Are Cross-Sectional Predictors Good Market-Level Predictors?

JOSEPH ENGELBERG, R. DAVID MCLEAN, JEFFREY PONTIFF,
AND MATTHEW C. RINGGENBERG*

October 2018

ABSTRACT

Firm-level variables that predict cross-sectional stock returns, such as price-to-earnings and book-to-market, are often aggregated and used to predict time-series market returns. We extend this literature and limit the data-snooping bias by using a near-complete population of the literature's cross-sectional return predictors. Our tests reject the null of no predictability at the annual horizon in-sample. Moreover, we find the literature has ignored several cross-sectional variables–such as change in asset turnover and co-skewness–that contain strong in-sample predictability. When we consider out-of-sample testing, however, we find little evidence that cross-sectional predictors make good market-level predictors.

Keywords: Return predictability, data snooping, statistical bias, market risk premium.

JEL Code: G00, G14, L3, C1

---

* Engelberg is at the University of California, San Diego, McLean is at Georgetown University, Pontiff is at Boston College, and Ringgenberg is at the University of Utah. The authors thank Mike Cooper, Owen Lamont, and conference and seminar participants at the 2018 Society for Financial Studies Cavalcade and the University of Utah. All errors are our own. Comments welcome. © 2016-2018.

Is the market risk premium predictable? Financial research has strived to answer this question going back, at least, to Dow (1920). “Data-mining” or “data-snooping” (Lo and MacKinlay, 1990) poses a well-recognized obstacle in obtaining an answer--research is more likely to be published in academic journals if it rejects the no-predictability null, and thus, statistically significant predictors in the literature will have a disproportionately high Type one error rate. Harvey, Liu, and Zhou (2015) demonstrate this using a wide range of cross-sectional predictors. Thus, many apparent predictors featured in the literature will fail to predict outside of the original sample.

When choosing candidate right-hand-side variables to predict the market risk premium, academics often draw from the well of cross-sectional predictors. For example, Baker and Wurgler (2000) cite the extensive market-timing literature, which relates equity issuance to returns in the cross-section, and suggest “that issuers try to time *both* their idiosyncratic return and the market return.” They go on to find that aggregate equity issuance predicts aggregate market returns. Similar arguments for aggregating cross-sectional predictors for the purpose of predicting the market risk premium are found in Campbell and Shiller (1988) with P/E ratios, Seyhun (1988) with insider trading, Chordia et al. (2002) with order imbalance, and Rapach et al. (2016) with short interest. The purpose of this paper is to evaluate cross-sectional predictors such as these in light of the data mining bias. Are the aforementioned predictors simply well-chosen among a much larger set of predictors? Or are there cross-sectional predictors that aggregate to give true market-level predictability?

To answer these questions, we borrow a list of 97 cross-sectional predictors from McLean and Pontiff (2016) that is essentially the population of firm-level characteristics that have demonstrated cross-sectional predictability in the academic literature. The benefit of examining return predictability from this cross-sectional list is that doing so avoids issues of self-selection, thereby reducing the data-snooping problem.

For each of our 97 predictors, we calculate cross-sectional averages of the firm-specific values each month to get a single, monthly value. The resulting database has a possibility of 582 different predictive variables (equal-weighted and value-weighted; raw, first-differenced, and linearly de-trended), which we use to examine their market-level predictability. We perform both in-sample and out-of-sample tests. Like other papers in the time-series literature, our in-sample tests use the entire sample of data and estimate a single parameter estimate from a time-series regression of the market risk premium on the predictor. Our out-of-sample tests consist of rolling regressions that ask whether a relation between a predictor and the market risk-premium over the last 10 years is useful for predicting the market's risk premium over the subsequent year.

At first blush, it appears that cross-sectional predictors are good market-level predictors in-sample. When we consider the 582 predictive variables, we find 552 of them are stationary and 198 of those 552 predictors (about 36%) predict 1-year ahead market returns in an OLS regression with coefficients that are significant at the 10% level or better. The strength of this result is strongly related to the horizon of predictability: when considering 1-month ahead market returns only 78 of the 552 predictors (about 14% of them) are significant. We also find several cross-sectional predictors – such as firm age, profit margin and change in asset turnover – that are among the best performers for time-series predictability, which the existing time-series literature has yet to discover. For example, the value-weighted change in asset turnover can predict the market risk premium with an R-squared of 26%. To the best of our knowledge, aggregate asset turnover--or any of its variants--have not been previously proposed as a predictor of market risk premia.

Since we examine 552 predictors, however, we expect that some variables will have spurious predictability. To address this issue we perform White's (2000) reality check bootstrap (RCB). This procedure assesses whether the best forecast in the group performs better than the efficient-market benchmark by chance, given we've examined 552 predictors. It is generally thought to be better than either the Bonferroni or Holm (1979) tests, which do not consider the dependence structure of multiple predictors and thus have low power (Romano and Wolfe (2005)).[1] The RCB tests the null hypothesis that none of the predictors in a group perform better than the benchmark against the alternative that the best predictor in the group outperforms it.

When we consider the full set of 552 predictors and perform the RCB test for predictability of market returns at the 1-month, 3-month, 6-month, and 12-month horizons, we are only able to reject the null of no-predictability at the 12-month horizon. However, at the 12-month horizon there are a number of cross-sectional anomaly variables that predict the equity risk premium with an R-squared of 10% or better, including asset turnover, co-skewness, sales-growth, and size. Thus, after accounting for a potential data-snooping bias, the results suggest that cross-sectional anomaly variables do contain information about the equity risk premium in an in-sample regression framework.

Following McLean and Pontiff (2016), we also consider four sub-categories of cross-sectional predictors: event, fundamental, market, and valuation. Valuation predictors are defined as those variables that are constructed from accounting numbers and market prices. Many valuation predictors have received attention in the existing market risk premium literature, especially dividend-to-price and earnings-to-price ratios. Ironically, using the RCB, we reject the null of no-predictability at the annual horizon for each sub-class separately, *except for valuation*.

[1] Both the Bonferroni and Holm (1979) tests lack power in that they tend not to reject the null enough when the null is true.

In other words, among all four sub-categories of cross-sectional predictors, the valuation predictors are the only category that does not appear to forecast the equity risk premium, despite their popularity in the extant literature. The *valuation* result offers a comparison to Lewellen (2004), who investigates whether valuation measures documented in earlier literature remains robust in various subsamples. Lewellen is able to reject the null of predictability for dividend yield in both subsamples, although his results for earnings-to-price and book-to-market are subsample specific. Our consideration of data-snooping for *valuation* measure predictability offers a cautionary interpretation of the apparent predictability from previous literature.

Among event-based predictors the best performer is *change in asset turnover*, among fundamental-based predictors the best performer is *sales growth*, and among market-based predictors the best performer is *co-skewness*. Among the three sub-classes that are significant at the annual horizon, we also find improved predictability when those predictors are value-weighted (rather than equal-weighted) and linearly detrended. For example, among the 38 predictors that are fundamentals-based and constructed by value-weighting and linearly de-trending, the RCB p-values are below 5% at the 3-month, 6-month, and 12-month horizons.

We also find that, among the 97 predictors, the 10 anomaly variables that exhibit the best cross-sectional predictability also have particularly good in-sample time-series predictability. The RCB p-values among this group are significant at the 3-month, 6-month and 12-month horizons. In other words, the best cross-sectional predictors tend to also be strong time-series predictors.

Things look bleaker for cross-sectional predictors, however, when we consider out-of-sample forecasting. Among the 414 stationary predictors with a long enough history to do out-of-sample forecasting, 3% significantly predict market returns at the 1-month horizon and 20%

predict market returns at the 12-month horizon (compared with 14% and 36% in-sample). Moreover, the RCB p-value is above .05 for the entire sample and in nearly every sub-sample. Put differently, we find little evidence of return predictability in out-of-sample regressions. The notable exceptions are linearly de-trended market predictors; here we find some evidence of return predictability, after accounting for a data-snooping bias. In particular, we find that *size*, *short interest* (Rapach et al. (2016)), and *long-term reversals* exhibit excellent predictability both in-sample and out-of-sample. These findings suggest that cross-sectional anomaly variables in the market sub-category do contain some information about the systematic component of returns.

Overall, our paper makes a number of contributions. First, our results provide new insight into the nature of return predictability, both in the cross-section and the time-series. By aggregating cross-sectional anomalies into time series variables, we are able to understand whether cross-sectional anomaly variables contain information about both the idiosyncratic component and systematic component of returns. Our results suggest they largely contain *idiosyncratic* information, with linearly-detrended market predictors as a notable exception. Moreover, we also contribute to the extensive literature on predicting the equity risk premium. While Goyal and Welch (2008) show that 14 popular time-series variables do not significantly predict returns in out-of-sample tests, subsequent papers have documented evidence of return predictability using firm-level variables aggregated across stocks (e.g., Hirshleifer, Hou, and Teoh (2009), Rapach, Ringgenberg, and Zhou (2016)). Our results extend these findings by showing that several other cross-sectional anomalies can be aggregated to form time series predictors. Finally, because we use the White (2000) RCB procedure and we examine a nearly complete sample of 97 cross-sectional predictors, our analyses are able to account for data snooping concerns that impact inference in time-series return predictability tests. Our paper is the first to apply the White (2000) procedure to such a large sample of candidate predictors. As such, our results provide novel evidence on the challenges of predicting the equity risk premium.

The remainder of this paper proceeds as follows: Section I briefly describes the existing literature and outlines the theoretical relation between cross-sectional anomaly variables and time series return predictability. Section II describes the data used in this study. Section III characterizes our findings and Section IV concludes.

I. Background

Financial researchers have examined the predictability of stock returns for over a century (e.g., Gibson (1906)) and a large literature has documented evidence of predictability in the cross-section of stock returns. A separate literature has examined the predictability of the equity risk-premium using time-series predictive variables. Yet, to date, these two literatures have evolved relatively independently. We connect these two literatures by creating a sample of time-series predictors based on 97 well-known cross-sectional predictors.

A. Time Series Return Predictability

A number of papers find in-sample evidence of time series return predictability, but out-of-sample evidence is rare, suggesting that many predictors are the result of data snooping (i.e., overfitting). For example, Bossaerts and Hillion (1999) use model selection criteria from the statistics literature to choose candidate predictors, which allows them to partially avoid data snooping biases, yet they find that the resulting predictors are unable to forecast returns in out-of-sample tests. Similarly, Goyal and Welch (2008) examine 14 popular predictors from the existing literature and find that they fail to forecast the equity risk premium in out-of-sample tests. Cooper and Gulen (2006) note that researchers have many different choices regarding the specification of predictability tests, including the predictor variables, the estimation periods, and the assets being forecasted. They perform specification searches across these parameters and find that return predictability results are highly sensitive to these parameter choices. More recently, Bartsch, Dichtl, Drobetz, and Neuhierl examine a wide variety of possible permutations of the predictors in Goyal and Welch (2008) and the technical predictors in Neely, Rapach, Tu, and Zhou (2014) and find that most out-of-sample performance for these variables appears to be due to data snooping.

In light of the poor performance of many predictive variables in out-of-sample tests, researchers have focused on developing methodologies that are robust to data snooping concerns. Foster, Smith, and Whaley (1997) develop a procedure to account for data snooping biases when evaluating the fit of predictive regressions. White (2000) develops a reality check bootstrap (RCB) to account for data snooping biases that result from specification searches, and Sullivan, Timmermann, and White (1999) show how to apply the RCB procedure using a set of technical trading rules. To the best of our knowledge, we are the first to apply the White (2000) RCB to a large set of predictive variables derived from existing academic studies.

B. Cross-sectional Return Predictability

While the literature on time series return predictability has generally found that most predictors fail to perform in out-of-sample tests, a large literature finds evidence of return predictability in the cross-section of stocks. Consequently, a number of recent papers have examined the multi-dimensionality of the cross-section of returns. McLean and Pontiff (2016) examine 97 anomalies in the extant literature and find many anomalies exhibit lower returns following publication in an academic journal. Green, Hand, and Zhang (2017) examine 94 firm characteristics simultaneously and find that only 12 of them reliably contain information about returns. Harvey, Liu and Zhu (2016) note there are more than 300 factors in the academic literature and Harvey and Liu (2016) examine these factors after accounting for potential data snooping. Finally, Hou, Xue, and Zhang (2017) replicate more than 400 different anomalies from the accounting and finance literature and find that many of the anomalies do not predict returns in their sample.

B. The Information in Cross-Sectional Anomaly Variables

While there are extensive literatures on both time series predictability and cross-sectional predictability, they have largely evolved independently, with a few notable exceptions. Several papers show that firm-level anomalies aggregate to market-wide predictors including Pontiff and Schall (1998) with book-to-market ratios, Campbell and Shiller (1988) with P/E ratios, and Chordia et al. (2002) with insider trading. More recently, Hirshleifer, Hou, and Teoh (2009) find that firm-level accruals and cash-flow, when aggregated across stocks, contain information about market returns, and Wen (2017) shows that aggregate asset growth predicts market returns. Finally, Rapach, Ringgenberg and Zhou (2016) show that firm level short interest aggregates to form one of the strongest known predictors of market returns.

C. The Information in Idiosyncratic Anomaly Variables

In this paper, our goal is to understand the relation between cross-sectional return predictability and time-series return predictability, more generally. While it may seem natural that cross-sectional return predictors should aggregate across assets to generate time-series return predictors, it is possible to have one without the other.[2] Specifically, note that cross-sectional anomaly variables could predict returns because they forecast either the idiosyncratic component of returns or the systematic component of returns. As such, cross-sectional return predictors do not necessarily aggregate to form good time series predictors. To see this, define a variable $X_{i,t}^{idio}$ as an idiosyncratic anomaly if it forecasts the *idiosyncratic* portion of stock returns for asset *i* on date *t+1*. Using the market model, we can write the return on asset *i* as:

$$R_{i,t} = R_f + \beta_i\left(R_{m,t} - R_f\right) + \varepsilon_{i,t},$$

where $R_{i,t}$ is the return on stock *i* on date *t*, R_f is the risk-free rate, $R_{m,t}$ is the market return, and is the idiosyncratic portion of stock *i*'s return. We define an idiosyncratic anomaly as a variable $X_{i,t-1}^{idio}$ that satisfies $\gamma_1 \neq 0$ in a linear regression of the form: [3]

$$\varepsilon_{i,t} = \gamma_0 + \gamma_1 X_{i,t-1}^{idio} + \omega_{i,t}, \tag{1}$$

where $\varepsilon_{i,t}$ is the abnormal return from the market model (Sharpe (1964); Lintner (1965)). In other words, an idiosyncratic anomaly, by definition, forecasts the portion of asset *i*'s return that is *not* explained by aggregate market movements. However, while $X_{i,t-1}^{idio}$ does contain information about individual stock returns, it will not aggregate to generate time-series return predictability. To see this, aggregate equation (2) across assets all *N* stocks in the economy and multiply both sides by $\frac{mc_{i,t}}{\sum_i^N mc_{i,t}}$, where $mc_{i,t}$ is the market capitalization of stock *i* on date *t*:

[2] Indeed, several existing papers find that firm-level relations do not hold at the aggregate level. Kothari, Lewellen, and Warner (2006) document a negative relation between returns and earnings surprise at the aggregate level, in contrast to the positive relation documented at the firm-level. Similarly, Hirshleifer, Hou, and Teoh (2009) find that the relation between accruals and returns changes sign between firm-level and aggregate-level analyses.

[3] For simplicity, we ignore the sign of the abnormal return and define an anomaly as any variable that predicts abnormal returns in either direction.

$$\frac{mc_{i,t}}{\sum_i^N mc_{i,t}} \sum_{i=1}^N [R_{i,t} - R_f - \beta_i(R_{m,t} - R_f)] = \bar{\gamma_0} + \bar{\gamma_1} \overline{X_{\iota,t-1}^{\iota d\iota o}}, \tag{2}$$

where the bar above a variable denotes the value-weighted mean. It is simple to show that the left-hand side of equation (3) is equal to zero. Thus, the value-weighted idiosyncratic anomaly variable $\overline{X_{\iota,t-1}^{\iota d\iota o}}$ contains *no* information about aggregate market returns.

D. The Information in Systematic Anomaly Variables

While idiosyncratic anomalies contain no information about aggregate market returns, it is possible to have a variable that predicts information in the cross-section that does contain information about the aggregate risk-premium. Define a variable $X_{i,t-1}^{syst}$ as a systematic anomaly if it forecasts the *systematic* portion of stock returns for asset *i* on date *t+1*. Thus, define a systematic anomaly as a variable $X_{i,t-1}^{syst}$ that satisfies $\gamma_1 \neq 0$ in a linear regression of the form:

$$\beta_i(R_{m,t} - R_f) = \gamma_0 + \gamma_1 X_{i,t-1}^{syst} + \omega_{i,t}. \tag{3}$$

Because the market beta is 1, it is easy to show that the left-hand side of equation (4) implies a direct linear relation between the predictor variable and the market risk-premium. Notice also that this relation goes in both directions: if a time-series predictor is constructed from individual assets, it *must* contain information about the systematic portion of individual asset returns. Thus, unless the idiosyncratic portion of returns exactly offsets the systematic portion of returns, we should expect any aggregate return predictor constructed from individual asset characteristics to also price individual assets. This implication provides additional economic information to test the validity of proposed predictors. In other words, when evaluating predictors constructed from individual characteristics, we should focus on the subsample of individual characteristics that contain information about individual asset returns. Accordingly, in the rest of the paper, we examine the aggregate information in a subset of 97 anomalies that have been previously shown to contain information about individual asset returns (McLean and Pontiff (2016)).

II. Data

To examine the relation between cross-sectional anomaly variables and the equity risk premium, we combine daily data from the Center for Research in Security Prices ("CRSP") and Compustat. We first construct aggregate time series variables for each of the 97 cross-sectional anomalies in McLean and Pontiff (2016).[4] For each variable, we construct six possible time series predictors. We start by constructing two different versions of each aggregate predictor: (1) we calculate the equal-weighted mean of the variable across all stocks and (2) we calculate the value-weighted mean of the variable across all stocks. We then test each predictor for a unit root using an Augmented Dickey-Fuller (1979) test. Because some of the resulting time-series variables are non-stationary, we then apply two different de-trending procedures: (1) we calculate the first-difference of each variable and (2) we calculate deviations from a linear trend model for each variable. For the linear trend, we estimate a model of the form:

$$x_t = a + bt + u_t \text{ for } t = 1, \dots, T, \tag{5}$$

for each predictor variable x_t and time period t. We take the fitted residual, $\hat{u}_t$, as our de-trended measure. By construction, $\hat{u}_t$ has a mean of zero and we normalize it to have a standard deviation of one.[5]

[4] See the Internet Appendix to McLean and Pontiff (2016), available on the *Journal of Finance* website, for a detailed overview of the construction of these variables.
[5] For our in-sample analyses, we estimate the linear trend model using all available data. For our out-of-sample analyses, we estimate the trend model only using data available at each point in time to avoid a look-ahead bias.

The resulting database has a possibility of 582 different predictive variables (equal-weighted and value-weighted; raw, first-differenced, and linearly de-trended).[6] Of these, 552 variables are stationary; we examine only stationary variables in all of our predictive regressions. Finally, we combine the 552 time series predictors with data on the equity risk premium. We calculate the equity risk premium as the log return on the S&P 500 index minus the log return on a one-month Treasury bill.[7]

III. Results

In this section, we examine whether cross-sectional anomalies, in general, contain information about the equity risk premium. We start by examining in-sample tests that use the entire sample of data and estimate a single parameter estimate from a time-series regression of the market risk premium on the predictor. We then examine out-of-sample tests that use rolling regressions to test whether a variable is useful for predicting the future equity risk premium, using only information available at each date.

A. In-Sample Tests

As discussed in Section II, above, we start with 97 anomaly variables from the existing literature and using these, we form 552 candidate predictor variables. As in McLean and Pontiff (2016), the sample length for each predictive variable depends on data availability. Some predictors have data available as far back as 1926, while other variables have samples that start more recently. In our in-sample tests, we allow the length of the time series examined to vary, depending on data availability.

---

[6] 582 = 97 anomalies × 2 (equal or value weighted) × 3 (raw, first-differenced, or linearly de-trended).
[7] We download this data from Amit Goyal's website (http://www.hec.unil.ch/agoyal/).

For each variable, we run predictive regression models of the form:

$$r_{t:t+h} = \alpha + \beta x_t + \varepsilon_{t:t+h} \text{ for } t = 1, \ldots, T - h, \tag{6}$$

where $r_{t:t+h} = (1/h)(r_{t+1} + \cdots + r_{t+h})$, r_t is the continuously compounded S&P 500 return for month t from CRSP including dividends and excess of the monthly risk-free rate from Goyal and Welch (2008), x_t is the predictor variable, and h denotes the forecast horizon. We examine four different forecast horizons: one month ahead, one quarter ahead, one-half year ahead, and one year ahead (i.e., h = 1, 3, 6, or 12). For each predictor at each forecast horizon, the regression is run using all available data, leading to a single parameter estimate (β) that measures the predictive ability of the candidate variable at that horizon.

The results are shown in Tables 1 and 2. In Table 1, we provide a summary of the performance of the candidate predictors, broken out by various sub-categories. Panel A displays the results from all 552 predictors, as well as six sub-categories: (1) *Popular*, defined as the 10 most commonly used anomalies in the cross-sectional literature based on Google Scholar Citations; (2) *Best Cross-sectional*, defined as the 10 most statistically significant anomalies in the cross-sectional literature; (3) *Event*, defined as predictors that are based on events within the firm or external events that affect the firm; (4) *Fundamental*, defined as predictors constructed from financial statement data and analysts' expectations of financial statement data; (5) *Market*, defined as predictors that can be constructed using only financial data; and (6) *Valuation*, defined as predictors that are ratios where one of the numbers reflects a market value and the other reflects fundamentals.

For the entire sample of predictors, we find evidence of return predictability. At the one-month ahead forecasting horizon, 78 of the 552 predictors are statistically significant at the 10% level or better. This number increases monotonically as the forecast horizon increases, to 138, 193, and 227 at the 3-month, 6-month, and 12-month horizon. When we examine the six sub-categories defined above, the results are similar. For each sub-category, between 30% and 40% of the predictors are statistically significant at the 12-month horizon.

We then examine the results broken out by the different methodologies used to construct the aggregate predictor variables. Specifically, in panels B and C, we examine the results when we calculate the aggregate predictor using a value-weighted average or an equal-weighted average, respectively. The results are largely consistent across the two methodologies: in general, between 30% and 50% of the predictors are statistically significant, with the value-weighted predictors showing slightly more statistical significance than the equal-weighted predictors. We also examine sub-categories broken out by the different de-trending procedures we use to construct the aggregate predictors. In panels D and E, we display results for the first-difference of value-weighted and equal-weighted predictors, respectively, while in panels F and G, we display results for the linearly de-trended value-weighted and equal-weighted predictors, respectively. Here, we find significant differences between the results. In panels D and E, which examine the first-difference of the predictors, we find very little evidence of return predictability. In all but one case, less than 10% of the predictors are statistically significant. The one exception is the market category of the first-difference of equal-weighted predictors (the second to last row in panel E). However, both the value-weighted and equal-weighted version of the linearly de-trended predictors show strong evidence of return predictability. In panel F (linearly de-trended value weighted predictors), between 57% and 78% of the predictors are statistically significant at the 10% level or better. While the equal-weighted results in panel G are slightly weaker, they too show strong evidence of return predictability. Overall, the results in Table 1 provide suggestive evidence that cross-sectional anomaly variables can be used to construct aggregate return predictors.

In Table 2, we display the detailed parameter estimates for the best predictor at each horizon.[8] We report heteroskedasticity and autocorrelation robust *t*-statistics computed using a wild bootstrapped with 500 draws to account for the Stambaugh (1999) bias and the fact that the model uses overlapping observations when h > 1 (Goetzmann and Jorion (1993); Hodrick (1992); Nelson and Kim (1993)).

In general, the table shows strong evidence of return predictability. At the one-month horizon, a number of the predictors are statistically significant. *Sales Growth*, with an R-squared of 1.88% shown in column (3), appears to be the best predictor at the one-month horizon. Although the R-squared values might seem small in absolute magnitude, Zhou (2010) notes that R-squared values from predictive regressions are typically small in absolute magnitude as stock returns are difficult to forecast. Accordingly, Zhou (2014) builds on the insights of Ross (2005) to construct a mathematical bound on the maximum R-squared that can exist under no arbitrage conditions. Using consumption growth rates as a state variable, Zhou's bounds imply a maximum R-squared at the monthly horizon of between 0.079% and 0.177%.[9] In light of this, the return predictability documented in column (3) of Table 2 seems economically significant.

Similarly, we find strong evidence of return predictability at the three-month, six-month, and twelve-month horizons. At the twelve-month horizon, a number of the predictors have impressive R-squared values exceeding 10%, including *Size*, *Price*, *Sales Growth*, *Asset Growth*,

---

[8] We estimate six different models for each cross-sectional anomaly variable (equal-weighted and value-weighted for raw, first-differenced, and linearly de-trended values of the predictor). For each cross-sectional anomaly variable, Table 2 displays the best performing of the six estimates, as measured by the R-squared.

[9] The bounds developed in Zhou (2010) depend on the choice of a state variable. We do not take a stance on state variables in this paper, we simply note that the R-squared values we document appear economically meaningful relative to the example bounds presented in Zhou (2010).

Co-Skewness, and *Change in Asset Turnover*. Moreover, these predictors are statistically and economically significant at all forecasting horizons. Taken together, the results suggest that cross-sectional anomaly variables may forecast cross-sectional returns, at least in part, because they contain information about the equity risk premium.

A.1. White RCB Test of In-Sample Results

Of course, Table 2 shows the *best* results for each of the 97 cross-sectional anomaly variables (across the six different specifications considered for each variable). As such, the results are subject to concerns of data snooping. Fortunately, White (2000) develops a reality check bootstrap (RCB) to correct for data-snooping that is particularly suited to application in return predictability regressions. In Table 3, we apply the White (2000) RCB procedure to account for data snooping bias in a multiple hypothesis testing.

Of course, there are other ways to adjust p-values for the bias that results from multiple hypothesis testing. The White (2000) procedure has an advantage over the Bonferroni or Holm (1979) approach in that it uses a bootstrap procedure to estimate the dependence structure of the individual p-values. In contrast, the Bonferroni and Holm (1979) approaches assume the worst case dependence structure. As such, they are overly conservative in that they do not reject the null hypothesis enough when the null is true. The White (2000) RCB approach uses a stationary block bootstrap (Politis and Romano (1994)) to ensure that the bootstrap accurately estimates the sampling distribution.[10] The procedure then compares the mean squared forecast error for each predictor and compares it to the mean squared forecast error from a benchmark model which

[10] We use the bootstrap procedure in Politis and Romano (1994) which preserves the underlying stationarity of the data. For each predictor, we use 1000 replications with a mean block size of 5. Sullivan, Timmerman, and White (1999) show the procedure is robust to various block size choices.

uses the prevailing mean return as the forecast of next period's equity risk premium. Finally, the procedure examines whether the best performing model is sufficiently better than the benchmark model after accounting for the number of different models that were examined.

We are the first paper to use this procedure to evaluate the predictability of the market's risk premium. The closest related paper is Sullivan, Timmerman, and White (1999), which applies the White procedure to examine the daily returns of technical trading strategies. Sullivan et al. find no evidence that trading strategies can generate portfolio performance that out-performs a benchmark.

The results from the White (2000) RCB for our in-sample regressions are shown in Table 3. As in Table 1, the first panel (Panel A) displays the results when all of the predictive variables are considered, while Panels B and C display the results when we consider only value-weighted or equal-weighted predictors, respectively. Finally, Panel D and Panel F display results when we consider only the first-difference of value-weighted or equal-weighted predictors, respectively, while Panel F and Panel G display results when we consider only linearly de-trended value-weighted or equal-weighted predictors, respectively.

Interestingly, the White (2000) procedure suggests a different interpretation of the results in Table 1 and Table 2. In particular, after accounting for data snooping, many of the results do not show statistically significant evidence of return predictability. In Panel A, we fail to reject the null of no predictably at the one-month forecasting horizon when we consider all of the predictors. We also fail to reject the null of no predictability for each of the six sub-categories (*Popular, Best Cross-Sectional, Event, Fundamental, Market*, and *Valuation*). However, as the forecasting horizon increases, the evidence shifts. At the three-month and six-month horizons we do see return predictability for the *Popular*, *Best Cross-sectional*, and *Fundamental* sub-categories, and at the twelve-month horizon we see weak evidence of return predictability for the entire sample of predictors. Moreover, all of the sub-categories except *Valuation* show statistically significant evidence of return predictability. This finding is interesting in light of the existing literature's focus on *Valuation* measures as return predictors (e.g., dividend-to-price and earnings-to-price ratios). Our findings suggest these predictors are the only category that do not forecast the equity risk premium after accounting for data snooping biases.

In the remaining panels of Table 3, we present the results of the White (2000) RCB procedure across the different methodologies we used to construct our time series predictors. As with Table 1, the results show a clear pattern: value-weighted predictors are consistently better than equal-weighted predictors. Similarly, linearly de-trended predictors are consistently better the first-differences predictors. Overall, the results suggest that value-weighted predictors do contain information about the future equity risk premium. In other words, our in-sample tests suggest that cross-sectional anomaly variables contain some information about the systematic component of returns.

B. Out-of-Sample Tests

In the preceding section, we found in-sample evidence that cross-sectional anomaly variables could be aggregated across assets to form time series return predictors. However, a number of papers note that in-sample tests may overstate predictability due to the use of information that was not known ex-ante (e.g., Cooper, Gutierrez, Marcum (2005), Goyal and Welch (2008)). Accordingly, in this section, we revisit each of our tests using out-of-sample forecasting regressions. We again run predictive regressions of the form:

$$r_{t:t+h} = \alpha_t + \beta_t x_t + \varepsilon_{t:t+h} \text{ for } t = 1, \ldots, T-h, \quad (7)$$

where $r_{t:t+h} = (1/h)(r_{t+1} + \cdots + r_{t+h})$, r_t is the excess return on the S&P500, and x_t is the predictor variable. However, now we estimate the model separately for each time period, using only information that was available at each date. As such, we estimate new parameter estimates for α_t and β_t at each point in time. If the relation between the predictor variable and the equity risk premium is stable over time, then this out-of-sample approach should give the same results as the in-sample analysis discussed in section III.A. However, if the between the predictor variable and the equity risk premium is not stable, the out-of-sample tests may lead to a different conclusion.

As previously discussed, the sample length for each predictive variable depends on data availability. For the out-of-sample tests, we require a minimum of 10 years of data to train the model before we make our first forecast. To ensure we have an adequate time series to train the model and evaluate the resulting forecasts, we require that each cross-sectional anomaly variable has at least 25 years of data to be included in the out-of-sample analyses. This sample filter eliminates several of the anomaly variables from consideration; of the 582 possible specifications, we are left with 414 that are stationary and have long enough time series.

As before, we start by summarizing the results across all specifications. Table 4 provides a summary of the performance of the candidate predictors. Panel A summarizes the results from all 414 predictors, as well as six sub-categories (*Popular, Best Cross-sectional, Event, Fundamental, Market,* and *Valuation*). Recall that in Table 1, the in-sample evidence was strongest at the annual horizons, however, even at the 1-month horizon approximately 14% of the predictors were statistically significant at the 10% level or better. In contrast, the out-of-sample evidence is much weaker. At the one-month horizon, only 2% of the variables are statistically significant and while this number increases monotonically as the forecast horizon increases, it is still only 19% at the twelve-month horizon. When we examine the individual sub-categories, the results do not look much better. Again, the *Valuation* sub-category appears to be the worst predictor, despite its popularity in the extant literature.

In the remaining panels of Table 4, we examine the out-of-sample results broken out by the different methodologies used to construct the aggregate predictor variables. Specifically, in panels B and C, we examine the results when we calculate the aggregate predictor using a value-weighted average or an equal-weighted average, respectively. In panels D and E, we display results for the first-difference of value-weighted and equal-weighted predictors, respectively, while in panels F and G, we display results for the linearly de-trended value-weighted and equal-weighted predictors, respectively. While there are some variables that predict returns in each of these panels, the number of significant predictors remains low relative to the total number of variables considered.

In Table 5, we display the detailed results for the best predictor at each forecasting horizon. To make inferences, we report the time-series mean of the betas from these calculations in addition to the out-of-sample R^2_{OS} statistic defined as in Campbell and Thompson (2008).[11] To calculate the out-of-sample R^2_{OS}, we use the prevailing mean equity risk premium, at each date, as our benchmark model and we use the Clark and West (2007) statistic to assess statistical significance.[12]

Interestingly, the detailed results in Table 5 show that several of the best out-of-sample predictors were also good in-sample predictors. Specifically, the out-of-sample tests find that *Size*, *Price*, *Rating Down*, *Short Interest*, and *Change in Asset Turnover* are strong predictors of

[11] We use the unconstrained out-of-sample R-squared from Campbell and Thompson (2008) (i.e., we do not impose any sign restrictions).
[12] Formally, we test the null hypothesis that the mean square forecast error (MSFE) from the baseline model is less than or equal to the MSFE from the predictive model versus the alternative hypothesis that the MSFE from the benchmark model is greater than the MSFE from the predictive model (H_0: $R^2_{OS} \leq 0$ against (H_A: $R^2_{OS} > 0$).

the future equity risk premium. Because the out-of-sample R_{OS}^2 statistic measures the proportional reduction in mean squared forecast error that results from using the predictor (relative to the benchmark model), its magnitude is also useful for interpreting the economic significance of these findings. For the five variables listed above, the out-of-sample R_{OS}^2 statistic suggests economically large return predictability. In particular, *Change in Asset Turnover* has impressive forecasting ability at the three-month, six-month, and twelve-month horizons with R_{OS}^2 values of 6%, 16%, and 28%, respectively. Taken together, it is tempting to conclude that some cross-sectional predictors can be used to form aggregate predictors, however, our out-of-sample analyses have considered more than 400 different specifications. Accordingly, in the next section we apply the White (2000) RCB procedure to account for the large number of specifications considered.

B.1. White RCB Test of Out-of-Sample Results

As before, we ask whether our out-of-sample tests show evidence of return predictability after accounting for possible data snooping biases. To do this, we again use the White (2000) procedure. The results are shown in Table 6. The first panel (Panel A) displays the results when all of the predictive variables are considered, while Panels B and C display the results when we consider only value-weighted or equal-weighted predictors, respectively. Finally, Panel D and Panel F display results when we consider only the first-difference of value-weighted or equal-weighted predictors, respectively, while Panel F and Panel G display results when we consider only linearly de-trended value-weighted or equal-weighted predictors, respectively.

When we applied the White (2000) RCB procedure to our in-sample estimates (Table 3) we found little evidence of return predictability at short horizons, but strong evidence of return predictability at longer horizons. However, the results in Table 6 show little evidence of predictability at *any* horizon. Across all sub-categories and all panels, we find no statistical significance at the one-month horizon, and only sub-category in one panel (*Market* in Panel F) is statistically significant at the three-month horizon. At the six-month horizon, only two categories (*Market* in Panel F and Panel G) are significant. Finally, at the twelve-month horizon, we find weak evidence of return predictors when we consider all predictors (row one in Panel A), all *Market* predictors (row six in Panel A), and when we consider linearly de-trended predictors (Panels F and G). Taken together, the findings suggest that most cross sectional anomaly variables cannot be aggregated to form time series return predictors. Put differently, the results imply that most cross-sectional anomaly variables are able to forecast in the cross-section of returns because they contain information about the idiosyncratic component of returns, not the systematic component of returns. The notable exception to this appears to be the *Market* category of anomalies. For these variables, there is some evidence that they contain information about the systematic component of returns at longer horizons (six-months to one-year ahead).

IV. Conclusion

There is a large literature examining the cross-sectional determinants of stock returns. Similarly, countless time series variables have been proposed as predictors of the equity risk premium. Yet, to date, these literatures have evolved largely independently. We connect the two literatures by forming time series predictor variables out of 97 cross-sectional anomaly variables from the existing academic literature. The resulting dataset allows us to explore the relation between the cross-sectional determinants of stock returns and the time series determinants. More generally, the analyses provide new information on the nature of return predictability. While it may seem natural that cross-sectional variables should contain information about aggregate market returns, we show that this does not have to be the case. Crucially, the relation depends on the nature of cross-sectional return predictability: if a firm-level variable contains only information about the idiosyncratic component of returns, then it will not be possible to generate a variable that significantly predicts the equity risk premium by simply aggregating the firm-level variable across assets.

We find mixed evidence on the relation between cross-sectional predictors and time series predictors. Using in-sample regressions, we find evidence that cross-sectional variables, when aggregated, do forecast the future equity risk premium. However, this result becomes weaker once we use the White (2000) RCB procedure to account for the data snooping bias arising from the plethora of predictive variables being considered. Moreover, when we examine out-of-sample forecasting regressions, we find little evidence of return predictability. Overall, the results suggest that cross-sectional anomaly variables contain information about the idiosyncratic component of returns, but not the systematic component.

REFERENCES

Baker, M. and Wurgler, J., 2000, The Equity Share in New Issues and Aggregate Stock Returns, *The Journal of Finance* 55, 2219-2257.

Bartsch, Viktoria-Sophie, Hubert Dichtl, Wolfgang Drobetz, and Andreas Neuhierl, 2017, Data Snooping in Equity Premium Prediction, *Working Paper*.

Bay, Raymond, 1978, Anomalies in relationships between securities' yields and yield-surrogates, *Journal of Financial Economics* 6, 103-126.

Berk, Jonathan B, 1995, A Critique of Size-Related Anomalies, *Review of Financial Studies* 8, 275-286.

Bossaerts, P., and Pierre Hillion, 1999, Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?, *Review of Financial Studies* 12, 405-428.

Campbell, John Y. and Robert J. Shiller, 1988, The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors, *Review of Financial Studies* 1, 195-228.

Campbell, J.Y. , Thompson, S.B., 2008. Predicting excess stock returns out of sample: can anything beat the historical average? *Review of Financial Studies* 21, 1509-1531.

Chordia, Tarun, Richard Roll and Avanidhar Subrahmanyam, 2002, Order Imbalance, Liquidity and Market Returns, *Journal of Financial Economics* 65, 111-130.

Clark, T.E., West, K.D., 2007. Approximately normal tests for equal predictive accuracy in nested models, *Journal of Econometrics* 138, 291–311.

Cochrane, John H., 2008, The Dog That Did Not Bark: A Defense of Return Predictability, Review of Financial Studies 21, 1533-1575.

Cooper, Michael, and Hyseyin Gulen, Is Time-Series-Based Predictability Evident in Real Time?, Journal of Business 79, 1263-1292.

Cowles, Alfred, 1933, Can Stock Market Forecasters Forecast?, *Econometrica* 1, 309-324.

Dickey, D.A., and W.A. Fuller, 1979, Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74: 427-431.

Dow, C. H., 1920, Scientific Stock Speculation, *The Magazine of Wall Street*.

Foster, F. Douglas, Tom Smith, and Robert E. Whaley, 1997, Assessing Goodness-of-Fit of Asset Pricing Models: The Distribution of the Maximal R^2, *Journal of Finance* 52, 591-607.

Gibson, Thomas, 1906, The Pitfalls of Speculation. The Moody Corporation, New York.

Green, J., J. R. Hand and X. F. Zhang, 2017, "The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns, forthcoming *Review of Financial Studies*.

Goetzmann, W.N. , Jorion, P. , 1993. Testing the predictive power of dividend yields. *Journal of Finance* 48, 663–679.

Goyal, Amit, and Ivo Welch, 2008, A Comprehensive Look at the Empirical Performance of Equity Premium Prediction, *Review of Financial Studies* 21, 1455-1508.

Harvey, Campbell R., Yan Liu, and Heqing Zhu, 2016, …and the Cross-Section of Expected Returns, *Review of Financial Studies* 29, 5-68.

Harvey, Campbell R., and Yan Liu, 2015, Lucky Factors, *Working Paper*.

Hodrick, R.J. , 1992. Dividend yields and expected stock returns: alternative procedures for inference and measurement, *Review of Financial Studies* 5, 357-386.

Hou, K., C. Xue, and L. Zhang, 2017, Replicating Anomalies, *Working Paper*.

Lewellen, J., 2004, Predicting returns with financial ratios, Journal of Financial Economics 74, 209-235.

Lo, Andrew W., and A. Craig MacKinlay, Data-Snooping Bias in Tests of Financial Asset Pricing Models, *Review of Financial Studies* 3, 431-467.

McLean, R. David, and Jeffrey Pontiff, 2016, Does Academic Publication Destroy Stock Return Predictability?, *Journal of Finance* 71, 5-32.

Neely, C.J., D.E. Rapach, J. Tu, and G. Zhou, 2014, Forecasting the Equity Risk Premium: The Role of Technical Indicators, *Management Science* 60, 1772–1791.

Nelson, C.R. , Kim, M.J. , 1993. Predictable stock returns: the role of small sample bias, *Journal of Finance* 48, 641-661.

Politis, D., and J. Romano, 1994, The Stationary Bootstrap, *Journal of the American Statistical Association* 89, 1303-1313.

Pontiff, Jeffrey, and Lawrence Schall, 1998, Book-to-market as a predictor of market returns, *Journal of Financial Economics* 49, 141-160.

Rapach, David E., Matthew C. Ringgenberg, Guofu Zhou, 2016, Aggregate Short Interest and Return Predictability, *Journal of Financial Economics* 121, 46-65.

Ross, S.A., 2005, Neoclassical Finance. Princeton University Press.

Schwert, G.W., 2003. Anomalies and market efficiency. In: Constantinides, G.M., Harris, M., Stultz, R. (Eds.), Handbook of the Economics of Finance, 1B. Elsevier, Amsterdam, pp. 939-974.

Seyhun, H. Nejat, 1988, The Information Content of Aggregate Insider Trading, *The Journal of Business* 61, 1-24.

Stambaugh, R.F., 1999, Predictive regressions, *Journal of Financial Economics* 54, 375–421.

Sullivan, R., A. Timmermann, and H. White, 1999, Data-snooping, technical trade rule performance, and the bootstrap, The Journal of Finance 54, 1647-1691.

White, H., 2000, A Reality Check for Data Snooping, *Econometrica* 68, 1097-1126.

Zhou, G., 2010, How much stock return predictability can we expect from an asset pricing model?, *Economic Letters* 108, 184-186.

Table 1
Summary of In-Sample Anomaly Performance

The table displays a count of the number of predictive variables that are statistically significant at the 10% level or better. For each anomaly, we estimate an in-sample predictive regression of the form:

$$r_{t:t+h} = \alpha + \beta x_t + \varepsilon_{t:t+h} \text{ for } t = 1, \ldots, T-h,$$

where $r_{t:t+h} = (1/h)(r_{t+1} + \cdots + r_{t+h})$, r_t is the continuously compounded S&P 500 return for month t from CRSP including dividends and excess of the monthly risk-free rate from Goyal and Welch (2008), h indicates the forecast horizon in months, and x_t is one of the 97 predictor variables. To construct aggregate predictors out of cross-sectional anomalies, we calculate the value-weighted and equal-weighted mean across all stocks on each date. Because some of the resulting variables are non-stationary, we also examine the first-difference of these predictors and we examine a linearly de-trended version of these predictors. These permutations result in 582 possible predictors (552 of them are stationary). The table shows the number of predictors that are statistically significant relative to the total number of variables considered. Panel A summarizes the results for all predictors, Panel B examines the value-weighted predictors, Panel C examines the equal-weighted predictors, Panel D examines the first-difference of value-weighted predictors, Panel E examines the first-difference of equal-weighted predictors, Panel F examines linearly de-trended value-weighted predictors, and Panel G examines linearly de-trended equal-weighted predictors. In each panel, the first row displays results for all variables in that panel, and the remaining rows examine sub-samples formed on the ten most popular cross-sectional predictors (*Popular*), the ten most statistically significant cross-sectional predictors (*Best Cross-Sectional*), and four different groupings based on the categories in McLean and Pontiff (2015): (1) *Event*, (2) *Fundamental*, (3) *Market*, and (4) *Valuation*. See Section 3.C of the text for a detailed discussion of these categorizations.

(1)	**(2)**	**(3)**	**(4)**	**(5)**
	Number significant predictors / Total number predictors			
Predictive Variable	***h*=1**	***h*=3**	***h*=6**	***h*=12**
Panel A: All Predictors				
All Predictors	78/552	138/552	193/552	227/552
Popular	12/52	8/52	11/52	20/52
Best Cross-sectional	19/73	21/73	25/73	22/73
Event	18/187	42/187	68/187	75/187
Fundamental	24/165	48/165	62/165	71/165
Market	31/134	39/134	45/134	60/134
Valuation	5/66	9/66	18/66	21/66
Panel B: VW Predictors				
All Predictors	49/278	82/278	114/278	128/278
Popular	8/26	7/26	6/26	12/26
Best Cross-sectional	11/37	13/37	13/37	13/37
Event	11/99	26/99	41/99	44/99
Fundamental	16/83	28/83	41/83	45/83
Market	20/67	23/67	24/67	29/67
Valuation	2/29	5/29	8/29	10/29
Panel C: EW Predictors				
All Predictors	29/274	56/274	79/274	99/274
Popular	4/26	1/26	5/26	8/26
Best Cross-sectional	8/36	8/36	12/36	9/36
Event	7/88	16/88	27/88	31/88
Fundamental	8/82	20/82	21/82	26/82
Market	11/67	16/67	21/67	31/67
Valuation	3/37	4/37	10/37	11/37

Table 1 (continued)
Summary of In-Sample Anomaly Performance

Panel D: First Difference of VW Predictors				
All Predictors	9/97	10/97	9/97	6/97
Popular	2/10	1/10	0/10	0/10
Best Cross-sectional	2/10	2/10	0/10	0/10
Event	2/33	4/33	4/33	3/33
Fundamental	0/27	1/27	2/27	1/27
Market	6/24	3/24	2/24	1/24
Valuation	1/13	2/13	1/13	1/13
Panel E: First Difference of EW Predictors				
All Predictors	10/96	9/96	8/96	9/96
Popular	2/10	0/10	1/10	1/10
Best Cross-sectional	2/10	1/10	1/10	1/10
Event	2/32	4/32	4/32	2/32
Fundamental	2/27	2/27	1/27	1/27
Market	5/24	2/24	3/24	5/24
Valuation	1/13	1/13	0/13	1/13
Panel F: Linearly De-trended VW Predictors				
All Predictors	28/127	50/127	74/127	85/127
Popular	4/11	4/11	4/11	8/11
Best Cross-sectional	6/19	8/19	10/19	11/19
Event	7/45	15/45	26/45	29/45
Fundamental	11/38	19/38	26/38	30/38
Market	9/30	13/30	15/30	18/30
Valuation	1/14	3/14	7/14	8/14
Panel G: Linearly De-trended EW Predictors				
All Predictors	14/131	37/131	50/131	65/131
Popular	1/11	1/11	2/11	5/11
Best Cross-sectional	3/19	5/19	7/19	7/19
Event	3/42	10/42	16/42	22/42
Fundamental	4/39	14/39	15/39	18/39
Market	5/30	10/30	12/30	18/30
Valuation	2/20	3/20	7/20	7/20

Table 2
Best In-Sample Predictive Regression Results for each Anomaly

The table reports the mean of the ordinary least squares estimate of β and the R^2 statistic from in-sample predictive regression models of the form:

$$r_{t:t+h} = \alpha + \beta x_t + \varepsilon_{t:t+h} \text{ for } t = 1, \ldots, T-h,$$

where $r_{t:t+h} = (1/h)(r_{t+1} + \cdots + r_{t+h})$, r_t is the continuously compounded S&P 500 return for month t from CRSP including dividends and excess of the monthly risk-free rate from Goyal and Welch (2008), h indicates the forecast horizon in months, and x_t is the predictor variable in the first column. For each predictor, we report the version that has the best R^2. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		***h*=1**		***h*=3**		***h*=6**		***h*=12**	
	Predictive Variable	$\hat{\beta}$	R^2	$\hat{\beta}$	R^2	$\hat{\beta}$	R^2	$\hat{\beta}$	R^2
1	e/p	-0.14	0.10	-0.12	0.23	-0.14*	0.56	-0.17***	1.54
2	m/b	-0.05	0.01	0.10	0.14	-0.03	0.03	-0.07	0.25
3	size	-0.55***	1.00	-0.57***	2.85	-0.55***	5.56	-0.58***	11.60
4	momentum	-0.12	0.08	-0.16	0.37	-0.13*	0.49	-0.13**	0.96
5	UMD-reversal	0.14	0.10	0.16	0.39	-0.24***	1.58	-0.22***	2.66
6	LT reversal	-0.50***	0.82	-0.52***	2.41	-0.52***	5.03	-0.46***	7.65
7	ST reversal	0.54***	0.94	0.14	0.18	0.06	0.06	0.06	0.11
8	dividends	0.24	0.30	0.25**	0.99	0.29***	2.37	0.25***	3.68
9	price	-0.56***	1.10	-0.59***	3.35	-0.60***	7.57	-0.55***	12.50
10	leverage	-0.08	0.03	-0.11	0.20	0.00	0.00	0.05	0.15
11	volume	-0.18	0.18	-0.25**	0.96	-0.11	0.33	-0.07	0.27
12	investment	0.16	0.12	0.17	0.40	0.17*	0.80	0.09	0.44
13	accruals	0.33*	0.55	0.33***	1.61	0.33***	2.95	0.20***	2.28
14	debt issues	0.18	0.17	0.18	0.46	0.16*	0.68	0.20***	2.02
15	repurchases	-0.34*	0.59	-0.34***	1.65	-0.35***	3.31	-0.32***	5.39
16	seo	0.11	0.06	0.20	0.57	0.20**	1.08	0.11	0.67
17	mergers	-0.25	0.33	-0.24**	0.87	-0.28***	2.14	-0.26***	3.79
18	spinoffs	-0.47**	1.16	-0.08	0.11	-0.09	0.26	0.01	0.01
19	z-score	-0.42*	0.92	-0.31**	1.40	-0.24**	1.63	-0.19***	1.88
20	o-score	-0.42*	0.91	-0.16	0.35	-0.09	0.23	0.13*	0.91
21	noa	-0.36**	0.68	-0.38***	2.23	-0.37***	3.86	-0.35***	6.74
22	R&D/MV	0.37*	0.71	0.32***	1.60	0.29***	2.51	0.3***	5.31
23	Marketing/MV	0.32	0.54	0.25**	0.99	0.18**	0.96	0.05	0.13
24	Sales Growth	-0.61***	1.88	-0.56***	4.47	-0.50***	6.80	-0.47***	11.69
25	CF/MV	-0.09	0.04	-0.06	0.05	-0.03	0.03	0.08	0.33
26	52-Week High	-0.17	0.16	-0.13	0.27	-0.12	0.39	-0.11**	0.64
27	Beta	0.31*	0.32	0.10	0.09	0.06	0.07	0.04	0.07
28	PEAD	-0.35*	0.63	-0.21*	0.67	-0.20**	1.19	-0.09	0.48
29	Asset Growth	-0.41**	0.83	-0.42***	2.59	-0.43***	5.06	-0.44***	10.29
30	Share issues DT	0.13	0.09	-0.13	0.26	-0.22***	1.33	-0.19***	1.85
31	Share issues PW	0.13	0.09	-0.13	0.26	-0.22***	1.33	-0.19***	1.85
32	Idio. Risk	0.25	0.32	-0.28**	1.11	-0.25***	1.67	-0.24***	3.18
33	Herfindahl	-0.35*	0.64	-0.26***	1.09	-0.19**	1.02	-0.2***	2.24
34	VAR(Volume)	0.17	0.16	0.13	0.27	0.03	0.03	0.02	0.02
35	Exch. Switch	-0.25	0.32	-0.17	0.41	-0.14*	0.50	-0.13**	0.92
36	ROE	0.11	0.07	-0.12	0.20	-0.15*	0.65	-0.18***	1.80
37	Volume / MV	0.18	0.16	0.20	0.61	0.21**	1.20	0.15**	1.21
38	Amihud	-0.49**	1.21	-0.27**	1.08	-0.18*	0.92	-0.16**	1.38
39	Volume Trend	-0.17	0.16	-0.15	0.36	-0.14*	0.60	-0.14***	1.15
40	VAR(CF)	0.35	0.65	0.34***	1.71	0.32***	3.02	0.30***	5.03
41	Div. Initiation	-0.18	0.17	0.18	0.49	0.18*	0.89	0.11	0.63

Table 2 (continued)
Best In-Sample Predictive Regression Results for each Anomaly

42	Div. Omission	-0.08	0.03	-0.12	0.23	-0.05	0.07	-0.09	0.42
43	Pension Funding	0.38**	0.46	0.37***	1.20	0.35***	2.20	0.34***	4.08
44	Vol-Momentum	-0.36*	0.65	-0.26*	0.96	-0.27***	2.00	-0.28***	4.37
45	G-Score	-0.17	0.15	-0.23**	0.78	-0.22***	1.31	-0.16***	1.42
46	G-Score 2	0.17	0.14	0.10	0.16	-0.07	0.13	-0.06	0.23
47	R&D Increases	0.19	0.18	0.19	0.53	0.19**	1.03	0.08	0.41
48	Rating Down	0.24	0.29	-0.25**	0.87	-0.29***	2.33	-0.37***	8.09
49	Moment-Ratings	-0.20	0.21	0.08	0.09	0.11	0.32	-0.07	0.30
50	Work. Capital Δ	-0.15	0.11	-0.17	0.40	-0.10	0.28	-0.06	0.20
51	N. Op. Assets Δ	-0.02	0.00	-0.15	0.37	-0.11	0.33	0.06	0.23
52	Age	-0.38**	0.82	-0.36***	2.01	-0.35***	3.60	-0.38***	8.07
53	Age-Momentum	-0.25	0.21	-0.07	0.04	-0.09	0.14	-0.13**	0.59
54	IPOs	-0.09	0.04	-0.13	0.27	-0.13	0.47	-0.14**	1.08
55	Seasonality	-0.34	0.59	-0.20	0.61	-0.19**	1.04	0.14**	1.13
56	Max	0.18	0.18	-0.19**	0.56	-0.10	0.27	-0.08*	0.41
57	ΔSales-ΔInvent	0.36**	0.42	0.19*	0.32	-0.09	0.14	0.10**	0.35
58	ΔCpX-ΔIndCpX	-0.21	0.23	-0.26**	0.99	-0.27***	2.12	-0.21***	2.72
59	ΔSales-ΔSG&A	0.13	0.08	0.18	0.49	-0.21**	1.29	-0.19***	2.07
60	Ent. Comp B/P	0.15	0.11	0.16	0.40	0.19**	1.05	0.12**	0.86
61	Lev. Comp B/P	-0.14	0.11	0.11	0.17	0.13*	0.49	0.08	0.35
62	Total XFIN	-0.19	0.19	-0.23**	0.79	-0.11	0.32	0.10*	0.59
63	Grwth LTNOA	-0.41**	0.85	-0.40***	2.26	-0.36***	3.44	-0.33***	5.85
64	Grwth Inventory	0.27	0.37	-0.15	0.33	-0.15*	0.65	-0.16***	1.36
65	Rev. Surprises	0.25	0.31	0.26**	1.01	0.28***	2.08	0.21***	2.33
66	Operating Lev.	-0.28	0.39	-0.27*	1.05	-0.41***	4.85	-0.33***	6.00
67	M/B & Accruals	-0.08	0.03	-0.23**	0.82	0.05	0.09	0.06	0.21
68	Mom &Reversal	0.33	0.55	-0.12	0.22	-0.18*	0.92	0.19***	2.02
69	Lag(Momentum)	0.41**	0.86	0.34***	1.73	0.28***	2.16	0.19***	2.07
70	Short Interest	0.46***	0.69	-0.24**	0.51	-0.24***	1.09	-0.29***	2.81
71	Enterprise Multp	0.59***	1.83	-0.51***	3.97	-0.54***	8.69	-0.42***	10.43
72	Sustain Grwth	-0.09	0.04	0.09	0.14	0.03	0.03	-0.06	0.20
73	Co-skewness	-0.43**	0.97	-0.46***	3.16	-0.47***	6.37	-0.48***	12.78
74	% Total Accrual	0.44**	1.05	0.34***	1.82	0.24***	1.69	0.16***	1.47
75	% Op. Accrual	-0.27	0.41	-0.11	0.19	-0.17	0.84	-0.17**	1.60
76	Earnings Consist	-0.34	0.64	-0.43***	2.95	-0.45***	5.94	-0.43***	10.03
77	Ind. Momentum	-0.10	0.05	-0.14	0.29	-0.16*	0.66	0.08	0.31
78	Δ Recommend	-0.15	0.11	-0.17	0.43	-0.22***	1.37	-0.24***	3.20
79	Forecast Disp.	-0.35	0.65	-0.29*	1.18	-0.20	1.00	-0.10	0.42
80	Up Forecast	0.46**	1.09	0.34***	1.75	0.33***	3.17	0.27***	4.11
81	Down Forecast	-0.04	0.01	-0.10	0.13	-0.17*	0.84	-0.06	0.19
82	G Index	0.24	0.30	0.14	0.30	0.24**	1.61	0.21***	2.38
83	Sales/Price	-0.13	0.09	-0.21	0.66	-0.15	0.57	-0.17**	1.48
84	Asset Turnover	0.25	0.32	0.15	0.34	0.13	0.45	-0.09	0.46
85	ΔAsset Turnover	0.64***	2.07	0.67***	6.71	0.7***	14.30	0.67***	25.60
86	Profit Margin	-0.12	0.07	-0.23	0.76	0.22**	1.43	0.39***	8.72
87	ΔProfit Margin	0.37	0.70	0.46***	3.18	0.47***	6.28	0.41***	9.41
88	Tax	0.18	0.16	-0.22	0.72	-0.18*	0.98	-0.14**	1.06
89	Gross Profit	0.06	0.02	-0.04	0.02	-0.08	0.15	0.03	0.06
90	Profitability	-0.27	0.38	-0.26***	1.05	-0.28***	2.28	-0.26***	3.72
91	IPO no R&D	-0.27	0.38	-0.26***	1.05	-0.28***	2.28	-0.26***	3.72
92	Analyst Value	-0.16	0.13	-0.12	0.20	0.08	0.18	0.05	0.13
93	Dividends	0.13	0.08	-0.18	0.47	0.23**	1.52	0.17**	1.50
94	Org. Capital	0.13	0.09	0.17	0.46	0.14	0.61	0.05	0.17
95	ΔRec+ Accrual	-0.25	0.20	0.01	0.00	0.00	0.00	0.00	0.00
96	Age IPO	0.37*	0.68	0.36***	1.92	0.31***	2.63	0.24***	3.00
97	Spreads	-0.36*	0.70	-0.30**	1.38	-0.12	0.41	-0.10	0.54

Table 3

White Reality Check of In-Sample Anomaly Performance

The table displays p-values from a White (2000) reality check bootstrap of the in-sample predictability of anomaly variables. The null hypothesis is that the predictive models, when considered together, do not predict future excess returns. To calculate the reality check, we first estimate predictive regressions of the form:

$$r_{t:t+h} = \alpha + \beta x_t + \varepsilon_{t:t+h} \text{ for } t = 1, \ldots, T-h,$$

where $r_{t:t+h} = (1/h)(r_{t+1} + \cdots + r_{t+h})$, r_t is the continuously compounded S&P 500 return for month t from CRSP including dividends and excess of the monthly risk-free rate from Goyal and Welch (2008), h indicates the forecast horizon in months, and x_t is one of the 97 predictor variables. For each predictive regression, we use the estimated coefficients to forecast future excess returns, and we compare the forecast errors from these predictive models to the forecast errors from a benchmark model that uses the prevailing mean return as the forecast. We bootstrap the test statistics using the stationary bootstrap procedure in Politis and Romano (1994) with 1000 replications and a mean blocksize of 5. The table displays the p-values when we consider *All Predictors*, as well as when we consider the ten most popular cross-sectional predictors (*Popular*) and the ten most statistically significant cross-sectional predictors (*Best Cross-Sectional*). We also examine p-values for four different groupings of the predictor variables using the categories in McLean and Pontiff (2015): (1) *Event*, (2) *Fundamental*, (3) *Market*, and (4) *Valuation*. See Section 3.C of the text for a detailed discussion of these categorizations. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

(1)	(2)	(3)	(4)	(5)
	P-values from White Reality Check			
Predictive Variable	***h*=1**	***h*=3**	***h*=6**	***h*=12**
Panel A: All Predictors				
All Predictors	0.36	0.22	0.23	0.05*
Popular	0.22	0.05*	0.03**	0.01**
Best Cross-sectional	0.25	0.04**	0.03**	0.01**
Event	0.43	0.18	0.20	0.04**
Fundamental	0.12	0.05*	0.04**	0.02**
Market	0.52	0.21	0.11	0.03**
Valuation	0.17	0.22	0.20	0.19
Panel B: VW Predictors				
All Predictors	0.36	0.13	0.12	0.03**
Popular	0.21	0.05**	0.03**	0.01**
Best Cross-sectional	0.19	0.04**	0.02**	0.01**
Event	0.32	0.16	0.16	0.03**
Fundamental	0.09*	0.04**	0.03**	0.02**
Market	0.51	0.20	0.11	0.03**
Valuation	0.24	0.53	0.46	0.29
Panel C: EW Predictors				
All Predictors	0.24	0.30	0.22	0.24
Popular	0.24	0.63	0.44	0.42
Best Cross-sectional	0.42	0.48	0.43	0.40
Event	0.56	0.38	0.19	0.05*
Fundamental	0.58	0.38	0.30	0.14
Market	0.26	0.51	0.37	0.30
Valuation	0.16	0.21	0.20	0.18

Table 3 (continued)
White Reality Check of In-Sample Anomaly Performance

Panel D: First Difference of VW Predictors				
All Predictors	0.43	0.47	0.23	0.19
Popular	0.31	0.19	0.34	0.37
Best Cross-sectional	0.41	0.18	0.36	0.34
Event	0.34	0.31	0.14	0.16
Fundamental	0.52	0.26	0.12	0.11
Market	0.41	0.34	0.44	0.31
Valuation	0.21	0.27	0.15	0.31
Panel E: First Difference of EW Predictors				
All Predictors	0.20	0.33	0.26	0.27
Popular	0.15	0.28	0.15	0.24
Best Cross-sectional	0.26	0.4	0.41	0.30
Event	0.5	0.42	0.11	0.15
Fundamental	0.53	0.50	0.20	0.28
Market	0.20	0.43	0.30	0.35
Valuation	0.15	0.17	0.36	0.21
Panel F: Linearly De-trended VW Predictors				
All Predictors	0.22	0.11	0.11	0.02**
Popular	0.15	0.04**	0.03**	0.01**
Best Cross-sectional	0.14	0.03**	0.02**	0.01**
Event	0.20	0.15	0.15	0.03**
Fundamental	0.07*	0.03**	0.03**	0.02**
Market	0.39	0.17	0.09*	0.03**
Valuation	0.29	0.45	0.45	0.29
Panel G: Linearly De-trended EW Predictors				
All Predictors	0.65	0.27	0.21	0.22
Popular	0.55	0.56	0.41	0.39
Best Cross-sectional	0.35	0.43	0.42	0.41
Event	0.43	0.35	0.17	0.05*
Fundamental	0.41	0.31	0.27	0.12
Market	0.61	0.39	0.32	0.26
Valuation	0.31	0.20	0.20	0.18

Table 4
Summary of Out of Sample Anomaly Performance

The table displays a count of the number of predictive variables that are statistically significant at the 10% level or better. For each anomaly, we estimate an out of sample predictive regression of the form:

$$r_{t:t+h} = \alpha + \beta x_t + \varepsilon_{t:t+h} \text{ for } t = 1, \ldots, T-h,$$

where $r_{t:t+h} = (1/h)(r_{t+1} + \cdots + r_{t+h})$, r_t is the continuously compounded S&P 500 return for month t from CRSP including dividends and excess of the monthly risk-free rate from Goyal and Welch (2008), h indicates the forecast horizon in months, and x_t is one of the 97 predictor variables. To construct aggregate predictors out of cross-sectional anomalies, we calculate the value-weighted and equal-weighted mean across all stocks on each date. Because some of the resulting variables are non-stationary, we also examine the first-difference of these predictors and we examine a linearly de-trended version of these predictors. These permutations result in 582 possible predictors. Of these 582, 414 of them are stationary and have a long enough time series to estimate an out of sample regression. The table shows the number of predictors that are statistically significant relative to the total number of variables considered. Panel A summarizes the results for all predictors, Panel B examines the value-weighted predictors, Panel C examines the equal-weighted predictors, Panel D examines the first-difference of value-weighted predictors, Panel E examines the first-difference of equal-weighted predictors, Panel F examines linearly de-trended value-weighted predictors, and Panel G examines linearly de-trended equal-weighted predictors. In each panel, the first row displays results for all variables in that panel, and the remaining rows examine sub-samples formed on the ten most popular cross-sectional predictors (*Popular*), the ten most statistically significant cross-sectional predictors (*Best Cross-Sectional*), and four different groupings based on the categories in McLean and Pontiff (2015): (1) *Event*, (2) *Fundamental*, (3) *Market*, and (4) *Valuation*. See Section 3.C of the text for a detailed discussion of these categorizations.

(1)	**(2)**	**(3)**	**(4)**	**(5)**
	Number significant predictors / Total number predictors			
Predictive Variable	***h*=1**	***h*=3**	***h*=6**	***h*=12**
Panel A: All Predictors				
All Predictors	11/414	39/414	70/414	82/414
Popular	0/40	6/40	11/40	11/40
Best Cross-sectional	0/53	8/53	11/53	11/53
Event	2/142	10/142	19/142	24/142
Fundamental	3/120	15/120	15/120	18/120
Market	6/102	12/102	33/102	38/102
Valuation	0/50	2/50	3/50	2/50
Panel B: VW Predictors				
All Predictors	8/211	26/211	44/211	50/211
Popular	0/20	4/20	6/20	6/20
Best Cross-sectional	0/27	5/27	7/27	7/27
Event	2/75	6/75	13/75	15/75
Fundamental	2/61	12/61	11/61	13/61
Market	4/51	7/51	18/51	21/51
Valuation	0/24	1/24	2/24	1/24
Panel C: EW Predictors				
All Predictors	3/203	13/203	26/203	32/203
Popular	0/20	2/20	5/20	5/20
Best Cross-sectional	0/26	3/26	4/26	4/26
Event	0/67	4/67	6/67	9/67
Fundamental	1/59	3/59	4/59	5/59
Market	2/51	5/51	15/51	17/51
Valuation	0/26	1/26	1/26	1/26

Table 4
Summary of Out of Sample Anomaly Performance

Panel D: First Difference of VW Predictors				
All Predictors	3/97	3/97	13/97	13/97
Popular	0/10	1/10	3/10	3/10
Best Cross-sectional	0/10	0/10	1/10	1/10
Event	1/33	1/33	3/33	3/33
Fundamental	0/27	1/27	2/27	2/27
Market	2/24	1/24	7/24	8/24
Valuation	0/13	0/13	1/13	0/13
Panel E: First Difference of EW Predictors				
All Predictors	2/96	4/96	12/96	11/96
Popular	0/10	1/10	3/10	3/10
Best Cross-sectional	0/10	1/10	1/10	1/10
Event	0/32	2/32	3/32	2/32
Fundamental	1/27	0/27	1/27	1/27
Market	1/24	2/24	8/24	8/24
Valuation	0/13	0/13	0/13	0/13
Panel F: Linearly De-trended VW Predictors				
All Predictors	2/60	10/60	15/60	18/60
Popular	0/5	1/5	2/5	2/5
Best Cross-sectional	0/9	3/9	5/9	5/9
Event	1/21	3/21	6/21	6/21
Fundamental	1/16	4/16	4/16	5/16
Market	0/14	2/14	4/14	6/14
Valuation	0/9	1/9	1/9	1/9
Panel G: Linearly De-trended EW Predictors				
All Predictors	0/60	4/60	7/60	11/60
Popular	0/5	0/5	1/5	1/5
Best Cross-sectional	0/9	1/9	2/9	2/9
Event	0/21	2/21	2/21	4/21
Fundamental	0/16	1/16	1/16	1/16
Market	0/14	0/14	3/14	5/14
Valuation	0/9	1/9	1/9	1/9

Table 5
Best Out of Sample Predictive Regression Results for each Anomaly

The table reports the mean of the ordinary least squares estimate of β and the Campbell and Thompson (2008) R^2_{OS} statistic from out of sample predictive regression models of the form:

$$r_{t:t+h} = \alpha + \beta x_t + \varepsilon_{t:t+h} \text{ for } t = 1, \ldots, T-h,$$

where $r_{t:t+h} = (1/h)(r_{t+1} + \cdots + r_{t+h})$, r_t is the continuously compounded S&P 500 return for month t from CRSP including dividends and excess of the monthly risk-free rate from Goyal and Welch (2008), h indicates the forecast horizon in months, and x_t is the predictor variable in the first column. For each predictor, we report the version that has the best R^2_{OS}. $\hat{\beta}$ is the time-series mean of the coefficient estimates for each predictor. The Campbell and Thompson R^2_{OS} statistic is calculated as 1 minus the proportional reduction in mean squared forecast error (MSFE) at the h-month horizon for a predictive regression forecast of the S&P 500 log excess return based on the predictor variable in the first column vis-a-vis the prevailing mean benchmark forecast. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		***h*=1**		***h*=3**		***h*=6**		***h*=12**	
	Predictive Variable	$\hat{\boldsymbol{\beta}}$	R^2_{OS}	$\hat{\boldsymbol{\beta}}$	R^2_{OS}	$\hat{\boldsymbol{\beta}}$	R^2_{OS}	$\hat{\boldsymbol{\beta}}$	R^2_{OS}
1	e/p	0.06	-1.22	0.01	-1.87***	0.01	-2.35	-0.03	0.12
2	m/b	0.02	-1.29	0.03	-1.71	0	-0.71***	-0.03	-0.74
3	size	0.09	-1.35	0	0.37***	0.01	0.65***	0.02	0.74***
4	momentum	-0.01	-0.74***	-0.01	-0.57***	-0.01	-0.84***	-0.02	-0.53
5	UMD-reversal	0	-0.57***	0.02	-0.53***	-0.02	0.35	-0.02	1.05*
6	LT reversal	0	-0.17***	0	0.4	-0.01	2.29**	-0.01	7.67**
7	ST reversal	-0.03	-0.72	0	0.39	0.03	0.7*	0	1.91**
8	dividends	-1.83	-0.6***	-0.56	0.1	-0.54	0.18	0.1	0.53
9	price	0.07	0.35	-0.01	0.84***	-0.02	1.67	-0.01	4.76
10	leverage	-0.02	-1.09***	-0.04	-2.45***	-0.01	-3.91	0.02	-6.48
11	volume	0.03	-1.09***	0.05	-0.03	0.03	-0.38	0.01	0.76
12	investment	-0.01	-0.36	-0.02	0.38	-0.01	1.22	-0.03	2.64
13	accruals	0.29	-0.25	0.22	0.4	0.19	0.63	0.12	1.08
14	debt issues	0.17	-1.13***	0.16	-2.58***	0.05	-4.3***	0.07	-7.09***
15	repurchases	-0.03	-0.39***	-0.04	1.98***	-0.06	6.17*	-0.08	14.39
16	seo	0.01	-0.6	0.37	-0.03***	0.27	1.08	0.1	1.81***
17	mergers	-0.43	-0.44***	-0.13	-0.21***	-0.21	1.98***	-0.23	4.2
18	spinoffs	-1.42	-1.53***	-0.19	-1.68***	-0.21	-2.2	0.07	-2.23***
19	z-score	0	0.12***	0.01	1.05***	0.01	0.63***	0.01	1.04***
20	o-score	0	0.69***	0	-0.24***	0	-0.01***	0	1.37***
21	noa	0.02	-0.49	-0.02	1.7	-0.05	6.17	-0.07	11.56
22	R&D/MV	2.64	-0.51	0.51	-0.33	0.69	0.01	0.3	1.3
23	Marketing/MV	3.3	-0.23***	3.18	0.47	2.79	0.67	0.09	0.52
24	Sales Growth	-0.01	1.09	-0.01	3.89	-0.01	5.68	-0.01	10.08
25	CF/MV	-0.16	-0.55	-0.13	-0.77	0	0.62	0.13	2
26	52-Week High	-0.04	-0.31***	-0.02	-0.1***	-0.01	0.3	0	0.64
27	Beta	0.16	0.04	0.1	-0.15	0.04	0.98***	0	3.52***
28	PEAD	0	-0.43	0	-0.08	0	1.05	0	0.22
29	Asset Growth	-0.03	-0.22	-0.06	2.95	-0.09	7.53	-0.1	14.42
30	Share issues DT	-0.03	-0.97***	-0.06	-1.38***	-0.17	-1.52***	-0.18	-4.02***
31	Share issues PW	0.06	-0.97***	-0.06	-1.38	-0.17	-1.52	-0.18	-4.02
32	Idio. Risk	0.12	-0.61	-0.2	-1.14***	-0.28	-0.32***	0.01	1.01
33	Herfindahl	0	-1.18***	0	0.24***	0	0.17***	0	1.28***
34	VAR(Volume)	0.58	-0.17***	0.43	-0.07***	0.1	0.09***	0.02	-0.31***
35	Exch. Switch	-0.45	-0.46***	-0.07	0.61*	-0.09	0.92	0	0***
36	ROE	1.33	-0.38	0.87	0.09	0.42	0.63	-0.94	1.6
37	Volume / MV	0.05	-0.69***	0.04	0.12***	0.03	1.16***	0.01	1.85
38	Amihud	-0.08	0.07	0.03	0.13	0	0.4	-0.01	0.46
39	Volume Trend	-0.11	-0.45	-0.13	-0.58	0.15	-0.79	0	1.68**
40	VAR(CF)	6.47	0.14***	-1.77	0.7	1.62	3.42*	1.28	4.97
41	Div. Initiation	-4.32	-0.92	0.67	-0.07***	3.02	0.54	1.35	1.46

Table 5 (continued)

Best Out of Sample Predictive Regression Results for each Anomaly

42	Div. Omission	-0.86	-0.67***	-1.35	-0.72***	-1.09	-1.58***	-0.69	-5.15
43	Pension Funding	0.37	0.03***	0.38	0.96	0.39	0.97	0.05	0.87***
44	Vol-Momentum	-0.89	-0.54	-0.38	-0.13	0.08	-0.28***	0.13	-0.04
45	G-Score	-0.01	-0.47***	-0.01	0.13***	-0.01	0.5***	-0.01	1.17
46	G-Score 2	0.03	-0.79	0.02	-0.49	0	-0.11	0.01	0.84
47	R&D Increases	-0.01	0.05	-0.01	1.48	-0.01	3.06	-0.01	1.03
48	Rating Down	1.8	-0.57***	0.2	-0.07	0.01	1.28***	0.01	8.92***
49	Moment-Ratings	0.01	-0.99***	-0.09	-0.92***	-0.09	-0.73***	0	-0.29
50	Work. Capital Δ	-0.01	-0.86	-0.02	-1.01	-0.01	-0.54	-0.01	0.16
51	N. Op. Assets Δ	0.27	-0.07	0.26	1.04*	0.27	2.43*	0.15	2.48
52	Age	-0.07	0.28	-0.05	1.82	-0.03	3.84	-0.01	10.46
53	Age-Momentum	0	-0.69	0	-0.6	0	0.5	0	3.04
54	IPOs	-0.01	-0.91***	-0.01	-0.64***	0	-0.09***	0	0.66
55	Seasonality	1.36	-1.26	-0.51	-0.95	-0.43	-0.63	-0.71	0.41***
56	Max	0	-0.51***	-0.04	0.44	-0.01	1.14	-0.03	2.62*
57	ΔSales-ΔInvent	-0.03	0.21	-0.07	1.11	-0.13	1.59***	-0.07	1.84**
58	ΔCpX-ΔIndCpX	-0.02	-0.6***	-0.05	0.07	-0.06	1.54	-0.02	3.81
59	ΔSales-ΔSG&A	0	-0.58***	0	0***	-0.01	2.58	-0.01	8.2**
60	Ent. Comp B/P	0.04	-0.65	-0.05	-0.48	0.11	0.83***	0.09	2.82
61	Lev. Comp B/P	-0.02	-1.13***	-0.05	-1.55***	0	-0.56***	0.03	-0.82***
62	Total XFIN	-0.02	-0.7	-0.11	0.28*	-0.07	0.67*	0.01	1.9
63	Grwth LTNOA	-0.14	-0.37	-0.09	1.39	-0.04	2.56	-0.12	5.9
64	Grwth Inventory	-0.25	-0.32***	-0.13	0.07	-0.19	1.76	-0.2	6.22
65	Rev. Surprises	-1.72	-0.36	-2.9	-0.28	0.35	0.36	-0.5	2.26
66	Operating Lev.	0	-0.91	0	-0.34	0	5.2***	0	5.81***
67	M/B & Accruals	-0.18	-1.01	-0.21	-1.98	-0.16	-0.79***	0.09	-0.68
68	Mom &Reversal	0.57	0.12	-0.54	-0.21***	-0.25	1.31***	-0.51	1.77***
69	Lag(Momentum)	0.01	-0.33	0.01	0.1	0.03	0.9	0.05	2.26
70	Short Interest	-0.02	-0.19***	-0.01	0.9**	-0.01	2.13**	-0.01	4.55***
71	Enterprise Multp	5.23	0.58***	-0.68	7.53***	-0.56	17.48***	-0.65	21.74
72	Sustain Grwth	0	-0.95***	0.01	-1.05***	0	-0.44***	0	-0.43
73	Co-skewness	-0.03	-0.59	-0.03	1.52	-0.03	5.74	-0.02	14.59
74	% Total Accrual	0.29	0.52***	0.19	1.37	0.16	1.17	-0.04	1***
75	% Op. Accrual	-0.01	-1.28***	0.01	-2.1	0	-2.79***	0	-4.7***
76	Earnings Consist	0.02	-1.88	0.02	-2.35	0	-2.35***	0	-4.29***
77	Ind. Momentum	0.01	-1.1***	0.02	-2.04***	-0.17	-2.46	-0.02	-5.49***
78	Δ Recommend	0.07	-1.68***	0	-1.41	-0.01	-1.08	-0.02	0.76
79	Forecast Disp.	-0.07	-0.46***	0.04	-0.33	-0.08	-0.15***	-0.03	0.73***
80	Up Forecast	0.08	-0.11***	0.28	0.73	0.06	-0.45***	0.04	1.06***
81	Down Forecast	-0.01	-0.73	0	-0.38	0	-0.11	0	0.87
82	G Index	0.01	-0.48	0.01	-0.28	-0.01	0.47	0	1.07
83	Sales/Price	0.01	-1.84***	-0.03	-2.12	-0.02	-2.08	-0.02	-1.11
84	Asset Turnover	0.02	-0.46***	0.02	0.07	0	-0.17***	0	0.91
85	ΔAsset Turnover	0.02	0.08***	0.02	6.41***	0.02	16.01***	0.02	28.92***
86	Profit Margin	0.01	-0.84***	-0.01	0.22***	0.01	0.74	0.01	9.67*
87	ΔProfit Margin	-0.05	-1.24***	0.15	1.44***	0.11	3.97***	0.05	5.13***
88	Tax	0.02	-0.67***	0.04	0.27***	0.02	0.61***	0.02	0.95***
89	Gross Profit	-0.02	-0.78	-0.01	-1.15***	-0.01	-1.16	0.02	-1.61
90	Profitability	0.06	-0.39	-0.2	-0.24***	-0.15	0.69	-0.02	1.64
91	IPO no R&D	0.06	-0.39***	-0.2	-0.24	-0.15	0.69	0.03	1.64
92	Analyst Value	0.01	-0.85	0.01	-0.61	0.01	0.42	0.01	0.94
93	Dividends	0.1	-0.51	-0.76	0.25***	0.27	1	0.28	2.58
94	Org. Capital	0.01	-0.54	-0.01	-0.45	0.01	-0.03	0.01	1.01
95	ΔRec+ Accrual	-0.01	-0.35	0	0.22	0	1.02	0	1.71
96	Age IPO	0.1	1.04	0.1	4.64	0.1	6.74	0.09	9.61*
97	Spreads	-0.02	0.44	-0.01	0.41	0	-0.18***	0	0.71

Table 6
White Reality Check of Out of Sample Anomaly Performance

The table displays p-values from a White (2000) reality check bootstrap of the out of sample predictability of anomaly variables. The null hypothesis is that the predictive models, when considered together, do not predict future excess returns. To calculate the reality check, we first estimate predictive regressions of the form:

$$r_{t:t+h} = \alpha + \beta x_t + \varepsilon_{t:t+h} \text{ for } t = 1, \ldots, T-h,$$

where $r_{t:t+h} = (1/h)(r_{t+1} + \cdots + r_{t+h})$, r_t is the continuously compounded S&P 500 return for month t from CRSP including dividends and excess of the monthly risk-free rate from Goyal and Welch (2008), h indicates the forecast horizon in months, and x_t is one of the 97 predictor variables. For each predictive regression, we use the estimated coefficients to forecast future excess returns, and we compare the forecast errors from these predictive models to the forecast errors from a benchmark model that uses the prevailing mean return as the forecast. We bootstrap the test statistics using the stationary bootstrap procedure in Politis and Romano (1994) with 1000 replications and a mean blocksize of 5. We examine p-values when we consider *All Predictors*, as well as when we consider the ten most popular cross-sectional predictors (*Popular*) and the ten most statistically significant cross-sectional predictors (*Best Cross-Sectional*). We also examine p-values for four different groupings of the predictor variables using the categories in McLean and Pontiff (2015): (1) *Event*, (2) *Fundamental*, (3) *Market*, and (4) *Valuation*. See Section 3.C of the text for a detailed discussion of these categorizations. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

(1)	(2)	(3)	(4)	(5)
	P-values from White Reality Check			
Predictive Variable	***h*=1**	***h*=3**	***h*=6**	***h*=12**
Panel A: All Predictors				
All Predictors	0.91	0.44	0.34	0.08*
Popular	0.61	0.87	0.78	0.52
Best Cross-sectional	0.49	0.79	0.70	0.45
Event	0.86	0.60	0.57	0.53
Fundamental	0.54	0.25	0.26	0.11
Market	0.84	0.63	0.27	0.06*
Valuation	0.91	0.92	0.87	0.85
Panel B: VW Predictors				
All Predictors	0.87	0.39	0.32	0.08*
Popular	0.75	0.81	0.69	0.44
Best Cross-sectional	0.64	0.73	0.65	0.36
Event	0.84	0.58	0.55	0.50
Fundamental	0.39	0.18	0.24	0.10
Market	0.92	0.57	0.26	0.06*
Valuation	0.90	0.86	0.79	0.79
Panel C: EW Predictors				
All Predictors	0.83	0.92	0.72	0.53
Popular	0.41	0.84	0.76	0.66
Best Cross-sectional	0.38	0.77	0.71	0.60
Event	1.00	0.93	0.75	0.63
Fundamental	1.00	0.98	0.91	0.72
Market	0.60	0.77	0.58	0.37
Valuation	0.89	0.92	0.88	0.86

Table 6 (continued)
White Reality Check of Out of Sample Anomaly Performance

Panel D: First Difference of VW Predictors				
All Predictors	0.96	1.00	0.92	0.76
Popular	0.62	0.81	0.53	0.42
Best Cross-sectional	0.53	0.69	0.48	0.42
Event	0.99	0.92	0.61	0.54
Fundamental	1.00	0.91	0.59	0.45
Market	0.78	0.99	0.85	0.66
Valuation	0.89	0.85	0.76	0.72
Panel E: First Difference of EW Predictors				
All Predictors	0.70	0.99	0.95	0.82
Popular	0.27	0.72	0.62	0.50
Best Cross-sectional	0.28	0.63	0.47	0.39
Event	1.00	0.90	0.59	0.54
Fundamental	0.98	0.81	0.68	0.42
Market	0.44	0.97	0.91	0.76
Valuation	0.84	0.90	0.81	0.76
Panel F: Linearly De-trended VW Predictors				
All Predictors	0.40	0.37	0.16	0.08*
Popular	0.85	0.30	0.30	0.01***
Best Cross-sectional	0.99	0.68	0.67	0.38
Event	0.19	0.36	0.35	0.36
Fundamental	0.57	0.32	0.18	0.06*
Market	0.77	0.06*	0.01***	0.00***
Valuation	0.96	0.91	0.93	0.90
Panel G: Linearly De-trended EW Predictors				
All Predictors	0.81	0.68	0.46	0.32
Popular	0.93	0.21	0.14	0.00***
Best Cross-sectional	0.98	0.66	0.55	0.38
Event	0.87	0.67	0.48	0.41
Fundamental	0.99	0.94	0.95	0.98
Market	0.44	0.24	0.08*	0.04**
Valuation	0.91	0.85	0.87	0.82