The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXXIV, NO. 2 • APRIL 2019

Sticky Expectations and the Profitability Anomaly

JEAN-PHILIPPE BOUCHAUD, PHILIPP KRÜGER, AUGUSTIN LANDIER, and DAVID THESMAR*

ABSTRACT

We propose a theory of 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 model predictions: (1) analysts are on average too pessimistic regarding the future profits of high-profit firms, (2) the profitability anomaly is stronger for stocks that are followed by stickier analysts, and (3) the profitability anomaly is stronger for stocks with more persistent profits.

A NUMBER OF STUDIES IN THE asset pricing literature document the existence of stock return predictability: Several stock-level characteristics beyond market betas significantly predict future stock returns. However, the origin of such abnormal returns, and how they can exist in equilibria without being arbitraged away, is subject to debate. One strand of the literature interprets abnormal returns as risk premia (see, for instance, Cochrane (2011)), which implies that they are only seemingly abnormal. Other authors attribute abnormal returns to behavioral biases combined with limits to arbitrage (see, e.g., Barberis and Thaler (2003) or references therein, such as Daniel, Hirshleifer, and Subrahmanyam (1998, 2001) and 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 nonrational market participants typically take the form of non-Bayesian expectations grounded in the psychology literature (see, e.g., Hong and Stein (1999) or Barberis, Shleifer, and Vishny (1998)).

In this paper, we focus on the profitability anomaly, whereby stocks with high profitability ratios tend to outperform on a risk-adjusted basis (Novy-Marx

*Jean-Philippe Bouchaud is with Capital Fund Management. Philipp Krüger is with the University of Geneva and Swiss Finance Institute. Augustin Landier is with HEC Paris. David Thesmar is with MIT Sloan and CEPR. We thank Nick Barberis, Bruno Biais, and Eric So, as well as the two referees and Stefan Nagel for their very detailed and constructive feedback. We are also grateful to seminar audiences at AFA, UC Berkeley, and NBER. Disclosure statement: Bouchaud is chairman of Capital Fund Management. Krüger has nothing to disclose. Landier and Thesmar were doing research at Capital Fund Management when they started this paper.

DOI: 10.1111/jofi.12734

(2013, 2015)). Recent research suggests that profitability is one of the stock return anomalies that has the largest economic significance. The corresponding long-short arbitrage strategy features high Sharpe ratios, no crash risk (Lemperiere et al. (2017)), 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, Simon, and Thesmar (2015)). Our goal in this paper is to test whether 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 Coibion and Gorodnichenko (2015) show, this model has the advantage of nesting rational expectations as a particular case and offers a simple way to measure expectation stickiness using the link between forecast errors and past forecast revisions. It can thus be easily applied 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. The model therefore 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. Directly observable expectations contained in financial analysts' EPS forecasts is a natural setting to study how beliefs of market participants deviate from rational expectations. Analysts are professional forecasters whose forecasts are not cheap talk, which mitigates the legitimate concerns about subjective answers found in surveys (see Bertrand and Mullainathan (2001)). However, we do 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 forecasts.

The data are consistent with key cross-sectional predictions of the model. First, we expect analysts to systematically underestimate future profits when current profits are high. Second, we expect the profitability anomaly 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. Fourth, the previous three predictions should also hold for two significant signals besides the profitability level, namely, earnings momentum (profit change) and returns momentum (past returns). All of these predictions are robust outcomes of the model, and they all hold in the data. They thus support our interpretation of this anomaly.

Our analysis's greatest contribution is to the behavioral finance literature, which documents both under- and overreaction of analyst forecasts. There is an old tradition of papers on investor underreaction. Abarbanell and Bernard (1992) find that analysts underreact to past earnings, in line with our own results. Ali, Klein, and Rosenfeld (1992) find a similar result for annual earnings forecasts. Like us, such positive serial correlation is most often interpreted in the literature as a sign that analysts are underreacting in a non-Bayesian manner when setting expectations of future earnings (see, e.g., Ali, Klein, and Rosenfeld (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 forecast errors is compatible with Bayesian updating if analysts do not know the true data-generating process for earnings but rather learn about it slowly. Consistent with this hypothesis, Mikhail, Walther, and Willis (2003) find that analysts with more experience underreact 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 regressions 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 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 document the slow diffusion of information in markets (see, e.g., Hong, Lim, and Stein (2000), Hou (2007)).

The literature also provides abundant evidence of overreaction. For instance, Debondt and Thaler (1990) document patterns of overreaction by looking at analyst revisions. Most relevant to our present work is a group of papers starting with Debondt and Thaler (1985) and Lakonishok, Shleifer, and Vishny (1994) that seek to explain the value premium by extrapolating beliefs. La Porta (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 be glamour stocks and to have low expected returns. Alti and Tetlock (2014) calibrate a model in which overreaction and overconfidence distort agents' expectations of firm productivity. Weber (2018) 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, Ma, and Shleifer (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 the time sequence of forecasts and on expectations of long-term growth and returns. These papers differ from ours in two respects. First, they seek to explain 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 that link analyst forecast errors with well-known signals that predict returns. Brav, Lehavy, and Michaely (2005) find that systematic expectation errors are consistent with a large number of signals used to forecast returns, but they do not attempt to put economic structure on expectation dynamics. Also, Engelberg, McLean, and Pontiff (2016) document that predictable returns in various anomalies are concentrated around earnings announcements and days on which significant news is revealed. Such a prediction is consistent with our setup, but we do not explore this avenue in our paper.

In terms of theoretical asset pricing models, an important strand of the behavioral literature focuses on explaining the value, momentum, and postearnings announcement drift anomalies. Most related to our work are papers that propose non-Bayesian theories of belief dynamics that can explain these anomalies. Barberis, Shleifer, and Vishny (1998) propose a model in which 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 in which two types of traders coexist: (1) traders who trade on news and (2) trend followers. The interaction between these two sets of 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 setup with only one type of risk-neutral agent. We directly measure the stickiness of analysts' beliefs 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 investors' beliefs.

In its methodology, our paper is also related to the recent macroeconomic literature on expectation formation. The model of expectations dynamics that we use is analyzed by Coibion and Gorodnichenko (2012), who originally apply it to professional inflation forecasts. In Mankiw and Reis (2002), agents also update beliefs infrequently due to fixed costs, which leads in turn to sticky prices.

The rest of the paper is organized as follows. Section I 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 stickiness of analysts' beliefs, and the dynamics of their forecast errors. Section II describes the data. Section III presents 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. Second, we test structural predictions of the model. Section IV use Monte Carlo simulations to test the robustness of our results. Finally, Section V concludes.

I. Model

A. Expectation Stickiness

We start by analyzing a model with expectation dynamics that 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 macroeconomic literature on information rigidity (see Mankiw and Reis (2002) or Reis (2006)). We use the notation of Coibion and Gorodnichenko (2012, 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 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. The term $E_t\pi_{t+h}$ denotes the 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 underreaction ($0 < \lambda < 1$) and overreaction ($\lambda < 0$) (shown, for instance, by Greenwood and Shleifer (2014) and Gennaioli, Ma, and Shleifer (2015)). When applied to consensus forecasts, this structure of forecasts can be made consistent with models of Bayesian learning of private information (Coibion and Gorodnichenko (2015)). When applied to individual-level forecasts, however, it can only come from non-Bayesian underreaction (below, we show that the data favor this type of explanation).

As noted by Coibion and Gorodnichenko (2012, 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:

PROPOSITION 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:

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

(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 the Appendix. ■

These two relations can be readily 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 that is 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 into forecasts, so that 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,

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

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 could view s_t as 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, assumption (4), assumption (5), and normality require that profits follow an ARMA(1,1) process.

The definition of expectation (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 signal informativeness, we know that $E_{t-k}\pi_{t+1} = \rho^k s_{t-k}$, so that forecasts should be written as follows:

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

The econometrician does not observe the signal s_t, but observes profits π_t. Thus, 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.

PROPOSITION 2 (Past profits predict future forecast errors): *Assuming that 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:*

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

PROOF: See the Appendix. ■

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, its conditional expectation 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 to adjust their beliefs, they underestimate this persistence, which leads to predictable forecast errors. The term $\frac{\sigma_u^2}{\sigma_u^2+(1-\rho^2)\sigma_\epsilon^2}$ 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 a large and 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-bones asset pricing model in which we assume that all investors are risk-neutral and have the same expectation stickiness parameter λ. This is an extreme assumption that is designed to focus on our key effects. A natural extension would be a limit of arbitrage model in which rational, risk-averse arbitrageurs trade against the sticky investors. Our qualitative predictions would carry through in such a setup, 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 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 expectations are sticky in the sense of equation (1), prices and returns are functions of past signals:*

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

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

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

To interpret the first formula, note that $P_t^\star = \frac{1}{1+r-\rho}s_t$, which is the price that prevails when $\lambda = 0$, that is, 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 comes directly from the definition of returns. This equation confirms that past signals predict returns, provided that $\lambda \neq 0$. 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. In this case, high returns may arise from temporary profit shocks ϵ_{t+1} as well as innovation in 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.

PROPOSITION 3 (Belief stickiness and stock-market anomalies): *When agents are risk-neutral and expectations are sticky in the sense of equation (1), then at the steady state, noting that $m = \frac{1-\lambda}{1+r-\rho}$:*

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

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

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

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

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

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

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

PROOF: See the Appendix. ■

That items 1 through 3 of Proposition 3 hold in the data has been shown in the extensive empirical literature on asset pricing. Novy-Marx (2013) shows that the Sharpe ratio of the profitability anomaly is high, while Landier, Simon, and Thesmar (2015) document that it is indeed a large anomaly, in the sense that large amounts can be invested in it without being eaten up 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 formulas in items (1), (2), and (3) are consistent with the results derived in the formation of profit expectations. This is because past profit, profit change, and past returns contain information about future profits that has not been fully incorporated into current prices. We notice two interesting properties. First, if expectations are rational ($\lambda = 0$), then neither past profits (levels or changes) nor past returns can forecast future returns. Second, sticky expectations have the power to explain the profitability anomaly *if and only if* the signal is persistent. This ties again to the intuition that slow updating is not a major source of mispricing when recent news is 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 build on existing literature on the profitability anomaly by testing 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 λ reinforces the anomaly: intuitively, stickier beliefs reinforce the relationship between past profits, change in profits, or returns and future stock returns. Second, the proposition also shows that signal persistence (higher ρ) increases the strength of these anomalies. This comes from the above-mentioned fact that higher persistence makes slow expectations a larger source of mistakes about the future. This is because a current signal about future profit has a larger impact on actual value when persistence is higher. The scope for underreaction is therefore higher.

II. Data

A. 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.

We use these detailed analyst-by-analyst forecasts to calculate the firm-level consensus EPS forecast. In other words, we do not use the consensus forecast from the I/B/E/S Summary History File as it is not known how I/B/E/S decides whether to include an individual analyst-level forecast in its calculation of the consensus. In particular, the I/B/E/S consensus could contain stale

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

information, which we want to avoid using. To compute the forecasts for one-, two-, and three-year-ahead earnings for 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 use an upper limit of 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. Using the relatively short period of 45 days also maximizes the scope for forecast errors and biases, and it ensures that as little material information for year t as possible has been released. 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. However, since this does not happen very often (less than 2% of the cases), our consensus is based mainly 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, Malloy, and Scherbina (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 different numbers of shares outstanding. To address 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 based on the EPS forecast sample.

This data set is an annual panel of firms. It has about 54,000 observations for most variables, and some 16,000 when we require the presence of three-year-ahead forecasts (which we use in one specification). We use it to investigate the determinants of forecast errors (Propositions 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 (Proposition 3).

B. Stock Returns

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

[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.

Table I

Summary Statistics

This table reports 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 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-over-year change in the operating cash flows to assets ratio. $mom_{f,t}$ is the usual momentum signal, that is, 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
$(\pi_{f,t} - F_{t-1}\pi_{f,t})/P_{f,t-2}$	54,090	−0.006	0.028	−0.130	−0.014	−0.001	0.005	0.126
$(\pi_{f,t} - F_{t-2}\pi_{f,t})/P_{f,t-2}$	54,062	−0.015	0.044	−0.225	−0.032	−0.007	0.005	0.207
$(F_{t-1}\pi_{f,t} - F_{t-2}\pi_{f,t})/P_{f,t-2}$	54,090	−0.009	0.029	−0.134	−0.020	−0.004	0.004	0.126
$(F_{t-2}\pi_{f,t} - F_{t-3}\pi_{f,t})/P_{f,t-3}$	15,632	−0.006	0.031	−0.145	−0.017	−0.003	0.006	0.138
$(\pi_{f,t-1} - \pi_{f,t-2})/P_{f,t-2}$	45,545	0.002	0.034	−0.149	−0.010	0.004	0.014	0.157
$(\pi_{f,t-2} - \pi_{f,t-3})/P_{f,t-3}$	39,272	0.003	0.034	−0.150	−0.009	0.005	0.015	0.157
$cf_{f,t}$	51,710	0.079	0.107	−0.599	0.035	0.082	0.132	0.699
$\Delta cf_{f,t}$	51,038	−0.001	0.072	−0.381	−0.029	−0.001	0.027	0.381
$mom_{f,t}$	33,636	0.123	0.431	−0.991	−0.131	0.088	0.316	2.567

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 flow 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, Frazzini, and Pedersen (2014), Landier, Simon, and Thesmar (2015)). One possible explanation is that cash flow is a better measure of a firm's fundamental value, consistent with the idea that the difference between cash flow and earnings predicts returns (Sloan (1996)).
(2) Δ Cash flow (Δ*cf*) 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 that accounting data are available *after* the 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 be 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. To form our portfolios, we restrict ourselves to the 3,000 largest stocks. As is standard in the literature, we measure size as stock market capitalization in the previous June and we use ranks that are calculated in each month. We also exclude penny stocks by requiring, at portfolio formation, that the previous month's closing price exceed \$5. We then compute the returns of equally weighted portfolios for each of the five quintile portfolios, as well as the long-short Q5 − Q1 portfolio. In Panel A, we report 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.

As previous literature shows, the three signals do indeed forecast returns, and predictability is robust to risk adjustment. In Panel D, the monthly four-factor alpha of the long-short portfolio is 70 bps using the cash flow signal (t-statistic 3.6), and 18 bps using the change in cash flow signal (t-statistic 2.9). In Panel C, the three-factor alpha (thus excluding the momentum factor control) on momentum returns is 123 bps monthly (t-statistic 3.7).

Table II

Profitability and Momentum in the I/B/E/S Sample

This table reports excess returns (Panel A), CAPM (Panel B), Fama and French (1993) three-factor (Panel C), and Carhart (1997) four-factor (Panel D) alphas 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 the cumulative firm-level return between months $t-12$ and $t-2$. The cash flow 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 is announced. 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 are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

	Q1 (1)	Q2 (2)	Q3 (3)	Q4 (4)	Q5 (5)	Q5–Q1 (6)
			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)

III. Earnings Forecasts and Sticky Beliefs: Testing the Model

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

A. Proposition 1: Measuring Stickiness

A.1. Pooled Analysis

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

To implement this test, we calculate the forecast revision, which we define as the change in the consensus earnings forecast 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 by $P_{f,t-2}$. The forecast revision for firm f'2 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 $(\pi_{f,t} - F_{t-1}\pi_{f,t})/P_{f,t-2}$.

Before running regressions, we first provide a graphical illustration of the data. In Figure 1, we show the forecast error as a function of forecast revisions. We sort all observations into 20 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 the average forecast error $(\pi_{f,t} - F_{t-1}\pi_{f,t})/P_{f,t-2}$ and the average forecast revision for each of the 20 ordered bins. The figure shows a strong monotonic relationship between the forecast error and the revision. We next run the statistical analysis and estimate the following regression, where the time unit t is the fiscal year:

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

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 increase, extrapolators are on average optimistic, that is, their forecast error $\pi_{f,t} - F_{t-1}\pi_{f,t}$ should be negative.

We report the regression results in Table III. In column (1) of Panel A, we directly estimate equation (8), setting $c = 0$. We find that $b = 0.165$, which means

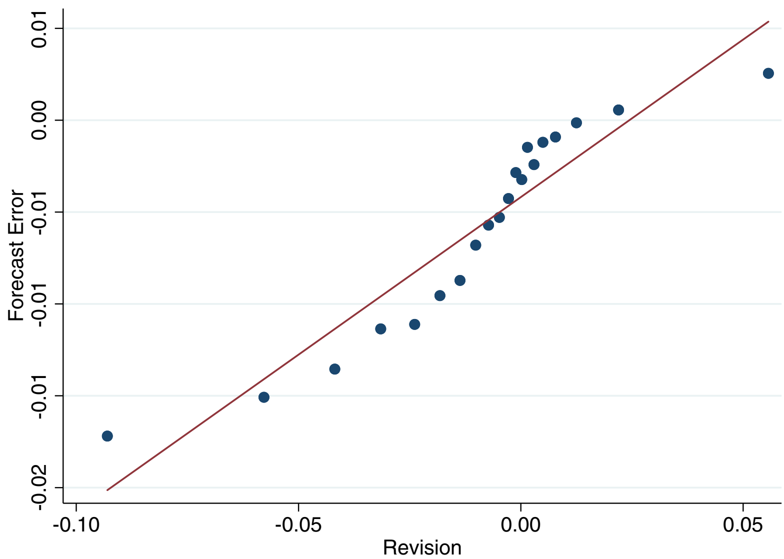


Figure 1. Forecast errors and forecast revisions. This figure shows the forecast errors as a function of forecast revisions. We sort observations into 20 bins of forecast revision $(F_{t-1}\pi_{f,t} - F_{t-2}\pi_{f,t})/P_{f,t-2}$ and calculate the average forecast error (defined as $(\pi_{f,t} - F_{t-1}\pi_{f,t})/P_{f,t-2}$) and the average forecast revision for each of the 20 ordered bins. (Color figure can be viewed at wileyonlinelibrary.com)

$\lambda = 0.14$. These results suggest that, at the quarterly frequency, the weight of lagged forecasts is given by $0.14^{\frac{1}{4}} = 0.6$, which is very similar to what Coibion and Gorodnichenko (2015) find for quarterly revisions of inflation forecasts (they find $\lambda \approx 0.55$). Hence, our estimation of stickiness is in the ballpark of recent estimates coming from macroeconomic forecasts made by independent forecasters and not by security analysts. In column (2), we separately include the two components of the revision, and find that their absolute values do not differ much, which is reassuring. 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 exists ($c < 0$) but is insignificant. As a result, controlling for extrapolation marginally increases the stickiness coefficient, but not significantly so.

In Panel B of Table III, we use an alternative strategy to estimate λ, that is based on the dynamics of forecasts revisions (equation (3) in Proposition 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 in theory—if the expectation model (1) is true—b is equal to λ.

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 $t-2$, that is, $(F_{t-2}\pi_{f,t} - F_{t-3}\pi_{f,t})/P_{f,t-3}$. Standard errors are double-clustered at the firm-year level. t-statistics are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

	(1)	(2)	(3)
Panel A: Dependent Variable: $(\pi_{f,t} - F_{t-1}\pi_{f,t})/P_{f,t-2}$			
$(F_{t-1}\pi_{f,t} - F_{t-2}\pi_{f,t})/P_{f,t-2}$	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: $(F_{t-1}\pi_{f,t} - F_{t-2}\pi_{f,t})/P_{f,t-3}$			
$(F_{t-2}\pi_{f,t} - F_{t-3}\pi_{f,t})/P_{f,t-3}$	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

When testing this prediction, we have to rely on analysts' EPS forecasts for three fiscal years ahead, which decreases our sample size substantially: We keep only about one-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 of 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, however, is 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 Levels

In this section, we extend the methodology used in the previous subsection 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, analysts who follow more industries should 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 (respectively, 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- (firm-) level stickiness to observable analyst (firm) characteristics.

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

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

Next, using the relation $\lambda_a = b_a/(1 + b_a)$ implied by the model, we can back out the analyst-level stickiness using the regression coefficient b_a from the above equation. Panel A of Table IV presents 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 valid only at the consensus level. At the forecaster level, forecast errors are unpredictable. This is because, in their paper, they consider two models of expectation formation at the individual level that 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.

In total we are able to estimate 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 of Table III. The mean analyst-level stickiness λ_a is estimated using approximately 23 observations (mean $N_{\lambda_a} =$ 22.96). Note also that more than 25% of analysts have a negative λ_a, that

Table IV

Summary Statistics (λ and ρ at the Analyst and Firm Levels)

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 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(Two-Digit Standard Industrial Classification) 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 cash flow 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 an 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	6,938	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}	6,885	22.96	27.68	2.00	2.00	2.00	4.00	11.00	31.00	62.00	85.00	151.00
Experience	6,938	6.25	4.38	0.00	1.29	1.71	2.86	5.09	8.48	12.89	15.52	20.18

(Continued)

Table IV—*Continued*

Panel A: Analyst-Level												
	Count	Mean	*SD*	Min	p5	p10	p25	p50	p75	p90	p95	Max
Firm experience	6,908	2.25	1.37	0.00	0.67	0.82	1.19	1.91	3.00	4.18	4.97	7.64
Industry experience	6,799	4.44	2.97	0.00	1.03	1.32	2.09	3.70	6.07	8.81	10.56	14.28
Covered industries	6,887	3.19	1.97	1.00	1.00	1.00	1.71	2.72	4.19	5.90	7.16	11.68
Covered firms	6,913	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												
λ_f	5,916	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}	5,916	8.87	6.71	2.00	2.00	2.00	3.00	7.00	13.00	20.00	24.00	26.00
ρ_f	5,916	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}	5,916	13.09	7.46	2.00	3.00	4.00	7.00	11.00	19.00	25.00	26.00	26.00
Firm size	5,899	6.10	1.82	−0.74	3.32	3.85	4.78	5.99	7.29	8.48	9.28	13.34
EPS volatility	5,842	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	5,689	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	5,897	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	5,897	0.07	0.02	0.01	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.13

is, they "overreact" to recent information. This finding is consistent with the results of Coibion and Gorodnichenko (2015) at the consensus level, but not with their interpretation, because in the two-expectation formation models that 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 stems directly from underreaction 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). We again 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, which is estimated using nine years of data. Again, the stickiness parameter estimated at the firm level is quite similar to what we obtain 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 nonsignificant.

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

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

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

In Panel A, we report results on the determinants of analyst-level stickiness. Stickiness tends to decrease with the analyst's years of experience (columns (1) to (3)), but this result is insignificant after we control for the number of firms and industries covered by the analyst. In columns (4) to (6), results on the number of firms followed and number of industries covered are loosely consistent with a bounded rationality explanation whereby analysts can form better forecasts by specializing on fewer sectors or by extracting the industry component of profitability. Assume, for instance, that the persistent signal

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 entire 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 are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Dependent Variable λ_a (Analyst Level)						
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)						
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

in our model is industry-specific while the shock ϵ is firm-specific. The job of analysts thus consists of extracting the industry-specific signal. While analysts are sticky by default, they can undo their bias by following more firms in a given sector, which improves their ability to observe the industry specific signal, or by following fewer industries. This story helps rationalize the results in columns (4) to (6). Increasing the number of firms for a given number of industries reduces stickiness: it is easier to learn about the industry signal. In contrast,

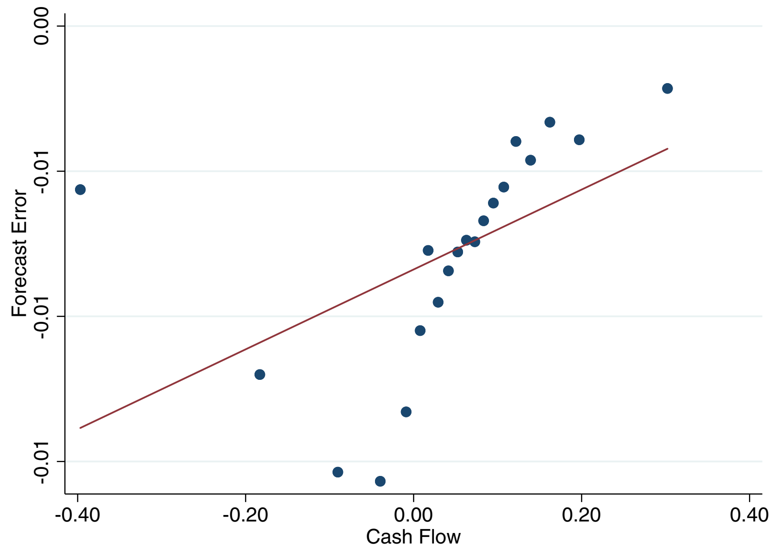


Figure 2. Forecast error and cash flows. This figure shows forecast error as a function of past cash flows. We sort observations into 20 ordered bins of the previous fiscal year's operating cash flows to assets ratio. For each of the 20 ordered groups, we then calculate both average previous year's cash flows-to-assets ratio and current fiscal year's average forecast error. (Color figure can be viewed at wileyonlinelibrary.com)

increasing the number of industries increases stickiness: not only is it harder to learn about industry signals, but the cognitive burden of following several industries also increases.

In Panel B, we report the results from the firm-level regressions. We 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 finding is loosely consistent with a learning model whereby analysts invest in noisy signals of EPS. If EPS is fundamentally noisy, signals are less informative and analysts update their forecasts less frequently.

B. Proposition 2: Past Profits Predict Forecast Errors

Proposition 2 suggests that if expectations are sticky, past profits should predict forecast errors, that is, 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. In Figure 2, we provide graphical evidence on this theoretical prediction. To do so, we first sort observations into 20 bins of previous fiscal year-end operating cash flows over assets. We then calculate both average previous fiscal year-end operating cash flows over assets and average current forecast error for each of the 20 ordered bins.

Figure 2 shows that the relationship between forecast errors and cash flows is positive, which suggests that, in forming their EPS forecasts, analysts do not sufficiently account for current earnings information as measured by operating cash flows.

To test this relationship more formally, we next regress forecast errors on the cash flow signal cf. Our model predicts that the signals Δcf and *mom* should also predict forecast errors in the same direction. This is because they both contain information about future profits that has not been fully incorporated into the expectations of sticky forecasters. We therefore run the 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. The term $\pi_{f,t}$ denotes the firm's realized EPS, which we normalize using the stock price at the fiscal year-end lagged twice, that is, P_{t-2}. The term $F_{t-h}\pi_{f,t}$ denotes the consensus EPS forecast formed over the 45 days following the announcement of cf_{t-h}. We allow error terms to be correlated over 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 zero mean conditional on the 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 Proposition 3, we provide a structural interpretation of the coefficient b_{t-h}.

We allow for a nonzero constant a, which captures the fact that expectations may have a constant positive bias as found in prior literature (see, e.g., Hong and Kacperczyk (2010), Guedj and Bouchaud (2005), or Hong and Kubik (2003)). In other words, we do not analyze the average positive bias of analysts in this paper, but rather the cross-section of their bias conditional on firm characteristics and the dynamics of their bias over time. Results from regressions of the type in equation (14) are reported in Table VI.

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 nonrational and that analysts tend to underreact to some persistent signals that predict future profits. One possible interpretation is to simply view past signals cf, Δcf, and *mom* as measures of the signal itself. But our model is more general, in that it does not require that cash flows or returns be the only neglected signals.

C. Proposition 3: Relating Anomalies to Structural Parameters

C.1. Anomalies Are Stronger for Firms Followed by Sticky Analysts

We now test the link made in Proposition 3 between the stickiness of the analysts covering a firm (λ_f) and the strength of the profitability and momentum

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, that is, $(\pi_t - F_{t-2}\pi_t)/P_{t-2}$. *cf* is Compustat item *oancf* divided by item *at*. Δcf is the year-over-year difference in *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 are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

	(1)	(2)	(3)
Panel A: Dependent Variable $(\pi_{f,t} - F_{t-1}\pi_{f,t})/P_{f,t-2}$			
$cf_{f,t-1}$	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 $(\pi_{f,t} - F_{t-2}\pi_{f,t})/P_{f,t-2}$			
$cf_{f,t-2}$	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

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 a direct test of our theory because it 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 is plausible that the marginal investor anchors her beliefs to some extent on analyst forecasts.

To test the prediction that the strength of the profitability and momentum anomalies depend 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 terciles is -0.23, 0.13, and

0.41, respectively. It thus turns out that firms in the second and third terciles 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 tend to have, by and large, negative values (so that forecasts about their profits are more 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.

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 sort firms first into terciles of the stickiness parameter λ_f and second 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 just use 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 monthly alpha of the long-short cash flow strategy is equal to 102 bps for the stickiest stocks (t-statistic 4.9). In contrast, the cash flow alpha for the least sticky stocks is only 51 bps (t-statistic 2.4). The difference between the two is highly significant: the t-statistic of the differential return between most and least sticky long-short portfolios 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. The profitability momentum strategy is not significant for the least sticky stocks, with an alpha of 4 bps (t-statistic 0.5), but very strongly significant for the stickiest ones, with an alpha of 31 bps (t-statistic 3.9). The returns spread is marginally significant with a t-statistic of 2.7. Last, portfolio strategies based on returns momentum are strongly consistent with our prediction. Momentum has a (three-factor) alpha of 1.51 (4.8) for the stickiest stocks, versus 111 bps (t-statistic 3.3), and the spread has a t-statistic of 2.64. In all cases, focusing on sticky stocks significantly boosts the risk-adjusted return of the strategy.

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 primary reason is that when cash flows are highly persistent, slower updating leads to larger mistakes. To test this prediction, we perform portfolio tests similar to those carried out above.

Table VII
Anomalies Sorted by Expectation Stickiness λ_f

Panels A and B of this table report Carhart (1997) four-factor alphas for equally weighted portfolios that are double-sorted on λ_f and the level of (and change in) 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 report 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 flow signal *(cf)* in Panel A, the change in cash flow signal (Δ*cf*) in Panel B, and momentum (*mom*) in Panel C. Standard errors are adjusted for heteroskedasticity and autocorrelations up to 12 lags. t-statistics are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

	Q1 (1)	Q2 (2)	Q3 (3)	Q4 (4)	Q5 (5)	Q5 − Q1 (6)
			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)

First, we measure each firm's cash flow persistence ρ_f. We do so by 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 flow signal.

The median cash flow persistence is about $\rho_f \approx 0.22$ and 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 sort firms first into terciles of ρ_f and second into quintiles of the cash flow, change in cash flow, and momentum signals. The median ρ_f in the first, second, and third terciles is -0.10, 0.28, and 0.62, respectively.

In Panels A and B of Table VIII, we report Carhart (1997) alphas of portfolios double-sorted on ρ and the cash flow–based signals (cf and Δcf). In Panel C, we display Fama and French (1993) alphas for portfolios double-sorted 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 is equal to 27 bps per month (t-statistic of 2.2) for the cash flow signal. Cash flow change yields a monthly alpha spread of 43 bps (t-statistic 3.5). Returns momentum has a monthly alpha spread of 34 bps (t-statistic 2.1). In all cases, focusing on the most persistent stocks significantly boosts the risk-adjusted returns of strategies, with a t-statistic of 4.3 (cash flow), 3.7 (change in cash flow), and 3.8 (returns momentum).

IV. Robustness

A potential concern with our results arises from the fact that we use the entire 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. To focus on reasonably long-term expectations—arguably most susceptible to behavioral biases—and to avoid seasonality concerns, we employ 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 Panel B of Table IV). We thus need the entire time series of forecasts to estimate λ_f with reasonable precision. The downside of this approach, however, is 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 hardwire a correlation between our stickiness parameter and returns.

To address this concern, we use simulations that allow us 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

Table VIII

Anomalies Sorted by Persistence ρ_f

Panels A and B of this table report Carhart (1997) four-factor alphas for equally weighted portfolios that are double-sorted on ρ_f and the level of (and change in) 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 report 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 flow signal *(cf)* in Panel A, the change in cash flow signal (Δ*cf*) in Panel B, and momentum (*mom*) in Panel C. Standard errors are adjusted for heteroskedasticity and autocorrelations up to 12 lags. *t*-statistics are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

	(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)

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 signals and profit:

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

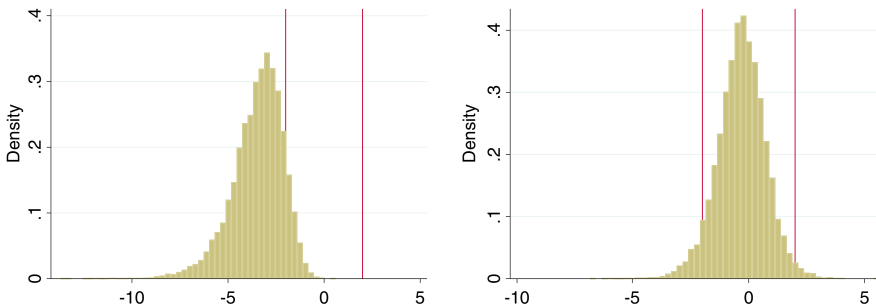
The idea is to then simulate data generated by this model *under the null hypothesis that expectations are rational*, that is, that $\lambda = 0$. Under rational expectations, as shown in Lemma 1, realized dollar returns are given by $R_{t+1} = \epsilon_{t+1} + u_{t+1}/(1 + r - \rho)$. While actual stickiness 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 that 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 $\frac{\widehat{\text{cov}}(\epsilon_{t+1}, u_t)}{\widehat{\text{var}} u_t}$. Even though it is zero on average by design, there may be significant dispersion in the simulated data if the number of years per firm is low. In this setting, we thus ask whether a financial econometrician who estimates 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 as follows. 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 $r = 0.03$. 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 (0.19 and 0.05, respectively, as shown in Panel B of Table IV). To generate a third relation, we impose the condition that the R^2 of the regression of π_{t+1} on s_t is equal to 0.7.[4] For each firm in our sample, we then estimate λ by regressing profit expectation errors ϵ_{t+1} on the expectation update u_t using the entire 20-year period as we do in the paper. We then implement a double-sort on the simulated data that is similar to the procedure that follows for the 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). Then, for each of these quintiles of $\hat{\lambda}$, we compute the realized returns of a long-short

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

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

where R^2 is the explanatory power of s_t on π_{t+1} in a linear regression. This calibration leads to $\sigma_\epsilon = 0.027$, $\sigma_u = 0.022$, and $\rho = 0.27$.



Panel A. Cash flow signal (cf) Panel B. Δ Cash flow signal (Δcf)

Figure 3. *t*-statistics of double-sorts under the null of rational expectations: Results from simulations. These histograms present the distribution of t-statistics from double-sorts by profitability and stickiness for 10,000 simulations under the null hypothesis that expectations are *not* sticky. Each simulation works as follows. 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 the 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 for a given 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 return of a long-short portfolio in which stocks are weighted by their rank in terms of past profitability, which is normalized to range between -0.5 and $+0.5$. 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: $R_t^q = \sum_q \beta_q \mathbf{1}_q + \nu_t$. We retrieve the t-statistic on β_5. We repeat this procedure 100 times and present the histograms of the 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 $\pi_t - \pi_{t-1}$. (Color figure can be viewed at wileyonlinelibrary.com)

portfolio in which stocks are weighted by their rank in terms of each of the two profitability signals, normalized to the range $(-0.5, +0.5)$. 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 five portfolios R_t^q, one per quintile of $\hat{\lambda}$. Let $q = 1, \ldots, 5$ be the index on this quintile. We next regress these returns on $\hat{\lambda}$ quintile dummies, $R_t^q = \sum_{q \geq 2} \beta_q \mathbf{1}_q + \nu_t$ using the first quintile as a reference, and retrieve the t-statistic on β_5. We repeat this procedure 10,000 times. In this model economy, returns are unpredictable and expectations are not sticky. Any significant relationship between $\hat{\lambda}$ and profitability anomaly returns would thus have to come from the look-ahead bias in $\hat{\lambda}$, which we estimate using the entire period and therefore using future expectation errors.

In Figure 3, we present histograms of the resulting t-statistics. In Panel A, we use past profits as the portfolio-sorting variable, and in Panel B, we use past profit changes. In 10,000 simulations, we do not find one single simulation where the t-statistic for cash flows ends up being greater than 2. For the cash flow change signal, the t-statistic is greater than 2 in 1.5% of the simulations. Hence, 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, for both signals, the average t-statistic across simulations is *negative*: -3.3 for the straight cash flow signal, and -0.3 for the cash flow change. The look-ahead bias thus tends to generate a relationship contrary to what we find in the data.

The intuition for this result is similar to the Kendall (1954) analysis of autocorrelation bias in small samples. In our 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 typically happens when the firm has dates when $\epsilon_{f,t+1}$ and $u_{f,t}$ are both above average *and* dates when they are both below average: the coexistence of such observations 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 both $u_{f,t}$ and $\epsilon_{f,t+1}$ to be relatively likely to be negative at future dates $t > T$, and therefore future returns—which are a combination of both effects—to be lower than average. High profit stocks with a high measured $\hat{\lambda}$ are therefore 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 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, across simulations, we expect on average a slightly negative t-statistic for the double-sort, although it is rarely significant.

V. Conclusion

In this paper, we propose a model that predicts that one of the most economically significant stock return anomalies, namely, the profitability anomaly, 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 that are followed by analysts characterized by more sticky expectations or for firms subject to more persistent profits. The theoretical predictions are borne out by the data. When we explore cross-sectional determinants of the expectation stickiness measure that we propose in this paper, we find 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.

Initial submission: October 31, 2016; Accepted: January 30, 2018
Editors: Stefan Nagel, Philip Bond, Amit Seru, and Wei Xiong

Appendix: Proofs

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 plug this back into prices,

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

We can then compute dollar returns as

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

■

Proof of Proposition 2: First notice that $\text{cov}(s_{t-k}, s_t) = \rho^k \text{ var}(s_t)$.
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 stationary random variables centered on zero, we can write the conditional expectations as simple projections.

- For $k > 0$:

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

  because $\text{var}(s_t) = \rho^2\text{var}(s_t) + \sigma_u^2$.
- For $k = 0$:

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

So,

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

The second prediction follows directly from

$$\begin{aligned} E_t\,(\pi_{t+1}|\,\pi_t) &= E(s_t|\pi_t) \\ &= \frac{\text{cov}(s_t, \pi_t)}{\text{var}(\pi_t)}\pi_t \\ &= \frac{\text{cov}(s_t, s_{t-1})}{\text{var}(\pi_t)}\pi_t \\ &= \rho\frac{\sigma_u^2}{\sigma_u^2 + (1-\rho^2)\sigma_\epsilon^2}\pi_t. \end{aligned}$$

■

PROOF OF PROPOSITION 3: We know that prices and returns are given by

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

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

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

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

Profitability Anomaly: Using the above expression, we first compute the covariance between future returns and past profitability.

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

We conclude by using

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

Earnings Momentum: We then compute the covariance between future returns and past profitability change.

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

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

Thus,

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

Momentum: Finally, we compute the covariance between past and future returns.

The covariance between consecutive returns is given by

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

We replace the values of the coefficients on the a's into the above equation, to obtain

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

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

REFERENCES

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

Ali, Ashiq, April Klein, and James Rosenfeld, 1992, Analysts' use of information about permanent and transitory earnings components in forecasting annual EPS, *The Accounting Review* 67, 183–198.

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

Asness, Cliff, Andrea Frazzini, and Lasse Pedersen, 2014, Quality minus junk, SSRN Working Paper.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny, 1998, A model of investor sentiment, *Journal of Financial Economics* 49, 307–343.

Barberis, Nicholas, and Richard Thaler, 2003, A survey of behavioral finance, in G. Constanides, M. Harris, and R.M. Stulz, eds.: *Handbook of the Economics of Finance*, Vol. 1 (Elsevier/North Holland, Amsterdam).

Bertrand, Marianne, and Sendhil Mullainathan, 2001, Do people mean what they say? Implications for subjective survey data, *American Economic Review* 91, 67–72.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer, 2017, Diagnostic expectations and stock returns, NBER Working Paper No. w23863.

Brav, Alon, Reuven Lehavy, and Roni Michaely, 2005, Using expectations to test asset pricing models, *Financial Management* 34, 31–64.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

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

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

Coibion, Olivier, and Yuriy Gorodnichenko, 2015, Information rigidity and the expectations formation process: A simple framework and new facts, *American Economic Review* 105, 2644–2678.

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

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, 2001, Overconfidence, arbitrage, and equilibrium asset pricing, *Journal of Finance* 56, 921–965.

Debondt, Werner, and Richard Thaler, 1985, Does the stock market overreact? *Journal of Finance* 40, 793–805.

Debondt, Werner, and Richard Thaler, 1990, Do security analysts overreact? *American Economic Review* 80, 52–57.

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

Engelberg, Joseph, David McLean, and Jeffrey Pontiff, 2016, Anomalies and news, SSRN Working Paper.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Gennaioli, Nicola, Yueran Ma, and Andrei Shleifer, 2015, Expectations and investment, NBER Working Paper No. w21260.

Greenwood, Robin, and Andrei Shleifer, 2014, Expectations of returns and expected returns, *Review of Financial Studies* 27, 714–746.

Guedj, Olivier, and Jean-Philippe Bouchaud, 2005, Experts' earning forecasts: Bias, herding and gossamer information, *International Journal of Theoretical and Applied Finance* 8, 933–946.

Hirshleifer, David, 2001, Investor psychology and asset pricing, *Journal of Finance* 56, 1533–1597.

Hong, Harrison, and Marcin Kacperczyk, 2010, Competition and bias, *The Quarterly Journal of Economics* 125, 1683–1725.

Hong, Harrison, and Jeffrey D. Kubik, 2003, Analyzing the analysts: Career concerns and biased earnings forecasts, *Journal of Finance* 58, 313–351.

Hong, Harrison, Terence Lim, and Jeremy Stein, 2000, Bad news travel slowly: Size, analyst coverage, and the profitability of momentum strategies, *Journal of Finance* 55, 265–295.

Hong, Harrison, and Jeremy C. Stein, 1999, A unified theory of underreaction, momentum trading, and overreaction in asset markets, *Journal of Finance* 54, 2143–2184.

Hou, Kewei, 2007, Industry information diffusion and the lead-lag effect in stock returns, *Review of Financial Studies* 20, 1113–1138.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

Kendall, Maurice G., 1954, Note on bias in the estimation of autocorrelation, *Biometrika* 41, 403–404.

La Porta, Rafael, 1996, Expectations and the cross-section of stock returns, *Journal of Finance* 51, 1715–1742.

Lakonishok, Josef, Andrei Shleifer, and Robert Vishny, 1994, Contrarian investment, extrapolation and risk, *Journal of Finance* 49, 1541–1578.

Landier, Augustin, Guillaume Simon, and David Thesmar, 2015, The capacity of trading strategies, HEC Paris Research Paper No. FIN-2015-1089.

Lemperiere, Yves, Cyril Deremble, Trung-Tu Nguyen, Philip Seager, and Jean-Philippe Bouchaud, 2017, Risk premia: Asymmetric tail risks and excess returns, *Quantitative Finance* 17, 1–14.

Mankiw, Gregory, and Ricardo Reis, 2002, Sticky information versus sticky prices: A proposal to replace the new Keynesian Philips curve, *The Quarterly Journal of Economics* 117, 1295–1328.

Markov, Stanimir, and Ane Tamayo, 2006, Predictability in financial analyst forecast errors: Learning or irrationality? *Journal of Accounting Research* 44, 725–761.

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

Novy-Marx, Robert, 2013, The other side of value: The gross profitability premium, *Journal of Financial Economics* 108, 1–28.

Novy-Marx, Robert, 2015, Fundamentally, momentum is fundamental momentum, NBER Working Paper No. 20984.

Reis, Ricardo, 2006, Inattentive producers, *Review of Economic Studies* 73, 793–821.

Robinson, David, and Denys Glushkov, 2006, A note on IBES unadjusted data, Wharton Research Data Services Working Paper.

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

Weber, Michael, 2018, Cash flow duration and the term structure of equity returns, Technical Report, NBER. *Journal of Financial Economics*, 128, 486–503.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Replication code