Anomalies and News[Ψ]

Joseph Engelberg

R. David McLean

and

Jeffrey Pontiff

September 3, 2016

Abstract

Using a sample of 97 stock return anomalies, we find that anomaly returns are 7 times higher on earnings announcement days and 2 times higher on corporate news days. Anomaly variables also predict analyst earnings forecast errors: analysts' earnings forecasts are too low for anomaly-longs, and too high for anomaly-shorts. We develop and conduct several unique data mining tests, and find that data mining cannot explain our findings. Our results support the view that anomaly returns are the result of biased expectations, which are at least partially corrected upon news arrival.

Keywords: News, cross-sectional return predictability, earnings announcements, market efficiency, biased expectations, expectational errors.

JEL Code: G00, G14, L3, C1.

[Ψ] Engelberg (jengelberg@ucsd.edu) is at UCSD, McLean (rmclean2@depaul.edu) is at DePaul, and Pontiff (pontiff@bc.edu) is at Boston College. McLean is grateful to the Keeley Chair for financial support. We thank Mark Bradshaw, Gene Fama, Juhani Linnainmaa, Peter Nyberg, Philip Strahan, Tuomo Vuolteenaho, and Bohui Zhang for helpful comments, and seminar participants at Auburn, the Asian Bureau of Finance and Economic Research, Berkeley, Boston College, UC Irvine, Georgetown, Michigan State, Cornell, DePaul, Northeastern, UC Riverside, University of Toronto, University of Utah, University of South Carolina, Arizona State, ESCP, Yale, Temple, the Tinbergen Institute, Arrowstreet Capital and the University of Washington, and conference participants at the University of Miami Behavioral Finance Conference. We thank Lauren Vollon for excellent research assistance.

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 research goes back to at least Ball and Brown (1968) and Blume and Husick (1973), yet more than four decades later years later academics still disagree on what causes this return predictability.

There are three popular explanations for cross-sectional predictability. First, predictability could be the result of cross-sectional differences in risk, reflected in discount rates (see Fama (1991, 1998)). In this framework, cross-sectional return predictability is expected because return differences simply reflect ex-ante differences in discount rates that were used to value the stocks. There are no surprises here: what happens with average returns ex-post was *expected* by rational investors ex-ante (e.g., Fama and French (1992, 1996)).

The second explanation comes from behavioral finance, and argues that return predictability reflects mispricing (e.g., Barberis and Thaler (2003)). For example, the marginal investor may have biased expectations of cash flows, and the anomaly variables are correlated with these mistakes among the cross-section of stocks. When new information arrives, investors update their beliefs, which corrects prices and creates the return-predictability.

A third explanation for return predictability is data mining. As Fama (1998) points out, academics have likely tested thousands of variables, so it is not surprising to find that some of them predict returns in-sample, even if in reality none of them do.[1]

Each of these three explanations have different predictions for how firm-specific news arrival affects cross-sectional return predictability. For example, standard risk-based models typically do not predict return differences on news days. On the other hand, behavioral models based on biased expectations typically predict larger anomaly returns when news arrives and corrects the erroneous expectations.

Therefore, in order to differentiate between the explanations of cross-sectional return predictability, we compare predictability on days where firm-specific information is publicly released to days where we do not observe news. The cross-sectional predictors used are the 97 anomaly variables studied in McLean and Pontiff (2016), each of which 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 or days with a Dow Jones news items.

We find that anomaly returns are 7 times higher on earnings announcement days and 2 times higher on Dow Jones news days. We find similar effects on both the long and short sides, i.e., anomaly-shorts have lower returns and anomaly-longs have higher returns on news days. Moreover, anomaly variables predict analyst earnings forecast errors; analysts' earnings forecasts are too low for anomaly-longs,

[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 (2015), McLean and Pontiff (2016), and Roberts and Linnainmaa (2016).

and too high for anomaly-shorts. We discuss how our results relate to each of the three explanations of cross-sectional return predictability below.

Systematic risk. Our main tests include day-fixed effects, so the effects that we document cannot be explained by systematically higher or lower risk for all stocks on news and earnings announcement days. When we examine both the long and short side of anomaly portfolios separately, we find that anomaly returns are 5.9 times *higher* on earnings day for long-side stocks and 10 times *lower* for short-side stocks. If these returns reflect priced risk, then the underlying asset pricing model would require some stocks to have discount rates that are 5.5 times higher on earnings announcement days and other stocks to be 10 times less risky on earnings announcement days. Then, after the announcements, risk would return back to the pre-announcement level.

We also include specifications that model time-varying risk exposure to the market portfolio or to a factor that is based on an aggregate anomaly portfolio. Neither of these specifications changes the tenor of our results. Anomalies also perform worse on days on which macroeconomic news is announced. This is difficult to reconcile with the idea that anomaly returns reflect compensation for consumption risk, which ought to heighten on days when macroeconomic news is expected (Savor and Wilson, 2013).[2]

[2] A common theme in rational expectations models links discount rates to consumption risk, with investors demanding higher returns for assets that perform worse when macroeconomic events reduce consumption choices. Recent examples of such papers include Bansal, Dittmar, and Lundblad (2005), Parker and Julliard (2005), Yogo (2006), Jagganathan and Wang (2007), Hansen, Heaton, and Li (2008), Savov (2011), and Dittmar and Lundblad (2015).

It is well known that stock returns are unconditionally higher on earnings announcement days (see Franzini and Lamont, 2006). This is not the effect that we document; our main specifications control for this through the use of earnings announcement dummy variables, i.e., we find that anomaly-long (anomaly-short) returns are higher (lower) on earnings and news days *after controlling for the fact that stock returns are higher on earnings announcement days*.

Savor and Wilson (2016) develop a model to explain why stock returns are higher on earnings announcement days. In their model, an earnings day return premium occurs because the rational representative agent infers market earnings information from individual earnings announcements, which makes the betas of announcing firms higher. As we mention above, we control for higher stock returns on announcement days, so Savor and Wilson's model does not explain our findings.[3] Moreover, Savor and Wilson (2016) do not have implications for abnormal returns on *unexpected* Dow Jones News days, nor can their framework explain the relation between anomaly variables and analyst forecast errors, which we also document.

Wu, Zhang, and Zhang (2010) and Liu and Zhang (2014) argue that investment-based models imply higher risk premiums on earnings announcement days. Their argument is that in the simplest investment-based model returns are equivalent to return-on-assets, which become known on the earnings announcement day. These models are not equipped to make predictions about

[3] If Savor and Wilson's model were to explain our earnings day results with respect to anomaly returns, it must be that anomaly-longs are highly informative about aggregate earnings, and anomaly-shorts are completely uninformative. This seems farfetched; if anything, we expect stocks on the short-anomaly side, e.g., large stocks, growth stocks, liquid stocks, etc., to be more informative about market-wide earnings.

unexpected non-earnings news or analyst earnings forecast errors. Moreover, investment-production models are still risk-based, and our anomaly variable performs worse when market returns are high and during macroeconomic announcements. Our results are also unaffected by various controls for systematic risk.

Mispricing due to biased-expectations. Prominent 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. To illustrate this intuition, we consider a simple representative agent model (further elaborated in the appendix) with an agent that has biased expectations about future cashflows 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) cashflow expectations have negative (positive) news-day returns. The earnings announcement day and news day returns that we document are consistent with this intuition.

We further assess the impact of biased expectations by examining earnings forecasts of sell-side analysts. If analysts have biased expectations regarding anomaly stocks, then their forecasts should be too optimistic for stocks on the short side of anomaly portfolios and too pessimistic for stocks on the long side of anomaly portfolios. This is precisely what we find; for stocks in the long leg of anomaly portfolios, analysts' forecasts are too low, and for stocks in the short leg, analysts' forecasts are too high.

Data mining. Although our earnings announcement day and news day results are inconsistent with risk-based explanations, they could be consistent with data mining. This is because stocks with high (low) ex-post returns over a given period are more likely to also have high (low) returns on news days and earnings announcement days during the same period. Put differently, stocks with high ex-post monthly returns probably had good news during that month, which explains why the returns were high that month. We show that this is the case; stocks with high (low) monthly returns in month *t*, regardless of their anomaly portfolio membership, tend to also have especially high (low) returns on earnings announcement days and on Dow Jones news days during month *t*. Similarly, we also show that stocks with high (low) monthly returns that also announced earnings in that month had analysts' earnings forecasts that were too low (high) during the same month.

To address the data mining issue 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. We also build an out-of-sample anomaly variable that is constructed solely with out-of-sample anomalies. We show that the returns of out-of-sample anomalies are also higher on earnings days and news days, and that out-of-sample anomaly variables predict analyst forecast errors.

Previous Literature. Our paper builds on previous studies, which show that the returns of specific anomalies are higher on earnings announcement days (e.g., Bernard and Thomas (1989), Ball and Kothari (1991), Chopra, Lakonishok and Ritter (1992), La Porta et al. (1994), Sloan (1996), and Jegadeesh and Titman (1993)). Our findings are also related to Edelen, Ince, and Kadlec (forthcoming), who show that institutions tend to take the wrong positions in stocks that eventually end up in anomaly portfolios, and that such trading activities portend higher anomaly returns in general and on earnings announcement days.

Our paper differs from the previous 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 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 errors. Our forecast error results are important because they are not subject to the joint-hypothesis problem and are in agreement with our news and earnings announcement findings.

Previous studies do not consider how data mining could affect earnings announcement day effects with respect to anomaly returns. We show that spurious anomaly strategies 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, i.e., these papers may be able to rule out risk-based explanations, but they cannot rule out data mining. We deal with Fama's (1998) conjecture by developing the first information day data-mining test, the results of which allow us to rule out the possibility that our results are entirely driven by data mining.

1. Sample and Data

We begin our sample with 97 cross-sectional anomalies studied in McLean and Pontiff (2016). These anomalies 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. All of the variables 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 short side of each anomaly strategy. 16 of our 97 anomalies are indicator variables (e.g., credit rating downgrades). For these cases, there is only a long or short side, based on the binary value of the indicator. We remake the anomaly portfolios each month. As in McLean and Pontiff (2016), the sample selection for each anomaly follows the original study. So, if a study only uses NYSE firms, then we only create that anomaly variable for NYSE firms.

We obtain earnings announcement dates from the Compustat quarterly database. Compustat reports the earnings announcement day, 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 stories 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 1979-2013. These news data are also used in Tetlock (2010, 2011) and Engelberg, Reed, and Ringgenberg (2012). We merge this 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 off information days.

For consistency, we conduct all of our tests during the period 1979-2013, which is the period for which we have news data. We also 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.

1.1. Sample Descriptive Statistics

Table 1 provides some descriptive statistics for our sample, which consists of 40,165,651 firm-day observations for the period 1979-2013. Each observation is in the CRSP daily return database with reported stock returns and a stock price greater than \$5. Among these observations, 16% have Dow Jones news stories, while 1.2% have earnings announcements reported in Compustat.

There is overlap between the news days and the earnings announcement days. Of the 489,966 earnings announcement days, 256,745, or 52%, are also Dow Jones news days. This is, however, a small percentage of the total news days. The total number of news days is 6,453,258 so only 4% of these are also earnings announcements that are reported in Compustat. It could be that Dow Jones stories cover a significant number of earnings announcements not covered in Compustat, so 4% is a lower bound on the percentage of news stories that likely reflect earnings announcements. Table 2 provides descriptive descriptions of the portfolio variables.

3. Main Results

3.1 Anomaly Returns On and Off Information Days

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

$$R_{i,t} = \alpha_t + \beta_1 Net_{i,t} + \beta_2 Net_{i,t} \times Eday_{i,t} + \beta_3 Net_{i,t} \times Nday_{i,t} + \beta_4 Eday_{i,t} + \beta_5 Nday_{i,t} + \sum_{j=1}^{10} \gamma_j Lag\ Return_{i,t-j} + \sum_{j=1}^{10} \delta_j Lag\ Return^2{}_{i,t-j} + \sum_{j=1}^{10} \rho_j Volume_{i,t-j} + \varepsilon_{i,t} \quad (1)$$

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. $Net_{i,t}$ is our aggregate anomaly variable, which we describe in more detail below. *Net* is measured at the beginning of each month and returns are measured 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. Returns are multiplied by 10,000. Thus, each unit of return is one basis point.

The variables *Eday* and *Nday* are dummy variables equal to 1 on earnings and news days for firm *i* and zero otherwise. Our hypotheses are tested with the interaction term: i.e., are anomaly returns higher on information days?

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

To construct *Net* for each firm-month observation we sum up the number of long side (*Long*) and short side (*Short*) anomaly portfolios that the observation belongs to. *Net* is the difference between *Long* and *Short*: *Net* = *Long* – *Short*. Summary statistics for *Net, Long,* and *Short* are provided in Table 2. The average stock is in 8.61 long portfolios and 9.23 short portfolios. If the portfolios were solely based on 97 random quintile groupings, we would expect long and short to equal 19.4 (97 x 0.20). Our counts are lower since some characteristics are indicator variables. Thus, they lack either a long or short side and, following the original study, some characteristics are only constructed for a subset of stocks (for example, NYSE stocks). For characteristics that are subset based, stocks that fall out of the subset are not assigned to a long or short side. The mean value for *Net* is -0.61, the maximum is 32, and the minimum is -36.

With respect to the above regression equation, market efficiency (in the absence of data mining and changes in risk exposure) suggests that the interaction terms should be zero: i.e., anomaly returns should not be any stronger on information days as compared to other days. This is because, in the rational expectations framework, return-predictability is explained by ex-ante differences in discount rates, which should not change in a predictable manner on firm-specific information days.

In contrast, the biased expectations framework suggests that the coefficient for the interaction between *Net* and the earnings and news day dummies should be positive, or that anomaly returns should be greater when new information is released. This is because, in the biased expectations framework, return predictability is the result of ex-post releases of information that cause investors to update their expectations, which were systematically biased ex-ante.

Panel A of Table 3 reports the regression results. As mentioned above, returns are expressed in basis points. In Panel A, we define the information day as a 1-day window, while in Panel B we use a 3-day window: i.e., days t-1, t, and t+1.

The first regression presents results that do not include the lagged volume, lagged return, and lagged squared return controls, whereas all the other regressions include these controls. Since our estimation uses millions of observations, omission and inclusion of correlated variables may cause drastic changes in statistical influence. A comparison of the first two regressions shows that this is not the case. The controls absorb variation, in that the standard errors in the second specification shrink slightly, but the slope coefficients remain similar. In the second regression in Panel A, the *Net* coefficient is 0.259, while the *Net x Earnings Announcement* interaction coefficient is 1.980. Taken together, the coefficients show that for a *Net* value of 10 (about 1½ standard deviations) expected returns are higher by 2.59 basis points on non-earnings announcement days, and by an additional 19.8 basis points on earnings announcement days. Put differently, anomaly returns for a *Net* value of 10 are in total 22.39 on earnings announcement days, which is 8.6 times higher than anomaly returns on non-earnings announcement days. The *Net x News Day* interaction coefficient is 0.317, showing that anomaly returns are 2.2 times higher on news days that are not also earnings announcement days, which is also a sizeable effect. The coefficients show an incremental unit of *Net* is associated with an extra 2.556 basis points of return, on earnings days that are also Dow Jones news days. This is almost 10 times higher the estimate on non-news days. All of the coefficients are significant at the 1% level.

In the third regression reported in Panel A, we replace the day-fixed effect with a day-information event fixed effect. That is, for a given day t, all of the firms with news or earnings announcements share one intercept and all of the firms without news or earnings announcements share another. In this regression, the comparison is between two firms that both have a news story or earnings announcement on the same day, but have different values of *Net*. The coefficients in this regression are very similar to those in the second regression. The *Net* coefficient is 0.254, while the earnings day and news day interactions are 1.967 and 0.0392 respectively.

In the next few regressions, we dig deeper into the idea that systematic risk can explain anomaly returns. Savor and Wilson (2016) show that exposure to systematic risk increases on earnings release days. We therefore consider specifications that model day-specific changes in risk exposure. We add either an anomaly factor (*Factor*) or a market portfolio factor (*Market*) to our regressions, and interact each with the information day dummies*.* *Factor* is the daily long-short portfolio return for a portfolio that is long in the top 20% percentile of *Net* and short in the bottom 20% percentile of *Net*. The coefficients for *Factor* and *Market* reflect the average stock's beta with respect to each portfolio. Following Shanken (1990), interactions are used to broaden our specification to consider time-series and cross-sectional variation in beta. The information day interactions with *Factor* and *Market* show whether these betas increase on information days. Finally, we include an interaction between *Net* and *Market*, which allows market beta to vary based on *Net*.

In the regression reported in the fourth column, the coefficient for *Factor* is -0.931 and statistically significant. Hence, when *Factor* has high returns, expected stock returns are lower. We discuss this negative coefficient in detail below following our discussion of the other regressions reported in Panel A. The coefficient for the interaction between *Factor* and the earnings announcement day dummy is 0.030 and not significant. The interaction between *Factor* and the news day dummy is -0.461 and significant. This means that if a stock has a news announcement, its exposure to *Factor* becomes even more negative. Note that this is the opposite result that we find with *Net*; if a stock has a higher value of *Net*, its expected return is higher and this effect is greater on news days. The results with *Factor* therefore contradict the idea that the *Net* results are explained by covariance with some underlying risk; including *Factor* does not affect the *Net* coefficient, and the *Factor* and *Net* coefficients are of the opposite sign.

The regression in the fifth column replaces *Factor* with the CRSP value-weighted index minus the risk free rate as a proxy for the market portfolio. This specification tells us whether controlling for market risk changes our inferences. The coefficients for *Net*, the *Net* earnings day, and news day interactions are still 0.249, 1.992, and 0.236 respectively. These values are similar to those reported in regression 2. Thus, the results for *Net* cannot be explained entirely by market risk. The coefficient for the market portfolio is 0.737, whereas the market portfolio earnings day and news day interactions are 0.033 and 0.319, respectively. These results make sense; the average beta is close to one, and beta increases earnings announcement and news days.

The sixth regression is like the fifth, but it includes an interaction between *Net* and the market portfolio and a three-way interaction between *Net*, the market portfolio, and the earnings announcement and news dummies. The interaction between *Net* and the market portfolio is negative and significant; stocks with higher values of *Net* have lower market beta, i.e., they are less risky. The results show that for every unit increase in *Net*, market beta falls by -0.023. Moreover, this effect is greater on earnings announcement days; for every unit increase in *Net*, market beta falls by an additional -0.003 on earnings announcement days, although the coefficient is insignificant. With respect to news days, market beta increases by 0.004 for each unit increase in *Net*.

Returning to the negative coefficient on *Factor* reported in regression 4, this follows from two effects. First, previous literature finds that long-anomaly stocks tend to have lower betas than short-anomaly stocks (we also show this in regression 5, i.e., beta declines with *Net*).[4] The difference in betas encourages *Factor* to be negatively correlated with the market. Second, our estimation is equal-weighted, so the estimation is sensitive to the performance of small market capitalization stock returns in excess of large market capitalization stock returns. If we control for this effect the slope on *Factor* becomes positive and marginally significant. In untabulated results, we estimate a regression that includes *Factor*, the value-weighted market portfolio, and a portfolio that reflects the difference between the value-weighted and equal-weighted market portfolios. In this specification the coefficient for *Factor* is positive. The coefficients for *Net* and its interactions with earnings day and news day dummies are virtually unchanged; they are the same as those reported throughout Table 3, so nowhere does any type of risk control seem to affect *Net* or its information-day interactions.

The results in Panel B, which study news and earnings announcement returns over 3-day windows, are similar. The information day coefficients are smaller as compared to Panel A, which is to be expected because Panel B uses 3-day

---

[4] Examples include; higher book-to-market and higher earnings-to-price stocks have low betas (Fama and French, 1992), higher momentum stocks have lower betas (Jegadeesh and Titman, 1993), higher idiosyncratic stocks earn lower returns despite the fact that they have higher betas (Ang, Hodrick, Xing, and Zhang, 2006), and firms that repurchase shares experience positive abnormal returns and diminished betas (Grullon and Michaely, 2004).

windows. Yet, there are still significantly higher returns on information days and these effects and are unchanged in the presence of various fixed effects and controls for market factor and market risk.

The coefficients reported in both panels document substantially higher returns on both earnings days and news-not-earnings days. The earnings day result is consistent with Franzini and Lamont (2006). We do not know of previous research that has documented our news-not-earnings day finding—such news days are also associated with positive stock price reactions.

3.2. Estimating Separate Long and Short Anomaly Effects

In Table 4, we remove the *Net* variable from the regressions and replace it with *Long* and *Short*, which, as we explain above, are the sums of the number of long-side and short-side anomaly portfolios that the stock belongs. Using *Long* and *Short* separately allows us to examine whether the effects of information are different for the long and short sides of anomalies. We use the lagged controls described in the previous section in both of the regressions reported in Table 4 along with day fixed effects.

The first regression in Table 4 uses the 1-day announcement window. In this regression, the *Long* coefficient is 0.396, while the *Long x Earnings Announcement* interaction coefficient is 2.192, showing that long-side anomaly returns are 554% higher on earnings announcement days. The news day interaction is 0.048 and not significant. Hence, on the long side, the effects are largely from earnings announcements.

The effects on the short side are even stronger. The *Short* coefficient is -0.193, while the *Short x Earnings Announcement* interaction coefficient is -1.962, showing that the incremental impact of short anomalies on earnings announcement days is more than 10 times that of a typical day. The news day interaction is -0.605 and highly significant.

Various authors (for example, Miller, 1977) argue that if short-selling imposes extra-costs on short sellers, overvaluation situations will be more frequent than undervaluation situations. On the surface, the symmetry of the long and short interactions runs counter to such an argument. The overall effect is that, on earnings days that are also news days, the overall short coefficient is -0.193 + -1.962+ -0.605 = -2.760, whereas the overall long coefficient is 0.396 + 2.192 + 0.048 = 2.636. One reason is that short-specific costs are holding costs, and thus proportional to holding period length (Pontiff, 1996). In this case, we expect the incremental costs of shorting around earnings announcements and other expected news days to be minor.

In column 2, we replace the 1-day window with a 3-day window for the news and earnings announcements. The results are similar. The magnitudes are smaller, which is to be expected with the longer window, however, the signs and significance of the coefficients are unchanged.

Taken together, the results in Tables 3 and 4, and Figure 1 are consistent with the idea that mispricing and, specifically, biased expectations play an important role in explaining cross-sectional return predictability. The long side of anomaly strategies tends to do especially well on days when new information is released, whereas stocks on the short side have especially low returns on days when information is released. Investors seem to be expecting too much from the short side firms and too little from the long side firms.

The results here are very different than what we should observe in an efficient market where investors have rational expectations. In the rational expectations world, cross-sectional differences in stock returns are explained by cross-sectional differences in expected returns.

3.3 Do the Effects vary Across Anomaly Types?

In this section of the paper, we ask whether the type of information used to create the anomaly affects the results in the previous section. McLean and Pontiff (2016) categorize anomalies into four different types: (i) Event; (ii) Market; (iii) Valuation; and (iv) Fundamentals. The categorization 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 constructed with financial statement data and nothing else. Leverage, taxes, and accruals are fundamental anomalies.

We construct the same *Net* variable as before, only we sum up the portfolio memberships within each of the four groups. As in the previous tables, the regressions include time fixed effects, the lagged control variables used in the previous tables, controls for the market factor interacted with information day, and standard errors clustered on time.

We report the results from these tests in Table 5. Panel A reports the results from the regression, while Panel B reports the results from linear restriction tests that compare the effects among the four anomaly types.

The regression in Panel A shows that all four of the anomaly types have significantly higher returns on earnings announcement days. Hence, the results in the previous tables are not driven by a few anomalies or just one type of anomaly; instead, the effects are common across all types of anomalies. With respect to news days, 3 of the 4 anomaly types have positive and significant interactions. Fundamental anomalies have a negative and significant interaction. The coefficient for fundamental anomalies is 0.001, whereas the news day interaction is -0.004.

Panel B tests whether the interactions vary across the anomaly types. One salient result is that market anomalies, which are based solely on prices, returns, variance of returns, and trading volume, have the lowest earnings day effects but the highest news day effects. Valuation anomalies, which are based on ratios of price to fundamentals, have the highest earnings day effects, although the difference relative to fundamental anomalies is not statistically significant. These findings suggest a possible relation to Stein (2009), which bifurcates trading strategies of sophisticated traders into strategies that are "anchored" in fundamental information and "unanchored" in market-based information. Tables 6 shows anchored, fundamental-based strategies have sharp increases in predictability when earnings are released and minor or no increases in predictability when non-earnings news is released. Unanchored, market-based strategies, have a modest increase on earnings release days, and pronounced increases on general news days. Although Stein's model is stylized and not designed to predict our findings, this suggests a role for future research.

3.4 Robustness: Day of the Week, Macroeconomic News, and the Endogeneity of News

In this section of the paper we explore alternative explanations for our findings. Specifically, we ask whether day of the week effects, macroeconomic news announcements, and the possibility that extreme returns cause news (rather than the other way around) can explain some or all of our findings.

Day of the Week Effects. Birru (2016) 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 (2016) argues that these patterns are consistent with studies in the psychology literature, which show that mood increases from Thursday to Friday and decreases on Monday.

In order to test whether such day of the week effects influence our results we estimate a specification where we interact the *Long* and *Short* 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 5. 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 5 are very similar to those reported in column 1 of Table 4, which is the same regression but excluding the Monday and Friday interactions.

We do not classify our anomalies into speculative and non-speculative legs like Birru does, so our results may not be directly comparable to his, however we do find evidence of day of the week effects with our anomaly variables. Both *Long* and *Short* perform better on Mondays, and *Long* performs worse on Friday. 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 has not been shown in the literature. On explanation for the Monday effect is that there is more information impounded into prices on Monday. News is released over the weekend, however investors cannot trade until Monday. Hence, Monday itself can be thought of as a type of "news day" indicator. These results are therefore consistent with the other findings in our paper, i.e., anomalies perform better on days in which new information gets incorporated into prices.

Macroeconomic News. Savor and Wilson (2013) find that market returns are higher on days for which macroeconomic news about inflation unemployment, or interest rates is scheduled for announcement. They argue that their results reflect compensation for higher risk that is associated with such announcements. Their story is that although investors do not know what the news will be, they do know when there will be news, as these macroeconomic announcements are scheduled well in advance, and the dates are public. If asset prices are affected by macroeconomic news, then the risk associated with holding securities will be higher around macroeconomic announcements, and rational investors should anticipate this effect. Thus, stock returns should be higher on days which macroeconomic news is announced.

In order to control for the effects of macroeconomic news we estimate a specification that interacts the *Long* and *Short* anomaly variables with a macro news dummy. The macro news dummy is the same variable that is used in Savor and Wilson (2013). The macro news dummy is equal 1 if there is a scheduled announcement regarding inflation, employment, or interest rates, and zero otherwise. We report the results from this specification in the second column in Table 5. 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 in Table 5 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, i.e., short positions have negative alphas on macro news days. These findings are consistent with the results in Table 3, which show that anomaly returns are lower when market returns are higher, i.e., the aggregate anomaly variable that we use produces a strategy with a negative beta. These findings make it very difficult to explain anomalies with a risk-based framework, as anomaly returns are lowest on days which risk premiums should be highest.

Endogeneity of News. It could be the case that extreme returns cause news stories. If this is the case, the anomaly-news day interactions we document might not reflect news being impounded into asset prices, but instead reflect news stories being 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 *Long* and *Short* with the contemporaneous daily stock return squared. The slope on the interaction measures whether *Long* and *Short* perform differently on extreme return days.

We report the results from this test in the third column of Table 5. 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.

3.5. What Portion of Abnormal Returns are Earned on Information Days?

In this section, we decompose each anomaly's return into returns earned on information days and returns earned on non-information days. In Panel A, we define an information day as the 3-day window around either an earnings announcement or news story. This decomposition allows us to place a lower bound on the importance of information releases. As we explain before, this is a lower bound because it is well-documented that earnings announcements are persistent and produce drifts in stock returns, and because there can be information about the firm that is released but not covered by Dow Jones.

For each firm-day observation, we first measure the firm's abnormal return as the firm's return minus the value-weighted market return on the same day. Then, for each anomaly portfolio, we sum up all of the abnormal returns on information days and on non-information days separately. We also count the number of information firm-days and the number of non-information firm-days in each anomaly portfolio. This exercise allows us to say what percentage of an anomaly's return is earned on information days and what percentage is earned on non-information days.

As an example, consider an anomaly that over our sample period has 1,000 firm-day observations in total. Assume that 300 of these are information days. Assume that the abnormal firm-day returns in total sum to 5,000 basis points; 3,000 of which are earned on information days and 2,000 of which are earned on non-information days. This allows us to state that, for this anomaly, information days account for 30% of the total days and 60% of the total returns. We conduct this exercise of each of the anomaly portfolios in our sample and report the averages in Table 6.

We report results for the full 97-anomaly samples and for the four anomaly types. With respect to the full 97-anomaly sample, information days account for 34.5% of the firm-days on the long side and 80.1% of the returns. The results are similar on the short side. Information days account for 34.6% of the firm-days and

84.8% of the returns. These results are consistent with the previous tables.

The results are robust across the four different anomaly types. Among the anomaly types, the results are strongest for the market anomalies. Within this group of anomalies, on the short side, information days account for 33.6% of the firm day returns and 107.7% of the market returns. The price, bid ask spreads, volume, and Amihud illiquidity measure anomalies drive this effect, as the long side returns for these anomalies are almost entirely explained by returns on information days. This result is not salient in Tables 3 and 4, which reports results from tests that use an aggregate anomaly variable that mutes the effect of any single anomaly.

In Panel B we report results for the 3-day windows around earnings announcements only. We limit the sample to firms that have 4 earnings announcements days (in our data) during the year. The results are generally stronger than those in Panel A. On the long side earnings days are only 4.9% of the sample, but account for 17.2% of the returns. Similarly, on the short side, earnings days are 4.8% of the sample and account for 17.7% of the returns.

3.6. Analysts Earnings Forecast Errors

In this section of the paper we ask whether our anomaly variables predict analyst earnings forecast errors. The results thus far suggest that news day returns are inconsistent with risk-based asset pricing. We find 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 proxy for the expectations of investors, then analysts' 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 from IBES. It is the difference between a stock's last reported median sell-side forecast and the actual reported earnings (per IBES), divided by the closing stock price in the previous month. This variable is then winsorized at the 1st and 99th percentiles. We have data from IBES for the period 1983 through 2014. The biased expectations framework predicts that this variable will be negative for the long-side stocks (forecast too low) and positive for the short-side stocks (forecast too high). We merge the forecast data with our anomaly data and test whether anomaly portfolio membership can predict forecast error.

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 stock price. If there is only a single forecast, we set the standard deviation of the forecast equal to zero. We also include time fixed effects and cluster our standard errors on time. We do not include firm-level controls because the firm level variables that we would include are also anomalies (e.g., size, price, book-to-market).

We report the results from these tests in Table 8. We multiply the forecast error variable by 100 so that the coefficients are easier to read. We also divide *Long* and *Short* by 100. The first regression reports the findings for the full 97-anomaly samples. 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. These results share similarities with Edelen, Ince, and Kadlec (forthcoming), who show that earnings day anomaly returns are more pronounced when institutional investors are underinvested in anomaly-longs and overinvested in anomaly-shorts.

The effects are economically significant too. Our earnings forecast error variable has a mean value of 0.123 (not in tables). Table 2 shows that *Long* and *Short* have standard deviations of 5.07 and 5.94. Combining these statistics with the coefficients in Table 8, we see that a one standard deviation increase in *Long* results in a -0.043 decrease in expected earnings forecast error, whereas a one standard deviation increase in *Short* leads to a 0.032 increase in expected earnings forecast error.

Table 8 also reports the effects across the 4 anomaly groups. We see that in all four groups, the *Long* variable is negative and significant, showing that analysts' expectations are too pessimistic for firms in all types of short side anomaly portfolios. With respect to the *Short* variable, it is positive and significant for three of the anomaly groups, but insignificant for the market anomaly group. As we explain earlier, market anomalies include variables that are constructed only with market data, and include momentum, reversal and idiosyncratic risk.

Taken in their entirety, the results in Table 8 largely agree with the results in the other tables. Investors and analysts seem to be too pessimistic (optimistic) about the future earnings stocks in the long (short) side of anomaly portfolios. This bias is revealed in stock returns when firms announce earnings and other news, and in analysts' forecast errors.

3.7. Can Data Mining Explain Cross-Sectional Return Predictability?

Fama (1998) and Harvey, Lin, and Zhu (2014) stress 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 non-information days, and Dow Jones news days have a return standard deviation that is 30% greater than non-information days. Given that returns are so much more volatile on information days, an anomaly that is the result of data-mining would likely perform especially well on information days. We therefore conduct several different 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 test whether it is the case that a data-mined strategy performs especially well on information days. We test whether any firm with a high (low) return on month *t,* regardless of it being in an anomaly portfolio, would also have high (low) information day returns in month *t*. We test for this effect via the following regression equation:

$$R_{i,t} = \alpha_t + \beta_1 Monthly_{i,t} + \beta_2 Monthly_{i,t} \times Eday_{i,t} + \beta_3 Monthly_{i,t} \times Nday_{i,t} + \beta_4 Eday_{i,t} + \beta_5 Nday_{i,t} + \sum_{j=1}^{10} \gamma_j Lag\ Return_{i,t-j} + \sum_{j=1}^{10} \delta_j Lag\ Return^2{}_{i,t-j} + \sum_{j=1}^{10} \rho_j Volume_{i,t-j} + \varepsilon_{i,t} \qquad (2)$$

The above equation is essentially the same as equation (1), only we replace

Net with *Monthly*, which is the contemporaneous monthly stock return. The dependent variable is the daily stock return. The coefficient for *Monthly* will be positive and significant, i.e., firms with higher stock returns in a month also have higher stock returns during the days of that month. The interactions then 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 (2) are reported in the first column of Table 8. The results show that it is the case that when monthly returns are high (low) information day returns during that month are especially high (low). The coefficient for *Monthly* is 0.058, showing that a firm with a stock return of 10% in a given has an expected daily return of 0.58% 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 2.6. Hence, a data-mined strategy would also have extreme returns on information days, meaning that these types of test are not sufficient for ruling out data mining.

In column 2 of Table 8 we add *Net* and the interactions between *Net* and the earnings day and news day dummies along with *Monthly* and its interactions. The *Net* coefficient is negative and significant in this regression. Thus, after controlling for monthly returns, high *Net* stocks have lower expected returns on non-information days. The *Net* interactions with both the earnings day and news day dummies are positive and significant, showing that even after controlling for the effects of *Monthly* it is still the case that anomalies perform especially well on information days. This contradicts the idea that anomalies can be explained by data mining alone.

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

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

We report the results for *OOS* in column 3 of Table 8. This specification is like the specification defined in Equation (1), only *OOS* replaces *Net*. The results for *OOS* are similar to those with *Net*. Using *OOS*, we estimate that anomaly returns are 7 times higher on earnings announcement days, and more than 50% higher on Dow Jones News days. Given that *OOS* is built entirely from out-of-sample anomalies, it seems unlikely that these findings can be explained by data mining.[5]

Firm Size. A number of studies show that anomalies tend to be stronger in

[5] *OOS* and *Net* have a correlation of 0.76, so it reasonable that the two variables produce similar results.

small firms, illiquid firms, and firms with high idiosyncratic risk (see Pontiff, 1996, and Pontiff, 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 media market capitalization on day *t*, and estimate Equation (1) within each sample.

We report these size-partitioned results in columns 4 and 5 of Table 8. In column 4, which reports the results for large stocks, the *Net* coefficient is 0.289, and the *Net* earnings day interaction is 1.153, showing that anomaly returns are about 4x higher on earnings day among large stocks. In column 5, which reports the results for small stocks, the *Net* coefficient is 0.324, and the *Net* earnings day interaction is 3.843, showing that anomalies are almost 13 times higher on earnings day small large stocks. Similarly, the news day interaction is insignificant among large stocks, but shows an almost 3 times increase in anomaly returns among small stocks. What these results also show is 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), 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 arbitrage costs are greater among small stocks, under the mispricing theory of anomaly returns we would expect news to lead to larger corrections of mispricing for small stocks because there is more mispricing to correct.

3.7.1. Data Mining and Analyst Forecast Errors

The results in Table 8 show that analysts' earnings forecasts are too low for anomaly-longs and too high for anomaly-shorts, which is consistent with the idea that expectational errors 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 had positive earnings surprises and short stocks that have negative earnings surprises. It would be difficult to generate abnormal returns otherwise.

To test this hypothesis, we re-estimate the analyst forecast error regression reported in column 1 of Table 8, but use *Monthly* in place of Long and Short. These results are reported in column 1 of Table 9. Unlike the previous tables, the monthly return is not expressed in basis points, rather the normal convention is used, such that a unit return is 100%. The coefficient for *Monthly* is negative and significant, showing 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 error.

To control for this data-mining effect, we estimate a specification that includes *Monthly* along with *Long* and *Short*. The results for this specification are reported in column 2 of Table 9. As in column 1, *Monthly* is negative and significant, however *Long* is also negative and significant, and *Short* is positive and significant. These results are inconsistent with data-mining generating earnings forecast errors predictability by anomaly variables.

We further explore the possibility of data mining by replacing *Long* and *Short* with *OOS_Long* and *OOS_Short*. 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 9. The coefficient for *OOS_Long* is negative and significant, and the coefficient for *OOS_Short* is positive and significant. Hence, even for 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.

Conclusions

Evidence of cross-sectional return-predictability goes back more than four decades, yet to this day academics disagree about the cause. In this paper, we compare return predictability on news and non-news days, and provide evidence that is consistent with return predictability being caused by mispricing, and in particular, mispricing caused by biased expectations. Our findings are consistent with investors who 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, in turn, cause prices to change, which, in turn, causes the observed return predictability. Evidence from sell-side equity earnings forecasts dovetail with the stock return evidence: analysts overestimate the earnings for firms on the short-side of anomaly portfolios and underestimate earnings for firms on the long-side.

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 & Litov, Lubomir & Wachter, Jessica A. & Wurgler, Jeffrey, 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, N. and R. Thaler, 2003, “A Survey of Behavioral Finance,” in George Constantinides, Milton Harris, 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 and Thomas, 1989, Post-Earnings Announcement Drift: Delayed Price Response or Risk Premium, Journal of Accounting Research 27, 1-36.

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

Bonferroni, C. E. 1935, "Il calcolo delle assicurazioni su gruppi di teste." In Studi in Onore del Professore Salvatore Ortu Carboni. Rome: Italy, pp. 13-60.

Chopra, N., J. Lakonishok, and J. 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 marker overreaction and seasonality”, Journal of Finance 42, 557-81.

De Long, J.B., Shleifer, A., Summers, L.H., Waldmann, R.J., 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.

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.

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

Edelen, Roger, Ozgur Ince, and Gregoy Kadlec, forthcoming,"Institutional Investors and Stock Return Anomalies," Journal of Financial Economics.

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

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

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

Fama, E., and K. R. French, 1996, Multifactor Explanations of Asset Pricing Anomalies, Journal of Finance 51, 55–84.

Fama, E. 1998, Market Efficiency, Long-term Returns and Behavioral Finance, Journal of Financial Economics 49, 283-306.

Foster, G., C. Olsen, and T. Shevlin. 1984, Earnings Releases, Anomalies, and the Behavior of Security Returns, The Accounting Review, 574-603.

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

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, 2014, "… and the cross-section of expected returns," Review of Financial Studies, Forthcoming.

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.

Lakonishok J., A. Shleifer, and R. Vishny, 1994, Contrarian Investment, Extrapolation, and Risk, Journal of Finance 49, 1541-1578.

La Porta, R., J. Lakonishok, A. Shleifer, and R. Vishny. 1997. Good news for value stocks: further evidence on market efficiency. Journal of Finance 52, 859-874.

Liu, L. X., and L. Zhang, 2014, A neoclassical interpretation of momentum, 2014, 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," Annual 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, NJ: Princeton University Press.

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.

Savov, 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.

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 table reports the coefficients from regressions of daily returns on the aggregate anomaly variables *Long* and *Short*, dummies for 3-day windows around earnings announcements, interactions between *Long* and *Short* and the 3-day window dummies, and day fixed effects. Returns are expressed in basis points. *Long* and *Short* are defined in Table 2. The Figure plots the sum of the coefficients for the interactions and the coefficients for *Long and Short*, i.e., we plot the overall effect of *Long* and *Short* for each of the seven different 3-day windows.

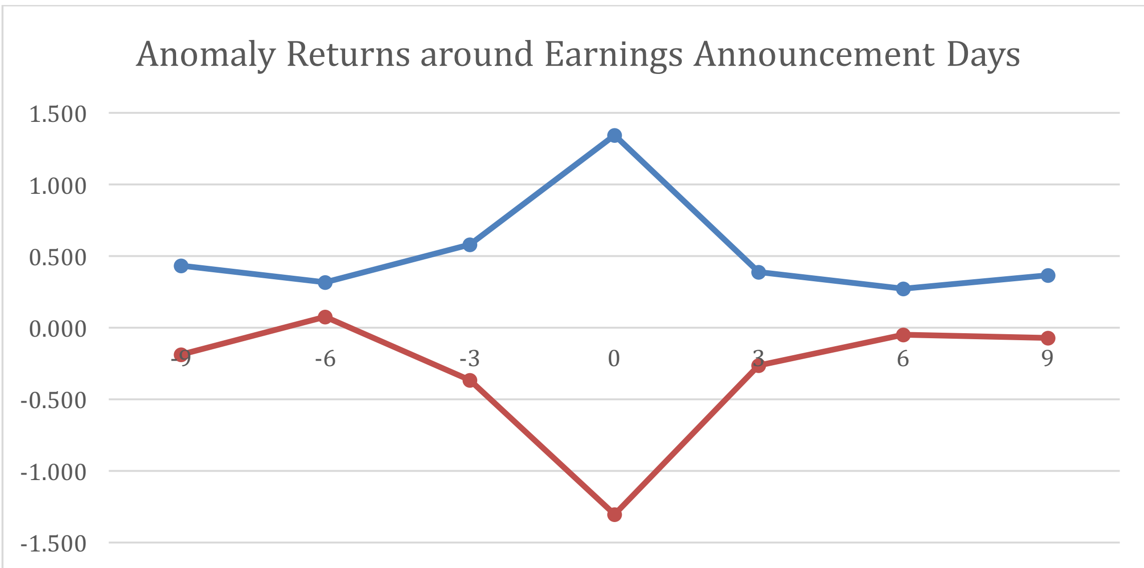

Table 1: Earnings Announcement and News Data

This table describes our sample in terms of earnings announcements and news releases. The unit of observation is at the firm-day level. To be included in our sample, a stock must have return data reported in both the CRSP monthly and daily stock returns databases, and have a stock price that is at least $5. We obtain earnings announcement dates from the Compustat quarterly database, and news announcements from the Dow Jones news archive. We define an earnings day or news day as the day of an earnings announcement or Dow Jones news release. If the announcement is made after hours then the following day is the event day. The sample period is from 1979-2013.

Number of Firm-Day Returns

	News Day		Total
Earnings Day	No	Yes	
No	33,510,434	6,223,007	39,733,441
Yes	256,745	230,251	486,996
Total	33,767,179	6,453,258	40,220,437

Percentage of Firm-Day Returns

	News Day		Total
Earnings Day	No	Yes	
No	83.4%	15.4%	98.8%
Yes	0.6%	0.6%	1.2%
Total	84%	16%	100%

Table 2: Descriptive Statistics for the Portfolio Variables

This table provides descriptive statistics for the anomaly variables. 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 long- and short-side of each anomaly strategy. 16 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. We remake the anomaly portfolios each month. For each firm-day observation, we sum up 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

Variable	Observations	Mean	Std. Dev.	25^{th} %ile	75^{th}%ile	Min	Max
Long	40,220,437	8.61	5.07	5	45	0	35
Short	40,220,437	9.21	5.93	4	13	0	45
Net	40,220,437	-0.61	6.10	-4	4	-36	32

Table 3: Anomaly Returns on Information Days vs. Off Information Days

This table reports results from a regression of daily returns on time-fixed effects, the *Net* anomaly variable, an information-day dummy variable, interactions between the *Net* and the information-day variables, and control variables (coefficients unreported). Returns are 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 sum up the number of long-side and short-side anomaly portfolios that the stock belongs to, thereby creating *Long* and *Short*. *Net* is equal to *Long* minus *Short*. We then merge this monthly dataset 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 1-day or 3-day window around an earnings announcement or news release, i.e., days t-1, t, and t+1. *Factor* is the returns of a portfolio that is long the stocks in the highest quintile of *Net* and short the stocks in the lowest quintile of *Net*. *Market Portfolio* is the return of the CRSP value-weighted portfolio. The sample period is from 1979-2013 and the sample contains 39,860,610 observations. The standard errors are clustered on time.

Table 3: (Continued)

	Panel A: 1-day Window					
Net	0.253	0.259	0.254	0.242	0.249	0.366
	(6.35)***	(6.97)***	(6.89)***	(6.28)***	(6.69)***	(13.22)***
*Net * Eday*	1.946	1.980	1.967	1.992	1.992	2.031
	(11.82)***	(12.11)***	(10.60)***	(12.29)***	(12.24)***	(13.01)***
*Net * Nday*	0.311	0.317	0.392	0.236	0.236	0.184
	(5.53)***	(5.77)***	(5.05)***	(3.33)***	(4.26)***	(4.02)***
Eday	20.7	20.2	21.6	19.9	20.7	20.7
	(20.01)***	(19.33)***	(11.90)	(12.14)***	(17.74)***	(17.76)***
Nday	14.5	15.0	19.8	0.118	10.6	10.6
	(22.12)***	(23.35)***	(9.80)	(7.53)***	(17.55)***	(17.78)***
Factor				-0.931		
				(38.02)***		
*Factor * Eday*				0.030		
				(0.61)		
*Factor * Nday*				-0.461		
				(11.20)***		
Market					0.737	0.726
					(114.86)***	(112.69)***
*Market * Eday*					0.033	0.030
					(2.04)**	(1.88)*
*Market * Nday*					0.319	0.309
					(34.38)***	(31.44)***
*Net * Market*						-2.340
						(31.99)***
*Net * Mrkt. * Eday*						-0.272
						(1.45)
*Net * Mrkt. * Nday*						0.438
						(5.20)***
Lagged Controls?	No	Yes	Yes	Yes	Yes	Yes
Fixed Effects?	Day	Day	Day * Event	None	None	None

Table 3 (Continued)

	Panel B: 3-day Window					
Net	0.238	0.245	0.246	0.238	0.240	0.363
	(6.31)***	(6.89)***	(6.95)***	(6.38)***	(6.73)***	(13.22)***
*Net * Eday*	0.984	1.020	0.935	1.080	1.046	1.053
	(11.04)***	(11.58)***	(11.87)***	(12.23)***	(11.93)***	(13.19)***
*Net * Nday*	0.188	0.197	0.225	0.072	0.123	0.073
	(3.82)***	(4.23)***	(4.74)***	(1.47)	(2.66)***	(2.10)**
Eday	8.2	8.4		7.0	8.3	8.5
	(15.41)***	(15.23)***		(5.46)***	(11.73)***	(11.73)***
Nday	9.8	10.3		7.9	6.5	6.7
	(18.29)***	(19.36)***		(6.46)***	(12.59)***	(12.64)***
Factor				-0.886		
				(36.64)***		
*Factor * Eday*				0.083		
				(1.93)*		
*Factor * Nday*				-0.414		
				(11.81)***		
Market					0.705	0.696
					(105.28)***	(104.63)***
*Market * Eday*					-0.004	-0.005
					(0.33)	(0.44)
*Market * Nday*					0.293	0.282
					(33.31)***	(32.69)***
*Net * Market*						-2.351
						(30.28)***
*Net * Mrkt. * Eday*						0.026
						(0.38)
*Net * Mrkt. * Nday*						0.340
						(3.99)***
Lagged Controls?	No	Yes	Yes	Yes	Yes	Yes
Fixed Effects?	No	Day	Day * Event	Day	Day	Day

Table 4: Long and Short Anomaly Returns on Information Days vs. Off Information Days

This table reports results from a regression of daily returns on time fixed effects, the *Long and Short* anomaly variables, an information day dummy variable, interactions between *Long and Short* and the information day variables, and control variables (coefficients unreported). Returns are 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. We also include as controls for market risk interactions between the information day dummies and the daily return of the market portfolio (info day x market), and this variable interacted with *Net* (info day x market x *Net*). To create the *Long and Short* anomaly variable we use the 97 cross-sectional anomalies studied in McLean and Pontiff (2016). For each stock-month observation, we sum up the number of long-side and short-side anomaly portfolios that the stock belongs to, thereby creating *Long* and *Short*. We then merge this monthly dataset 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 1-day or 3-day window around an earnings announcement or news release, i.e., days t-1, t, and t+1. The sample period is from 1979-2013 and the sample contains 39,860,610 observations. The standard errors are clustered on time.

Table 4 (Continued)

	1-day Window	3-Day Window
Long	0.369	0.367
	(10.84)***	(11.47)***
Short	-0.193	-0.169
	(4.24)***	(3.89)***
*Long * Eday*	2.192	1.010
	(9.89)***	(9.08)***
*Short * Eday*	-1.962	-1.090
	(10.49)***	(11.33)***
*Long * Nday*	0.048	0.087
	(0.88)	(1.34)
*Short * Nday*	-0.605	-0.386
	(9.39)***	(7.216)***
Nday	19.4	11.8
	(17.87)***	(13.64)***
Eday	18.2	9.3
	(6.67)***	(6.57)***
Day Fixed Effects?	Yes	Yes
Market Risk Controls?	Yes	Yes

Table 5: Robustness: Day of Week, Macro News, and the Endogeneity of News

This table reports results from a regression of daily returns on time fixed effects, the *Long and Short* anomaly variables, an information day dummy variable, interactions between *Long and Short* and the information day variables, and control variables (coefficients unreported). Returns are 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 *Long and Short* anomaly variable we use the 97 cross-sectional anomalies studied in McLean and Pontiff (2016). For each stock-month observation, we sum up the number of long-side and short-side anomaly portfolios that the stock belongs to, thereby creating *Long* and *Short*. We then merge this monthly dataset 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 1-day window around an earnings announcement or news release. In regression 1 we include interactions between *Long* and *Short* and Monday (*Mon*) and Friday (*Fri*). In regression 2 we interact *Long* and *Short* with a macro announcement dummy (*Mac*). Following Savor and Wilson (2013, 2016) *Mac* is equal to 1 if there is a news announcement regarding inflation, unemployment, or interest rates. The sample period is from 1979-2013 and the sample contains 39,860,610 observations. The standard errors are clustered on time.

Long	0.376	0.396	0.428
	(7.83)***	(10.48)***	(6.80)***
Short	-0.129	-0.229	-0.489
	(1.71)*	(4.39)***	(4.31)***
*Long * Eday*	2.215	2.184	2.059
	(9.81)***	(9.65)***	(7.42)***
*Short * Eday*	-1.991	-1.962	-2.974
	(9.39)***	(9.19)***	(7.36)***
*Long * Nday*	0.040	0.049	0.112
	(0.47)	(0.68)	(1.44)
*Short * Nday*	-0.598	-0.602	-0.787
	(7.52)***	(7.51)***	(8.05)***
*Long * Mon*	0.370		
	(3.79)***		
*Long * Fri*	-0.383		
	(3.85)***		
*Short * Mon*	-0.437		
	(3.22)***		
*Short * Fri*	0.091		
	(0.59)		
Nday	19.4	20.4	20.2
	(19.64)***	(19.50)***	(18.31)***
Eday	18.2	18.5	22.9
	(6.79)***	(6.79)***	(7.28)***
*Long * Macro*		-0.180	
		(1.52)*	
*Short * Macro*		0.318	
		(2.20)**	
*Long * Ret^2*			-0.001
			(0.27)
*Short * Ret^2*			0.027
			(2.61)***
Ret^2			-0.000
			(0.09)
Fixed Effects?	Yes	Yes	Yes
Market Risk Controls?	Yes	Yes	Yes

Table 6: The Effect of Information Across Anomaly Types

This table tests whether the effect of information on anomaly returns varies across different 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 are constructed with 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 each of the past 10 days for stock returns, stock returns squared, and trading volume (coefficients unreported). We also include, as controls for market risk, interactions between the information day dummies and the daily return of the market portfolio (info day x market), and this variable interacted with *Net* (info day x market x *Net*). Panel B reports the results of linear restriction tests that ask whether the various coefficients are equal or different. Returns are expressed in basis points. The sample period is from 1979-2013 and the sample contains 39,860,610 observations. The standard errors are clustered on time.

Table 6: (Continued)

Panel A: Regression Results

Market	0.370
	(4.42)**
*Market * Eday*	1.079
	(2.88)**
*Market * Nday*	1.319
	(9.92)**
Valuation	0.265
	(4.95)**
*Valuation * Eday*	3.242
	(8.35)**
*Valuation * Nday*	0.525
	(4.51)**
Fundamental	0.080
	(1.67)*
*Fundamental * Eday*	2.099
	(4.88)**
*Fundamental * Nday*	-0.365
	(3.77)**
Event	0.263
	(7.10)**
*Event * Eday*	2.121
	(6.21)**
*Event * Nday*	0.186
	(2.58)**
Eday	19.7
	(18.41)**
Nday	14.8
	(31.52)**
Day Fixed Effects?	Yes
Market Risk Controls?	Yes

Table 6: (Continued)

Panel B: Linear Restriction Tests

Earnings Day Tests	**Difference**	**p-value**
Market – Valuation = 0	-2.163	0.000
Market – Fundamental = 0	-1.020	0.053
Market – Event = 0	-1.042	0.043
Valuation – Fundamental = 0	0.243	0.067
Valuation – Event = 0	1.121	0.043
Fundamental – Event = 0	-0.022	0.970

News Day Tests	**Difference**	**p-value**
Market – Valuation = 0	0.794	0.000
Market – Fundamental = 0	1.684	0.000
Market – Event = 0	1.133	0.000
Valuation – Fundamental = 0	0.890	0.000
Valuation – Event = 0	0.339	0.0131
Fundamental – Event = 0	-0.551	0.000

Table 7: The Relative Importance of Information Days

In this Table, we document the relative importance of information days in explaining anomaly returns. For each firm-day observation, we first measure the firm's abnormal return as the firm's return minus the value-weighted market return on the same day. Then, for each anomaly portfolio, we sum up all of the abnormal returns on information days and on non-information days separately. We also count the number of days that are information days and the number of non-information days for each anomaly portfolio. This exercise allows us to say what percentage of an anomaly's days are information days and what percentage of the anomaly's returns is from information days. We conduct this exercise for each of the anomaly portfolios in our sample and report the average. We define an information day as the 3-day window around an earnings announcement or news release, i.e., days t-1, t, and t+1. Panel B considers just earnings announcement days. In Panel B we limit the sample to firms that have 4 earnings announcements days (in our data) during the year. The sample period is from 1979-2013.

Table 7: (Continued)

Panel A: Both Earnings Announcement Days and Dow Jones News Days					
Long Side	**Full Sample**	**Market**	**Valuation**	**Fundamental**	**Event**
Percentage of Days	0.345	0.319	0.326	0.358	0.367
Percentage of Returns	0.801	0.959	0.863	0.741	0.683

Short Side	**Full Sample**	**Market**	**Valuation**	**Fundamental**	**Event**
Percentage of Days	0.346	0.336	0.345	0.367	0.338
Percentage of Returns	0.848	1.077	0.747	0.766	0.766

Table 7: (Continued)

Panel B: Earnings Announcement Days Only					
Long Side	**Full Sample**	**Market**	**Valuation**	**Fundamental**	**Event**
Percentage of Days	0.049	0.050	0.049	0.049	0.048
Percentage of Returns	0.172	0.163	0.172	0.186	0.166

Short Side	**Full Sample**	**Market**	**Valuation**	**Fundamental**	**Event**
Percentage of Days	0.048	0.047	0.048	0.048	0.048
Percentage of Returns	0.177	0.215	0.153	0.155	0.177

Table 8: 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 the 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 1 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 equal to zero if *Single Forecast* is equal to 1. 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. Standard errors are clustered on time. The sample contains 345,913 observations.

	Full Anomalies Sample	Market	Valuation	Fundamental	Event
Long	-0.845	-0.378	-1.541	-0.466	-2.734
	(10.15)***	(1.66)*	(6.21)***	(2.50)***	(13.30)***
Short	0.531	-0.061	1.135	0.443	1.617
	(6.47)***	(0.31)	(4.56)***	(1.90)***	(9.55)***
Number of Estimates	-0.002	-0.001	-0.002	-0.001	-0.001
	(4.05)***	(1.99)*	(4.18)***	(2.52)***	(2.50)***
Single Forecast	0.337	0.330	0.330	0.330	0.331
	(24.41)***	(23.97)***	(24.20)***	(23.89)***	(23.89)***
Dispersion	69.856	70.068	70.030	69.816	69.579
	(25.57)***	(25.49)***	(25.34)***	(25.19)***	(25.31)***
Intercept	-0.040	-0.048	-0.048	-0.054	-0.044
	(4.31)***	(5.48)***	(6.11)***	(7.69)***	(5.30)***
Month Fixed Effects?	Yes	Yes	Yes	Yes	Yes

Table 9: Mispricing or Data Mining?

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 sum up the number of long-side and short-side anomaly portfolios that the stock belongs to, thereby creating *Long* and *Short*. *Net* is equal to *Long* minus *Short*. We then merge this monthly dataset 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 1-day window around an earnings announcement or news release. *Monthly* is the firm's contemporaneous monthly stock return. Out-of-sample (*OOS*) is like *Net*, only *OOS* constructed with anomalies that are out-of-sample, i.e., i.e., past the sample date of the original study to document the anomaly.. The final two columns report regressions estimated in samples of large and small stocks only, where large (small) stocks are those above (below) the media market capitalization on day *t*. The control variables include lagged values for each of the past 10 days for stock returns, stock returns squared, and trading volume. Returns are expressed in the typical fashion, such that a unit return is 100%. The sample period is from 1979-2013 and the sample contains 39,860,610 observations. The standard errors are clustered on time.

	(1) Monthly	(2) Net + Monthly	(3) Out-of-Sample	(4) Large Stocks Only	(5) Small Stocks Only
Net		-0.203		0.289	0.324
		(5.23)***		(7.62)***	(8.99)***
*Net * Eday*		0.746		1.153	3.843
		(4.49)***		(5.19)***	(15.37)***
*Net * Nday*		0.362		0.024	0.872
		(5.34)***		(0.59)	(9.24)***
Monthly	0.058	0.058			
	(118.23)**	(117.87)***			
*Monthly * Eday*	11.607	11.611			
	(38.31)**	(38.24)***			
*Monthly * Nday*	2.626	2.490			
	(21.54)**	(21.50)***			
OOS			0.404		
			(6.73)***		
*OOS * Eday*			2.934		
			(11.08)***		
*OOS * Nday*			0.269		
			(3.59)***		
Eday	-4.0	-3.2	22.0	5.3	7.8
	(2.92)**	(2.51)*	(20.89)***	(24.63)***	(3.58)***
Nday	7.4	6.9	13.8	24.5	29.5
	(10.82)**	(11.46)**	(23.18)***	(19.40)***	(26.51)***
Day Fixed Effects?	Yes	Yes	Yes	Yes	Yes
Market Risk Controls?	Yes	Yes	Yes	Yes	Yes

Table 10: Mispricing or 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 the 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 1 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 equal to zero if *Single Forecast* is equal to 1. *Monthly* is the firm's contemporaneous monthly stock return. *Long* and *Short* are the anomaly variables, which are defined in the previous tables. *OOS_Long* and *OOS_Short* are versions of *Long* and *Short* using out-of-sample anomalies only. For readability, we divide all of the anomaly variables by 100. The regressions include time-fixed effects. Standard errors are clustered on time. The sample contains 345,431 observations.

	Monthly	*Monthly + Long and Short*	*OOS_Long and OOS_Short*
Monthly	-0.943	-0.937	
	(17.21)***	(17.20)***	
Long		-0.789	
		(15.28)***	
Short		1.427	
		(22.30)***	
OOS_Long			-0.848
			(10.18)***
OOS_Short			0.530
			(6.46)***
Number of Estimates	-0.001	-0.002	-0.002
	(3.44)***	(4.87)***	(4.23)***
Single Forecast	0.330	0.337	0.337
	(24.07)***	(24.54)***	(24.41)***
Dispersion	68.669	68.552	69.963
	(25.20)***	(25.01)***	(25.35)***
Intercept	-0.033	-0.024	-0.040
	(4.71)***	(2.46)***	(4.31)***
Month Fixed Effects?	Yes	Yes	Yes

Appendix

Biased Expectations and Returns on News and Non-News Days

Consider a multi-period economy with three securities--a risk free security with perfectly elastic supply and the stocks of two risky firms. A representative, risk-neutral agent invests his endowment to maximize expected terminal period wealth. The investor incorrectly perceives expected future cashflows for both firms to be equal to zero, and thus, the price of each stock is determined by its periodically-announced accumulated cash.

High, h, and low, l, type firms are indexed by i. Each period's cashflow is $\gamma_t^i + \varepsilon_t^i$. The variables ε_t^i's, γ_t^h, and γ_t^l are independent random variables that have respective expected values of zero, M_γ, and -M_γ, where $M_\gamma > 0$. The γ_t^i's reflect cashflow shocks that the investor mistakenly thinks have zero means. This results in high-type firm's cashflows being underestimated and low-type firm's cashflows being overestimated.

The revealed accumulated cash is $\sum_{t=1}^{m}(\gamma_t^i + \varepsilon_t^i)$, where m is the last period the firm made an announcement. The assumptions of risk neutrality and perfectly elastic supply of the zero-return riskless asset, imply that the returns the risk-neutral agent "expects" are zero. In no-news periods the prices of risky stocks do not change and the risky stocks earn zero returns. In news periods, the return of each stock is the post-news price minus the price following the last news release. Denoting the period of the last news release as j, the time k return of each stock is $\sum_{t=j}^{k}(\gamma_t^i + \varepsilon_t^i)$, the expectation of which is $(k-j)M_\gamma$ for the high-type stock and $-(k-j)M_\gamma$ for the low-type stock. Thus, when news is revealed the high-type stock has positive expected returns and the low-type stock has negative expected returns. This result holds regardless of whether the news is anticipated. When no news is revealed both stocks have zero expected returns.