

NBER WORKING PAPER SERIES

THE HISTORY OF THE CROSS SECTION OF STOCK RETURNS

Juhani T. Linnainmaa
Michael R. Roberts

Working Paper 22894
<http://www.nber.org/papers/w22894>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2016

We thank Mike Cooper (discussant), Ken French, Travis Johnson (discussant), Mark Leary, Jon Lewellen, David McLean, and Jeff Pontiff for helpful discussions, and seminar and conference participants at University of Lugano, University of Copenhagen, University of Texas at Austin, SFS 2016 Finance Cavalcade, and Western Finance Association 2016 meetings for valuable comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2016 by Juhani T. Linnainmaa and Michael R. Roberts. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The History of the Cross Section of Stock Returns
Juhani T. Linnainmaa and Michael R. Roberts
NBER Working Paper No. 22894
December 2016
JEL No. G11,G12,G14

ABSTRACT

Using data spanning the 20th century, we show that most accounting-based return anomalies are spurious. When examined out-of-sample by moving either backward or forward in time, anomalies' average returns decrease, and volatilities and correlations with other anomalies increase. The data-snooping problem is so severe that even the true asset pricing model is expected to be rejected when tested using in-sample data. Our results suggest that asset pricing models should be tested using out-of-sample data or, when not feasible, by whether a model is able to explain half of the in-sample alpha.

Juhani T. Linnainmaa
USC FBE Dept.
Bridge Hall - 308, MC-0804
3670 Trousdale Parkway, Ste. 308
Los Angeles, CA 90089-0804
and NBER
juhani.linnainmaa@marshall.usc.edu

Michael R. Roberts
The Wharton School
University of Pennsylvania
3620 Locust Walk, #2319
Philadelphia, PA 19104
and NBER
mrrobert@wharton.upenn.edu

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. 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 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

¹The 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).

²These 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, Fama, and French (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.

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, during which the average returns are two (t -value = 0.14) and nine (t -value = 0.86) basis points per month, respectively. Between 1963 and 2014, these factors average 30 and 25 basis points per month with t -values of 3.60 and 3.41, respectively. The absence of these premiums stands in contrast to the value effect, which is statistically significant in the pre-1963 data (Fama and French 2006). A three-factor model adjustment increases the alpha of the profitability factor to 25 basis points per month (t -value = 1.90) in the pre-1963 data. This increase is due to the same negative correlation between profitability and value that is present in the post-1963 sample (Novy-Marx 2013). The three-factor model adjustment has no effect on the investment factor's insignificant alpha before 1963.

The attenuations of the investment and profitability premiums in the pre-1963 data are representative of most of the other 33 anomalies that we examine. Just eight out of the 36 (investment, profitability, value, and 33 others) earn average returns that are positive and statistically significant at the 5% level in the pre-1963 period. In the CAPM and three-factor models, the numbers of anomalies with statistically significant alphas are 8 and 14. These results are not due to lack of power. In most cases, the historical out-of-sample period is 37 years long, typically longer than the original study's sample period.

We measure the performance of the average anomaly during the pre-discovery, in-sample, and post-discovery periods to obtain precise estimates of the changes in average returns, Sharpe ratios, CAPM and three-factor model alphas, and information ratios from the CAPM and the three-factor model. All of these measures decrease sharply and statistically significantly when we move out of the original study's sample period by going either backward or forward in time. The average anomaly's Sharpe ratio and the information ratios from the CAPM and the three-factor model, for example, decrease between 60% to 77%. Anomalies perform significantly better during the in-sample period based on every metric; by contrast,

the differences between the post- and pre-discovery periods are rarely more than one standard error from zero, and never statistically significant.

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 the data point towards 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. The decrease in the average anomaly's information ratio typically exceeds that in its alpha, suggesting that anomalies are atypically safe in the original study's sample.

Our results do not suggest that all return anomalies are spurious. The average in-sample anomaly earns a CAPM alpha of 32 basis points per month (t -value = 10.87). The average alpha is 13 basis points (t -value = 4.42) per month for the pre-discovery sample and 14 basis points (t -value = 4.06) for the post-discovery sample. Although these estimates lie far below the in-sample numbers, they are highly statistically significant. 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.³ Those who assume that the cross section is immutable may

³This 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 positive. These tests, however, do not assist in identifying the skilled managers.

be disappointed. For example, using post-1963 data to construct the mean-variance efficient strategy from the market, size, value, investment, and profitability factors performs just the same as 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. However, 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.

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 and three-factor model alphas of 36 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 2015. 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.⁴ 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 back-filled information only for those firms that were deemed to be of the greatest interest to these analysts (Ball and Watts 1977).

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 Fama and French (1992) 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 $t - 1$.

⁴See, 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.

2.2 Coverage

Table 1 shows the number of firms in the CRSP database at five years intervals from 1925 through 1965. There are 490 (NYSE) firms on CRSP at the very beginning. The number of CRSP firms increases over time, reaching 1,113 firms in 1960. The large jump to 2,164 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 320 of the 998 NYSE firms. By 1965, which is the date by which Compustat is survivorship-bias free, the accounting data are available for 1,265 of the 2,164 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 345 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 320 firms, and the Moody's manuals have data for 462 additional firms. These manuals remain an important source even after 1962; in 1965, these manuals provide information for 294 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 2014.⁵ 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.

⁵We exclude year 2015 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 2015.

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 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 show that the results are both qualitatively and quantitatively the same when we exclude the pre-Securities Exchange Act era.⁶

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.⁷ So, real-world income statement and balance sheet information rarely line up exactly as they should under this ideal. We test the extent to which the

⁶Cohen, Polk, and Vuolteenaho (2003) also note that the pre-1936 data are congruent with the later data: “It is comforting, however, that our main regression results are robust to the choice between the 1928–1997 and 1936–1997 periods.” The timing convention in Cohen, Polk, and Vuolteenaho (2003) is such that their year 1936 observations use book values from 1937. In our subsample analysis, we start the return data in July 1938 so that, consistent with Cohen, Polk, and Vuolteenaho (2003), the book values of equity come from 1937.

⁷See endnote 1 in Ohlson (1995) for examples.

historical data conforms to this standard relative to the modern data. A test of conformity to clean-surplus accounting is 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 clean-surplus accounting, implied log-profitability equals

$$\text{implied log-profitability}_t = \log \left\{ \frac{(1 + R_t) \times \text{ME}_{t-1} - D_t}{\text{ME}_t} \times \frac{\text{BE}_t}{\text{BE}_{t-1}} - \left[1 - \frac{D_t}{\text{BE}_{t-1}} \right] \right\}, \quad (1)$$

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 .⁸ 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 the 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 1.07 (SE = 0.05), and the adjusted R^2 is 32%. In the post-1963 data, the slope is 0.64 (SE = 0.02), and the adjusted R^2 is 41%. In cross-sectional regressions, the average slope estimate is 1.00 for the pre-1963 sample and 0.75 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, we can 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 data than others, such as those based on the book value of equity. Book value of equity is potentially problematic because

⁸See, for example, Vuolteenaho (2002), Cohen, Polk, and Vuolteenaho (2003), and Nagel (2005).

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. With a finite number of firms, even a randomized factor's volatility is positive because, first, the portfolios do not perfectly diversify idiosyncratic risk and, second, because the firms in the high and low portfolios have slightly different fundamentals because of chance. We therefore test how volatile an actual anomaly is relative to its expected volatility under the null hypothesis that the anomalies sort on noise.

In section 4, we construct HML-like factors for 36 anomalies. The average anomaly factor's annualized return volatility is 9.4% in the pre-1963 data. The volatility of the average randomized factor, constructed from the same set of stocks as the actual factor, is 5.7% in the pre-1963 data, and so the amount of excess volatility is 3.6% (SE = 0.6%). In the post-1963 data, the amount of excess volatility is 7.1% – 3.1% = 3.9% (SE = 0.2%). The 0.3% difference between the pre- and post-1963 periods has a t -value of 0.4.⁹ The comparable amounts

⁹We estimate the standard errors for the excess volatilities and their difference by block bootstrapping the data by calendar month. We measure the volatility of each actual and randomized factor and then compute

of 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 adding the profitability and investment factors to the three-factor model. 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 36 anomalies.

3.1 Defining factors

Both Fama and French (2015) and Hou, Xue, and Zhang (2015) measure investment as the 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 alternatively known as the asset-growth anomaly (Cooper, Gulen, and Schill 2008).¹⁰ We follow Fama and French (2015) and construct HML-like profitability and investment factors by sorting stocks into six portfolios by size and profitability, or by size and investment. For example, we construct the following six portfolios at the end of each June using NYSE breakpoints to generate the investment factor:

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

the volatility and excess volatility of the average anomaly. We then resample the data with replacement and repeat the computations. The average randomized factor is more volatile in the pre-1963 data—5.7% versus 3.1%—because of the smaller number of stocks.

¹⁰We evaluate other investment-based anomalies in Section 4.

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 the average return on the two low investment portfolios minus the average return on the two high investment portfolios. 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 $\text{profitability}_t = (\text{revt}_t - \text{cogs}_t - \text{xsga}_t - \text{xint}_t) / \text{be}_t$. 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 high-profitability portfolios minus the average return on the two low-profitability portfolios. The size (SMB) and value (HML) factors are defined similarly (see Fama and French (1993)).

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 2015. 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 nearly identical but for investment that reveals a five-basis point difference. 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 compares the performance of the four factors between the pre- and post-1963 sample period. The pre-1963 sample period runs from July 1926 through June 1963 and the post-1963 sample period runs from July 1963 through December 2015. We further divide the pre-1963 sample into two subperiods. The early part runs from July 1926 through June 1938 and the

late part from July 1938 through June 1963. The Securities and Exchange Act had been in effect for two years by the time the late part begins (Cohen, Polk, and Vuolteenaho 2003).

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 0.43% with a t -value of 2.09—the premiums associated with the size, profitability, and investment factors are not. The average return on the size factor is 0.19% (t -value = 1.16), and those on the investment and profitability factors are 0.02% (t -value = 0.14) and 0.09% (t -value = 0.86).¹¹ 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.

Panel B of Table 3 shows that the absence of profitability and investment premiums is unlikely due to any lack of statistical power. The six portfolios are reasonably well diversified even during the early part of the pre-1963 sample. Over the entire pre-1963 sample, the average number of stocks per portfolio is always above 50. This amount of diversification, combined with the length of the sample period (37 years) gives us confidence that we should be able to detect return premiums when they exist. Moreover, the absence of the profitability and investment premiums is unlikely to be due to the fact that they use complex accounting 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 the value premium is statistically significant in the pre-1963 data.

¹¹In Table 3, we define the profitability factor without the SG&A term. Companies did not historically report these expenses, and so we construct the factor without them to maintain comparability throughout the 1926–2015 sample. This alternative profitability factor is superior to the original factor in the post-1963 sample—its t -value of 3.60 exceeds the t -value of 2.88 on the with-SG&A version—and so this change does not handicap the factor. This performance improvement is consistent with Ball, Gerakos, Linnainmaa, and Nikolaev (2015). They note that Compustat adds R&D expenses to XSGA even when companies report R&D expenses as a separate line item. An operating profitability measure’s predictive power increases substantially when SG&A is not used to compute the profitability measure or when the R&D expenses are removed from XSGA.

3.3 Cross sections of profitability and investment

Figure 2 shows how the cross sections of profitability and investment evolve between 1926 and 2015 by plotting these variables' decile breakpoints. The distribution of profitability changes over time. First, except for the 1940s, the distribution widens over time. Second, the Great Depression, World War II and, to a lesser extent, the recovery from the financial crisis, appear as shocks that shift the entire distribution. Panel B shows that asset growth (investment) is significantly more volatile than profitability, and its aggregate fluctuations—which register as shifts in the entire distribution—more pronounced.

In Table A2 in the Appendix, we show that the distributions of fundamentals are not related to the premiums earned by the profitability and investment factors. Whereas the value spread—that is, the difference in the average book-to-market ratios of value and growth firms—positively and statistically significantly predicts the value premium (Cohen, Polk, and Vuolteenaho 2003), neither profitability nor asset growth spread predicts their factors' return premiums. Figure A1 in the Appendix shows the distributions of the other anomalies that we study later and Table A2 shows that, in most cases, the anomaly spreads are unrelated to the factor premiums.

3.4 Alphas and subsample analysis

Panel C of Table 3 shows the CAPM alphas for the four factors and three-factor model alphas for the two factors, profitability and investment, that are not part of this model. These regressions are important from the investing viewpoint. A statistically significant alpha implies that the combination of the right-hand side factors is not mean-variance efficient; an investor could improve his Sharpe ratio by adding the left-hand side factor to his portfolio. From the asset pricing perspective, a statistically significant alpha implies that adding the left-hand side factor to the asset pricing model improves it (Barillas and Shanken 2015).

All four CAPM alphas are statistically insignificant during the entire pre-1963 period and during both subperiods. The insignificance of the value factor is consistent with Ang and Chen (2007), and its insignificance stems from value factor's positive market beta during this

period. In the three-factor model, the profitability factor is significant at the 10% level during the entire pre-1963 period, and at a 5% level during the later part of this period from July 1938 through June 1963. The three-factor model alpha is higher than the CAPM alphas because of the negative correlation between value and profitability (Novy-Marx 2013). The investment factor's three-factor model alpha, however, is lower than its CAPM alpha.

Figure 3 reports average returns for the same factors using rolling ten-year windows. For profitability and investment factors, we plot both the average returns on the standard factors as well as on the orthogonal components of these factors. A factor's orthogonal component in month t is equal to its alpha from the three-factor model regression plus the month- t residual. The time-series behavior of the value premium differs significantly from those of the other premiums. Whereas the value premium is positive almost throughout the full sample period except for the interruption towards the end of the 1990s during the Nasdaq episode, the other premiums are less stable.

The size factor performs poorly in the 1950s and 60s, and then again in the 90s, and it is too volatile to attain but fleeting periods of statistical significance. The investment premium is positive until 1950 after which point it turns and remains negative until the mid-1970s. The profitability premium is negative before 1950 and then again around 1980. However, the negative correlation between profitability and value is apparent throughout the entire 1926–2016 sample. Except for the very end of this long sample, the return on the orthogonal component of the profitability factor exceeds that on the profitability factor. Although the orthogonal component of profitability also suffers some losses, these down periods are shorter and milder than what they are without the value factor.

3.5 An investment perspective

The pre-1963 sample looks very different from the post-1963 data in terms of the profitability and investment premiums. Figure 4 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 2015. We report the Sharpe ratios for rolling ten-year windows.

The market’s Sharpe ratio for the entire 1926 through 2015 period is 0.42. It is slightly higher (0.46) for the pre-1963 sample than for the post-1963 sample (0.39). The optimal strategy’s Sharpe ratio for the post-1963 sample period is 1.07; by construction, this strategy is in-sample for this period. However, for the pre-1963 sample, the Sharpe ratio of this strategy is just 0.53, that is, almost the same as that of the market. Figure 4 shows that the optimal strategy rarely dominates the market portfolio by a wide margin in the pre-Compustat period; at the same time, the optimal strategy performs very poorly relative to the market 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 4 shows that the strategy that is (ex-post) optimal in the post-1963 data is 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 36 anomalies

4.1 Competing explanations for cross-sectional return anomalies

In this section, we use data on 36 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. Under the mispricing explanation, we expect the anomalies to grow stronger as we move backward in time. 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 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.

¹²See also French (2008).

¹³Researchers 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, Shtaubert, 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).

4.2 Defining anomalies

In this section, we compare how the returns on 36 accounting-based anomalies differ between the in-sample period (used in each original study) and out-of-sample periods that either pre-date 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 because the premiums are too noisy.

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 because many of these anomalies have already been taken to the pre-1963 period, often already in the original study. To the best of our knowledge, our list of anomalies is comprehensive: we include all accounting-based anomalies that can be replicated or reasonably approximated using the data from the Moody’s manuals.

We use the same definitions for all 36 anomalies—that is, value, profitability, investment, and the 33 additional anomalies—throughout the 1926–2015 sample period. For example, even though we could start using reported capital expenditures (CAPX) from Compustat to construct some of the 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 period. Table A1 in the Appendix describes these approximations and compares the average returns and the CAPM and three-

factor model alphas of the original definitions and the approximations.

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 of the two high portfolios minus the average of the two 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 exceed those of the high portfolios.

4.3 In-sample estimates

Table 5 reports the average monthly percent returns and the CAPM and three-factor model alphas for the 36 anomalies. We estimate the average returns and alphas 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).

Table 5 shows that 30 anomalies earn average returns that are positive and statistically significant at the 5% level. In the CAPM and the three-factor model, the numbers of positive and statistically significant anomalies are 32 and 29. Every anomaly is statistically significant at the 5% level in either the CAPM or the three-factor model. The differences between the average returns and alphas are sometimes large. The average return on the distress anomaly, for example, is 30 basis points per month (t -value = 2.4). However, because this anomaly covaries negatively with the market and HML factors (Campbell, Hilscher, and Szilagyi 2008), its CAPM and three-factor model alphas are considerably higher, 45 basis points (t -value = 4.09) and 49 basis points (t -value = 4.82) per month, respectively.

Some of the most impressive t -values belong to the composite anomalies. Piotroski’s F-score, which is a combination of 9 firm-quality signals, earns a three-factor model alpha of 60 basis points per month, which is statistically significant with a t -value of 7.48. Judged by the three-factor model alphas, gross profitability, net operating assets, and total external financing are among the best-performing non-composite anomalies.

Table 6 averages the estimates for the seven anomaly categories and includes an “all” category that includes all anomalies except for the two composite anomalies. The average anomaly is highly profitable during the in-sample period. Its average return is 28 basis points per month (t -value = 8.67); its CAPM alpha is 32 basis points (t -value = 10.83); and its three-factor model alpha is 26 basis points (t -value = 11.03).¹⁴

4.4 Pre-discovery out-of-sample estimates

Panel A of Table 7 reports average returns and alphas for the same 36 anomalies using data that predates the sample periods used in the original studies. The anomalies are significantly weaker during this out-of-sample period. Only 8 anomalies earn average returns that are positive and statistically significant at the 5% level. A total of 16 anomalies have either CAPM or three-factor model alphas that are statistically significant at the 5% level. Put differently, less than half of the anomalies that earn statistically significant alphas during the original sample periods do so in the pre-discovery sample.

The lack of significance is unlikely due to a lack of power. In many cases, the pre-discovery 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 36 anomalies is 12,443; the number of pre-discovery observations is 18,505; and the number of post-discovery observations is 7,299.

Among the best-performing anomalies are net working capital changes and all three distress

¹⁴We compute the point estimates and the standard errors in Table 6 as follows. We first compute each anomaly’s average return, CAPM alpha, and three-factor model alpha for the pre-discovery, in-sample, post-discovery periods. We then take the averages of these estimates either across all anomalies or by an anomaly category. These averages are the point estimates reported in Table 6. We then resample the data by drawing calendar months with replacement 10,000 times to get the bootstrapped distributions of average returns, CAPM alphas, and three-factor model alphas.

anomalies. Of the two composite anomalies, the profitability component of the quality-minus-junk factor (Asness, Frazzini, and Pedersen 2013) earns a three-factor model alpha that is statistically significant with a t -value of 2.30.

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 for the pre-discovery period. The significance of the net issuance anomaly over the modern, post-1963 sample period has been highlighted, for example, in Daniel and Titman (2006), Boudoukh, Michaely, Richardson, and Roberts (2007), Fama and French (2008), and Pontiff and Woodgate (2008). The last two of these studies, however, find no reliable evidence of this anomaly in the pre-1963 data. The estimates in Panel A of Table 7 suggest, in contrast to these null results, that the net share issues anomaly exists also in the pre-Compustat period. The reason for this difference appears to lie with the corrections to the number of shares data CRSP made in a project started in 2013.¹⁵

Table 6 shows that the average anomaly's average return, CAPM alpha, and three-factor model is 8 basis points (t -value = 2.18), 13 basis points (t -value = 4.27), and 16 basis points (t -value = 5.70), respectively, during the pre-discovery period. For some categories, the three-factor model alphas significantly differ from the average returns. In particular, the average profitability and distress anomalies earn negative average returns but, because of their negative covariances with the market and value factors, their three-factor model are positive and statistically significant with t -values of 3.54 and 4.41, respectively.

4.5 Post-discovery out-of-sample estimates

Panel B of Table 7 reports average returns and alphas for 34 out of the 36 anomalies for which we have at least five years of post-discovery data. This five-year threshold leaves out two anomalies, Operating profitability and QMJ: Profitability.

¹⁵See 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.”

The estimates for the post-discovery period are similar to those for the pre-discovery out-of-sample period. Of the 34 anomalies, only two earn average returns that are positive and statistically significant at the 5% level, and 10 earn either CAPM or three-factor model alphas that are significant at this level. Leverage is also statistically significant at the 5% level—its t -value is -3.72 —but its sign is the opposite of that for the in-sample period, and so we do not add it to the count of anomalies that “work.” Because the post-discovery period is often significantly shorter than either the in-sample or the pre-discovery period, the anomaly-level estimates are noisier than those in Table 5 and Panel A of Table 7.

Many of the anomalies that are statistically significant in the post-discovery data are not the same that are significant in the pre-discovery period. A notable exception is distress risk, which earns a monthly three-factor model alpha of 55 basis points (t -value = 3.28) after its discovery. This anomaly is therefore significant both before (t -value = 2.66) and after the original study’s sample period.

Table 6 shows that the average anomaly’s average return, CAPM alpha, and three-factor model is 10 basis points (t -value = 2.83), 14 basis points (t -value = 4.21), and 11 basis points (t -value = 3.97), respectively, during the post-discovery period. The pre-discovery and post-discovery data are in some agreement in terms of which anomaly categories reliably predict returns. Earnings quality and investment and growth categories have low three-factor model alphas both before and after their discovery, whereas profitability, financing, and distress categories earn high alphas. Over the seven anomaly categories, the correlation in three-factor model t -values, which are proportional to Sharpe ratios, is 0.54.

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

In Table 8 we compare the performance of the average anomaly between the pre-discovery, in-sample, and post-discovery periods. We measure average returns, Sharpe ratios, excess volatilities, and alphas and information ratios estimated from the CAPM and three-factor models. Excess volatility is defined, similar to Section 2.3, as an anomaly’s annualized stan-

dard deviation minus the standard deviation of its randomized version. Panel A uses the full data starting in July 1926. Panel B removes the pre-Securities and Exchange Act data and the initial two-year enforcement period (Cohen, Polk, and Vuolteenaho 2003) and starts the sample in July 1938.

The estimates show, for example, that the average anomaly earns 28 basis points (t -value = 8.67) per month during the sample period used in the original study, but just 8 basis points (t -value = 2.07) during the historical out-of-sample period and 10 basis points (t -value = 2.63) 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 -4.24 and -4.78 . These results do not materially change when we move to the CAPM or the three-factor model. The differences in CAPM alphas are -19 basis points (t -value = -4.55) and -18 basis points (t -value = -4.74). The differences in three-factor model alphas are -39 basis points (t -value = -5.76) and -40 basis points (t -value = -4.28).

The attractiveness of an anomaly as an investment depends on its volatility (or residual volatility) in addition to its alpha. Although the average anomaly earns a lower alpha out-of-sample, a simultaneous decrease in volatility could offset some of this effect. The estimates of the Sharpe and information ratios address this possibility. Sharpe ratio divides each anomaly's average return by its volatility, and information ratio divides its alpha by the standard deviation of its residuals. The Sharpe ratios and information ratios display the same pattern as average returns and alphas. They are statistically significantly higher during the in-sample period than what they are either before or after this in-sample period, and the estimates for the pre- and post-discovery samples are not statistically significantly different from each other.

The average anomaly's information ratio from the three-factor model is 0.63 during the original study's sample period, but just 0.24 for both the pre- and post-discovery periods. These differences, which are statistically significant with t -values of -5.09 and -4.26 , correspond to 61% decrease in the information when we move out-of-sample by going either backward or forward in time. The decreases in the three-factor model alphas, by contrast, are 40% (pre-discovery) and 61% (post-discovery) and so, if anything, the risk adjustment works

the “wrong way:” not only do anomalies earn high alphas during the original study’s sample period, but they are also less risky than what they are before this period.

The estimates in Panel A of Table 8 support the data-mining explanation for the return anomalies. Under this explanation, the in-sample period stands out as being special. The differences between the post- and pre-periods are generally less than one standard error away from each other; the in-sample period, by contrast, is different from both of them by at least four standard errors. These estimates show that the average returns and information ratios jump up as we enter the in-sample period and then, after this period ends, they fall back to levels that are statistically indistinguishable from those seen during the pre-sample period.

The estimates are not sensitive to removing the pre-Securities and Exchange Act data from the pre-discovery sample. Panel B of Table 8 removes the first 12 years of the data, but the differences between the in-sample period and the pre-discovery period, and those between the post-discovery and pre-discovery period, are quantitatively similar. Also in this sample the pre-period looks very similar to the post-period, and it is the in-sample period that stands out.

4.7 Changes in the correlation structure of returns

Data-mining does not necessarily work only through first moments. If the probability that an anomaly gets published depends on its statistical significance, then also the return processes’ higher moments may be distorted. An anomaly’s t -value is higher if its returns are less volatile. Researchers also often verify that a new candidate anomaly is not subsumed by known factors, such as size and value, or related anomalies. An anomaly is more likely to pass these tests if its correlation with other known anomalies are atypical. In this section, we measure the extent to which data-mining 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,

$$\begin{aligned} \text{anomaly}_{i,t} = & a + b_1 \times \text{in-sample index}_{-i,t} + b_2 \times \text{post-sample index}_{-i,t} + b_3 \times \text{post}_{i,t} \\ & + \text{post}_{i,t} \times (b_4 \times \text{in-sample index}_{-i,t} + b_5 \times \text{post-sample index}_{-i,t}) + e_{i,t}, \end{aligned} \quad (2)$$

where $\text{post}_{i,t}$ takes the value of one if anomaly i is in the post-discovery sample in month t and zero otherwise, $\text{in-sample index}_{-i,t}$ is the average return on all anomalies except anomaly i that are in-sample in month t , and $\text{post-sample index}_{-i,t}$ is the average return on all anomalies except anomaly i 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 (2) is similar to that estimated in McLean and Pontiff (2015) except that they (a) examine a different set of anomalies and (b) use the publication date as the cutoff. The interaction terms measure the changes in correlations as an anomaly moves from the in-sample period to the post-discovery sample upon its discovery.

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 0.59, but that against the post-sample index is just 0.09. However, when the anomaly moves out of sample, these correlations change to $0.59 + (-0.42) = 0.17$ and $0.09 + 0.46 = 0.54$. These changes are associated with t -values of -9.9 and 11.3 .

These estimates are consistent with arbitrageurs inducing comovement among already discovered anomalies (McLean and Pontiff 2015). If institutional investors 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 mispricings that these anomalies represent, but also by increasing their correlations.

The estimates from the second regression in Table 9, however, suggest an alternative explanation. This regression is the same as the first except that we swap the post-discovery

period for the pre-discovery period. That is, we examine how an anomaly’s correlation with other in-sample and pre-discovery anomalies changes when it “moves” from the in-sample period to the pre-discovery period. This specification is a natural counterfactual counterpart to the first regression because the threshold between the pre-discovery and the in-sample periods holds no special significance to arbitrageurs. 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 0.62, and that with the pre-discovery anomalies is 0.06. These correlations, however, change to 0.01 and 0.42 when the anomaly moves to the pre-discovery period. The differences between the in-sample and pre-discovery periods are statistically significant with t -values of -13.5 and 8.1 .

The similarity between the two regressions in Table 9 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. 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 an anomaly’s publication date. The second date holds no special significance under the data-mining explanation, but it is important if arbitrageurs learn about anomalies from published research.

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 36 accounting-based anomalies that we examine. At most 16 of these anomalies earn positive and statistically significant CAPM or three-factor model alphas in the pre-discovery period, and 10 do so in the post-discovery period. Because these anomaly-level estimates are noisy, we aggregate anomalies to get sharper estimates of the differences between the pre-discovery, in-sample, and post-discovery periods. Our results

suggest that data-snooping bias has had a significant impact on the discovery of the return anomalies. Average returns, Sharpe ratios, and alphas and information ratios from the CAPM and three-factor model all decrease sharply and statistically significantly when we move out of the original study’s sample period by moving either backward or forward in time. The information ratios from the three-factor model, for example, decrease by more than three-fifths when we move out-of-sample to either direction.

Our results suggest that many anomalies do not require explanations that invoke multi-dimensional risks or mispricing. Under the risk-based explanation, we would not expect the 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. Alphas and information ratios, however, 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 annualized three-factor model alpha of 3.1% (t -value = 11.6), our results imply that the correct standard by which to evaluate these models is by assessing their ability to explain approximately half of this alpha; depending on the asset pricing model, the out-of-sample alphas are between 40% to 60% of the in-sample alphas. Even the correct asset pricing model could not be expected to explain away the remainder, the spurious part of the 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 an 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. We lack quarterly data and some of the data items that would be needed to extend some anomalies back to 1926. Table A1 in the Appendix describes these approximations and compares the average returns and the CAPM and three-factor model alphas of the original definitions and the approximations.

A.1 Profitability

1. **Gross profitability** is defined as the revenue minus cost of goods sold, all divided by total assets: $\text{gross profitability}_t = (\text{rev}_t - \text{cogs}_t)/\text{at}_t$. 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: $\text{operating profitability}_t = (\text{rev}_t - \text{cogs}_t - \text{xsga}_t - \text{xint}_t)/\text{be}_t$. 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: $\text{return on assets}_t = \text{ib}_t/\text{at}_t$. 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: $\text{return on equity}_t = \text{ib}_t/\text{be}_t$. 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: $\text{profit margin}_t = \text{oiadp}_t/\text{rev}_t$. Soliman (2008) uses profit margin to predict returns using return data from 1984 to 2002.
6. **Change in asset turnover** is defined as the annual change in asset turnover, where asset turnover is revenue divided by total assets: $\text{change in asset turnover}_t = \Delta(\text{rev}_t/\text{at}_t)$. Soliman (2008) uses the change in asset turnover to predict returns between 1984 and 2002.

A.2 Earnings quality

7. **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.
8. **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.
9. **Net operating assets** represent the cumulative difference between operating income and free cash flow, scaled by lagged total assets, $\text{net operating assets}_t = [(\text{at}_t - \text{che}_t) - (\text{at}_t - \text{dlc}_t - \text{dltt}_t - \text{be}_t)]/\text{at}_{t-1}$. Hirshleifer, Hou, Teoh, and Zhang (2004) form trading strategies based on net operating assets using data from July 1964 through December 2002.

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

A.3 Valuation

11. **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 = be_t/mv_t . Fama and French (1992) use book-to-market ratio to predict returns using return data from July 1963 through December 1990.¹⁶
12. **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.
13. **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.
14. **Enterprise multiple** is a value measure used by practitioners: enterprise multiple_t = $(\text{mv}_t + \text{dlc}_t + \text{dltt}_t + \text{pstk}_t - \text{che}_t)/\text{oibdp}_t$, 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

¹⁶The 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 $\text{PSTK} + \text{CEQ}$. Third, if the above definitions cannot be used, but the book values of total assets (AT) and total liabilities (LT) exist, then we set the book value of equity equal to $\text{AT} - \text{LT}$. 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.

July 1963 through December 2009.

15. **Sales-to-price ratio** is defined as total sales divided by December market value of equity: $\text{sales-to-price ratio}_t = \text{rev}_t / \text{mv}_t$. 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

16. **Asset growth** is defined as the percentage change in total assets, $\text{asset growth}_t = \text{at}_t / \text{at}_{t-1} - 1$. Cooper, Gulen, and Schill (2008) examine the predictive power of asset growth using return data from July 1968 to June 2003.
17. **Growth in inventory** is defined as the change in inventory divided by the average total assets, $\text{growth in inventory}_t = \Delta \text{inv}_t / [(\text{at}_t + \text{at}_{t-1}) / 2]$. Thomas and Zhang (2002) use growth in inventory to predict stock returns using return data from 1970 to 1997.
18. **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: $\text{sales growth}_t = 5 * \text{rank}_t + 4 * \text{rank}_{t-1} + 3 * \text{rank}_{t-2} + 2 * \text{rank}_{t-3} + 1 * \text{rank}_{t-4}$. 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.
19. **Sustainable growth** is defined as the percentage change in the book value of equity, $\text{sustainable growth}_t = (\text{be}_t - \text{be}_{t-1}) / \text{be}_{t-1}$. Lockwood and Prombutr (2010) use return data from July 1964 through June 2007 to measure the predictive power of sustainable growth.
20. **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 $\text{CAPEX growth}_t =$

$(\text{capx}_t - (\text{capx}_{t-1} + \text{capx}_{t-2})/2)/[(\text{capx}_{t-1} + \text{capx}_{t-2})/2]$. 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.

21. **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.
22. **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.
23. **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.
24. **Investment-to-capital ratio** is defined as the ratio of capital expenditures to the lagged net value of plant, property, and equipment, investment-to-capital ratio $_t = \text{capx}_t/\text{ppent}_{t-1}$. Xing (2008) uses the investment-to-capital ratio to predict stock returns using return data from 1964 to 2003.
25. **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 $_t = (\Delta\text{ppent}_t + \Delta\text{inv}_t)/\text{at}_{t-1}$. Lyandres, Sun, and Zhang (2008) use the investment-to-assets ratio to predict returns from January 1970 through December 2005.

A.5 Financing

26. **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. McLean and Pontiff (2015) use Compustat variables dltis_t to measure debt issuances, and we use the same definition in the original-versus-approximation comparison in Table A1.
27. **Leverage** is defined as the ratio of long-term debt and the December book value of equity, 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.
28. **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 issuance $_t = \text{adjusted shrou}_t/\text{adjusted shrou}_{t-1}$, where adjusted shrou $_t = \text{shrou}_t * \text{cfacshr}_t$ 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 issuance $_t < 0$) and short positions in firms that issue shares (one-year share issuance $_t > 0$). Pontiff and Woodgate (2008) measure the predictive power of one-year share issuance using return data from 1932 through 2003. Because of the corrections to the historical CRSP data (see footnote 15), we set the beginning of the in-sample period as July 1968, the same as that in Daniel and Titman (2006).

29. **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.
30. **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 $\text{total external financing}_t = [(\text{adjusted shrou}_t / \text{adjusted shrou}_{t-1} - 1) * \text{me}_t + (\Delta \text{dlc}_t + \Delta \text{dltt}_t) - \text{dvc}_t] / \text{at}_t$, 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 $\text{total external financing}_t = (\text{sstk}_t - \text{prstk}_t + \text{dltis}_t - \text{dltr}_t + \text{dlcch}_t - \text{dv}_t) / \text{at}_t$, with dlcch_t set to zero when missing.

A.6 Distress

31. **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 100.¹⁷ Dichev (1998) uses Ohlson O-score to predict stock returns using return data from January 1981 through December 1995.
32. **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

¹⁷Because cpiind_t appears inside a log and because we predict the cross section of returns, this price-level adjustment washes out.

from 1946 through 1965: Altman’s Z-score $_t = 1.2 * (act_t - lct_t)/at_t + 1.4 * re_t/at_t + 3.3 * (ni_t + xint_t + txp_t)/at_t + 0.6 * me_t/lt_t + 1.0 * revt_t/at_t$, 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.

33. **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, Hilscher, and Szilagyi (2008, Table IV). The original study details the variable construction rules in Section I and the Appendix. Campbell, Hilscher, and Szilagyi (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

34. **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.
35. **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.

36. **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. and J. Chen (2007). CAPM over the long run: 1926–2001. *Journal of Empirical Finance* 14(1), 1–40.
- Ang, A., A. A. Shtaubert, 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., J. Gerakos, J. T. Linnainmaa, and V. Nikolaev (2015). Deflating profitability. *Journal of Financial Economics* 117(2), 225–248.
- 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.
- Barillas, F. and J. Shanken (2015). Which alpha? Working paper, Emory University.
- 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.
- Boudoukh, J., R. Michaely, M. Richardson, and M. R. Roberts (2007). On the importance of measuring payout yield: Implications for empirical asset pricing. *Journal of Finance* 62(2), 877–915.
- 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 (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 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.
- 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.
- 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.
- 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, 1925–2014.** 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 2014. The vertical lines indicate the dates on which the AMEX (1962) and Nasdaq (1972) stocks are added to CRSP.

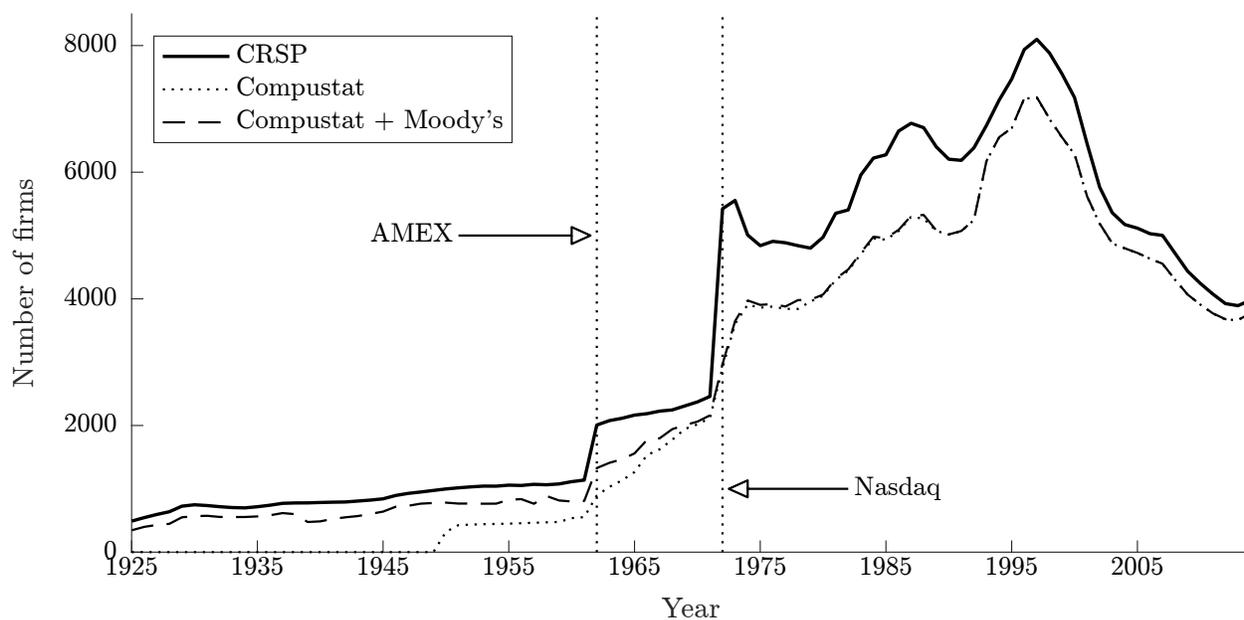


Figure 2: **Cross sections of operating profitability and asset growth, 1926–2015.** This figure displays the decile breakpoints for operating profitability (Panel A) and asset growth (Panel B) between 1926 and 2015. The thick red line corresponds to the distribution’s median. We compute the distributions at the end of June each year and use accounting data from the fiscal year that ended at least six months before. Operating profitability is the revenue minus cost of goods sold minus interest expense, all scaled by the book value of equity. Asset growth is the year-to-year percentage growth in the book value of total assets. The operating profitability and asset growth variables are those used to construct the profitability and investment factors in Panels B and C of Table 3.

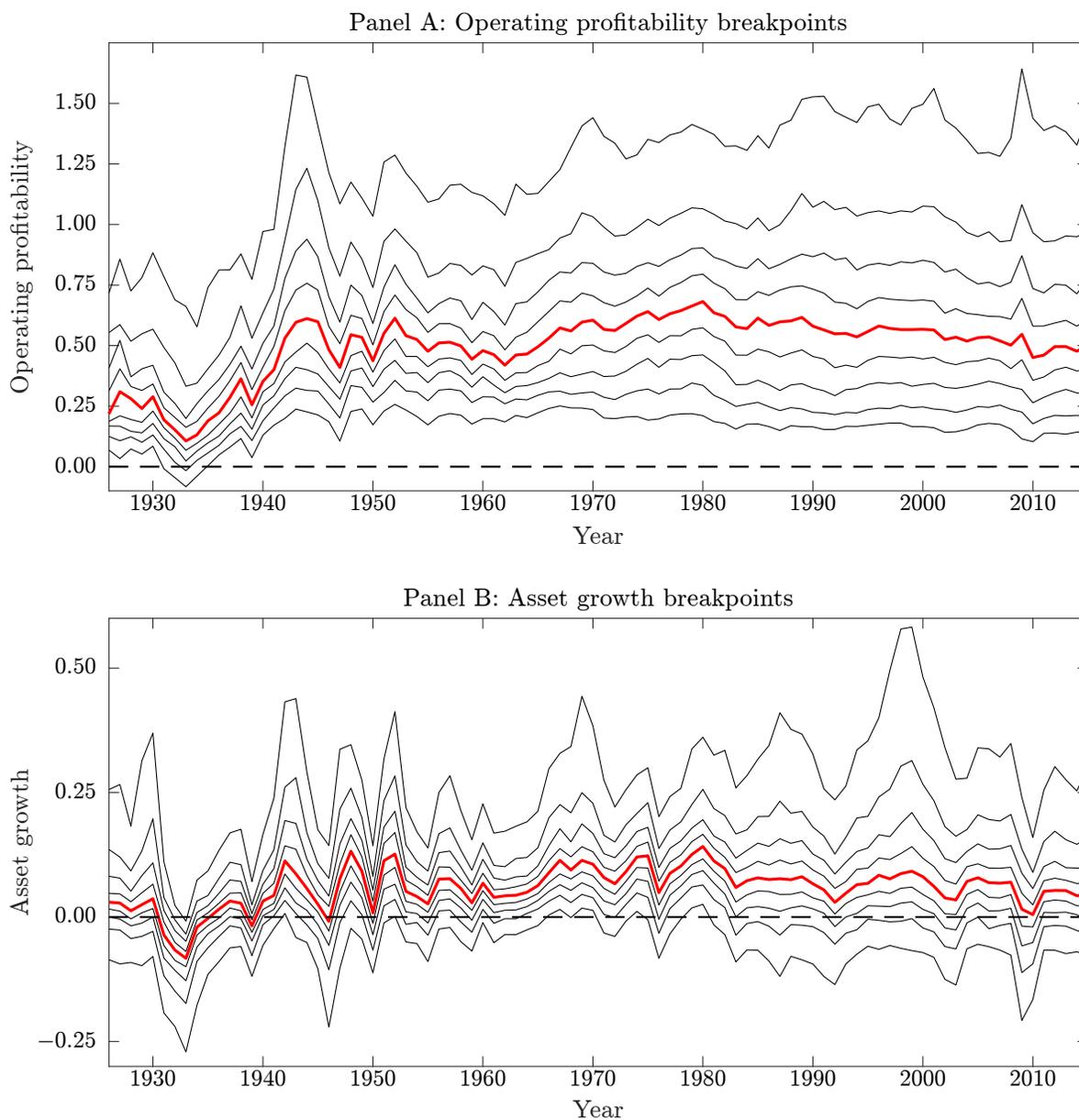
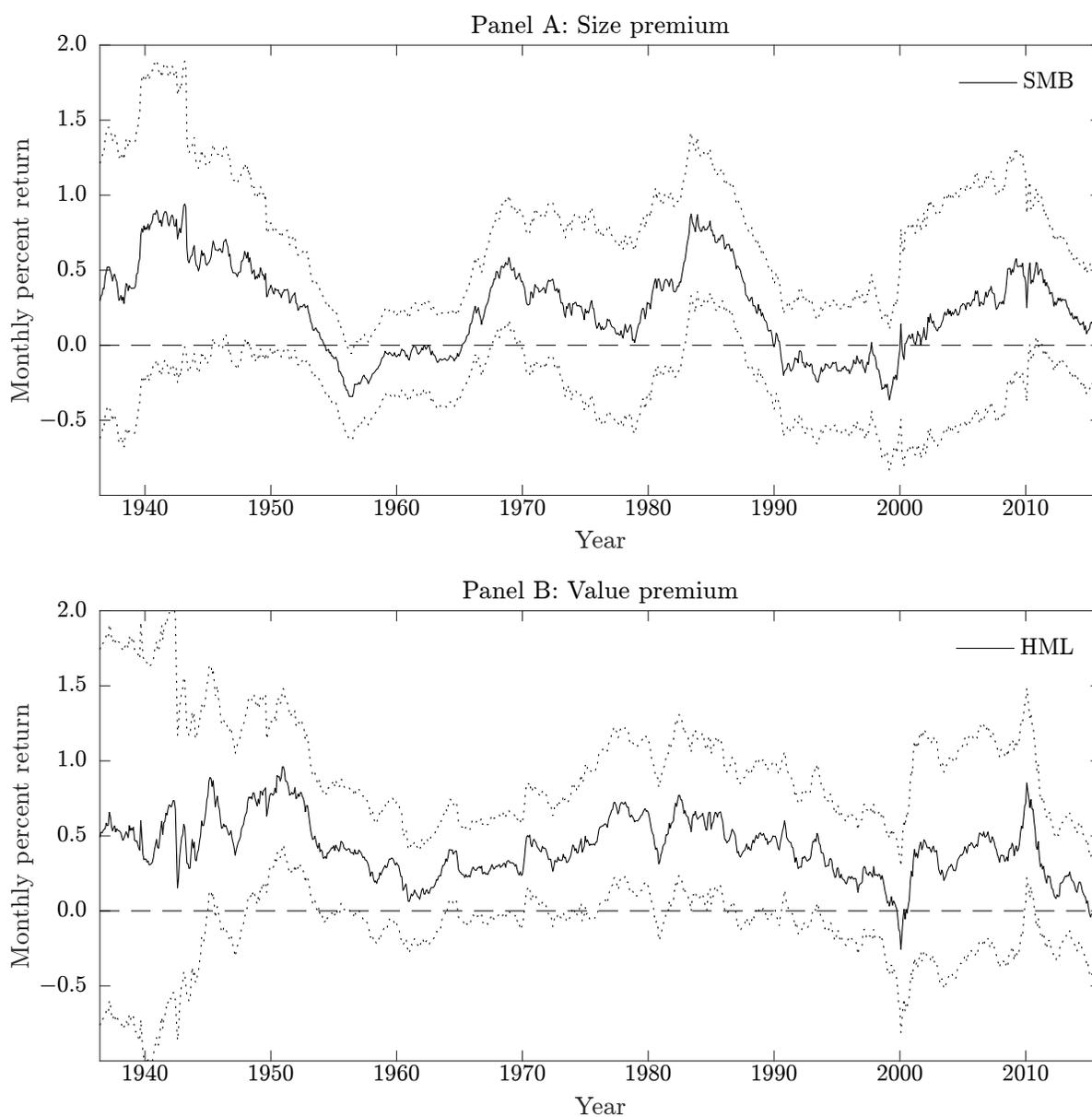
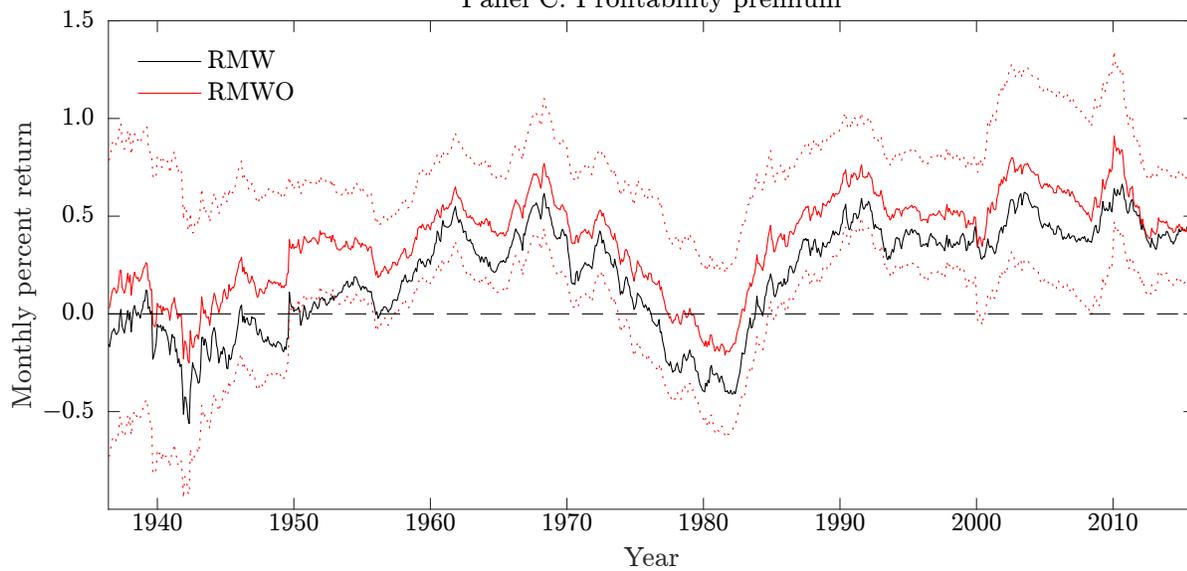


Figure 3: **Monthly percent returns on size, value, profitability, and investment factors, 1926–2015.** This figure reports rolling averages of monthly percent returns for the size (Panel A), value (Panel B), profitability (Panel C), and investment (Panel D) factors from July 1926 through December 2015. Each point represents the average return for a ten-year window up to the date indicated by the x -axis. The first point corresponds to June 1936, and it represents the average return from July 1926 through June 1936. The dotted lines denote the 95% confidence intervals. Panels C and D show average returns for the standard factors (RMW and CMA) and for the factor components that are orthogonal to the market, size, and value factors (RMWO and CMAO). A factor's orthogonal component in month t is equal to its alpha from the three-factor model regression plus the month- t residual. The confidence intervals in Panels C and D are for the orthogonal components.



Panel C: Profitability premium



Panel D: Investment premium

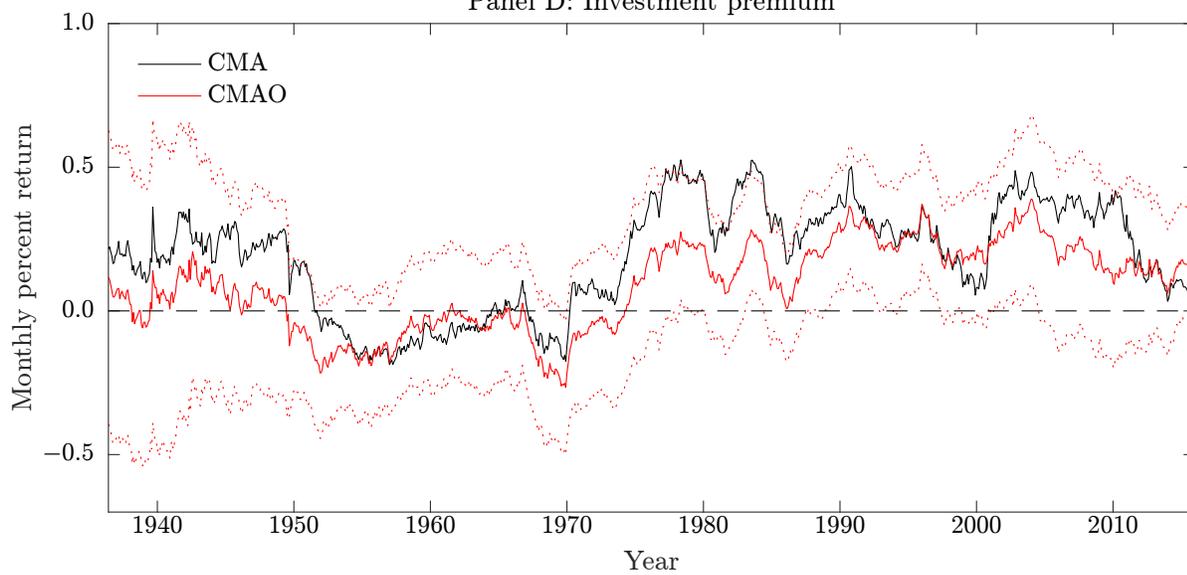


Figure 4: **Annualized Sharpe ratios for the market portfolio and an ex-post mean-variance efficient strategy for ten-year rolling windows, 1926–2015.** This figure reports Sharpe ratios for the market portfolio (thin line) and an ex-post mean-variance efficient strategy (thick line). Each point reports the annualized Sharpe ratio for a ten-year window up to the date indicated by the x -axis. The first point corresponds to June 1936, and it represents the Sharpe ratio from July 1926 through June 1936. The mean-variance efficient strategy is computed from the returns on the market, size, value, profitability, and investment factors from July 1963 through December 2015. This strategy is in-sample for the post-1963 period and out-of-sample for the pre-1963 period. The dashed segment 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.

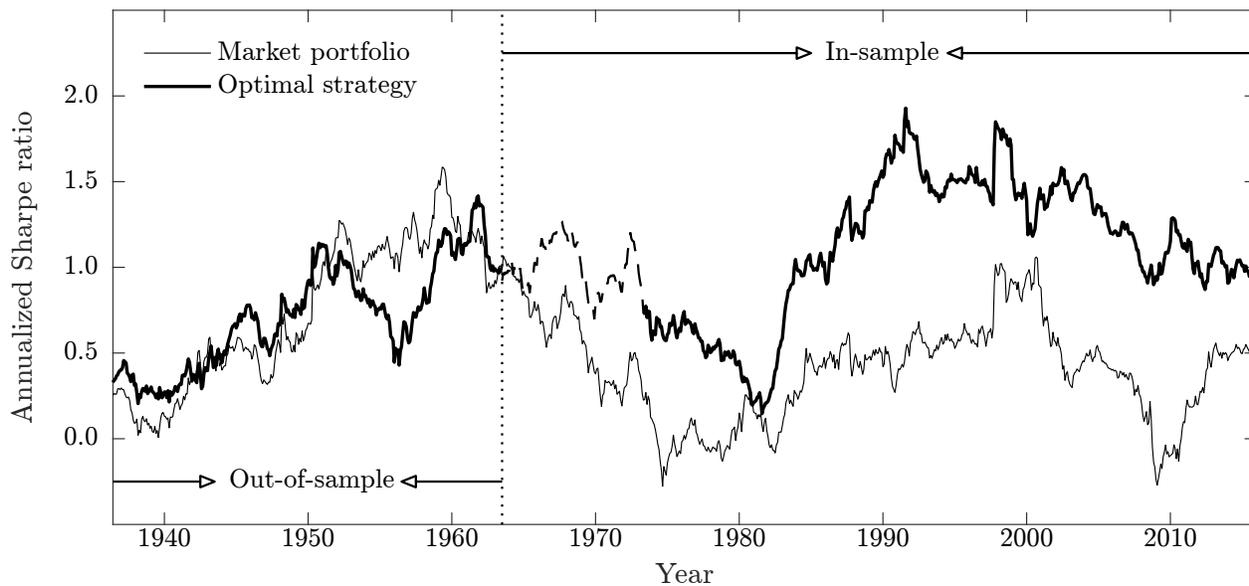


Table 1: Data coverage on CRSP, Compustat, and Moody’s Industrial and Railroad manuals, 1925–1965

This table reports the number of firms 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 these firms in either Compustat or Moody’s manuals. The bottom part of the table reports, for select data items, the number of firms covered by Compustat and the number of additional firms covered by the Moody’s manuals.

	Year								
	1925	1930	1935	1940	1945	1950	1955	1960	1965
	<u>Number of firms</u>								
CRSP	490	748	717	784	842	998	1,058	1,113	2,164
Compustat	0	0	0	0	0	320	450	537	1,265
Compustat + Moody's	345	568	564	486	640	782	828	800	1,559
	<u>Number of Compustat firms by data item</u>								
Income statement									
Revenue	0	0	0	0	0	319	450	537	1,258
Cost of goods sold	0	0	0	0	0	251	332	433	1,253
Depreciation	0	0	0	0	0	312	444	529	1,233
Interest	0	0	0	0	0	299	430	516	1,188
Net income	0	0	0	0	0	320	450	537	1,264
Balance sheet									
Total assets	0	0	0	0	0	319	449	537	1,264
Total liabilities	0	0	0	0	0	284	315	404	1,187
Common equity	0	0	0	0	0	0	0	3	702
Accounts payable	0	0	0	0	0	0	0	3	208
Receivables	0	0	0	0	0	319	383	465	1,154
Inventory	0	0	0	0	0	318	382	462	1,148
	<u>Number of additional firms from Moody's manuals by data item</u>								
Income statement									
Revenue	186	328	389	463	598	447	372	259	292
Cost of goods sold	64	187	264	416	544	398	338	245	269
Depreciation	227	465	529	479	633	442	365	257	253
Interest	152	347	316	254	345	279	259	195	195
Net income	342	550	553	485	640	458	377	264	292
Balance sheet									
Total assets	345	552	557	486	640	457	377	265	292
Total liabilities	345	568	564	486	640	465	376	264	293
Common equity	293	537	545	486	640	436	362	246	290
Accounts payable	338	547	552	483	638	437	363	245	282
Receivables	344	544	554	485	639	456	372	262	292
Inventory	340	542	549	479	632	451	371	263	286

Table 2: **Comparison to Fama-French Factors, 1963–2015**

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 2015. Panel B reports the correlations between our factors and those reported by Fama and French for the same sample period. The size (SMB) and value (HML) factors sort stocks into six portfolios by size and book-to-market at the end of each June and hold the value-weighted portfolios from the July of year t to the June of year $t + 1$. These sorts use the median NYSE breakpoint for size and the 30th and 70th percentile NYSE breakpoints for book-to-market. SMB is the average return on the three small-stock portfolios minus the average return on the three big-stock portfolios. HML is the average return on the two high-book-to-market portfolios minus the average return on the two low-book-to-market portfolios. The profitability factor (RMW) sorts stocks into six portfolios by size and operating profitability. Operating profitability is revenue minus the sum of the cost of goods sold, sales, general & administrative expenses, and interest scaled, all scaled by the book value of equity. RMW is the average return on the two high profitability (robust) portfolios minus that on the two low profitability (weak) portfolios. The investment factor (CMA) sorts stocks into six portfolios by size and growth in the book value of total assets. CMA is the average return on the two low investment (conservative) portfolios minus that on the two high investment (aggressive) portfolios.

Panel A: Monthly percent returns

	Our factors		Fama-French factors	
	Mean	t -value	Mean	t -value
Size (SMB)	0.22	1.85	0.23	1.86
Value (HML)	0.34	3.00	0.34	2.96
Profitability (RMW)	0.25	2.88	0.25	2.93
Investment (CMA)	0.25	3.38	0.30	3.77

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

Our factors	Fama-French factors			
	SMB	HML	RMW	CMA
Size (SMB)	0.998			
Value (HML)		0.994		
Profitability (RMW)			0.993	
Investment (CMA)				0.981

Table 3: Monthly percent returns and alphas on size, value, profitability, and investment portfolios and factors, 1926–2015

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). Profitability is revenue minus the sum of the cost of goods sold and interest expense, all divided by the book value of equity. The portfolios are constructed using the median NYSE breakpoint for size and the 30th and 70th percentile breakpoints for book-to-market, profitability, or asset growth. These portfolios are rebalanced annually at the end of June. Panel B reports the average number of stocks in each portfolio. Panel C reports monthly CAPM alphas for the four factors, and the three-factor model alphas for the profitability and investment factors. The pre-1963 sample period is divided into two segments, July 1926 through June 1938 and July 1938 through June 1963. In Panels A and C t -values for the average monthly percent returns and alphas are reported in parentheses.

Panel A: Monthly percent returns

Portfolio	Pre-1963 sample			
	July 1926 – June 1938	July 1938 – June 1963	July 1926 – June 1963	July 1963 – December 2015
<u>Portfolios sorted by size and book-to-market</u>				
Small Growth	0.76	1.21	1.06	0.90
Neutral	1.06	1.30	1.22	1.25
Value	1.45	1.59	1.55	1.38
Big Growth	0.81	0.99	0.93	0.88
Neutral	0.70	1.16	1.01	0.91
Value	0.94	1.50	1.32	1.07
Size factor	0.28 (0.67)	0.15 (1.07)	0.19 (1.16)	0.22 (1.85)
Value factor	0.41 (0.76)	0.44 (2.68)	0.43 (2.09)	0.33 (2.96)
<u>Portfolios sorted by size and profitability</u>				
Small Weak	0.88	1.32	1.18	0.93
Neutral	0.78	1.38	1.18	1.21
Robust	0.75	1.36	1.16	1.30
Big Weak	0.81	1.09	1.00	0.78
Neutral	0.76	1.08	0.98	0.87
Robust	0.68	1.24	1.06	1.02
Profitability factor	-0.13 (-0.31)	0.09 (0.65)	0.02 (0.14)	0.30 (3.60)
<u>Portfolios sorted by size and investment</u>				
Small Aggressive	0.84	1.24	1.11	0.99
Neutral	1.58	1.39	1.45	1.30
Conservative	1.06	1.38	1.27	1.31
Big Aggressive	0.69	1.09	0.96	0.86
Neutral	1.14	1.14	1.14	0.91
Conservative	0.86	1.04	0.98	1.03
Investment factor	0.19 (0.79)	0.04 (0.40)	0.09 (0.86)	0.25 (3.41)

Panel B: Average number of stocks in a portfolio

Portfolio	Pre-1963 sample			
	July 1926 – June 1938	July 1938 – June 1963	July 1926 – June 1963	July 1963 – December 2015
<u>Portfolios sorted by size and book-to-market</u>				
Small Growth	44.7	66.5	59.4	966.4
Neutral	111.0	179.6	157.3	1,017.0
Value	135.5	191.7	173.5	1,039.1
Big Growth	131.1	197.3	175.9	370.9
Neutral	123.7	172.3	156.5	309.4
Value	38.7	71.1	60.6	143.8
<u>Portfolios sorted by size and profitability</u>				
Small Weak	28.5	102.2	78.3	867.6
Neutral	28.8	107.1	81.7	1,190.2
Robust	25.4	80.8	62.8	943.6
Big Weak	20.7	71.2	54.8	249.3
Neutral	38.0	126.0	97.5	330.7
Robust	24.1	93.8	71.2	228.9
<u>Portfolios sorted by size and investment</u>				
Small Aggressive	46.8	87.1	74.0	943.7
Neutral	85.4	111.9	103.3	822.6
Conservative	89.0	125.6	113.8	1,014.7
Big Aggressive	86.8	108.2	101.3	287.0
Neutral	92.4	149.0	130.7	340.4
Conservative	43.4	69.0	60.7	167.7

Panel C: Monthly CAPM and three-factor model alphas

Factor	Pre-1963 sample			
	July 1926 – June 1938	July 1938 – June 1963	July 1926 – June 1963	July 1963 – December 2015
<u>CAPM alphas</u>				
SMB	0.17 (0.44)	-0.09 (-0.67)	0.02 (0.12)	0.12 (1.04)
HML	0.17 (0.40)	0.17 (1.11)	0.11 (0.61)	0.43 (3.99)
RMW	0.02 (0.06)	0.19 (1.26)	0.20 (1.34)	0.27 (3.18)
CMA	0.17 (0.71)	-0.02 (-0.19)	0.06 (0.52)	0.32 (4.70)
<u>Three-factor model alphas</u>				
RMW	0.06 (0.18)	0.30 (2.61)	0.25 (1.90)	0.35 (4.30)
CMA	0.12 (0.54)	-0.07 (-0.84)	0.02 (0.21)	0.14 (2.66)

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 approximations are described in Table A1. The bolded anomalies—operating profitability, book-to-market, and asset growth—are the profitability, value, and investment factors studied in detail in Section 3. The anomalies are described in the Appendix in Section A.

Category	No.	Anomaly	Original study	Original sample
Profitability	1	Gross profitability	Novy-Marx (2013)	1963–2010
	2	Operating profitability*	Fama and French (2015)	1963–2013
	3	Return on assets*	Haugen and Baker (1996)	1979–1993
	4	Return on equity*	Haugen and Baker (1996)	1979–1993
	5	Profit margin	Soliman (2008)	1984–2002
	6	Change in asset turnover	Soliman (2008)	1984–2002
Earnings quality	7	Accruals*	Sloan (1996)	1962–1991
	8	Earnings consistency	Alwathainani (2009)	1971–2007
	9	Net operating assets	Hirshleifer, Hou, Teoh, and Zhang (2004)	1964–2002
Valuation	10	Net working capital changes	Soliman (2008)	1984–2002
	11	Book-to-market	Fama and French (1992)	1963–1990
	12	Cash flow-to-price	Lakonishok, Shleifer, and Vishny (1994)	1968–1990
	13	Earnings-to-price	Basu (1977)	1957–1971
	14	Enterprise multiple*	Loughran and Wellman (2011)	1963–2009
Investment and growth	15	Sales-to-price	Barbee, Mukherji, and Raines (1996)	1979–1991
	16	Asset growth	Cooper, Gulen, and Schill (2008)	1968–2003
	17	Growth in inventory	Thomas and Zhang (2002)	1970–1997
	18	Sales growth	Lakonishok, Shleifer, and Vishny (1994)	1968–1990
	19	Sustainable growth	Lockwood and Prombutr (2010)	1964–2007
	20	Adjusted CAPX growth*	Abarbanell and Bushee (1998)	1974–1993
	21	Growth in sales – inventory	Abarbanell and Bushee (1998)	1974–1993
	22	Investment growth rate*	Xing (2008)	1964–2003
	23	Abnormal capital investment*	Titman, Wei, and Xie (2004)	1973–1996
	24	Investment-to-capital*	Xing (2008)	1964–2003
	25	Investment-to-assets	Lyandres, Sun, and Zhang (2008)	1970–2005
Financing	26	Debt issuance*	Spiess and Affleck-Graves (1999)	1975–1994
	27	Leverage	Bhandari (1988)	1948–1979
	28	One-year share issuance	Pontiff and Woodgate (2008)	1970–2003
	29	Five-year share issuance	Daniel and Titman (2006)	1968–2003
Distress	30	Total external financing*	Bradshaw, Richardson, and Sloan (2006)	1971–2000
	31	O-Score	Dichev (1998)	1981–1995
	32	Z-Score*	Dichev (1998)	1981–1995
Composite anomalies	33	Distress risk	Campbell, Hilscher, and Szilagyi (2008)	1963–2003
	34	Piotroski’s F-score	Piotroski (2000)	1976–1996
	35	M/B and accruals*	Bartov and Kim (2004)	1981–2000
	36	QMJ: Profitability	Asness, Frazzini, and Pedersen (2013)	1956–2012

Table 5: In-sample average returns and CAPM and three-factor model alphas for 36 anomalies

This table reports monthly average returns, CAPM alphas, and three-factor model alphas for the 36 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 use the 50 and 30/70 NYSE breakpoints. The return on the anomaly factor is the 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. The sample periods, which are the same as those used in the original studies, are reported in Table 4.

Category	Anomaly	Average return		CAPM		FF3	
		\bar{r}	$t(\bar{r})$	$\hat{\alpha}$	$t(\hat{\alpha})$	$\hat{\alpha}$	$t(\hat{\alpha})$
Profitability	Gross profitability	0.22	2.27	0.18	1.88	0.39	5.01
	Operating profitability	0.30	3.47	0.26	3.05	0.34	4.13
	Return on assets	0.07	0.52	0.07	0.52	0.44	4.50
	Return on equity	0.20	1.51	0.16	1.20	0.41	3.65
	Profit margin	0.40	2.09	0.57	3.31	0.41	3.31
Earnings quality	Change in asset turnover	0.44	4.41	0.47	4.82	0.47	4.94
	Accruals	0.23	2.76	0.29	3.66	0.21	2.77
	Earnings consistency	0.61	3.21	0.49	2.66	0.48	2.61
	Net operating assets	0.36	5.04	0.34	4.83	0.38	5.56
Valuation	Net working capital changes	0.19	2.08	0.21	2.34	0.23	2.46
	Book-to-market	0.43	3.08	0.50	3.80		
	Cash flow-to-price	0.65	4.17	0.71	4.91	0.22	3.58
	Earnings-to-price	0.47	3.19	0.58	4.17	0.35	3.82
	Enterprise multiple	0.42	4.07	0.50	5.28	0.21	3.27
	Sales-to-price	0.39	2.38	0.40	2.35	0.10	0.87
Investment and growth	Asset growth	0.38	3.97	0.46	5.36	0.22	3.36
	Growth in inventory	0.24	2.80	0.32	4.02	0.14	2.13
	Sales growth	0.17	1.58	0.22	2.10	0.01	0.07
	Sustainable growth	0.18	2.02	0.27	3.26	0.03	0.43
	Adjusted CAPX growth	0.20	3.05	0.22	3.43	0.13	2.06
	Growth in sales – inventory	0.31	4.00	0.30	3.85	0.32	3.90
	Investment growth rate	0.23	4.39	0.26	5.41	0.19	3.97
	Abnormal capital investment	0.20	3.45	0.20	3.49	0.15	2.66
	Investment-to-capital	0.16	1.38	0.31	3.20	0.05	0.74
	Investment-to-assets	0.26	3.11	0.32	4.07	0.15	2.12
Financing	Debt issuance	0.20	4.74	0.18	4.32	0.20	4.74
	Leverage	0.17	2.09	0.17	2.11	-0.01	-0.17
	One-year share issuance	0.30	3.40	0.38	4.71	0.24	3.61
	Five-year share issuance	0.23	2.82	0.29	3.94	0.18	2.88
	Total external financing	0.34	3.03	0.49	5.38	0.37	5.16
Distress	O-Score	0.21	1.77	0.25	2.12	0.44	4.15
	Z-Score	0.09	0.47	-0.08	-0.49	0.46	4.24
Composite anomalies	Distress risk	0.30	2.44	0.45	4.09	0.48	4.82
	Piotroski's F-score	0.48	5.64	0.54	6.45	0.60	7.49
	M/B and accruals	0.61	2.33	0.86	3.36	0.31	1.59
	QMJ: Profitability	0.24	3.26	0.29	3.98	0.44	7.03

Table 6: Average returns, CAPM alphas, and three-factor models for anomaly categories

This table reports monthly average returns and the CAPM and three-factor model alphas for anomalies classified by anomaly group. We estimate average returns, CAPM alphas, and three-factor model alphas for each anomaly using pre-discovery data, the data used in the original study, and post-discovery data. We report average measures by anomaly group. We compute the standard errors by block bootstrapping the data by calendar month 10,000 times. We exclude book-to-market from the list of anomalies because it is part of the three-factor model and it has been previously extended to pre-1926 data. The post-discovery sample excludes two anomalies, “Operating profitability” and “QMJ: Profitability,” with less than five years of data.

Anomaly category	Pre-discovery		Original		Post-discovery	
	EST	<i>t</i> -value	EST	<i>t</i> -value	EST	<i>t</i> -value
<u>Average returns</u>						
Profitability	-0.05	-0.41	0.27	3.89	0.06	0.62
Earnings quality	0.15	1.94	0.35	6.17	0.16	1.23
Valuation	0.17	1.59	0.48	5.10	0.23	1.74
Investment and growth	0.12	2.12	0.23	4.33	0.03	0.48
Financing	0.06	1.42	0.25	4.62	0.11	2.06
Distress	-0.02	-0.13	0.20	1.98	0.07	0.52
Composite	0.25	2.45	0.44	4.92	0.11	0.54
All except composite	0.08	2.09	0.28	8.62	0.10	2.68
<u>CAPM alphas</u>						
Profitability	0.11	1.27	0.29	4.11	0.19	2.24
Earnings quality	0.12	1.65	0.33	6.04	0.06	0.49
Valuation	0.20	2.34	0.55	6.01	0.32	2.44
Investment and growth	0.10	2.07	0.29	6.03	0.05	0.68
Financing	0.11	2.90	0.31	6.32	0.16	3.05
Distress	0.24	2.16	0.20	2.09	0.16	1.43
Composite	0.34	3.62	0.56	6.72	0.16	0.77
All except composite	0.13	4.42	0.32	10.87	0.14	4.06
<u>Three-factor model alphas</u>						
Profitability	0.21	3.34	0.41	7.42	0.17	2.33
Earnings quality	0.12	1.57	0.33	5.98	0.02	0.19
Valuation	0.18	2.28	0.22	4.51	0.14	2.12
Investment and growth	0.09	1.88	0.14	3.80	0.05	0.88
Financing	0.12	3.62	0.19	5.00	0.10	2.49
Distress	0.36	4.33	0.46	6.13	0.22	2.82
Composite	0.29	3.47	0.45	6.66	-0.10	-0.72
All except composite	0.16	5.78	0.26	11.50	0.10	3.66

Table 7: Pre- and post-discovery out-of-sample average returns and CAPM and three-factor model alphas for 36 anomalies

This table reports average returns and the CAPM and three-factor model alphas for the 36 anomalies described in Appendix A. Panel A reports estimates for the pre-discovery periods that pre-date the sample periods used in the original studies. Panel B uses data that have accumulated after the end of the original sample. We exclude two anomalies, “Operating profitability” and “QMJ: Profitability,” from Panel B because they have less than five years of post-discovery data.

Panel A: Pre-discovery estimates

Category	Anomaly	Average return		CAPM		FF3	
		\bar{r}	$t(\bar{r})$	$\hat{\alpha}$	$t(\hat{\alpha})$	$\hat{\alpha}$	$t(\hat{\alpha})$
Profitability	Gross profitability	0.00	-0.01	0.28	1.73	0.35	2.60
	Operating profitability	-0.01	-0.07	0.17	1.11	0.22	1.65
	Return on assets	-0.13	-0.94	0.06	0.48	0.23	2.64
	Return on equity	-0.08	-0.57	0.05	0.39	0.23	2.47
	Profit margin	-0.12	-1.02	0.01	0.05	0.14	1.60
	Change in asset turnover	0.07	0.84	0.09	1.14	0.11	1.44
Earnings quality	Accruals	0.10	1.01	0.10	1.00	0.07	0.79
	Earnings consistency	0.18	0.67	0.04	0.16	0.07	0.27
	Net operating assets	0.10	1.26	0.08	0.93	0.09	1.13
Valuation	Net working capital changes	0.23	3.56	0.27	4.22	0.24	3.78
	Book-to-market	0.43	2.08	0.10	0.59		
	Cash flow / price	0.14	1.17	0.12	0.98	0.10	0.91
	Earnings / price	0.11	0.54	0.28	1.47	0.33	1.95
	Enterprise multiple	0.17	1.06	0.24	1.54	0.22	1.43
	Sales / price	0.26	2.41	0.17	1.63	0.06	0.73
Investment and growth	Asset growth	0.06	0.59	0.03	0.27	-0.03	-0.30
	Growth in inventory	0.21	2.73	0.21	2.66	0.19	2.54
	Sales growth	0.08	0.52	0.02	0.12	-0.03	-0.20
	Sustainable growth	0.00	0.00	-0.10	-0.82	-0.15	-1.68
	Adjusted CAPX growth	0.14	1.76	0.06	0.80	0.03	0.40
	Growth in sales – inventory	0.08	0.66	0.14	1.20	0.21	2.02
	Investment growth rate	0.07	0.78	0.10	1.07	0.09	1.04
	Abnormal capital investment	0.10	0.92	0.12	1.09	0.15	1.43
	Investment-to-capital	0.19	2.13	0.18	2.00	0.17	1.88
	Investment-to-assets	0.25	3.15	0.25	3.07	0.22	2.91
Financing	Debt issuance	0.05	0.89	0.10	1.77	0.12	2.15
	Leverage	0.09	0.62	-0.02	-0.15	-0.04	-0.37
	One-year share issuance	0.20	2.51	0.24	2.95	0.26	3.25
	Five-year share issuance	0.08	1.15	0.15	2.16	0.18	2.68
	Total external financing	-0.14	-1.28	0.06	0.65	0.11	1.42
Distress	O-Score	0.00	0.02	0.24	2.29	0.35	4.17
	Z-Score	-0.14	-0.91	0.09	0.75	0.26	3.02
	Distress risk	0.07	0.31	0.38	1.83	0.45	2.66
Composite anomalies	Piotroski's F-score	0.06	0.61	0.14	1.34	0.18	1.87
	M/B and accruals	0.63	2.90	0.53	2.45	0.28	1.64
	QMJ: Profitability	0.04	0.18	0.36	1.72	0.42	2.30

Panel B: Post-discovery estimates

Category	Anomaly	Average return		CAPM		FF3	
		\bar{r}	$t(\bar{r})$	$\hat{\alpha}$	$t(\hat{\alpha})$	$\hat{\alpha}$	$t(\hat{\alpha})$
Profitability	Gross profitability	0.24	1.33	0.37	2.07	0.18	1.39
	Return on assets	0.19	1.21	0.36	2.60	0.42	3.34
	Return on equity	0.25	1.50	0.43	2.89	0.41	3.46
	Profit margin	-0.31	-1.84	-0.11	-0.78	-0.07	-0.61
	Change in asset turnover	-0.06	-0.67	-0.08	-0.77	-0.08	-0.88
Earnings quality	Accruals	0.15	1.69	0.13	1.38	0.13	1.44
	Earnings consistency	0.46	0.99	0.15	0.38	0.01	0.03
	Net operating assets	0.04	0.28	-0.02	-0.11	-0.03	-0.17
Valuation	Net working capital changes	0.00	0.04	-0.04	-0.39	-0.03	-0.37
	Book-to-market	0.23	1.26	0.35	2.03		
	Cash flow / price	0.24	1.25	0.42	2.38	0.17	2.00
	Earnings / price	0.36	2.56	0.52	4.08	0.20	2.55
	Enterprise multiple	-0.09	-0.57	-0.11	-0.64	0.03	0.27
Investment and growth	Sales / price	0.41	2.26	0.46	2.56	0.15	1.45
	Asset growth	0.05	0.48	0.03	0.26	0.04	0.37
	Growth in inventory	0.11	1.06	0.12	1.14	0.09	0.90
	Sales growth	-0.06	-0.50	0.06	0.56	-0.06	-0.72
	Sustainable growth	0.09	0.64	0.08	0.54	0.20	1.80
	Adjusted CAPX growth	0.02	0.21	0.07	0.93	0.03	0.42
	Growth in sales – inventory	0.11	1.29	0.11	1.24	0.13	1.51
	Investment growth rate	-0.15	-1.64	-0.12	-1.34	-0.11	-1.29
	Abnormal capital investment	0.00	0.00	0.01	0.05	0.00	-0.03
	Investment-to-capital	-0.02	-0.19	0.03	0.22	0.04	0.41
Financing	Investment-to-assets	0.17	1.15	0.11	0.76	0.17	1.25
	Debt issuance	0.13	1.55	0.08	0.96	0.14	2.43
	Leverage	0.09	0.64	0.15	1.04	-0.23	-3.72
	One-year share issuance	0.04	0.42	0.11	1.21	0.11	1.38
	Five-year share issuance	0.08	0.90	0.14	1.76	0.14	1.82
Distress	Total external financing	0.23	1.53	0.35	2.96	0.33	3.48
	O-Score	-0.08	-0.70	-0.01	-0.09	0.04	0.40
	Z-Score	-0.04	-0.23	-0.09	-0.51	0.09	0.91
Composite anomalies	Distress risk	0.33	1.27	0.59	2.86	0.55	3.28
	Piotroski's F-score	0.09	0.58	0.22	1.59	0.18	1.54
	M/B and accruals	0.13	0.36	0.10	0.29	-0.38	-1.54

Table 8: Measuring changes in average returns, alphas, and information ratios between pre-discovery, in-sample, and post-discovery periods

This table compares the performance of the average anomaly between the sample period used in the original study (original sample) and the periods that predate or follow the original sample period (pre- and post-discovery samples). Panel A uses data starting in July 1926. Panel B uses data back to July 1938. We estimate average monthly returns, variance ratios, annualized volatilities, CAPM alphas, and three-factor model alphas for 36 of the anomalies listed in the Appendix. The list of anomalies excludes book-to-market, and the post-discovery anomaly list additionally excludes the two anomalies, “Operating profitability” and “QMJ: Profitability,” with less than five years of post-discovery out-of-sample data. Variance ratio is the variance of factor returns divided by the return variance of a randomized factor. The randomized factor takes the same stocks as the actual factor and randomly sorts them into low, neutral, and high portfolios. Annualized information ratios from the CAPM and three-factor models equal $\sqrt{12} \times \hat{\alpha} / \hat{\sigma}_\varepsilon$, where $\hat{\sigma}_\varepsilon$ is the volatility of the regression residuals. We report t -values in parentheses. We compute the standard errors by block bootstrapping the data by calendar month 10,000 times.

Panel A: Full pre-1963 sample

Measure	Pre- discovery sample	Original sample	Post- discovery sample	Differences		
				Pre – Original	Post – Original	Post – Pre
Average returns						
Average return	0.08 (2.09)	0.28 (8.62)	0.10 (2.68)	–0.21 (–4.35)	–0.19 (–4.64)	0.02 (0.36)
Sharpe ratio	0.13 (2.64)	0.55 (8.74)	0.13 (1.82)	–0.43 (–5.51)	–0.42 (–4.66)	0.00 (0.01)
Excess volatility	3.69 (4.12)	3.46 (15.98)	3.64 (12.16)	0.23 (0.34)	0.21 (1.04)	–0.03 (–0.04)
CAPM						
Alpha	0.13 (4.42)	0.32 (10.87)	0.14 (4.06)	–0.19 (–4.63)	–0.18 (–4.58)	0.01 (0.21)
Information ratio	0.19 (4.11)	0.65 (10.43)	0.23 (2.87)	–0.46 (–6.03)	–0.43 (–4.52)	0.03 (0.37)
Three-factor model						
Alpha	0.16 (5.78)	0.26 (11.50)	0.10 (3.66)	–0.11 (–3.05)	–0.16 (–4.55)	–0.05 (–1.33)
Information ratio	0.24 (5.04)	0.63 (10.93)	0.24 (3.02)	–0.39 (–5.27)	–0.40 (–4.23)	–0.01 (–0.09)

Panel B: Pre-1963 sample without the pre-Securities and Exchange Act data

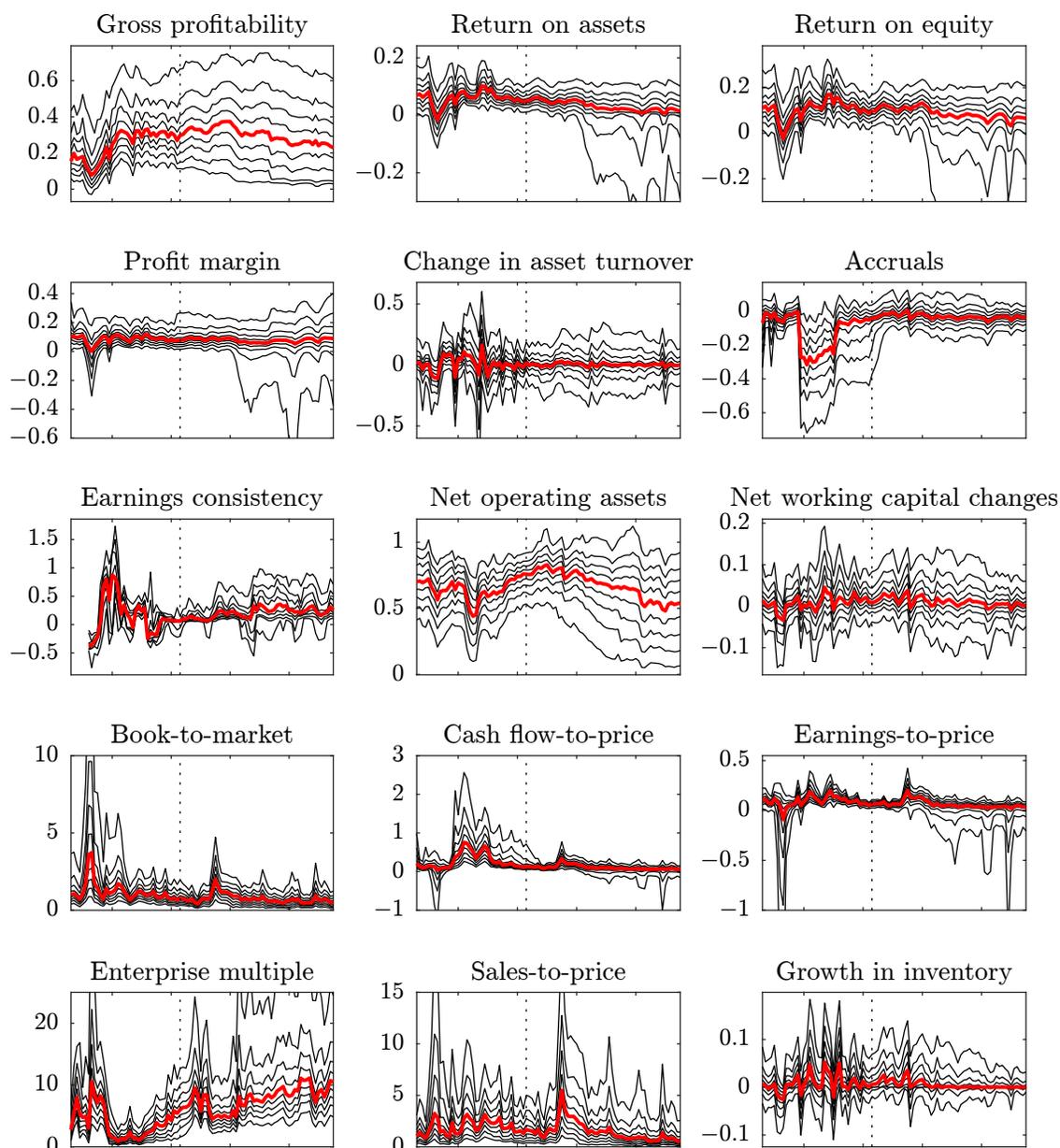
Measure	Pre- discovery sample	Original sample	Post- discovery sample	Differences		
				Pre – Original	Post – Original	Post – Pre
Average returns						
Average return	0.08 (2.81)	0.28 (8.59)	0.10 (2.66)	–0.20 (–4.51)	–0.19 (–4.64)	0.01 (0.24)
Sharpe ratio	0.16 (2.62)	0.55 (8.79)	0.13 (1.80)	–0.40 (–4.65)	–0.42 (–4.67)	–0.03 (–0.27)
Excess volatility	3.24 (2.48)	3.46 (15.94)	3.64 (12.09)	–0.22 (–0.24)	0.21 (1.03)	0.45 (0.51)
CAPM						
Alpha	0.12 (5.23)	0.32 (10.75)	0.14 (4.03)	–0.20 (–5.22)	–0.18 (–4.58)	0.02 (0.39)
Information ratio	0.23 (4.21)	0.65 (10.44)	0.23 (2.83)	–0.43 (–5.25)	–0.43 (–4.53)	0.00 (–0.02)
Three-factor model						
Alpha	0.15 (5.94)	0.26 (11.45)	0.10 (3.62)	–0.12 (–3.52)	–0.16 (–4.57)	–0.04 (–1.02)
Information ratio	0.29 (5.67)	0.63 (10.94)	0.24 (2.98)	–0.34 (–4.48)	–0.40 (–4.26)	–0.05 (–0.54)

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 returns on all other anomalies that are either in-sample or out-of-sample. In the first 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 and zero otherwise; in-sample $\text{index}_{-i,t}$ is the average return on all anomalies except anomaly i that are in-sample in month t ; and post-sample $\text{index}_{-i,t}$ is the average return on all anomalies except anomaly i that are in the post-discovery out-of-sample period in month t . In the second regression, $\text{pre}_{i,t}$ takes the value of one if anomaly i is in the pre-discovery out-of-sample period in month t and zero otherwise; in-sample $\text{index}_{-i,t}$ is the average return on all anomalies except anomaly i that are in-sample in month t ; and pre-sample $\text{index}_{-i,t}$ is the average return on all anomalies except anomaly i that are in the pre-discovery out-of-sample period in month t . The interaction terms measure the changes in correlations when an anomaly moves from being in sample to being out of sample. t -values are computed by clustering standard errors by calendar month.

Regressor	Coefficient	<i>t</i> -value
<u>Regression 1: In-sample versus post-discovery anomalies</u>		
Intercept	0.09	6.72
Main effects		
In-sample index _{-i,t}	0.59	16.13
Post-sample index _{-i,t}	0.09	8.67
Post _{i,t}	-0.10	-4.31
Interactions		
Post _{i,t} × In-sample index _{-i,t}	-0.42	-9.90
Post _{i,t} × Post-sample index _{-i,t}	0.46	11.28
Adjusted <i>R</i> ²		8.7%
<i>N</i>		15,152
<u>Regression 2: In-sample versus pre-discovery anomalies</u>		
Intercept	0.10	4.83
Main effects		
In-sample index _{-i,t}	0.62	15.16
Pre-sample index _{-i,t}	0.06	2.68
Pre _{i,t}	-0.05	-2.04
Interactions		
Pre _{i,t} × In-sample index _{-i,t}	-0.62	-13.54
Pre _{i,t} × Pre-sample index _{-i,t}	0.36	8.11
Adjusted <i>R</i> ²		4.0%
<i>N</i>		13,680

Figure A1: **Cross sections of anomaly variables, 1926–2015.** This figure displays the decile breakpoints for 28 of the anomalies described in the Appendix between 1926 and 2015. Figure 2 shows the decile breakpoints for two additional anomalies, operating profitability and asset growth, that are discussed in the main text. This figure does not include the five anomalies that are defined as discrete signals (debt issuance; market-to-book and accruals; one-year and five-year share issuances; Piotroski’s F-score) and QMJ: Profitability, which, by definition, has the same distribution in every cross section. The thick red line corresponds to the distribution’s median. We compute the distributions at the end of June each year and use accounting data from the fiscal year that ended at least six months before. The dotted vertical lines indicate the year 1963.



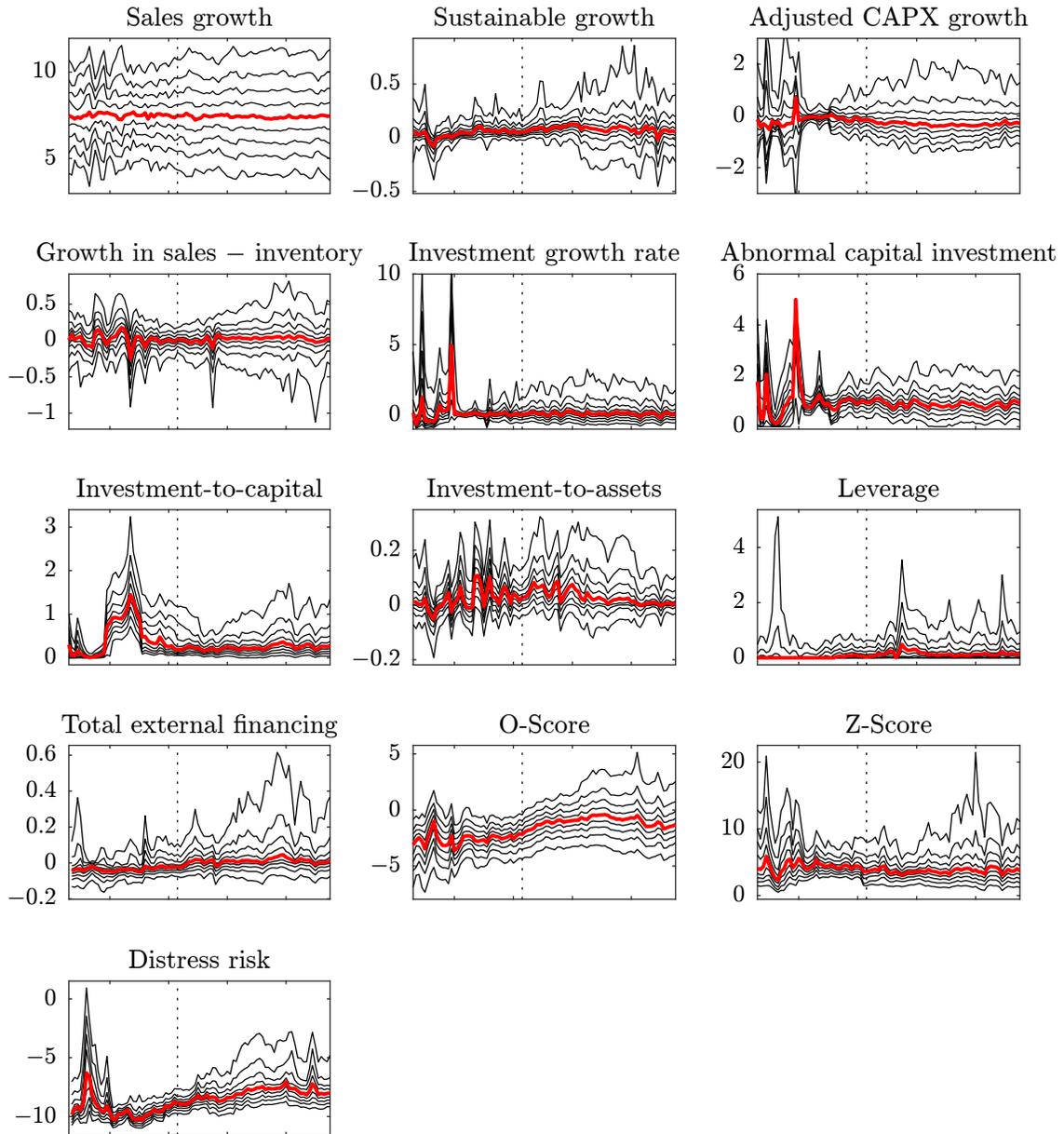


Table A1: Monthly average returns, CAPM alphas, and three-factor model alphas for anomalies and their approximations

This table shows the monthly average returns and the CAPM and three-factor model alphas for the 14 anomalies that we define differently from the original study. Moody's manuals do not contain SG&A expenses (XSGA), taxes payable (TXP), capital expenditures (CAPX), debt issuances (DLTIS), and retained earnings (RE). We leave out XSGA and TXP, approximate CAPX as the change in the plant, property, and equipment plus depreciation (or zero, whichever is greater), approximate debt issuance as the change in the outstanding short-term and long-term debt on the balance sheet, and approximate retained earnings by the sum of the retained earnings and capital surplus. We also approximate equity issuance as the number of adjusted shares outstanding times the market value of the company.

Anomaly	Approximation	Original			Approximation		
		\bar{r}	$\hat{\alpha}_{\text{capm}}$	$\hat{\alpha}_{\text{ff3}}$	\bar{r}	$\hat{\alpha}_{\text{capm}}$	$\hat{\alpha}_{\text{ff3}}$
Operating profitability	No SG&A	0.25 (2.88)	0.31 (3.56)	0.35 (4.26)	0.31 (3.52)	0.27 (3.08)	0.35 (4.13)
Return on assets	Annual data and annual rebalancing	0.51 (3.10)	0.46 (2.81)	0.87 (7.11)	0.05 (0.37)	0.05 (0.37)	0.41 (4.17)
Return on equity	Annual data and annual rebalancing	0.71 (4.50)	0.65 (4.13)	0.94 (7.05)	0.15 (1.12)	0.09 (0.67)	0.37 (3.29)
Accruals	No taxes payable	0.21 (2.46)	0.27 (3.31)	0.16 (2.26)	0.22 (2.50)	0.28 (3.35)	0.19 (2.51)
Enterprise multiple	Liquidation value of preferred stock	0.42 (4.02)	0.50 (5.28)	0.21 (3.31)	0.42 (4.04)	0.51 (5.28)	0.21 (3.31)
Adjusted CAPX growth	CAPX approximated	0.29 (4.18)	0.33 (4.79)	0.22 (3.37)	0.15 (2.39)	0.17 (2.71)	0.10 (1.59)
Investment growth rate	CAPX approximated	0.27 (4.38)	0.32 (5.46)	0.21 (3.87)	0.18 (3.60)	0.20 (4.07)	0.12 (2.62)
Abnormal capital investment	CAPX approximated	0.21 (3.24)	0.20 (3.19)	0.18 (2.76)	0.18 (3.09)	0.17 (2.89)	0.14 (2.37)
Investment-to-capital	CAPX approximated	0.14 (1.09)	0.31 (2.82)	0.00 (-0.04)	0.19 (1.56)	0.34 (3.29)	0.06 (0.79)
Debt issuance	Debt issuance from the balance sheet	0.04 (0.60)	0.03 (0.46)	0.10 (1.56)	0.15 (3.29)	0.13 (2.83)	0.17 (3.68)
Total external financing	Equity and debt issuances approximated	0.36 (3.11)	0.50 (5.01)	0.37 (4.66)	0.21 (2.19)	0.33 (3.95)	0.25 (3.60)
Altman's Z-score	No TXP; CAPS added to RE	0.08 (0.45)	-0.08 (-0.50)	0.45 (3.83)	0.10 (0.50)	-0.08 (-0.48)	0.45 (3.97)
Financial distress	Annual data	0.36 (2.25)	0.45 (2.86)	0.62 (4.15)	0.22 (1.47)	0.30 (2.14)	0.36 (2.57)
M/B and accruals	No taxes payable	0.66 (2.64)	0.90 (3.75)	0.40 (2.12)	0.69 (2.65)	0.93 (3.72)	0.41 (2.10)

Table A2: Predicting annual anomaly returns with anomaly spreads

This table reports estimates from OLS regressions that predict annual anomaly returns with anomaly spreads. Anomaly spread is computed from the same six portfolios that are used to construct the anomaly factors. In the case of book-to-market, for example, we compute average book-to-markets for the six portfolios, and define value spread as the average book-to-market of the two high portfolios minus the average book-to-market of the two low portfolios. These regressions do not include the five anomalies that are defined as discrete signals (debt issuance; market-to-book and accruals; one-year and five-year share issuances; Piotroski's F-score) or QMJ: Profitability, which, by definition, has the same distribution in every cross section. The bottom row reports the estimates from a pooled regression. This pooled regression uses a normalized spread, which is defined as the percentile rank of the spread within each anomaly. This pooled regression clusters standard errors by year.

Category	Anomaly	Intercept		Slope		Adjusted R^2
		EST	t -value	EST	t -value	
Profitability	Gross profitability	-0.12	-1.24	0.26	1.40	1.1%
	Operating profitability	-0.01	-0.20	0.02	1.18	0.4%
	Return on assets	-0.02	-0.38	0.07	0.31	-1.0%
	Return on equity	0.00	0.18	0.02	0.56	-0.8%
	Profit margin	-0.01	-0.61	0.00	0.11	-1.1%
	Change in asset turnover	0.03	1.35	-0.05	-0.86	-0.3%
Earnings quality	Accruals	0.03	2.10	-0.03	-0.76	-0.5%
	Earnings consistency	0.04	0.78	0.04	0.43	-1.1%
	Net operating assets	0.00	-0.02	0.05	0.79	-0.4%
	Net working capital changes	0.02	0.39	0.09	0.33	-1.0%
Valuation	Book-to-market	-0.01	-0.18	0.02	3.49	11.3%
	Cash flow-to-price	0.02	0.63	0.05	0.69	-0.6%
	Earnings-to-price	0.10	3.34	-0.21	-2.43	5.3%
	Enterprise multiple	0.01	0.19	0.00	1.21	0.5%
	Sales-to-price	-0.07	-2.20	0.02	4.05	14.9%
Investment and growth	Asset growth	-0.02	-0.76	0.09	2.22	4.3%
	Growth in inventory	0.01	0.44	0.14	0.45	-0.9%
	Sales growth	-0.12	-0.94	0.02	1.05	0.1%
	Sustainable growth	0.00	0.11	0.02	0.62	-0.7%
	Adjusted CAPX growth	0.02	0.68	0.00	0.05	-1.1%
	Growth in sales – inventory	0.02	1.95	0.00	0.15	-1.1%
	Investment growth rate	0.00	-0.20	0.01	1.51	1.4%
	Abnormal capital investment	0.00	-0.07	0.01	0.65	-0.7%
	Investment to capital	0.02	1.95	0.00	1.57	1.7%
	Investment-to-assets	-0.01	-0.18	0.17	1.18	0.4%
Financing	Leverage	-0.01	-0.28	0.01	1.46	1.3%
	Total external financing	-0.08	-2.64	0.22	3.21	9.6%
Distress	O-Score	-0.04	-0.54	0.01	0.52	-0.8%
	Z-Score	-0.01	-0.18	0.00	-0.16	-1.1%
	Distress risk	0.03	0.94	0.00	-0.06	-1.2%
Pooled regression		0.00	0.43	0.04	2.96	0.8%