Sticky Expectations and the Profitability Anomaly*

Jean-Philippe Bouchaud[†], Philipp Krüger[‡], Augustin Landier[§], David Thesmar[¶]

November 27, 2017

Abstract

We propose a theory of one of the most economically significant stock market anomalies, i.e. the "profitability" anomaly. In our model, investors forecast future profits using a signal and sticky belief dynamics. In this model, past profits forecast future returns (the profitability anomaly). Using analyst forecast data, we measure expectation stickiness at the firm level and find strong support for three additional predictions of the model: (1) analysts are on average too pessimistic regarding the future profits of high profit firms, (2) the profitability anomaly is stronger for stocks which are followed by stickier analysts, and (3) it is also stronger for stocks with more persistent profits.

*We thank Nick Barberis, Bruno Biais, Eric So as well as the two referees and the Editor for their very useful comments. We are also grateful to seminar audiences at UC Berkeley and NBER. This work was initiated while Landier and Thesmar were doing research at CFM.

[†]CFM

[‡]University of Geneva and Swiss Finance Institute

[§]HEC Paris

[¶]MIT Sloan and CEPR

I Introduction

The existence of stock-return predictability is a central theme in the asset pricing literature: several stock-level characteristics beyond market betas significantly predict future stock-returns. A long-lasting debate pertains to the origin of such abnormal returns and to how they can exist in equilibrium without being arbitraged away. One strand of the literature is focused on interpreting abnormal returns as risk premia (see, for instance, Cochrane (2011))–implying they are only seemingly abnormal–while other authors attribute them to behavioral biases combined with limits to arbitrage (see, e.g., Barberis and Thaler (2003) or references therein, such as Daniel et al. (1998, 2001) or Hirshleifer (2001)). Mispricing then relies on investors making systematic expectation errors, while rational arbitrageurs are unable to fully accommodate their demand because arbitrage is not risk-free. In this literature, the behavioral biases of the non-rational market-participants typically take the form of non-Bayesian expectations grounded in the psychology literature (see, e.g., Hong and Stein (1999) or Barberis et al. (1998)).

The focus of this paper is the "profitability" anomaly: stocks with high profitability ratios tend to outperform on a risk-adjusted basis (Novy-Marx, 2013, 2015). Profitability has recently emerged in the academic literature as one of the stock-return anomalies with the largest economic significance. The corresponding long-short arbitrage strategy features high Sharpe ratios, no crash risk (Lemperiere et al., 2015), and very high capacity due to the high persistence of the profitability signal (e.g., operating cash-flows to asset ratio) on which the strategy sorts stocks (Landier et al., 2015). Our goal in this paper is to test if the profitability anomaly can be directly related to a simple model of sticky expectations, in which investors update their beliefs too slowly.

We start by building a simple model in which risk-neutral investors price a stock, whose dividend is predictable with a persistent signal. These investors have "sticky" expectations. Each period, their expectations are given by λ times their previous belief and $1-\lambda$ times the rational expectation (i.e. the individual-level version of the consensus forecast model of Coibion and Gorodnichenko (2012, 2015)). As shown by Coibion and Gorodnichenko (2015), the model has the advantage of nesting rational expectations as a particular case and delivers a simple way of measuring expectation stickiness using the link between forecast errors and past forecast revisions. It can thus easily be taken to the data. When solving this simple model, we find that future stock returns can be forecasted using past profits and past changes in profits. Thus, the model provides a rationalization for the profitability anomaly. It also makes other predictions.

We test the predictions of the model using observed earnings per share (EPS) forecasts by financial analysts from I/B/E/S. Using directly observable expectations contained in financial analysts' EPS forecasts is a natural setting to study how beliefs of market participants potentially deviate from rational expectations. Analysts are professional forecasters and their forecasts are not cheap talk, which mitigates the legitimate skepticism for subjective answers found in surveys (see Bertrand and Mullainathan (2001)). We do, however, make the assumption that analysts data are representative of investors' expectations. Using these data, we find that the average forecaster puts an excess weight of 16% on earlier annual forecast.

The data are consistent with key cross-sectional predictions of the model. First, we expect that analysts systematically underestimate future profits when current profits are high. Second, the profitability anomaly is expected to be stronger for firms that are subject to stickier EPS forecasts. Third, firms with more persistent earnings should be more prone to the profitability anomaly. Additionally, these three predictions should also hold for two signals alternative to profitability level: earnings momentum (profit change) and returns momentum (past returns). All these predictions are robust outcomes of the model, and we find that they all hold in the data. They thus vindicate our interpretation of this anomaly.

Our analysis is mostly a contribution to the behavioral finance literature, which has documented both patterns of under- and overreaction on analyst forecasts. There is an old tradition of papers on investor underreaction. Abarbanell and Bernard (1992) find evidence that analysts under-react to past earnings, in line with our own results. Ali et al. (1992) find a similar result on annual earnings forecasts. Like us, such positive serial correlation is most often interpreted in the literature as a sign that analysts are under-reacting in a non-Bayesian manner when setting expectations of future earnings (see e.g. Ali et al. (1992) or Markov and Tamayo (2006) for a summary of the literature). An exception is Markov and Tamayo (2006), who argue that the positive autocorrelation of forecasts errors is compatible with Bayesian updating if analysts do not know the true generating process for earnings and slowly learn about the data generating process. Consistent with this hypothesis, Mikhail et al. (2003) find that analysts with more experience under-react less to prior earnings. To our knowledge, this literature does not establish a link between the persistence of forecast errors and the profitability anomaly. Also, our analyst-level regression are harder to reconcile with Bayesian learning. In addition, using the insight of Coibion and Gorodnichenko (2015), we propose a model of expectation formation where underreaction is captured by a single parameter, which we estimate. Finally, we add to the literature by documenting heterogeneity in analyst's biases at the firm level and by relating this heterogeneity to the intensity of stock-market anomalies. In this sense, our results are consistent with finance papers that have documented the slow diffusion of information in markets (see, e.g., Hong et al. (2000); Hou (2007)).

This underreaction tradition coexists with abundant evidence of overreaction. For instance, Debondt and Thaler (1990) document patterns of overreaction by looking at analyst revisions. Most related to our present work, there is an ecology of papers which seeks to explain the value premium with extrapolating beliefs starting with Debondt and Thaler (1985) and Lakonishok et al. (1994). Laporta (1996) and Bordalo et al. (2017) show that stocks with high expected growth (as measured by analyst consensus on long-term earnings growth) tend to (1) be glamour stocks and (2) have low expected returns. Alti and Tetlock (2014) calibrates a model where over-reaction and overconfidence distort agents' expectations of firm productivity. Weber (2016) documents abnormal returns of portfolios sorted on cash-flow duration and shows that this anomaly can be explained by extrapolation bias in analysts' long-term forecasts. Gennaioli et al. (2015) and Greenwood and Shleifer (2014) find that errors in CFO expectations of earnings growth are not rational and are compatible with a model of extrapolative expectations. They focus on time series of forecasts, and on expectations of long-term growth and returns. These papers differ from ours in two respects: First, they seek to explain a different anomaly (they focus on the value premium or the duration premium while we offer a theory of the profitability anomaly). Second, they find evidence of extrapolative behavior regarding long-term earnings growth forecasts, while we provide evidence of stickiness of near-term EPS forecasts. Consistent with this, Bordalo et al. (2017) run regressions similar to our Table III on both EPS forecasts (our focus here) and long-term growth forecasts (their focus), and confirm both our finding of stickiness in the short-run and their hypothesis of overreaction of long-run expectations.

Our results from Table VI also speak to a small number of papers who link analyst forecast errors with well-known signals that predict returns. Brav et al. (2005) find that systematic expectation errors are consistent with a large number of signals used to forecast returns, but do not attempt to put economic structure on expectation dynamics. Also, Engelberg et al. (2016) document that predictable returns in various anomalies are concentrated around earnings announcements and days on which significant news is revealed. Such a prediction would be consistent with our set-up, but we do not explore this avenur in our paper.

In terms of theoretical asset-pricing models, an important strand of the behavioral literature has focused on explaining the value, momentum, and post-earnings announcement drift anomalies. Most related to our work are papers which propose non-Bayesian theories of beliefs dynamics that can explain these anomalies. Barberis et al. (1998) propose a model where investors try to estimate whether prices are in a trending regime or a mean-reverting regime. This generates simultaneous short-term underreaction of stock prices to news and overreaction to a series of good or bad news. Hong and Stein (1999) develop a model where two types of traders co-exist: traders who trade on news and trend-followers. The interaction between these traders generates an equilibrium that exhibits both short-term momentum and long-term reversal. Because our paper focuses on the profitability anomaly, we use a simple non-Bayesian set-up with only one type of risk neutral agent. We directly measure analyst beliefs stickiness and test the comparative statics of the model which are highly constraining on the data: we show that the profitability anomaly is stronger for stocks where the measured stickiness of analyst forecasts is higher. This is an indirect validation of the assumption that biases in analyst forecasts about future profitability can be seen as being representative of beliefs of investors.

In its methodology, our paper is also related to the recent macro literature on expectation formation. The model of expectations dynamics that we use is analyzed in Coibion and Gorodnichenko (2012), which was originally applied to professional inflation forecasts. In Mankiw and Reis (2001), agents also update beliefs infrequently due to fixed costs, which in turn leads to sticky prices.

The rest of the paper is organized as follows: The next section lays out the model of Coibion and Gorodnichenko (2012) and adapts it to the context of firm-level characteristics with predictive power on future profits. We derive structural predictions that link the persistence and predictive power of these firm-level characteristics, the level of beliefs stickiness from analysts, and the dynamics of their forecast errors. Section **III** describes the data. Section **IV** gathers our empirical results: First, we document the predictability of returns, earnings, and forecast errors by several firm-level characteristics observable at the time of forecast formation. Secondly, we test structural predictions of the model. Section **V** uses Monte Carlo simulations to examine the robustness of our results and, finally, Section **VI** concludes.

II Model

A. Expectation stickiness

We start by analyzing a model with expectation dynamics which can be directly tested without further assumption on the data-generating process of the forecasted variable. We take our model of expectation dynamics from the macro literature on information rigidity (see Mankiw and Reis (2002) or Reis (2006)). We use notations from Coibion and Gorodnichenko (2012) and Coibion and Gorodnichenko (2015). Let $F_t\pi_{t+h}$ be the expectation formed at t about profits at $t+h$, which we denote as π_{t+h}. We assume that expectations are updated according to the following process:

$$F_t\pi_{t+h} = (1-\lambda)E_t\pi_{t+h} + \lambda F_{t-1}\pi_{t+h} \tag{1}$$

which is easy to interpret. $E_t\pi_{t+h}$ stands for rational expectation of π_{t+h} conditional on information available at date t. The coefficient λ indicates the extent of expectation "stickiness." When $\lambda = 0$, expectations are perfectly rational. When $\lambda > 0$, the forecaster insufficiently incorporates new information into her forecasts. This framework accommodates patterns of both under-reaction ($0 < \lambda < 1$) and overreaction ($\lambda < 0$) (shown for instance in Greenwood and Shleifer (2014) and Gennaioli et al. (2015)). When applied to consensus forecasts, this structure on forecasts can be made consistent with models of Bayesian learning with private information (Coibion and Gorodnichenko, 2015); When applied to individual level forecasts, however, it can only come from non-bayesian under-reaction (we show later that the data favor this type of explanation).

As noted by Coibion and Gorodnichenko (2012) and Coibion and Gorodnichenko (2015), this structure gives rise to straightforward testable predictions that are independent of the process underlying profits π_t and provide a direct measure of λ:

Prediction 1. ***Inferring stickiness from forecast dynamics (Coibion and Gorodnichenko, 2015)***

Assuming expectations are sticky in the sense of equation (1), then the following two closely linked relationships should hold:

1. *Forecast errors should be predicted by past revisions:*

$$E_t\left(\pi_{t+1} - F_t\pi_{t+1}\right) = \frac{\lambda}{1-\lambda}\left(F_t\pi_{t+1} - F_{t-1}\pi_{t+1}\right) \tag{2}$$

2. *Revisions are autocorrelated over time:*

$$E_{t-1}\left(F_t\pi_{t+1} - F_{t-1}\pi_{t+1}\right) = \lambda\left(F_{t-1}\pi_{t+1} - F_{t-2}\pi_{t+1}\right) \tag{3}$$

Proof. See Appendix A □

These two relations can readily be tested on expectations data without further assumption about the data-generating process of π_t. The intuition behind the first one is that forecast revisions contain some element of new information, only partially incorporated into expectations. As a result, revisions predict forecast errors. Quite elegantly, the regression coefficient is a simple transformation of the stickiness parameter λ. The second prediction pertains to the dynamics of forecast revisions. When expectations are sticky, information is slowly incorporated in forecasts, so that a positive news generates positive forecast revisions over several periods. This generates momentum in forecasts.

B. Earnings expectations

We now further assume that firm profits π_{t+1} can be predicted with a signal s_t, that is

$$\pi_{t+1} = s_t + \epsilon_{t+1}, \tag{4}$$

where ϵ_{t+1} is a noise term.

The signal is persistent, so that

$$s_{t+1} = \rho s_t + u_{t+1}, \tag{5}$$

where $\rho < 1$ and u_{t+1} is a noise term. One can think of s_t as a sufficient statistic capturing all public information useful to predict future profits. A particular case is to consider that s_t is simply equal to lagged profits or lagged cash-flows, but this is just a particular case. To obtain closed form solutions for conditional expectations, we also assume that ϵ_{t+1} and u_{t+1} follow a normal distribution, but the intuitions we derive in the paper do not hinge on this particular assumption. Note that, taken together, assumptions (4), (5) and normality impose that profits follow an ARMA(1,1) process.

The expectation definition (1) can be rewritten as:

$$F_t\pi_{t+1} = (1-\lambda)\sum_{k\geq 0}\lambda^k E_{t-k}\pi_{t+1}$$

Given our assumptions about the profit process and the signal informativeness, we know that $E_{t-k}\pi_{t+1} = \rho^k s_{t-k}$, so that forecasts should write:

$$F_t\pi_{t+1} = (1-\lambda)\sum_{k\geq 0}(\lambda\rho)^k s_{t-k} \tag{6}$$

The econometrician does not observe the signal s_t, but observes profits π_t. Thus, in order to implement our tests, we need to formulate a prediction about forecasts *conditional* on π_t. We do this in the following proposition, by showing that past profits predict future forecast errors:

Prediction 2. ***Past profits predict future forecast errors***

Assuming expectations are sticky in the sense of equation (1), and profits can be forecast using an autoregressive signal s_t, then earnings surprises should follow:

$$E_t\left(\pi_{t+1} - F_t\pi_{t+1}\middle|\,\pi_t\right) = \frac{\rho\lambda^2(1-\rho^2)}{1-\lambda\rho^2}\frac{\sigma_u^2}{\sigma_u^2+(1-\rho^2)\sigma_\epsilon^2}\pi_t$$

Proof. See Appendix B ∎

This equation is straightforward to interpret. If expectations are rational ($\lambda = 0$), the earnings surprise should be uncorrelated with past realizations of profits. In fact, it should be zero by definition of rationality. As soon as $\lambda > 0$, profits will positively predict future surprises, but only to the extent that the signal is persistent ($\rho > 0$). This happens because past profits need to be persistent to be indicative of future profits. Since investors are slow at adjusting their beliefs, they underestimate this persistence which leads to predictable forecast errors. The prefactor $\frac{\sigma_u^2}{\sigma_u^2+(1-\rho^2)\sigma_\epsilon^2}$ can be interpreted in a classic Bayesian manner as follows: When σ_ϵ is large, a high π_t is less likely to imply a high signal level and thus a large mistake. Conversely, when σ_u is large (fast moving signal), a high π_t is more likely to imply a high signal level that got high only recently, and thus implies a large mistake as expectations are still anchored in the past.

C. *Forecasting stock returns*

We now move from profits to returns. To simplify exposition, we set up a bare-bone asset pricing model: We assume that all investors are risk neutral and have the same expectation stickiness parameter λ. This is an extreme assumption designed to focus on our key effects. A natural extension would be a limits of arbitrage model where rational, risk averse, arbitrageurs trade against the sticky investors. Our qualitative predictions would carry out in such a set-up, although they would be partially attenuated by the presence of limited arbitrage.

Given our risk neutral pricing assumption, the stock price, just after receiving dividend π_t and observing signal s_t, is simply given by:

$$P_t = \sum_{k\geq 1} \frac{F_t \pi_{t+k}}{(1+r)^k} \tag{7}$$

Given that we know the process of profits and expectations updating, we can easily derive the prices and returns, defined as $R_{t+1} = (P_{t+1} + \pi_{t+1}) - (1+r)P_t$, as a function of past signals. This leads to the following, intermediate, result:

Lemma 1. *When agents are risk-neutral and expectation are sticky in the sense of Equation (1), prices and returns are functions of past signals:*

$$P_t = m \sum_{k\geq 0} (\lambda\rho)^k s_{t-k}$$
$$R_{t+1} = m u_{t+1} + \epsilon_{t+1} + \lambda(1+m\rho)s_t - (1-\lambda)(1+m\rho)\sum_{k\geq 1}(\lambda\rho)^k s_{t-k}$$

where $m = \frac{1-\lambda}{1+r-\rho}$.

To interpret the first formula, let us note $P_t^\star = \frac{1}{1+r-\rho}s_t$, which is the price that prevails when $\lambda = 0$, i.e. the rational price. Using this definition, we can rewrite price dynamics as

$$P_t = (1-\lambda)P_t^\star + \lambda\rho P_{t-1}.$$

Prices are equal to $1-\lambda$ times the rational price, and there is excess persistence of past prices, especially when ρ is large. The second equation directly comes from the definition of returns. This equation confirms that past signals predict returns, as long as $\lambda \neq 0$. If expectations are rational ($\lambda = 0$), then returns are given by $\frac{1}{1+r-\rho}u_{t+1} + \epsilon_{t+1}$ and have zero conditional mean: High returns in this case may arise from temporary profit shocks ϵ_{t+1}, as well as innovation on the signal u_{t+1}, which is multiplied by $\frac{1}{1+r-\rho}$ since the signal is persistent.

As with profit expectations, the econometrician does not observe the signal realization, so she cannot directly test the relationships in Lemma 1, but she observes past profits and past returns. Our third prediction is that future returns can be forecast using information available to the econometrician. In the following proposition, we describe these anomalies in terms of covariance of future returns with past predictive variables: in the rational case, this covariance should be null.

Prediction 3. ***Belief stickiness and stock-market anomalies***

When agents are risk-neutral and expectation are sticky in the sense of Equation (1), then, at the steady state, noting $m = \frac{1-\lambda}{1+r-\rho}$:

1. *Past profits predict future returns ("profitability"):*

$$cov(R_{t+1}, \pi_t) = (1+m\rho)\frac{\rho}{1-\lambda\rho^2}\lambda^2\sigma_u^2$$

2. *Increases in past profits predict future returns ("earnings momentum"):*

$$cov(R_{t+1}, \Delta\pi_t) = (1+m\rho)\frac{\rho}{1+\lambda\rho}\lambda^2\sigma_u^2$$

3. *Past returns predict future returns ("price momentum"):*

$$cov(R_{t+1}, R_t) = (1+m\rho)(m+\rho\lambda)\frac{\lambda\sigma_u^2}{1-\lambda^2\rho^2}$$

4. *All covariances $cov(R_{t+1}, \pi_t)$, $cov(R_{t+1}, \Delta\pi_t)$ and $cov(R_{t+1}, R_t)$ increase with ρ. They also increase with λ under the "near rational" approximation that $\lambda \ll 1$.*

Proof. See Appendix C □

That items 1–3 of Prediction 3 hold in the data has been shown in the large empirical literature on asset pricing. Novy-Marx (2013) shows the sharpe ratio of the profitability anomaly is high, while Landier et al. (2015) document that it is indeed a large anomaly, in the sense that large amounts can be invested in it without being erased by transaction costs. Novy-Marx (2015) documents that changes in earnings also forecast returns. That past returns forecast future returns in equity markets is well known since at least Jegadeesh and Titman (1993). The three formulas 1., 2. and 3. are consistent with the results derived on the formation of profit expectations. This happens because past profit, profit change or past returns contain information about future profits that has not been fully impounded into current prices. We notice two interesting properties. First, if expectations are rational ($\lambda = 0$), neither past profits (levels or changes) nor past returns can forecast future returns. Second, sticky expectations have the power of explaining the profitability anomaly *if and only if* the signal is persistent. This ties again to the intuition that slow updating is not a big source of mispricing when recent news are not informative about the future. It makes returns more volatile (bigger mistakes are made every period), but does not generate persistence.

In this paper, we go a step further than the existing literature on the profitability anomaly, and test the comparative statics suggested by the model on the cross-section of stock returns. First, when λ is small, the proposition shows that a higher value of λ reinforces the anomaly: quite intuitively, stickier beliefs reinforce the relationship between past profits, change in profits or returns, and future stock returns. Secondly, the proposition also shows that signal persistence (higher ρ) increases the strength of these anomalies. It comes from the abovementioned fact that higher persistence makes slow expectations a larger source of mistakes about the future. This is because current signal about future profit has a bigger impact on actual value when persistence is higher: The scope for underreaction is therefore higher.

III Data

A. Data construction

A.1. Analyst forecasts

To construct our sample of analyst expectations, we obtain analyst-by-analyst EPS forecasts from the I/B/E/S Detail History file (unadjusted). We retain all forecasts that were issued 45 days *after* an announcement of total fiscal year earnings. We focus on analyst EPS forecasts for the current fiscal year as well as forecasts for one and two fiscal years ahead.[1] If an analyst issues multiple forecasts for the same firm and the same fiscal year during this 45 day period, we retain only the first forecast.

Using these detailed analyst-by-analyst forecasts, we calculate the firm-level consensus EPS forecast ourselves. In other words, we do not use the consensus forecast from the I/B/E/S Summary History file, simply because it is not known how I/B/E/S decides on whether or not to include an individual analyst-level forecast in the calculation of the consensus. The I/B/E/S consensus could thus contain stale information, which we would like to avoid. To compute the one, two, and three year ahead forecasts for earnings of fiscal year t, that is $F_{t-h}\pi_t$ (with $h = 1, 2, 3$), we calculate the median of all forecasts submitted at most 45 days after the announcement of earnings for fiscal year $t - h$. We choose 45 days because this is the median time (across analysts) between the announcement of annual earnings and the issuance of their first forecast in the I/B/E/S Detail History file. Taking a relatively short period (45 days) also maximizes the scope for forecast errors and biases. At the same time, it ensures that as little material information for year t as possible has been released. In order to avoid staleness, we focus on forecasts that are actively submitted by analysts. A possible concern is that analysts "resubmit" old forecasts without changing the numbers. This does not happen very often (less than 2% of the cases). So our consensus is mainly based on "fresh forecasts" that are not artificially stale.

Next, we match actual reported EPS from the I/B/E/S unadjusted actuals file with the calculated consensus forecasts. As pointed out in prior research (see Diether et al. (2002); Robinson and Glushkov (2006)), problems can arise when actual earnings from the I/B/E/S unadjusted actuals file are matched with forecasts from the I/B/E/S unadjusted detail history file. These problems are due to stock splits occurring between the EPS forecast and the actual earnings announcement: if a split occurs between an analyst's forecast and the associated earnings announcement, the forecast and the actual EPS value may be based on a different number of shares outstanding. To deal with this issue, we use the CRSP cumulative adjustment factors to put the forecasts from the unadjusted detail history and the actual EPS from the unadjusted actuals on the same share basis. We retain all firm-level observations with fiscal years ending between 1989 and 2015. In Table I we report summary statistics for the main variables of the EPS forecast sample.

[Insert Table I about here.]

This dataset is an annual panel of firms. It has about 54k observations for most variables, and some 16k when we require the presence of 3 year ahead forecasts (which

[1] We identify forecasts for the different fiscal years by the means of the I/B/E/S Forecast Period Indicator variable *FPI*.

we use in one specification). We use it to investigate the determinants of forecast errors (predictions 1 and 2). We now turn to the construction of the panel of monthly stock returns, which we use to test our last set of predictions (prediction 3).

A.2. Stock Returns

To construct our panel of stock returns, we start with all firms in the monthly CRSP database between 1990 and 2015 having share codes 10 and 11. We keep only firms listed on NYSE, Amex, or Nasdaq[2] that can be matched with Compustat. We then match these data with our previously described dataset on analyst forecasts.[3]

For our portfolio analysis, we compute signals for profitability, profitability momentum, and price momentum in our sample:

1. **Cash-flows** *(cf)* is the net cash-flows from the firm's operating activities normalized by total assets. It is calculated as the ratio of Compustat items *oancf* and *at*. Cash-flows have been shown to be a very strong predictor of returns (see Asness et al. (2014), Landier et al. (2015)). One possible explanation is that cash-flows are a better measure of a firm's fundamental value, consistent with the idea that the difference between cash-flows and earnings predicts returns (Sloan, 1996).

2. Δ **Cash-flows** (Δcf) denotes the difference between the last available annual cash-flow to asset ratio (cf_t), and the value of this ratio in the previous fiscal year (cf_{t-1}). Such signals are sometimes referred to as "earnings momentum" (Novy-Marx, 2015).

3. **Momentum** (*mom*) is the cumulative firm-level return between months t-12 and t-2 as in Jegadeesh and Titman (1993).

We assume accounting data to be available *after* recorded earnings announcement, which we obtain from Compustat quarterly. Accounting profitability signals are updated in the month following a firm's fiscal year earnings announcement and remain valid until the month of the firm's next fiscal year earnings announcement. We thus require that two consecutive annual earnings announcements are available.

We check that the three anomalies are indeed present in our sample in Table II. For each of the three signals, we sort stocks each month into quintiles of the signal. At the point of portfolio formation, we restrict ourselves to the 3,000 largest stocks. As is standard in the literature, we measure size as stock market capitalization in last June and ranks are calculated in each month. We also exclude penny stocks by requiring, at

[2]Exchange codes 1,2 and 3

[3]We match I/B/E/S with CRSP/Compustat using CUSIP and keep only matches for which both the CUSIP and the *CUSIP dates* match in both CRSP/Compustat and I/B/E/S.

portfolio formation, that the previous month closing price exceeds \$5. We then compute equal weighted portfolio returns for each of the five quintile portfolios, as well as the long-short Q5-Q1 portfolio. In Panel A, we show excess returns without risk adjustment. We then regress portfolio returns on standard sets of risk factors. We use the CAPM (Panel B), the Fama and French (1993) three factor model (Panel C), and the Carhart (1997) four factor model, which includes a momentum factor (Panel D). Given that the factor model in Panel D includes a momentum risk-factor, we are not testing the returns of the momentum strategy in Panel D.

[Insert Table II about here.]

As shown in previous literature, the three signals indeed forecast returns, and predictability is robust to risk adjustment. In Panel D, the t-statistic for the long-short portfolio sorted on cash-flows is equal to 3.56^{***}. For Δcf, the significance is a bit weaker: 2.87^{***} (it is bigger than 3 for less conservative adjustments). In Panel C, long-short portfolio on momentum has a t-statistic of 3.68^{***}.

IV Earnings forecasts and sticky beliefs: testing the model

In this section, we now test the predictions derived from the model of sticky beliefs presented in Section **II**.

A. Prediction 1: measuring stickiness

A.1. Pooled analysis

We start by estimating equation (2), which links forecast errors with past forecast revisions. As shown by Coibion and Gorodnichenko (2015) – and recalled in Prediction 1 – this regression allows to directly recover the stickiness parameter λ without further assumption about the data-generating process of profits.

To implement this test, we calculate the forecast revision, which we define as the change in the consensus forecast of earnings for fiscal year t that was formed just after the announcement of fiscal year earnings $t-1$ (i.e., $F_{t-1}\pi_{f,t}$) with respect to the consensus earnings forecast for fiscal year earnings t that was formed just after the announcement of fiscal year earnings $t-2$ (i.e., $F_{t-2}\pi_{f,t}$). We normalize this revision of expectations by the stock price before the announcement of fiscal year earnings $t-2$, which we denote $P_{f,t-2}$. The forecast revision for firm $f's$ earnings in fiscal year t is thus defined as $(F_{t-1}\pi_{f,t} - F_{t-2}\pi_{f,t})/P_{f,t-2}$. Accordingly, we define the forecast error as the difference between total fiscal year earnings reported for fiscal year t and the consensus forecast for total fiscal year earnings that was formed just after the announcement of fiscal year earnings $t-1$, which we again normalize by $P_{f,t-2}$.The forecast error is thus $(\pi_{f,t} - F_{t-1}\pi_{f,t})/P_{f,t-2}$.

[Insert Figure 1 about here.]

Before running regressions, we first offer a graphical visualization of the data. In Figure 1, we show the forecast error as a function of forecast revisions. We sort all observations into twenty ordered bins of the forecast revision $(F_{t-1}\pi_{f,t} - F_{t-2}\pi_{f,t})/P_{f,t-2}$ and compute both average forecast error $(\pi_{f,t} - F_{t-1}\pi_{f,t})/P_{f,t-2}$ and average forecast revision for each of the twenty ordered bins. The figure shows a strong monotonic relationship between the forecast error and the revision. We then move to the statistical analysis, and estimate the following regression where the time unit t is the fiscal year:

$$\frac{\pi_{f,t} - F_{t-1}\pi_{f,t}}{P_{f,t-2}} = a + b \cdot \frac{F_{t-1}\pi_{f,t} - F_{t-2}\pi_{f,t}}{P_{f,t-2}} + c \cdot \frac{\pi_{f,t-1} - \pi_{f,t-2}}{P_{f,t-2}} + \epsilon_{f,t} \tag{8}$$

Our main specification has $c = 0$. As recalled in Proposition 1, the coefficient b can then be interpreted as a function of the stickiness parameter, so that $\lambda = b/(1+b)$. Error terms $\epsilon_{f,t}$ are allowed to be flexibly correlated within firm and within year. The negative coefficient $c < 0$ captures the presence of extrapolative bias. When profits go up, extrapolators are on average optimistic, i.e. their forecast error $\pi_{f,t} - F_{t-1}\pi_{f,t}$ should be negative.

[Insert Table III about here.]

We report regression results in Table III. In column (1) of Panel A, we directly estimate equation (8) setting $c = 0$. We find $b = 0.165$, which means $\lambda = 0.14$. This suggests that, at the quarterly frequency, the weight of lagged forecasts is given by $0.14^{\frac{1}{4}} = 0.6$, very similar to what Coibion and Gorodnichenko (2015) find for quarterly revisions of inflation forecasts (they find $\lambda \approx .55$). Hence, our estimation of stickiness is in the ballpark of recent estimates coming from macro forecasts made by independent forecasters instead of security analysts. In column (2), we include the two components of the revision separately, and find that their absolute values do not differ very much, which is a reassuring property. In column (3), we add the extrapolation parameter. The idea here is to (1) check that our estimate of λ is robust to controlling for extrapolation and (2) verify the presence of extrapolation in our data. We find that extrapolation is there ($c < 0$) but insignificant. As a result, controlling for extrapolation marginally increases the stickiness coefficient, but not significantly so.

In Panel B of Table III we use another strategy to estimate λ, which is based on the dynamics of forecasts revisions (equation (3) in Prediction 1). The idea of this second approach is that the change in forecasts at time t contains an "echo" of the previous change in forecasts. The strength of that "echo" provides a measure of λ. More formally, we estimate:

$$\frac{F_{t-1}\pi_{f,t} - F_{t-2}\pi_{f,t}}{P_{f,t-3}} = a + b \cdot \frac{F_{t-2}\pi_{f,t} - F_{t-3}\pi_{f,t}}{P_{f,t-3}} + \epsilon_{f,t}, \tag{9}$$

where b is in theory – i.e. if the expectation model (1) is true – equal to λ.

When testing this prediction, we have to rely on analysts' EPS forecasts for three fiscal years ahead, which makes our sample size drop substantially: We keep only about a third of the observations compared to Panel A where only two year ahead EPS forecasts are needed. In the data, the number of available analyst forecasts drops sharply with the forecast horizon. Despite this constraint, we find an estimate of λ equal to 0.06 (see Column (1), Panel B, Table III). This estimate is noisier but not significantly different from the one shown in Panel A. The similar magnitude of the two coefficients is reassuring because the two estimation strategies are quite different in nature. They provide two separate confirmations that our expectation model (1) holds. The estimation strategy in Panel B relies on the stickiness of expectations to be independent of the time distance to realization, which the strategy in Panel A, does not require. The second estimation procedure is, however, more fragile than the first one due to the smaller sample size imposed by the use of longer-term forecasts.

A.2. Stickiness at the analyst- and firm-level

In this section, we extend the methodology used in the previous subsection in order to estimate analyst- and firm-level stickiness parameters λ_a and λ_f. We then test whether certain analyst– and/or firm–level characteristics are correlated with higher levels of stickiness. For instance, if we interpret stickiness as resulting from time-constraints, we would expect analysts who follow more industries to exhibit stickier expectations as they are more constrained in the time they can allocate to revising forecasts. In a similar vein, more experienced analysts might be more inclined to process material information more quickly, leading to less sticky expectations.

To test predictions of this kind, we proceed in two steps. First, we separately estimate the stickiness parameter for each analyst a (resp. for each firm f). In doing so, we use *all* available observations at the analyst and firm-level. In a second step, we relate the cross section of analyst- (respectively firm-) level stickiness to observable analyst (respectively firm) characteristics.

Using the whole time-series of EPS forecasts for a given analyst a, we individually estimate the following regression for each analyst a

$$\frac{\pi_{f,t} - F_{a,t-1}\pi_{f,t}}{P_{f,t-2}} = a_a + b_a \cdot \frac{F_{a,t-1}\pi_{f,t} - F_{a,t-2}\pi_{f,t}}{P_{f,t-2}} + \epsilon_{a,f,t}. \tag{10}$$

Using the relation $\lambda_a = b_a/(1 + b_a)$ implied by the model, we can then back out the analyst level stickiness using the regression coefficient b_a from the above equation. Panel A of Table IV shows summary statistics for the parameter λ_a.

It is important to note that Equation (10) represents a significant departure from Coibion and Gorodnichenko (2015). In their paper, the link between forecast errors and revisions fleshed out in Equation (8) is only valid at the consensus level. At the forecaster level, forecast errors are unpredictable. This is because, in their paper, they consider only two models of expectation formation at the individual level which are close to rationality. Equation (10) assumes that, at the individual level, expectations are non-Bayesian. Hence forecast errors can be predicted with revisions at the individual level. This equation is not grounded in a psychological model of expectation formation (as for instance in Bordalo et al. (2017)): We think of it as an empirical equation designed to measure individual-level stickiness. We note, however, that if Coibion and Gorodnichenko (2015)'s interpretation of consensus data is correct, we should find that $b_a = 0$.

[Insert Table IV about here.]

In total we are able to estimate the analyst level stickiness for 6,938 analysts. The mean analyst–level stickiness is about 0.16, similar to what we obtained from the pooled estimation in Panel A, Table III. The mean analyst–level stickiness λ_a is estimated using about 23 observations (Mean $N_{\lambda_a} = 22.96$). Note also that more than 25 percent of analysts have a negative λ_a, i.e., they "overreact" to recent information. This finding is consistent with the results of Coibion and Gorodnichenko (2015) at the consensus level, but not consistent with their interpretation, because in the two expectation formation models they consider, the expectation errors at the individual forecaster-level cannot be predicted by past revisions. Our result suggests that the stickiness in consensus forecasts directly stems from under-reaction at the individual level (rather than. for instance, Bayesian updating with informational frictions).

We now repeat the same procedure at the firm–level, which amounts to estimating the stickiness parameter of the median analyst covering a firm (i.e., using the firm-level time series of consensus forecast errors and revisions). Again we use *all* observations that are available for a given firm to estimate the firm-level lambda. More specifically, we estimate

$$\frac{\pi_{f,t} - F_{t-1}\pi_{f,t}}{P_{f,t-2}} = a_f + b_f \cdot \frac{F_{t-1}\pi_{f,t} - F_{t-2}\pi_{f,t}}{P_{f,t-2}} + \epsilon_{f,t}, \tag{11}$$

and obtain the firm–level stickiness using the transformation $\lambda_f = b_f/(1+b_f)$. The mean firm–level stickiness λ_f is 0.13 and it is estimated using nine years of data. Again, the stickiness parameter estimated at the firm–level is quite similar to what was obtained in the pooled estimation. Similar to the distribution of λ_a, Panel B of Table IV shows that only a minority of firms displays evidence of overreaction: About 25% of the firms have a negative λ_f, though most of them are non-significant.

Next, we regress our estimated parameters λ_a (resp. λ_f) on analysts' (resp. firms') characteristics. Since we only have one observation per analyst, we use time-series averages of analyst (firm) characteristics during the sample period as explanatory variables. We estimate cross-sectional equations of the following type

$$\lambda_a = a + b \cdot x_a + \epsilon_a, \tag{12}$$

where x_a is, for instance, the average number of years an analyst has been forecasting earnings during the sample-period. We estimate similar kinds of regressions at the firm–level, that is

$$\lambda_f = a + b \cdot x_f + \epsilon_f, \tag{13}$$

where x_f denotes, for instance, the average firm size or average EPS volatility of the firm throughout the sample period. The results for both types of regressions are reported in Table V.

[Insert Table V about here.]

In Panel A, we report results on the determinants of analyst–level stickiness. We find that analysts covering a larger number of industries have stickier expectations, in line with a bounded rationality interpretation of the sticky forecasts model (see column (4), Panel A). Stickiness tends to decrease with the analyst's years of experience (columns (1)–(3)), but the result is insignificant once controlling for the number of firms and industries covered by the analyst.

In Panel B, we show the results from the firm–level regressions and find that stickiness is higher for firms with more volatile EPS, which can be interpreted as analysts "giving-up" on trying to make accurate forecasts for such firms. This is loosely consistent with a learning model where analysts invest in noisy signals of EPS. If EPS is fundamentally noisy, signals are less informative and analysts update their forecasts less frequently.

B. Prediction 2: Past profits predict forecast errors

Prediction 2 of the model suggests that if expectations are sticky, past profits should predict forecast errors, i.e. that forecasts of profitable firms should be, on average, pessimistic. This comes from the fact that, when analysts are sticky, not all good information about future profits has been incorporated into current forecasts. To provide graphical evidence supporting this theoretical prediction, we sort observations into twenty bins of previous fiscal year-end operating cash-flows over assets and calculate both average previous fiscal year-end operating cash-flows over assets and average current forecast error for each of the twenty ordered bins.

[Insert Figure 2 about here.]

Figure 2 shows a positive relationship between forecast errors and cash-flows, suggesting that analysts, in forming their EPS forecasts, do not sufficiently take into account current earnings information as measured by operating cash-flows.

To test this relationship more formally, we now regress forecast error on the cash-flow signal cf. Our model also predicts that the two other signals (Δcf and mom) should also predict forecast errors in the same direction. This happens because they both contain information about future profits that has not been fully incorporated into the expectations of sticky forecasters. Thus, we run the following regression:

$$\frac{\pi_{f,t} - F_{t-h}\pi_{f,t}}{P_{f,t-2}} = a + b_{t-h} \cdot s_{f,t-h} + \epsilon_{f,t} \tag{14}$$

for $h \in \{1, 2\}$. The variable $s_{f,t-h}$ corresponds to each of the three anomaly signals cf, Δcf, and mom that we consider in this paper. The time unit is the (fiscal) year. $\pi_{f,t}$ denotes the firm's realized EPS, which we normalize using the stock price at fiscal year-end lagged twice, that is P_{t-2}. $F_{t-h}\pi_{f,t}$ denotes the consensus EPS forecast formed in the 45 days after the announcement of cf_{t-h}. We allow for error terms to be correlated within time and within firm.

If expectations were formed rationally, expectation errors $(\pi_{f,t} - F_{t-h}\pi_{f,t})/P_{f,t-h}$ should have a zero mean conditional on information available at $t-h$. Cash-flows and prices at $t-1$ or $t-2$ are part of the information available to analysts when they form expectations about year t. If $b \neq 0$, then this suggests that forecasters underweight the information available in past profitability when forming their expectations. In our Prediction 3, we provide a structural interpretation of the coefficient b_{t-h}.

We allow for a non-zero constant a, which will capture the fact that expectations might have a constant positive bias as found in the literature (see e.g. Hong and Kacperczyk (2010), Guedj and Bouchaud (2005), or Hong and Kubik (2003)). In other words, we do not intend to analyze the average positive bias of analysts in this paper, but rather (1) the cross-section of their bias conditional on firm characteristics and (2) the dynamics of their bias over time. The results from regressions of the type of Equation 14 are reported in Table VI.

[Insert Table VI about here.]

We find that the forecast error is systematically positively related to all three signals. This finding is consistent with the idea that analyst expectations are non-rational, and that analysts tend to under-react to some persistent signals that predict future profits. One possible interpretation is to simply view past signals cf, Δcf and mom as measures as the signal itself. But our model is more general, in that it does not impose that cash-flows or returns be the only neglected signals.

C. *Prediction 3: relating anomalies to structural parameters*

C.1. *Anomalies are stronger for firms followed by sticky analysts*

We now test the link made in Prediction 3 between the stickiness of the analysts covering a firm (λ_f), and the strength of the profitability and momentum anomalies. The prediction of our theory is that when a firm is followed by stickier analysts, the three anomalies (profitability, change in profitability, and price momentum) should be more pronounced. This is quite a direct test of our theory because the test links asset prices to parameters of the model that are measured independently of stock-prices. Note that the underlying assumption is that the bias of analysts is also that of the marginal investor: if analysts were not representative of how the marginal investor is thinking, one would expect no link between analyst characteristics and stock prices. However, it seems quite plausible an assumption that the marginal investor anchors her beliefs at least to some extent on analyst forecasts. In that sense, our test is also a test that analyst expectations contain information about what investors believe, as in Engelberg et al. (2016).

To test the prediction that the strength of profitability and momentum anomalies depends on the extent to which a firm is covered by sticky analysts, we first sort stocks into terciles of the firm-level stickiness parameter λ_f. Note that the median λ_f in the first, second, and third tercile are -0.23, 0.13, and 0.41 respectively. It thus turns out that firms in the second and third tercile of the distribution of λ_f have mainly positive values (so they are subject to sticky expectations), whereas firms that fall in the first tercile of the λ_f distribution have, by and large, negative values (so that forecasts about their profits tend to be extrapolative). Within a tercile of λ_f, we sort firms into quintiles of profitability (cf), profitability momentum (Δcf), or momentum (mom). We then compute equally weighted returns of these double sorted portfolios and adjust them for risk using standard asset pricing techniques.

[Insert Table VII about here.]

Table VII displays alphas for portfolios that are double sorted on firm-level stickiness (λ_f) and cash-flows (Panel A), change in cash-flows (Panel B), and past returns (Panel C). In each month, we first sort firms into terciles of the stickiness parameter λ_f and then secondly into quintiles of the respective profitability or momentum signal. We then calculate equal-weighted returns for each of the portfolios. In Panels A and B, we use the four factor asset-pricing model of (Carhart, 1997). In Panel C, since the anomaly investigated is momentum itself, we are just using the three factors of the Fama and French (1993) asset pricing model. For each stickiness tercile, we report the alphas of each of the quintile portfolios as well as the long-short Q5-Q1 portfolio (18 portfolios). We then test whether the alpha of the Q5-Q1 portfolio in the highest λ_f tercile is greater than that in the lowest tercile (T3-T1). We find that the alpha of the long-short profitability strategy has a t-statistic of 4.94*** for the stickiest stocks, which is quite high. In contrast the t-statistic for the long-short strategy for the least sticky stocks is 2.42**. The differene between the two is highly significant: the t-statistic of the long-short portfolio consisting of the most and the least sticky stocks (i.e., the T3-T1 portfolio) is 3.18***. This result shows that compared to the least sticky stocks, the long-short profitability strategy is significantly stronger for the stickiest stocks. The effects are similar for the change in profitability strategy (Panel B), albeit slightly weaker statistically speaking (t-statistic=2.65***). Still, the alpha of the change in profitability strategy for the stickiest stocks has a t-statistic of 3.93***, far above significance levels recommended in the current asset pricing literature (see Harvey et al. (2014)). In contrast the profitability momentum strategy is not significant for the least sticky stocks (t-statistic=0.47). Portfolio strategies based on returns momentum give the same level of significance: the T3-T1 portfolio has a t-stat of 2.64***. Momentum of the stickiest firms has a t-stat of 4.79***.

C.2. Anomalies are stronger for firms with highly persistent cash-flows

Another prediction of our model is that the three anomalies should also be more pronounced for firms with more persistent cash-flows. The prime reason is that when cash-flows are highly persistent, slower updating leads to larger mistakes. To test this prediction, we thus perform portfolio tests similar to the ones carried out above.

In a first step, we measure each firm's cash-flows persistence ρ_f. We do so by individually estimating the following regression for each firm f

$$cf_{f,t} = a + \rho \cdot cf_{f,t-1} + \epsilon_{f,t}, \tag{15}$$

where $cf_{f,t}$ is the previously defined cash-flows signal.

The median cash-flows persistence is about $\rho_f \approx 0.22$ and it is estimated using 11 yearly observations (Median of $N_{\rho_f} = 11$) (see Panel B of Table IV). In a second step, we check that the profitability and momentum anomalies are indeed more pronounced among high ρ_f firms. To do so we first sort firms into terciles of ρ_f and secondly into quintiles of the cash-flows, change in cash-flows, and momentum signal. The median ρ_f in the first, second, and third tercile is -0.10, 0.28, and 0.62 respectively.

[Insert Table VIII about here.]

In Panel A and B of Table VIII, we report Carhart (1997) alphas of portfolios double sorted on ρ and the cash-flows based signals (cf and Δcf). In Panel C, we display Fama and French (1993) alphas for double sorted portfolios on ρ and mom. We generally find that alphas for all three anomalies are higher for firms with more persistent cash-flows, that is higher ρ_f. The difference between high and low persistence stocks has a t-stat of 2.18** for the cash-flows, 3.49*** for the cash-flows change, and 2.08** for the momentum signal. Focusing on the 33% stocks with the highest cash-flows persistence, the three strategies have t-statistics of 4.27 (cash-flows level), 3.71 (cash-flows change) and 3.76 (price momentum).

V Robustness

A potential concern with our results arises from the fact that we use the whole time series of firm–level consensus EPS forecasts to estimate stock-level expectation stickiness λ_f. This look-ahead bias is hard to avoid in our empirical design. In order to focus on reasonably long-term expectations –arguably most susceptible to behavioral biases– and to avoid seasonality concerns, we choose to use annual forecasts and realizations of EPS. Using annual forecasts limits us to using 11 observations to estimate the firm-level stickiness parameter for the median firm (see Table IV, Panel B). We thus need the entire time series of forecasts in order to estimate λ_f with reasonable precision. The downside of this approach is, however, that it forces us to include future forecasts and realizations of EPS in our estimate of λ_f. One might worry that such use of future information could hard-wire a correlation between our stickiness parameter and returns.

In this section, we address this concern. We use simulations in order to investigate how look-ahead bias in our estimation of λ_f affects our estimation procedure. We show that, under the assumptions of our model, look-ahead bias does not generate a spurious positive correlation between the returns to the profitability strategy and stickiness. In fact, the opposite is the case: Under the null of rational expectations, when past profits do not forecast returns, our procedure tends to generate the *opposite* relation to the one we observe in the data.

We implement the following procedure. We start from the same data generating process as in the model for signal and profit:

$$\pi_t = s_{t-1} + \epsilon_t$$
$$s_t = \rho s_{t-1} + u_t$$

The idea is then to simulate data generated by this model *under the null hypothesis that expectations are rational*, i.e. that $\lambda = 0$. Under rational expectations, as shown in Lemma 1, realized \$ returns are given by $R_{t+1} = \epsilon_{t+1} + u_{t+1}/(1+r-\rho)$. While the actual stickiness is by definition zero, it can be estimated by the econometrician by regressing profit expectation errors on forecast updates. In this rational case of our model, one can easily show that profit expectation errors are given by $\pi_{t+1} - E_t\pi_{t+1} = \epsilon_{t+1}$, and forecast revisions are given by $E_t\pi_{t+1} - E_{t-1}\pi_{t+1} = u_t$. Hence, the OLS estimate of stickiness that the econometrician obtains is given by $\frac{\widehat{cov}(\epsilon_{t+1},u_t)}{\widehat{var}u_t}$. Even though it is on average zero by design, there may be significant dispersion in the simulated data if the number of years per firm is low. In this setting, we then ask whether a financial econometrician who would estimate λ at the firm-level using the entire sample period would mechanically obtain that the profitability anomaly is stronger for stocks for which the estimated $\hat{\lambda}$ is higher.

Our Monte Carlo simulations work in the following way. In each round of simulation, we simulate a panel of 2,000 stocks – the approximate size of our sample – over 11 consecutive years – the median number of years per firm in our data. To calibrate the model, we set $r = .03$. To fix σ_ϵ, σ_u, and ρ, we need three relations. To get the first two relations, we require that the average persistence and volatility of π match the persistence and volatility of EPS/Total assets in the data (respectively .19 and .05 as shown in Table IV, Panel B). To generate a third relation, we impose that the R^2 of the regression of π_{t+1} on s_t is equal to .7.[4] For each firm in our sample, we then estimate λ by regressing profit expectation errors ϵ_{t+1} on expectation updating u_t using the entire 20 year period as we

[4]These three conditions determine σ_ϵ, σ_u and ρ uniquely via the relations:

$$\rho = \frac{\rho_\pi}{R^2}$$
$$\sigma_u = \sqrt{1-\rho^2}\sqrt{R^2}\sigma_\pi$$
$$\sigma_\epsilon = \sqrt{1-R^2}\sigma_\pi$$

do in the paper. We then implement on the simulated data a double sort similar to what the paper does on real data (results from Table VII). We first allocate each firm-year into a quintile of $\hat{\lambda}$ (each quintile thus contains 400 firms, with 12 observations each). For each of these quintiles of $\hat{\lambda}$, we then compute the realized returns of a long-short portfolio where stocks are weighted by their rank in terms of each of the two profitability signals, normalized to range between -.5 and +.5. The cash-flow signal is measured using the past profit realization π_t. The Δcf signal is given by $\pi_t - \pi_{t-1}$. For each anomaly, we thus obtain the time series of 5 portfolios R_t^q, one per quintile of $\hat{\lambda}$. Let $q = 1, ..., 5$ be the index on this quintile. We then regress these returns on $\hat{\lambda}$ quintile dummies: $R_t^q = \sum_{q \geq 2} \beta_q 1_q + \nu_t$ using the first quintile as a reference. We retrieve the t-stat on β_5. We repeat this procedure 100 times. In this model economy, returns are unpredictable and expectations are not sticky. Any significant relationship between $\hat{\lambda}$ and profitability anomaly returns would have to come from the look ahead bias in $\hat{\lambda}$, which we estimate using the entire period – and therefore using future expectation errors.

[Insert Figure 3 about here.]

In Figure 3, we report the histograms of the resulting t–statistics. In Panel (a), we use past profits as the portfolio-sorting variable, and in subfigure (b), we use past profit changes. In 2×100 simulations, we do not see one occurrence where the t-statistic (on cf of Δcf) is greater than 2. The look ahead bias induced by our estimation of λ is not strong enough to generate a statistically significant positive relationship between the returns of the profitability strategy and the estimated λ. In fact, in the pure past profit profitability signal (see Panel (a)), the look ahead bias tends to generate a relationship *opposite* to what we find in the data. The intuition for this can be described as follows: In this rational model, a stock f has a high $\hat{\lambda}_f$ if the regression coefficient of $\epsilon_{f,t+1}$ on $u_{f,t}$ is high. This happens typically when the firm has dates where $\epsilon_{f,t+1}$ and $u_{f,t}$ are both above average *and* dates where they are both below average: The coexistence of such data points produces the positive slope. Now, when past profits are known to be high at a given date T, this means that $u_{f,t-1}$ and $\epsilon_{f,t}$ are likely to be high for $t \leq T$. Thus, mechanically, knowing that $\hat{\lambda}_f$ is high, we expect $u_{f,t}$ and $\epsilon_{f,t+1}$ to be relatively likely to be both negative at future dates $t > T$, and therefore future returns – which are a combination of both effects – to be lower than average. Thus, high profit stocks with a high measured $\hat{\lambda}$ are mechanically expected to perform poorly in the future, if λ is computed using future information. This effect vanishes as the number of time periods goes to infinity. But with only 11 years, it is powerful enough to make the correlation

where R^2 is the explanatory power of s_t on π_{t+1} in a linear regression. This calibration leads to $\sigma_\epsilon = .027$, $\sigma_u = .022$ ans $\rho = .27$.

between estimated lambda and profit-based strategy returns significantly negative in most simulations (see Figure 3). The look ahead bias thus tends to bias the data against our findings. This countervailing force is also present in the Δcf anomaly, so that, on average, across simulations, we expect a slightly negative t stat for the double sort, although it is rarely significant.

VI Conclusion

In this paper, we propose a model that predicts that the profitability anomaly, one of the most economically significant stock return anomalies, arises if market participants update expectations of future profits too slowly, and if the level of profits can be predicted by persistent publicly observable signals. Assuming that financial analyst forecasts are representative of the beliefs of market participants, our theory suggests that the returns on this anomaly should be more pronounced for stocks which (1) are followed by analysts characterized by more sticky expectations or (2) firms subject to more persistent profits. The theoretical predictions are borne out by the data. We explore cross-sectional determinants of the expectation stickiness measure we propose in this paper. It turns out that less experienced analysts and busier analysts (i.e., those who follow more industries) tend to have stickier beliefs, in-line with a limited attention interpretation of our results.

References

Abarbanell, Jeffery S and Victor L Bernard, "Tests of analysts' overreaction/underreaction to earnings information as an explanation for anomalous stock price behavior," *The Journal of Finance*, 1992, *47* (3), 1181–1207.

Ali, Ashiq, April Klein, and James Rosenfeld, "Analysts' Use of Information about Permanent and Transitory Earnings Components in Forecasting Annual EPS," *The Accounting Review*, 1992, *67* (1), 183–198.

Alti, Aydoğan and Paul C Tetlock, "Biased beliefs, asset prices, and investment: A structural approach," *The Journal of Finance*, 2014, *69* (1), 325–361.

Asness, Cliff, Andrea Frazzini, and Lasse Pedersen, "Quality Minus Junk," 2014.

Barberis, Nicholas and Richard Thaler, "A survey of behavioral finance," *Handbook of the Economics of Finance*, 2003, *1*, 1053–1128.

_ , Andrei Shleifer, and Robert Vishny, "A model of investor sentiment," *Journal of financial economics*, 1998, *49* (3), 307–343.

Bertrand, Marianne and Sendhil Mullainathan, "Do people mean what they say? Implications for subjective survey data," *American Economic Review*, 2001, pp. 67–72.

Bordalo, Pedro, Nicola Gennaioli, Rafael Laporta, and Andrei Shleifer, "Diagnostic Expectations and Stock Returns," Technical Report 23863, NBER 2017.

Brav, Alon, Reuven Lehavy, and Roni Michaely, "Using Expectations to Test Asset Pricing Models," *Financial Management*, 2005, *34* (3), 31–64.

Carhart, Mark M, "On persistence in mutual fund performance," *The Journal of finance*, 1997, *52* (1), 57–82.

Cochrane, John, "Presidential address: Discount rates.," *Journal of Finance*, 2011, *66*, 1047–1108.

Coibion, Olivier and Yuriy Gorodnichenko, "What can survey forecasts tell us about information rigidities?," *Journal of Political Economy*, 2012, *120*, 116–159.

_ and _ , "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts," *American Economic Review*, 2015.

Daniel, Kent D, David Hirshleifer, and Avanidhar Subrahmanyam, "Overconfidence, arbitrage, and equilibrium asset pricing," *Journal of Finance*, 2001, pp. 921–965.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, "Investor psychology and security market under-and overreactions," *the Journal of Finance*, 1998, *53* (6), 1839–1885.

Debondt, Werner and Richard Thaler, "Does the Stock Market Overreact?," *Journal of Finance*, 1985, *40* (3), 793–805.

_ **and** _ , "Do security analysts overreact?," *The American Economic Review*, 1990, pp. 52–57.

Diether, Karl B, Christopher J Malloy, and Anna Scherbina, "Differences of opinion and the cross section of stock returns," *The Journal of Finance*, 2002, *57* (5), 2113–2141.

Engelberg, Joseph, David McLean, and Jeffrey Pontiff, "Anomalies and News," 2016. Working Paper University of California San Diego, Boston College, and University of Alberta.

Fama, Eugene F and Kenneth R French, "Common risk factors in the returns on stocks and bonds," *Journal of financial economics*, 1993, *33* (1), 3–56.

Gennaioli, Nicola, Yueran Ma, and Andrei Shleifer, "Expectations and Investment," 2015.

Greenwood, Robin and Andrei Shleifer, "Expectations of Returns and Expected Returns," *Review of Financial Studies*, 2014, *27* (3), 714–746.

Guedj, Olivier and Jean-Philippe Bouchaud, "Experts' Earning Forecasts: Bias, herding and Gossamer Iinformation," *International Journal of Theoretical and Applied Finance*, 2005, *8* (07), 933–946.

Harvey, Campbell, Yan Liu, and Heqing Zhu, "... and the Cross-Section of Expected Returns," 2014.

Hirshleifer, David, "Investor psychology and asset pricing," *Journal of finance*, 2001, pp. 1533–1597.

Hong, Harrison and Jeffrey D Kubik, "Analyzing the analysts: Career concerns and biased earnings forecasts," *Journal of Finance*, 2003, pp. 313–351.

_ **and Jeremy C. Stein**, "A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets," *The Journal of Finance*, 1999, *54* (6), 2143–2184.

_ **and Marcin Kacperczyk**, "Competition and bias," *The Quarterly Journal of Economics*, 2010, *125* (4), 1683–1725.

_ **, Terence Lim, and Jeremy Stein**, "Bad News Travel Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies," *Journal of Finance*, 2000, *55* (1), 265–295.

Hou, Kewei, "Industry Information Diffusion and the Lead-Lag Effect in Stock Returns," *Review of Financial Studies*, 2007, *20* (4), 1113–1138.

Jegadeesh, Narasimhan and Sheridan Titman, "Returns to buying winners and selling losers: Implications for stock market efficiency," *The Journal of finance*, 1993, *48* (1), 65–91.

Lakonishok, Josef, Andrei Shleifer, and Robert Vishny, "Contrarian Investment, Extrapolation and Risk," *Journal of Finance*, 1994, *49* (5), 1541–1578.

Landier, Augustin, Guillaume Simon, and David Thesmar, "The Capacity of Trading Strategies," 2015.

Laporta, Rafael, "Expectations and the Cross-Section of Stock Returns," *Journal of Finance*, 1996, *51* (5), 1715–1742.

Lemperiere, Yves, Cyril Deremble, Trung-Tu Nguyen, Philipp Seager, Marc Potters, and Jean-Philippe Bouchaud, "Risk Premia: Asymmetric Tail Risks and Excess Returns," *arXiv:1409.7720v3*, 2015.

Mankiw, Gregory and Ricardo Reis, "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Philips Curve," *Quarterly Journal of Economics*, 2002, *117* (4), 1295–1328.

Mankiw, N Gregory and Ricardo Reis, "Sticky information versus sticky prices: a proposal to replace the New Keynesian Phillips curve," Technical Report, National Bureau of Economic Research 2001.

Markov, Stanimir and Ane Tamayo, "Predictability in Financial Analyst Forecast Errors: Learning or Irrationality?," *Journal of Accounting Research*, 2006, *44* (4), 725–761.

Mikhail, Michael B., Beverly R. Walther, and Richard H. Willis, "The effect of experience on security analyst underreaction," *Accounting Information and Management*, 2003, *35* (1), 101–116.

Novy-Marx, Robert, "The Other Side of Value: The Gross Profitability Premium,," *Journal of Financial Economics*, 2013, *108* (1), 1–28.

_ , "Fundamentally, Momentum is Fundamental Momentum," 2015.

Reis, Ricardo, "Inattentive Producers," *Review of Economic Studies*, 2006, *73* (3), 793–821.

Robinson, David and Denys Glushkov, "A note on IBES unadjusted data," *Wharton Research Data Services*, 2006, pp. 1–15.

Sloan, Richard, "Do stock prices fully reflect information in accruals and cash flows about future earnings?," *The Accounting Review*, 1996, *71*, 289–315.

Weber, Michael, "Cash flow duration and the term structure of equity returns," Technical Report, National Bureau of Economic Research 2016.

Figures

Figure 1

Forecast errors and forecast revisions

This figure shows the forecast errors as a function of forecast revisions. We sort observations into twenty bins of forecast revision $(F_{t-1}\pi_{f,t} - F_{t-2}\pi_{f,t})/P_{f,t-2}$ and calculate average forecast error (defined as, $(\pi_{f,t} - F_{t-1}\pi_{f,t})/P_{f,t-2}$) and average forecast revision for each of the twenty ordered bins.

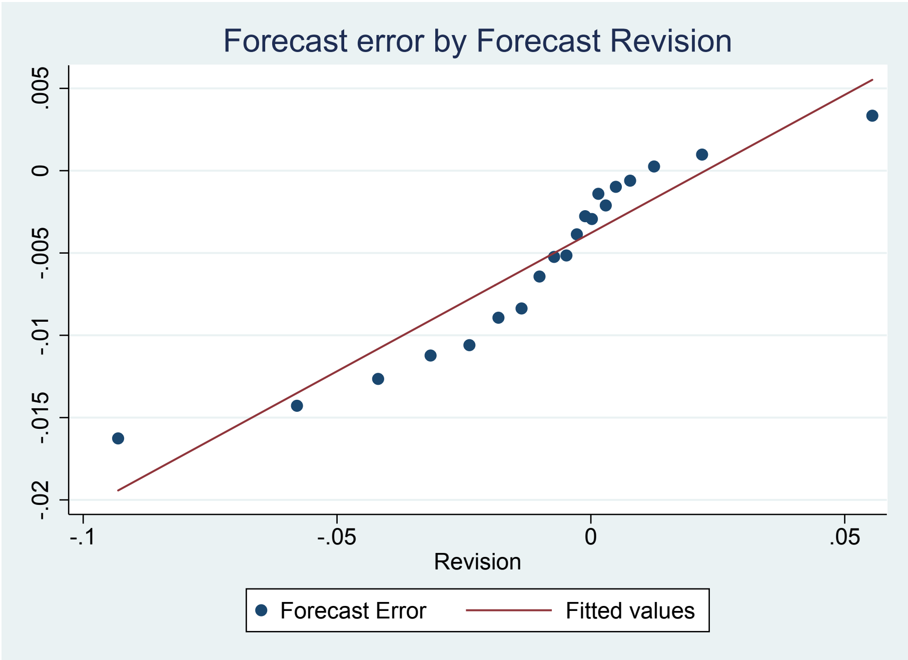


Figure 2

Forecast error and cash-flows

This figure shows forecast error as a function of past cash-flows. We sort observations into twenty ordered bins of the previous fiscal year's operating cash-flows to assets ratio. For each of the twenty ordered groups, we then calculate both average previous year's cash-flows to assets ratio and current fiscal year's average forecast error.

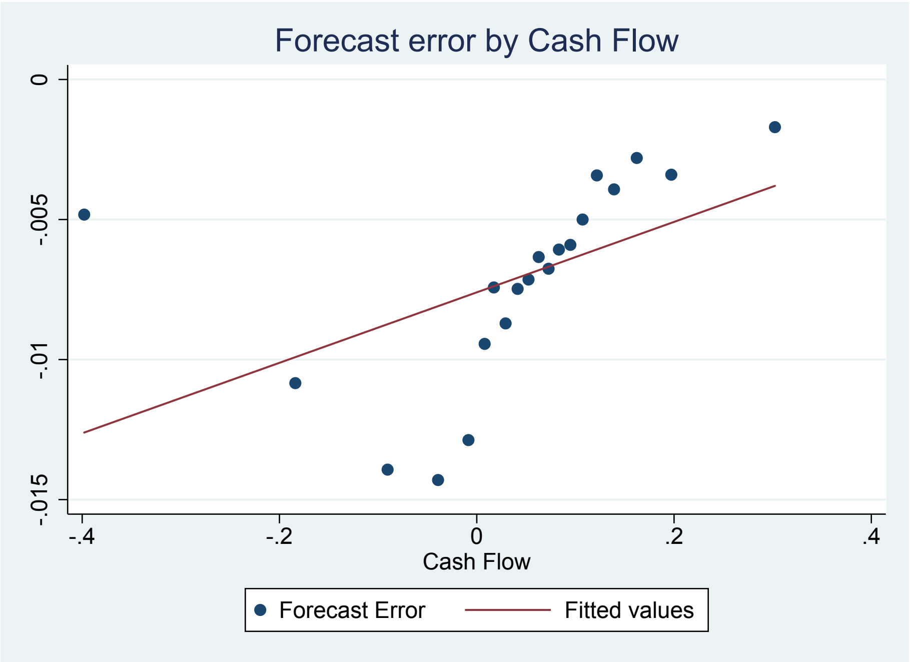


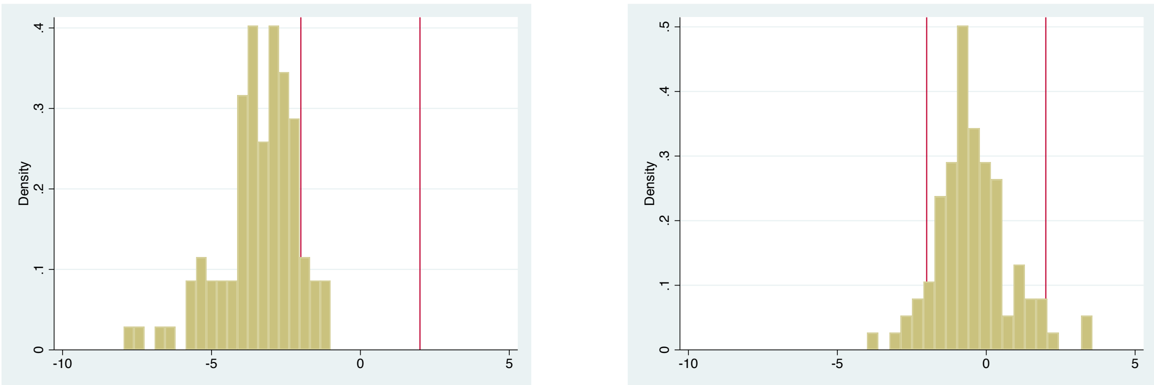
Figure 3

T-statistics of Double Sorts Under the Null of Rational Expectations
Results From Simulations

These histograms represent the distribution of t-stats from double sorts by profitability and stickiness, for 100 simulations, under the null hypothesis that expectations are *not* sticky. Each simulation works like this. For 2,000 firms over 12 years, we simulate our baseline model assuming $\lambda = 0$. Signals have persistence ρ and predict profits one period in advance:

$$\pi_{t+1} = s_t + \epsilon_{t+1}$$
$$s_t = \rho s_{t-1} + u_{t-1}$$

Expectations are fully rational (the true $\lambda = 0$), so realized returns are thus given by $\epsilon_{t+1} + u_{t+1}/(1+r-\rho)$. For each firm we then estimate a "stickiness" level λ by regressing profit expectation errors (given by ϵ_{t+1}) on expectation updating (u_t in this rational model). Since expectations are rational, the average $\hat{\lambda}$ is zero, but firm by firm, $\hat{\lambda}$ can be positive or negative. We then implement the double sort on stickiness and profitability. We first allocate each firm-year into a quintile of $\hat{\lambda}$ (each quintile thus contains 400 firms). For each of these quintiles of $\hat{\lambda}$, we then compute the realized returns of a long-short portfolio where stocks are weighted by their rank in terms of past profitability, normalized to range between -.5 and +.5. We obtain the time series of five profitability portfolios R_t^q, one per quintile q of firm-level $\hat{\lambda}$. We then regress these returns on λ quintile dummies: $R_t^q = \sum_q \beta_q 1_q + \nu_t$. We retrieve the t-stat on β_5. We repeat this procedure 100 times, and report the histograms of t-statistics below. Panel (a) uses as a profitability signal the past profit π_t of the firm. Panel (b) uses the change in past profit $\pi_t - \pi_{t-1}$.



(a) Cash-flow signal (cf)

(b) Δ Cash-flow signal (Δcf)

Tables

Table I

Summary statistics

This table shows summary statistics for the I/B/E/S earnings forecasts sample. $\pi_{f,t}$ is the actual EPS reported in I/B/E/S. $F_{t-1}\pi_{f,t}$, $F_{t-2}\pi_{f,t}$, and $F_{t-3}\pi_{f,t}$ are the one, two, and three year ahead consensus forecasts for earnings at date t, which we calculate as the median earnings forecast of all forecasts issued during the 45 days following the respective fiscal year earnings announcement at $t-1$, $t-2$, and $t-3$. $(\pi_{f,t} - F_{t-1}\pi_{f,t})/P_{f,t-2}$, $(\pi_{f,t} - F_{t-2}\pi_{f,t})/P_{f,t-2}$, and $(\pi_{f,t} - F_{t-3}\pi_{f,t})/P_{f,t-3}$ are the forecast errors with respect to the one, two, and three year ahead earnings forecast. $P_{f,t-n}$ denotes the stock price at fiscal year end $t-n$. $(F_{t-1}\pi_{f,t} - F_{t-2}\pi_{f,t})/P_{f,t-2}$ and $(F_{t-2}\pi_{f,t} - F_{t-3}\pi_{f,t})/P_{f,t-3}$ are the forecast revisions of the one and two year ahead earnings forecasts. $(\pi_{f,t-1} - \pi_{f,t-2})/P_{f,t-2}$ is the trend in earnings. cf is the ratio between operating cash-flows (Compustat item *oancf*) divided by total assets (item *at*). $\Delta cf_{f,t}$ is year-on-year change in the operating cash-flows to assets ratio. $mom_{f,t}$ is the usual momentum signal, i.e., the cumulative firm-level return between months t-12 and t-2. To reduce the impact of outliers, all variables are trimmed by removing observations for which the value of a variable deviates from the median by more than five times the interquartile range.

	count	mean	sd	min	p25	p50	p75	max
$(\pi_{f,t} - F_{t-1}\pi_{f,t})/P_{f,t-2}$	54090	-0.006	0.028	-0.130	-0.014	-0.001	0.005	0.126
$(\pi_{f,t} - F_{t-2}\pi_{f,t})/P_{f,t-2}$	54062	-0.015	0.044	-0.225	-0.032	-0.007	0.005	0.207
$(F_{t-1}\pi_{f,t} - F_{t-2}\pi_{f,t})/P_{f,t-2}$	54090	-0.009	0.029	-0.134	-0.020	-0.004	0.004	0.126
$(F_{t-2}\pi_{f,t} - F_{t-3}\pi_{f,t})/P_{f,t-3}$	15632	-0.006	0.031	-0.145	-0.017	-0.003	0.006	0.138
$(\pi_{f,t-1} - \pi_{f,t-2})/P_{f,t-2}$	45545	0.002	0.034	-0.149	-0.010	0.004	0.014	0.157
$(\pi_{f,t-2} - \pi_{f,t-3})/P_{f,t-3}$	39272	0.003	0.034	-0.150	-0.009	0.005	0.015	0.157
$cf_{f,t}$	51710	0.079	0.107	-0.599	0.035	0.082	0.132	0.699
$\Delta cf_{f,t}$	51038	-0.001	0.072	-0.381	-0.029	-0.001	0.027	0.381
$mom_{f,t}$	33636	0.123	0.431	-0.991	-0.131	0.088	0.316	2.567

Table II

Profitability anomaly in the IBES sample

This table displays excess returns (Panel A), CAPM (Panel B), Fama and French (1993) three factor (Panel C), and Carhart four factor (1997) alphas (Panel D) for quintile portfolios, which are constructed based on the level of operating cash-flows *(cf)*, the change in operating cash-flows *(Δcf)*, or momentum *(mom)*. Excess returns and alphas are in percentage. Cash-flows *(cf)* is defined as Compustat item *oancf* divided by item *at*. Δcf is the change in *cf* since the previous earnings announcement. *mom* is cumulative firm-level return between months t-12 and t-2. The cash-flows signal is updated in the month following the month of a firm's announcement of fiscal year earnings, which we obtain from Compustat quarterly. The signal is valid until the month in which the next fiscal year earnings are announced. Q5-Q1 is the long–short portfolio which is long the 20% of firms with the highest values of the respective signal (Fifth quintile) and short the 20% of firms with the lowest values (First quintile). Portfolios are equally-weighted. The sample period runs from 1990 to 2013. Standard errors are adjusted for heteroskedasticity and autocorrelations up to 12 lags. t–statistics in parentheses. (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)

	(1) Q1	(2) Q2	(3) Q3	(4) Q4	(5) Q5	(6) Q5-Q1
Panel A: Excess returns						
cf	0.55	0.73**	0.88***	0.97***	1.11***	0.56**
	(1.35)	(2.35)	(3.22)	(3.62)	(4.14)	(2.33)
Δcf	0.84**	0.77***	0.71***	0.91***	1.04***	0.20***
	(2.52)	(2.75)	(2.70)	(3.29)	(3.24)	(2.83)
mom	0.43	0.65**	0.80***	1.00***	1.44***	1.01***
	(1.14)	(2.24)	(3.28)	(3.93)	(3.66)	(2.89)
Panel B: CAPM						
cf	-0.27	0.06	0.25	0.34*	0.45**	0.72***
	(-1.26)	(0.33)	(1.41)	(1.78)	(2.41)	(3.14)
Δcf	0.08	0.13	0.14	0.28	0.29	0.21***
	(0.41)	(0.75)	(0.78)	(1.65)	(1.48)	(2.94)
mom	-0.40*	0.03	0.25	0.44**	0.70***	1.10***
	(-1.76)	(0.15)	(1.46)	(2.51)	(2.60)	(3.48)
Panel C: FF1993						
cf	-0.28*	-0.07	0.13	0.23**	0.38***	0.66***
	(-1.84)	(-1.03)	(1.59)	(2.36)	(3.33)	(3.23)
Δcf	0.02	0.01	-0.00	0.17**	0.23**	0.21***
	(0.22)	(0.13)	(-0.06)	(2.35)	(2.28)	(3.04)
mom	-0.53***	-0.12	0.13	0.34***	0.70***	1.23***
	(-3.11)	(-1.11)	(1.46)	(4.36)	(3.30)	(3.68)
Panel D: Carhart						
cf	-0.24	0.01	0.20**	0.30***	0.46***	0.70***
	(-1.54)	(0.09)	(2.48)	(3.36)	(3.73)	(3.56)
Δcf	0.11	0.09	0.06	0.24***	0.29**	0.18***
	(1.06)	(1.28)	(0.68)	(3.40)	(2.52)	(2.87)

Table III

Estimating expectation stickiness

In column (1), we regress the one year forecast error $(\pi_{f,t} - F_{t-1}\pi_{f,t})/P_{f,t-2}$ on the forecast revision between dates $t-1$ and $t-2$, that is $(F_{t-1}\pi_{f,t} - F_{t-2}\pi_{f,t})/P_{f,t-2}$. In column (2) we regress the forecast error on the individual components of the forecast revision. In column (3) we add the past trend in profits to capture potential extrapolative patterns. In Panel B, we use the forecast revision at date $t-1$ that is $(F_{t-1}\pi_{f,t} - F_{t-2}\pi_{f,t})/P_{f,t-2}$ as the dependent variable and regress it on the forecast revision at date $t-2$, i.e. $(F_{t-2}\pi_{f,t} - F_{t-3}\pi_{f,t})/P_{f,t-3}$. Standard errors are double clustered at the firm–year level. t–statistics in parentheses. (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)

Panel A: Dependent variable: $(\pi_{f,t} - F_{t-1}\pi_{f,t})/P_{f,t-2}$			
	(1)	(2)	(3)
$(F_{t-1}\pi_{f,t} - F_{t-2}\pi_{f,t})/P_{f,t-2}$	0.165***		0.176***
	(10.28)		(9.99)
$F_{t-1}\pi_{f,t}/P_{f,t-2}$		0.156***	
		(9.65)	
$F_{t-2}\pi_{f,t}/P_{f,t-2}$		-0.201***	
		(-11.30)	
$(\pi_{f,t-1} - \pi_{f,t-2})/P_{f,t-2}$			-0.011
			(-0.83)
Observations	54,090	54,090	45,545
R^2	0.030	0.036	0.032
Panel B: Dependent variable: $(F_{t-1}\pi_{f,t} - F_{t-2}\pi_{f,t})/P_{f,t-3}$			
	(1)	(2)	(3)
$(F_{t-2}\pi_{f,t} - F_{t-3}\pi_{f,t})/P_{f,t-3}$	0.063**		0.087**
	(2.27)		(2.33)
$F_{t-2}\pi_{f,t}/P_{f,t-3}$		0.048	
		(1.61)	
$F_{t-3}\pi_{f,t}/P_{f,t-3}$		-0.103***	
		(-3.76)	
$(\pi_{f,t-2} - \pi_{f,t-3})/P_{f,t-3}$			-0.027
			(-1.25)
Observations	16,118	16,118	14,646
R^2	0.005	0.015	0.008

Table IV

Summary statistics (λ and ρ at the analyst- and firm-level)

In Panel A, we report summary statistics for several analyst–level variables. The parameter λ_a is the analyst–level stickiness parameter obtained from running the following regression

$$\frac{\pi_{f,t} - F_{a,t-1}\pi_{f,t}}{P_{f,t-2}} = a_a + b_a \cdot \frac{F_{a,t-1}\pi_{f,t} - F_{a,t-2}\pi_{f,t}}{P_{f,t-2}} + \epsilon_{a,f,t}$$

separately for each analyst a. $F_{a,t-h}\pi_{f,t}$ is the first EPS forecast issued by analyst a for firm f in the 45 days after the announcement of earnings for fiscal year $t-h$. In estimating this regression, we use all available observations at the analyst-level. This regression identifies the analyst-level stickiness parameter λ_a using forecasts issued by the same analyst for different firms. The stickiness parameter λ_a is simply the transformation $\lambda_a = b_a/(1+b_a)$ of the coefficient b_a. N_{λ_a} is the number of analyst–level observations used to identify λ_a. *Experience* is the difference between the current year and the year an analyst first appeared in the I/B/E/S database. *Firm experience* is the firm–specific experience of an analyst, which we define as the difference between the current year and the year in which an analyst has issued an EPS forecast for a given firm for the first time. *Industry experience* is the number of years an analyst has been forecasting earnings for the SIC2 industry to which the firm belongs. *Covered industries* is the number of SIC2 industries covered by the analyst. Analogously *Covered firms* is the number of firms an analyst covers. We calculate the time-series averages of these analyst-characteristics during the sample period.

In Panel B, we report descriptive statistics for firm–level variables and parameters obtained from estimating the regression

$$\frac{\pi_{f,t} - F_{t-1}\pi_{f,t}}{P_{f,t-2}} = a_f + b_f \cdot \frac{F_{t-1}\pi_{f,t} - F_{t-2}\pi_{f,t}}{P_{f,t-2}} + \epsilon_{f,t}$$

for each firm separately. This regression identifies the firm–level stickiness parameter λ_f by using the entire history of consensus forecasts and errors. λ_f is simply the transformation $\lambda_f = b_f/(1+b_f)$ of coefficient b_f in the above regression. N_{λ_f} is the number of firm–level observations used to identify the λ_f stickiness parameter. ρ_f is obtained from estimating $cf_{f,t} = a_f + \rho_f \cdot cf_{f,t-1} + \epsilon_{f,t}$ for each firm, where cf is *oancf/at*. N_{ρ_f} is the number of observations used for estimating the cash-flows persistence at the firm–level. *Firm size* is *ln(assets)*. *EPS volatility* is the standard deviation of EPS at the firm-level. *Firm level forecast dispersion* is the standard deviation of analyst forecasts issued for the firm. *Within industry EPS (forecast) dispersion* is the dispersion of EPS (forecasts) within a SIC2 industry. We calculate time series averages over the sample period of the firm-level variables. All variables are trimmed by removing observations for which the value of a variable deviates from the median by more than five times the interquartile range.

Panel A: Analyst-level

	count	mean	sd	min	p5	p10	p25	p50	p75	p90	p95	max
λ_a	6938	0.16	0.56	-2.26	-0.78	-0.42	-0.05	0.18	0.40	0.66	0.91	2.61
N_{λ_a}	6885	22.96	27.68	2.00	2.00	2.00	4.00	11.00	31.00	62.00	85.00	151.00
Experience	6938	6.25	4.38	0.00	1.29	1.71	2.86	5.09	8.48	12.89	15.52	20.18

Continued on next page.

Table IV – continued from previous page

	count	mean	sd	min	p5	p10	p25	p50	p75	p90	p95	max
Firm experience	6908	2.25	1.37	0.00	0.67	0.82	1.19	1.91	3.00	4.18	4.97	7.64
Industry experience	6799	4.44	2.97	0.00	1.03	1.32	2.09	3.70	6.07	8.81	10.56	14.28
Covered industries	6887	3.19	1.97	1.00	1.00	1.00	1.71	2.72	4.19	5.90	7.16	11.68
Covered firms	6913	12.29	6.43	1.00	2.63	4.13	7.98	11.88	15.78	19.97	23.11	51.94

Panel B: Firm-level

	count	mean	sd	min	p5	p10	p25	p50	p75	p90	p95	max
λ_f	5916	0.13	0.68	-2.67	-1.02	-0.57	-0.11	0.15	0.39	0.76	1.16	2.97
N_{λ_f}	5916	8.87	6.71	2.00	2.00	2.00	3.00	7.00	13.00	20.00	24.00	26.00
ρ_f	5916	0.19	0.49	-3.29	-0.56	-0.36	-0.08	0.22	0.49	0.70	0.83	3.45
N_{ρ_f}	5916	13.09	7.46	2.00	3.00	4.00	7.00	11.00	19.00	25.00	26.00	26.00
Firm size	5899	6.10	1.82	-0.74	3.32	3.85	4.78	5.99	7.29	8.48	9.28	13.34
EPS volatility	5842	0.05	0.04	0.00	0.01	0.02	0.02	0.04	0.07	0.10	0.13	0.32
Firm level forecast dispersion	5689	0.12	0.11	0.00	0.02	0.03	0.05	0.08	0.15	0.26	0.37	0.66
Within industry forecast dispersion	5897	0.06	0.02	0.01	0.03	0.04	0.04	0.05	0.06	0.07	0.09	0.11
Within industry EPS dispersion	5897	0.07	0.02	0.01	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.13

Table V

Explaining λ_a and λ_f

In Panel A, we relate the analyst-level stickiness parameter λ_a to various cross-sectional analyst characteristics. The cross-sectional characteristics are time-series averages over the whole sample-period. In Panel B, we relate the firm-level stickiness parameter λ_f to various cross-sectional firm characteristics. For variable definitions, see Table IV. Standard errors account for heteroskedasticity. t–statistics in parentheses. (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)

Panel A: Dependent variable λ_a (Analyst-level)

	(1)	(2)	(3)	(4)	(5)	(6)
Experience	-0.005***					-0.002
	(-3.20)					(-0.61)
Firm experience		-0.019***				-0.012*
		(-4.26)				(-1.65)
Industry experience			-0.010***			-0.001
			(-4.64)			(-0.13)
Covered industries				0.011***		0.020***
				(3.44)		(5.05)
Covered firms					-0.003***	-0.005***
					(-2.73)	(-4.25)
Constant	0.185***	0.197***	0.200***	0.116***	0.191***	0.196***
	(14.43)	(14.24)	(15.02)	(8.85)	(11.67)	(9.34)
Observations	6,938	7,054	6,890	7,036	7,063	6,716
R^2	0.001	0.002	0.003	0.002	0.001	0.007

Panel B: Dependent variable λ_f (Firm-level)

	(1)	(2)	(3)	(4)	(5)	(6)
Firm size	-0.010**					-0.007
	(-2.04)					(-1.26)
EPS volatility		2.210***				2.460***
		(8.93)				(8.82)
Firm level forecast dispersion			-0.037			-0.134
			(-0.42)			(-1.33)
Within industry forecast dispersion				-2.563***		-3.210*
				(-4.42)		(-1.81)
Within industry EPS dispersion					-2.010***	-0.221
					(-3.47)	(-0.12)
Constant	0.193***	0.016	0.132***	0.275***	0.273***	0.248***
	(5.89)	(1.05)	(9.77)	(8.44)	(6.71)	(3.70)
Observations	6,009	5,940	5,788	6,007	6,007	5,737
R^2	0.001	0.015	0.000	0.004	0.002	0.021

Table VI

Forecast errors and anomaly signals

In this table we present the results from regressing firm-level EPS forecast errors on profitability and momentum signals. The dependent variable in Panel A is the forecast error based on the consensus forecast for the current fiscal year earnings, that is $(\pi_{f,t} - F_{t-1}\pi_{f,t})/P_{f,t-2}$. Analogously, the dependent variable in Panel B is the forecast error with respect to the consensus forecast that was issued in the previous fiscal year, i.e., $(\pi_t - F_{t-2}\pi_t)/P_{t-2}$. *cf* is Compustat item *oancf* divided by item *at*. Δcf is the year-on-year difference in the *cf*. *mom* is the cumulative firm-level return between months t-12 and t-2 relative to the month t in which earnings are announced. Standard errors are double clustered at the firm–year level. t–statistics in parentheses. (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)

Panel A: Dependent variable $(\pi_{f,t} - F_{t-1}\pi_{f,t})/P_{f,t-2}$			
	(1)	(2)	(3)
$cf_{f,t-1}$	0.018***		
	(6.31)		
$\Delta cf_{f,t-1}$		0.016***	
		(5.96)	
$mom_{f,t-1}$			0.006***
			(7.97)
Observations	63,547	61,166	39,290
R^2	0.027	0.024	0.037

Panel B: Dependent variable $(\pi_{f,t} - F_{t-2}\pi_{f,t})/P_{f,t-2}$			
	(1)	(2)	(3)
$cf_{f,t-2}$	0.040***		
	(7.75)		
$\Delta cf_{f,t-2}$		0.017***	
		(3.96)	
$mom_{f,t-2}$			0.007***
			(5.14)
Observations	52,614	47,443	34,083
R^2	0.036	0.030	0.040

Table VII

Anomalies sorted by expectation stickiness λ_f

Panels A and B of this table show Carhart (1997) four factor alphas for equally-weighted portfolios that are double sorted on λ_f and the level (and change) of cash-flows *cf* (Δcf). Panel C displays Fama and French (1993) three factor alphas for equally-weighted portfolios that are double sorted on λ_f and momentum (*mom*). We first sort stocks into terciles of the firm-level stickiness parameter λ_f. Within a tercile of the stickiness parameter, we sort firms into quintiles of cash-flows *(cf)*, change in cash-flows (Δcf) or past returns (*mom*). We also show the alphas for the Q5-Q1 long-short portfolios as well as the differences in alphas between the high stickiness (T3) and low stickiness (T1) portfolios. We display results for the cash-flows signal *(cf)* in Panel A, the change in cash-flows *(Δcf)* signal in Panel B, and momentum (*mom*) in Panel C. Standard errors are adjusted for heteroskedasticity and autocorrelations up to 12 lags. t–statistics in parentheses. (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)

	(1) Q1	(2) Q2	(3) Q3	(4) Q4	(5) Q5	(6) Q5-Q1
Panel A: Cash-flows (cf)						
T1	-0.18	0.03	0.21**	0.26**	0.33**	0.51**
	(-1.00)	(0.34)	(2.29)	(2.30)	(2.37)	(2.42)
T2	0.12	0.16*	0.31***	0.41***	0.59***	0.47**
	(0.76)	(1.74)	(3.50)	(4.50)	(4.39)	(2.40)
T3	-0.58***	-0.18*	0.12	0.20*	0.44***	1.02***
	(-3.56)	(-1.81)	(1.39)	(1.80)	(3.74)	(4.94)
T3 - T1	-0.40**	-0.21**	-0.09	-0.06	0.11	0.51***
	(-2.36)	(-2.45)	(-0.99)	(-0.78)	(1.11)	(3.18)
Panel B: Change in cash-flows (Δcf)						
T1	0.12	0.10	0.09	0.19**	0.16	0.04
	(0.89)	(1.21)	(0.86)	(2.27)	(1.19)	(0.47)
T2	0.32***	0.23***	0.21***	0.29***	0.55***	0.23**
	(2.71)	(2.78)	(2.81)	(3.48)	(3.79)	(2.21)
T3	-0.10	-0.11	-0.08	0.17**	0.21**	0.31***
	(-1.00)	(-1.09)	(-0.65)	(2.03)	(2.10)	(3.93)
T3 - T1	-0.22**	-0.21*	-0.17*	-0.02	0.05	0.27***
	(-2.19)	(-1.96)	(-1.81)	(-0.19)	(0.43)	(2.65)
Panel C: Momentum (mom)						
T1	-0.51***	-0.08	0.10	0.32***	0.60**	1.11***
	(-3.08)	(-0.68)	(1.06)	(3.44)	(2.38)	(3.28)
T2	-0.20	-0.01	0.24**	0.39***	0.79***	0.99***
	(-1.00)	(-0.08)	(2.49)	(4.51)	(3.61)	(2.76)
T3	-0.87***	-0.25**	0.05	0.34***	0.65***	1.51***
	(-4.94)	(-1.97)	(0.41)	(3.41)	(3.56)	(4.79)
T3 - T1	-0.36***	-0.17*	-0.05	0.01	0.05	0.41***
	(-3.16)	(-1.87)	(-0.57)	(0.12)	(0.30)	(2.64)

Table VIII

Anomalies sorted by persistence ρ_f

Panels A and B of this table show Carhart (1997) four factor alphas for equally-weighted portfolios that are double sorted on ρ_f and the level (and change) of cash-flows *cf* (Δcf). Panel C displays Fama and French (1993) three factor alphas for equally-weighted portfolios that are double sorted on ρ_f and momentum (*mom*). We first sort stocks into terciles of the firm-level persistence parameter ρ_f. Within a tercile of the persistence parameter, we sort firms into quintiles of cash-flows *(cf)*, change in cash-flows (Δcf) or past returns (*mom*). We also show the alphas for the Q5-Q1 long-short portfolios as well as the differences in alphas between the high persistence (T3) and low persistence (T1) portfolios. We display results for the cash-flows signal *(cf)* in Panel A, the change in cash-flows *(Δcf)* signal in Panel B, and momentum (*mom*) in Panel C. Standard errors are adjusted for heteroskedasticity and autocorrelations up to 12 lags. t–statistics in parentheses. (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)

	(1) Q1	(2) Q2	(3) Q3	(4) Q4	(5) Q5	(6) Q5-Q1
Panel A: Cash-flows (cf)						
T1	-0.34**	-0.11	0.14	0.23**	0.30**	0.64***
	(-2.27)	(-0.89)	(1.38)	(2.38)	(2.36)	(3.48)
T2	-0.09	0.06	0.19**	0.38***	0.42***	0.51**
	(-0.60)	(0.66)	(2.20)	(3.72)	(3.25)	(2.48)
T3	-0.26*	0.04	0.27***	0.33***	0.65***	0.91***
	(-1.69)	(0.49)	(2.78)	(3.65)	(4.61)	(4.27)
T3 - T1	0.08	0.15	0.13*	0.10	0.35***	0.27**
	(0.86)	(1.12)	(1.71)	(1.58)	(4.61)	(2.18)
Panel B: Change in cash-flows (Δcf)						
T1	0.15	0.03	-0.03	0.04	0.09	-0.06
	(1.38)	(0.36)	(-0.22)	(0.41)	(0.85)	(-0.61)
T2	0.03	0.21**	0.04	0.29***	0.39***	0.36***
	(0.29)	(2.55)	(0.38)	(3.27)	(2.82)	(3.79)
T3	0.07	0.07	0.14*	0.37***	0.45***	0.37***
	(0.57)	(0.83)	(1.92)	(4.65)	(3.42)	(3.71)
T3 - T1	-0.08	0.04	0.16*	0.33***	0.36***	0.43***
	(-0.73)	(0.43)	(1.80)	(3.87)	(4.55)	(3.49)
Panel C: Momentum (mom)						
T1	-0.55***	-0.18	0.05	0.27**	0.48**	1.03***
	(-3.26)	(-1.50)	(0.54)	(2.49)	(2.55)	(3.39)
T2	-0.52***	-0.06	0.21**	0.32***	0.69***	1.21***
	(-2.98)	(-0.48)	(2.23)	(4.68)	(3.29)	(3.52)
T3	-0.49**	-0.13	0.14	0.41***	0.88***	1.37***
	(-2.59)	(-1.12)	(1.64)	(5.31)	(3.66)	(3.76)
T3 - T1	0.06	0.06	0.08	0.14*	0.40***	0.34**
	(0.56)	(0.65)	(1.32)	(1.87)	(3.28)	(2.08)

APPENDIX : PROOFS

A Proof of Proposition 1

Our goal here is to compute prices and returns. Start from the definition of sticky expectations:

$$
\begin{aligned}
F_t(\pi_{t+k}) &= (1-\lambda)\sum_{j\geq 0}\lambda^j E_{t-j}\pi_{t+k} \\
&= (1-\lambda)\rho^{k-1}\sum_{j\geq 0}\lambda^j\rho^j s_{t-j}
\end{aligned}
$$

We can then plug this back into prices:

$$
\begin{aligned}
P_t &= \sum_{k\geq 1}\frac{F_t\pi_{t+k}}{(1+r)^k} \\
&= \sum_{k\geq 1}\frac{1}{(1+r)^k}((1-\lambda)\rho^{k-1}\sum_{j\geq 0}\lambda^j\rho^j s_{t-j}) \\
&= \sum_{j\geq 0}\sum_{k\geq 1}\frac{1}{(1+r)^k}((1-\lambda)\rho^{k-1}\lambda^j\rho^j s_{t-j}) \\
&= \sum_{j\geq 0}\frac{1-\lambda}{1+r}[\sum_{k\geq 0}\frac{\rho^k}{(1+r)^k}](\lambda^j\rho^j s_{t-j}) \\
&= \sum_{j\geq 0}\frac{1-\lambda}{1+r}[\frac{1}{1-\rho/(1+r)}](\lambda^j\rho^j s_{t-j}) \\
&= \frac{1-\lambda}{1+r-\rho}\sum_{j\geq 0}\lambda^j\rho^j s_{t-j}
\end{aligned}
$$

Finally, we can compute dollar returns as:

$$
\begin{aligned}
R_{t+1} &= P_{t+1} + \pi_{t+1} - (1+r)P_t \\
&= ms_{t+1} + s_t + \epsilon_{t+1} - zm\sum_{k\geq 0}(\lambda\rho)^k s_{t-k}
\end{aligned}
$$

B Proof of Prediction 2

First notice that $Cov(s_{t-k}, s_t) = \rho^k Var(s_t)$.
From Equation (6):

$$
E_t\left(F_t\pi_{t+1}|\,\pi_t\right) = (1-\lambda)\sum_{k\geq 0}(\lambda\rho)^k E_t(s_{t-k}|\pi_t)
$$

Since s_t and π_t are Gaussian stationnary random variables centered on zero, we can write the conditional expectations as simple projections.

- for $k > 0$:

$$
\begin{aligned}
E_t(s_{t-k}|\pi_t) &= \frac{Cov(s_{t-k}, \pi_t)}{Var(\pi_t)}\pi_t \\
&= \frac{Cov(s_{t-k}, s_{t-1} + \epsilon_t)}{Var(s_t) + \sigma_\epsilon^2}\pi_t \\
&= \frac{Cov(s_{t-(k-1)}, s_t)}{Var(s_t) + \sigma_\epsilon^2}\pi_t \\
&= \rho^{k-1}\frac{Var(s_t)}{Var(s_t) + \sigma_\epsilon^2}\pi_t \\
&= \rho^{k-1}\frac{\sigma_u^2}{\sigma_u^2 + (1-\rho^2)\sigma_\epsilon^2}\pi_t
\end{aligned}
$$

  because $Var(s_t) = \rho^2 Var(s_t) + \sigma_u^2$.

- for $k = 0$:

$$
\begin{aligned}
E_t(s_t|\pi_t) &= \frac{Cov(s_t, \pi_t)}{Var(\pi_t)}\pi_t \\
&= \frac{Cov(s_t, s_{t-1} + \epsilon_t)}{Var(s_t) + \sigma_\epsilon^2}\pi_t \\
&= \rho\frac{Var(s_t)}{Var(s_t) + \sigma_\epsilon^2}\pi_t \\
&= \rho\frac{\sigma_u^2}{\sigma_u^2 + (1-\rho^2)\sigma_\epsilon^2}\pi_t
\end{aligned}
$$

So:

$$
\begin{aligned}
E_t\left(F_t\pi_{t+1}|\,\pi_t\right) &= (\rho + \lambda\rho\sum_{k\geq 0}\lambda^k\rho^{2k})(1-\lambda)\frac{\sigma_u^2}{\sigma_u^2 + (1-\rho^2)\sigma_\epsilon^2}\pi_t \\
&= (1-\lambda)\rho(1 + \frac{\lambda}{1-\lambda\rho^2})\frac{\sigma_u^2}{\sigma_u^2 + (1-\rho^2)\sigma_\epsilon^2}\pi_t
\end{aligned}
$$

The second prediction follows directly from:

$$
\begin{aligned}
E_t\left(\pi_{t+1}|\,\pi_t\right) &= E(s_t|\pi_t) \\
&= \frac{Cov(s_t, \pi_t)}{Var(\pi_t)}\pi_t \\
&= \frac{Cov(s_t, s_{t-1})}{Var(\pi_t)}\pi_t \\
&= \rho\frac{\sigma_u^2}{\sigma_u^2 + (1-\rho^2)\sigma_\epsilon^2}\pi_t
\end{aligned}
$$

C Proof of Prediction 3

We know that prices and returns are given by the following formulas:

$$P_t = m.\sum_{k\geq 0} (\lambda\rho)^k s_{t-k}$$

$$R_{t+1} = ms_{t+1} + s_t + \epsilon_{t+1} - zm\sum_{k\geq 0} (\lambda\rho)^k s_{t-k}$$

where $m = \frac{1-\lambda}{1+r-\rho}$ and $z = 1 + r - \rho\lambda$. It is useful to note that $zm = (1-\lambda)(1+m\rho)$ and replace it in the above expression.

Note $a_k = cov\,(R_{t+1}, s_{t-k})$. After some tedious algebra, we can prove that:

$$a_k = (1+m\rho)\frac{\lambda\sigma_u^2}{1-\lambda\rho^2}(\lambda\rho)^k$$

A. *Profitability Anomaly*

$$\begin{aligned} cov(R_{t+1}, \pi_t) &= cov(R_{t+1}, s_{t-1}) \\ &= a_1 \\ &= \sigma_s^2\left[m\rho^2 + \rho - (1-\lambda)(1+m\rho)\left(\rho + \frac{\lambda\rho}{1-\lambda\rho^2}\right)\right] \\ &= (1+m\rho)\lambda\rho\sigma_s^2\left(1 - \frac{1-\lambda}{1-\lambda\rho^2}\right) \end{aligned}$$

And we conclude by using

$$\sigma_s^2 = \frac{\sigma_u^2}{1-\rho^2}.$$

B. *Earnings momentum*

We need to compute $cov(R_{t+1}, \Delta\pi_t)$. Quite simply:

$$cov(R_{t+1}, \Delta\pi_t) = a_1 - a_2$$

Thus:

$$cov(R_{t+1}, \Delta\pi_t) = (1+m\rho)(1-\lambda\rho)\frac{\lambda^2\rho\sigma_u^2}{1-\lambda^2\rho^2}$$

C. *Momentum*

The covariance between consecutive returns is given by:

$$cov(R_{t+1}, R_t) = ma_0 + a_1 - zm\sum_{k\geq 0} (\lambda\rho)^k a_{k+1}$$

We inject the values of the a's coefficients into the above equation, and obtain:

$$cov(R_{t+1}, R_t) = (1+m\rho)(m+\rho\lambda^2)\frac{\lambda\sigma_u^2}{1-\lambda\rho^2}$$

which immediately shows that momentum is positive as soon as $\lambda > 0$.