(1 - aK(1+\theta)\right)\left(x_{i,t} - bx_{i,t-1} - a\hat{f}_{i,t-1}\right)$$

The revision depends in part on the surprise relative to diagnostic expectations, namely on $x_{i,t+1} - bx_{i,t} - a\hat{f}^{\theta}_{i,t}$. But even in the absence of surprising earnings, namely when $x_{i,t} - bx_{i,t-1} = a\hat{f}_{i,t-1}$, beliefs about fundamentals are updated. This is partially due to mean reversion of fundamentals (i.e. the second term $-(1-a)\hat{f}_{i,t}$). But it is also due to the waning of over-reaction to previous shocks (i.e. the third term $x_{i,t} - bx_{i,t-1} - a\hat{f}_{i,t-1}$). For HLTG stocks, both forces point toward a downward revision of believed fundamentals, regardless of the current news received, while for LLTG stocks the opposite holds. This leads to systematic mean reversion of LTG estimates for these portfolios. In contrast, no systematic mean reversion should be expected under adaptive expectations.

To test this prediction, we consider the change in LTG around earnings announcement dates. We rank earnings of all firms surprises into deciles and follow LTG revisions for each decile. The results are reported in Figure 10 below.

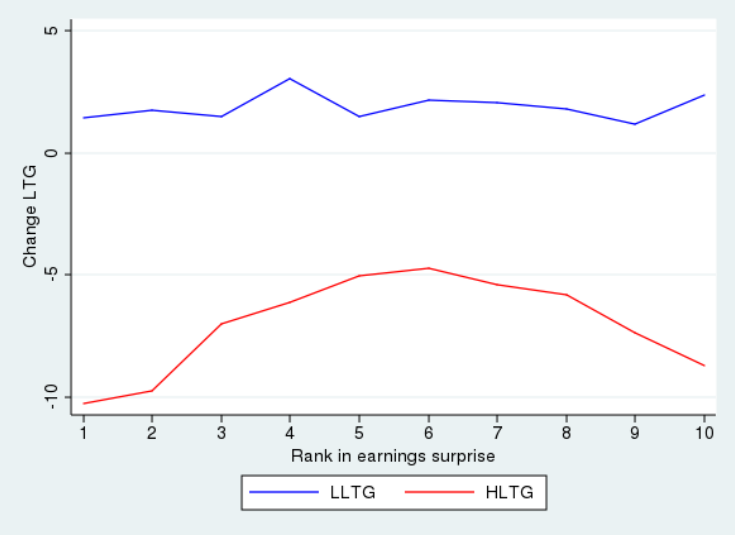


Figure 10. Evolution of Analysts' Beliefs in Response to Earnings' Announcements. For each analyst *j*, firm *i*, and fiscal year *t*, we compute the difference between the first-available LTG forecast made during the 45-90 day window following the earnings' announcement for *t-1* and the first-available LTG forecast made during the 45-90 day window following the earnings' announcement for *t*. We rank observations into deciles based on the ratio of the forecasting error for earnings per share in year *t* to the stock price when that forecast was made. We measure forecasting errors using the first-available forecasts for earnings per

share during the 45-90 day window following the earnings' announcement for *t-1*. The Figure reports the sample average for all observations, for portfolios HLTG and LLTG.

The data show strong evidence of systematic mean reversion of expectations. Regardless of the experienced earnings surprise, expectations about HLTG firms deteriorate while those about LLTG ones improve. This implies that the reversal of Figure 3 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 fully consistent with the forward looking nature of diagnostic expectations.[20]

VII. Conclusion.

This paper revisits what since Shiller (1981) has been perhaps the most basic challenge to rational asset pricing, namely 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. LaPorta (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

[20] In the Appendix, we show that adaptive expectations predict no over-reaction to news after the persistence of the earnings process is accounted for. In particular, after controlling for current levels x_t, the adaptive forecast revision $(x^a_{t+1} - x^a_t)$ 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. Table C.1 in Appendix C reports the coefficient on forecast revision obtained when controlling for current growth in earnings in Equation (10). The coefficients on forecast revision become larger than those estimated in Table 2, but they remain mostly negative and statistically significant.

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. Most important, the theory yields a number of strong empirical predictions, which have not been considered before, but which we have brought to the data. Although some puzzles remain, the evidence is supportive of the proposed theory.

Of course, this is just a start. The approach to expectation formation we have proposed 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. One could also perform a quantitative evaluation of the model. 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.

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:

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

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

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

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:

$$-(f - \hat{f}_{i,t})^2 + \theta(\hat{f}_{i,t} - a\hat{f}_{i,t-1})(2f - a\hat{f}_{i,t-1} - \hat{f}_{i,t})$$
$$= -\left(f - \left(\hat{f}_{i,t} + \theta(\hat{f}_{i,t} - a\hat{f}_{i,t-1})\right)\right)^2 + O(\hat{f}_{i,t}, \hat{f}_{i,t-1})$$

Taking normalization into account, we find

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

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

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

■

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:

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

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:

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

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

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

is positive. For convenience, rewrite this as:

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

It is useful to rewrite the first term as:

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

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

$$\begin{aligned} cov\left(x_{i,t} - x_{i,t-1}, a(1-K)\hat{f}_{i,t-2} + K\left(f_{i,t-1} + \epsilon_{i,t-1}\right)\right) \\ = a(1-K)cov\left(x_{i,t} - x_{i,t-1}, \hat{f}_{i,t-2}\right) - (1-b)Kcov\left(x_{i,t-1}, f_{i,t-1}\right) \\ - (1-b)Kvar(\epsilon) + aKvar\left(f_{i,t}\right) \end{aligned}$$

We can show that:

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

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

$$cov\left(x_{i,t} - x_{i,t-1}, a(1-K)\hat{f}_{i,t-2} + K\left(f_{i,t-1} + \epsilon_{i,t-1}\right)\right)$$

$$> var\left(f_{i,t}\right)\left[\frac{aK}{1-a^2(1-K)} - (1-b)\left(K + a^2(1-K) + \frac{ab}{1-ab}\right)\right]$$

$$-(1-b)Kvar(\epsilon)$$

So putting the two terms together we find

$$var(\epsilon)\left[\frac{A}{1+b} - (1-b)KB\right]$$

$$+ var\left(f_{i,t}\right)\left[\frac{A}{1+b}\frac{1-a}{1-ab}\right.$$

$$\left. + B\left[\frac{aK}{1-a^2(1-K)} - (1-b)\left(K + a^2(1-K) + \frac{ab}{1-ab}\right)\right]\right]$$

A sufficient condition that makes both terms positive is:

$$\vartheta_h K' - \varphi_h - (1-b)(1+a)\vartheta_h K > 0$$

This condition is easier to satisfy for large b (it is trivially satisfied when $b = 1$), and for large θ. Moreover, for $b > a/(1+a)$, this condition is also easier to satisfy for larger a.■

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

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

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

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

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

$$\mathbb{E}\left(x_{i,t+h} - x_{i,t} - LTG_{i,t} | HLTG_t\right) = \vartheta_h \mathbb{E}\left(\hat{f}_{i,t} - \hat{f}_{i,t}^{\theta} | HLTG_t\right) =$$

$$-\vartheta_h K\theta \mathbb{E}\left(x_{i,t} - bx_{i,t-1} - a\hat{f}_{i,t-1} | HLTG_t\right) = -\vartheta_h K\theta \mathbb{E}\left(\eta_{i,t} + \epsilon_{i,t} | HLTG_t\right).$$

The expectation $\mathbb{E}\left(\eta_{i,t} + \epsilon_{i,t} | HLTG_t\right)$ 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:

$$\mathbb{E}\left(LTG_{i,t+s} | HLTG_t\right) = \mathbb{E}\left(-\varphi_h x_{i,t+s} + \vartheta_h \hat{f}_{i,t+s}^{\theta} | HLTG_t\right) =$$

$$\mathbb{E}\left(-\varphi_h x_{i,t+s} + \vartheta_h \hat{f}_{i,t+s} + \vartheta_h \left(\hat{f}_{i,t+s}^{\theta} - \hat{f}_{i,t+s}\right) | HLTG_t\right) =$$

$$\mathbb{E}\left(-\varphi_h x_{i,t+s} + \vartheta_h \hat{f}_{i,t+s} + \vartheta_h \left(\hat{f}_{i,t+s}^{\theta} - \hat{f}_{i,t+s}\right) | HLTG_t\right) =$$

$$\mathbb{E}\left(-\varphi_h x_{i,t+s} + \vartheta_h \hat{f}_{i,t+s} + \vartheta_h K\theta \left(g_{i,t+s} - a\hat{f}_{i,t+s-1}\right) | HLTG_t\right) =$$

$$-\varphi_h b^s x_{i,t} + \vartheta_h a^s \hat{f}_{i,t},$$

where the last inequality follows from the fact that within the HLTG group of stocks, $g_{i,t+s} - a\hat{f}_{i,t+s-1}$ is one 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, namely for $\theta > 0$. 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:

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

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

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

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

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

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

$$\mathbb{E}_{t}^{\theta}\left(D_{i,t+s}\right)=\mathbb{E}_{t}^{\theta}(e^{x_{i,t+s}})=e^{(1-\varphi_s)x_{i,t}+\vartheta_s\hat{f}_{i,t}^{\theta}+\frac{1}{2}Var_t(x_{t+s})}.$$

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

$$x_{i,t}+LTG_{i,t}>x_{i,t-1}+LTG_{i,t-1}$$

on average in HLTG. We have:

$$\mathbb{E}\left[x_{i,t} + LTG_{i,t} - x_{i,t-1} - LTG_{i,t-1} | LTG_{i,t} = \lambda\right]$$

$$= \frac{cov\left(x_{i,t} + LTG_{i,t} - x_{i,t-1} - LTG_{i,t-1}, LTG_{i,t}\right)}{var\left(LTG_{i,t}\right)} \lambda.$$

The numerator reads

$$cov\left(x_{i,t} + LTG_{i,t} - x_{i,t-1} - LTG_{i,t-1}, LTG_{i,t}\right)$$

$$= cov\left(x_{i,t} - x_{i,t-1}, LTG_{i,t}\right) + cov\left(LTG_{i,t} - LTG_{i,t-1}, LTG_{i,t}\right)$$

The first term is positive under the assumptions of Proposition 2. The second term can be written

$$var\left(LTG_{i,t}\right) - cov\left(LTG_{i,t-1}, LTG_{i,t}\right) = var\left(LTG_{i,t}\right) - \rho_{t,t-k}\sqrt{var\left(LTG_{i,t}\right)}\sqrt{var\left(LTG_{i,t-1}\right)}$$

$$= var\left(LTG_{i,t}\right)\left[1 - \rho_{t,t-k}\right] > 0.$$

Thus, $\mathbb{E}\left[x_{i,t} + LTG_{i,t} - x_{i,t-1} - LTG_{i,t-1} | LTG_{i,t} \geq \lambda_h\right]$ is positive, which proves the result.■

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

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

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

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

where, as usual, we assume that lnR is large enough that the sum converges. Given lognormality, we have that:

$$\mathbb{E}_t^{\theta}\left(P_{i,t+1}+D_{i,t+1}\right)=\sum\nolimits_{s\geq 0} e^{\mathbb{E}_t^{\theta}\left(x_{i,t+1+s}\right)-slnR+\frac{1}{2}Var_t\left(x_{i,t+1+s}\right)},$$

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 $\mathbb{E}_t^{\theta}\left(x_{i,t+1+s}\right)$, but not the perceived variance $Var_t\left(x_{i,t+1+s}\right)$. In particular:

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

This implies that:

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

$$=K\theta\left(g_{i,t}-a\hat{f}_{i,t-1}\right)\sum\nolimits_{s\geq 0} a^{s+1}\frac{1-(b/a)^{s+1}}{1-(b/a)}e^{\mathbb{E}_t^{\theta}\left(x_{i,t+1+s}\right)-slnR+\frac{1}{2}Var_t\left(x_{i,t+1+s}\right)}$$

Under diagnostic expectations, pre-formation news drive mispricing. HLTG stocks experience positive surprises before formation, namely $\left(g_{i,t}-a\hat{f}_{i,t-1}\right)>0$. As a result, the diagnostic expectation $\mathbb{E}_t^{\theta}\left(P_{i,t+1}+D_{i,t+1}\right)$ 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

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

The forecast error in the denominator reads

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

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

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

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

$$-\varphi_h\left(x_{i,t} - x_{i,t-1}\right) + \vartheta_h a(1 - K')\left(\hat{f}_{i,t-1} - \hat{f}_{i,t-2}\right) + \vartheta_h K'\left(x_{i,t} - bx_{i,t-1} - a\hat{f}_{i,t-1}\right)$$
$$- \vartheta_h K'\left(x_{i,t-1} - bx_{i,t-2} - a\hat{f}_{i,t-2}\right)$$

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

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

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

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

This condition is slightly weaker than that of Proposition 2.■

References

Adam, Klaus, and Albert Marcet, and Juan Pablo Nicolini. 2016. "Stock Market Volatility and Learning." *Journal of Finance* 71(1): 33 – 82.

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. 2016. "Diagnostic Expectations and Credit Cycles." *Journal of Finance*, forthcoming.

Bouchaud, Jean-Philippe, Philipp Krueger, Augustin Landier, and David Thesmar. 2016. "Sticky Expectations and the Profitability Anomaly." Working paper.

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. "Overpriced Winners." 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.

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.

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.

Michaely, Roni, and Kent Womack. 1999. "Conflict of Interest and the Credibility of Underwriter Analyst Recommendations." *Review of Financial Studies* 12 (4), 653-686.

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 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.

Appendix.

A. Robustness of LLTG-HLTG Return Differential

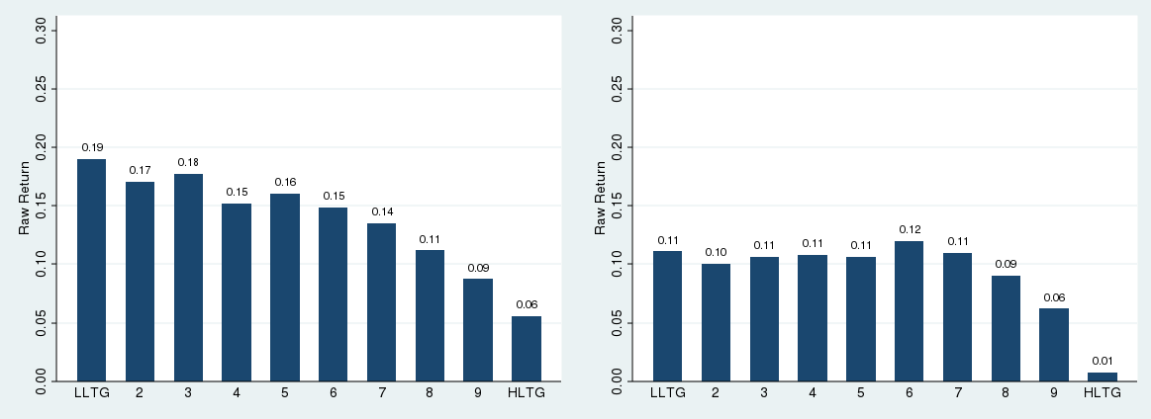


Figure A.1. The return differential between LLTG and HLTG stocks holds in the period 1982-1997 (left panel) and 1998-2014 (right panel).

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

	LTG decile										
	1	2	3	4	5	6	7	8	9	10	1-10
β	0.78	0.93	1.00	1.06	1.11	1.18	1.28	1.37	1.46	1.51	-0.7

Table A.2 – LLTG-HLTG Return differential holds across momentum buckets

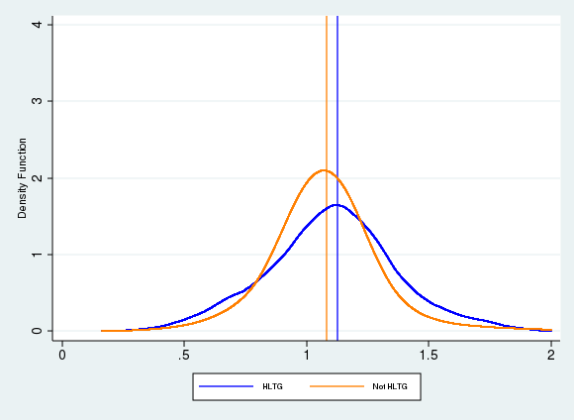
	Rank of Returns in the Previous 6 Months			
LTG	Bottom 30%	Middle	Top 30%	Top-Bottom
1	9.8%	15.5%	17.7%	7.8%
2	9.2%	15.3%	14.0%	5%
3	9.8%	15.0%	15.4%	6%
4	9.7%	14.0%	13.3%	4%
5	8.9%	14.9%	14.7%	6%
6	10.3%	14.4%	14.3%	4%
7	7.1%	12.5%	15.5%	8%
8	5.6%	11.2%	13.0%	7%
9	5.8%	8.0%	8.6%	3%
10	-1.0%	4.6%	8.2%	9%
1-10	10.9%	10.9%	9.4%	-1.4%

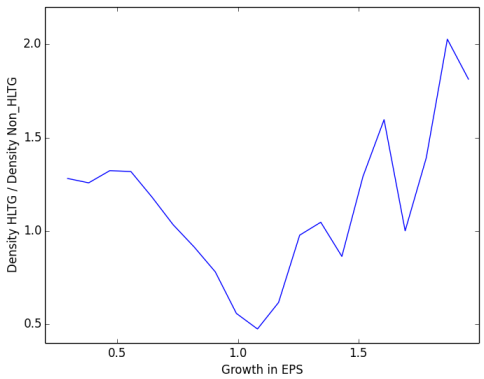
Table A.3 – Factor Regressions

LTG Portfolio									
1	2	3	4	5	6	7	8	9	10
Panel A: 1 Factor Model									
0.7834[a]	0.9251[a]	1.0045[a]	1.0417[a]	1.0627[a]	1.1423[a]	1.1958[a]	1.3131[a]	1.4813[a]	1.6512[a]
(0.0275)	(0.0275)	(0.0284)	(0.0260)	(0.0241)	(0.0288)	(0.0304)	(0.0358)	(0.0466)	(0.0581)
0.4084[a]	0.2282[c]	0.2463[c]	0.1392	0.1427	0.1307	0.0205	-0.1755	-0.4086[c]	-0.7590[a]
(0.1237)	(0.1235)	(0.1276)	(0.1168)	(0.1083)	(0.1292)	(0.1365)	(0.1607)	(0.2092)	(0.2610)
Panel B: 3 Factor Models									
0.8731[a]	0.9965[a]	1.0727[a]	1.0689[a]	1.0621[a]	1.0745[a]	1.1089[a]	1.1717[a]	1.2513[a]	1.3360[a]
(0.0167)	(0.0160)	(0.0184)	(0.0171)	(0.0163)	(0.0178)	(0.0210)	(0.0241)	(0.0302)	(0.0380)
0.3755[a]	0.4611[a]	0.4566[a]	0.5212[a]	0.5312[a]	0.7350[a]	0.7216[a]	0.8226[a]	0.9664[a]	1.0151[a]
(0.0244)	(0.0234)	(0.0270)	(0.0250)	(0.0238)	(0.0260)	(0.0308)	(0.0353)	(0.0443)	(0.0556)
0.6718[a]	0.6391[a]	0.6205[a]	0.4651[a]	0.3379[a]	0.1455[a]	0.0458	-0.1515[a]	-0.4853[a]	-0.8635[a]
(0.0249)	(0.0239)	(0.0276)	(0.0255)	(0.0243)	(0.0266)	(0.0314)	(0.0360)	(0.0452)	(0.0568)
0.1285[c]	-0.0388	-0.0130	-0.0565	-0.0006	0.0652	-0.0038	-0.1192	-0.2158[c]	-0.4105[b]
(0.0711)	(0.0681)	(0.0787)	(0.0729)	(0.0694)	(0.0759)	(0.0897)	(0.1028)	(0.1291)	(0.1621)

B. Kernel of Truth

B.1: EPS growth as a measure of performance





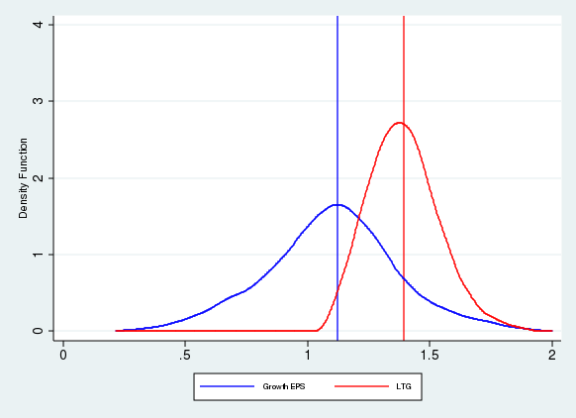


Figure B.1. Long term annualized EPS growth for the HLGT portfolio (top left panel). Representativeness of EPS growth for HLTG (top right panel). EPS growth and LTG forecasts for HLTG (bottom panel).

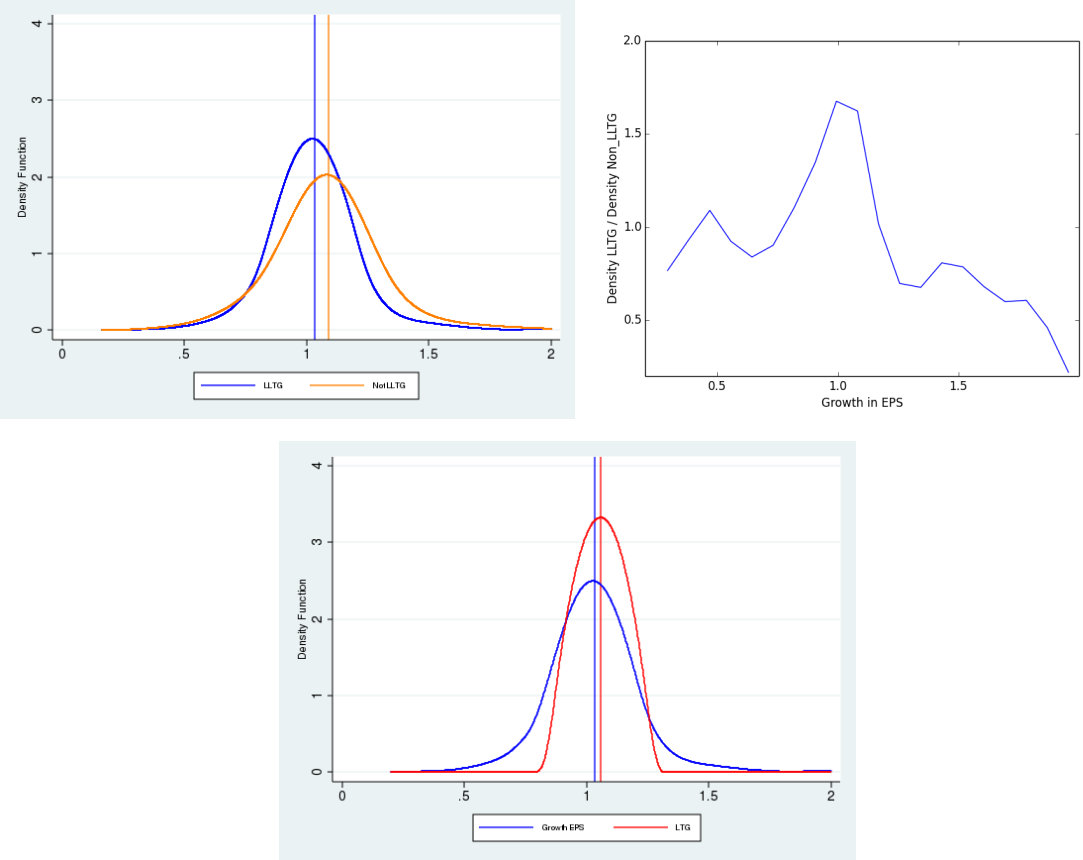


Figure B.1. Long term annualized EPS growth for the LLGT portfolio (top left panel). Representativeness of EPS growth for LLTG (top right panel). EPS growth and LTG forecasts for LLTG (bottom panel).

B.2: Robustness: RMC growth as a measure of performance

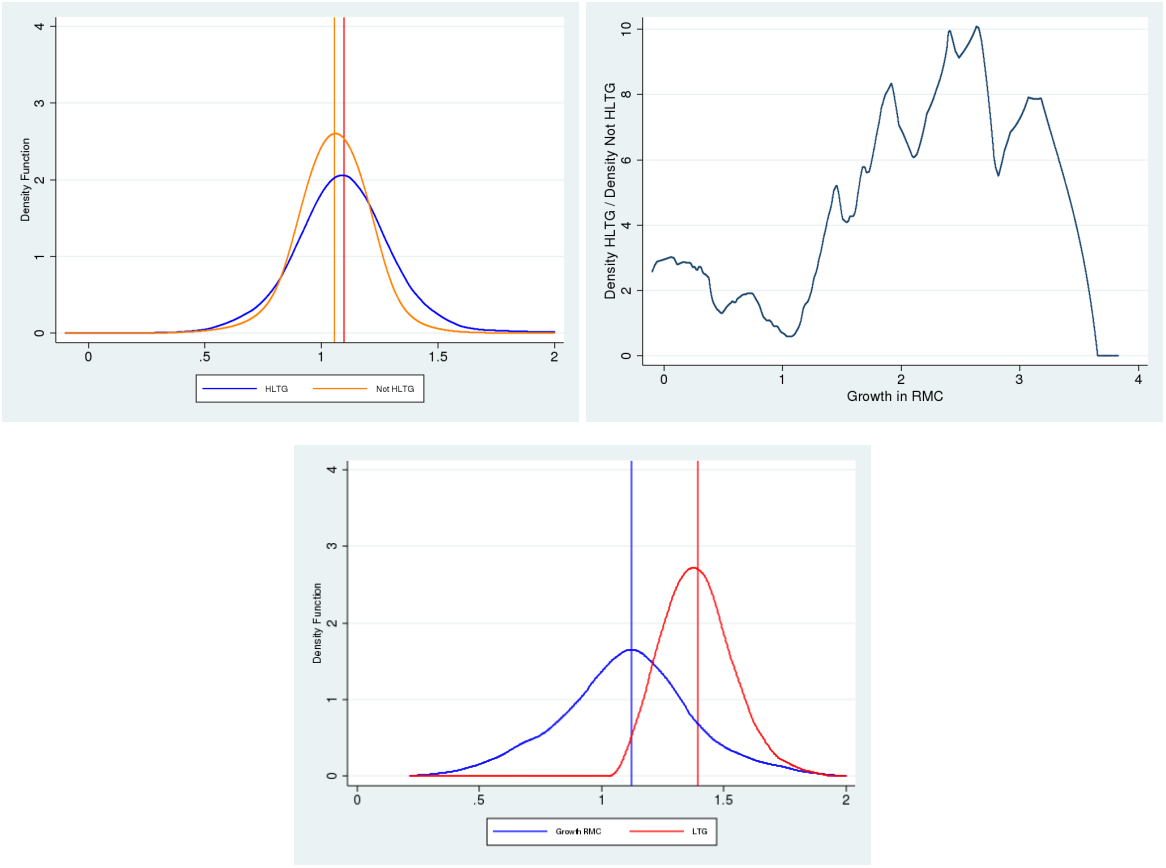


Figure B.2 Long term annualized growth of revenues minus cost of goods sold (RMC) for HLGT portfolio (top left panel). Representativeness of RMC growth for HLTG stocks (top right panel). Growth in RMC and LTG forecasts for HLTG (bottom panel).

C. Coibion and Gorodnichenko Analysis

C.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 C.1: Forecast Errors

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

	Dependent Variable		
	$(eps_{t+3} / eps_t)^{1/3}$-$LTG_t$	$(eps_{t+4} / eps_t)^{1/4}$-$LTG_t$	$(eps_{t+5} / eps_t)^{1/5}$-$LTG_t$
LTG_t-LTG_{t-1}	0.0332	-0.0733	-0.1372[b]
	(0.0725)	(0.0660)	(0.0589)
LTG_t-LTG_{t-2}	-0.0875	-0.1435[b]	-0.1842[a]
	(0.0641)	(0.0691)	(0.0545)
LTG_t-LTG_{t-3}	-0.0956	-0.1184[c]	-0.1701[a]
	(0.0578)	(0.0627)	(0.0517)

C.2 Underreaction for forecasts at short time horizons

Table C.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.

	$(eps_{t+1}-E_t eps_{t+1}) / sps_{t-1}$	$(eps_{t+3}-E_t eps_{t+3}) / sps_{t-1}$	$(eps_{t+5}-E_t eps_{t+2}*(1+LTG_t)^3) / sps_{t-1}$
LTG_t-LTG_{t-1}	0.0839	-0.3226[b]	-0.5918[a]
	(0.0554)	(0.1312)	(0.1435)
LTG_t-LTG_{t-2}	0.1629[a]	0.1262[c]	-0.0227
	(0.0275)	(0.0677)	(0.0772)
LTG_t-LTG_{t-3}	0.0825[a]	-0.0664	-0.2145[b]
	(0.0195)	(0.0532)	(0.0919)