12-2014

The Long of it: Odds That Investor Sentiment Spuriously Predicts Anomaly Returns

Robert F. Stambaugh
University of Pennsylvania

Jianfeng Yu

Yu Yuan

Follow this and additional works at: http://repository.upenn.edu/fnce_papers

Part of the Finance Commons, and the Finance and Financial Management Commons

Recommended Citation

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/fnce_papers/207
For more information, please contact repository@pobox.upenn.edu.
The Long of it: Odds That Investor Sentiment Spuriously Predicts Anomaly Returns

Abstract
Extremely long odds accompany the chance that spurious-regression bias accounts for investor sentiment’s observed role in stock-return anomalies. We replace investor sentiment with a simulated persistent series in regressions reported by Stambaugh, Yu, and Yuan (2012), who find higher long-short anomaly profits following high sentiment, due entirely to the short leg. Among 200 million simulated regressors, we find none that support those conclusions as strongly as investor sentiment. The key is consistency across anomalies. Obtaining just the predicted signs for the regression coefficients across the 11 anomalies examined in the above study occurs only once for every 43 simulated regressors.

Keywords
investor sentiment, anomalies, spurious regressors

Disciplines
Finance | Finance and Financial Management

This journal article is available at ScholarlyCommons: http://repository.upenn.edu/fince_papers/207
The long of it: Odds that investor sentiment spuriously predicts anomaly returns

by*

Robert F. Stambaugh, Jianfeng Yu, and Yu Yuan

February 16, 2014

Abstract

Extremely long odds accompany the chance that spurious-regression bias accounts for investor sentiment’s observed role in stock-return anomalies. We replace investor sentiment with a simulated persistent series in regressions reported by Stambaugh, Yu and Yuan (2012), who find higher long-short anomaly profits following high sentiment, due entirely to the short leg. Among 200 million simulated regressors, we find none that support those conclusions as strongly as investor sentiment. The key is consistency across anomalies. Obtaining just the predicted signs for the regression coefficients across the 11 anomalies examined in the above study occurs only once for every 43 simulated regressors.

JEL classifications: G12, G14, C18
Keywords: investor sentiment, anomalies, spurious regressors

* We are grateful for the comments from an anonymous referee. Author affiliations/contact information:


Yu: Associate Professor of Finance, The Carlson School of Management, University of Minnesota, 321 19th Avenue South, Suite 3-122, Minneapolis, MN 55455, phone 612-625-5498, email jianfeng@umn.edu.

Yuan (corresponding author): Associate Professor, Shanghai Advanced Institute of Finance, 211 West Huaihai Road, Shanghai, China, 200030, and Fellow, Wharton Financial Institutions, phone: +86-21-62932114, email: yyuan@saif.sjtu.edu.cn.
1. Introduction

Caution is warranted when inferring that a highly autocorrelated variable can predict asset returns. One reason is the possibility of a “spurious” regressor: If the unobserved expected return on an asset is time-varying and persistent, another persistent variable having no true relation with return can appear to predict return in a finite sample. Ferson, Sarkissian, and Simin (2003) demonstrate how the potential for such regressors complicates the task of assessing return predictors, and they explain how the underlying mechanism relates to the spurious regression problem analyzed by Yule (1926) and Granger and Newbold (1974). Ferson et al. also explain how data mining interacts with the problem of spurious regressors. When the potential for spurious regressors exists (i.e., a persistent time-varying expected return), data mining produces an especially greater chance of finding a series that appears to predict returns but does so only spuriously.

The stronger is the prior motivation for entertaining a series as a return predictor, the weaker is the concern that its apparent predictive ability is spurious. One quantity with strong prior motivation as a return predictor is market-wide investor sentiment. At least as early as Keynes (1936), numerous authors have considered the possibility that a significant presence of sentiment-driven investors can cause prices to depart from fundamental values, thereby creating a component of future returns that corrects such mispricing. Baker and Wurgler (2006) and Stambaugh, Yu and Yuan (2012), among others, find that investor sentiment and/or consumer confidence exhibits an ability to predict returns on various classes of stocks and investment strategies. These studies also refine the prior motivation of investor sentiment as a predictor. For example, Baker and Wurgler (2006) argue that sentiment should play a stronger role among stocks that are more difficult to value. In support of that hypothesis, they find sentiment exhibits greater ability to predict returns on small stocks, young stocks, high volatility stocks, unprofitable stocks, non-dividend-paying stocks, extreme growth stocks, and distressed stocks. Stambaugh, Yu, and Yuan (2012) hypothesize that when market-wide sentiment is combined with Miller’s (1977) argument about the effects of short-sale impediments, overpricing due to high sentiment is more likely than underpricing.

---

1 A regressor with prior motivation also often violates the spurious-regressor setting in Ferson, Sarkissian, and Simin (2003), wherein the regressor bears no relation to return. Instead, the innovation in the regressor is often correlated with contemporaneous return, whether or not the regressor predicts future return. Such a correlation is especially likely for a regressor that is a valuation ratio, such as dividend yield. The finite-sample bias that arises in such a setting is analyzed by Stambaugh (1999).

due to low sentiment. Their results support that argument, in that sentiment predicts profits on the short legs of a large set of anomaly-based long-short strategies, whereas sentiment exhibits no ability to predict long-leg profits.

Despite the prior motivation for the properties that investor sentiment exhibits empirically as a predictor of anomaly returns, one might nevertheless be concerned that sentiment is simply a spurious predictor. Such a concern might be prompted, for example, by the results of Novy-Marx (2013b), who reports that returns on various subsets of anomalies can apparently be predicted by seemingly unlikely variables such as sunspots and planetary positions.\(^3\) This study assesses the odds that investor sentiment’s observed ability as a predictor can be achieved by a spurious regressor. We focus on the role of consistency across multiple return series and hypotheses. To understand the value of consistency, suppose the true expected returns across a number of portfolios possess some independent variation, but each expected return’s true correlation with investor sentiment has the same sign. The greater the number of portfolios, the more difficult it becomes to find a spurious regressor that will exhibit finite-sample predictive ability consistently across portfolios comparable to that of investor sentiment. Our setting for exploring the role of consistency is that of Stambaugh, Yu, and Yuan (2012). That study examines 11 different anomalies and finds consistent results across those anomalies in support of three hypotheses: (i) a positive relation between current sentiment and future long-short return spreads, (ii) a negative relation between current sentiment and future short-leg returns, and (iii) no relation between current sentiment and future long-leg returns. We simply ask how likely it is that such hypotheses are supported as strongly by a randomly generated spurious regressor used in place of investor sentiment.

Out of 200 million simulated regressors, we find none that jointly support the three hypotheses in Stambaugh, Yu, and Yuan (2012) as strongly as investor sentiment. The odds are still quite long if one looks at just one of the three hypotheses. For example, comparably strong and consistent support for the first hypothesis—a positive relation between sentiment and the long-short return spread—occurs once in every 28,500 simulated regressors. For the second hypothesis—a negative relation between sentiment and short-leg returns—comparable support occurs once in every 105,000 regressors. If one sets aside any consideration of strength (t-statistics) and simply looks at the signs of regression coefficients dictated by the first two hypotheses, even then only one in every 43 simulated regressors achieves the consistency exhibited with investor sentiment.

\(^3\)Indeed, a preliminary version of that study presented such results in the context of spurious regressors.
2. Empirical setting and simulation results

The empirical setting we analyze here focuses on the main set of regression results reported by Stambaugh, Yu, and Yuan (2012), hereafter SYY. That study estimates the regression,

\[ R_{i,t} = a + bS_{t-1} + cMKT_t + dSMB_t + eHML_t + u_t, \] (1)

where \( R_{i,t} \) is the excess return in month \( t \) on an anomaly strategy’s long leg, short leg, or the difference, \( S_{t-1} \) is the level of the investor-sentiment index of Baker and Wurgler (2006) at the end of month \( t - 1 \), and \( MKT_t, SMB_t, \) and \( HML_t \) are the returns on month \( t \) on the three stock-market factors defined by Fama and French (1993). SYY examine 11 anomalies documented previously in the literature:

1. Failure probability (Campbell, Hilscher, and Szilagyi, 2007)
2. Distress (Ohlson, 1980)
4. Composite equity issues (Daniel and Titman, 2006)
5. Total accruals (Sloan, 1996)
7. Momentum (Jegadeesh and Titman, 1993)
8. Gross profitability (Novy-Marx, 2013a)
9. Asset growth (Cooper, Gulen, and Schill, 2008)
11. Investment-to-assets (Titman, Wei, and Xie, 2004, and Xing, 2008)

As in SYY, the sample period is from August 1965 through January 2008 for all but anomaly (1), whose data begin in December 1974, and anomalies (2) and (10), whose data begin in January 1972. For each anomaly, SYY examine the long-short strategy using deciles 1 and 10 of a sort based on the anomaly variable, with the long leg being the decile with the highest average return. SYY also examine a combination strategy that takes equal positions across the long-short strategies constructed in any given month.

The coefficient of interest in equation (1) is \( b \). SYY (cf. table 5) report results of estimating \( b \) for each of the 11 anomalies, as well as the combination strategy, in three sets of regressions that relate to the three hypotheses explored in that study. For the first hypothesis, \( R_{i,t} \) is the long-short return difference, and the estimate \( \hat{b} \) has the predicted
positive sign for all 11 anomalies. The t-statistic for \( \hat{b} \), based on the heteroskedasticity-consistent standard error of White (1980), ranges from 0.22 to 3.38 across the individual anomalies and equals 2.98 for the combination strategy. For the second hypothesis, \( R_{i,t} \) is the short-leg return, and \( \hat{b} \) has the predicted negative sign for all 11 anomalies. The t-statistic ranges from \(-1.11\) to \(-3.58\) across the individual anomalies and equals \(-3.01\) for the combination strategy. The third hypothesis, in which \( R_{i,t} \) is the long-leg return, predicts \( b \) should be roughly zero. In these regressions, the signs of \( \hat{b} \) are mixed across the individual anomalies (7 positive, 4 negative), with t-statistics ranging from \(-2.07\) to \(1.44\), and the combination strategy has a t-statistic of \(0.15\). When viewed collectively across the estimated 36 regressions (12 for each hypothesis), the SYY results appear to present fairly strong support for all three hypotheses explored.

In this study, we ask how likely it is that a spurious predictor would support the three SYY hypotheses as strongly as investor sentiment. We randomly generate a predictor series \( x_t \), use it to replace \( S_t \), and then re-estimate equation (1) for the same 36 regressions summarized above. That procedure is repeated 200 million times. Each predictor series \( x_t \) is generated as a first-order autoregressive process with normal innovations and autocorrelation equal to 0.988, which equals the sample autocorrelation of \( S_t \) adjusted for the first-order bias correction in Marriott and Pope (1954) and Kendall (1954).

### 2.1. Joint comparisons of t-statistics

To judge whether \( x_t \) supports a given hypothesis as strongly as \( S_t \), we ask whether the t-statistics for \( \hat{b} \), viewed jointly across anomalies, are as favorable to the hypothesis as those produced using \( S_t \). To determine this condition in the case of the first hypothesis, for which \( R_{i,t} \) is the long-short return difference, define \( \bar{t}_i^S \) as the \( i \)-th highest t-statistic for \( \hat{b} \) among the 11 anomalies when \( S_t \) is used. Similarly define \( \bar{t}_i^x \) as the \( i \)-th highest t-statistic for \( \hat{b} \) among the 11 anomalies when \( x_t \) is used. Let \( t_C^S \) denote the t-statistic for the combination strategy when \( S_t \) is used, and let \( t_C^x \) denote the corresponding t-statistic when \( x_t \) is used. Then \( x_t \) supports the first hypothesis \((b > 0)\) as strongly as \( S_t \) if \( \bar{t}_i^x \geq \bar{t}_i^S \) for \( i = 1, \ldots, 11 \) and \( t_C^x \geq t_C^S \).

Only once in every 28,500 generated \( x_t \) series, on average, is the first hypothesis supported as strongly by \( x_t \) as by \( S_t \). This result is reported in the last row of the first column of Table 1. The other rows display the frequencies with which fewer of the above inequalities are satisfied. For example, the first row of the same column reports that at least one of the 11
values of $\bar{t}_i^x$ exceeds the corresponding value of $\bar{t}_i^S$ once in each 22 generated $x_t$ series. The sharp increase in values as one moves down the column illustrates the dramatic effect of requiring consistency across multiple anomalies. Just finding an $x_t$ for which more than half of the $\bar{t}_i^x$ values exceed the corresponding $\bar{t}_i^S$ values happens only once in every 833 $x_t$ series. The next-to-last row reports that, for just the combination strategy, the t-statistic obtained with $x_t$ exceeds that obtained with $S_t$ once in every 67 series.

The odds for a spurious regressor become even longer when considering the second hypothesis, as we see from the second column of Table 1. That hypothesis is supported as strongly by $x_t$ as it is by $S_t$ only once in every 105,000 series. The inequality conditions here are essentially just the reverse of those earlier, since $R_{i,t}$ is now the short-leg return and the prediction is instead that $b < 0$. Let $\bar{t}_i^S$ denote the i-th lowest t-statistic for $\hat{b}$ when $S_t$ is used, and let $\bar{t}_i^x$ denote the i-th lowest t-statistic when $x_t$ is used. Then $x_t$ supports the second hypothesis as strongly as $S_t$ if $\bar{t}_i^x \leq \bar{t}_i^S$ for $i = 1, \ldots, 11$ and $t_C \leq t_C^x$. As with the first hypothesis, the effects of requiring consistency across the separate regressions are dramatic. Even for just the single regression with the combination strategy, however, obtaining a negative t-statistic greater in magnitude than that obtained with $S_t$ occurs only once in every 169 series.

The third hypothesis is that $b = 0$. In order for that hypothesis to be supported at least as strongly by a randomly generated $x_t$ as it is by $S_t$, we require $x_{t-1}$ to be as consistently weak as $S_{t-1}$ in its ability to predict $R_{i,t}$, now defined as the long-leg return. For this case, let $|t|_i^S$ denote the i-th smallest t-statistic in absolute value when $S_t$ is used, and let $|t|_i^x$ denote the i-th smallest t-statistic in absolute value when $x_t$ is used. Then $x_t$ supports the third hypothesis as strongly as $S_t$ if $|t|_i^x \leq |t|_i^S$ for $i = 1, \ldots, 11$ and $|t_C^x| \leq |t_C^S|$.

While the odds for a spurious regressor improve when considering just the third hypothesis, they are still rather long. Again we see the effect of consistency when requiring the absence of an apparent relation with the regressor. Only once in every 919 randomly generated $x_t$ series do we find one that is as consistently unsuccessful in predicting long-leg returns.

Of course, the story does not end with simply considering each of the three hypotheses in isolation. As SYY explain, these hypotheses arise as a set of joint implications, developed by combining the presence of market-wide swings in sentiment with the argument in Miller (1977) that short-sale impediments allow overpricing to be more prevalent than underpricing. The final two columns report the frequencies with which a spurious regressor $x_t$ supports more than one hypothesis as strongly as $S_t$, where comparable support of each individual
hypothesis is judged as before. Only one spurious regressor out of 468,000 supports the first
two hypotheses as strongly as investor sentiment. When we look for a spurious regressor that
supports all three hypotheses as strongly as investor sentiment, we actually find none among
200 million simulated series. When confining the exercise to just the single regressions using
the combination strategy, we still find that only one spurious regressor out of every 6,580
simultaneously supports each of the three hypotheses as strongly as investor sentiment.

2.2. Joint-comparison benchmarks

As the above analysis illustrates, the consistency of results across multiple anomalies and
hypotheses makes it especially unlikely that such results are produced by a spurious regressor.
While simultaneous joint comparisons reveal the importance of consistency, they can also
make interpreting the strength of the results less straightforward. Each number in Table 1
essentially gives the reciprocal of the probability under the “null hypothesis”—a spurious
predictor—of obtaining a sample outcome at least as extreme as the one actually observed
using the sentiment series $S_t$. However, when the comparison involves a vector of statistics,
as opposed to a single statistic, the corresponding probability can be fairly low even if the
sample outcome is considerably less extreme than the sample outcome that was actually
observed. If considerably less extreme outcomes also have low probabilities under the null,
then it becomes difficult to interpret the low probability associated with outcomes more
extreme than the actual outcome.\footnote{We are grateful to the referee for raising this issue.}

Interpreting the values in Table 1 becomes easier in the presence of benchmark values
that reflect what one expects the values in Table 1 to be when the actual sentiment series $S_t$ is
replaced by a truly spurious predictor. Table 2 contains such benchmark values, computed by
replacing the t-statistics based on the sentiment series $S_t$ with t-statistics based on a spurious
regressor $y_t$. That is, rather than tabulating how often a spurious regressor $x_t$ supports the
SYY hypothesis as well as the actual series $S_t$, we tabulate how often a spurious regressor
$x_t$ does as well as another spurious regressor $y_t$. A new series $y_t$ is drawn for each draw of
the series $x_t$.

Consider, for example, the frequency with which a spurious regressor $x_t$ jointly supports
the three SYY hypotheses across all anomalies as strongly as the actual regressor $S_t$. Recall
from Table 1 that we find this frequency to be less than one in 200 million. When $S_t$ is
replaced by a truly spurious regressor $y_t$, we see from the bottom-right entry in Table 2 that
one spurious regressor $x_t$ out of about 71 supports the three SYY hypotheses as strongly as $y_t$. In other words, the Table 2 value of 71 is a benchmark for interpreting the Table 1 value of 200 million: it is what one expects the Table 1 value to be if $S_t$ is truly spurious. Dividing the Table 1 value by the Table 2 value gives what might be characterized as the “effective” value of the former. For example, dividing 200 million by 71 gives an effective value of about 2.8 million—still very large. Similar comparisons to Table 2 can be made for other values in Table 1. For example, recall from Table 1 that only one spurious regressor out of 468,000 supports the first two SYY hypotheses as strongly as $S_t$. The corresponding benchmark value in Table 2 is 4.4, and dividing 468,000 by 4.4 still gives over 106,000. In general we see that, while the joint-comparison issue is important, interpreting the Table 1 values in light of the Table 2 benchmarks still yields the overall conclusion that the SYY results are extremely unlikely if $S_t$ is a spurious regressor.

2.3. Additional comparisons

To judge whether a spurious regressor supports the SYY hypotheses as strongly as the actual investor sentiment series, one must define “supports as strongly.” While the definition employed above in Tables 1 and 2 seems a reasonable way to capture the consistency of results across anomalies, there are of course alternative definitions. For example, we could instead examine the $k$ least favorable t-statistics for a given hypothesis, comparing those produced by $x_t$ to those obtained using $S_t$. To illustrate, let $k = 1$ and consider the first hypothesis, which predicts $b > 0$ when $R_{i,t}$ is the long-short return difference. The lowest t-statistic produced by $S_t$ among the 11 anomalies is equal to 0.22, and less than one $x_t$ series out of every 50 produces a minimum t-statistic greater than that value. For the second hypothesis, which predicts $b < 0$ when $R_{i,t}$ is the short-leg return, the weakest t-statistic using $S_t$ is -1.11, and only one $x_t$ in every 2,300 produces a weakest statistic less than -1.11. Now let $k = 2$, and note that the second-lowest t-statistic produced by $S_t$ for the first hypothesis equals 0.76. Only one $x_t$ series out of every 163 produces a lowest t-statistic greater than 0.22 as well as a second-lowest t-statistic greater than 0.76. With hypothesis 2, for only one $x_t$ out of 10,000 are the two weakest t-statistics more favorable to the hypothesis than the two weakest t-statistics using $S_t$. Proceeding through additional $k$ values and the remaining third hypothesis would produce a table in the same format as Table 1, with entries in the final three rows identical to those in Table 1 and larger entries in the first ten rows, corresponding to longer odds.\(^5\) Thus, comparing the weakest results across the individual anomalies would

\(^5\)To see this, note that the $k$-th row of Table 1 reports the frequency with which any $k$ of the ordered t-statistics using $x_t$ is as favorable to the given hypothesis as are the corresponding ordered t-statistics using
deliver a similar message as Table 1, if anything even more strongly.

Of course, conducting joint comparisons of weakest results raises the same benchmarking issue discussed in the previous subsection. That is, an alternative version of Table 1 based on comparing weakest results could be accompanied by the corresponding weakest-result version of Table 2. For example, when \( k = 1 \), the alternative Table 1 values of 50 and 2,300 reported above for the first and second hypotheses have corresponding “effective” values of 25 and 1,150 when divided by the values that would appear in the alternative version of Table 2. Similarly, when \( k = 2 \), the alternative Table 1 values of 163 and 10,000 reported above have corresponding effective values of 70 and 4,367. As before, the low frequencies still seem low when interpreted in the context of joint comparisons.

Another approach that to some degree captures consistency across anomalies is simply comparing median t-statistics. For example, across the 11 individual anomalies as well as the combination strategy, the median t-statistic for the first hypothesis equals 2.41 using \( S_t \), and one \( x_t \) out of every 1,650 produces a median t-statistic as large. For the second hypothesis, the median t-statistic using \( S_t \) equals -2.57, and one \( x_t \) out of every 1,186 produces a median t-statistic greater in negative magnitude. Only one \( x_t \) out of every 7,103 produces median t-statistics that are simultaneously as favorable to both hypotheses. For the third hypothesis, the median absolute t-statistic using \( S_t \) is 0.46. One \( x_t \) out of every 15 produces a median absolute t-statistic that low, but only one \( x_t \) out of 562,000 does so while simultaneously producing median statistics as favorable to the first two hypotheses as those obtained using \( S_t \). The effective frequency of such an outcome is still less than one out of 123,000 if one adjusts for the joint-comparison issue in the same manner as discussed earlier.

The average t-statistic across anomalies says little about consistency across anomalies. Nevertheless, it appears rather unlikely that a spurious regressor can produce even comparably favorable average t-statistics. For example, the averages of the SYY-reported t-statistics across the 11 anomalies and the combination strategy are 2.14 and -2.38 for the first and second hypotheses, respectively. The average absolute value of the SYY-reported t-statistics is 0.69 for the third hypothesis. An average t-statistic supporting the first hypothesis as strongly (i.e., greater than 2.14) is produced by one \( x_t \) out of every 554. An average t-statistic supporting the second hypothesis as strongly (i.e., less than -2.38) occurs for one \( x_t \) out of every 1,393. Average t-statistics simultaneously supporting both hypotheses as strongly occur once every 2,412. An \( x_t \) producing that simultaneous support for the first hypothesis is then unlikely in the context of joint comparisons.

\( S_t \). The \( k \)-th row of the alternative table would consider instead the least favorable \( k \) t-statistics, constituting only a subset of the outcomes included in the frequency in Table 1.
two hypotheses while also being as favorable to the third hypothesis—delivering an average absolute t-statistic less than 0.69—occurs only once in every 237,000. Adjusting for the joint comparison issue still leaves that effective frequency at less than one in every 53,000.

Finally, fairly unlikely is just the possibility that a spurious regressor would give $\hat{b}$'s with the predicted signs consistently across all anomalies. Table 3 reports the frequencies with which a spurious regressor gives the predicted sign across anomalies for the long-short difference (first hypothesis) and the short-leg return (second hypothesis). For the first hypothesis, one in every 25 spurious regressors gives the predicted positive sign for all 11 anomalies. For the second hypothesis, the frequency of getting the predicted negative sign for all 11 anomalies is one in every 21. A spurious predictor that produces all 22 coefficients with the predicted signs, as does investor sentiment, occurs only once in every 43 randomly generated regressors.

3. Conclusions

It appears to be extremely unlikely that the observed role of investor sentiment in stock-return anomalies can be filled by a spurious regressor. Out of 200 million simulated regressors, we find none. These very long odds—seemingly no better than those attached to winning the Powerball Jackpot with a single play—reflect the consistency with which investor sentiment produces results across multiple anomalies for the three SYY hypotheses. Simultaneous support of the SYY hypotheses is important, by itself, in that the odds of a spurious regressor supporting them as strongly as investor sentiment are only 1 in 6,580 even when all of the anomalies are combined into a single long-short strategy. It is the consistency across the individual anomalies, however, that raises the highest hurdle for a spurious regressor to clear in order to play the role of investor sentiment.

---

6Powerball is a multi-state lottery in which the odds of a single combination of numbers claiming a share of the top “Jackpot” prize are roughly 1 in 175 million.
Table 1
Number of Randomly Generated Predictors Required to Obtain One Predictor That Produces Results as Strong as Investor Sentiment

The table reports the reciprocal of the frequency with which a randomly generated predictor $x_t$ produces results as strong as investor sentiment $S_t$ when $x_t$ replaces $S_t$ in the regression,

$$R_{i,t} = a + bS_{t-1} + cMKT_t + dSMB_t + eHML_t + u_t,$$

where $R_{i,t}$ is the excess return in month $t$ on an anomaly’s long leg, short leg, or the difference, $S_t$ is the level of the investor-sentiment index of Baker and Wurgler (2006), and $MKT_t$, $SMB_t$ and $HML_t$ are the three stock-market factors defined in Fama and French (1993). The predictor $x_t$ is generated as a first-order autoregression with autocorrelation equal to 0.988, the bias-corrected estimate of the autocorrelation of $S_t$.

Let $\bar{t}_i^S$ denote the $i$-th highest t-statistic for $\hat{b}$ (the estimate of $b$) among the 11 anomalies when $S_t$ is used, and let $\bar{t}_i^x$ denote the $i$-th highest t-statistic when $x_t$ is used. Let $\underline{t}_i^S$ denote the $i$-th lowest t-statistic for $\hat{b}$ when $S_t$ is used, and let $\underline{t}_i^x$ denote the $i$-th lowest t-statistic when $x_t$ is used. Let $|t|_i^S$ denote the $i$-th smallest t-statistic in absolute value when $S_t$ is used, and let $|t|_i^x$ denote the $i$-th smallest t-statistic in absolute value when $x_t$ is used. The row for $j$ anomalies reflects the frequency with which the following conditions are satisfied:

$\bar{t}_i^x \geq \bar{t}_i^S$ occurred at least $j$ times among $i = 1, \ldots, 11$, in the long-short column.

$\underline{t}_i^x \leq \underline{t}_i^S$ occurred at least $j$ times among $i = 1, \ldots, 11$, in the short-leg column.

$|t|_i^x \leq |t|_i^S$ occurred at least $j$ times among $i = 1, \ldots, 11$, in the long-leg column.

The “combination” row reflects the frequencies with which a simulated predictor produces t-statistics satisfying the above inequalities when $R_{i,t}$ is an equally weighted combination of the 11 anomaly strategies. The final row reflects the frequencies with which the above inequalities are satisfied for 11 anomalies as well as the combination strategy. The last two columns reflect the frequencies with which the inequalities are satisfied jointly across the previous columns.

<table>
<thead>
<tr>
<th>Comparisons</th>
<th>(1) Long–Short</th>
<th>(2) Short Leg</th>
<th>(3) Long Leg</th>
<th>(1) and (2)</th>
<th>(1), (2), and (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 anomaly</td>
<td>22</td>
<td>39</td>
<td>1.2</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2 anomalies</td>
<td>57</td>
<td>77</td>
<td>1.5</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>3 anomalies</td>
<td>124</td>
<td>146</td>
<td>1.9</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>4 anomalies</td>
<td>251</td>
<td>288</td>
<td>2.6</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>5 anomalies</td>
<td>469</td>
<td>616</td>
<td>3.7</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>6 anomalies</td>
<td>833</td>
<td>1,310</td>
<td>5.4</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>7 anomalies</td>
<td>1,460</td>
<td>2,950</td>
<td>8.5</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>8 anomalies</td>
<td>2,570</td>
<td>5,700</td>
<td>14</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>9 anomalies</td>
<td>4,740</td>
<td>11,400</td>
<td>25</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>10 anomalies</td>
<td>10,000</td>
<td>28,400</td>
<td>51</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>11 anomalies</td>
<td>28,500</td>
<td>105,000</td>
<td>143</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Combination</td>
<td>67</td>
<td>169</td>
<td>13</td>
<td>221</td>
<td>6,580</td>
</tr>
<tr>
<td>11 plus the combination</td>
<td>28,500</td>
<td>105,000</td>
<td>919</td>
<td>468,000</td>
<td>$&gt; 200,000,000^a$</td>
</tr>
</tbody>
</table>

$^a$There were zero cases obtained among the 200,000,000 predictors randomly generated.
Table 2
Benchmark Number of Randomly Generated Predictors Required to Obtain One Predictor That Produces Results as Strong as Another Random Predictor

The table reports the reciprocal of the frequency with which a randomly generated predictor $x_t$ produces results as strong as another randomly generated predictor $y_t$ when $x_t$ and $y_t$ replace $S_t$ in the regression,

$$R_{i,t} = a + bS_{t-1} + cMKT_t + dSMB_t + eHML_t + u_t,$$

where $R_{i,t}$ is the excess return in month $t$ on an anomaly’s long leg, short leg, or the difference, $S_t$ is the level of the investor-sentiment index of Baker and Wurgler (2006), and $MKT_t$, $SMB_t$ and $HML_t$ are the three stock-market factors defined in Fama and French (1993). The predictor $x_t$ and $y_t$ are generated as a first-order autoregression with autocorrelation equal to 0.988, the bias-corrected estimate of the autocorrelation of $S_t$.

Let $\bar{t}_i^x$ denote the $i$-th highest t-statistic for $\hat{b}$ (the estimate of $b$) among the 11 anomalies when $y_t$ is used, and let $\bar{t}_i^y$ denote the $i$-th highest t-statistic when $x_t$ is used. Let $\underline{t}_i^y$ denote the $i$-th lowest t-statistic for $\hat{b}$ when $y_t$ is used, and let $\underline{t}_i^x$ denote the $i$-th lowest t-statistic when $x_t$ is used. Let $|t|_i^y$ denote the $i$-th smallest t-statistic in absolute value when $y_t$ is used, and let $|t|_i^x$ denote the $i$-th smallest t-statistic in absolute value when $x_t$ is used. The row for $j$ anomalies reflects the frequency with which the following conditions are satisfied:

- $\bar{t}_i^x \geq \bar{t}_i^y$ occurred at least $j$ times among $i = 1, \ldots, 11$, in the long-short column.
- $\underline{t}_i^x \leq \underline{t}_i^y$ occurred at least $j$ times among $i = 1, \ldots, 11$, in the short-leg column.
- $|t|_i^x \leq |t|_i^y$ occurred at least $j$ times among $i = 1, \ldots, 11$, in the long-leg column.

The “combination” row reflects the frequencies with which a simulated predictor produces t-statistics satisfying the above inequalities when $R_{i,t}$ is an equally weighted combination of the 11 anomaly strategies. The final row reflects the frequencies with which the above inequalities are satisfied for 11 anomalies as well as the combination strategy. The last two columns reflect the frequencies with which the inequalities are satisfied jointly across the previous columns.

<table>
<thead>
<tr>
<th>Comparisons</th>
<th>(1) Long–Short</th>
<th>(2) Short Leg</th>
<th>(3) Long Leg</th>
<th>(1) and (2)</th>
<th>(1), (2), and (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 anomaly</td>
<td>1.4</td>
<td>1.4</td>
<td>1.1</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2 anomalies</td>
<td>1.6</td>
<td>1.5</td>
<td>1.2</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>3 anomalies</td>
<td>1.7</td>
<td>1.7</td>
<td>1.3</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>4 anomalies</td>
<td>1.8</td>
<td>1.8</td>
<td>1.5</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>5 anomalies</td>
<td>1.9</td>
<td>1.9</td>
<td>1.7</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>6 anomalies</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>7 anomalies</td>
<td>2.1</td>
<td>2.1</td>
<td>2.4</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>8 anomalies</td>
<td>2.3</td>
<td>2.3</td>
<td>3.0</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>9 anomalies</td>
<td>2.5</td>
<td>2.5</td>
<td>3.9</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>10 anomalies</td>
<td>2.8</td>
<td>2.8</td>
<td>5.8</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>11 anomalies</td>
<td>3.5</td>
<td>3.5</td>
<td>11.5</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Combination</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.2</td>
<td>4.4</td>
</tr>
<tr>
<td>11 plus the combination</td>
<td>3.5</td>
<td>3.5</td>
<td>16.3</td>
<td>4.4</td>
<td>70.8</td>
</tr>
</tbody>
</table>

11
The table reports the reciprocal of the frequency with which a randomly generated predictor $x_t$ produces an estimate of $b$ with the predicted sign when $x_t$ replaces $S_t$ in the regression,

$$R_{i,t} = a + bS_{t-1} + cMKT_t + dSMB_t + eHML_t + u_t,$$

where $R_{i,t}$ is the excess return in month $t$ on an anomaly’s long leg, short leg, or the difference, $S_t$ is the level of the investor-sentiment index of Baker and Wurgler (2006), and $MKT_t, SMB_t$ and $HML_t$ are the three stock-market factors defined in Fama and French (1993). The predictor $x_t$ is generated as a first-order autoregression with autocorrelation equal to 0.988, the bias-corrected estimate of the autocorrelation of $S_t$.

The row for $j$ anomalies reflects the frequency with which a simulated predictor produces an estimate of $b$ for at least $j$ anomalies with the predicted sign (positive in the long-short column and negative in the short-leg column). The “combination” row reflects the frequency with which a simulated predictor produces an estimate of $b$ with the predicted sign when $R_{i,t}$ is an equally weighted combination of the 11 anomaly strategies. The last column reflects the frequencies with which the predicted signs are obtained jointly across the previous columns.

<table>
<thead>
<tr>
<th>Comparisons</th>
<th>(1) Long-Short</th>
<th>(2) Short Leg</th>
<th>(1) and (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 anomaly</td>
<td>1.0</td>
<td>1.1</td>
<td>–</td>
</tr>
<tr>
<td>2 anomalies</td>
<td>1.1</td>
<td>1.1</td>
<td>–</td>
</tr>
<tr>
<td>3 anomalies</td>
<td>1.3</td>
<td>1.3</td>
<td>–</td>
</tr>
<tr>
<td>4 anomalies</td>
<td>1.4</td>
<td>1.4</td>
<td>–</td>
</tr>
<tr>
<td>5 anomalies</td>
<td>1.7</td>
<td>1.7</td>
<td>–</td>
</tr>
<tr>
<td>6 anomalies</td>
<td>2.0</td>
<td>2.0</td>
<td>–</td>
</tr>
<tr>
<td>7 anomalies</td>
<td>2.5</td>
<td>2.5</td>
<td>–</td>
</tr>
<tr>
<td>8 anomalies</td>
<td>3.3</td>
<td>3.3</td>
<td>–</td>
</tr>
<tr>
<td>9 anomalies</td>
<td>4.9</td>
<td>4.9</td>
<td>–</td>
</tr>
<tr>
<td>10 anomalies</td>
<td>8.8</td>
<td>8.5</td>
<td>–</td>
</tr>
<tr>
<td>11 anomalies</td>
<td>25</td>
<td>21</td>
<td>–</td>
</tr>
<tr>
<td>Combination</td>
<td>2.0</td>
<td>2.0</td>
<td>2.2</td>
</tr>
<tr>
<td>11 plus the combination</td>
<td>25</td>
<td>21</td>
<td>43</td>
</tr>
</tbody>
</table>
References


