Anomalies and News

JOSEPH ENGELBERG,

R. DAVID MCLEAN,

and

JEFFREY PONTIFF*

ABSTRACT

Using a sample of 97 stock return anomalies, we find that anomaly returns are 50% higher on corporate news days and six times higher on earnings announcement days. These results could be explained by dynamic risk, mispricing due to biased expectations, or data mining. We develop and conduct several unique tests to differentiate between these three explanations. Our results are most consistent with the idea that anomaly returns are driven by biased expectations, which are at least partly corrected upon news arrival.

Academic research shows that a large number of observable firm characteristics can predict the cross-section of stock returns (see Fama (1998), Nagel (2013), and McLean and Pontiff (2016)). This "anomalies" research goes back to at least Ball and Brown (1968) and Blume and Husic (1973), yet more than four decades later, academics still disagree on what causes this return predictability.

There are three popular explanations for cross-sectional predictability. First, return predictability could be a result of cross-sectional differences in risk, which are reflected in discount rates (e.g., Fama (1991, 1998)). Under this explanation, cross-sectional return predictability is expected because return differences reflect ex-ante differences in the discount rates used to value the stocks.

Second, return predictability could reflect mispricing (e.g., Barberis and Thaler (2003)). For example, the marginal investor may have biased expectations about cash flows, and the anomaly variables are correlated with these mistakes across stocks. Under this view, when new

2

information arrives, investors update their beliefs, which in turn corrects prices and leads to return predictability.

Third, return predictability could be due to data mining. As Fama (1998) points out, academics have likely tested thousands of variables, in which case it would not be surprising to find that some of them predict returns in-sample, even if in reality none of them do.[1]

In this paper we differentiate between these three explanations of cross-sectional return predictability by comparing predictability on days when firm-specific information is publicly released to days when we do not observe news. We use the 97 anomaly variables studied in McLean and Pontiff (2016), as each of these variables has been reported to predict the cross-section of stock returns in a published academic study. Days with firm-specific information releases are defined as earnings announcements days or days with a Dow Jones news item.

We find that anomaly returns are 50% higher on corporate news days and six times higher on earnings announcement days.[2] We find similar effects on both the long and short sides, that is, anomaly-shorts have lower returns and anomaly-longs have higher returns on information days. These effects appear to be related to firm-specific news, as anomaly returns are not higher on days with macroeconomic news. The findings are also not explained by a day-of-the week effect or by extreme returns causing news, as anomaly returns are not elevated on

---

[1] Recognition of a "multiple testing bias" in all types of empirical research dates at least back to Bonferroni (1935) and is stressed more recently in the finance literature by Harvey, Lin, and Zhu (2016), McLean and Pontiff (2016), and Linnainmaa and Roberts (2017).

[2] Stock returns are unconditionally higher on earnings announcement days (Franzini and Lamont (2006)). Savor and Wilson (2016) attempt to explain this fact. This is not the effect that we document nor the one we want to explain; our main specifications control for this effect through the use of earnings announcement dummy variables. We find that anomaly-long (anomaly-short) returns are higher (lower) on earnings and news days while controlling for the fact that stock returns are higher on earnings announcement days.

3

extreme return days that do not also have news. We discuss how our results relate to each of the three explanations of cross-sectional return predictability below.

Systematic risk. A standard, static risk-factor model (e.g., Fama and French (1993), and Carhart (1997)) constructs a stock's expected return as a product of its systematic risk exposures ("betas") and their corresponding risk premiums. In these factor models, a firm-specific news event does not change a stock's expected return because it is unrelated to the time-invariant betas and risk premiums. Our finding of predictably higher anomaly returns on information days is at odds with static risk-factor models.

However, our results could be consistent with dynamic-risk models, which allow for time-varying risk premia and time-varying betas. Papers in this spirit include Patton and Verardo (2012), who find that a stock's beta with respect to the market portfolio is higher on earnings announcement days and explain this finding with a dynamic-learning model, and Savor and Wilson (2016), who develop a dynamic risk-based model to explain why stock returns are higher on earnings announcement days.

We consider both time-varying risk premia and time-varying betas as explanations for our asymmetric result that anomaly-longs have higher returns on news days and anomaly-shorts have lower returns on news days. We show that time-varying risk premia cannot explain our result. When we include a variety of fixed effects to absorb any daily variation in risk factors, our results are virtually unchanged.

When we consider time-varying betas, however, we obtain mixed results. Specifically, when we examine the market betas of anomaly stocks, we find no asymmetric effects for anomaly-longs and anomaly-shorts, whereas when we consider time-varying exposure to an aggregate anomaly factor, we do find an asymmetry: anomaly-longs increase their factor betas, and anomaly-shorts decrease their factor betas, on earnings days. This could explain why

4

anomaly-longs have higher returns on news days while anomaly-shorts have lower returns on news days. However, even after controlling for these changing betas on earnings days, we still find that anomaly-longs have higher returns and anomaly-shorts have lower returns on earnings days. The inclusion of the betas has virtually no impact on the magnitude or statistical significance of the earnings-day effect.

It could be the case that our aggregate anomaly factor is not the "right" risk factor, and if we had the "right" factor it would explain our results. However, it is important to note what kind of dynamic risk model would be necessary to generate our findings. Using quintile portfolios, we find that on earnings days, anomaly returns are 8.7 times *higher* for long-side stocks, and 7.53 times *lower* for short-side stocks. If these returns reflect priced risk, then the underlying asset pricing model would require some stocks to *ex-ante* have discount rates that are 8.7 times higher on earnings announcement days, and other stocks to *ex-ante* have discount rates that are 7.53 times lower on earnings announcement days. Then, after the earnings announcements, risk would return back to its pre-announcement level.

Mispricing due to biased expectations. Under the biased expectations explanation, investors are too optimistic about some stocks and too pessimistic about others, and anomaly variables are correlated with these biases. When new information arrives in the form of an earnings announcement or other news story, investors update their beliefs, resulting in a correction to the stock price. To illustrate this intuition, we consider a simple representative agent model (further elaborated in the Internet Appendix[3]) with an agent that has biased expectations about future cash flows that are corrected with the arrival of public cash flow news. The end result is that firms for which the agent has overly optimistic (pessimistic) cash

---

[3] The Internet Appendix is available on the *Journal of Finance* website.

flow expectations have negative (positive) news-day returns. The earnings announcement-day and news-day returns that we document are consistent with this intuition.

To better test the idea that biased expectations explain stock return anomalies, we also study analyst forecast errors. We find that analysts' earnings forecasts are too low for anomaly-longs and too high for anomaly-shorts, that is, analysts' earnings forecasts are too optimistic for anomaly-shorts and too pessimistic for anomaly-longs. This result is consistent with biased expectations as an explanation for stock return anomalies, but not risk, as it is difficult to see how even dynamic betas can explain why analysts' earnings forecasts are biased. However, as we explain below, our result concerning analyst forecast errors does not rule out data mining as an explanation for stock returns anomalies.

The idea that biased expectations can explain stock return anomalies can be traced to Basu (1977), DeBondt and Thaler (1985), and La Porta, et al. (1997), who argue that biased expectations can explain long-term reversal and value strategies. More recent models of stock return anomalies that are based on biased expectations include Barberis, Shleifer, and Vishny (1998) and Daniel, Hirshleifer, and Subrahmanyam (1998, 2001). In these models, long-term reversal and price-to-fundamental anomalies are caused by biased expectations about future cash flows and a price correction that occurs when new information is made public.

Data mining. Our finding that anomaly returns are higher on information days could be consistent with data mining, as could our finding that anomaly variables predict analysts' forecast errors. To see why, consider the fact that stocks with high (low) ex-post returns over a given period are likely to also have had high (low) returns on news days and earnings announcement days during the same period. Put differently, stocks with high (low) monthly returns over the last month probably had good (bad) news during the month, which explains why the returns were high (low) in that month.

6

We show that this intuition is supported empirically: stocks with high (low) monthly returns in month t, regardless of their anomaly portfolio membership, tend to have especially high (low) returns on earnings days and on news days during month t. Similarly, we show that stocks with high (low) monthly returns in month t that also announced earnings in month t had analyst earnings forecasts that turned out to be too low (high) during the same month.

To address the data-mining explanation, we conduct several tests. First, we re-estimate our main daily regression tests while controlling for the contemporaneous monthly stock return and its relation with earnings-day and news-day returns. We find that, even after controlling for monthly returns, anomaly returns are still high on news days and earnings days, and anomaly variables still predict analyst forecast errors. Second, we build an out-of-sample anomaly variable that is constructed only with out-of-sample anomalies. We find that the anomaly returns are significant out-of-sample and higher on earnings days and news days. Moreover, the out-of-sample anomaly variable predicts analyst forecast errors. These findings challenge the idea that data mining alone can explain why anomaly returns are higher on information days.

Previous Literature. Our paper builds on earlier studies that report higher anomaly returns on earnings announcement days (e.g., Bernard and Thomas (1989), Ball and Kothari (1991), Chopra, Lakonishok and Ritter (1992), La Porta et al. (1997), Sloan (1996), Jegadeesh and Titman (1993), and Bradshaw, Richardson, and Sloan (2006)). Variations of the literature include Edelen, Ince, and Kadlec (2016), who show that institutional trading is related to anomaly returns around earnings announcement days, and Wu, Zhang, and Zhang (2010) and Liu and Zhang (2014), who argue that investment-based models imply higher risk premiums on earnings announcement days.

Our paper differs from this literature in several ways. First, we investigate not only earnings announcement days but also more than 6 million news days that do not coincide with

7

Compustat earnings announcements. We use a broad set of 97 anomalies that not only gives us more statistical power than previous studies, but also allows us to draw novel comparisons between categories of anomalies. Our paper is also the first to relate such a broad set of anomalies to analyst forecast error. Our forecast error results are important because they are not subject to the joint-hypothesis problem and are in agreement with our daily stock return findings.

In addition, previous studies do not consider whether data mining can generate higher anomaly returns on announcement days. Our results suggest that spurious anomaly strategies will most likely also have higher returns on news days and earnings announcement days. This finding means that previous studies that relate earnings announcements to anomaly returns do not address Fama's (1998) data-mining conjecture. We test Fama's (1998) conjecture by developing a series of data-mining analyses that allow us to rule out the possibility that our results are driven entirely by data mining.

Finally, our study links the stock return anomaly literature to a literature on dynamic risk. To the best of our knowledge, we are the first to make this connection. As we mention above, the dynamic risk frameworks of Patton and Verardo (2012) and Savor and Wilson (2016) can explain why stock returns are higher on earnings announcement days, but these papers do not attempt to explain stock return anomalies. Our findings suggest that these frameworks could be useful to researchers that want to explain anomalies with risk.

The paper is organized as follows. In Section I we introduce our data. In Section II we document stylized facts regarding cross-sectional returns on information days. Section III considers whether risk can explain differences in return predictability between information and noninformation days. Section IV examines the role of biased expectations. Section V considers data mining explanations. Section VI concludes.

8

I. Sample and Data

We begin our sample with 97 cross-sectional anomalies studied in McLean and Pontiff (2016) that are drawn from 80 studies published in peer-reviewed finance, accounting, and economics journals. Each of the anomaly variables has been reported to predict the cross-section of stock returns, and can be constructed with data from CRSP, Compustat, or IBES.

To create the anomaly portfolios, stocks are sorted each month on each of the anomaly characteristics. We define the extreme quintiles as the long and the short sides of each anomaly strategy. Sixteen of our 97 anomalies are indicator variables (e.g., credit rating downgrades); in these cases, there is only a long or a short side, based on the binary value of the indicator. We reconstruct the anomaly portfolios each month. As in McLean and Pontiff (2016), the sample selection for each anomaly follows the original study. Thus, if a study uses only NYSE firms, we create that anomaly variable only for NYSE firms.

We obtain earnings announcement dates from the Compustat quarterly database. Compustat reports the day or the earnings announcement but not the time. Many firms report earnings after the market closes. In these cases, the information will be reflected in the stock return on the following day (CRSP returns are from close to close). We therefore examine the firm's trading volume scaled by market trading volume for the day before, the day of, and the day after the reported earnings announcement date. We define the day with the highest volume as the earnings announcement day.

We obtain news story dates from the Dow Jones news archive. Dow Jones reports both the date and time of its news stories. This archive contains all news stories from Dow Jones newswire and all *Wall Street Journal* stories for the period 1979:06 to 2013. These news data

9

are also used in Tetlock (2010, 2011) and Engelberg, Reed, and Ringgenberg (2012), who report the frequency of news categories in the archive. Popular categories include mergers and acquisitions, earnings news and projections, analysts' comments and rankings, insider buying and selling, personnel appointments, and dividend news.

We merge these news data and the earnings announcement data with daily stock return data, so we can test whether anomaly returns are higher on information days as compared to non-information days. For consistency, we conduct all of our tests during the period 1979:06 to 2013, which is the period for which we have news data. We exclude stocks with prices under $5. These low-priced stocks are excluded from many of the anomaly portfolios to begin with, and low-priced stocks are less likely to have news or earnings announcement data.

Sample Descriptive Statistics. Table I provides descriptive statistics for our sample, which consists of 45,975,693 firm-day observations for the period 1979:06 to 2013. Each observation is in the CRSP daily return database with reported stock returns and a stock price greater than $5 at the end of the previous trading day. Among these observations, 14.5% have Dow Jones news stories, while 1.1% have earnings announcements reported in Compustat.

(Table I goes here)

There is overlap between the news days and the earnings announcement days. Of the 509,720 earnings announcement days, 235,444, or 46%, are also Dow Jones news days. However, this is a small percentage of the total news days. The total number of news days is 6,629,300, so only 3.6% of these are also earnings announcements that are reported in Compustat. To the extent that Dow Jones stories cover a significant number of earnings announcements not covered in Compustat, 3.6% is a lower bound on the percentage of news stories that likely reflect earnings announcements.

10

Table II provides descriptive statistics for the anomaly variables. Our primary anomaly variable is *Net*. To construct *Net* for each firm-month observation, we sum the number of long-side (*Long*) and short-side (*Short*) anomaly portfolios that the observation belongs to. Recall that we form long and short portfolios by placing stocks into quintiles based on monthly rankings of each anomaly variable. *Net* is the difference between *Long* and *Short*, that is, *Net = Long – Short*. Table II shows that the average stock is in 4.52 long portfolios and 5.52 short portfolios. If the portfolios were based solely on 97 random quintile groupings, we would expect both long and short to equal 19.4 (97 x 0.20). Our counts are lower since some characteristics are indicator variables, and thus, lack either a long or a short side. Also, following the original study, some variables are constructed only for a subset of stocks (for example, NYSE stocks). For anomaly variables that are subset based, stocks that fall out of the subset are not assigned to a long or short side. For more information on the construction of the anomaly variables, see the Internet Appendix of McLean and Pontiff (2016). The mean value for *Net* is -0.71, the maximum value is 33, and the minimum value is -38.

(Table II goes here)

II. The Stylized Facts: Anomaly Returns on Information Days

A. Anomaly Returns on Information Days versus Noninformation Days

In this section we report our main findings. In our first set of tests, we estimate the following regression equation:

11

$$R_{i,t} = \alpha_t + \beta_1 Net_{i,t} + \beta_2 Net_{i,t} \times Eday_{i,t} + \beta_3 Net_{i,t} \times Nday_{i,t} + \beta_4 Eday_{i,t} + \beta_5 Nday_{i,t}$$

$$+ \sum_{j=1}^{10} \gamma_j Lag\ Return_{i,t-j} + \sum_{j=1}^{10} \delta_j Lag\ Return^2_{i,t-j} + \sum_{j=1}^{10} \rho_j Volume_{i,t-j}$$

$$+ \varepsilon_{i,t} \qquad\qquad\qquad (1)$$

The regression includes day fixed effects (α_t). In the above equation, $R_{i,t}$ is the daily return of stock i on day t, and $Net_{i,t}$ is our aggregate anomaly variable, which we describe in more detail below. We measure *Net* at the beginning of each month and stock returns on each day throughout the month. Thus, although news such as earnings announcements may affect future values of *Net* for a given stock, the value of *Net* that we use in our regressions remains the same throughout a month. The variable *Eday* (*Nday*) is a dummy variable equal to one on earnings (news) days for firm i and zero otherwise. Interaction terms indicate whether anomaly returns are higher on information days.

We include as controls lagged values of over the last 10 days of returns, volatility (return squared), and volume. For brevity, we do not report these coefficients. Given the large number of observations used in most of our regressions, the inclusion of the 30 lagged values enables us to compare saturated and nonsaturated regressions to assess the robustness of our results. When we run specifications without these controls, we find that the results are virtually identical. This comparison gives us more confidence in the robustness of our findings.

Table III reports the regression results. As mentioned above, returns are expressed in basis points. In Panel A, we define the information day as a one-day window, while in Panel B we use a three-day window, that is, days *t*-1, *t*, and *t+*1. To facilitate interpretation, we multiply returns by 10,000, so that each unit of return is equal to one basis point.

(Table III goes here)

12

The first regression does not include the lagged volume, lagged return, and lagged squared return controls. Since our estimation uses millions of observations, omission and inclusion of correlated variables may cause changes in statistical significance. A comparison of the first two regressions shows that this is not the case. Instead, the controls absorb variation, in the sense that the standard errors in the second specification, which includes the lagged controls, shrink slightly but the slope coefficients remain similar.

In the second regression in Panel A, the coefficient on *Net* is 0.384, while that on *Net x Earnings Announcement* is 2.164. The coefficients show that for a *Net* value of 10 (about 1½ standard deviations), expected returns are 3.84 basis points higher on non-earnings announcement days and 21.64 basis points higher on earnings announcement days. Put differently, anomaly returns for a *Net* value of 10 are 25.48 on earnings announcement days, which is 6.3 times higher than anomaly returns on non-earnings announcement days.

The coefficient on *Net x News Day* is 0.178, which shows that anomaly returns are about 50% higher on news days. For example, a stock with a *Net* value of 10 has an expected return that is 3.84 basis points higher on non-news days, compared with 5.62 basis points higher on news days.

In the third regression reported in Panel A, we replace the day fixed effects with day-information event fixed effects. This specification has four separate intercepts. Specifically, for a given day *t*, all firms with news share one intercept, all firms with earnings announcements share another intercept, all firms with both news and earnings announcements share a third intercept, and all firms without news or earnings announcements share a fourth intercept. In this regression, the comparison is between two firms that have the same information event on the same day but have different values of *Net*. The coefficients in this regression are similar to

13

those in the second regression: the coefficient on *Net* is 0.363, while the earnings-day and news-day interactions are 2.127 and 0.300, respectively.

The results in Panel B, which report results on news and earnings announcement returns over three-day windows, are similar. The information-day coefficients are smaller compared to Panel A, which shows that most information is reflected in prices the day it is released. The news-day interactions remain positive but are insignificant. Taken together with the results in Panel A, these results suggest that the effect of news on anomalies is reflected in prices almost fully on the day the information is released.

The coefficients reported in both panels document unconditionally higher returns on both earnings days and news days. The earnings-day result is consistent with Franzini and Lamont (2006). However, our anomaly results are completely different from the results in Franzini and Lamont (2006). We show that *anomaly returns* are elevated on information days, and that this effect holds after controlling for the fact that regular stock returns are higher on earnings announcement days. Note also that our result is asymmetric (we discuss this in more detail below). For stocks with negative values of *Net*, anomaly returns are *lower* on earnings days.

B. Estimating Separate Long and Short Anomaly Effects

In Table IV, we remove *Net* from the regressions and replace it with *High Net* and *Low Net* dummy variables. The dummies are based on quintiles constructed via daily sorts on *Net*. Using *High Net* and *Low Net* allows us to examine whether the information effects we document above are different for the long and short sides of anomalies. We use the lagged controls

14

described in the previous section in both regressions reported in Table IV along with day fixed effects.

(Table IV goes here)

Relative to Table III, which uses the continuous *Net* variable, the portfolio results in Table IV are sharper. The first regression in Table IV uses the one-day announcement window. In this regression, the coefficient on *High Net* is 0.018, while that on the *High Net x Earnings Announcement* interaction is 0.139. These results indicate that long-side anomaly returns are 872% higher on earnings announcement days. The news-day interaction is 0.031, which shows that long-side anomaly returns are 272% higher on news days.

The effects on the short side are similar. The coefficient on *Low Net* is -0.017, while that on the *Low Net x Earnings Announcement* interaction is -0.111, which indicates that short-side anomaly returns are 753% lower on earnings announcement days. The news-day interaction is -0.041, which shows that short-side anomaly returns are 341% lower on news days.

In column (2), we replace the one-day window with a three-day window for the news and earnings announcements. The results are again similar. The magnitudes are smaller, which is to be expected with the longer window, but the signs and significance of the coefficients are unchanged.

Figure 1 explores the dynamics of these effects before, on, and after an earnings announcement. We plot the coefficients from a regression of daily returns regressed on *High Net* and *Low Net* interacted with three-day windows surrounding the earnings announcement. The figure clearly displays an asymmetric effect that does not reverse. In particular, anomaly-shorts have lower abnormal returns during the three-day announcement window, while anomaly-longs have higher abnormal returns during the same period. Both sides have milder abnormal

15

returns before and after the announcement. Thus, there is no reversal of the announcement effect; instead, things seem to return back to "normal."

(Figure 1 goes here)

C. Do the Effects vary Across Anomaly Types?

We next ask whether the type of information used to create the anomaly affects the dynamics of its return around information days. Put differently, we ask whether the results in Tables III and IV hold for different types of anomalies or are instead limited to certain types of anomalies. To categorize anomalies, we follow McLean and Pontiff (2016), who classify anomalies into four different types: (i) Event, (ii) Market, (iii) Valuation, and (iv) Fundamentals. The classification is based on the information needed to construct the anomaly.

Event anomalies are based on events within the firm, external events that affect the firm, and changes in firm performance. Examples of event anomalies include share issues, changes in financial analyst recommendations, and unexpected increases in R&D spending. Market anomalies are anomalies that can be constructed using only financial data, such as volume, prices, returns, and shares outstanding. Momentum, long-term reversal, and market value of equity are included in our sample of market anomalies.

Valuation anomalies are ratios, where one of the numbers reflects a market value and the other reflects fundamentals. Examples of valuation anomalies include sales-to-price and market-to-book. Finally, fundamental anomalies are those that are based only on financial statement data. Leverage, taxes, and accruals are fundamental anomalies.

In this analysis, we construct the same *Net* variable as before, but now we sum portfolio memberships within each of the four groups. As in the previous tables, the regressions include

16

time fixed effects, the lagged control variables used in the previous tables, and standard errors clustered on time.

In Table V we report regression results separately for each anomaly type. The results show that the earnings-day and news-day effects documented in the earlier tables are pervasive. In particular, *Event*, *Market*, and *Valuation* anomalies all perform better on earnings days and news days. The *Fundamental* anomalies are much stronger on earnings days, but weaker on news days. As we note above, *Fundamental* anomalies are based entirely on accounting information, which may explain why they are so affected by earnings.

(Table V goes here)

With respect to the earnings-day interactions, all of the anomaly-long portfolios have positive earnings-day interactions, and three of the four anomaly-short portfolios have negative earnings-day interactions. The exception is the short-side *Market* anomaly portfolio, which has a positive earnings-day interaction. The *Market* anomalies are based only on market data, and no accounting data, so this may explain why they are more affected by news than earnings announcements.

With respect to the news-day interactions, on the short side all of the anomaly groups have negative and significant interactions with the exception of *Fundamental* anomalies, which have a positive, and marginally significant interaction. On the long side, *Market* anomalies have a positive and significant interaction, the *Valuation* and *Event* interactions are insignificant, and *Fundamental* anomalies have a negative and significant interaction. Overall, the results show that all types of anomalies are stronger on news days and earnings days, with some differences that seem to be related to the information used to construct the anomaly variable.

D. Firm-Specific News or a Different Explanation?

17

In this section we try to better understand whether higher anomaly returns on information days reflect a reaction to firm specific news, or can instead be explained by some other effect. Specifically, we ask whether day-of-the-week effects, macroeconomic news announcements, or reverse causality (perhaps extreme returns cause news) can explain why anomaly returns are higher on information days.

Day-of-the-Week Effects. Birru (2017) finds that anomalies for which the long leg is the speculative leg perform better on Fridays, and anomalies for which the short leg is the speculative leg perform better on Mondays. Birru (2017) argues that these patterns are consistent with studies in the psychology literature, that show that mood increases from Thursday to Friday and decreases on Monday.

To test whether such day-of-the-week effects influence our results, we estimate a specification in which we interact the *High Net* and *Low Net* anomaly variables with Monday and Friday dummy variables and the news-day and earnings-day dummy variables. We report these results in the first column of Table VI. The results show that including the Monday and Friday interactions has virtually no effect on the earnings day and news-day interactions, as the coefficients reported in column 1 of Table VI are very similar to those reported in column (1) of Table IV, which is the same regression but excluding the Monday and Friday interactions.

(Table VI goes here)

We do not classify our anomalies into speculative and nonspeculative legs like Birru (2017) does, and thus our results may not be directly comparable to his, but we do find evidence of day-of-the-week effects with our anomaly variables. Both *High Net* and *Low Net* perform better on Mondays, and worse on Friday, although the Friday effects are not significant. The Monday effect is quite strong: anomaly returns are more than twice as strong on Monday as compared to other days of week. To the best of our knowledge, this result has not been

18

previously shown in the literature. One explanation for the Monday effect is that more information is impounded into prices on Monday, as news is released over the weekend but investors cannot trade until Monday. Hence, Monday itself can be thought of as a type of "news-day" indicator. These results are therefore consistent our result that anomalies perform better on days when new information gets incorporated into prices.

Macroeconomic News. Savor and Wilson (2013) find that market returns are higher on days when macroeconomic news about inflation, unemployment, or interest rates is scheduled for announcement. They argue that their results reflect compensation for the increased risk that is associated with such announcements. It could be the case that investors infer macroeconomic news from earnings announcements and corporate news stories, and that this in turn explains why anomaly returns are higher on information days.

To control for the effects of macroeconomic news, we estimate a specification that interacts the *High Net* and *Low Net* anomaly variables with a macro news dummy. Specifically, the macro news dummy is the same variable as that used in Savor and Wilson (2013). The macro news dummy is equal to one if there is a scheduled announcement regarding inflation, employment, or interest rates, and zero otherwise. We report the results from this specification in column (2) of Table VI. The results show that the inclusion of the macroeconomic news interactions has virtually no effect on the earnings-day and news-day interactions. The results also show that anomaly portfolios perform significantly worse on macroeconomic news days. The long-side returns are half as large on macro news days, and short-side returns are actually positive, that is short positions have negative alphas on macro news days. These findings do not support the idea that anomaly returns are higher on days with earnings announcements and corporate news because investors infer macroeconomic news from these firm-specific information events.

19

Reverse Causality. It could be the case that extreme returns cause news stories. If this is the case, then the anomaly-news day interactions we document might not reflect news being impounded into asset prices, but instead reflect news stories written about high and low stock returns. This effect cannot explain our anomaly earnings announcement interactions (stock returns do not cause firms to report earnings), which are significantly larger than the news-day interactions. Nonetheless, we address the possibility that this framework can explain our news-day interactions by interacting *High Net* and *Low Net* with the contemporaneous daily stock return squared. The slope on the interaction measures whether *High Net* and *Low Net* perform differently on extreme return days.

We report the results from this test in column (3) of Table VI. The results show that including the extreme return interactions has almost no effect on our news-day and earnings-day interactions. The one exception is that anomaly-shorts perform better on news days and earnings days if the extreme return interactions are included. Anomaly-shorts also perform worse on extreme return days, whereas anomaly-longs are unaffected. The results here do not support the idea that anomaly returns are higher on information days because news stories are being written about extreme returns.

III. Risk as an Explanation for the Anomaly-News Findings

In this section, we examine whether risk-based explanations can explain why anomaly returns are higher on earnings days and news days.

Dynamic Risk and Mispricing Framework. Consider the expected return of a stock that is on the long side of a portfolio based on our *Net* variable. Assume that the returns associated with *Net* can be explained entirely by exposure to systematic risk. In the typical static-factor

20

model, we could write the expected return for such a stock as,

$$E(r_{i,t}) = rf + Beta_i * E(RiskPremium). \qquad (2)$$

Could such a framework explain the stylized facts presented in the paper so far? Recall that Table III shows that *Net* predicts higher returns, and that this return predictability is elevated on earnings announcement days. Note that in equation (2), beta and the expected value of the risk premium are both time-invariant, so this static framework cannot explain our findings.

We can modify equation (2) by letting beta and the expected value of the risk premium vary overtime, as we do in equation (3):

$$E(r_{i,t}) = rf + Beta_{i,t} * E(RiskPremium_t). \qquad (3)$$

In equation (3) both the risk premium and individual firm betas can vary over time. This might explain our findings if either (i) risk premiums increase on information days, or (ii) beta changes on information days. If high (low) *Net* stocks have high (low) betas, then an increase in the risk premium on information days would result in an increase in the return spread between high and low *Net* stocks. Alternatively, it could be the case that when information is released, beta increases for high *Net* stocks and decreases (or increases much less) for low *Net* stocks. If investors expect this and know the announcement date ahead of time, then it could also account for the elevated spread between high and low *Net* stocks on information days.

In Table VII we test whether dynamic risk premiums or dynamic betas can explain why anomaly returns are higher on information days. The previous tables show that the effects are more salient on earnings days than on news days, so for parsimony we only include earnings days in these tests.

21

(Table VII goes here)

As we explain above, one reason that anomaly returns could be higher on earnings days is that risk premiums are elevated on earnings days, and high (low) *Net* stocks have high (low) betas, so a spike in risk premiums leads to a widening in return differentials between high and low *Net* stocks. To address this possibility, in regression (1), we regress daily stock returns on the *High Net* and *Low Net* dummies, an earnings-day dummy, interactions between the earnings-day dummy and the high and low *Net* dummies, and day fixed effects. The day fixed effects mean that our coefficients reflect differences in expected returns *across stocks on the same day*.

In regression (1), the coefficients on *High Net* and *High Net x Earnings Day* are both positive and significant, whereas those on *Low Net* and *Low Net x Earnings Day* are both negative and significant. The coefficients indicate that if there are two high *Net* stocks on day *t*, and one has an earnings announcement while the other does not, then the stock with the earnings announcement has a return that is almost 10 times higher. Similarly, if there are two low *Net* stocks on day *t*, and one has an earnings announcement while the other does not, the stock with the earnings announcement has a return that is about five times lower. These results clearly cannot be caused by a daily change in risk premiums, as we are comparing returns *across stocks with the same beta (Net) on the same day*.

Regression (2) takes things a step further, including *Day x Hi Net* and *Day x Low Net* fixed effects. In this test, on each day *t*, high *Net* stocks have their own intercept, low *Net* stocks have their own intercept, and the rest of the stocks have their own intercept. We find that in this specification as well, the coefficient on *High Net x Earnings Day* is positive and significant while that on *Low Net x Earnings Day* coefficient is negative and significant. Thus, if we compare two high *Net* stocks on day *t*, the one with the earnings announcement has the higher return, while if

22

we compare two low *Net* stocks on an earnings day, the one with the earnings announcement has a lower return, even after accounting for the average return in each group on each day (via the fixed effect). These results again contradict the idea that anomaly returns are higher on earnings days because risk premiums are higher, as the fixed effects absorb any changes in the risk premium. To summarize, we take stocks with the same level of *Net* (high or low) and show that there are large differences *across these stocks on the same day due* to earnings announcements.

In the next few regressions, we examine whether betas can change on earnings announcement days. We add either a market portfolio factor (*Market*) or an anomaly factor (*Factor*) to our regressions, and interact each with the information-day dummies*.* The variable *Market* is the daily return on the CRSP value-weighted index, while *Factor* is the daily return on a portfolio that is long in stocks ranked in top 20% percentile of *Net* and short in stocks ranked in the bottom 20% percentile of *Net*. The coefficients on *Factor* and *Market* reflect the average stock's beta with respect to each portfolio.

Following Shanken (1990), we use interactions to consider both time-series and cross-sectional variations in beta. In regression (3), we include interactions between each of the *Net* variables and *Market*, as well as three-way interactions that include each of the *Net* variables, *Market*, and the earnings-day dummy. In regression (4) we perform a similar regression, but replace *Market* with *Factor*. We include day fixed effects in both regressions, and thus coefficients for *Market* and *Factor* are not estimated.

In regression (3), we see that, when compared to regression (1), the coefficients on *High Net* and *Low Net* and the coefficients on the earnings-day interactions are virtually the same. Hence, controlling for market beta and the fact that market beta can be elevated on earnings days does not seem to explain our findings. We also see that the coefficient on *High Net x Market*

23

is negative, whereas that on *Low Net x Market* is positive. What this shows is that high *Net* stocks have *lower* market betas than low *Net* stocks do, that is *Net* produces a portfolio that has a *negative* beta. This finding is consistent with several earlier studies, which also show that stocks on the long (short) side of anomaly portfolios have higher (lower) betas.[4] The results further show that for both high and low *Net* stocks, betas are not significantly higher on earnings announcement days. The coefficient on *Earnings Day x Market* is positive and significant, which shows that on average, across all stocks, betas are higher on earnings days, consistent with Patton and Verardo (2012) and Savor and Wilson (2016), but this effect is not different for high or low *Net* stocks. Overall, the results in regression (3) do not support the idea that elevated market betas can explain why anomaly returns are higher on earnings days.

Regression (4) is like regression (3), except that it replaces *Market* with *Factor*, which is a portfolio that is long in high *Net* stocks and short in low *Net* stocks. Here again we see that, compared to regression (1), the coefficients on *High Net* and *Low Net* and the coefficients on the earnings day interactions are virtually the same. Hence, controlling for beta with respect to *Factor* and the fact that *Factor* can change on earnings days does not seem to explain why anomaly returns are higher on earnings days. We also see that the interaction between *Factor* and the earnings-day dummy is negative and significant. What this means is that if a stock has an earnings announcement, its covariance with *Factor* is lower.

Regression (4) further shows that the interaction between *High Net* and *Factor* is positive and significant (this is partly mechanical), as is the three-way interaction between *High Net*, *Factor*, and the earnings-day dummy. Similarly, we see that the interaction between *Low*

---

[4] For example, high book-to-market and high earnings-to-price stocks have low betas (Fama and French (1992)), high momentum stocks have low betas (Jegadeesh and Titman (1993)), high idiosyncratic stocks earn lower returns despite the fact that they have higher betas (Ang, et al. (2006)), and firms that repurchase shares experience positive abnormal returns and reductions in betas (Grullon and Michaely (2004)).

24

Net and *Factor* is negative and significant (this is also partly mechanical), as is the three-way interaction between *Low Net*, *Factor*, and the earnings-day dummy. We thus find that beta with respect to *Factor* is higher (lower) for high (low) *Net* stocks, and that this effect is stronger on earnings days. However, even after all of the above controls, it is still the case that high (low) *Net* stocks have higher (lower) returns and that this effect is stronger on earnings days. It follows that beta with respect to *Factor* cannot fully account for our findings, but it could still be the case that if we had the "right" factor, beta would account for our findings.

Overall, the results in Table VII provide evidence that factor betas are elevated on earnings days. Thus, for a risk-based model to explain anomaly returns, it would need to have this feature. Static models and models that only allow risk premiums to vary with time are therefore not likely to be able to explain anomaly returns. Instead, a risk-based model needs to explicitly allow for dynamic betas that can change dramatically when firm-specific information is released if it is going to explain stock return anomalies.

IV. Biased Expectations

In this section of the paper we take a closer look at the idea that biased expectations can explain why anomaly returns are higher on information days. Under the biased expectations framework, investors are too optimistic about some stocks and too pessimistic about others, and the anomaly variables reflect these biases. As an example, stocks with high past sales growth have low stock returns (Lakonishok, Shleifer, and Vishny (1994)). A possible reason for this is that investors naively extrapolate past sales growth into the future, whereas in reality there is a good deal of mean reversion. Hence, high sales growth firms subsequently report lower sales and earnings than investors expect, resulting in low stock returns on days that this information is released.

25

Models of stock return anomalies that are based on biased expectations include Barberis, Shleifer, and Vishny (1998) and Daniel, Hirshleifer, and Subrahmanyam (1998, 2001). In these models, long-term reversal and price-to-fundamental anomalies are caused by biased expectations about future cash flows and a price correction that occurs when new information is made public. Here we ask whether the ideas in these papers can be extended to explain anomalies in general.

To further study the role of biased expectations in a setting that cannot be affected by risk (static or dynamic), we consider analyst earnings forecast errors.[5] No framework that we know of links risk, especially dynamic risk, to analyst forecast errors, which reflect mistakes on the part of sell-side analysts. The results in this paper thus far show that when new information is released, anomaly-longs have higher returns and anomaly-shorts have lower returns. If biased expectations explain these effects, and if analysts' earnings forecasts are correlated with investors' expectations, then analyst earnings forecasts should be too low (high) for stocks on the long (short) side of anomaly portfolios.

Our analyst earnings forecast error variable is a summary variable constructed using data from IBES. Specifically, it is given as the difference between a stock's last reported median sell-side forecast and actual reported earnings (per IBES), divided by the closing stock price in the previous month:

$$Forecast\ Error_t = \frac{Earnings\ Forecast_t - Actual\ Earnings_t}{Price_{t-1}}.$$

This variable is winsorized at the 1st and 99th percentiles. The biased expectations framework

---

[5] Analyst forecast errors are particularly useful as a proxy for expected cash flows. Analysts also generate recommendations and price targets. Engelberg, McLean, and Pontiff (2017) study the extent to which information from anomaly variables is reflected in analyst enthusiasm for a stock.

26

predicts that this variable is negative for long-side stocks (forecast too low) and positive for short-side stocks (forecast too high). We obtain data from IBES for the period 1983 through 2014. We merge the forecast data with our anomaly data and test whether anomaly portfolio membership can predict forecast errors.

We control for the number of analysts making earnings forecasts, whether there is only a single forecast, and the standard deviation of the forecast scaled by the stock price. If there is only a single forecast, we set the standard deviation of the forecast to zero.

We report the results from these tests in Table VIII. We multiply the forecast error variable by 100 so that the coefficients are easier to read. The first regression reports findings for the full 97-anomaly sample. The regression coefficients show that analyst forecasts are too high for stocks in the short side of anomaly portfolios and too low for stocks in the long side of anomaly portfolios. Both of these effects are statistically significant.

(Table VIII goes here)

The effects are also economically significant. The regression intercept is 0.028, and the regression coefficients show that the effect of being in the *High Net* portfolio is a forecast error that is -0.045 lower, while the effect of being in the *Low Net* portfolio is a forecast error that is 0.017 higher. These effects are large and support the idea that earnings forecast surprises have a sizeable impact on anomaly returns on earnings announcement days.

Table VIII also reports the effects across the four anomaly groups. We see that in all four groups, the *High Net* variable is negative and significant for three of the anomaly groups, while the *Low Net* variable is positive and significant for three of the anomaly groups. The exception in both cases corresponds to the *Market* anomalies, for which the opposite results obtain. As we explain earlier, market anomalies include variables based only on market data, and include

27

momentum, reversal, and idiosyncratic risk. With *Market* anomalies, analyst forecasts are too high for the longs and too low for the shorts. The evidence here is consistent with that in Table V, which shows that *Market* anomalies perform better on news days but on not earnings days.

In sum the results in Table VIII largely agree with those in the other tables above. Investors and analysts seem to be too pessimistic (optimistic) about future earnings of stocks in the long (short) side of anomaly portfolios. This bias is revealed in stock returns when firms announce earnings and other news. This finding is consistent with the biased expectations explanation for anomaly returns, but not risk. However, as we explain below, these results may not contradict data mining as an explanation for anomaly returns, so further tests are needed before concluding that biased expectations play a role in anomaly returns.

V. Data Mining

Fama (1998), McLean and Pontiff (2016), and Harvey, Liu, and Zhu (2016), suggest that data mining could explain a good deal of cross-sectional return predictability. In our sample, earnings-day returns have a return standard deviation that is twice that of noninformation days, and Dow Jones news days have a return standard deviation that is 30% greater than noninformation days. Given that returns are so much more volatile on information days, an anomaly that is a result of data mining would likely perform especially well on information days. Accordingly, we conduct several tests of the hypothesis that the information-day effects documented in this paper can be explained by data mining.

Data-Mined Strategies and Information-Day Returns. We first examine whether a data-mined strategy performs especially well on information days. To do so, we test whether a firm with a high (low) return in month *t,* regardless of whether it is in an anomaly portfolio, would

28

also have high (low) information-day returns in month *t*. Specifically, we estimate the following regression equation:

$$R_{i,t} = \alpha_t + \beta_1 Monthly_{i,t} + \beta_2 Monthly_{i,t} \times Eday_{i,t} + \beta_3 Monthly_{i,t} \times Nday_{i,t} + \beta_4 Eday_{i,t}$$

$$+ \beta_5 Nday_{i,t} + \sum_{j=1}^{10} \gamma_j Lag\ Return_{i,t-j} + \sum_{j=1}^{10} \delta_j Lag\ Return^2{}_{i,t-j}$$

$$+ \sum_{j=1}^{10} \rho_j Volume_{i,t-j} + \varepsilon_{i,t}. \qquad (4)$$

The above equation is the same as equation (1), except that we replace *Net* with *Monthly*, the contemporaneous monthly stock return. The dependent variable is the daily stock return. We expect the coefficient on *Monthly* to be positive and significant, that is, firms with higher stock returns in a month to also have higher stock returns during the days of that month. The interactions test whether, after controlling for the effects of *Monthly*, stocks with high (low) monthly stock returns also have especially (high) low information-day returns during that month.

The results for our estimation of equation (4) are reported in column (1) of Table IX. We find that it is indeed the case that when monthly returns are high (low), information-day returns during the month are especially high (low). The coefficient on *Monthly* is 5.225, which shows that a firm with a stock return of 10% in a given month has an expected daily return of 0.5225% during that month. If the day has an earnings announcement, the expected return increases by a factor of 11.6. If the day has a Dow Jones news story, the expected return increases by a factor of 2.3. Hence, a data-mined strategy would also have extreme returns on

29

information days.

(Table IX goes here)

In column (2) of Table IX we add *Net* and the interactions between *Net* and the earnings-day and news-day dummies along with *Monthly* and its interactions. The coefficient on *Net* is negative and significant in this regression. Thus, after controlling for monthly returns, high *Net* stocks have *lower expected returns on noninformation days*. This means that if we have two stocks with same return in month *t*, *the anomaly stock earns more of its return on information days*. This result does not support the idea that anomaly returns are caused by data mining: if anomaly returns reflect data mining, then we would expect the same daily return patterns between anomaly stocks and non-anomaly stocks with the same monthly return.

The *Net* interactions with both the earnings- and news-day dummies are positive and significant, which shows that even after controlling for the effects of *Monthly*, anomalies still perform especially well on information days. This result contradicts the idea that anomalies can be explained by data mining alone.

This result also contradicts the idea that extreme stock returns cause news, which we discuss earlier in the paper. If extreme returns cause news, then this should be the case for both anomaly firms and nonanomaly firms. Yet we find that the effect of news on stock returns is stronger for anomaly firms, even after controlling for the level of monthly returns.

Critically, our test does not estimate the portion of anomaly returns that come from data mining; rather, it estimates the incremental return that data mining cannot explain. As an analogy, suppose a student may have been given a "cheat sheet" for an exam, so that if he followed it he would get a 75 on the exam. If we observe him getting a 95 on the exam, we can reject the hypothesis that his exam score was due purely to cheating. However, we cannot

30

identify the points due to cheating--he may have scored a 95 entirely on his own merit, or he may have used the "cheat sheet" for 75 of the points. Similarly, our tests in Table IX allow us to reject the null of pure data mining, but we cannot say what portion, if any, of our results comes from data mining.

Firm Size. A number of studies show that anomalies tend to be stronger in small firms, illiquid firms, and firms with high idiosyncratic risk (see Pontiff (1996, 2006)). We can think of no reason why spurious anomalies should be stronger in small firms. We therefore split our sample into small and large firms, where large (small) stocks are those above (below) the median market capitalization on day t, and estimate equation (1) within each sample. In these specifications we continue to control for the *Monthly* stock return, and therefore compare the coefficients to those reported in column (2).

We report these size-partitioned results in columns (3) and (4) of Table IX. The coefficient on *Net* is insignificant in both specifications, which shows that within size groups, after we control for *Monthly*, the daily return of a high or low *Net* stock is not different on a noninformation day. In column (3), which reports the results for large stocks, the *Net* earnings-day interaction is 0.715, while in column (4) the *Net* earnings-day interaction is 3.067, or four times higher, for small stocks. Similarly, the news-day interaction is insignificant among large stocks but positive and significant among small stocks. These results also show that virtually all of the difference in anomaly returns between large and small stocks occurs on information days. Data mining does not predict such dramatic differences between large and small stocks, but mispricing theories, which require limits to arbitrage, do. For example, Pontiff (1996, 2006), Shleifer and Vishny (1997), and Pedersen (2015) all argue that the size of the market inefficiency should be related to the cost of correcting that inefficiency. Given that arbitrage costs are greater among small stocks, under the mispricing theory of anomaly returns we expect

31

news to lead to larger corrections of mispricing for small stocks because there is more mispricing to correct.

Out-of-Sample Predictability. An alternative way to get at the data-mining question is to focus only on anomalies after the sample period in the study that first documented the anomaly. We therefore build the out-of-sample anomaly variable, (*OOS)*, which is constructed similarly to *Net* except that *OOS* only uses anomalies in months after the end date of the original sample. As an example, the sample period for the accrual anomaly (Sloan (1996)) is 1962 to 1991. With *OOS*, we begin to measure the accrual anomaly in 1992, whereas with *Net* we begin using accruals in 1979 (the first year for which we have news data).

We report the results for *OOS* in column (5) of Table IX. This specification is that in equation (1), except that here *OOS* replaces *Net*. The results for *OOS* are similar to those for *Net*. Specifically, using *OOS*, we estimate that anomaly returns are almost 5.6 times higher on earnings announcement days and 50% higher on Dow Jones News days. As a comparison, in regression (3) in Table III we see that news-day and earnings-day effects are virtually the same.

Data Mining and Analyst Forecast Errors. The results in Table VIII show that analysts' earnings forecasts are too low for anomaly-longs and too high for anomaly-shorts, which is consistent with the idea that biased expectations are what create anomaly returns. Yet this finding is also consistent with data mining as an explanation for anomaly returns. A spurious anomaly is likely just by chance to be long in stocks that have positive earnings surprises and short in stocks that have negative earnings surprises. It would be difficult to generate abnormal returns otherwise.

To further explore this idea, we reestimate the analyst forecast error regression reported in column (1) of Table VIII, but use *Monthly* in place of *High Net* and *Low Net*. These results are reported in column (1) of Table X. The coefficient on *Monthly* is negative and

32

significant, which shows that analyst forecasts were too low for stocks with high returns and too high for stocks with low returns. These findings suggest that virtually any variable that predicts returns *in-sample*, be it spurious or authentic, would most likely also predict analysts' forecast errors.

(Table X goes here)

To control for this data-mining effect, we estimate a specification that includes *Monthly* along with *High Net* and *Low Net*. The results for this specification are reported in column (2) of Table X. As in column (1), *Monthly* is negative and significant, however *High Net* is also negative and significant and *Low Net* is positive and significant. These results are inconsistent with the idea that data mining explains earnings forecast error predictability by anomaly variables.

We next explore the possibility of data mining by replacing *High Net* and *Low Net* with *High OOS* and *Low OOS*. The *OOS* variables are constructed entirely with anomalies that are out-of-sample, which makes it unlikely that results with the *OOS* variables can be explained by data mining. We report these results in column (3) of Table X. The coefficient on *High OOS* is negative and significant, while that on *Low OOS* is positive and significant. Hence, even out-of-sample anomaly variables can predict analysts' forecast errors. This finding is difficult to reconcile with risk or data mining, but is fully consistent with mispricing.

VI. Conclusions

Evidence of cross-sectional return predictability goes back more than four decades, yet to this day academics disagree about its cause. In this paper we compare return predictability on news days to non-news days, and find that anomaly returns are elevated on news days. We document this using a sample of 97 anomalies. This finding is robust across different types of

33

anomalies, it is not explained by day-of-the-week effects, nor is it explained by anomalies simply being greater on high volatility days. The results likely reflect firm-specific news, as anomaly returns are not higher on days when macroeconomic news is announced.

Although earlier studies conclude that higher anomaly returns on earnings days reflect mispricing, we show that this need not be the case. Such results could also reflect dynamic risk or data mining. We show that anomaly betas are elevated on earnings announcement days, so future studies should consider this feature. That being said, our finding of higher anomaly returns on earnings days is not affected by the inclusion of dynamic betas. With respect to data mining, out-of-sample anomalies, which are likely not explained by data mining, exhibit higher returns on earnings days and news days and also predict analyst forecast errors, so it is unlikely that data mining can explain our findings.

Our findings are most consistent with the idea that investors have overly optimistic expectations about the cash flows of some firms and overly pessimistic expectations about the cash flows of other firms. When new information is released, investors revise their biased beliefs, which causes prices to change, and in turn, the observed return predictability. Evidence from sell-side equity earnings forecasts dovetail with the stock return evidence: analysts overestimate earnings for firms on the short side of anomaly portfolios and underestimate earnings for firms on the long side.

34

REFERENCES

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *Journal of Finance* 61, 259-298.

Baker, Malcolm, Lubomir Litov, Jessica A. Wachter, and Jeffrey Wurgler, 2010, Can mutual fund managers pick stocks? Evidence from their trades prior to earnings announcements, *Journal of Financial and Quantitative Analysis* 45, 1111-1131.

Ball, Raymond, and Phillip Brown, 1968, An empirical evaluation of accounting income numbers, *Journal of Accounting Research* 6, 159-178.

Ball, Raymond, and S.P. Kothari, 1991, Security returns around earning announcements, *The Accounting Review* 66, 718-738.

Bansal, Ravi, Robert F. Dittmar, and Christian Lundblad, 2005, Consumption, dividends, and the cross-section of equity returns, *Journal of Finance* 60, 1639–1672.

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 George Constantinides, Milton Harris, and Rene Stulz, eds., *Handbook of the Economics of Finance* (North-Holland).

Basu, S., 1977, Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient markets hypothesis, *Journal of Finance* 32, 663-682.

Bernard Vincent, and Jacob Thomas, 1989, Post-earnings announcement drift: Delayed price response or risk premium, *Journal of Accounting Research* 27, 1-36.

Birru, Justin, 2017, Day or the week and the cross-section of returns, *Journal of Financial Economics*, forthcoming.

Blume, Marshal E., and Frank Husic, 1973, Price, beta, and exchange listing, *Journal of Finance* 28, 283-299.

Bonferroni, Carlo E., 1935, The calculation of insurance on groups of heads, *Studies in Honor of Professor Salvatore Ortu Carboni* (Rome), 13-60.

Bradshaw, Mark T., Scott A. Richardson, and Richard G. Sloan, 2006, The relation between corporate financing activities, analysts' forecasts and stock returns, *Journal of Accounting and Economics* 42, 53-85.

Carhart, Mark M., 1997, On Persistence in Mutual Fund Performance, *Journal of Finance* 52, 57-82.

Chopra, Navin, Josef Lakonishok, and Jay Ritter, 1992, Measuring abnormal performance: Do stocks overreact? *Journal of Financial Economics* 31, 235-268.

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

De Bondt, Werner, and Richard Thaler, 1987, Further evidence of stock market overreaction and seasonality, *Journal of Finance* 42, 557-581.

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.

36

De Long, J. Bradford, Andrei Shleifer, and Lawrence H. Summers, 1990, Noise trader risk in financial markets, *Journal of Political Economy* 98, 703–738.

Dittmar, Robert, and Christian Lundblad, 2015, Firm characteristics, consumption risk, and firm-level risk exposures, Working paper, University of Michigan.

Duan, Ying, Gang Hu, and R. David McLean, 2009, When is stock-picking likely to be successful? Evidence from mutual funds, *Financial Analysts Journal* 65, 55-65.

Duan, Ying, Gang Hu, and R. David McLean, 2009, Costly arbitrage and idiosyncratic risk: Evidence from short sellers, *Journal of Financial Intermediation* 19, 564-579.

Edelen, Roger, Ozgur Ince, and Gregoy Kadlec, 2016, Institutional investors and stock return anomalies, *Journal of Financial Economics* 119, 472-488.

Engelberg, Joseph, R. David McLean, and Jeffrey Pontiff, 2017, Analysts and anomalies, Working paper, Boston College.

Engelberg, Joseph, Adam Reed, and Matthew Ringgenberg, 2012, How are shorts informed?: Short sellers, news, and information processing, *Journal of Financial Economics* 105, 260-278.

Fama, Eugene, F., 1976, *The Theory of Finance* (Basic Books, New York, NY).

Fama, Eugene F., 1991, Efficient capital markets: II, *Journal of Finance* 46, 1575-1617.

Fama, Eugene F., 1998, Market efficiency, long-term returns and behavioral finance, *Journal of Financial Economics* 49, 283-306.

Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427-465.

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

Fama, Eugene F., and Kenneth R. French, 1996, Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51, 55–84.

Foster, George, Chris Olsen, and Terry Shevlin, 1984, Earnings releases, anomalies, and the behavior of security returns, *The Accounting Review* 59, 574-603.

Frazzini Andrea, and Owen Lamont, 2006, The earnings announcement premium and trading volume, NBER Working paper 13090.

Grullon, Gustavo, and Roni Michaely, 2004, The information content of share repurchase programs, *Journal of Finance* 59, 651-680.

Hansen, Lars Peter, John C Heaton, and Nan Li, 2008, Consumption strikes back? Measuring long-run risk, *Journal of Political Economy* 116, 260–302.

Harvey, Campbell R., Yan Liu, and Heqing Zhu, 2016, … and the cross-section of expected returns, *Review of Financial Studies* 29, 5-68.

Jagannathan, Ravi, and Yong Wang, 2007, Lazy investors, discretionary consumption, and the cross-section of stock returns, *Journal of Finance* 62, 1623–1661.

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

38

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

La Porta, Rafael, Josef Lakonishok, Andrei Shleifer, and Robert Vishny, 1997, Good news for value stocks: Further evidence on market efficiency, *Journal of Finance* 52, 859-874.

Linnainmaa, Juhani, and Michael Roberts, 2018, The history of the cross section of stock returns, *Review of Financial Studies* .

Liu, Laura X., and Lu Zhang, 2014, A neoclassical interpretation of momentum, *Journal of Monetary Economics* 67, 109-128.

McLean, R. David, and Jeffrey Pontiff, 2016, Does academic research destroy stock return predictability? *Journal of Finance* 71, 5-32.

Miller, Edward, 1977, Risk, uncertainty, and divergence of opinion, *Journal of Finance* 32, 1151-1168.

Nagel, Stefan, 2013, Empirical cross-sectional asset pricing, A*nnual Review of Financial Economics* 5, 167-199.

Parker, Jonathan, and Christian Julliard, 2005, Consumption risk and the cross section of expected returns, *Journal of Political Economy* 113, 185–222.

Pedersen, Lasse, 2015, *Efficiently Inefficient: How Smart Money Invests and Market Prices Are Determined*, (Princeton University Press, Princeton, NJ).

Patton, Andrew J. and Michela Verardo, 2012, Does beta move with News? Firm-specific information flows and learning about predictability, Review of Financial Studies 25, 2789-2839.

Pontiff, Jeffrey, 1996, Costly arbitrage: Evidence from closed-end funds, *Quarterly Journal of Economics* 111, 1135-1151.

Pontiff, Jeffrey, 2006, Costly arbitrage and the myth of idiosyncratic risk, *Journal of Accounting and Economics* 42, 35-52.

Savor, Pavel, and Mungo Wilson, 2013, How much do investors care about macroeconomic risk? Evidence from scheduled economic announcements, *Journal of Financial and Quantitative Analysis* 48, 343–375.

Savor, Pavel, and Mungo Wilson, 2016, Earnings announcements and systematic risk, *Journal of Finance* 71, 83-138.

Savoy, Alexi, 2011, Asset pricing with garbage, *Journal of Finance* 66, 177-201.

Shanken, Jay, 1990, Intertemporal asset pricing, *Journal of Econometrics* 45, 99-120.

Shleifer, Andrei, and Robert W. Vishny, 1992, Liquidation values and debt capacity: A market equilibrium approach, *Journal of Finance* 47, 1343-1366.

Shleifer, Andrei, and Robert W. Vishny, 1997, The limits of arbitrage, *Journal of Finance* 52, 35-55.

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

Stein, Jeremy C., 2009, Sophisticated investors and market efficiency, *Journal of Finance* 64, 1517-1548.

Tetlock, Paul C., 2010, Does Public Financial News Resolve Asymmetric Information? *Review of Financial Studies* 23, 3520-3557.

Tetlock, Paul C., 2011, All the News That's Fit to Reprint: Do Investors React to Stale Information? *Review of Financial Studies* 24, 1481-1512.

Wu, Jin, Lu Zhang, and X. Frank Zhang, 2010, The q-theory approach to understanding the accrual anomaly, *Journal of Accounting Research* 48, 177-223.

Yogo, Motohiro, 2006, A consumption-based explanation of expected stock returns, *Journal of Finance* 61, 539-580.
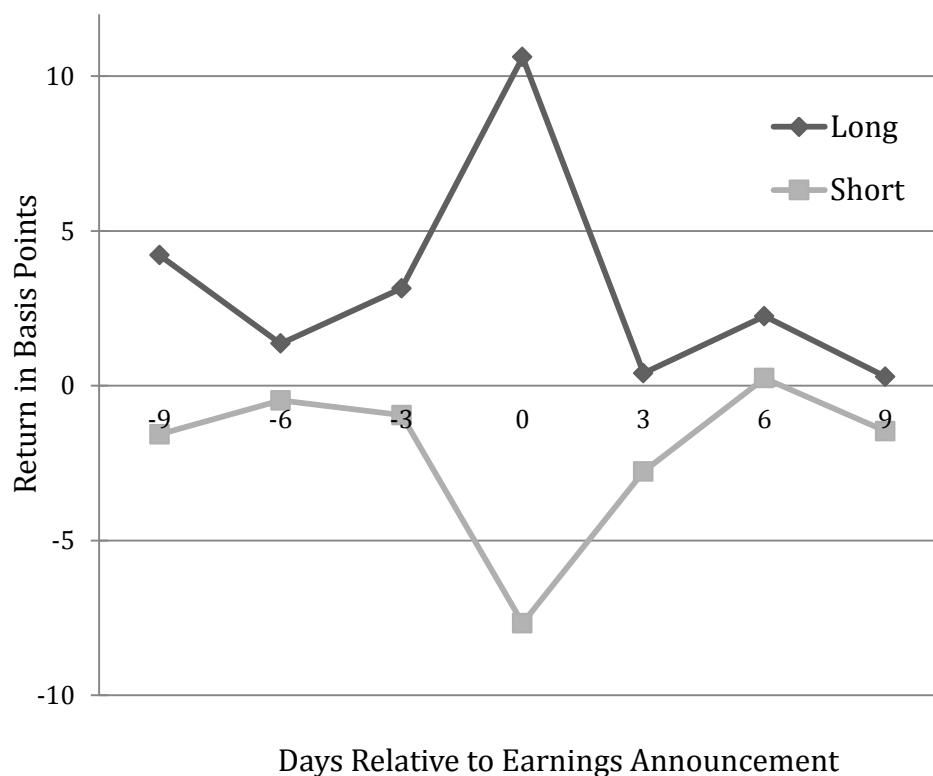
Figure 1. Anomaly returns around earnings announcement days. This figure plots the coefficients from regressions of daily returns on the *High Net* and *Low Net* dummy variables, dummies for three-day windows around earnings announcements, interactions between *High Net* and *Low Net* and the three-day window dummies, and day fixed effects. Returns are expressed in basis points. *High Net* and *Low Net* are defined in Table II. The figure plots the sum of the coefficients for the interactions and the coefficients for *High Net* (darker line) and *Low Net* (lighter line), that is, we plot the overall effect of *High Net* and *Low Net* for each of the seven three-day windows.

42

Table I

Earnings Announcement and News Data

This table describes our sample in terms of earnings announcements and news releases. The unit of observation is the firm-day level. To be included in our sample, a stock must have a daily stock retur, and a stock price that is at least $5 as of yesterday's close. We obtain earnings announcement dates from the Compustat quarterly database and news announcements from the Dow Jones news archive. If the announcement is made after hours, then the following trading day is the event day. The sample period is 1979 to 2013.

	News Day		Total
	(1)	(2)	(3)

Panel A: Number of Firm-Day Returns

Earnings Day	No	Yes	
No	38,679,894	6,393,856	45,073,750
Yes	274,276	235,444	509,720
Total	38,954,170	6,629,300	45,583,470

Panel B: Percentage of Firm-Day Returns

Earnings Day	No	Yes	
No	84.9%	14.0%	98.9%
Yes	0.6%	0.5%	1.1%
Total	85.5%	14.5%	100%

43

Table II

Descriptive Statistics for the Portfolio Variables

This table provides descriptive statistics for the anomaly variables used in this study. We use the 97 cross-sectional anomalies studied in McLean and Pontiff (2016). Each month, stocks are sorted on each anomaly characteristic (e.g., size, book-to-market, accruals, etc.). We use the extreme quintiles to define the long side and short side of each anomaly strategy. Sixteen of our 97 anomalies are indicator variables (e.g., credit rating downgrades); for these anomalies, there is only a long or short side, based on the binary value of the indicator. For each firm-day observation, we sum the number of long-side and short-side anomaly portfolios that the firm belongs to; this creates the variables *Long* and *Short*. The variable *Net* is equal to *Long–Short*.

				Aggregate Anomaly Variables		
			45, 583,470 Observations			
Variable	Mean	Std. Dev.	25th %ile	75th %ile	Min	Max
Long	4.52	4.46	1	7	0	37
Short	5.52	4.78	2	8	0	45
Net	-0.71	4.19	-3	1	-38	33

44

Table III

Anomaly Returns on Earnings Days and News Days

This table reports results from a regression of daily returns on time fixed effects, the *Net* anomaly variable, earnings day and news day dummy variables, interactions between *Net* and the information-day variables, and control variables (coefficients unreported). Daily return, the dependent variable, is expressed in basis points. The control variables include lagged values for each of the past 10 days for stock returns, stock returns squared, and trading volume. To create the *Net* anomaly variable, we use the 97 cross-sectional anomalies studied in McLean and Pontiff (2016). For each stock-month observation, we create by summing the number of long-side and short-side anomaly portfolios that the stock belongs to, *Long* and *Short*. *Net* is equal to *Long* minus *Short*. We then merge this monthly data set with daily stock return data from CRSP and with daily indicators for earnings announcement days and Dow Jones News stories, which we refer to as information days. We define an earnings day (Eday) or news day (Nday) as the one-day or three-day window around an earnings announcement or news release, that is., days *t*-1, *t*, and *t*+1. Standard errors are clustered on time. *t*-statistics are in parentheses. The sample period is 1979:6 to 2013:12.

Table III (Continued)

	Panel A: One-day Window			Panel B: Three-day Window		
Net	0.367	0.384	0.363	0.387	0.408	0.385
	(5.77)***	(6.06)***	(5.69)***	(6.02)***	(6.35)***	(6.03)***
*Net * Eday*	2.098	2.164	2.127	1.051	1.115	1.083
	(8.32)***	(8.54)***	(8.43)***	(7.86)***	(8.25)***	(7.14)***
*Net * Nday*	0.170	0.178	0.300	0.038	0.042	0.120
	(2.92)***	(3.04)***	(3.08)***	(0.71)	(0.78)	(1.30)
Eday	0.084	0.090		0.034	0.040	
	(8.24)***	(8.71)***		(6.48)***	(7.48)***	

45

Nday	0.117	0.124		0.081	0.087	
	(17.48)***	(19.06)***		(14.65)***	(16.18)***	
Lagged Controls ?	No	Yes	Yes	No	Yes	Yes
Fixed Effects?	Day	Day	Day * Event	Day	Day	Day * Event

Table IV

Long and Short Anomaly Returns on Earnings Days and News Days

This table reports results from a regression of daily returns on the *High Net* and *Low Net* dummy variables, information-day dummy variables, interactions between *High Net* and *Low Net* and the information-day variables, control variables (coefficients unreported), and time fixed effects, Daily return, the dependent variable, is expressed in basis points. The controls include lagged values for each of the past 10 days for stock returns, stock returns squared, and trading volume. To create the *High Net* and *Low Net* anomaly variable, we use the 97 cross-sectional anomalies studied in McLean and Pontiff (2016). We sort firms each month on our aggregate anomaly variable, *Net*, and classify them into high and low quintiles. We then merge this monthly data set with daily stock return data from CRSP and with daily indicators for earnings announcement days and Dow Jones News stories, which we refer to as information days. We define an earnings day (Eday) or news day (Nday) as the one-day or three-day window around an earnings announcement or news release, that is, days t-1, t, and t+1. The sample period is 1979:6 to 2013:12. Standard errors are clustered on time. t-statistics are in parentheses.

	One-day Window	Three-day Window
	(1)	(2)
High Net	0.018	0.016
	(6.79)***	(6.62)***
Low Net	-0.017	-0.016
	(3.91)***	(3.88)***
*High Net * Eday*	0.139	0.086
	(6.22)***	(8.14)***
*Low Net * Eday*	-0.111	-0.052
	(4.96)***	(4.56)***
*High Net * Nday*	0.031	0.017
	(5.51)	(3.70)
*Low Net * Nday*	-0.041	-0.028
	(6.81)***	(5.30)***

47

Eday	0.068	0.024
	(5.51)***	(3.89)***
Nday	0.126	0.029
	(19.80)***	(17.42)***
Day Fixed Effects?	Yes	Yes

Table V

The Effect of Information Across Anomaly Types

This table tests whether the effect of information on daily anomaly returns varies across types of anomalies. To conduct this exercise, we split our anomalies into the four groups created in McLean and Pontiff (2016): (i) Event, (ii) Market, (iii) Valuation, and (iv) Fundamentals. Event anomalies are those based on corporate events or changes in performance. Examples of event anomalies are share issues, changes in financial analyst recommendations, and unexpected increases in R&D spending. Market anomalies are anomalies that can be constructed using only financial data, such as volume, prices, returns, and shares outstanding. Momentum, long-term reversal, and market value of equity (size) are included in our sample of market anomalies. Valuation anomalies are ratios, where one of the numbers reflects a market value and the other reflects fundamentals. Examples of valuation anomalies include sales-to-price and market-to-book. Fundamental anomalies are those that based only on financial statement data and nothing else. Leverage, taxes, and accruals are fundamental anomalies. The regressions include time fixed effects and controls for lagged values for stock returns, stock returns squared, and trading volume over each of the past 10 days (coefficients unreported). The dependent variable, daily stock return, is expressed in basis points. The sample period is 1979:6 to 2013:12. Standard errors are clustered on time. t-statistics are in parentheses.

	arket	Valuation	Fundamental	Event
High Net	0.000	0.008	0.009	0.013
	(0.02)	(3.26)***	(3.61)***	(5.69)***
Low Net	-0.005	-0.021	-0.021	-0.015
	(1.31)	(4.76)***	(5.51)***	(4.83)***
*High Net * Eday*	0.078	0.121	0.028	0.105
	(3.17)***	(5.98)***	(1.36)	(4.81)***
*Low Net * Eday*	0.052	-0.118	-0.140	-0.082
	(2.59)***	(4.85)***	(6.05)***	(3.80)***
*High Net * Nday*	0.085	-0.000	-0.013	-0.001

49

	(12.38)***	(0.03)	(2.90)***	(0.12)
Low Net * Nday	-0.059	-0.013	0.010	-0.020
	(10.31)***	(2.09)**	(1.75)*	(3.95)***
Eday	0.039	0.067	0.068	0.098
	(3.17)***	(5.41)***	(5.33)***	(7.55)***
Nday	0.124	0.124	0.123	0.120
	(18.85)***	(19.03)***	(19.32)***	(18.64)***
Day Fixed Effects?	Yes	Yes	Yes	Yes

Table VI

Do the Results Reflect Firm-Specific Information?

This table reports results from a regression of daily returns on time fixed effects, the *High Net* and *Low Net* anomaly variables, an information-day dummy variable, interactions between *High Net* and *Low Net* and the information-day variables, and control variables (coefficients unreported). The dependent variable, daily stock return, is expressed in basis points. The controls include lagged values for stock returns, stock returns squared, and trading volume over each of the past 10 days. To create the *High Net* and *Low Net* anomaly variables, we use the 97 cross-sectional anomalies studied in McLean and Pontiff (2016). We sort firms each month on our aggregate anomaly variable, *Net*, and classify them into high and low quintiles. We define an earnings day (Eday) or news day (Nday) as the one-day window around an earnings announcement or news release. In regression (1) we include interactions between *High Net* and *Low Net* and Monday (*Mon*) and Friday (*Fri*). In regression (2) we interact *High Net* and *Low Net* with a macro announcement dummy (*Mac*). Following Savor and Wilson (2013, 2016), *Mac* is equal to one if there is a news announcement regarding inflation, unemployment, or interest rates. The sample period is 1979:6 to 2013:12. Standard errors are clustered on time. *t*-statistics are in parentheses.

Table VI (Continued)

	(1)	(2)	(3)
High Net	0.016	0.020	-0.001
	(4.73)***	(7.27)***	(0.11)
Low Net	-0.011	-0.018	-0.025

50

	(1.88)*	(3.86)***	(5.37)***
*High Net * Eday*	0.140	0.139	0.078
	(6.28)***	(6.21)***	(2.67)***
*Low Net * Eday*	-0.113	-0.111	-0.158
	(5.03)***	(4.96)***	(6.21)***
*High Net * Nday*	0.030	0.031	0.029
	(5.40)***	(5.49)***	(4.81)***
*Low Net * Nday*	-0.042	-0.041	-0.050
	(6.98)***	(6.80)***	(7.88)***
*High Net * Mon*	0.015		
	(2.01)**		
*High Net * Fri*	-0.005		
	(0.78)		
*Low Net * Mon*	-0.022		
	(1.68)*		
*Low Net * Fri*	-0.010		
	(0.88)		
Eday	0.067	0.068	-0.010
	(5.50)***	(5.51)***	(0.70)
Nday	0.126	0.126	0.104
	(20.01)***	(19.80)***	(15.91)***
*High Net * Macro*		-0.019	
		(2.39)**	
*Low Net * Macro*		0.009	
		(0.63)	
*High Net * Ret²*			0.001
			(2.05)**

51

*Low Net * Ret²*			0.001
			(2.01)**
Ret²			0.004
			(10.07)***
Day Fixed Effects?	Yes	Yes	Yes

Table VII

Anomaly Returns and Dynamic Risk on Earnings Announcement Days

In this table we examine whether discount rates and betas change on earnings announcement days, and whether any such effects can explain why anomaly returns are higher on earnings days. We sort firms each month on our aggregate anomaly variable, *Net*, and classify them into high and low quintiles. We create dummy variables for high *Net* stocks and for low *Net* stocks, as well as for earnings announcement days. We examine two sources of risk. *Market* is the daily realization of the CRSP value-weighted return index. *Factor* is the daily realization of a portfolio that is long high *Net* stocks and short low *Net* stocks. We interact both sources of risk with the earnings announcement dummies, and also include 3-way interactions between each of the high and low *Net* variables, each source of risk, and the earnings-day dummy. The sample period is 1979:6 to 2013:12. Standard errors are clustered on time. *t*-statistics are in parentheses.

Table VII (Continued)

	(1)	(2)	(3)	(4)
High Net (H)	0.024		0.024	0.024
	(9.18)***		(13.71)***	(10.14)***
Low Net (L)	-0.016		-0.016	-0.016
	(3.53)***		(9.47)***	(5.06)***
High x Eday	0.147	0.157	0.156	0.148
	(6.52)***	(6.62)***	(7.03)***	(6.61)***
Low x Eday	-0.127	-0.139	-0.139	-0.124
	(5.65)***	(6.07)***	(6.75)***	(5.70)***
Eday (E)	0.108	0.109	0.106	0.107
	(8.64)***	(8.64)***	(8.58)***	(8.75)***
H x Factor (F)				0.291
				(54.87)***
L x Factor (F)				-0.701

				(135.03)***
H x F x E				0.176
				(4.41)***
L x F x E				-0.217
				(4.47)***
Eday x Factor				-0.196
				(7.41)***
H x Market (M)			-0.086	
			(18.94)**	
L x Market (M)			0.298	
			(41.58)**	
H x M. x E			-0.057	
			(2.39)*	
L x M x E			-0.002	
			(0.09)	
Eday x Market			0.166	
			(9.91)**	
Fixed Effects	Day	Day x High x Low	Day	Day

Table VIII

Analysts' Earnings Forecast Errors

In this table, we test whether anomalies are related to analysts' earnings forecast errors. The dependent variable is analysts' earnings forecast error, which is measured as the median earnings forecast minus actual reported earnings (per IBES), scaled by last month's closing stock price. This variable is then winsorized at the 1st and 99th percentiles. We use the median quarterly earnings forecast from the latest IBES statistical period, or the last date that IBES computed its summary statistics for the firm's earnings forecasts. *Number of Estimates* is the number of analysts issuing forecasts. *Single Forecast* is a dummy equal to one if only one analyst makes a forecast for the firm and zero otherwise. *Dispersion* is the

54

standard deviation of the forecasts scaled by stock price. We set dispersion to zero if *Single Forecast* is equal to one. The variables *Long* and *Short* and the different anomaly samples are defined in the previous tables. For readability, we divide *Long* and *Short* by 100. The regressions include time fixed effects. The sample contains 345,913 observations. Standard errors are clustered on time. *t*-statistics are in parentheses.

	Full Sample	Market	Valuation	Fundamental	Event
	(1)	(2)	(3)	(4)	(5)
High Net	-0.045	0.020	-0.007	-0.008	-0.031
	(5.40)***	(1.77)*	(2.69)***	(4.02)***	(12.82)***
Low Net	0.017	-0.031	0.022	0.012	0.020
	(2.36)**	(3.64)***	(7.57)***	(4.40)***	(10.42)***
Number of Estimates	-0.004	-0.003	-0.005	-0.004	-0.004
	(9.61)***	(6.37)***	(10.33)***	(8.49)***	(8.51)***
Single Forecast	0.280	0.270	0.276	0.275	0.278
	(20.54)***	(19.56)***	(20.49)***	(20.28)***	(20.32)***
Dispersion	59.613	59.312	59.148	59.205	59.274
	(19.57)***	(19.38)***	(19.35)***	(19.24)***	(19.45)***
Intercept	0.028	0.025	0.012	0.018	0.031
	(3.39)***	(3.08)***	(1.31)	(2.33)**	(3.17)***
Month FE's?	Yes	Yes	Yes	Yes	Yes

Table IX

Data Mining Tests

In this table we conduct several tests of the hypothesis that anomaly returns can be explained by data mining. To create the *Net* anomaly variable, we use the 97 cross-sectional anomalies studied in McLean and Pontiff (2016). For each stock-month observation, we create *Long* and *Short* by summing the number of long-side and short-side anomaly portfolios that the stock belongs to. *Net* is equal to *Long* minus *Short*. We then merge this monthly data set with daily stock return data from CRSP and with daily indicators for earnings announcement days and Dow Jones News stories, which we refer to as information days. We define an earnings day (Eday) or news day (Nday) as the one-day window around an earnings

55

announcement or news release. *Monthly* is the firm's contemporaneous monthly stock return. Out-of-sample (*OOS*) is like *Net*, but is constructed with anomalies that are out-of-sample, that is, past the sample date of the study that originally documents the anomaly. Columns (3) and (4) report regressions estimated in samples of large and small stocks only, where large (small) stocks are those above (below) the median market capitalization on day t. The control variables include lagged values for stock returns, stock returns squared, and trading volume over each of the past 10 days. The dependent variable, daily return, is expressed in basis points. The sample period is 1979:6 to 2013:12. Standard errors are clustered on time. t-statistics are in parentheses.

	Monthly	Net + Monthly	Large Stocks Only	Small Stocks Only	Out-of-Sample
	(1)	(2)	(3)	(4)	(5)
Net		-0.137	-0.044	0.061	
		(2.23)**	(0.80)	(1.08)	
*Net * Eday*		0.738	0.715	3.067	
		(3.19)***	(2.54)*	(8.24)**	
*Net * Nday*		0.237	0.020	0.241	
		(4.13)***	(0.36)	(2.84)**	
Monthly	5.225	5.227	5.348	5.127	
	(129.53)***	(129.19)***	(96.02)**	(132.29)**	
*Monthly * Eday*	11.620	11.612	11.949	11.125	
	(43.38)***	(43.35)***	(31.12)**	(35.68)**	
*Monthly * Nday*	2.284	2.282	1.867	2.869	
	(25.19)***	(25.17)***	(22.93)**	(18.28)**	
OOS					0.384
					(6.05)***
*OOS * Eday*					2.165
					(8.54)***
*OOS * Nday*					0.179
					(3.03)***
Eday	-0.118	-0.111	-0.006	-0.272	0.090

56

	(8.98)***	(8.55)***	(0.37)	(16.57)**	(8.71)***
Nday	0.060	0.062	0.044	0.078	0.124
	(8.85)***	(9.35)***	(9.65)**	(8.55)**	(18.91)***
Day Fixed Effects?	Yes	Yes	Yes	Yes	Yes

Table IX (Continued)

Table X

Data Mining? Evidence from Analysts Forecast Errors

In this table we test whether anomalies are related to analysts' earnings forecast errors. The dependent variable is analysts' earnings forecast error, which is measured as the median earnings forecast minus actual reported earnings (per IBES), scaled by last month's closing stock price. This variable is then winsorized at the 1st and 99th percentiles. We use the median quarterly earnings forecast from the latest IBES statistical period, or the last date that IBES computed its summary statistics for the firms' earnings forecasts. *Number of Estimates* is the number of analysts issuing forecasts. *Single Forecast* is a dummy equal to one if only one analyst makes a forecast for the firm and zero otherwise. *Dispersion* is the standard deviation of the forecasts scaled by stock price. We set dispersion to zero if *Single Forecast* is equal to one. *Monthly* is the firm's contemporaneous monthly stock return. *High Net* and *Low Net* are the anomaly dummy variables, which are defined in the previous tables. *High OOS* and *Low OOS* are versions of *High Net* and *Low Net* created using out-of-sample anomalies only. For readability, we divide all of the anomaly variables by 100. The regressions include time fixed effects. The sample contains 345,431 observations. Standard errors are clustered on time. *t*-statistics are in parentheses.

	Monthly	Monthly + High Net and Low Net	High OOS and Low OOS
	(1)	(2)	(3)
Monthly	-1.560	-1.557	
	(21.62)***	(21.58)***	
Long		-0.030	
		(3.52)***	
Short		0.013	

57

			(1.76)*
OOS_Long			-0.046
			(5.56)***
OOS_Short			0.016
			(2.22)**
Number of Estimates	-0.004	-0.004	-0.004
	(7.49)***	(8.29)***	(9.64)***
Single Forecast	0.272	0.276	0.279
	(20.46)***	(20.69)***	(20.47)***
Dispersion	58.402	58.440	59.260
	(19.54)***	(19.57)***	(19.54)***
Intercept	0.040	0.043	0.029
	(5.34)***	(5.39)***	(3.57)***
Month Fixed Effects?	Yes	Yes	Yes