The History of the Cross Section of Stock Returns*

Juhani T. Linnainmaa Michael Roberts

February 2016

Abstract

Using accounting data spanning the 20th century, we show that most accounting-based return anomalies are spurious. When we take anomalies out-of-sample by moving either backwards or forwards in time, their average returns decrease and volatilities increase. These patterns emerge because data-snooping works through t-values, and an anomaly's t-value is high if its average return is high or volatility low. The average anomaly's in-sample Sharpe ratio is biased upwards by a factor of three. The data-snooping problem is so severe that we would expect to reject even the true asset pricing model when tested using in-sample data. Our results suggest that asset pricing models should be tested using out-of-sample data or, if not not feasible, that the correct standard by which to judge a model is its ability to explain half of the in-sample alpha.

*Juhani Linnainmaa is with the University of Chicago Booth School of Business and NBER and Michael Roberts is with the Wharton School and NBER. We thank Ken French, Jon Lewellen, David McLean, and Jeff Pontiff for helpful discussions, and seminar participants at University of Lugano, University of Copenhagen, and University of Texas at Austin for valuable comments.

1 Introduction

Asset pricing research continues to uncover new anomalies at an impressive rate. Harvey, Liu, and Zhu (2015) document 314 factors identified by the literature, with the majority being identified during the last 15 years. McLean and Pontiff (2015) study the out-of-sample performance of 97 variables that previous research has identified as significant predictors of the cross section of stock returns. Cochrane (2011) summarizes the state of the literature by noting: "We thought 100% of the cross-sectional variation in expected returns came from the CAPM, now we think that's about zero and a zoo of new factors describes the cross section."

We examine cross-sectional anomalies in stock returns using hand-collected accounting data extending back to the start of the 20th century. Specifically, we investigate three potential explanations for these anomalies: unmodeled risk, mispricing, and data-snooping. Each of these explanations generate different testable implications across three eras encompassed by our data: (1) pre-sample data existing before the discovery of the anomaly, (2) in-sample data used to identify the anomaly, and (3) post-sample data accumulating after identification of the anomaly.

The anomalies on which we focus rely on accounting data, which, except for the value effect, have been largely unavailable prior to 1963 when the popular Compustat database becomes free of backfill bias.[1] We amass comprehensive accounting data from Moody's manuals from 1918 through the 1960s, and merge these data with the Compustat and CRSP records.[2] To our knowledge, the

[1] The 1963 date holds special significance only because Standard and Poor's created Compustat in 1962. Although Standard and Poor's collected historical data going back to 1947, they did so only for some of the surviving firms (Ball and Watts 1977).

[2] These same historical accounting data have previously been used in Graham, Leary, and Roberts (2014, 2015). This data collection project resembles that undertaken in Davis, Fama, and French (2000) except that, whereas Davis et al. (2000) collect information on the book value of equity, we collect the complete income statements and balance sheets. The initial Davis, Fama, and French (2000) study used data on industrial firms, but they subsequently extended the data collection efforts to cover both industrials and non-industrials. These data are provided by Ken French at `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library/det_historical_be_data.html`.

final database provides the most comprehensive look at returns and fundamentals from the start of the CRSP database in 1926 to today. Importantly, its coverage of publicly traded firms is similar before and after 1963, and our tests indicate that the quality of the pre-1963 data is comparable to that of the post-1963 data.

We first characterize the returns earned by the profitability and investment factors in the pre-1963 period. Our focus on these factors is motivated by Fama and French (2015) and Hou, Xue, and Zhang (2015) who show that these factors, in concert with the market and size factors, capture much of the cross-sectional variation in stock returns. We find no statistically reliable premiums on the profitability and investment factors in the pre-1963 sample period. The average returns on these factors are -11 (t-value $= -0.68$) and 10 (t-value $= 0.89$) basis points per month from July 1926 through June 1963. In the 1963 through 2014 data, each factor averages 26 basis points per month with t-values of 3.03 and 3.55. The absence of these premiums stands in contrast to the value effect, which is statistically significant also in the pre-1963 data (Fama and French 2006). We also show that returns on profitability and investment factors are close to zero *throughout* the pre-1963 period, suggesting that our findings are unlikely a result of realized returns falling short of expectations.[3]

The findings for investment and profitability premiums in the pre-1963 data are representative of most of the other 36 anomalies that we examine. Just seven out of the 38 earn average returns that are positive and statistically significantly at the 5% level in the pre-1963 period, and nine anomalies have CAPM alphas that are significant at this level. These results are not due to lack of power. In most cases, the historical out-of-sample period is 37 years long, and therefore typically longer than the original study's sample period. Additionally, the average alpha is significantly lower during the pre-discovery out-of-sample period than what it is during the original study's sample. Relative to

[3]The investment premium appears briefly before World War II only to disappear until 1970. The profitability premium does not emerge until 1980.

the original study's sample period, the average CAPM alpha is 61% lower between July 1926 and the beginning of the in-sample period.

At first glance, these findings are consistent with data-snooping as the anomalies are clearly sensitive to the choice of sample period. In contrast, if the anomalies are a consequence of multidimensional risk that is not accurately accounted for by the empirical model (i.e., unmodeled risk), then we would have expected them to be similar across periods, absent structural breaks in the risks that matter to investors. Similarly, if the anomalies are a consequence of mispricing, then we would have expected them to be larger during the pre-discovery sample period when limits to arbitrage, such as transaction costs (Hasbrouck 2009), were greater. Both of these implications are counterfactual.

Other features of data also point towards the data-snooping, whose bias works through t-values. An effect is deemed a return anomaly if its t-value is high. Because t-values are proportional to average excess returns scaled by volatilities, anomalies' in-sample returns are "too high" and their volatilities "too low" if data-snooping bias matters. Therefore, if anomalies are selected because of their t-values, average returns and volatilities should correlate positively—low return anomalies should be less volatile and vice versa.

Even without collecting out-of-sample data, an examination of the in-sample return processes of the 38 anomalies suggests that data-snooping might be an issue. The cross-sectional correlation between these anomalies' average returns and return standard deviations is 0.57 (p-value = 0.0002). Further, the average anomaly becomes less profitable and more volatile when we look either backward or forward in time relative to the original study's sample period. Although a risk-return tradeoff could account for the positive in-sample correlation, it is difficult to explain why both the average returns and volatilities change when moving outside the original sample period.

Our results do not suggest that all return anomalies are spurious. An equal-weighted portfolio of in-sample anomalies earns a CAPM alpha of 31 basis points per month (t-value = 11.16). A similar portfolio of pre-discovery anomalies earns an alpha of 12 basis points per month (t-value = 4.43). Investors, however, face the uncertainty of not knowing which anomalies are real and which are spurious, and so they need to treat them with caution.[4] Those who assume that the cross section is immutable may be disappointed. For example, using post-1963 data to construct the mean-variance efficient strategy from the market, size, value, investment, and profitability factors leads to *underperformance* relative to the market portfolio when applied to the pre-1963 sample.

Our findings suggest that asset pricing models should not be evaluated by their ability to explain anomalies' in-sample returns. If data-snooping works only through first moments, then a general rule of thumb by which to judge asset pricing models using only in-sample data is their ability to explain half of the alpha associated with the anomaly. Unfortunately, data-snooping can distort estimates of the return processes' higher moments. We find that the correlation structure of anomalies differs significantly between the in- and out-of-sample periods, suggesting that estimates of factor loadings may be biased as well. Indeed, recent research by McLean and Pontiff (2015) argues that learning by arbitrageurs from academic research leads to increased comovement. Interestingly, we find the same pattern when we move out-of-sample by moving *backward* in time. That is, before its discovery, an anomaly correlates more with other yet-to-be-discovered anomalies than it correlates with those anomalies that are already in-sample. Thus, because data-mining bias affects many facets of returns—averages, volatilities, and correlations—it is best to test asset pricing models out of sample.

[4]This problem is the analogous to that highlighted in the mutual fund literature. Kosowski, Timmermann, Wermers, and White (2006), Barras, Scaillet, and Wermers (2010), Fama and French (2010), and Linnainmaa (2013), for example, develop tests that adjust for the multiple-comparisons problem, and estimate that the fraction of actively managed mutual funds that can beat the market is small but non-zero. These tests, however, do not assist in identifying the skilled managers.

The rest of the paper is organized as follows. Section 2 describes our data sources and the coverage of publicly traded firms. Section 3 compares the returns earned by the profitability and investment factors between the modern (post-1963) and the pre-1963 sample period. Section 4 compares the average returns and CAPM alphas of 38 accounting-based anomalies between the original study's sample period and the pre- and post-discovery out-of-sample periods. Section 5 concludes.

2 Data

2.1 Data sources

We use data from four sources. First, we obtain monthly stock returns and shares-outstanding data from the Center for Research in Securities Prices (CRSP) database from January 1926 through December 2014. We exclude securities other than common stocks (share codes 10 and 11). CRSP includes all firms listed on the New York Stock Exchange (NYSE) since December 1925, all firms listed on the American Stock Exchange (AMEX) since 1962, and all firms listed on the NASDAQ since 1972. Although stocks also traded on regional exchanges before 1962, CRSP does not cover the other venues.[5] We take delisting returns from CRSP; if a delisting return is missing and the delisting is performance-related, we impute a return of −30% (Shumway 1997).

Second, we take annual accounting data from Standard and Poor's Compustat database. These data begin in 1947 for some firms, but become more comprehensive in 1962. Standard and Poor's established Compustat in 1962 to serve the needs of financial analysts, and backfilled information only for those firms that were deemed to be of the greatest interest to these analysts (Ball and Watts 1977).

[5]See, for example, Gompers and Lerner (2003). They note that firms that end up on the NYSE had often been trading as public companies on regional exchanges long before obtaining the NYSE listing.

Third, we add accounting data from Moody's Industrial and Railroad manuals. We collect information for all CRSP firms going back to 1918. These same data have previously been used in Graham, Leary, and Roberts (2014, 2015). .

Fourth, we add to our data the historical book value of equity data provided by Ken French. These are the data initially collected by Davis, Fama, and French (2000) for industrial firms, but later expanded to include non-industrial firms. We use the same definition of book value of equity as Ken French throughout this study.

In constructing our final database, we make the typical assumption that accounting data are available six months after the end of the fiscal year (Fama and French 1993). In most of our analyses, we construct factors using annual rebalancing. When we sort stocks into portfolios at the end of June in year t, we therefore use accounting information from the fiscal year that ended in year $t-1$.

2.2 Coverage

Table 1 shows the number of firms in the CRSP database at five years intervals from 1925 through 1965. There are 503 (NYSE) firms on CRSP at the very beginning. The number of CRSP firms increases over time, reaching 1,140 firms in 1960. The large jump to 2,251 firms in 1965 is due to the introduction of AMEX in 1962.

The second line shows the number of firms for which Compustat provides any accounting information. There is no information until 1947, and by 1950 the data are available for 324 of the 1,018 NYSE firms. By 1965, which is the date by which Compustat is survivorship-bias free, the accounting data are available for 1,301 of the 2,251 firms. The third line shows the number of firms for which we have accounting information either from Compustat or Moody's Industrial and Railroad manuals. The number of firms with accounting information starts at 354 in 1925 and increases over time as the number of firms listed on the NYSE expands. The Moody's manuals are an important source of information even after Compustat comes online. In 1950, Compustat has data for 324 firms, and the Moody's manuals have data for 476 additional firms. These manuals remain an important source even after 1962; in 1965, these manuals provide information for 308 additional firms. That is, although Compustat is free of a backfill bias as of 1963, it is not comprehensive. Figure 1 plots firm counts for CRSP, Compustat, and the combination of Compustat and Moody's from 1925 through 2013.[6] This figure illustrates that the final database that combines Compustat with Moody's manuals has similar coverage of CRSP firms both before and after 1963.

The lower part of Table 1 disaggregates data coverage by data item. This breakdown shows that the coverage of the Compustat data varies by data item. Accounts Payable, for example, is missing for almost all firms in the pre-1962 (backfill) period. This lack of coverage is, in part, due to the fact that not all firms reported this item in the 1960s and before. Even with the Moody's manuals, this item is missing for most firms. By contrast, almost all firms that provide any accounting information report revenue, net income, and total assets.

2.3 Data quality

Limitations in data quality could distort measurements of return anomalies. These anomalies could appear weaker or be absent if the historical data contain errors or if individual firms use different accounting standards. Because of the central importance of data quality, in this section we describe four considerations and tests that indicate that the quality of the pre-1963 data is comparable to that of the post-1963 data.

First, in terms of the accounting standards, the important historical date is the enactment of the

[6] We exclude year 2014 from this graph because, as of the time this study was undertaken, most firms' accounting information was not yet available for the fiscal year that ended in 2014.

Securities Exchange Act of 1934. The purpose of this act was to ensure the flow of accurate and systematic accounting information, and researchers typically consider the accounting information reliable after this date. Cohen, Polk, and Vuolteenaho (2003), for example, discuss the Securities Exchange Act in detail and, based on their analysis of the historical SEC enforcement records, use the post-1936 data on the book value of equity in their main tests. They characterize the first two years after the enactment of the act as an initial enforcement period, and drop these years from the sample. Although our data start in 1926 for many anomalies, we confirm that the results are both qualitatively and quantitatively the same in the post-1936 data.[7]

Second, we can compare the two parts of the sample by testing how closely the accounting data conform to clean-surplus accounting. Under clean-surplus accounting, the change in book value of equity equals earnings minus dividends (Ohlson 1995). Clean-surplus accounting is a central idea in accounting theory because it requires that the changes in assets and liabilities pass through the income statement. However, even under the generally accepted accounting principles (GAAP), some transactions can circumvent the income statement and affect the book value of equity directly,[8] and so real-world income statement and balance sheet information rarely line up exactly as they should under this ideal. A study of the extent to which firms conform to clean-surplus accounting is therefore a joint test of two issues that are relevant for the validity of the accounting information: (a) errors in Moody's manuals and (b) firms' tendencies to circumvent the income statement. We implement this test by comparing how closely *implied profitability*, computed using the clean-surplus formula, tracks the profitability that firms report on their income statements. Specifically, under

[7]Cohen, Polk, and Vuolteenaho (2003) also note that the pre-1936 data are not incongruent with the other data: "It is comforting, however, that our main regression results are robust to the choice between the 1928–1997 and 1938–1997 periods." Please note that the Cohen et al. (2003) timing convention is such that their year 1938 observations use book values from 1937.

[8]See endnote 1 in Ohlson (1995) for examples.

clean-surplus accounting, implied (log-)profitability equals

$$\text{implied profitability}_t = \log\{[(1+R_t)\text{ME}_{t-1} - \text{D}_t]\text{BE}_t/(\text{ME}_t\text{BE}_{t-1}) - [1 - \text{D}_t/\text{BE}_{t-1}]\}, \qquad (1)$$

where R_t is the total stock return over fiscal year t, ME_t and BE_t are the market and book values of equity at the end of fiscal year t, and D_t is the sum of dividends paid over fiscal year t.[9] This formula adjusts the change in the book value of equity for dividends, share repurchases, and share issuances to back out the implied earnings. The income-statement profitability is the net income reported for fiscal year t divided by the book value of equity at the end of fiscal year $t-1$.

We estimate annual panel regressions of implied log-profitability on log-return on equity using pre- and post-1963 data. We adjust standard errors by clustering by year. In the pre-1963 data, the slope on log-return on equity is 1.05 (SE = 0.05), and the adjusted R^2 is 33.6%. In the post-1963 data, the slope is 0.65 (SE = 0.02), and the adjusted R^2 is 41.6%. In cross-sectional regressions, which are equivalent to estimating weighted panel regressions with year fixed effects, the average slope estimate is 0.99 for the pre-1963 sample and 0.76 for the post-1963 sample. The comparable conformity to clean-surplus accounting suggests that the historical data are accurate, and that the typical firm does not circumvent the income statement to a significantly different degree in the pre-1963 data than in the post-1963 data.

Third, our results suggests that the quality of the data and the differences in accounting standards are probably not a major concern. To see why, note that we could place anomalies in an approximate order based on how sensitive they are to the quality of the accounting data. We believe that some anomalies, such as those based on the growth in total assets or sales, are more robust to noise in

[9] See, for example, Vuolteenaho (2002), Cohen, Polk, and Vuolteenaho (2003), and Nagel (2005).

data than others, such as those based on the book value of equity. Book value of equity is potentially problematic because it is the sum of retained earnings adjusted for dividends and net stock issues, and so it is affected by both data quality and variation in accounting standards. Nevertheless, the value premium (which is based on the book value of equity) is one of the anomalies that exists in the pre-1963 data (Fama and French 2006); in section 4, we show that the asset and sales growth anomalies, by contrast, are absent.

Fourth, our results also suggest more directly that the pre-1963 accounting data are of high quality and reflect differences in firm fundamentals. Specifically, the return anomalies we construct from these data are significantly more volatile than what they would be if the data were either noisy or irrelevant for describing firms' return processes. To see the connection, suppose that accounting variable X is unrelated to fundamentals either because the data are of poor quality or because firms follow different accounting standards. In this case, if we sort firms into portfolios by X, the average firms in the high and low portfolios will be similar in every dimension. With an infinite number of firms, the firms in these portfolios will be of the same size, have the same (true) market beta, and so forth. The two portfolios would therefore earn identical returns—because even idiosyncratic risk disappears as the number of firms grows—and therefore the volatility of the *high-minus-low strategy* would be zero. In a finite sample, this strategy's volatility is positive because, in finite samples, some change variation in fundamentals remains, and because the portfolios are not perfectly diversified. We can, however, test how volatile an actual anomaly is relative to its expected volatility under the null hypothesis that we construct the anomalies by sorting on noise.

In section 4, we construct HML-like factors for 38 anomalies. The average anomaly's return variance is 5.6 times that of a randomized factor in the pre-1963 data. We construct this randomized factor from the same set of stocks as the actual factor but, instead of creating the high and low portfolios based on the real firm characteristics, we assign stocks into the high and low portfolios at random. In the post-1963 data, this variance ratio is 6.6.[10] This "excess" volatility suggests that the historical accounting data measure differences in firm fundamentals to the same extent as they measure them in the post-1963 data.

3 Profitability and investment factors

We begin by measuring the pre-1963 performance of the profitability and investment factors. We focus on these factors because of their prominence in recent empirical asset pricing work. Both Fama and French (2015) and Hou, Xue, and Zhang (2015) propose modifying the three-factor model by adding profitability and investment factors, and by possibly eliminating the value factor. This section's detailed analysis of the profitability and investment factors sets the stage for Section 4 in which we analyze returns on a total of 38 anomalies.

3.1 Defining factors

Both Fama and French (2015) and Hou, Xue, and Zhang (2015) measure investment as the rate of change in the book value of total assets over the previous fiscal year. Using the Compustat variable names, this measure is defined as $\text{investment}_t = \text{at}_t/\text{at}_{t-1}$. This measure is also alternatively known as the asset-growth anomaly (Cooper, Gulen, and Schill 2008).[11] We follow Fama and French (2015) and construct HML-like profitability and investment factors by sorting stocks into six portfolios by

[10] The standard errors for the pre- and post-1963 ratios, computed by block bootstrapping the data by calendar month, are 0.57 and 0.26.

[11] We evaluate other investment-based anomalies in Section 4. Fama and French (2001, p. 16) motivate using the growth in total assets as a measure of investment as follows: "Some readers express a preference for capital expenditures (roughly the change in long-term assets), rather than the change in total assets, to measure investment. Our view is that short-term assets are investments. Just as they invest in machines, "firms invest in cash, accounts receivable, and inventory to facilitate their business activities. And when cash is retained for future long-term investments, the resources for these investments are committed when the cash is acquired."

size and profitability, or by size and investment. For example, to generate the investment factor, we construct the following six portfolios at the end of each June using NYSE breakpoints:

	Investment		
Size	**Low** (30%)	**Neutral** (40%)	**High** (30%)
Small (50%)	Small-Conservative	Small-Neutral	Small-Aggressive
Big (50%)	Big-Conservative	Big-Neutral	Big-Aggressive

We then hold these value-weighted portfolios from July of year t to the end of June of year $t+1$. The investment factor, called CMA for "conservative minus aggressive" in Fama and French (2015), is defined as:

$$\text{Investment factor} =$$

$$\frac{1}{2}\left(\text{Small-Conservative} + \text{Big-Conservative}\right) - \frac{1}{2}\left(\text{Small-Aggressive} + \text{Big-Aggressive}\right). \quad (2)$$

We follow Fama and French (2015) and measure profitability as operating profits over book value of equity. Using the Compustat variable names, this measure is defined as $\text{profitability}_t = (\text{revt}_t - \text{cogs}_t - \text{xsga}_t - \text{xint}_t)/\text{be}_t$. Because companies did not historically separately report sales, general, and administrative (SG&A) expenses from cost of goods sold, we set SG&A to zero when it is not reported. Similar to the construction of the investment factor, we sort stocks into six portfolios at the end of June of year t, and compute value-weighted returns on these portfolios from July of year t to the end of June of year $t+1$. The profitability factor, called RMW for "robust minus weak" in Fama and French (2015), is then defined as the average return on the two "robust profitability" portfolios minus the average return on the two "weak profitability" portfolios,

$$\text{Profitability factor} = \frac{1}{2}\,(\text{Small-Robust} + \text{Big-Robust}) - \frac{1}{2}\,(\text{Small-Weak} + \text{Big-Weak})\,. \qquad (3)$$

Finally, we define the size and value factors similar to Fama and French (1993), that is, by sorting stocks into six portfolios based on size and book-to-market, and defining the factors as the differences in the average returns of the small and big portfolios (size) or those of the high book-to-market and low book-to-market portfolios (value).

Table 2 compares our size, value, profitability, and investment factors to the corresponding Fama-French factors using the common sample period from July 1963 through December 2014. This is the modern sample period used in most asset pricing studies because of the availability of the Compustat data. In Panel A, we report average monthly percent returns for these factors as well as the t-values associated with these averages. The average returns on these factors are fairly close to each other. In our data, the estimated monthly premiums on the size, value, profitability, and investment factors are 23 basis points (size; t-value = 1.91), 36 basis points (value; t-value = 3.14), 25 basis points (profitability; t-value = 3.03), and 26 basis points (investment; t-value = 3.55). Panel B shows that the correlations between our factors and the Fama-French factors are high. Even the lowest correlation, which is the between the two investment factors, is 0.98. The reason for the small discrepancy between our numbers and those in Fama and French (2015) is that the Compustat-CRSP mapping used in Fama and French (2015) includes more firms than the standard mapping provided by CRSP.

3.2 Portfolio and factor returns

Table 3 Panel A reports average returns for the three sets of portfolios that are used to construct the size, value, profitability, and investment factors. We divide the sample period into two main parts: the "pre-1963" sample period runs from July 1926 through June 1963 and the "post-1963" sample period runs from July 1963 through December 2014. We further divide the pre-1963 sample into two subperiods of 18.5 years (222 months) each. The first half runs from July 1926 through December 1944 and the second from January 1945 through June 1963.

The estimates for the pre-1963 sample period differ significantly from those for the post-1963 sample period. Although the value premium is significant over the 1926–1963 period—the estimated monthly premium is 0.42% with a t-value of 2.04—the premiums associated with the size and investment factors are not. The average return on the size factor is 0.18% (t-value = 1.13), and that on the investment factor is 0.10% (t-value = 0.89). The profitability premium does not exist in this pre-1963 sample. Over the same July 1926–June 1963 period, the profitability factor earns an average return of −0.11% (t-value = −0.68).

The average returns on the portfolios that are used to construct the profitability and investment factors show that these insignificant estimates are not confined to either big or small stocks. Among both small and big stocks, the most profitable stocks on average earn lower returns than the least profitable stocks, although insignificantly so. The investment premium is positive among both small and big stocks, but these premiums are too small and the factor returns too noisy to push the premium on the investment factor into statistical significance.

Table 3 Panel B shows that the absence of profitability and investment premiums is unlikely due to any lack of statistical power. Consistent with Table 1 and Figure 1, the average number of stocks per portfolio is high even during the first half of the pre-1963 sample. Over the entire pre-1963 sample, the average number of stocks per portfolio is always greater than 50. This amount of diversification, combined with the length of the sample period (37 years) gives us confidence that we should be able to identify return premiums when they exist. Moreover, the absence of the profitability and investment premiums is unlikely to be due to the fact that they are computed through complicated accounting-based measures. The investment premium, for example, is based on the growth in total assets. Although there may be noise in this measure, the amount of such noise is probably less than that in book value of equity, yet value premium is statistically significant in the pre-1963 data.

3.3 Subsample analysis

The first two columns in Table 3 show that the first half of the pre-1963 sample (from July 1926 through December 1944) is not meaningfully different from the second half (from January 1945 through June 1963). The negative overall return on the profitability factor is due to its poor performance during the first half. It earns a small positive premium during the second half, but this premium is not statistically significant at conventional levels. The investment factor, on the other hand, earns a positive premium only during the first half of the pre-1963 sample. But, similar to the profitability factor, the estimated premium is statistical significant neither during the first nor the second half.

These subsample estimates suggest that investors towards the beginning of the century were unlikely to require a premium for investing in profitable companies that follow conservative investment policies. It could, however, be that investors required such premiums, but that the realized factor returns were negative or statistically insignificant because some of the risks associated with such stocks were realized during the July 1926 through June 1963 sample period. If so, we would expect to find statistically significant premiums when we divide the sample—at least if the negative return shocks were more prevalent during the first or the second half of the pre-1963 sample. The estimates in Table 3 do not seem to support this explanation.

Figure 2 reports average returns for the same factors using rolling ten-year windows. The first point in each panel, for example, corresponds to June 1936, and it reports the average monthly percent return from July 1926 through June 1936. The dotted lines in these graphs indicate 95% confidence intervals. The time-series behaviors of the size and value premiums differ significantly from those of the profitability and investment premiums. The size premium is positive except for two (extended) periods of time, one in the 1950s and the other from 1980s to 1990s. The value premium is positive almost throughout the full sample period except for a sharp (but brief) interruption towards the end of the 1990s during the Nasdaq episode.

The profitability premium, on the other hand, is thoroughly absent from the historical data. The rolling-average estimate of the profitability premium is negative until 1950, after which point it remains close to zero all the way to the 1980s. That is, the profitability premium emerges in full only sometime in the early 1980s. The investment premium behaves similarly. The average investment premium is positive until 1945—but the return realizations relatively noisy—after which point it remains close to zero until the 1970s. That is, the investment premium emerges sometime in the early 1970s.

3.4 An investment perspective

The pre-1963 sample looks very different from the post-1963 data when it comes to the profitability and investment premiums. Figure 3 illustrates this dissimilarity by reporting annualized Sharpe ratios for the market portfolio and an optimal strategy that trades the market, size, value, profitability, and investment factors. We construct the mean-variance efficient strategy using the modern sample period that runs from July 1963 through December 2014. We report the Sharpe ratios for rolling ten-year windows.

The market's Sharpe ratio for the entire 1926 through 2014 period is 0.42. It is slightly higher (0.46) for the pre-1963 sample than for the modern, post-1963 sample (0.39). The Sharpe ratio for the optimal strategy for the modern sample period is 1.16; by construction, this strategy is in-sample for this period. However, for the pre-1963 sample, the Sharpe ratio of this strategy is just 0.40, that is, lower than that of the market. Figure 3 shows that the optimal strategy rarely dominates holding the market portfolio in the pre-Compustat period; at the same time, the optimal strategy performs very poorly relative to the market portfolio in particular in the 1950s.

This computation illustrates that our view of what matters in the cross section of stocks greatly depends on where we look. An assumption that the cross section is immutable is poor at least when it comes to the profitability and investment factors. Figure 3 shows that the strategy that is (ex-post) optimal in the post-1963 data is wholly unremarkable in the pre-1963 data. Moreover, this computation suggests that investors could not have known in real-time in June 1963—at least on the basis of any historical return data—that this particular combination of size, value, profitability, and investment factors would perform so well relative to the market over the next 50 years.

4 Assessing the pre-discovery and post-discovery performance of 38 anomalies

4.1 Competing explanations for cross-sectional return anomalies

In this section, we use data on 38 anomalies to investigate the extent to which they are driven by three potential mechanisms—unmodeled risk, mispricing, and data-snooping. The first mechanism, unmodeled risk, asserts that cross-sectional return anomalies come about because stock risks are multidimensional. If the Sharpe (1964)-Lintner (1965) capital asset pricing model is not the true data-generating model, an anomaly might represent a deviation from the CAPM. The most prominent examples of this argument are the value and size effects. Fama and French (1996) suggest that the value effect is a proxy for relative distress and that the size effect is about covariation in small stock returns that, while not captured by the market returns, is compensated in average returns. The same argument can be made for any return anomaly. The joint hypothesis problem states that it may be our imperfect model that misprices assets and not the investors. Under the risk explanation, we expect the in-sample period to resemble the out-of-sample period, assuming that there are no structural breaks in the risks that matter to investors.

The second mechanism, mispricing, asserts that investor irrationality combined with limits to arbitrage causes asset prices to deviate from fundamentals. Lakonishok, Shleifer, and Vishny (1994), for example, suggest that value strategies are not fundamentally riskier, but that the value effect emerges because the typical investor's irrational behavior induces mispricing. The joint hypothesis problem applies here as well. Under the mispricing explanation, we expect the anomalies to grow stronger as we move backward in time. The reason is that trading costs were almost twice as high in the 1920s than in the 1960s (Hasbrouck 2009, Figure 3), and so would-be arbitrageurs would have had less power to attack mispricing.[12]

The third mechanism, data snooping, suggests that some, if not all, return anomalies are spurious. If researchers try enough many trading strategies, some of these experiments produce impressive t-statistics, even though the anomaly is entirely sample-specific. If an initial study exhausts all available data, it is difficult to address data-mining concerns except by waiting for additional data to accumulate.[13] The data-snooping explanation suggests that the in-sample period is different from the periods that predate and follow the original study's sample period.

4.2 Defining anomalies

The cross section of stock returns is full of other anomalies that are largely independent of the size, value, profitability, and investment factors. Are most anomalies similar to the profitability and investment factors in that they are largely absent from the historical data, or do some anomalies persist throughout the entire sample period? In this section, we compare how the returns on 38 accounting-based anomalies differ between the in-sample period (used in each original study) and out-of-sample periods that either predate or follow the in-sample period.

The benefit of analyzing a large number of anomalies is the increase in statistical power. Consider, for example, the investment and profitability premiums of Section 3. Although we cannot reject the null hypothesis that these premiums are zero in the pre-1963 period, we also cannot reject the null hypothesis that these premiums differ between the pre-1963 and post-1963 periods. The premiums are too noisy for us to reject this null.

[12] See also French (2008).

[13] Researchers can also turn to other markets or asset classes for additional evidence (see, for example, Fama and French (1998) and Asness, Moskowitz, and Pedersen (2013)) or, in some cases, examine securities excluded from the initial study (see, for example, Barber and Lyon (1997) and Ang, Shtauber, and Tetlock (2013)). Jegadeesh and Titman (2001) is a prime example of a paper that analyzed data that had accumulated after the initial study; in this case, the original momentum study of Jegadeesh and Titman (1993).

Table 4 lists the additional anomalies that we study along with references to the original studies and the original sample periods. The starting point for our list is McLean and Pontiff (2015). We add to their list a few anomalies that have been documented after that study. We describe each anomaly in detail in the Appendix. All these anomalies use accounting information and, therefore, with the exceptions of book-to-market and net share issuances, have not been extended to the pre-Compustat sample. We group similar anomalies into seven categories: profitability, earnings quality, valuation, investment and growth, financing, distress, and composite anomalies. In our classification, composite anomalies, such as Piotroski's (2000) F-score, are anomalies that combine multiple anomalies into one. We do not examine "return-based" anomalies such as momentum, idiosyncratic volatility, and low beta, because many of these anomalies have already been taken to the pre-1963 period, often already in the original study. Moreover, many return-base anomalies rely on high turnover (Novy-Marx and Velikov 2015); accounting-based anomalies, by contrast, rebalance annually, and so the inferences about their profitability are not influenced as much our estimates of their trading costs.

We use the same definitions for all 38 anomalies—that is, value, profitability, investment, and the 35 additional anomalies—throughout the sample period. For example, even though we could start using reported capital expenditures (CAPX) from Compustat to construct some of the growth and investment anomalies, we always approximate these expenditures by the annual change in the plant, property, and equipment plus depreciation. By using constant definitions, we ensure that the estimates are comparable over the entire 1926–2014 sample period.

We construct similar HML-like factors for each of the additional anomalies as what we constructed for the profitability and investment anomalies in Section 3. That is, we sort stocks into six portfolios at the end of June of year t by size and each anomaly variable, and then compute value-weighted returns on these portfolios from July of year t to June of year $t+1$. The exceptions are the debt and net issuance anomalies. The debt issuance anomaly takes short positions in firms that issue debt and long positions in all other firms. The net issuance anomalies take short positions in firms that issue equity and long positions in firms that repurchase equity. We compute the return on each anomaly as the average return of the "high" portfolios minus the average return of the "low" portfolios. We reverse the high and low labels for those anomalies for which the original study indicates that the average returns of the low portfolios exceeds that of the high portfolio.

4.3 In-sample estimates

Table 5 reports the average monthly percent returns and CAPM alphas and betas for the 38 anomalies. We estimate each measure using the same sample period as that used in the original study. The returns on the value, profitability, and investment factors are reported on rows labeled book-to-market (value), operating profitability (profitability), and asset growth (investment).

Out of 38 anomalies, 33 earn average returns that are positive and statistically significant at the 10% level. In the CAPM, the number of positive and statistically significant anomalies increases to 35 because most anomalies—30 out of 38—correlate negatively with the market, sometimes significantly so. Consider, for example, the distress anomaly of Campbell, Hilscher, and Szilagyi (2008). The average return on this distress factor is 30 basis points per month (t-value = 2.4). However, because this anomaly's CAPM beta is -0.3, its CAPM alpha is considerably higher, 44 basis points per month (t-value = 4.03).

Some of the most impressive t-values belong to the composite anomalies. Piotroski's F-score, which is a combination of 9 signals of firm quality, generates a factor that earns a CAPM alpha of 52 basis points per month, which is statistically significant with a t-value of 6.44. Among the best non-composite anomalies are change in asset turnover, net operating assets, debt issuance, and distress risk.

4.4 Historical out-of-sample estimates

Table 6 reports average monthly percent returns and CAPM alphas and betas for the same 38 anomalies using data that predates the sample periods used in the original studies. In Panel A, we use return data up to one month prior to the beginning of the original study's sample period; in Panel B, we stop the historical sample either in June 1963 or one month prior to the beginning of the original study's sample, whichever is earlier.

The anomalies are significantly weaker for the historical out-of-sample period. In Panel A, 9 anomalies earn average returns that are positive and statistically significantly different from zero at the 10% level. Evaluated by their CAPM alphas, 12 anomalies are statistically significant. Put differently, just one third of the anomalies that earn statistically significant alphas during the original sample periods do so in the pre-discovery sample. In Panel B, the number of statistically significant anomalies is 7 (average returns) and 8 (CAPM alphas).

The lack of significance is probably not due to a lack of power. In many cases, Panel B's historical sample period is 37 years long—and therefore often longer than that used in the original study. The total number of monthly in-sample observations for the 38 anomalies is 12,773; but the number of monthly pre-sample observations is 19,831. Among the most significant anomalies in Panel B are net working capital changes, growth in inventory, and investment-to-assets ratio. Among the composite anomalies, both Mohanram's G-score and the profitability component of the quality-minus-junk factor (Asness, Frazzini, and Pedersen 2013) perform well.

One noteworthy anomaly is that related to net share issues. Both the one- and five-year versions of this anomaly are statistically significant at the 5% level with t-values of 2.95 and 2.16. The significance of the net issuance anomaly over the modern, post-1963 sample period has been highlighted, for example, in Daniel and Titman (2006), Fama and French (2008), and Pontiff and Woodgate (2008). The last two of these studies, however, find no reliable evidence of a net issuances anomaly in the pre-1963 data. These studies could investigate the pre-Compustat era because also CRSP provides shares-outstanding information for the pre-1963 period. In contrast to these null results, the estimates in Table 6 suggest that the net share issues anomaly exists also in the pre-Compustat period. The reason our estimates differ from those in the earlier studies appears to lie with the corrections to the number of shares data CRSP made in a project started in 2013.[14]

4.5 Post-discovery out-of-sample estimates

Table 6 reports the average monthly percent returns and CAPM alphas and betas for 36 out of the 38 anomalies using data that have accumulated after the original study's sample period. We exclude both the operating profitability and quality-minus-junk: profitability anomalies because they have no more than two years of post-discovery data.

The estimates for the post-discovery period are similar to those for the pre-discovery out-of-sample period. Of the 36 anomalies, 5 earn average returns that are statistically significant at the 10% level, and 14 anomalies earn statistically significant CAPM alphas. However, many of the anomalies that are statistically significant in these data are not the same that are significant in the pre-discovery period. A noteworthy exception is Mohanram's G-score, which earns a monthly CAPM alpha of 44 basis points (t-value = 3.42) after its discovery. Besides this anomaly, only gross profitability and five-year share issuance anomalies are statistically significant both in the pre- and post-discovery

[14] See `http://crsp.com/files/images/release_notes/mdaz_201306.pdf` and `http://crsp.com/files/images/release_notes/mdaz_201402.pdf`. Ken French also highlights the repercussions of these changes at `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`: "The file [CRSP] released in January 2015…incorporates over 4000 changes that affect 400 Permnos. As a result, many of the returns we report for 1925–1946 change in our January 2015 update and some of the changes are large."

sample periods.

4.6 Comparing the performance of accounting-based anomalies: Historical out-of-sample period, original sample, and post-discovery sample

In Table 8 we measure the average returns, volatilities, CAPM alphas, and Sharpe ratios of 32 anomalies using the original study's sample period (original sample), between July 1926 and the beginning of the in-sample period (pre-discovery sample), and between the end of the in-sample period and December 2014 (post-discovery sample). We exclude the composite anomalies and the operating profitability and quality-minus-junk: profitability from this analysis.

We compute standard errors by block bootstrapping monthly anomaly returns by calendar month to preserve the covariance structure of returns. The first row reports the average monthly returns for the anomalies during the three sample periods, and the differences between the in-sample period and the pre- and post-discovery periods. The estimates show, for example, that the average anomaly earns 28 basis points (t-value = 9.15) per month during the sample period used in the original study, but just 9 basis points (t-value = 2.73) during the historical out-of-sample period and 11 basis points (t-value = 3.6) after the end of the original sample. The differences in average returns between the original period and pre- and post-discovery periods are significant with t-values of -4.64 and -3.95. These results do not materially change when we adjust for market risk. The corresponding differences in CAPM alphas are -19 basis points (t-value = -4.95) and -15 basis points (t-value = -3.80).

In addition to the difference in average returns, the volatilities of the anomalies are also different. The annualized volatility of the average anomaly is 9.4% during the pre-discovery period; 6.4% during the original study's in-sample period; and 6.9% during the post-discovery period. The differences between the in-sample period and the pre- and post-discovery samples are statistically significant with t-values of 4.24 and 2.19. This increase in volatility is consistent with data-snooping bias contaminating the distribution of in-sample returns. Because data-mining works through t-values, both high average return and low volatility make it more likely that a particular factor is deemed as a return anomaly.

This volatility mechanism shows up also in the anomalies' in-sample returns as a positive correlation between the average return and volatility. In a cross-sectional regression of the average monthly return against the standard deviation of month returns, the slope coefficient is 0.12 with a t-value of 4.1. We estimate this regression using just 38 data points, where each data point consists of the in-sample average returns and standard deviations. This positive correlation between average returns and volatility may emerge because of an adding-up constraint: the sample consists of anomalies that (mostly) have high t-value. Therefore, if an anomaly has a relatively low average return, it must also have a lower volatility to qualify as an anomaly. Although the positive correlation could also emerge naturally from a risk-return tradeoff—riskier anomalies earn high average returns as a compensation for risk—this explanation would not explain why (most) anomalies are absent from the out-of-sample data, and why volatilities increase when we move out-of-sample by going either backwards or forwards in time.

The changes in average returns and CAPM alphas do not provide an accurate measure of the change in the attractiveness of the average anomaly because their risks change over time. Even if the average returns or alphas were not different between the in- and out-of-sample periods, the anomalies would be less attractive out-of-sample because of their higher volatilities.

In Table 8, we address this issue by estimating changes in annualized Sharpe ratios and information ratios. In the excess returns specification, Sharpe ratio is the annualized return over the annualized standard deviation. In the CAPM specification, we compute the information ratio by multiplying the t-value associated with the monthly CAPM alpha by $\sqrt{12}/\sqrt{T}$. These information ratios are *marginal* Sharpe ratios that measure the instantaneous rate of change in the Sharpe ratio of an investor who is currently holding the market. These estimates confirm that, if anything, the average anomaly would have been even less attractive before the beginning of the sample used in the original study.

The average anomaly's information ratio is 0.63 during the original study's sample period, but just 0.18 and 0.26 during the pre- and post-discovery out-of-sample periods. These differences, which are statistically significant with t-values of -6.52 and -4.19, correspond to 71% and 59% decreases in the information ratios depending on whether we move out-of-sample by going backwards or forwards in time. These estimates suggest that the risk adjustment works the "wrong way:" not only do anomalies earn high alphas during the sample period used in the original study, but the factors diversify out more risk than what they do outside this period.

The estimates in Table 8 support the data-mining explanation for the return anomalies. Under this explanation, the in-sample period stands out as being special. The estimates in the last column show that, except for the volatility of the average factor, the differences between the post- and pre-periods are at most one standard error away from each other. Put differently, in the data, the average returns and Sharpe ratios jump up as we enter the in-sample period and then, after this period ends, these quantities fall back to levels that are statistically indistinguishable from those seen during the pre-sample period.

4.7 Changes in the correlation structure of returns

The average anomaly's Sharpe ratio is lower in the out-of-sample periods that surround the original study's sample period because the average returns are lower and volatilities higher. These differences suggest that some anomalies are sample-specific or, interpreted from an econometrics viewpoint, that the in-sample returns do not constitute a representative sample. In this section, we measure the extent to which data-mining also alters the correlation structure of returns.

The first column of Table 9 shows estimates from a panel regression that explains anomaly returns using the average returns on all other anomalies,

$$\begin{aligned}\text{anomaly}_{i,t} &= a + b_1 * \text{post}_{i,t} + b_2 * \text{in-sample index}_{-i,t} + b_3 * \text{post-sample index}_{-i,t} \\ &\quad + b_4 * \text{post}_{i,t} * \text{in-sample index}_{-i,t} + b_5 * \text{post}_{i,t} * \text{post-sample index}_{-i,t} + e_{i,t}, \quad (4)\end{aligned}$$

where $\text{post}_{i,t}$ takes the value of one if anomaly i is in the post-discovery sample in month t, in-sample $\text{index}_{-i,t}$ is the average return over all other anomalies (that is, except anomaly i) that are in-sample in month t, and post-sample $\text{index}_{-i,t}$ is the average return over all other anomalies that are in the post-discovery sample in month t. We cluster standard errors by calendar month to account for the correlated errors in the cross sections. The specification (4) is similar to that estimated in McLean and Pontiff (2015) except that they (1) use a different set of anomalies and (2) define they cutoff based on the publication date.

The estimates in Table 9 suggest that anomalies correlate more with other already discovered anomalies after their discovery. When an anomaly is in-sample, its slope coefficient against the in-sample index is 0.482, but that against the post-sample index is just 0.092. However, when the anomaly most out-of-sample, these correlations change to $0.482 + (-0.373) = 0.109$ and $0.092 +$

$0.575 = 0.668$. These changes are associated with t-values of -7.6 and 21.9.

These estimates are consistent with arbitrageurs inducing comovement among already discovered anomalies. If entities such as hedge funds trade many anomalies at the same time, their capital in- and out-flows may translate into correlated price pressures in the anomalies' long and short legs. This regression therefore suggests that the post-discovery decrease in average returns may, in part, be due to sophisticated money flowing into the anomalies, thereby correcting the mispricing that these anomalies represented, but also by increasing their correlations.

The estimates in the second column of Table 9 suggest an alternative explanation, which is that data-mining may also affect the correlation structure of returns. This regression is the same as that in the first column except that we swap the post-discovery out-of-sample period for the the pre-discovery out-of-sample period. The estimates in this column are similar to those in the first column, although the sequence of events is reversed. When an anomaly is in-sample, its correlation with other in-sample anomalies is 0.640, and that with the pre-discovery out-of-sample anomalies is 0.077. These correlations, however, change to 0.018 and 0.426 when the anomaly "moves" to the pre-discovery out-of-sample period. The differences between the in-sample and pre-discovery periods are statistically significant with t-values of -15.4 and 8.1.

The similarity between the pre-discovery and post-discovery estimates suggest that the higher post-discovery correlations are not necessarily driven by arbitrageurs. Rather, the correlations may increase because data-mining affects the entire return process, including the correlations with other anomalies. One explanation is that a new anomaly is more likely to be published if its t-value is high *and* if it is not subsumed by other known anomalies. Anomalies may therefore also be selected, in part, based on how they correlate with other anomalies and factors.

Although the increase in correlations appears to be another consequence of the data-mining process, our results do not imply that arbitrageurs have no influence over the return processes. In particular, McLean and Pontiff (2015) show that the average returns decrease not only after the end of the in-sample period but also later after the anomaly is published in an academic study. The second date holds no special significance under the data-mining explanation, but it is important under the mispricing-plus-arbitrageurs explanation.

5 Discussion and conclusions

The investment and profitability premiums are largely absent from the cross section of stock returns before 1963, and so are most of the 38 accounting-based anomalies that we examine. At most 12 of these anomalies earn positive and statistically significant returns also in the out-of-sample periods that surround the samples used in each of the original studies. The typically anomaly therefore appears to be spurious. The similarity between the pre-discovery and post-discovery out-of-sample periods suggests that data-snooping bias has had a significant impact on the discovery of return anomalies.

We estimate that the average anomaly's CAPM alpha is 61% lower in the sample that predates that used in the original study. Moreover, data snooping appears to distort return processes beyond the sample average. Anomalies are less volatile during the in-sample period, and their correlations with other anomalies are atypical. An anomaly's correlation with other out-of-sample anomalies increases when we move to either the post-discovery or *or* pre-discovery out-of-sample period.

These findings are important. Our results suggest that many anomalies do not require explanations that invoke multidimensional risks or mispricing. Under the risk-based explanation, we would not expect each study's original sample period to stand out in the way it does in the data. And under the behavioral explanation, we would expect the average anomaly to become more profitable as we move backward in time because of the higher trading costs. In the data, however, both the average returns and CAPM alphas are lower in the pre-discovery sample.

Our results may be good news for asset pricing models. Instead of having to explain away an average annual CAPM alpha of 3.7%, our results imply that the correct standard by which to evaluate these models is by assessing their ability to explain half of this alpha. However, because data snooping affects all facets of return processes—averages, volatilities, and correlations with other anomalies and factors—it will be difficult to correct test statistics even approximately for the effects of data-snooping bias. A preferred approach, when feasible, would be to test asset pricing models using out-of-sample data such as ours.

Future research can benefit from the new historical sample to gain additional insights into asset prices. Two questions permeate most of the empirical asset pricing literature. The first relates to identifying a parsimonious empirical asset pricing model that provides a passable description of the cross section of average returns; the second is about delineating between the risk-based and behavioral explanations for the many anomalies. Both lines of research can greatly benefit from the power afforded by additional 37 years of pristine data.

A Anomalies

In this appendix, we define the anomalies examined in Section 4. When applicable, we state the formulas using the Compustat item names. For those anomalies that are computed through a process involving multiple steps, we refer to the studies that describe the implementation in detail. We also indicate the first study that used each variable to explain the cross section of stock returns, and the sample period used in that study. When applicable, we use McLean and Pontiff (2015) to identify the first study. We state both the year and month when the months are provided in the original study; if not, we state the year, and assume that the sample begins in January and ends in December. The sample period refers to the sample in which the study uses the anomaly variable to predict returns.

A.1 Profitability

1. **Gross profitability** is defined as the revenue minus cost of goods sold, all divided by total assets: gross profitability$_t$ = $(\text{revt}_t - \text{cogs}_t)/\text{at}_t$. Novy-Marx (2013) examines the predictive power of gross profitability using return data from July 1963 through December 2010.

2. **Operating profitability** is defined as the revenue minus cost of goods sold, SG&A, and interest, all divided by book value of equity: operating profitability$_t$ = $(\text{revt}_t - \text{cogs}_t - \text{xsga}_t - \text{xint}_t)/\text{be}_t$. Fama and French (2015) construct a profitability factor based on operating profitability using return data from July 1963 through December 2013.

3. **Return on assets** is defined as the earnings before extraordinary items, divided by total assets: return on assets$_t$ = ib_t/at_t. Haugen and Baker (1996) use return on assets to predict returns between 1979 and 1993.

4. **Return on equity** is defined as the earnings before extraordinary items, divided by the book

value of equity: return on $\text{equity}_t = \text{ib}_t/\text{be}_t$. Haugen and Baker (1996) use return on equity to predict returns between 1979 and 1993.

5. **Profit margin** is defined as the earnings before interest and taxes, divided by sales: profit $\text{margin}_t = \text{oiadp}_t/\text{revt}_t$. Soliman (2008) uses profit margin to predict returns using return data from 1984 to 2002.

6. **Revenue surprise** is defined as year-to-year growth in revenue-per-share, divided by the standard deviation of this growth over the previous four years: revenue $\text{surprise}_t = \Delta(\text{revt}_t/\text{shares}_t)/\text{SD}(\Delta(\text{revt}_{t-j}/\text{shares}_{t-j}))$, where shares_t is the number of shares outstanding taken from CRSP ($\text{shrout}_t/1000$), if available, or from Compustat (csho_t), and $\text{SD}(\Delta(\text{revt}_{t-j}/\text{shares}_{t-j}))$ is the standard deviation over lags $1 \geq j \geq 4$. This measure is an adapted version of the one that Jegadeesh and Livnat (2006) use to predict returns between 1987 and 2003. Jegadeesh and Livnat (2006) define their measure using quarterly data; our definition adapts their definition to annual data.

7. **Change in asset turnover** is defined as the annual change in asset turnover, where asset turnover is revenue divided by total assets: change in asset $\text{turnover}_t = \Delta(\text{revt}_t/\text{at}_t)$. Soliman (2008) uses the change in asset turnover to predict returns between 1984 and 2002.

A.2 Earnings quality

8. **Accruals** is the non-cash component of earnings divided by the average total assets: $\text{accruals}_t = (\Delta\text{act}_t - \Delta\text{che}_t - \Delta\text{lct}_t - \Delta\text{dlc}_t - \Delta\text{txp}_t - \text{dp}_t)/((\text{at}_{t-1} + \text{at}_t)/2)$, where Δ denotes the change from fiscal year $t-1$ to t. Sloan (1996) uses data from 1962 to 1991 to examine the predictive power of accruals.

9. **Earnings consistency** is the geometric average changes in (scaled) earnings to price over the previous five years; see Alwathainani (2009) for the formulas and sample selection rules. Alwathainani (2009) uses earnings consistency to predict stock returns using data from January 1971 through December 2007.

10. **Net operating assets** represent the cumulative difference between operating income and free cash flow, scaled by lagged total assets, net operating assets$_t$ = $[(\mathrm{at}_t - \mathrm{che}_t) - (\mathrm{at}_t - \mathrm{dlc}_t - \mathrm{dltt}_t - \mathrm{be}_t)]/\mathrm{at}_{t-1}$. Hirshleifer, Hou, Teoh, and Zhang (2004) form trading strategies based on net operating assets using data from July 1964 through December 2002.

11. **Net working capital changes** is another measure of accruals: net working capital changes$_t$ = $[\Delta(\mathrm{act}_t - \mathrm{che}_t) - \Delta(\mathrm{lct}_t - \mathrm{dlc}_t)]/\mathrm{at}_t$. Soliman (2008) uses net working capital changes to predict stock returns using return data from 1984 to 2002.

12. **Taxes-to-income ratio** is defined as taxes payable divided by net income, taxes-to-income ratio$_t$ = $\mathrm{txp}_t/\mathrm{ni}_t$. Lev and Nissim (2004) measure the predictive power of the taxes-to-income ratio using data from 1973 through 2000. Lev and Nissim (2004) suggest that the gap between book earnings and taxable income may be informative about the quality of earnings.

A.3 Valuation

13. **Book-to-market ratio** is defined as the book value of equity divided by the December market value of equity: book-to-market ratio$_t$ = $\mathrm{be}_t/\mathrm{mv}_t$. Fama and French (1992) use book-to-market ratio to predict returns using return data from July 1963 through December 1990.[15]

[15] The book value of equity is computed as follows. First, we set the book value of equity equal to stockholders' equity (SEQ) if this data item exists. This is also the data item collected by Davis, Fama, and French (2000) for the pre-1963 data. Second, if SEQ is missing but both common equity (CEQ) and the par value of preferred stock (PSTK) exist, then we set the book value of equity equal to PSTK + CEQ. Third, if the above definitions cannot be used, but

14. **Cash flow-to-price ratio** is defined as the income before extraordinary items plus depreciation, all scaled by the December market value of equity: cash flow-to-price ratio$_t$ = $(\text{ib}_t + \text{dp}_t)/\text{mv}_t$. Lakonishok, Shleifer, and Vishny (1994) use cash flow-to-price ratio in tests that use return data from May 1968 through April 1990.

15. **Earnings-to-price ratio** is defined as the income before extraordinary items divided by the December market value of equity: earnings-to-price ratio$_t$ = ib_t/mv_t. Basu (1977) measures the predictive power of earnings-to-price ratio using data from April 1957 through March 1971.

16. **Enterprise multiple** is a value measure used by practitioners: enterprise multiple$_t$ = $(\text{mv}_t + \text{dlc}_t + \text{dltt}_t + \text{pstkrv}_t - \text{che}_t)/\text{oibdp}_t$, where mv_t is the end-of-June (that is, portfolio formation date) market value of equity. Loughran and Wellman (2011) compare the predictive power of enterprise multiple to that of book-to-market using return data from July 1963 through December 2009.

17. **Sales-to-price ratio** is defined as total sales divided by December market value of equity: sales-to-price ratio$_t$ = $\text{revt}_t/\text{mv}_t$. Barbee, Mukherji, and Raines (1996) compare the predictive power of sales-to-price to those of book-to-market and debt-to-equity ratio using return data from 1979 through 1991.

A.4 Growth and investment

18. **Asset growth** is defined as the percentage change in total assets, asset growth$_t$ = $\text{at}_t/\text{at}_{t-1} - 1$. Cooper, Gulen, and Schill (2008) examine the predictive power of asset growth using return

the book values of total assets (AT) and total liabilities (LT) exist, then we set the book value of equity equal to AT − LT. If the book value of equity is now non-missing, we adjust it by subtracting the redemption, liquidation, or par value of preferred stock—in that order, depending on data availability. Lastly, we add deferred taxes (TXDITC) and subtract postretirement benefits (PRBA) when these items exist.

data from July 1968 to June 2003.

19. **Growth in inventory** is defined as the change in inventory divided by the average total assets, growth in inventory$_t$ = Δinvt$_t$/[(at$_t$ + at$_{t-1}$)/2]. Thomas and Zhang (2002) use growth in inventory to predict stock returns using return data from 1970 to 1997.

20. **Sales growth** is constructed by ranking firms each year by sales rank and by computing the weighted average sales growth rank over the previous five years: sales growth$_t$ = $5 *$ rank$_t$ $+ 4 *$ rank$_{t-1}$ $+ 3 *$ rank$_{t-2}$ $+ 2 *$ rank$_{t-3}$ $+ 1 *$ rank$_{t-4}$. The ranks are computed using data on firms with six years of sales data. Lakonishok, Shleifer, and Vishny (1994) measure the predictive power of sales growth using return data from May 1968 through April 1990.

21. **Sustainable growth** is defined as the percentage change in the book value of equity, sustainable growth$_t$ = (be$_t$ − be$_{t-1}$)/be$_{t-1}$. Lockwood and Prombutr (2010) use return data from July 1964 through June 2007 to measure the predictive power of sustainable growth.

22. **Adjusted CAPX growth** is the industry-adjusted increase in capital expenditures over its average value over the previous two years, all scaled by the average value over the previous two years. First, unadjusted CAPEX growth is computed as CAPEX growth$_t$ = (capx$_t$ − (capx$_{t-1}$ + capx$_{t-2}$)/2)/[(capx$_{t-1}$ + capx$_{t-2}$)/2]. The industry adjustment subtracts the average CAPEX growth of the firms with the same two-digit SIC code. Abarbanell and Bushee (1998) use industry-adjusted CAPEX growth as a return predictor using return data from 1974 through 1993.

23. **Growth in sales minus inventory** is the difference between sales growth and inventory growth. Sales growth is the increase in sales over its average value over the previous two

years, all scaled by the average value over the previous two years; inventory growth is the increase in inventory over its average value over the previous two years, all scaled by the average value over the previous two years; growth in sales minus inventory$_t$ = $[\text{revt}_t - (\text{revt}_{t-1} + \text{revt}_{t-2})/2]/[(\text{revt}_{t-1} + \text{revt}_{t-2})/2] - [\text{invt}_t - (\text{invt}_{t-1} + \text{invt}_{t-2})/2]/[(\text{invt}_{t-1} + \text{invt}_{t-2})/2]$. Abarbanell and Bushee (1998) use growth in sales minus inventory to predict returns from 1974 through 1993.

24. **Investment growth rate** is the percentage change in capital expenditures, investment growth rate$_t$ = $\text{capx}_t/\text{capx}_{t-1} - 1$. Xing (2008) uses investment growth rate to construct an investment factor using return data from 1964 to 2003.

25. **Abnormal capital investment** is defined as capital expenditures scaled by revenues, scaled by the average of this ratio over the previous three years: abnormal capital investment$_t$ = $(\text{capx}_t/\text{revt}_t)/\{[(\text{capx}_{t-1}/\text{revt}_{t-1}) + (\text{capx}_{t-2}/\text{revt}_{t-2}) + (\text{capx}_{t-3}/\text{revt}_{t-3})]/3\}$. Titman, Wei, and Xie (2004) measure the predictive power of abnormal capital investment using return data from July 1973 through June 1996.

26. **Investment-to-assets ratio** is defined as the change in the net value of plant, property, and equipment plus the change in inventory, all scaled by lagged total assets, investment-to-assets ratio$_t$ = $(\Delta\text{ppent}_t + \Delta\text{invt}_t)/\text{at}_{t-1}$. Lyandres, Sun, and Zhang (2008) use the investment-to-assets ratio to predict returns from January 1970 through December 2005.

A.5 Financing

27. **Debt issuance** is defined as indicator variable that takes the value of one if the sum of short- and long-term debt on the balance sheet in year t exceeds this sum in year $t-1$, debt issuance$_t$ =

$\mathbf{1}_{\text{dlc}_t+\text{dltt}_t>\text{dlc}_{t-1}+\text{dltt}_{t-1}}$. Spiess and Affleck-Graves (1999) measure debt offerings using the *Investment Dealers' Digest Directory of Corporate Financing* over the period 1975–1989 as the source, and measure the performance of debt issuers and non-issuers using return data from February 1975 through December 1994.

28. **Leverage** is defined as the ratio of long-term debt and the December book value of equity, $\text{leverage}_t = \text{dltt}_t/\text{me}_t$. Bhandari (1988) uses leverage (defined as total assets minus book value of equity, all divided by the market value of equity) as a return predictor using return data from 1948 to 1979.

29. **One-year share issuance** is the log-change in the split-adjusted number of shares outstanding from fiscal year $t-1$ to t, one-year share $\text{issuance}_t = \text{adjusted shrout}_t/\text{adjusted shrout}_{t-1}$, where adjusted $\text{shrout}_t = \text{shrout}_t * \text{cfacshr}_t$ from CRSP or, if missing or zero, $1000 * \text{csho}_t * \text{ajex}_t$ from Compustat. The number of shares from CRSP are measured at the fiscal-year ends. The share issuance factor takes long positions in firms that repurchase shares (one-year share $\text{issuance}_t < 0$) and short positions in firms that issue shares (one-year share $\text{issuance}_t > 0$). Pontiff and Woodgate (2008) measure the predictive power of one-year share issuance in return data from 1932 through 2003.

30. **Five-year share issuance** is the log-change in the split-adjusted number of shares outstanding from fiscal year $t-5$ to t. The share issuance factor takes long positions in firms that repurchase shares and short positions in firms that issue shares. Daniel and Titman (2006) examine the predictive power of five-year share issuance using return data from July 1968 through December 2003.

31. **Total external financing** is the sum of net share issuance and net debt issuance minus

cash dividends, all scaled by total assets. We compute this measure from the balance sheet as total external financing$_t$ = [(adjusted shrout$_t$/adjusted shrout$_{t-1}-1)*\text{me}_t+(\Delta\text{dlc}_t+\Delta\text{dltt}_t)-\text{dvc}_t]/\text{at}_t$, where the first term approximates share issuance and me_t is the fiscal year-end market value of equity. Bradshaw, Richardson, and Sloan (2006) use a measure of total external financing computed from the statement of cash flows to predict returns using return data from 1971 through 2000. Their measure is defined as total external financing$_t$ = $(\text{sstk}_t - \text{prstkc}_t + \text{dltis}_t - \text{dltr}_t + \text{dlcch}_t - \text{dv}_t)/\text{at}_t$, with dlcch_t set to zero when missing.

A.6 Distress

32. **Ohlson's O-score** is a measure of distress. It is the fitted value from a logistic regression that Ohlson (1980) estimates to explain bankruptcies using data from 1970 and 1976. The fitted values of this regression are given by $\text{O-score}_t = -1.32-0.407*\log(\text{at}_t/\text{cpiind}_t)+6.03*\text{lt}_t/\text{at}_t-1.43*(\text{act}_t-\text{lct}_t)/\text{at}_t+0.076*\text{lct}_t/\text{act}_t-1.72*\mathbf{1}_{\text{lt}_t>\text{at}_t}-2.37*\text{ib}_t/\text{at}_t-1.83*\text{oiadp}_t/\text{lt}_t+0.285*\mathbf{1}_{\text{ib}_t<0\&\text{ib}_{t-1}<0}-0.521*(\text{ib}_t-\text{ib}_{t-1})/(|\text{ib}_t|+|\text{ib}_{t-1}|)$, where cpiind_t is the consumer price index normalized so that 1968 value is 100.[16] Dichev (1998) uses Ohlson O-score to predict stock returns using return data from January 1981 through December 1995.

33. **Altman's Z-score** is another measure of distress. It is the fitted value from a discriminant function that Altman (1968) estimates to predict bankruptcies among 66 companies from 1946 through 1965: Altman's $\text{Z-score}_t = 1.2*(\text{act}_t-\text{lct}_t)/\text{at}_t+1.4*\text{re}_t/\text{at}_t+3.3*(\text{ni}_t+\text{xint}_t+\text{txp}_t)/\text{at}_t+0.6*\text{me}_t/\text{lt}_t+1.0*\text{revt}_t/\text{at}_t$, where me_t is the December market value of equity. Dichev (1998) predicts returns using Altman's Z-score from January 1981 through December 1995.

[16] Because cpiind_t appears inside a log and because we predict the cross section of returns, this price-level adjustment washes out.

34. **Distress risk** is yet another measure of distress. It is the fitted value from a logistic regression that Campbell, Hilscher, and Szilagyi (2008) estimate using data for the period from 1963 through 2003 to predict failures. We use the logit-regression estimates for the 12-month horizon reported in Campbell et al. (2008, Table IV). The original study details the variable construction rules in Section I and the Appendix. Campbell et al. (2008) measure the relation between distress and stock returns using data from 1981 through 2003. We consider the full 1963–2003 period to be the in-sample period.

A.7 Composite anomalies

35. **Piotroski's F-score** is a score that ranges from 0 to 9, constructed by taking the sum of nine binary signals that measure financial performance. The signals are based on income, accruals, ratios of current assets and current liabilities, and so forth. Piotroski (2000) describes the construction of the score in detail, and predicts returns on high book-to-market stocks using return data from 1976 to 1996. We compute Piotroski's F-score for all firms.

36. **Mohanram's G-score** is a score that ranges from 0 to 8, constructed by taking the sum of eight binary signals that measure financial performance. The measure is similar to Piotroski's score except that it replaces some of the signals in Piotroski (2000) with others, and defines the signals on an industry-adjusted basis. An R&D signal, for example, is set one one when a firm's R&D is greater than the contemporaneous median in the same industry. Mohanram (2005) predicts returns on low book-to-market firms using return data from 1979 to 2001. We compute Mohanram's G-score for all firms. We use a score that ranges from 0 to 6 because the other two binary signals use R&D and advertising items that do not exist in the historical NYSE data.

37. **Market-to-book and accruals** is constructed by combining information on book-to-market ratios and accruals. The market-to-book and accruals-signal is set to one for firms that are both in the highest book-to-market quintile and the lowest accruals quintile, and it is set to zero for firms that are both in the lowest book-to-market quintile and in the highest accruals quintile. These quintiles are based on NYSE breakpoints. Bartov and Kim (2004) use this composite anomaly to predict returns between May 1981 and April 2000.

38. **Quality-minus-junk: Profitability** is the profitability component of the quality-minus-junk measure of Asness, Frazzini, and Pedersen (2013). This measure is a combination of six profitability signals: gross profitability, return on equity, return on assets, cash flow to assets, gross margin, and accruals. This composite measure is constructed by transforming each signal into a z-score based on the cross-sectional averages and standard deviations, and by taking the average of the resulting six z-scores. Asness, Frazzini, and Pedersen (2013) examine the predictive power of this composite measure using return data from 1956 to 2012.

REFERENCES

Abarbanell, J. S. and B. J. Bushee (1998). Abnormal returns to a fundamental analysis strategy. *Accounting Review 73*(1), 19–45.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance 23*(4), 589–609.

Alwathainani, A. M. (2009). Consistency of firms' past financial performance measures and future returns. *British Accounting Review 41*, 184–196.

Ang, A., A. A. Shtauber, and P. C. Tetlock (2013). Asset pricing in the dark: The cross-section of OTC stocks. *Review of Financial Studies 26*(12), 2985–3028.

Asness, C., T. J. Moskowitz, and L. H. Pedersen (2013). Value and momentum everywhere. *Journal of Finance 68*(3), 929–985.

Asness, C. S., A. Frazzini, and L. H. Pedersen (2013). Quality minus junk. AQR Capital Management working paper.

Ball, R. and R. Watts (1977). Reply to Salamon and Smith. *Journal of Finance 32*(5), 1802–1808.

Barbee, Jr., W. C., S. Mukherji, and G. A. Raines (1996). Do sales-price and debt-equity explain stock returns better than book-market and firm size? *Financial Analysts Journal 52*(2), 56–60.

Barber, B. M. and J. D. Lyon (1997). Firm size, book-to-market ratio, and security returns: A holdout sample of financial firms. *Journal of Finance 52*(2), 875–883.

Barras, L., O. Scaillet, and R. Wermers (2010). False discoveries in mutual fund performance: Measuring luck in estimated alphas. *Journal of Finance 65*(1), 179–216.

Bartov, E. and M. Kim (2004). Risk, mispricing, and value investing. *Review of Quantitative Finance and Accounting 23*(4), 353–376.

Basu, S. (1977). Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *Journal of Finance 32*(3), 663–682.

Bhandari, L. C. (1988). Debt/equity ratio and expected common stock returns: Empirical evidence. *Journal of Finance 43*(2), 507–528.

Bradshaw, M. T., S. A. Richardson, and R. G. Sloan (2006). The relation between corporate financing activities, analysts' forecasts and stock returns. *Journal of Accounting and Economics 42*(1–2), 53–85.

Campbell, J. Y., J. Hilscher, and J. Szilagyi (2008). In search of distress risk. *Journal of Finance 63*(6), 2899–2939.

Cochrane, J. H. (2011). Presidential address: Discount rates. *Journal of Finance 66*(4), 1047–1108.

Cohen, R. B., C. Polk, and T. Vuolteenaho (2003). The value spread. *Journal of Finance 58*(2), 609–641.

Cooper, M. J., H. Gulen, and M. J. Schill (2008). Asset growth and the cross-section of stock returns. *Journal of Finance 63*(4), 1609–1651.

Daniel, K. and S. Titman (2006). Market reactions to tangible and intangible information. *Journal of Finance 61*(4), 1605–1643.

Davis, J. L., E. F. Fama, and K. R. French (2000). Characteristics, covariances, and average returns: 1929 to 1997. *Journal of Finance 55*(1), 389–406.

Dichev, I. A. (1998). Is the risk of bankruptcy a systematic risk? *Journal of Finance 53*(3), 1131–1147.

Fama, E. F. and K. R. French (1992). The cross-section of expected stock returns. *Journal of Finance 47*(2), 427–465.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns of stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Fama, E. F. and K. R. French (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance 51*(1), 55–84.

Fama, E. F. and K. R. French (1998). Value versus growth: The international evidence. *Journal of Finance 53*(6), 1975–1999.

Fama, E. F. and K. R. French (2001). Disappearing dividends: Changing firm characteristics or lower propensity to pay? *Journal of Financial Economics 60*(1), 3–43.

Fama, E. F. and K. R. French (2006). The value premium and the CAPM. *Journal of Finance 61*(5), 2163–2185.

Fama, E. F. and K. R. French (2008). Average returns, B/M, and share issues. *Journal of Finance 63*(6), 2971–2995.

Fama, E. F. and K. R. French (2010). Luck versus skill in the cross section of mutual fund returns. *Journal of Finance 65*(5), 1915–1947.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

French, K. R. (2008). Presidential address: The cost of active investing. *Journal of Finance 63*(4), 1537–1573.

Gompers, P. A. and J. Lerner (2003). The really long-run performance of initial public offerings: The pre-Nasdaq evidence. *Journal of Finance 58*(4), 1355–1392.

Graham, J. R., M. T. Leary, and M. R. Roberts (2014). Taxes and corporate capital structure in the U.S. Wharton working paper.

Graham, J. R., M. T. Leary, and M. R. Roberts (2015). A century of capital structure: The leveraging of corporate America. *Journal of Financial Economics 118*(3), 658–683.

Harvey, C. R., Y. Liu, and H. Zhu (2015). . . . and the cross-section of expected returns. *Review of Financial Studies 29*(1), 5–68.

Hasbrouck, J. (2009). Trading costs and returns for u.s. equities: Estimating effective costs from daily data. *Journal of Finance 64*(3), 1445–1477.

Haugen, R. A. and N. L. Baker (1996). Commonality in the determinants of expected stock returns. *Journal of Financial Economics 41*(3), 401–439.

Hirshleifer, D., K. Hou, S. H. Teoh, and Y. Zhang (2004). Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics 38*, 297–331.

Hou, K., C. Xue, and L. Zhang (2015). Digesting anomalies: An investment approach. *Review of Financial Studies 28*(3), 650–705.

Jegadeesh, N. and J. Livnat (2006). Revenue surprises and stock returns. *Journal of Accounting and Economics 41*(1–2), 147–171.

Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance 48*(1), 65–91.

Jegadeesh, N. and S. Titman (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *Journal of Finance 56*(2), 699–720.

Kosowski, R., A. Timmermann, R. Wermers, and H. White (2006). Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis. *Journal of Finance 61*(6), 2551–2595.

Lakonishok, J., A. Shleifer, and R. Vishny (1994). Contrarian investment, extrapolation and investment risk. *Journal of Finance 49*(5), 1541–1578.

Lev, B. and D. Nissim (2004). Taxable income, future earnings, and equity values. *Accounting Review 79*(4), 1039–1074.

Linnainmaa, J. (2013). Reverse survivorship bias. *Journal of Finance 68*(3), 789–813.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics 47*(1), 13–37.

Lockwood, L. and W. Prombutr (2010). Sustainable growth and stock returns. *Journal of Financial Research 33*(4), 519–538.

Loughran, T. and J. W. Wellman (2011). New evidence on the relation between the enterprise multiple and average stock returns. *Journal of Financial and Quantitative Analysis 46*(6), 1629–1650.

Lyandres, E., L. Sun, and L. Zhang (2008). The new issues puzzle: Testing the investment-based explanation. *Review of Financial Studies 21*(6), 2825–2855.

McLean, R. D. and J. Pontiff (2015). Does academic research destroy stock return predictability? *Journal of Finance*. Forthcoming.

Mohanram, P. S. (2005). Separating winners from losers among low book-to-market stocks using financial statement analysis. *Review of Accounting Studies 10*(2), 133–170.

Nagel, S. (2005). Short sales, institutional investors and the cross-section of stock returns. *Journal of Financial Economics 78*(2), 277–309.

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

Novy-Marx, R. and M. Velikov (2015). A taxonomy of anomalies and their trading costs. *Review of Financial Studies 29*(1), 104–147.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research 18*(1), 109–131.

Ohlson, J. A. (1995). Earnings, book values, and dividends in equity valuation. *Contemporary Accounting Research 11*(2), 661–687.

Piotroski, J. D. (2000). Value investing: The use of historical financial statement information to separate winners from losers. *Journal of Accounting Research 38*, 1–41.

Pontiff, J. and A. Woodgate (2008). Share issuance and cross-sectional returns. *Journal of Finance 63*(2), 921–945.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance 19*(3), 425–442.

Shumway, T. (1997). The delisting bias in CRSP data. *Journal of Finance 52*(1), 327–340.

Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review 71*(3), 289–315.

Soliman, M. T. (2008). The use of DuPont analysis by market participants. *Accounting Review 83*(3), 823–853.

Spiess, D. K. and J. Affleck-Graves (1999). The long-run performance of stock returns following debt offerings. *Journal of Financial Economics 54*(1), 45–73.

Thomas, J. K. and H. Zhang (2002). Inventory changes and future returns. *Review of Accounting Studies 7*(2), 163–187.

Titman, S., K. C. J. Wei, and F. Xie (2004). Capital investments and stock returns. *Journal of Financial and Quantitative Analysis 39*(4), 677–700.

Vuolteenaho, T. (2002). What drives firm-level stock returns? *Journal of Finance 57*(1), 233–264.

Xing, Y. (2008). Interpreting the value effect through the q-theory: An empirical investigation. *Review of Financial Studies 21*(4), 1767–1795.

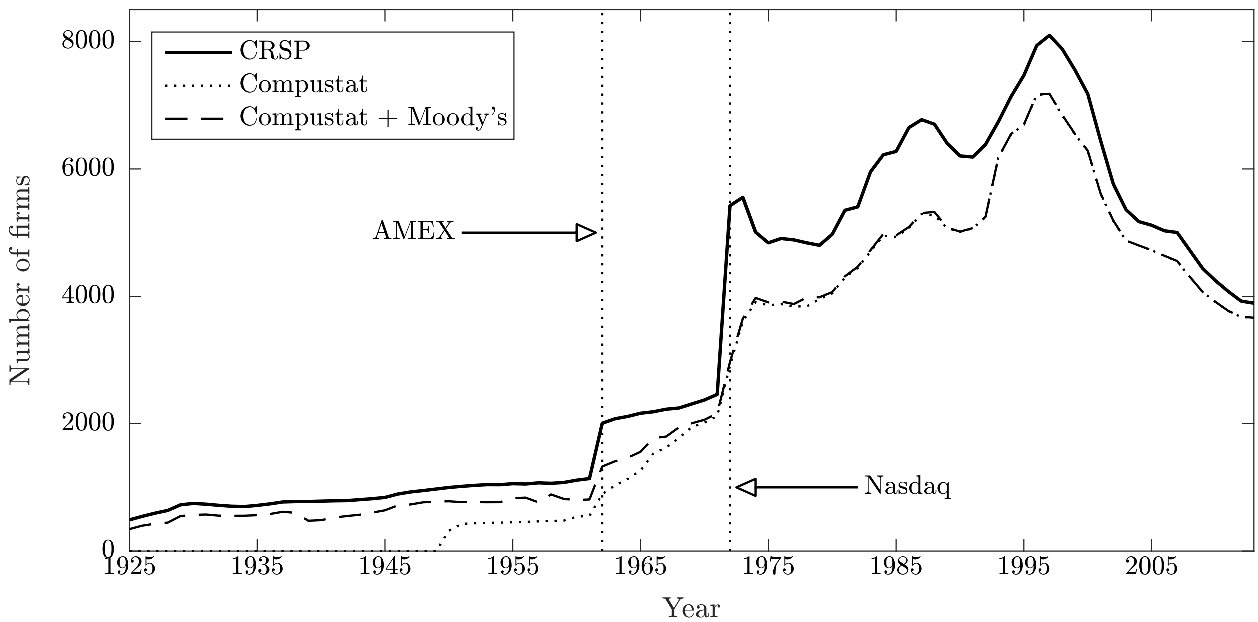
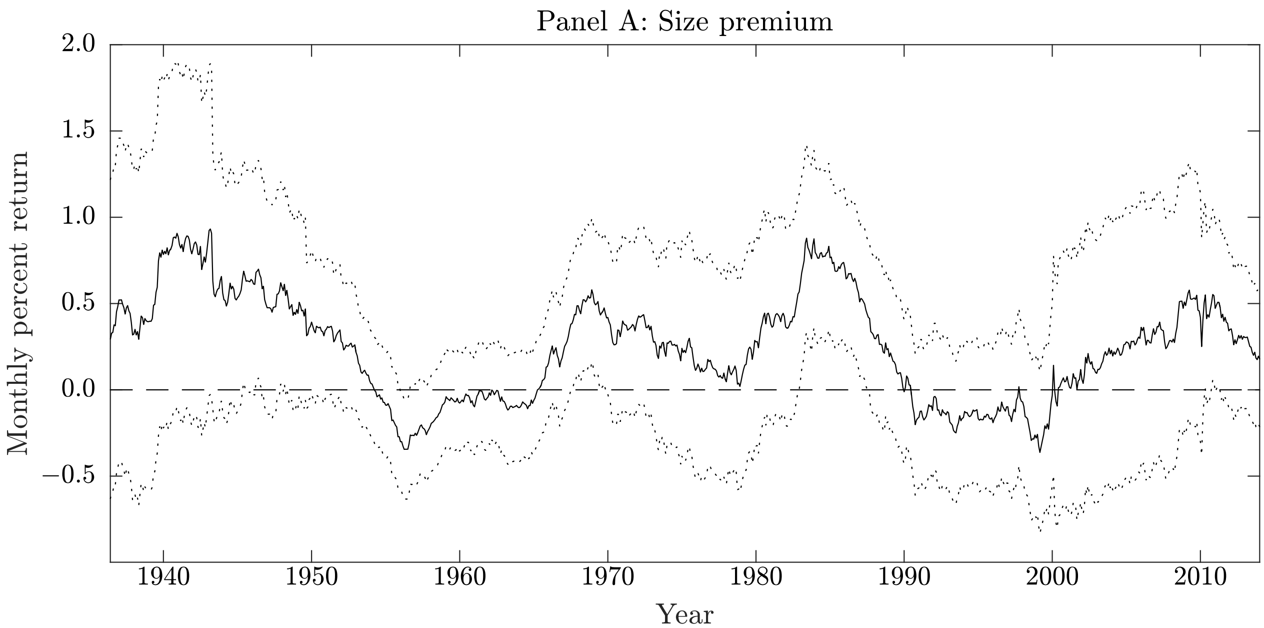
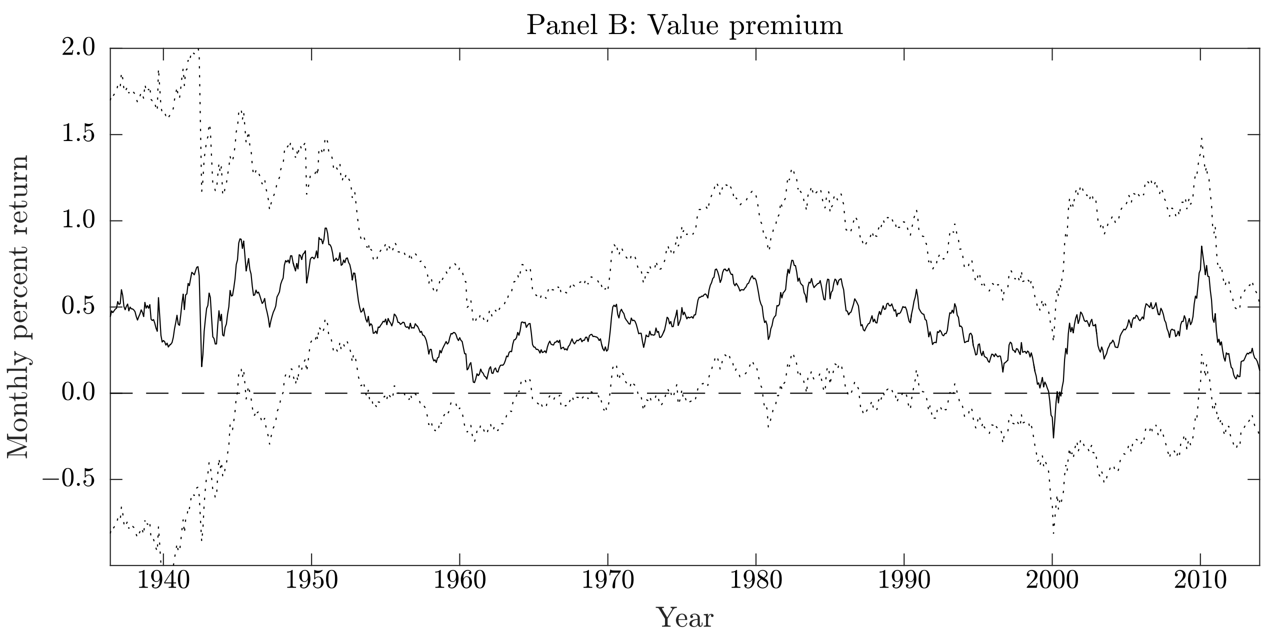


Figure 1: **Number of firms in CRSP, Compustat, and Moody's Industrial and Railroad Manuals, 1925–2013.** This figure shows the number of firms in the Center for Research in Securities Prices (CRSP) database; the number of these firms in Standard and Poor's Compustat database; and the number of these firms in either Compustat or Moody's Industrial and Railroad manuals between 1925 and 2013. The vertical lines indicate the dates on which AMEX (1962) and Nasdaq (1972) stocks are added to the CRSP.


Panel A: Size premium
2.0
1.5
1.0
0.5
0.0
−0.5
Monthly percent return
1940
1950
1960
1970
1980
1990
2000
2010
Year


Panel B: Value premium
2.0
1.5
1.0
0.5
0.0
−0.5
Monthly percent return
1940
1950
1960
1970
1980
1990
2000
2010
Year

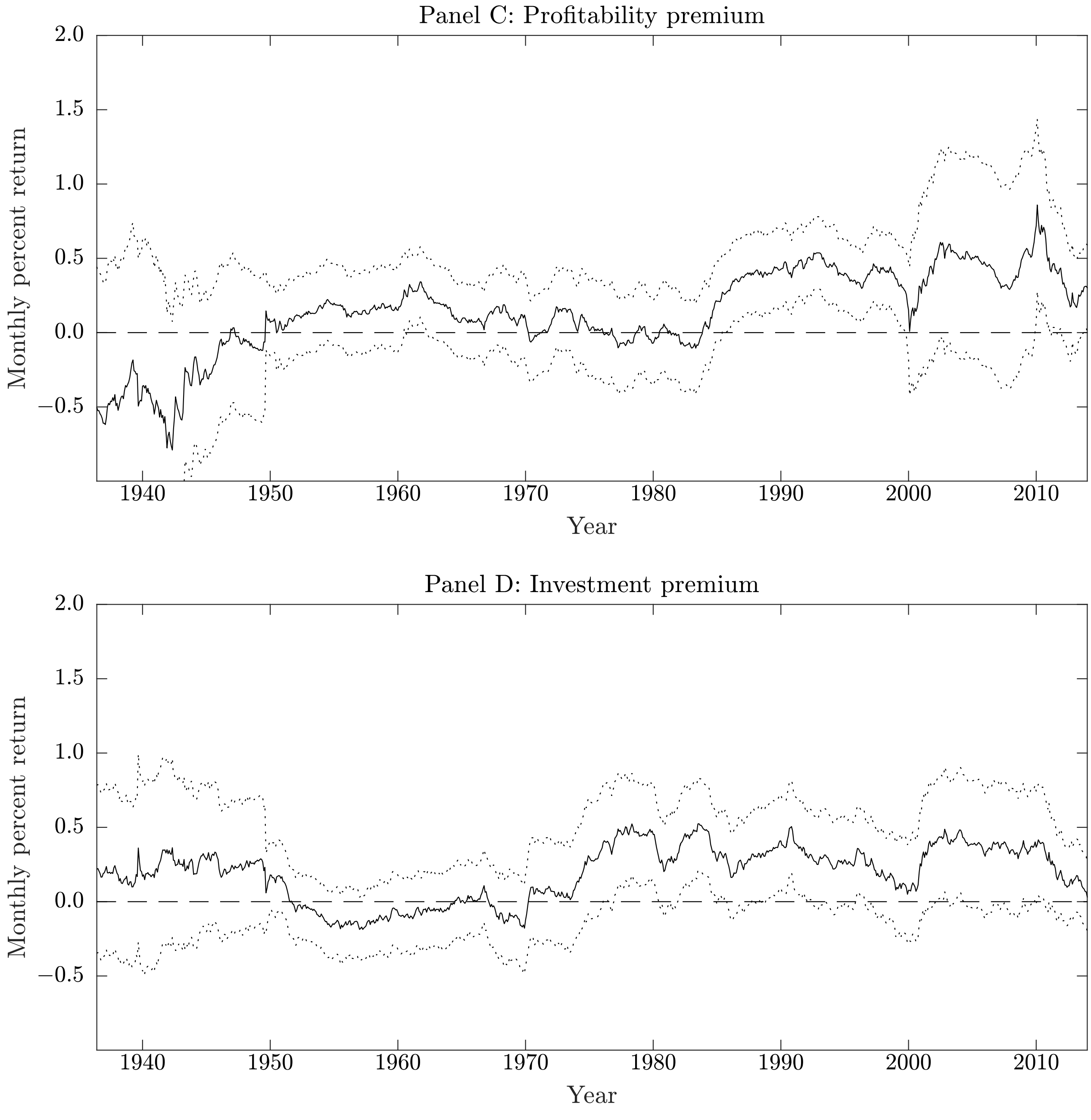


Figure 2: **Monthly percent returns on size, value, profitability, and investment factors, 1926–2014.** This figure reports rolling averages of monthly percent returns for the size (Panel A), value (Panel B), profitability (Panel C), and investment (Panel D) factors from July 1926 through December 2014. Each point reports the average return for a ten-year window up to the date indicated by the x-axis. The first point, for example, corresponds to June 1936, and it represents the average return from July 1926 through June 1936. The dotted lines indicate the 95% confidence intervals around the averages.

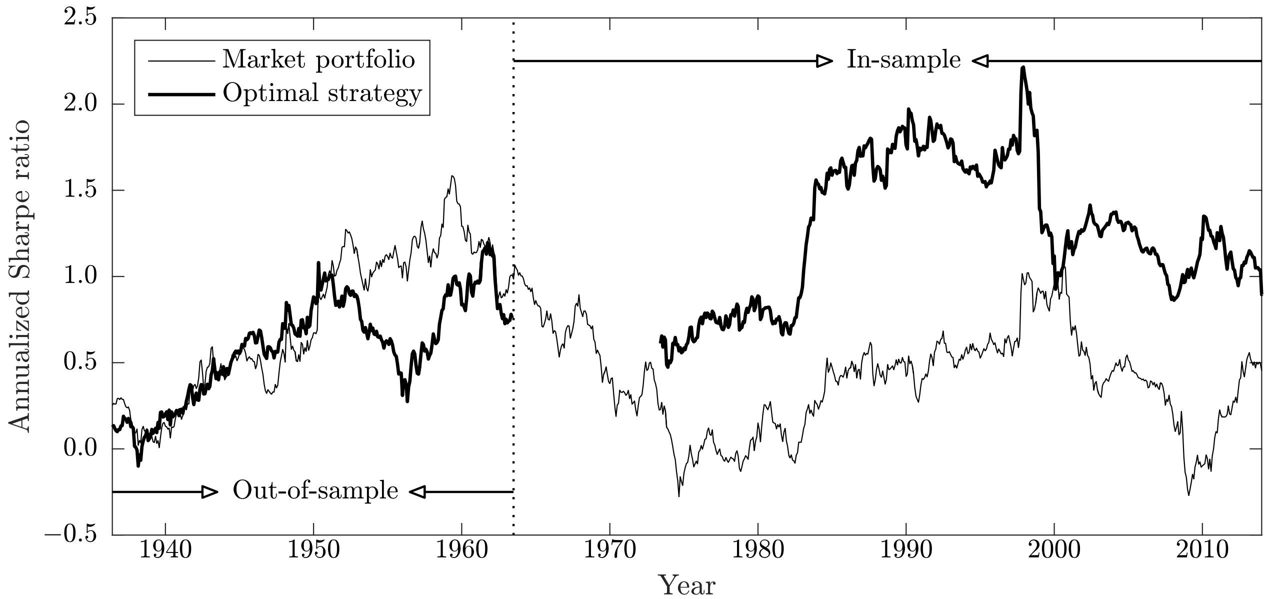


Figure 3: **Annualized Sharpe ratios for the market portfolio and an ex-post mean-variance efficient strategy for ten-year rolling windows, 1926–2014.** This figure reports Sharpe ratios for the market portfolio (thin line) and an ex-post mean-variance efficient strategy (thick line) for ten-year rolling windows. Each point reports the Sharpe ratio for a ten-year window up to the date indicated by the x-axis. The first point, for example, corresponds to June 1936, and it represents the Sharpe ratio from July 1926 through June 1936. We annualized Sharpe ratios computed from monthly returns. The mean-variance efficient strategy is computed using returns from July 1963 through December 2014. It trades the market, size, value, profitability, and investment factors. This strategy is in-sample for the post-1963 period and out-of-sample for the pre-1963 period. The gap in the thick line from July 1963 through May 1973 corresponds to a period during which the optimal strategy is partly in-sample and partly out-of-sample due to the use of ten-year rolling windows.

Table 1: Data coverage: CRSP, Compustat, and Moody's Industrial and Railroad manuals, 1925–1965

This table summarizes the amount of information available on the Center for Research in Security Prices Database, Standard and Poor's Compustat database, and Moody's Industrial and Railroad manuals between 1925 and 1965. The top part of the table shows the number of firms in CRSP; the number of these firms in Compustat; and the number of firms in either Compustat or Moody's manuals. The bottom part of the table reports separately by data item the number of firms covered by Compustat and the number of additional firms covered by Moody's manuals.

	Year								
	1925	1930	1935	1940	1945	1950	1955	1960	1965
					Number of firms				
CRSP	503	775	739	802	860	1,018	1,081	1,140	2,251
Compustat	0	0	0	0	0	324	456	546	1,301
Compustat + Moody's	354	580	575	491	646	790	839	811	1,609
					Number of Compustat firms by data item				
Income statement									
Revenue	0	0	0	0	0	323	456	546	1,294
Cost of goods sold	0	0	0	0	0	255	336	439	1,287
Depreciation	0	0	0	0	0	316	450	538	1,269
Interest	0	0	0	0	0	302	435	525	1,222
Net income	0	0	0	0	0	324	456	546	1,300
Balance sheet									
Total assets	0	0	0	0	0	323	455	546	1,300
Total liabilities	0	0	0	0	0	289	320	411	1,220
Shareholders' equity	0	0	0	0	0	0	0	4	729
Accounts payable	0	0	0	0	0	0	0	4	227
Receivables	0	0	0	0	0	323	387	472	1,189
Inventory	0	0	0	0	0	322	386	469	1,183
					Number of additional firms from Moody's manuals by data item				
Income statement									
Revenue	191	333	395	468	603	451	376	261	304
Cost of goods sold	67	189	266	419	549	399	340	246	279
Depreciation	231	471	537	484	638	445	368	258	264
Interest	155	353	323	256	347	282	262	197	207
Net income	349	561	564	490	646	463	382	266	305
Balance sheet									
Total assets	191	333	395	468	603	451	376	261	304
Total liabilities	67	189	266	419	549	399	340	246	279
Shareholders' equity	231	471	537	484	638	445	368	258	264
Accounts payable	155	353	323	256	347	282	262	197	207
Receivables	349	561	564	490	646	463	382	266	305
Inventory	352	562	568	491	646	462	382	267	305

Table 2: Comparison to Fama-French Factors, 1963–2014

Panel A shows the average monthly percent returns and the associated t-values for the size, value, profitability, and investment factors between July 1963 through December 2014. Panel B reports the correlations between our factors and those reported by Fama and French for the same sample period. We construct size (SMB) and value (HML) factors by sorting stocks into six portfolios by size and book-to-market at the end of each June and by holding the value-weighted portfolios from the July of year t to the June of year $t+1$. These sorts use 50/50 breakpoints for size and 30/40/30 NYSE breakpoints for BE/ME. SMB is the average return on the three "small" portfolios minus that on the three "big" portfolios. HML is the average return on the two "high" portfolios minus that on the two "low" portfolios. The profitability factor (RMW) is constructed by sorting stocks into six portfolios by size and profitability. We compute profitability as revenue minus cost of goods sold and interest scaled by the book value of equity; Fama and French also subtract sales, general & administrative expenses. RMW is the average return on the two "robust" profitability portfolios minus that on the two "weak" profitability portfolios. The investment factor (CMA) is constructed by sorting stocks into six portfolios by size and growth in total assets. CMA is the average return on the two "conservative" investment portfolios minus that on the two "aggressive" investment portfolios.

Panel A: Monthly percent returns

	Our factors		Fama-French factors	
	Mean	t-value	Mean	t-value
Size (SMB)	0.23	1.91	0.24	1.92
Value (HML)	0.36	3.14	0.36	3.15
Profitability (RMW)	0.25	3.03	0.25	2.89
Investment (CMA)	0.26	3.55	0.32	4.03

Panel B: Correlations between our factors and Fama-French factors

	Fama-French factors			
Our factors	SMB	HML	RMW	CMA
Size (SMB)	0.998			
Value (HML)		0.994		
Profitability (RMW)			0.990	
Investment (CMA)				0.980

Table 3: Monthly percent returns on size, value, profitability, and investment portfolios and factors, 1926–2014

Panel A reports average monthly percent returns for the size, value, profitability, and investment factors, and for value-weighted portfolios that are used to construct these factors. The size and value factors are constructed by sorting stocks into portfolios by size and book-to-market; the profitability factor sorts stocks by size and profitability; and the investment factor by size and investment (asset growth). The portfolios are constructed using 50/50 and 30/40/30 NYSE breakpoints and they are rebalanced annually at the end of each June. Panel B reports the average number of stocks in each portfolio. The first two columns divide the pre-1963 sample period (from July 1926 through June 1963) into two segments of equal length (222 months). The last column reports the estimates for the modern ("Compustat") sample period from July 1963 through December 2014. t-values for the average monthly percent returns are reported in parentheses.

Panel A: Monthly percent returns

		Pre-1963 sample			Modern sample
		Jul 1926 – Dec 1944	Jan 1945 – Jun 1963	Jul 1926 – Jun 1963	Jul 1963 – Dec 2014
		Portfolios sorted by size and book-to-market			
Small	Growth	1.04	1.07	1.05	0.92
	Neutral	1.30	1.17	1.24	1.28
	Value	1.68	1.37	1.52	1.42
Big	Growth	0.76	1.10	0.93	0.88
	Neutral	0.80	1.23	1.01	0.94
	Value	1.16	1.47	1.31	1.10
Size factor		0.43	−0.06	0.18	0.23
		(1.42)	(−0.54)	(1.13)	(1.91)
Value factor		0.52	0.33	0.42	0.36
		(1.33)	(2.28)	(2.04)	(3.14)
		Portfolios sorted by size and profitability			
Small	Weak	1.50	1.22	1.36	1.01
	Neutral	1.16	1.24	1.20	1.24
	Robust	1.13	1.23	1.18	1.33
Big	Weak	1.04	1.10	1.07	0.78
	Neutral	0.92	1.21	1.07	0.87
	Robust	0.70	1.39	1.04	0.99
Profitability factor		−0.36	0.15	−0.11	0.26
		(−1.22)	(1.50)	(−0.68)	(3.03)
		Portfolios sorted by size and investment			
Small	Aggressive	1.04	1.16	1.10	1.01
	Neutral	1.57	1.31	1.44	1.33
	Conservative	1.32	1.23	1.27	1.36
Big	Aggressive	0.73	1.19	0.96	0.87
	Neutral	0.99	1.31	1.15	0.93
	Conservative	0.99	0.98	0.98	1.05
Investment factor		0.27	−0.08	0.10	0.26
		(1.38)	(−0.83)	(0.89)	(3.55)

Panel B: Average number of stocks in a portfolio

		Pre-1963 sample			Modern sample
		Jul 1926 – Dec 1944	Jan 1945 – Jun 1963	Jul 1926 – Jun 1963	Jul 1963 – Dec 2014
			Portfolios sorted by size and book-to-market		
Small	Growth	48.8	70.0	59.4	972.2
	Neutral	122.9	192.2	157.5	1,019.4
	Value	145.9	201.0	173.4	1,042.4
Big	Growth	144.0	210.0	177.0	370.5
	Neutral	132.9	179.7	156.3	309.3
	Value	44.2	77.2	60.7	143.3
			Portfolios sorted by size and profitability		
Small	Weak	74.6	128.8	101.7	1,477.8
	Neutral	68.7	119.7	94.2	910.0
	Robust	57.9	105.6	81.8	628.7
Big	Weak	46.3	84.8	65.6	185.1
	Neutral	92.8	165.0	128.9	328.6
	Robust	62.8	106.9	84.9	294.7
			Portfolios sorted by size and investment		
Small	Aggressive	56.2	92.9	74.5	946.6
	Neutral	86.8	119.9	103.4	823.9
	Conservative	87.2	139.8	113.5	1,019.4
Big	Aggressive	83.5	120.5	102.0	287.3
	Neutral	98.0	163.3	130.6	339.7
	Conservative	50.6	71.5	61.1	167.0

Table 4: Defining return anomalies

This table lists the return anomalies examined in this study, the paper that first used each variable to predict the cross-section of returns, and the sample period used in that study. An asterisk denotes an anomaly that is defined differently from the initial study due to the lack of either quarterly data or some data items. The bolded anomalies—value, profitability, and investment—are those studied in detail in Section 3. The anomalies and the approximations are described in the Appendix.

Anomaly group	Anomaly	Original study	Original sample period
Profitability	Gross profitability	Novy-Marx (2013)	1963–2010
	Operating profitability	Fama and French (2015)	1963–2013
	Return on assets	Haugen and Baker (1996)	1979–1993
	Return on equity	Haugen and Baker (1996)	1979–1993
	Profit margin	Soliman (2008)	1984–2002
	Revenue surprises*	Jegadeesh and Livnat (2006)	1987–2003
	Change in asset turnover	Soliman (2008)	1984–2002
Earnings quality	Accruals	Sloan (1996)	1962–1991
	Earnings consistency	Alwathainani (2009)	1971–2007
	Net operating assets	Hirshleifer, Hou, Teoh, and Zhang (2004)	1964–2002
	Net working capital changes	Soliman (2008)	1984–2002
	Tax / income	Lev and Nissim (2004)	1973–2000
Valuation	**Book-to-market**	Fama and French (1992)	1963–1990
	Cash flow / price	Lakonishok, Shleifer, and Vishny (1994)	1968–1990
	Earnings / price	Basu (1977)	1957–1971
	Enterprise multiple	Loughran and Wellman (2011)	1963–2009
	Sales / price	Barbee, Mukherji, and Raines (1996)	1979–1991
Investment and growth	**Asset growth**	Cooper, Gulen, and Schill (2008)	1968–2003
	Growth in inventory	Thomas and Zhang (2002)	1970–1997
	Sales growth	Lakonishok, Shleifer, and Vishny (1994)	1968–1990
	Sustainable growth	Lockwood and Prombutr (2010)	1964–2007
	Adjusted CAPX growth	Abarbanell and Bushee (1998)	1974–1993
	Growth in sales − inventory	Abarbanell and Bushee (1998)	1974–1993
	Investment growth rate	Xing (2008)	1964–2003
	Abnormal capital investment	Titman, Wei, and Xie (2004)	1973–1996
	Investment / assets	Lyandres, Sun, and Zhang (2008)	1970–2005
Financing	Debt issuance*	Spiess and Affleck-Graves (1999)	1975–1994
	Leverage	Bhandari (1988)	1948–1979
	One-year share issuance	Pontiff and Woodgate (2008)	1970–2003
	Five-year share issuance	Daniel and Titman (2006)	1968–2003
	Total external financing	Bradshaw, Richardson, and Sloan (2006)	1971–2000
Distress	O-Score	Dichev (1998)	1981–1995
	Z-Score	Dichev (1998)	1981–1995
	Distress risk	Campbell, Hilscher, and Szilagyi (2008)	1963–2003
Composite anomalies	Piotroski's F-score	Piotroski (2000)	1976–1996
	Mohanram's G-score*	Mohanram (2005)	1979–2001
	M/B and accruals	Bartov and Kim (2004)	1981–2000
	QMJ: Profitability	Asness, Frazzini, and Pedersen (2013)	1956–2012

Table 5: In-sample average returns and CAPM estimates for 38 anomalies

This table reports average returns and monthly CAPM alphas and betas for the 38 anomalies described in Appendix A. Every anomaly is constructed as an HML-like factor by sorting stocks first into six portfolios by size and the anomaly variable at the end of each June. The sorts are done using 50/50 and 30/40/30 NYSE breakpoints. The return on the anomaly factor is the value-weighted average return on the two high portfolios minus that on the two low portfolios. The "high" and "low" labels are chosen based on the original study so that the stocks in the "high" portfolio earn higher returns than those in the "low" portfolios. Column "In-sample period" reports the estimation period, which is the same as that used in each of the original studies indicated in the Appendix and Table 4.

Anomaly group	Anomaly	In-sample period	Average return $\bar{r}$	$t(\bar{r})$	CAPM $\hat{\alpha}$	$t(\hat{\alpha})$	β_{mkt}
Profitability	Gross profitability	1963–2010	0.22	2.32	0.18	1.91	0.09
	Operating profitability	1963–2013	0.25	2.88	0.31	3.56	−0.11
	Return on assets	1979–1993	0.07	0.52	0.07	0.52	0.00
	Return on equity	1979–1993	0.20	1.51	0.16	1.20	0.06
	Profit margin	1984–2002	0.39	2.09	0.57	3.31	−0.27
	Revenue surprises	1987–2003	0.20	1.95	0.24	2.37	−0.07
	Change in asset turnover	1984–2002	0.43	4.36	0.47	4.76	−0.06
Earnings quality	Accruals	1962–1991	0.21	2.46	0.27	3.31	−0.12
	Earnings consistency	1971–2007	0.62	3.30	0.51	2.77	0.22
	Net operating assets	1964–2002	0.35	4.93	0.33	4.70	0.04
	Net working capital changes	1984–2002	0.20	2.13	0.22	2.40	−0.04
	Tax / income	1973–2000	0.13	1.51	0.14	1.72	−0.03
Valuation	Book-to-market	1963–1990	0.44	3.13	0.51	3.85	−0.20
	Cash flow / price	1968–1990	0.66	4.18	0.72	4.94	−0.21
	Earnings / price	1957–1971	0.50	3.37	0.58	4.13	−0.16
	Enterprise multiple	1963–2009	0.24	2.75	0.27	3.22	−0.09
	Sales / price	1979–1991	0.40	2.42	0.40	2.39	0.00
Investment and growth	Asset growth	1968–2003	0.39	4.09	0.47	5.49	−0.19
	Growth in inventory	1970–1997	0.25	2.86	0.32	4.08	−0.14
	Sales growth	1968–1990	0.18	1.62	0.22	2.14	−0.14
	Sustainable growth	1964–2007	0.18	2.06	0.27	3.30	−0.19
	Adjusted CAPX growth	1974–1993	0.19	2.87	0.22	3.21	−0.04
	Growth in sales − inventory	1974–1993	0.31	4.04	0.30	3.86	0.02
	Investment growth rate	1964–2003	0.27	4.39	0.32	5.47	−0.10
	Abnormal capital investment	1973–1996	0.20	3.22	0.20	3.17	0.00
	Investment / assets	1970–2005	0.26	3.18	0.32	4.13	−0.12
Financing	Debt issuance	1975–1994	0.20	4.73	0.18	4.31	0.03
	Leverage	1948–1979	0.19	2.20	0.20	2.30	−0.02
	One-year share issuance	1970–2003	0.30	3.40	0.38	4.71	−0.15
	Five-year share issuance	1968–2003	0.23	2.82	0.29	3.94	−0.15
	Total external financing	1971–2000	0.36	3.11	0.50	5.02	−0.24
Distress	O-Score	1981–1995	0.26	2.23	0.30	2.57	−0.06
	Z-Score	1981–1995	0.09	0.46	−0.08	−0.49	0.26
	Distress risk	1963–2003	0.30	2.40	0.44	4.03	−0.30
Composite anomalies	Piotroski's F-score	1976–1996	0.46	5.57	0.52	6.44	−0.08
	Mohanram's G-score	1979–2001	0.34	2.94	0.47	4.47	−0.19
	M/B and accruals	1981–2000	0.66	2.63	0.90	3.75	−0.29
	QMJ: Profitability	1956–2012	0.24	3.05	0.29	3.78	−0.10

Table 6: Pre-discovery out-of-sample average returns and CAPM estimates for 38 anomalies

This table reports average returns and monthly CAPM alphas and betas for the 38 anomalies described in Appendix A. Panel A uses return data from 1926 up to the beginning of the in-sample period reported in Table 4. Panel B uses the return data from the beginning up to the earlier of two dates: June 1963 or the beginning of the in-sample period.

Panel A: Return data from July 1926 to the beginning of the in-sample period

Anomaly group	Anomaly	Pre-discovery period	Average return $\bar{r}$	Average return $t(\bar{r})$	CAPM $\hat{\alpha}$	CAPM $t(\hat{\alpha})$	CAPM β_{mkt}
Profitability	Gross profitability	1926–1963	0.00	−0.01	0.28	1.73	−0.33
	Operating profitability	1926–1963	−0.06	−0.40	0.11	0.76	−0.20
	Return on assets	1926–1978	−0.13	−0.94	0.06	0.47	−0.29
	Return on equity	1926–1978	−0.07	−0.56	0.05	0.40	−0.19
	Profit margin	1926–1984	−0.12	−1.03	0.01	0.05	−0.20
	Revenue surprises	1931–1987	0.05	0.52	0.11	1.08	−0.08
	Change in asset turnover	1926–1984	0.07	0.95	0.10	1.23	−0.03
Earnings quality	Accruals	1926–1962	0.11	1.08	0.10	1.01	0.01
	Earnings consistency	1932–1970	0.17	0.66	0.04	0.16	0.18
	Net operating assets	1926–1964	0.10	1.20	0.06	0.72	0.05
	Net working capital changes	1926–1983	0.23	3.47	0.27	4.15	−0.06
	Tax / income	1926–1972	0.15	1.86	0.15	1.79	0.00
Valuation	Book-to-market	1926–1963	0.43	2.13	0.12	0.71	0.36
	Cash flow / price	1926–1968	0.18	1.55	0.15	1.28	0.04
	Earnings / price	1926–1957	0.11	0.54	0.28	1.47	−0.19
	Enterprise multiple	1926–1963	0.23	1.73	0.17	1.32	0.06
	Sales / price	1926–1978	0.26	2.41	0.17	1.64	0.14
Investment and growth	Asset growth	1926–1968	0.06	0.59	0.03	0.28	0.04
	Growth in inventory	1926–1969	0.21	2.70	0.21	2.64	0.00
	Sales growth	1926–1968	0.08	0.52	0.02	0.12	0.08
	Sustainable growth	1927–1964	−0.05	−0.37	−0.15	−1.19	0.12
	Adjusted CAPX growth	1926–1973	0.10	1.22	0.01	0.16	0.13
	Growth in sales − inventory	1926–1973	0.08	0.66	0.14	1.20	−0.09
	Investment growth rate	1926–1963	0.08	0.91	0.11	1.23	−0.03
	Abnormal capital investment	1926–1973	0.09	0.86	0.11	1.02	−0.03
	Investment / assets	1926–1969	0.25	3.13	0.24	3.05	0.01
Financing	Debt issuance	1926–1975	0.05	0.92	0.10	1.80	−0.07
	Leverage	1926–1947	0.09	0.62	−0.02	−0.15	0.15
	One-year share issuance	1927–1970	0.20	2.51	0.24	2.95	−0.05
	Five-year share issuance	1931–1968	0.08	1.15	0.15	2.16	−0.07
	Total external financing	1927–1970	−0.14	−1.30	0.05	0.62	−0.26
Distress	O-Score	1926–1980	0.00	0.02	0.24	2.29	−0.35
	Z-Score	1926–1980	−0.12	−0.84	0.11	0.85	−0.34
	Distress risk	1927–1963	0.03	0.13	0.33	1.60	−0.35
Composite anomalies	Piotroski's F-score	1926–1975	0.07	0.66	0.14	1.41	−0.11
	Mohanram's G-score	1926–1978	0.08	0.80	0.24	2.72	−0.24
	M/B and accruals	1926–1981	0.61	2.91	0.51	2.41	0.12
	QMJ: Profitability	1926–1956	0.13	0.55	0.44	2.14	−0.34

Panel B: Return data from July 1926 to min(June 1963, the beginning of the in-sample period)

Anomaly group	Anomaly	Pre-discovery period	Average return $\bar{r}$	$t(\bar{r})$	CAPM $\hat{\alpha}$	$t(\hat{\alpha})$	β_{mkt}
Profitability	Gross profitability	1926–1963	0.00	−0.01	0.28	1.73	−0.33
	Operating profitability	1926–1963	−0.06	−0.40	0.11	0.76	−0.20
	Return on assets	1926–1963	−0.21	−1.11	0.07	0.45	−0.33
	Return on equity	1926–1963	−0.05	−0.27	0.15	0.91	−0.23
	Profit margin	1926–1963	−0.09	−0.57	0.07	0.48	−0.20
	Revenue surprises	1931–1963	−0.07	−0.42	−0.01	−0.08	−0.06
	Change in asset turnover	1926–1963	0.00	−0.03	0.02	0.19	−0.03
Earnings quality	Accruals	1926–1962	0.11	1.08	0.10	1.01	0.01
	Earnings consistency	1932–1963	0.37	1.23	0.26	0.86	0.12
	Net operating assets	1926–1963	0.10	1.19	0.06	0.74	0.05
	Net working capital changes	1926–1963	0.17	2.10	0.21	2.49	−0.04
	Tax / income	1926–1963	0.12	1.21	0.13	1.28	−0.01
Valuation	Book-to-market	1926–1963	0.43	2.13	0.12	0.71	0.36
	Cash flow / price	1926–1963	0.17	1.33	0.14	1.08	0.04
	Earnings / price	1926–1957	0.11	0.54	0.28	1.47	−0.19
	Enterprise multiple	1926–1963	0.23	1.73	0.17	1.32	0.06
	Sales / price	1926–1963	0.18	1.33	0.05	0.40	0.15
Investment and growth	Asset growth	1926–1963	0.10	0.89	0.06	0.54	0.04
	Growth in inventory	1926–1963	0.22	2.66	0.21	2.48	0.02
	Sales growth	1926–1963	0.12	0.66	0.04	0.24	0.09
	Sustainable growth	1927–1963	−0.04	−0.34	−0.14	−1.12	0.12
	Adjusted CAPX growth	1926–1963	0.15	1.45	0.03	0.34	0.14
	Growth in sales − inventory	1926–1963	0.03	0.21	0.11	0.75	−0.09
	Investment growth rate	1926–1963	0.08	0.91	0.11	1.23	−0.03
	Abnormal capital investment	1926–1963	0.11	0.78	0.13	0.97	−0.04
	Investment / assets	1926–1963	0.26	2.90	0.24	2.75	0.01
Financing	Debt issuance	1926–1963	0.02	0.28	0.09	1.29	−0.08
	Leverage	1926–1947	0.09	0.62	−0.02	−0.15	0.15
	One-year share issuance	1927–1963	0.18	1.92	0.22	2.34	−0.05
	Five-year share issuance	1931–1963	0.04	0.54	0.12	1.50	−0.08
	Total external financing	1927–1963	−0.22	−1.67	0.01	0.06	−0.26
Distress	O-Score	1926–1963	0.05	0.28	0.38	2.73	−0.38
	Z-Score	1926–1963	−0.16	−0.76	0.21	1.32	−0.43
	Distress risk	1927–1963	0.03	0.13	0.33	1.60	−0.35
Composite anomalies	Piotroski's F-score	1926–1963	0.01	0.08	0.10	0.78	−0.11
	Mohanram's G-score	1926–1963	0.06	0.42	0.28	2.46	−0.26
	M/B and accruals	1926–1963	0.52	1.74	0.21	0.71	0.26
	QMJ: Profitability	1926–1956	0.13	0.55	0.44	2.14	−0.34

Table 7: Post-discovery out-of-sample average returns and CAPM estimates for 36 anomalies

This table reports average returns and monthly CAPM alphas and betas for 36 anomalies described in Appendix A. Anomaly returns are measured using data that have accumulated after the sample used in the original study. Each anomaly's return series starts one month after the end of the original sample and ends in December 2014. We exclude operating profitability and quality-minus-junk: profitability from the list of anomalies because they have no more than two years of post-discovery data.

Anomaly group	Anomaly	Post-discovery period	Average return $\bar{r}$	Average return $t(\bar{r})$	CAPM $\hat{\alpha}$	CAPM $t(\hat{\alpha})$	CAPM β_{mkt}
Profitability	Gross profitability	2011–2014	0.16	0.80	0.35	1.72	−0.15
	Return on assets	1994–2014	0.20	1.22	0.37	2.63	−0.28
	Return on equity	1994–2014	0.26	1.51	0.45	2.94	−0.30
	Profit margin	2003–2014	−0.35	−2.00	−0.13	−0.88	−0.27
	Revenue surprises	2004–2014	0.01	0.11	0.12	1.01	−0.16
	Change in asset turnover	2003–2014	−0.06	−0.58	−0.08	−0.74	0.02
Earnings quality	Accruals	1992–2014	0.20	2.14	0.17	1.83	0.04
	Earnings consistency	2008–2014	0.53	1.04	0.18	0.41	0.48
	Net operating assets	2003–2014	−0.04	−0.26	−0.10	−0.65	0.08
	Net working capital changes	2003–2014	−0.05	−0.58	−0.09	−0.96	0.05
	Tax / income	2001–2014	0.03	0.27	0.01	0.12	0.04
Valuation	Book-to-market	1991–2014	0.28	1.50	0.41	2.29	−0.19
	Cash flow / price	1990–2014	0.29	1.45	0.49	2.65	−0.30
	Earnings / price	1971–2014	0.38	2.65	0.54	4.20	−0.31
	Enterprise multiple	2010–2014	0.07	0.51	0.03	0.19	0.03
	Sales / price	1992–2014	0.47	2.51	0.52	2.80	−0.09
Investment and growth	Asset growth	2003–2014	0.11	0.96	0.08	0.72	0.04
	Growth in inventory	1998–2014	0.12	1.08	0.13	1.18	−0.02
	Sales growth	1990–2014	−0.05	−0.42	0.08	0.68	−0.19
	Sustainable growth	2007–2014	0.17	1.15	0.16	1.03	0.03
	Adjusted CAPX growth	1994–2014	0.19	2.10	0.23	2.67	−0.07
	Growth in sales − inventory	1994–2014	0.10	1.17	0.10	1.13	0.00
	Investment growth rate	2004–2014	−0.06	−0.56	−0.06	−0.53	0.00
	Abnormal capital investment	1996–2014	0.08	0.70	0.04	0.37	0.07
	Investment / assets	2006–2014	0.16	0.96	0.09	0.57	0.10
Financing	Debt issuance	1995–2014	0.14	1.53	0.08	0.94	0.08
	Leverage	1980–2014	0.12	0.82	0.18	1.24	−0.10
	One-year share issuance	2004–2014	0.03	0.30	0.10	1.09	−0.11
	Five-year share issuance	2004–2014	0.09	0.99	0.15	1.85	−0.10
	Total external financing	2001–2014	0.31	1.89	0.45	3.50	−0.30
Distress	O-Score	1996–2014	−0.03	−0.22	0.04	0.31	−0.10
	Z-Score	1996–2014	−0.07	−0.45	−0.12	−0.76	0.08
	Distress risk	2004–2014	0.31	1.14	0.61	2.77	−0.46
Composite anomalies	Piotroski's F-score	1997–2014	0.17	0.97	0.32	2.05	−0.27
	Mohanram's G-score	2002–2014	0.26	1.53	0.44	3.42	−0.32
	M/B and accruals	2000–2014	0.40	1.13	0.40	1.13	−0.01

Table 8: Measuring the effects of data mining on the in-sample anomaly returns

This table reports estimates that compare the performance of the average anomaly between the sample period used in the original study (original sample), between July 1926 and the beginning of the in-sample period (pre-discovery sample), and between the end of the in-sample period and December 2014 (post-discovery sample). We estimate average monthly returns, annualized volatilities, CAPM alphas, and information ratios from the CAPM for 33 anomalies listed in the Appendix. We also exclude composite anomalies and those anomalies (operating profitability and quality-minus-junk: profitability) that have no more than two years of post-discovery out-of-sample data. The information ratio ("marginal Sharpe ratio") from the CAPM is $t(\hat{\alpha})$ multiplied by $\sqrt{12}/\sqrt{T}$, where T is the length of the estimation period in months. We report t-values in parentheses. These t-values are computed by block bootstrapping monthly anomaly returns by calendar month.

				Differences		
Measure	Pre-discovery sample	Original sample	Post-discovery sample	Pre − original	Post − original	Post − pre
			Excess returns			
Average return	0.09	0.28	0.11	−0.19	−0.16	0.03
	(2.59)	(8.56)	(3.31)	(−4.10)	(−4.40)	(0.56)
Volatility	9.4%	6.4%	6.9%	3.1%	0.5%	−2.5%
				(4.36)	(2.19)	(−3.53)
Sharpe ratio	0.14	0.54	0.17	−0.40	−0.38	0.03
	(3.33)	(8.70)	(2.69)	(−5.55)	(−4.81)	(0.38)
			CAPM			
Alpha	0.12	0.31	0.16	−0.19	−0.15	0.04
	(4.42)	(10.92)	(4.90)	(−4.69)	(−3.96)	(0.93)
Information ratio	0.18	0.63	0.26	−0.45	−0.37	0.08
	(4.59)	(10.51)	(3.93)	(−6.17)	(−4.44)	(1.02)

Table 9: Changes in the correlation structure of returns

This table reports estimates from two panel regressions that explain monthly anomaly returns using the average monthly returns on all other anomalies that are either in-sample or out-of-sample. In the first column's regression, $\text{post}_{i,t}$ takes the value of one if anomaly i is in the post-discovery out-of-sample period in month t, in-sample $\text{index}_{-i,t}$ is the average return over all other anomalies (that is, except anomaly i) that are in-sample in month t, and post-sample $\text{index}_{-i,t}$ is the average return over all other anomalies that are in the post-discovery out-of-sample period in month t. The second column's regression is the same except that it swaps the post-discovery sample with the pre-discovery sample, and redefines the explanatory variables accordingly. Standard errors are clustered by calendar month.

Specification			
In-sample versus post-discovery out-of-sample		In-sample versus pre-discovery out-of-sample	
Regressor	Estimate	Regressor	Estimate
Intercept	0.001	Intercept	0.640
	(7.07)		(17.54)
$\text{Post}_{i,t}$	−0.001	$\text{Pre}_{i,t}$	0.000
	(−4.52)		(−1.85)
In-sample $\text{index}_{-i,t}$	0.482	In-sample $\text{index}_{-i,t}$	0.640
	(10.76)		(17.54)
Post-sample $\text{index}_{-i,t}$	0.092	Pre-sample $\text{index}_{-i,t}$	0.077
	(8.19)		(3.78)
$\text{Post}_{i,t} *$ In-sample $\text{index}_{-i,t}$	−0.373	$\text{Pre}_{i,t} *$ In-sample $\text{index}_{-i,t}$	−0.622
	(−7.64)		(−15.37)
$\text{Post}_{i,t} *$ Post-sample $\text{index}_{-i,t}$	0.575	$\text{Pre}_{i,t} *$ Pre-sample $\text{index}_{-i,t}$	0.349
	(21.94)		(8.07)
Adjusted R^2	8.1%	Adjusted R^2	4.3%
N	15,962	N	15,672