The History of the Cross-Section of Stock Returns

Juhani T. Linnainmaa
University of Southern California and NBER

Michael R. Roberts
The Wharton School, University of Pennsylvania and NBER

Using data spanning the twentieth century, we show that the majority of accounting-based return anomalies, including investment, are most likely an artifact of data snooping. When examined out-of-sample by moving either backward or forward in time, the average returns and Sharpe ratios of most anomalies decrease, whereas their volatilities and correlations with other anomalies increase. The few anomalies that do persist out-of-sample correlate with the shift from investment in physical capital to intangible capital and the increasing reliance on debt financing over the twentieth century. (*JEL* G12, G14)

Received November 25, 2016; editorial decision September 23, 2017 by Editor Andrew Karolyi. Authors have an Internet Appendix, which is available on the Oxford University Press Web site next to the link to the final published paper online.

An "anomaly" in the context of financial economics typically refers to the rejection of an asset pricing model, such as the capital asset pricing model (CAPM) or the three-factor model of Fama and French (1993). Academic research has uncovered a large number of such anomalies—314 according to Harvey et al. (2015)—with the majority being identified during the last 15 years.

For each hypothesis test resulting in a rejection of a particular model, there is a test size quantifying the probability of a false rejection, that is, Type I error. Although the test size for any one test is low, often 5%, the test size quickly grows with the number of tests (e.g., Bickel and Doksum 1977). Multiple comparison methods exist to address this issue (e.g., White 2000), but these

We thank Jules van Binsbergen, Mike Cooper (discussant), Ken French, Chris Hrdlicka (discussant), Travis Johnson (discussant), Andrew Karolyi (editor), Mark Leary, Jon Lewellen (discussant), David McLean, Toby Moskowitz, Robert Novy-Marx, Christian Opp, Jeff Pontiff, Nick Roussanov, and Rob Stambaugh for helpful discussions and seminar and conference participants at Carnegie Mellon University, Georgia State University, Northwestern University, University of Lugano, University of Copenhagen, University of Texas at Austin, American Finance Association 2017, NBER Behavioral Finance 2017, SFS 2016 Finance Cavalcade, and Western Finance Association 2016 meetings for valuable comments. Supplementary data can be found on *The Review of Financial Studies* Web site. Send correspondence to Michael Roberts, The Wharton School, University of Pennsylvania, 3620 Locust Walk, #2319, Philadelphia, PA 19104; telephone: (215) 573-9780. E-mail: mrrobert@wharton.upenn.edu.

doi:10.1093/rfs/hhy030 Advance Access publication April 19, 2018

methods rely on the reporting of *all* tests to adjust test sizes and p-values. Because there is no mechanism to ensure authors report all hypothesis tests in published papers, let alone unpublished papers, statistical solutions have their limitations.

To understand these limitations and identify the mechanisms behind observed anomalies, we perform an out-of-sample analysis by gathering comprehensive accounting data extending back to the start of the twentieth century. More precisely, we merge data collected from Moody's manuals between 1918 and 1970 with Standard & Poor's (S&P) Compustat database and the Center for Research in Security Prices (CRSP). The result is comprehensive coverage of returns and accounting information—balance sheet and income statement data—from 1926 to 2016.

After performing a number of tests aimed at ensuring the quality of the Moody's data, we investigate the behavior of accounting-based anomalies across three eras encompassed by our data:

1. "In-sample" denotes the sample frame used in the original discovery of an anomaly (e.g., 1970 to 2004).
2. "Pre-sample" denotes the sample frame occurring prior to the in-sample period (e.g., 1926 to 1969).
3. "Post-sample" denotes the sample frame occurring after the in-sample period (e.g., 2005 to 2016).

Each anomaly we study spans these three eras, though the start and end dates for each era vary by anomaly.

The breadth of our sample frame enables us to examine competing explanations for the existence of these anomalies based on (1) unmodeled risk, (2) mispricing, and (3) data-snooping. Each of these mechanisms generate different testable implications across different time-periods encompassed by our sample.

To illustrate our empirical approach and broader findings, we first characterize the return behavior of the profitability and investment factors during the pre-sample period of 1926 to 1962. Our focus on these factors is motivated by findings in Fama and French (2015, 2017) and Hou et al. (2015) showing their importance, in combination with the market and size factors, for capturing cross-sectional variation in stock returns.

We find no economically or statistically significant premiums on the profitability and investment factors in the pre-sample period, during which the average returns are negative one (t-value = −0.06) and six (t-value = 0.64) basis points (bps) per month, respectively. The CAPM the alphas are 17 (t-value = 1.11) and three (t-value = 0.33) bps per month. The Fama and French (1993) three-factor model alphas are 22 (t-value = 1.66) and −2 (t-value = −0.25) bps per month. Only when we omit pre-1938 data from the pre-sample period do we find a statistically significant three-factor model alpha of 30 basis points (t-value = 2.60) for profitability. All other results are unchanged for this subperiod. These results for the pre-sample period stand in stark contrast to those found with in-sample data where profitability and investment factors average a statistically significant 27 and 40 bps per month, and the corresponding CAPM and three-factor alphas are similarly significant.

The investment premium's attenuation in the pre-sample data is representative of most of the other anomalies that we examine. Twenty-eight of the thirty-six anomalies we study earn average returns that are statistically insignificant during the pre-sample period. Twenty-eight CAPM alphas and twenty of the three-factor model alphas are statistically insignificant as well.

This insignificance is unlikely due to low power. For most anomalies, the pre-sample period is 37 years, typically longer than the original study's sample period. Additionally, average returns, Sharpe ratios, CAPM and three-factor model alphas and information ratios all decrease between 50% and 70%. The partial correlation between each anomaly and other anomalies increases from 0.06 to 0.56. The post-sample results are similar, consistent with the findings in McLean and Pontiff (2016).

We also find that the choice of in-sample start date is particularly important for the significance of an anomaly. By exploiting the asynchronicity of the in-sample start dates across anomalies, we show that the rise of anomalies does not coincide with a specific time period or economic event. Rather, anomalies are present only during the precise window in which they were originally discovered. Small perturbations to the in-sample start date of each anomaly lead to significant reductions, ranging from 50% to 75%, in the average return and alphas.

In total, our results show that the anomalous behavior of most anomalies is a decidedly in-sample phenomenon, consistent with data-snooping. In addition to large abnormal returns, we find low in-sample volatilities that correlate positively with those returns. Coupled with low correlations among the anomalies, t-values from multivariate regressions tend to be inflated in-sample.

In contrast, explanations based on mispricing and unmodeled risk face several challenges in reconciling our results. Most frictions responsible for limits to arbitrage (e.g., transaction costs, liquidity, and information availability) declined over the last century (French 2008; Hasbrouck 2009; Jones 2002). Even if this decline has not been constant, the nonmonotonicity of our results and variation across characteristics requires an explanation based on transient fads for those anomalies that do not persist, and seemingly permanent fads for those that do.

Perhaps more challenging is the sensitivity of anomalies to the precise sample start date. To reconcile these findings, one must rely on structural breaks in the risks that investors care about or transient fads that occur on an almost-annual basis and operate through a variety of different accounting signals.

Data-snooping does not explain the few anomalies that persist out-of-sample. While a complete investigation into the economic mechanisms behind these surviving anomalies is beyond the scope of this study, they do exhibit interesting temporal variation in light of broader economic changes. In the pre-sample period, significant anomalies are based on real investment (inventory or capital), financial distress, accounting returns (equity and assets), and equity financing. In the post-sample period, financial distress and accounting returns remain significant but real investment and equity financing do not. Instead, income statement measures (e.g., sales and earnings) and the total amount of external financing are significant. This shift coincides with the shift from tangible to intangible capital investment (Bond et al. 2000), and the growing importance of debt as a source of financing (Graham et al. 2015).

A number of studies exploit out-of-sample testing strategies to investigate anomalies. Several studies examine anomalies in domestic equity markets by exploiting sample periods prior to (e.g., Jaffe et al. 1989; Wahal 2016) or after (e.g,. Jegadeesh and Titman 2001; Schwert 2003; Chordia et al. 2014; McLean and Pontiff 2016) the original sample period used to discover the anomaly. Other studies have used international markets (Fama and French 1998), other asset classes (Asness et al. 2013a), and other securities (Barber and Lyon 1997) as out-of-sample tests. What distinguishes our study is (1) its breadth, which enables us to better gauge the extent of potential data-snooping, and (2) the combination of pre- and post-sample data, which provides for more powerful tests of competing hypotheses and a richer set of results to guide future work.

Other studies explore the statistical implications of data snooping for anomalous returns (e.g., Lo and MacKinlay 1990; Sullivan, Timmermann, and White 1999, 2001; Harvey et al. 2015; Yan and Zheng 2017). Our study complements these by providing a useful benchmark against which in-sample statistical corrections can be compared. When we apply the t-values corrected for multiple comparisons, as proposed by Harvey et al. (2015), we find 17 significant Fama-French three-factor model alphas compared to 16 found in the pre-sample period. More interesting, ten anomalies are in the intersection of these two sets of robust anomalies. Thus, while statistical corrections may provide a greater degree of reassurance that one's findings are not a statistical artifact, they cannot replace the new information obtained from out-of-sample data.

1. Data

1.1 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 2016. 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. We take delisting returns from CRSP; if a delisting return is missing and the delisting is performance-related, we impute a return of −30% for NYSE and Amex firms (Shumway 1997) and −55% for Nasdaq firms (Shumway and Warther 1999).

Second, we take annual accounting data from Standard and Poor's Compustat database. This data begins in 1947 for some firms, but become more comprehensive in 1962. Standard and Poor's established Compustat in 1962 to serve the needs of financial analysts, and backfilled information only for those firms that were deemed to be of the greatest interest to these analysts (Ball and Watts 1977). The result is significantly sparser coverage prior to 1963 for a selected sample of successful firms.

Third, we add accounting data from Moody's Industrial and Railroad manuals (Moody's). We collect information for all CRSP firms going back to 1918. These same data have previously been used in Graham et al. (2015, 2016). These data do not include financials or utilities.

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 et al. (2000) for industrial firms, but later expanded to include nonindustrial 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 conventional 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 $t-1$.

The appendix provides detailed discussions of the data and variable construction.

1.2 Firm coverage

Table 1 shows the number of firms in the CRSP database at five years intervals from 1925 through 1965. The number of CRSP firms increases over time from 490 in 1925 to 1,113 in 1960. The large jump to 2,164 firms in 1965 is due to the inclusion of AMEX-listed firms in 1962. Because Moody's does not cover financials and utilities, we exclude these firms henceforth except when mentioned otherwise. The number of nonfinancials and nonutilities on CRSP increases from 458 to 1,905 from 1925 through 1965.

The third 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 313 of the 881 NYSE nonfinancials and nonutilities. By 1965, the date by which Compustat became survivorship-bias free, the accounting data are available for 59% of firms. The fourth line

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

	Year						
	1925	1935	1945	1950	1955	1960	1965
	Number of firms						
CRSP							
Full database	490	717	842	998	1,058	1,113	2,164
Excl. financials & utilities	458	665	778	881	926	964	1,905
Compustat	0	0	0	313	376	456	1,115
Compustat + Moody's	345	564	640	776	756	721	1,406
	Number of Compustat firms by data item						
Income statement							
Revenue	0	0	0	312	376	456	1,115
Cost of goods sold	0	0	0	250	329	429	1,109
Depreciation	0	0	0	310	375	455	1,105
Interest	0	0	0	292	356	436	1,050
Net income	0	0	0	313	376	456	1,115
Balance sheet							
Total assets	0	0	0	312	375	456	1,114
Total liabilities	0	0	0	277	308	394	1,043
Common equity	0	0	0	0	0	2	561
Accounts payable	0	0	0	0	0	2	180
Receivables	0	0	0	312	375	456	1,112
Inventory	0	0	0	311	374	454	1,111
	Number of additional firms from Moody's manuals by data item						
Income statement							
Revenue	186	389	598	448	374	261	291
Cost of goods sold	64	264	544	399	340	247	268
Depreciation	227	529	633	443	367	259	252
Interest	152	316	345	280	261	197	194
Net income	342	553	640	459	379	266	291
Balance sheet							
Total assets	345	557	640	458	379	267	291
Total liabilities	345	564	640	466	377	266	292
Common equity	293	545	640	436	362	246	290
Accounts payable	338	552	638	437	363	245	282
Receivables	344	554	639	457	374	264	291
Inventory	340	549	632	452	373	265	285

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 with and without financials (SIC codes 6000–6999) and utilities (SIC codes 4900–4999); the number of these firms in Compustat; and the number of these firms in either Compustat or Moody's manuals. Except for the "full database" row, the sample excludes financials and utilities. 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.

shows that when we add Moody's to Compustat, we get a marked increase in coverage before and after 1965. At the start of CRSP in 1925, Moody's coverage is approximately 75%. In 1965, when Compustat is free of backfill bias, coverage of CRSP firms increases from 59% with Compustat alone to 74%.

Figure 1 presents a visual representation of these coverage patterns throughout our entire sample. The figure highlights the increased coverage coming with Moody's data through 1970. It also reveals consistent coverage of

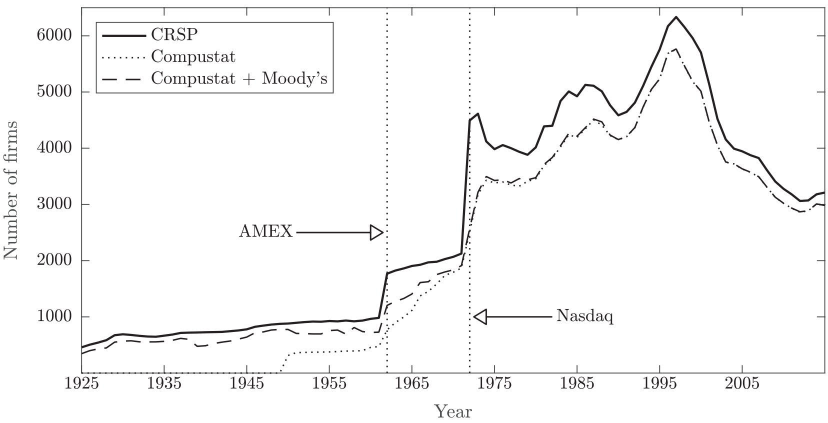

Figure 1
Number of firms in CRSP, Compustat, and Moody's Industrial and Railroad Manuals, 1925–2014
This figure shows the number of firms in the Center for Research in Securities Prices (CRSP) database; the number of these firms in Standard and Poor's Compustat database; and the number of these firms in either Compustat or Moody's Industrial and Railroad manuals between 1925 and 2014. All samples exclude financial firms (SIC codes 6000–6999) and utilities (SIC codes 4900–4999). The vertical lines represent the dates on which the AMEX (1962) and Nasdaq (1972) stocks were added to CRSP. The sample excludes financials and utilities.

CRSP, but for the years immediately after the addition of AMEX and NASDAQ firms.[1]

The lower part of Table 1 shows that coverage varies significantly by data item in the pre-1965 era. This variation in coverage reflects coarser data in the earlier parts of the sample because of the greater discretion in firms' reporting to Moody's. That said, coverage of balance sheet data is near 100% for most major accounts throughout the sample period, and income statement coverage improves quickly after 1930.

1.3 Data quality

We perform several tests to gauge the quality of the Moody's data and ensure its consistency with more recent Compustat data. These tests complement those performed by Graham et al. (2015). To conserve space, the results discussed in this section are not tabulated.

We first note that the Securities Exchange Act of 1934 was enacted in 1934 to ensure the flow of accurate and systematic accounting information. Cohen et al. (2003), discuss the Act in detail and, based on their analysis of the historical SEC enforcement records, determine that post-1936 accounting data is of sufficiently high quality to employ in empirical analysis. They characterize the first two years after the enactment of the act as an initial enforcement period, and drop these years from their sample. Although our data start in 1926 for many

[1] We exclude year 2016 from this graph because, at the time this study was undertaken, most firms' accounting information was not yet available for the fiscal year that ended in 2016.

anomalies, we repeat all of our analysis excluding pre-1936 data. We discuss some results in the body but relegate the rest to the appendix.[2]

Second, we compare the extent to which the accounting data from pre- and post-1963 conform to clean-surplus accounting. Under clean-surplus accounting, the change in assets and liabilities must pass through the income statement, implying that the change in the book value of equity equals earnings minus dividends (Ohlson 1995). Mathematically, this relation can be expressed as

$$\text{Clean-surplus ROE}_t = \log\left\{\frac{(1+R_t)\times \text{ME}_{t-1}-\text{D}_t}{\text{ME}_t}\times\frac{\text{BE}_t}{\text{BE}_{t-1}}-\left[1-\frac{\text{D}_t}{\text{BE}_{t-1}}\right]\right\}, \quad (1)$$

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.[3]

We test this relation by estimating panel regressions of clean-surplus ROE on reported ROE using annual data,

$$\text{Clean-surplus ROE}_{it} = b_1 \text{Reported ROE}_{it} + b_2 \text{Post}_t + b_3 \text{Post}_t \times \text{Reported ROE}_{it} + \mu_i + \varepsilon_{it}, \quad (2)$$

where μ_i are firm fixed effects and Post_t identifies post-1963 data. This regression measures how closely *clean-surplus ROE* tracks reported ROE before and after 1963.

Under the null that firms adhere to clean-surplus accounting, the slope coefficient on reported profitability is equal to one. We note that a test of conformity to clean-surplus accounting is a joint test of two hypotheses: (a) errors in Moody's manuals and (b) firms' tendencies to circumvent the income statement. Even under the generally accepted accounting principles (GAAP), some transactions can circumvent the income statement and affect the book value of equity directly.[4] So real-world income statement and balance sheet information rarely line up exactly as they should under this ideal.

The results reveal that the pre-1963 slope on Reported ROE is 1.04 (SE = 0.06). The interaction between Post_t and Reported ROE is −0.32 (SE = 0.06),

[2] Cohen et al. (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 et al. (2003) is such that their year 1938 observations use book values from 1937. In our subsample analysis, we start the return data in July 1938 so that, consistent with Cohen et al. (2003), the book values of equity come from 1937.

[3] 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 $t-1$. See, for example, Vuolteenaho (2002), Cohen et al. (2003), and Nagel (2005).

[4] One such example is foreign currency translations. See endnote 1 in Ohlson (1995) for others.

implying that the post-1963 slope on reported ROE is 0.71. (The standard errors (SE) are clustered by year.) The difference in slopes shows that clean-surplus accounting is violated more frequently in the more recent data, most likely due to an increase in transactions that circumvent this relation. Foreign transactions by U.S. firms have increased over the last 100 years and, as noted above, this is one type of transaction that will break the clean-surplus relation. However, this is not an indictment of data quality, only a change in corporate behavior over the last 100 years. The unit coefficient on the pre-1963 data is comforting.

Finally, we test the relative volatility of anomalies in the pre-1963 era against that in the post-1963 era. More precisely, we separately estimate in each era the return volatility for each anomaly, which is constructed as an "HML-like" factor discussed below. Because return volatility may vary over time, we use as a control group the return volatility from randomly constructed factors in which we randomly assign the same number of firms to the low and high portfolios.

The average anomaly factor's annualized return volatility is 9.7% in the pre-1963 data. The volatility of the average random factor factor is 5.9%, implying that the excess volatility is 3.8% (SE = 0.6%). In the post-1963 data, the excess volatility is $7.2\% - 3.3\% = 3.9\%$ (SE = 0.2%). The -0.1% difference between the pre- and post-1963 periods is statistically insignificant with a t-value of -0.3.[5] The comparable amounts of excess volatility across the eras suggests that the historical accounting data measure differences in firm fundamentals to the same extent as they measure them in the post-1963 data.

2. 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, 2017) and Hou et al. (2015) add 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 3 in which we analyze returns on a total of 36 anomalies.

2.1 Defining factors

Following Fama and French (2015), we define profitability as revenue less cost of goods sold (COGS); selling, general, and administrative expenses (SG&A); and interest expense scaled by book equity. Investment is defined as the change in total assets over the year scaled by the start of year assets. We construct HML-like profitability and investment factors by sorting stocks into six portfolios by

[5] We 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 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—5.9% versus 3.3%—because of the smaller number of stocks.

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

	Investment		
Size	Low	Neutral	High
Small	Small-Conservative	Small-Neutral	Small-Aggressive
Big	Big-Conservative	Big-Neutral	Big-Aggressive

The breakpoint for size is the NYSE median and the breakpoints for investment are the 30th and 70th NYSE percentiles. Holding these portfolios fixed from July of year t to the end of June of year $t+1$, we compute the value-weighted monthly returns to each of the six portfolios. 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. Profitability, called RMW for "robust minus weak" in Fama and French (2015), is computed in a similar manner as the average return on the two high-profitability portfolios minus the average return on the two low-profitability portfolios. Size (SMB) and value (HML) factors are defined as in 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 2016. We add financials and utilities back into the sample for this table to make the numbers comparable. In panel A, we report average monthly percentage returns for these factors as well as corresponding t-statistics for the null hypothesis that the mean is equal to zero. The average returns on these factors are nearly identical but for investment that reveals a six-basis point difference. Panel B shows that the correlations between our factors and the Fama-French factors are nearly perfect. The lowest correlation, which is between the two investment factors, is 0.98.[6]

2.2 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 2016. 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

[6] The small discrepancies between our numbers and those in Fama and French (2015) are due to the more inclusive Compustat-CRSP mapping used in Fama and French (2015) relative to that provided by CRSP. Additionally, Fama and French bring their data to the permanent company (permco) level, as opposed to the permanent number (permno) level.

Table 2
Comparison to Fama-French Factors, 1963–2016

Panel A: Monthly percent returns

	Our factors		Fama-French factors	
	Mean	t-value	Mean	t-value
Size (SMB)	0.23	1.92	0.23	1.86
Value (HML)	0.36	3.23	0.37	3.35
Profitability (RMW)	0.25	2.94	0.24	2.74
Investment (CMA)	0.25	3.49	0.31	3.91

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

	Fama-French factors			
Our factors	SMB	HML	RMW	CMA
Size (SMB)	0.996			
Value (HML)		0.982		
Profitability (RMW)			0.982	
Investment (CMA)				0.980

Panel A shows the average monthly percentage returns and the associated t-values for the size, value, profitability, and investment factors between July 1963 through December 2016. Panel B reports the correlations between our factors and those reported by Fama and French for the same sample period. The sample in this table includes financials (SIC codes 6000–6999) and utilities (SIC codes 4900–4999). 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 the 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 by operating profitability. Operating profitability is revenue minus the sum of the cost of goods sold; sales, general, and 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 by 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.

had been in effect for two years by the time the late part begins (Cohen et al. 2003).

Although the value premium is significant over the 1926–1963 period—the estimated monthly premium is 0.45% with a t-value of 2.06—the premiums associated with the size, profitability, and investment factors are statistically and economically insignificant. The average return on the size factor is 0.18% (t-value = 1.11), and those on the profitability and investment factors are 0.02% (t-value = 0.14) and 0.09% (t-value = 0.80).[7] The average returns on the portfolios used to construct the profitability and the investment factors show that these insignificant estimates are not confined to either big or small stocks.

[7] In 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 1926–2016 sample. This alternative profitability factor is superior to the original factor in the post-1963 sample—its t-value of 3.09 exceeds the t-value of 2.94 on the with-SG&A version—and so this change does not handicap the factor. This performance improvement is consistent with Ball et al. (2016).

Table 3
Monthly percent returns and alphas on size, value, profitability, and investment portfolios and factors, 1926–2016

Panel A: Monthly percent returns

	Pre-1963 sample			
Portfolio	1926:7 –1938:6	1938:7 –1963:6	1926:7 –1963:6	1963:7 –2016:12
	Portfolios sorted by size and book-to-market			
Small Growth	0.73	1.17	1.03	0.88
Neutral	1.04	1.32	1.23	1.29
Value	1.41	1.62	1.55	1.40
Big Growth	0.80	0.99	0.93	0.88
Neutral	0.73	1.16	1.02	0.96
Value	0.88	1.51	1.31	1.13
Size factor	0.26	0.15	0.18	0.20
	(0.62)	(1.05)	(1.11)	(1.54)
Value factor	0.38	0.48	0.45	0.38
	(0.66)	(2.87)	(2.06)	(3.40)
	Portfolios sorted by size and profitability			
Small Weak	0.88	1.32	1.18	0.96
Neutral	0.78	1.38	1.19	1.24
Robust	0.75	1.37	1.17	1.31
Big Weak	0.81	1.09	1.00	0.83
Neutral	0.76	1.10	0.99	0.90
Robust	0.68	1.23	1.05	1.03
Profitability factor	−0.13	0.09	0.02	0.28
	(−0.31)	(0.64)	(0.14)	(3.09)
	Portfolios sorted by size and investment			
Small Aggressive	0.84	1.23	1.10	0.98
Neutral	1.58	1.36	1.43	1.34
Conservative	1.06	1.36	1.27	1.33
Big Aggressive	0.69	1.11	0.97	0.90
Neutral	1.14	1.12	1.13	0.94
Conservative	0.86	1.04	0.98	1.08
Investment factor	0.19	0.03	0.09	0.26
	(0.79)	(0.32)	(0.80)	(3.28)

(continued)

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.

2.3 Cross-sections of profitability and investment

Figure 2 shows how the cross-sections of profitability and investment evolve between 1926 and 2016 by plotting these variables' decile breakpoints. Clear from the figures is time variation in both distributions. Panel A shows a widening of the profitability distribution over time. The Great Depression, World War II,

Table 3
Continued

Panel B: Average number of stocks in a portfolio

Portfolio	Pre-1963 sample 1926:7 –1938:6	1938:7 –1963:6	1926:7 –1963:6	1963:7 –2016:12
	Portfolios sorted by size and book-to-market			
Small Growth	41.5	64.3	56.9	783.4
Neutral	105.7	162.3	143.9	803.9
Value	127.5	170.0	156.2	861.5
Big Growth	124.5	174.3	158.1	319.3
Neutral	115.6	156.4	143.2	242.9
Value	36.7	68.0	57.9	99.7
	Portfolios sorted by size and profitability			
Small Weak	28.5	102.1	78.3	853.9
Neutral	28.8	106.1	81.1	850.4
Robust	25.4	80.1	62.4	736.5
Big Weak	20.7	70.2	54.2	199.3
Neutral	38.0	125.5	97.1	264.7
Robust	24.1	93.5	71.0	190.2
	Portfolios sorted by size and investment			
Small Aggressive	46.8	84.0	71.9	772.2
Neutral	85.4	108.5	101.0	642.5
Conservative	89.0	117.8	108.5	849.3
Big Aggressive	86.8	102.8	97.6	242.6
Neutral	92.4	140.9	125.2	268.3
Conservative	43.4	68.2	60.2	131.0

Panel C: Monthly CAPM and three-factor model alphas

Factor	Pre-1963 sample 1926:7 –1938:6	1938:7 –1963:6	1926:7 –1963:6	1963:7 –2016:12
	CAPM alphas			
SMB	0.15	−0.09	0.01	0.08
	(0.39)	(0.68)	(0.07)	(0.62)
HML	0.12	0.23	0.11	0.44
	(0.27)	(1.42)	(0.59)	(3.99)
RMW	0.02	0.19	0.20	0.30
	(0.06)	(1.25)	(1.33)	(3.33)
CMA	0.17	−0.02	0.05	0.33
	(0.71)	(−0.19)	(0.49)	(4.39)
	Three-factor model alphas			
RMW	0.06	0.30	0.25	0.35
	(0.18)	(2.60)	(1.90)	(3.81)
CMA	0.12	−0.07	0.02	0.11
	(0.54)	(−0.82)	(0.16)	(1.96)

Panel A reports average monthly percentage returns for the size, value, profitability, and investment factors, and for the value-weighted portfolios 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 percentage returns and alphas are reported in parentheses.

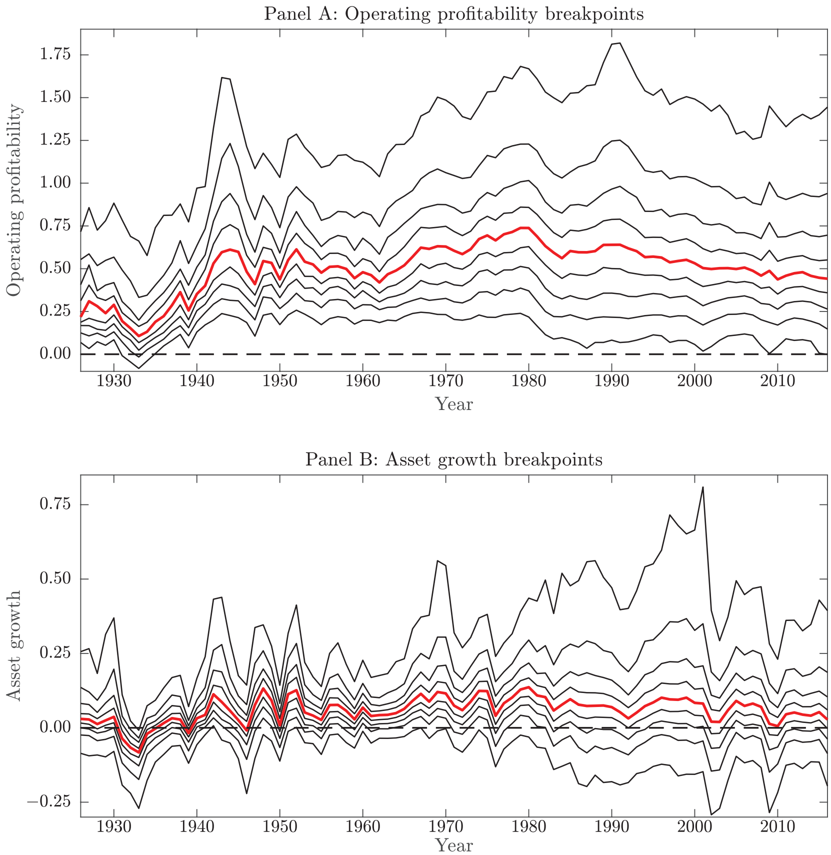

Figure 2
Cross-sections of operating profitability and asset growth, 1926–2016
This figure displays the decile breakpoints for operating profitability (panel A) and asset growth (panel B) between 1926 and 2016. The thick red line represents the distribution's median. We compute the distributions at the 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.

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, are more pronounced.

2.4 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 and represent mean-variance spanning tests. 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 2017).

All four CAPM alphas are 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 (panel B) significantly differs from those of the other premiums. Whereas the value premium is positive almost throughout the full sample period except for the interruption toward the end of the 1990s during the peak of the tech bubble, the other premiums are less stable.

The size factor (panel A) performs poorly in the 1950s and 1960s, and then again in the 1990s. It is too volatile to attain but fleeting periods of statistical significance. The investment premium (panel D) is positive until 1950 after which point it turns and remains negative until the mid-1970s. The profitability premium (panel C) 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 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.

2.5 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 2016. We report the Sharpe ratios for rolling ten-year windows.

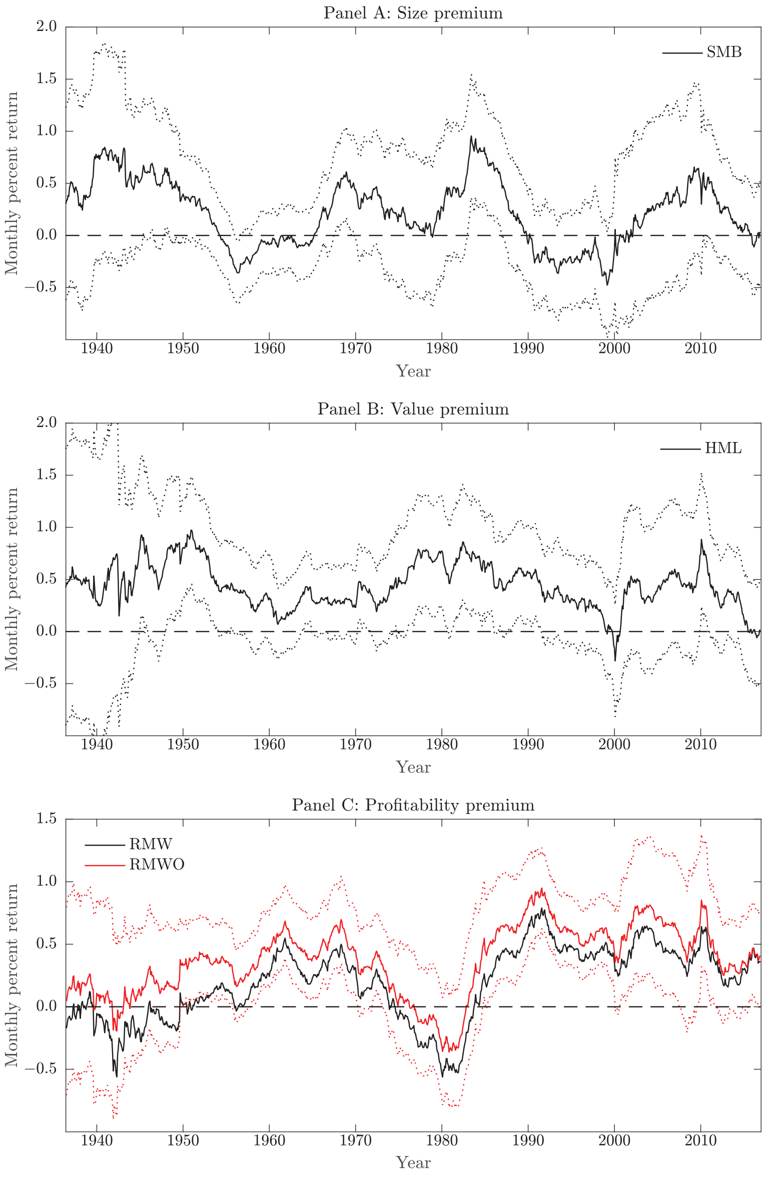

Figure 3
Monthly percentage returns on size, value, profitability, and investment factors, 1926–2016
This figure reports rolling averages of monthly percentage returns for the size (panel A), value (panel B), profitability (panel C), and investment (panel D) factors from July 1926 through December 2016. 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 represents the average return from July 1926 through June 1936. The dotted lines denote the 95% confidence intervals. Panels C and D show the average returns for the standard factors (RMW and CMA) and for the factor components 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.

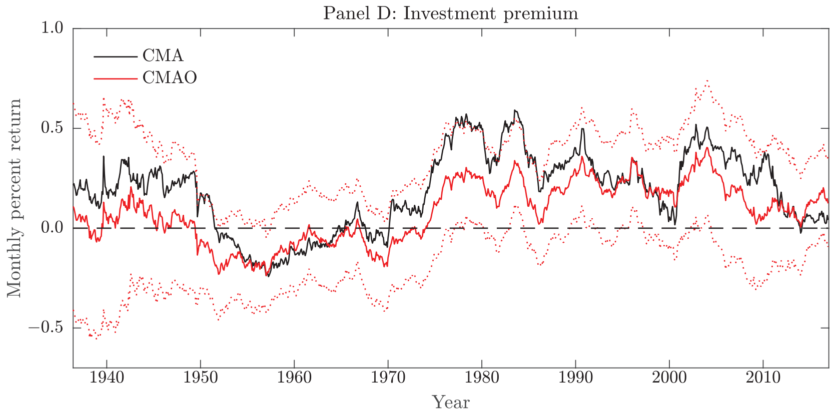

Figure 3
Continued

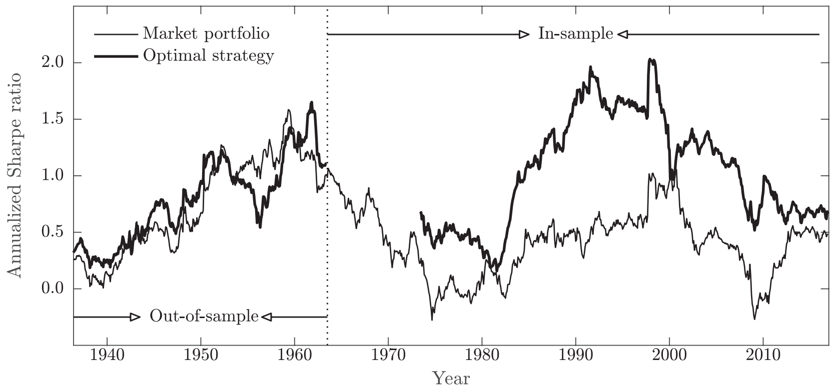

Figure 4
Annualized Sharpe ratios for the market portfolio and an ex-post mean-variance efficient strategy for ten-year rolling windows, 1926–2016
This figure reports Sharpe ratios for the market portfolio (thin line) and an ex post mean-variance efficient strategy (thick line). Each point represents 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 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 2016. This strategy is in-sample for the post-1963 period and out-of-sample for the pre-1963 period. The missing 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.

The market's Sharpe ratio for the entire 1926 through 2016 period is 0.42. It is slightly higher (0.46) for the pre-1963 sample than for the post-1963 sample (0.40). The optimal strategy's Sharpe ratio for the post-1963 sample period is 0.97; by construction, this strategy is in-sample for this period. However, for the pre-1963 sample, the Sharpe ratio of this strategy is just 0.55, 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 poorly relative to the market in the 1950s.

This computation illustrates that one's view of what matters in the cross-section of stocks depends critically on where one looks. The cross-section of stock returns is not immutable, especially with respect to the profitability and investment factors. Figure 4 shows that the optimal strategy (ex post) for 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.

3. Anomaly Performance: The Rest of the "Zoo"

Cochrane (2011) describes the large number of anomalies as a "zoo." This section examines the performance of 36 anomalies, the maximum number possible given the limitations of our data. Before doing so, we motivate our analysis with a discussion of the competing hypotheses and empirical challenges.

3.1 Competing explanations for cross-sectional return anomalies

The first hypothesis—unmodeled risk—asserts that cross-sectional return anomalies come about because stock risks are multidimensional and previous empirical attempts to reduce that dimensionality lead to model misspecification. For example, 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. Arguments similar in spirit can be made for other return anomalies.

The empirical implication of this hypothesis is that the choice of sample period should be irrelevant for the significance of an anomaly, absent a structural break in the risks that matter to investors. Figure 2 clearly shows that the cross-sections of corporate characteristics have undergone fairly significant changes during the last century.[8] Thus, any change in the significance of anomalies may simply represent a shift in the relevance of the underlying risk driving that anomaly.

[8] In the Internet Appendix, we expand on Figure 2 and plot the breakpoints for all anomaly variables.

The second mechanism—mispricing—asserts that investor irrationality combined with limits to arbitrage causes asset prices to deviate from fundamentals. Lakonishok et al. (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.

Limits to arbitrage (Shleifer and Vishny 1997) enable mispricing to persist. The market frictions responsible for these limits have arguably ameliorated over the last century. Trading costs were almost twice as high in the 1920s than in the 1960s (Hasbrouck 2009, Figure 3), and information availability and the computing power to process that information have increased dramatically. Consequently, arbitrageurs' ability to attack mispricing has improved over time (French 2008).

Like unmodeled risk, mispricing may be dynamic. Any fads or sentiment giving rise to an anomaly could be transient. This transiency poses an identification challenge similar to that posed by structural breaks. We address both identification threats in our empirical analysis.

The third mechanism—data snooping—suggests that anomalies are an artifact of chance error. All hypothesis tests come with a probability of Type I error governed by the size of the test, typically 5%. Consequently, if one performs enough hypothesis tests without appropriately adjusting for the composite nature of the tests, 5% will be significant solely due to chance error (e.g., White 2000). Thus, the data-snooping hypothesis implies that the in-sample performance of anomalies is unique: a lucky draw driven by sample error as opposed to economics.

It is worth emphasizing that data-snooping works through all relevant sample moments. To be considered an anomaly, a factor must have a sufficiently large t-statistic, for example, greater than 1.96. Thus, a large average return is insufficient to be considered an anomaly. Rather, the average return must be large relative to its variation (i.e., standard error). Related, the return must have sufficient independent variation to not be subsumed by existing factors in a multivariate regression setting.

While statistical adjustments to t-statistics do exist (see Harvey et al. (2015) for a discussion), they have their limitations. Researchers need not report all of their tests in published work, and anomaly tests in unpublished work go unreported. With our out-of-sample data, we are able to mitigate these concerns while investigating the performance of in-sample adjustments.

3.2 Defining anomalies

Table 4 lists the 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 (2016). We add to their list a few anomalies that have

Table 4
Defining return anomalies

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	Net operating assets	Hirshleifer, Hou, Teoh, and Zhang (2004)	1964–2002
	9	Net working capital changes	Soliman (2008)	1984–2002
Valuation	10	**Book-to-market**	Fama and French (1992)	1963–1990
	11	Cash flow / price	Lakonishok, Shleifer, and Vishny (1994)	1968–1990
	12	Earnings / price	Basu (1977)	1957–1971
	13	Enterprise multiple*	Loughran and Wellman (2011)	1963–2009
	14	Sales / price	Barbee, Mukherji, and Raines (1996)	1979–1991
Investment and growth	15	**Asset growth**	Cooper, Gulen, and Schill (2008)	1968–2003
	16	Growth in inventory	Thomas and Zhang (2002)	1970–1997
	17	Sales growth	Lakonishok, Shleifer, and Vishny (1994)	1968–1990
	18	Sustainable growth	Lockwood and Prombutr (2010)	1964–2007
	19	Adjusted CAPX growth*	Abarbanell and Bushee (1998)	1974–1993
	20	Growth in sales − inventory	Abarbanell and Bushee (1998)	1974–1993
	21	Investment growth rate*	Xing (2008)	1964–2003
	22	Abnormal capital investment*	Titman, Wei, and Xie (2004)	1973–1996
	23	Investment to capital*	Xing (2008)	1964–2003
	24	Investment-to-assets	Lyandres, Sun, and Zhang (2008)	1970–2005
Financing	25	Debt issuance*	Spiess and Affleck-Graves (1999)	1975–1994
	26	Leverage	Bhandari (1988)	1948–1979
	27	One-year share issuance	Pontiff and Woodgate (2008)	1970–2003
	28	Five-year share issuance	Daniel and Titman (2006)	1968–2003
	29	Total external financing*	Bradshaw, Richardson, and Sloan (2006)	1971–2000
Distress	30	O-score	Dichev (1998)	1981–1995
	31	z-score*	Dichev (1998)	1981–1995
	32	Distress risk	Campbell, Hilscher, and Szilagyi (2008)	1963–2003
Other	33	Industry concentration	Hou and Robinson (2006)	1951–2001
Composite anomalies	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

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 of returns 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 the Internet Appendix. The bold anomalies—operating profitability, book-to-market, and asset growth—are the profitability, value, and investment factors studied in detail in Section 2. The appendix describes these anomalies.

been documented after that study. The appendix describes each anomaly in detail.

For ease of reference, we group anomalies into eight categories: profitability, earnings quality, valuation, investment and growth, financing, distress, other, and composite anomalies. In our classification, composite anomalies, such as Piotroski's (2000) F-score, are anomalies that combine multiple anomalies into one. To our knowledge, our list of anomalies is comprehensive given our data limitations.

We also examine "price-based" anomalies, such as short-term reversals and medium-term momentum, over our sample period. Because their pre-1963 performance has either already been documented or could have been documented given existing data availability, these results are presented in our Internet Appendix for completeness. In contrast, the anomalies on which we focus could not have been investigated prior to their availability in the Compustat database.

We use the same definitions for all 36 anomalies—that is, value, profitability, investment, and the 33 additional anomalies—throughout the 1926–2016 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. In the Internet Appendix, we describe these approximations and compare the average returns and the CAPM and three-factor model alphas of the original definitions and the approximations.

We construct HML-like factors for all of the anomalies in the same manner as was done for profitability and investment above. The exceptions are the debt and net issuance anomalies. The debt issuance anomaly (Spiess and Affleck-Graves 1999) 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.

3.3 Anomaly performance by sample period

3.3.1 Individual anomalies. Table 5 presents anomaly performance results. Specifically, the average monthly percentage returns and the CAPM and three-factor model alphas for each of the HML-like anomaly returns are presented for each of the three eras. Recall that these returns are based on two-way sorts: size and the anomaly.[9] The side-by-side presentation is intended to ease comparisons and highlight differences across the three eras. For any one anomaly, statistical power is limited even in our sample. Our subsequent analysis aggregates anomalies to construct more powerful hypothesis tests.

In-sample, all of the anomalies have statistically significant average returns, CAPM alphas, or three-factor model alphas. Twenty-eight of the thirty-six 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 27. 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

9 In the Internet Appendix we report results for each anomaly based on univariate sorts.

Table 5
Average returns and CAPM and three-factor model alphas for 36 anomalies

Anomaly		Average return		CAPM		FF3	
		Avg.	t	$\hat{\alpha}$	t	$\hat{\alpha}$	t
				Profitability			
Gross profitability	Pre	0.00	−0.01	0.28	1.74	0.35	2.61
	In	0.26	2.76	0.30	3.10	0.43	4.68
	Post	0.23	0.83	0.63	2.65	0.57	2.80
Operating profitability	Pre	−0.01	−0.06	0.17	1.11	0.22	1.66
	In	0.27	2.93	0.29	3.13	0.33	3.52
	Post	.	.	.	.	.	.
Return on assets	Pre	−0.15	−1.03	0.05	0.45	0.23	2.68
	In	0.18	1.18	0.21	1.40	0.53	4.44
	Post	0.23	1.29	0.48	3.12	0.47	3.65
Return on equity	Pre	−0.08	−0.63	0.04	0.31	0.22	2.37
	In	0.22	1.56	0.18	1.28	0.45	3.55
	Post	0.20	1.09	0.43	2.67	0.39	3.05
Profit margin	Pre	−0.11	−0.94	0.00	0.04	0.15	1.71
	In	0.27	1.56	0.39	2.38	0.31	2.25
	Post	−0.14	−0.88	0.04	0.30	0.07	0.59
Change in asset turnover	Pre	0.07	0.83	0.09	1.14	0.11	1.43
	In	0.49	4.49	0.52	4.72	0.55	5.03
	Post	−0.05	−0.52	−0.06	−0.55	−0.07	−0.67
				Earnings quality			
Accruals	Pre	0.10	1.02	0.10	1.01	0.08	0.81
	In	0.25	2.82	0.31	3.75	0.22	2.83
	Post	0.18	1.87	0.14	1.52	0.16	1.67
Net operating assets	Pre	0.11	1.37	0.09	1.07	0.10	1.29
	In	0.30	3.80	0.29	3.71	0.38	5.03
	Post	0.11	0.80	0.08	0.61	0.07	0.59
Net working capital changes	Pre	0.26	3.98	0.30	4.64	0.27	4.21
	In	0.21	2.06	0.23	2.27	0.26	2.48
	Post	0.01	0.12	−0.02	−0.18	−0.01	−0.13
				Valuation			
Book-to-market	Pre	0.44	2.05	0.10	0.58	−0.02	−0.64
	In	0.46	3.22	0.51	3.73	0.01	0.15
	Post	0.31	1.72	0.36	2.00	0.00	0.05
Cash flow / price	Pre	0.17	1.51	0.15	1.26	0.13	1.13
	In	0.69	4.38	0.74	4.93	0.23	3.43
	Post	0.23	1.12	0.40	2.01	0.10	0.88
Earnings / price	Pre	0.12	0.62	0.30	1.57	0.35	2.05
	In	0.43	2.85	0.53	3.70	0.30	3.11
	Post	0.37	2.62	0.53	3.94	0.20	2.13
Enterprise multiple	Pre	0.17	1.08	0.24	1.56	0.22	1.44
	In	0.36	3.06	0.46	4.21	0.12	1.70
	Post	−0.14	−0.80	−0.21	−1.11	−0.20	−1.36
Sales / price	Pre	0.25	2.35	0.17	1.65	0.06	0.69
	In	0.34	1.87	0.39	2.15	0.07	0.55
	Post	0.39	1.94	0.49	2.39	0.10	0.87
				Investment and growth			
Asset growth	Pre	0.06	0.64	0.03	0.33	−0.02	−0.25
	In	0.40	3.75	0.48	4.97	0.19	2.66
	Post	0.05	0.45	0.01	0.06	−0.01	−0.05
Growth in inventory	Pre	0.20	2.70	0.20	2.61	0.18	2.49
	In	0.29	3.24	0.36	4.26	0.20	2.61
	Post	0.18	1.69	0.17	1.63	0.15	1.42

(continued)

Table 5
Continued

Anomaly		Average return Avg.	t	CAPM $\hat{\alpha}$	t	FF3 $\hat{\alpha}$	t
Sales growth	Pre	0.11	0.68	0.04	0.28	−0.01	−0.04
	In	0.24	2.06	0.28	2.55	0.02	0.27
	Post	−0.10	−0.81	−0.01	−0.04	−0.18	−2.01
Sustainable growth	Pre	0.05	0.38	−0.04	−0.35	−0.10	−1.02
	In	0.15	1.57	0.24	2.60	−0.05	−0.64
	Post	0.13	0.89	0.11	0.79	0.17	1.37
Adjusted CAPX growth	Pre	0.14	1.78	0.06	0.86	0.03	0.43
	In	0.21	2.70	0.24	3.12	0.13	1.71
	Post	0.03	0.27	0.09	0.92	0.03	0.30
Growth in sales - inventory	Pre	0.09	0.76	0.15	1.31	0.22	2.13
	In	0.40	4.93	0.39	4.76	0.42	4.97
	Post	0.10	1.04	0.10	1.09	0.14	1.61
Investment growth rate	Pre	0.07	0.76	0.09	1.05	0.09	1.01
	In	0.26	4.29	0.30	5.13	0.21	3.67
	Post	−0.09	−0.94	−0.06	−0.66	−0.06	−0.70
Abnormal capital investment	Pre	0.10	0.94	0.12	1.13	0.15	1.44
	In	0.21	3.30	0.20	3.22	0.16	2.46
	Post	−0.03	−0.33	−0.02	−0.21	−0.02	−0.24
Investment to capital	Pre	0.20	2.24	0.19	2.13	0.18	2.01
	In	0.18	1.49	0.31	2.96	0.03	0.37
	Post	0.05	0.36	0.06	0.43	0.05	0.50
Investment-to-assets	Pre	0.26	3.24	0.25	3.16	0.23	3.00
	In	0.27	3.24	0.34	4.26	0.19	2.44
	Post	0.23	1.84	0.19	1.51	0.20	1.67
				Financing			
Debt issuance	Pre	0.05	0.82	0.10	1.76	0.11	2.07
	In	0.18	3.99	0.19	4.01	0.19	4.06
	Post	0.11	1.50	0.08	1.05	0.14	2.28
Leverage	Pre	0.08	0.48	−0.08	−0.63	−0.11	−0.95
	In	0.21	2.65	0.19	2.33	0.01	0.12
	Post	0.18	1.28	0.19	1.36	−0.18	−2.22
One-year share issuance	Pre	0.26	2.47	0.34	3.25	0.37	3.70
	In	0.27	2.71	0.38	4.21	0.20	2.80
	Post	0.04	0.31	0.13	1.29	0.13	1.44
Five-year share issuance	Pre	0.09	1.20	0.18	2.43	0.20	2.77
	In	0.24	2.39	0.34	3.67	0.17	2.42
	Post	0.12	1.24	0.20	2.24	0.20	2.40
Total external financing	Pre	−0.13	−1.16	0.07	0.81	0.12	1.58
	In	0.34	2.61	0.52	4.82	0.33	4.21
	Post	0.19	1.06	0.34	2.31	0.30	2.66
				Distress			
Ohlson's O-score	Pre	−0.05	−0.33	0.21	2.01	0.32	3.84
	In	0.27	1.97	0.35	2.71	0.51	4.41
	Post	−0.08	−0.68	0.04	0.33	0.09	0.83
Altman's z-score	Pre	−0.23	−1.53	0.02	0.18	0.19	2.28
	In	0.14	0.82	0.07	0.40	0.59	4.74
	Post	−0.09	−0.53	−0.06	−0.36	0.14	1.28
Distress risk	Pre	0.08	0.36	0.44	2.35	0.50	3.37
	In	0.41	2.45	0.63	4.43	0.58	4.82
	Post	0.14	0.60	0.45	2.62	0.44	3.01
				Other			
Industry concentration	Pre	0.04	0.27	−0.02	−0.10	−0.05	−0.32
	In	0.09	1.06	0.01	0.17	0.18	2.18
	Post	−0.13	−0.77	−0.21	−1.33	−0.17	−1.29

(continued)

Table 5
Continued

Anomaly		Average return Avg.	t	CAPM $\hat{\alpha}$	t	FF3 $\hat{\alpha}$	t
				Composite anomalies			
Piotroski's F-score	Pre	0.07	0.67	0.14	1.41	0.19	1.92
	In	0.47	5.20	0.53	5.92	0.61	7.08
	Post	0.11	0.66	0.27	1.85	0.21	1.69
M/B and accruals	Pre	0.56	2.54	0.45	2.08	0.20	1.18
	In	0.62	2.34	0.86	3.33	0.25	1.25
	Post	0.36	1.00	0.30	0.86	−0.20	−0.74
QMJ: Profitability	Pre	0.04	0.17	0.36	1.72	0.42	2.29
	In	0.22	2.70	0.31	4.06	0.47	7.09
	Post	.	.	.	.	.	.

This table reports monthly average returns, CAPM alphas, and three-factor model alphas for the 36 anomalies described in the appendix. 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, except for anomalies such as debt issuance in which indicator variables determine the portfolio assignments. 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. We report the estimates for pre-sample, in-sample, and post-sample periods. In-sample period is the period used in the original study as reported in Table 4. Pre-sample period is the period that predates the sample period used in the original study. Post-sample period is the sample that has accumulated after the end of the in-sample period. The post-sample estimates are not reported for operating profitability and QMJ profitability because they have less than five years post-sample data.

large. The average return on the distress anomaly, for example, is 39 basis points per month (t-value = 2.34). However, because this anomaly covaries negatively with the market and HML factors (Campbell et al. 2008), its CAPM and three-factor model alphas are considerably higher, 60 basis points (t-value = 4.26) and 56 basis points (t-value = 4.69) per month, respectively.

Out-of-sample, most anomalies are significantly weaker both economically and statistically. In the pre-sample period, eight of the average returns and CAPM alphas, and sixteen of the three-factor model alphas are statistically significant. A total of 17 anomalies have either CAPM or three-factor model alphas that are statistically significant. 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.

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 et al. (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 Table 5 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. As Ken French notes, "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."[10]

The estimates in the post-sample period are similar to those for the pre-sample period. Of the 34 anomalies with post-sample data, only one earns an average return that is positive and statistically significant at the 5% level, and 12 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 -2.22—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 post-sample estimates are noisier than their in-sample and pre-sample counterparts.

The overlap between significant anomalies in the pre-sample period and significant anomalies in the post-sample period is modest. Accounting returns (on assets and equity) and distress risk are two sets of anomalies that are significant across both periods, at least with respect to the three-factor model. However, anomalies related to physical investment (e.g., inventories and capital expenditures) and equity financing are highly significant in the pre-sample period but insignificant in the post-sample period. In contrast, anomalies related to income statement measures (e.g., sales and earnings) and total financing are significant in the post-sample period but insignificant in the pre-sample period.

This pattern is interesting when considered in light of the changes in corporate behavior occurring over the last century. During this time frame, the U.S. economy underwent a transformation from a manufacturing and capital intensive production economy to a more service-oriented economy with a greater reliance on intangible assets, such as human and intellectual capital (Bond et al. 2000). Much intangible investment is expensed in R&D and SG&A accounts so the income statement of the second half of our sample embeds a significant amount of investment (Peters and Taylor 2017). Contemporaneously, the financing preferences of U.S. companies underwent a potentially more dramatic shift from equity financing to debt financing (Graham et al. 2015).

In combination with our anomaly results, these facts suggest that during the pre-sample period, investment and equity financing may have contained relevant information about future cash flows or discount rates when the economy was reliant on tangible investments and equity financing. However, when productivity became more reliant on intangible assets and a mix of debt and equity financing, the information content of physical investment and equity

10 Ken French discusses the repercussions of these changes at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. See also http://crsp.com/files/images/release_notes/mdaz_201306.pdf and http://crsp.com/files/images/release_notes/mdaz_201402.pdf for details on the project.

Table 6
Measuring changes in average returns, alphas, and information ratios between pre-, in-, and post-sample periods

Measure	Pre-sample	In-sample	Post-sample	Differences Pre − In	Post − In	Post − Pre
Panel A: Full pre-1963 sample						
			Average returns			
Average	0.08	0.29	0.09	−0.21	−0.20	0.00
return	(2.21)	(7.01)	(1.72)	(−3.78)	(−3.69)	(0.03)
Sharpe	0.15	0.54	0.13	−0.39	−0.42	−0.03
ratio	(3.38)	(7.57)	(1.52)	(−4.71)	(−4.14)	(−0.30)
			CAPM			
Alpha	0.15	0.34	0.17	−0.20	−0.18	0.02
	(4.80)	(9.75)	(3.50)	(−4.27)	(−3.44)	(0.38)
Information	0.22	0.66	0.27	−0.43	−0.40	0.04
ratio	(5.08)	(9.72)	(2.99)	(−5.43)	(−3.83)	(0.43)
			Three-factor model			
Alpha	0.17	0.27	0.12	−0.10	−0.15	−0.05
	(6.42)	(10.12)	(3.19)	(−2.57)	(−3.44)	(−1.10)
Information	0.28	0.60	0.25	−0.32	−0.35	−0.03
ratio	(6.35)	(9.91)	(2.86)	(−4.26)	(−3.46)	(−0.32)
Panel B: Pre-1963 sample without the pre-Securities and Exchange Act data						
			Average returns			
Average	0.09	0.29	0.09	−0.20	−0.20	0.00
return	(2.86)	(7.01)	(1.72)	(−3.96)	(−3.69)	(−0.04)
Sharpe	0.19	0.54	0.13	−0.36	−0.42	−0.06
ratio	(2.94)	(7.57)	(1.52)	(−3.88)	(−4.14)	(−0.55)
			CAPM			
Alpha	0.14	0.34	0.17	−0.20	−0.18	0.03
	(5.89)	(9.75)	(3.50)	(−4.84)	(−3.44)	(0.53)
Information	0.27	0.66	0.27	−0.39	−0.40	0.00
ratio	(4.77)	(9.72)	(2.99)	(−4.52)	(−3.83)	(−0.01)
			Three-factor model			
Alpha	0.16	0.27	0.12	−0.11	−0.15	−0.04
	(6.50)	(10.12)	(3.19)	(−3.06)	(−3.44)	(−0.78)
Information	0.34	0.60	0.25	−0.26	−0.35	−0.08
ratio	(6.17)	(9.91)	(2.86)	(−3.30)	(−3.46)	(−0.76)

This table compares the performance of the average anomaly between the sample period used in the original study (in-sample) and the periods that predate or follow the original sample period (pre- and post-sample). Panel A uses data starting in July 1926. Panel B uses data as of July 1938. We estimate average monthly returns and CAPM and three-factor model alphas for the 36 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. 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.

financing declined while that of income statement accounts and total financing increased.

3.3.2 Average anomalies. In Table 6 we compare the performance of the average HML-like anomaly between the pre-sample, in-sample, and post-sample periods. We measure average returns, Sharpe ratios, and alphas and information ratios estimated from the CAPM and three-factor models. 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 et al. 2003) and starts the sample in July 1938.

Panel A shows the average anomaly earns 29 basis points (t-value = 7.01) per month during the sample period used in the original study, but just 8 basis points (t-value = 2.21) during the historical out-of-sample period and 9 basis points (t-value = 1.72) after the end of the original sample. The differences in average returns between the original period and pre- and post-sample periods are significant with t-values of -3.78 and -3.69. We find similar results when we condition on either the market return or Fama-French three factors.

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. The Sharpe ratio divides each anomaly's average return by its volatility, and the 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-sample samples are not statistically significantly different from each other.

The average anomaly's information ratio from the three-factor model is 0.60 during the original study's sample period, but just 0.28 for the pre-sample period and 0.25 for the post-sample periods. These differences, which are statistically significant with t-values of -4.26 and -3.46, correspond to 53% to 59% decrease in the information ratio 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 36% (pre-sample) and 54% (post-sample) and so, if anything, the risk adjustment works the "wrong way": not only do anomalies earn higher alphas during the original study's sample period but they are also less risky than what they are before this period.

Panel B shows the estimates are not sensitive to removing the pre-Securities and Exchange Act data from the pre-sample sample. The differences between the in-sample period and the pre-sample period, and those between the post-sample and pre-sample period, are quantitatively similar. The pre-period looks very similar to the post-period, and it is the in-sample period that stands out.

4. Inferences

Most anomalies' in-sample behavior is markedly different from their out-of-sample behavior. In this section, we investigate why.

4.1 Sample selection sensitivity

At first glance, our findings are consistent with data-snooping. Table 5 shows most anomalies exhibited economically large and statistically significant average returns and alphas in-sample. However, out-of-sample—pre- or post-sample—most anomalies are economically and statistically insignificant. Table 6 shows the differences across performance metrics between the post- and pre-sample periods are generally less than one standard error away from each other; the in-sample period, in contrast, is different from both of them by at least four standard errors.

However, as previously discussed, dynamic considerations enable one to rationalize our findings with both alternative hypotheses: unmodeled risk and mispricing. There were a number of important macroeconomic events occurring in the 1960s and 1970s, when most in-sample periods begin (e.g., Kennedy's and Johnson's accelerated federal spending, expansion of the Vietnam war, the oil embargo, and a growing trade deficit). Any of these, among other, events could have created a shift in the risks that matter for investors. The release of Compustat in 1962 may have altered the cost of information acquisition for investors in a manner that affected trading and prices. A similar argument can be made for the difference between in-sample and post-samples.

By the same token, mispricing need not be static. Stocks may fall in and out of favor with investors, representing transient fads, consistent with our findings (e.g., Shiller 1984; Camerer 1989).

To better understand the relevance of these alternatives, we investigate the importance of the precise in-sample start date for anomaly performance. Specifically, we ask: What would anomaly performance been had the in-sample start date been something earlier, such as 1963, 1964,..., 1973. Equivalently, we ask whether a structural break occurring during one or more of these years brought about anomalies, most of whose in-sample start dates occur in the 1960s and 1970s.

We choose 1963 as a starting point for two reasons. First, Compustat was released and largely free from back-fill bias in 1963. Second, the in-sample period in the original study identifying each anomaly could have started in 1963 given the availability of Compustat. We stop in 1973 because there are few anomalies with in-sample start dates sufficiently removed from 1973 to statistically identify the effects of this change.

To test our hypotheses, we estimate the following regression:

$$\text{anomaly}_{it} = \beta_0 + \beta_1 \text{Pre-sample}_{it} + \mu_i + \varepsilon_{it}, \tag{3}$$

where (i,t) indexes (anomaly, month), μ_i is an anomaly fixed effect, and ε_{it} is an error term. Pre-sample$_{it}$ is an indicator equal to one if the anomaly-month observation falls in the time period before the start-date of the anomaly's in-sample start date. We restrict the sample to include anomaly-month observations from July of 1963 (1964, 1965,..., 1973) to the end of the in-sample period.

Table 7
Estimating the effect of the start date on average anomaly returns and CAPM and three-factor model alphas

	Average return		CAPM alpha		FF3 alpha		
Start year	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	No. of obs.
1963	0.30	−0.15	0.36	−0.18	0.27	−0.14	14,793
	(6.77)	(−2.16)	(10.07)	(−2.97)	(10.35)	(−3.18)	
1964	0.30	−0.15	0.36	−0.19	0.28	−0.13	14,385
	(6.77)	(−1.93)	(10.06)	(−2.86)	(10.40)	(−2.68)	
1965	0.30	−0.13	0.36	−0.17	0.28	−0.11	13,977
	(6.78)	(−1.58)	(10.07)	(−2.42)	(10.40)	(−2.27)	
1966	0.30	−0.13	0.37	−0.17	0.28	−0.12	13,569
	(6.84)	(−1.46)	(10.15)	(−2.26)	(10.45)	(−2.21)	
1967	0.31	−0.09	0.37	−0.13	0.29	−0.12	13,161
	(6.89)	(−0.98)	(10.21)	(−1.84)	(10.50)	(−2.25)	
1968	0.31	−0.07	0.37	−0.13	0.29	−0.13	12,753
	(6.90)	(−0.80)	(10.20)	(−2.07)	(10.45)	(−2.31)	
1969	0.31	−0.11	0.38	−0.17	0.29	−0.18	12,345
	(6.77)	(−1.21)	(10.09)	(−2.53)	(10.44)	(−3.04)	
1970	0.30	−0.22	0.37	−0.24	0.29	−0.25	11,937
	(6.53)	(−2.68)	(9.91)	(−3.37)	(10.22)	(−4.12)	
1971	0.31	−0.22	0.37	−0.26	0.29	−0.28	11,532
	(6.63)	(−2.65)	(9.88)	(−3.78)	(10.02)	(−4.60)	
1972	0.32	−0.21	0.38	−0.26	0.29	−0.31	11,136
	(6.64)	(−2.18)	(9.86)	(−3.24)	(9.86)	(−4.52)	
1973	0.31	−0.24	0.38	−0.28	0.31	−0.27	10,740
	(6.38)	(−2.20)	(9.70)	(−3.31)	(10.17)	(−3.51)	

This table reports estimates from panel regressions that estimate the effect of the start date on average anomaly returns and CAPM and three-factor model alphas. The dependent variable is an anomaly's return, the constant plus residual from the CAPM regression, or the constant plus residual from the three-factor model regression. The beginning of the sample period is indicated in the first column. For each anomaly, we include data up to the end of the in-sample period. The regressions only include anomalies whose in-sample periods begin after the first column's start date. The regression is

$$\text{anomaly}_{it} = \beta_0 + \beta_1 \text{Pre-sample}_{it} + \mu_i + \varepsilon_{it},$$

where (i,t) indexes (anomaly, month), Pre-sample_{it} is an indicator equal to one if the anomaly-month observation falls in the time period before the start-date of the anomaly's in-sample start date, μ_i is an anomaly fixed effect, and ε_{it} is a error term. From the sample, we exclude ple the book-to-market anomaly and the two anomalies that are annual versions of originally quarterly anomalies (return on assets and return on equity). Standard errors are clustered by calendar month.

For example, consider the asset growth (i.e., investment) anomaly. The in-sample period is from 1968 to 2003. When investigating the impact of extending the sample back to 1963, the estimation sample for the asset growth anomaly runs from July of 1963 through June of 2003. When investigating the impact of starting in 1964, the estimation sample runs from July of 1964 through June of 2003. And so on. We pool all anomaly-month observations and condition out any between-variation with anomaly fixed effects. Therefore, identification of β_1 comes from anomaly-month observations whose in-sample start date begins after the start of the sample. We cluster standard errors by calendar month to account for correlated errors in the cross-sections.

Each row in Table 7 presents estimation results for a different subsample of anomaly-month observations whose start year is denoted in the first column. The first two columns show the results when the dependent variable is the anomaly's HML-like average return. We see that the constant, which measures the average in-sample anomaly return, is always significant and ranges between 30 and 32 basis points.

More interesting is the economically large and often significant coefficient on the pre-sample indicator. Looking at the first row, "1963", we note that extending the in-sample period back to 1963 leads to a 15 basis point or 52% reduction in the average anomaly return. Looking across the rows we see a similar pattern. Extending the sample back in time even by a modest amount leads to sharp reductions in average returns. For anomalies whose in-sample start date is after 1974, the reductions are even larger—24 basis points or approximately 77%.

The CAPM and three-factor model columns tell a similar, if not more compelling, story. The middle columns, for example, show estimation results of Equation (3), where the dependent variable is constructed from the sum of the residuals and intercept from a regression of anomaly returns on the market factor. The result is a y-variable that is orthogonal with respect to market risk. The rightmost columns present analogous results where the dependent variable is constructed as the sum of the residuals and intercept from a regression of anomaly returns on the market, size, and value factors.

These results present a challenge for theories predicated on unmodeled risk in combination with structural breaks in the risks that matter for investors. Such a theory would have to contain a large number of distinct breaks in the 1960s and 1970s, each of which operates through a different accounting-based signal.[11] Likewise, transient fads require reconciling the large number of fads moving in and out of favor at distinct points in time, as opposed to broader changes in investor tastes.

This analysis begs the question: if earlier data was available at the time, why did researchers not use it? We reviewed the original articles for discussions of the sample frame to answer this question. In most instances, the in-sample start date is determined by a desire to ensure a consistent sample frame throughout the original study. For example, some studies use additional data sources that have a later start date than CRSP and Compustat (e.g., Haugen and Baker 1996; Lyandres et al. 2008; Soliman 2008). Others perform additional analysis that requires multiple years of lagged data (e.g., Lakonishok et al. 1994; Cooper et al. 2008). Thus, consistency of sample period throughout the original study appears to be the primary motivation, though the lack of robustness to small changes in the sample frame remains.

4.2 Out-of sample validation versus statistical adjustments

An alternative to out-of-sample validation is statistical adjustments based on multiple hypothesis testing frameworks. As previously mentioned,

[11] In so far as some signals are redundant, the onus is somewhat lessened but still requires reconciling the asynchronicity that we observe.

Table 8
Out-of-sample evaluation versus statistical adjustments

Anomaly	Significant in pre-sample	In-sample t-value > 3	Intersection
Gross profitability	❀	❀	❀
Operating profitability		❀	
Return on assets	❀	❀	❀
Return on equity	❀	❀	❀
Change in asset turnover		❀	
Net operating assets		❀	
Net working capital changes	❀		
Cash flow / price		❀	
Earnings / price	❀	❀	❀
Growth in inventory	❀		
Growth in sales − inventory	❀	❀	❀
Investment growth rate		❀	
Investment to capital	❀		
Investment-to-assets	❀		
Debt issuance	❀	❀	❀
One-year share issuance	❀		
Five-year share issuance	❀		
Total external financing		❀	
O-score	❀	❀	❀
z-score	❀	❀	❀
Distress risk	❀	❀	❀
Piotroski's F-score		❀	
QMJ: Profitability	❀	❀	❀
Count	16	17	10

This table reports three sets of anomalies. The first set in the leftmost column contains anomalies identified as having statistically significant Fama-French three-factor model alphas in the pre-sample period. The second set in the middle column contains anomalies identified as having statistically significant Fama-French three-factor model alphas based on a t-statistic threshold of 3.00, as suggested by Harvey et al. (2015). The third set in the rightmost column contains the intersection of the first two sets.

these adjustments have limitations in light of the accompanying reporting requirements. Nonetheless, recent studies have proposed statistically motivated rules of thumb for mitigating the effects of data-snooping. Harvey et al. (2015), for example, suggest employing a test statistic cutoff of 3.00, as opposed to the conventional 1.96 for a two-sided test at the 5% level. We apply this approach to our set of anomalies and compare the results with those obtained from out-of-sample tests.

The results of this comparison are presented in Table 8. The listed anomalies corresponds to the union of two sets. The first consists of anomalies whose pre-sample Fama-French three-factor model alphas are statistically significant at conventional levels. Table 5 reveals 16 such anomalies. The second set consists of anomalies whose in-sample t-value is greater than 3.00. Table 5 reveals 17 such anomalies. Casual observation reveals a clear positive correlation between the anomalies identified as robust by both approaches. However, the intersection between these two sets is 10, suggesting a significant number of anomalies in which one approach identifies an anomaly as robust while the other does not. Therefore, statistical adjustments are not a substitute for the additional information contained in out-of-sample tests.

4.3 Volatility across the eras

Because data-snooping works through t-values, it is instructive to examine second moments. For example, the relative volatilities of the anomalies vary substantially across eras. We measure this effect by estimating a panel regression,

$$(\text{anomaly}_{it} - \bar{r}_{it})^2 = \beta_0 + \beta_1 \text{In-sample}_{it} + \beta_2 \text{Post-sample}_{it} + \mu_t + \varepsilon_{it}, \qquad (4)$$

where anomaly_{it} is the return on anomaly i in month t, $\bar{r}_{it}$ is the anomaly's average return, estimated separately for the pre-, in-, and post-periods (hence the t subscript), In-sample_{it} and Post-sample_{it} are indicators equal to one if the anomaly-month observation is in the in- or post-sample period and μ_t is a month fixed effect. We include month fixed effects to control for the large market-wide changes in market and idiosyncratic volatility over the twentieth century (Campbell et al. 2001). In this regression, with standard errors clustered by month, the slope estimates for the in- and post-sample indicators are -0.61 (t-value = -2.88) and 0.38 (t-value = 0.79), respectively. These parameter estimates indicate that the average anomaly's volatility is 9.19% per year during the pre-sample period, but 8.78% when in-sample.

This decrease in volatility is consistent with data-snooping bias contaminating the distribution of in-sample returns. Because data-mining works through t-values, both high average return and low volatility make it more likely that a particular factor is deemed a return anomaly.

4.4 Comovement across the eras

In addition, an anomaly's t-value is higher if its returns have more independent variation. Researchers 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 is atypically low.

We use the empirical specification of McLean and Pontiff (2016) as a start point to investigate comovement and its implications. In particular, an important economic message of McLean and Pontiff (2016) is that investors trade on signals discovered by academic research. This correlated trading increases comovement among anomalies leading to a reduction in anomaly t-values after publication (i.e., in our post-sample period).

Using anomaly-month observations, we estimate the following regression:

$$\begin{aligned} \text{anomaly}_{i,t} = a + b_1 \text{in-sample index}_{-i,t} \\ + b_2 \text{post-sample index}_{-i,t} + b_3 \text{post}_{i,t} \\ + \text{post}_{i,t} \times \big(b_4 \text{in-sample index}_{-i,t} \\ + b_5 \text{post-sample index}_{-i,t}\big) + e_{i,t}, \end{aligned} \qquad (5)$$

where $\text{post}_{i,t}$ takes the value of one if anomaly i is in the post-sample in month t and zero otherwise, $\text{in-sample index}_{-i,t}$ is the average return on all anomalies

Table 9
Changes in the correlation structure of returns

Regressor	Coefficient	t-value
Regression 1: In-sample versus post-sample anomalies		
Intercept	0.05	4.54
Main effects		
In-sample index$_{-i,t}$	0.74	33.98
Post-sample index$_{-i,t}$	0.08	7.46
Post$_{i,t}$	−0.06	−2.23
Interactions		
Post$_{i,t}$ × In-sample index$_{-i,t}$	−0.53	−13.74
Post$_{i,t}$ × Post-sample index$_{-i,t}$	0.46	11.19
Adjusted R^2		17.9%
N		15,152
Regression 2: In-sample versus pre-sample anomalies		
Intercept	0.07	4.35
Main effects		
In-sample index$_{-i,t}$	0.74	28.90
Pre-sample index$_{-i,t}$	0.07	3.42
Pre$_{i,t}$	−0.04	−2.09
Interactions		
Pre$_{i,t}$ × In-sample index$_{-i,t}$	−0.69	−22.72
Pre$_{i,t}$ × Pre-sample index$_{-i,t}$	0.48	13.68
Adjusted R^2		9.3%
N		13,650

This table reports estimates from panel regressions of monthly anomaly returns on the average returns on all other anomalies that are either in-sample or out-of-sample. In the first regression, post$_{i,t}$ takes the value of one if anomaly i is in the post-sample period in month t and zero otherwise; in-sample index$_{-i,t}$ is the average return on all anomalies, except anomaly i, in-sample in month t; and post-sample index$_{-i,t}$ is the average return on all anomalies, except anomaly i, in the post-sample period in month t. In the second regression, pre$_{i,t}$ takes the value of one if anomaly i is in the pre-sample period in month t and zero otherwise; in-sample index$_{-i,t}$ is the average return on all anomalies, except anomaly i, in-sample in month t; and pre-sample index$_{-i,t}$ is the average return on all anomalies, except anomaly i, in the pre-sample period in month t.

except anomaly i that are in-sample in month t, and post-sample index$_{-i,t}$ is the average return on all anomalies, except anomaly i, in the post-sample in month t.

The sample is selected to include either in-sample or post-sample anomaly-month observations. (We exclude all pre-sample observations.) We cluster standard errors by calendar month to account for the correlated errors in the cross-sections. The specification (5) is similar to that estimated in McLean and Pontiff (2016) 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-sample upon its discovery.

The Regression 1 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 0.74, and that against the post-sample index is just 0.08. When the anomaly is post-sample, this pattern reverses. The slope coefficient against the in-sample index is $0.74+(-0.53)=0.22$, and that against the post-sample index is $0.08+0.46=0.55$.

These changes in slope coefficients are highly significant with corresponding t-values of -13.7 and 11.2. Thus, these results show that comovement among anomalies increases as they get discovered (i.e., moving from in-sample to post-sample), consistent with the notion that institutional investors learn from and trade on anomalies discovered by academics (McLean and Pontiff 2016).

Of course, these results beg the question of why anomalies are so highly correlated in-sample, before their discovery. McLean and Pontiff (2016) argue that this is due to a common source, such as investor sentiment. But, if this is the case, then it is unclear why this common source would disappear, to be replaced entirely by institutional trading on the anomaly, when an anomaly enters the post-sample period. In other words, it seems that an equally valid null hypothesis is that the factor structure behind the anomalies in the in-sample period persists into the post-sample period, and institutional trading increases comovement above and beyond what is implied by the factor structure. In the context of the regression results, the null hypothesis is not that $b_5 = 0$, but rather that $b_5 = b_1$, which we fail to reject.

A statistical issue unrelated to economics may also account for the pattern in correlations in Table 9. Suppose that anomalies share a common factor, as suggested by Lee et al. (1991), Barberis and Shleifer (2003), and Barberis et al. (2005). This common factor may generate the pattern reported in Table 9 because of how the "diversification" of the in- and post-sample indices changes over time. In the beginning of the sample, the post-sample index is poorly diversified compared to the in-sample index; and toward the end of the sample, it is the other way around. Prior to 1971, no anomaly is in the post-sample; from 1971 to 1990, at most two anomalies are in the post-sample. During this period, the post-sample index is therefore a much noisier measure of the common factor than the in-sample index. These counts reverse toward the end of the sample. In 2004, for example, only seven anomalies are in the in-sample index while 29 are in the post-sample index. Now, by construction, the *average* anomaly always belongs to the better-populated index. When the average anomaly is in-sample, most other anomalies are also in-sample; and when it moves into the post-sample, most other anomalies are also in post-sample. Table 9 tells us that correlations change when an anomaly moves into post-sample because, at this point, the post-sample index becomes less noisy.

To emphasize this point, we reestimate Equation (5) with two modifications. First, we select the sample to include anomaly-month observations in the pre-sample and in-sample periods. (We exclude post-sample observations.) Second, we replace the post-sample indicator with a pre-sample indicator. In accordance with the inferences in McLean and Pontiff (2016), we should not observe any change in the partial correlations as we move from in-sample to pre-sample because institutional investors cannot trade against anomalies that have yet to be discovered. In other words, the coefficients on the interaction terms should be zero.

In contrast, the Regression 2 estimates in Table 9 are indistinguishable from those found in the first specification. To ease comparisons, we repeat the corresponding estimates from the first regression in parentheses below. When an anomaly is in-sample, its slope coefficient against the in-sample index is 0.74 (0.74), and that against the pre-sample index is just 0.07 (0.08). When the anomaly is in its pre-sample period, this pattern reverses. The slope coefficient against the in-sample index is 0.05 (0.22), and that against the pre-sample index is 0.55 (0.55). These changes in slope coefficients are highly significant with corresponding t-values of −22.7 (−13.7) and 13.7 (11.2).

In sum, arbitrageurs may well learn from academic research. However, the empirical tests from this regression have no power to distinguish this hypothesis from data-snooping or, more simply, the presence of a factor structure in anomalies. In light of the magnitudes of the t-statistics, most of the increased correlation in the pre- and post-sample eras is likely due to the common factor structure among the anomalies.

5. Discussion and Conclusions

It is worth repeating that a number of anomalies are robust across two or more eras in our sample. Accounting returns and distress risk persist across all three eras, depending on the benchmark model. Others, such as capital expenditures and modes of financing, persist only across two eras but in a manner consistent with their relevance for future cash flows or discount rates. When we also consider the robust price-based anomalies, for example, short-term reversals and medium-term momentum, our study offers further confirmation that some anomalies are likely "real."

However, we also show that factors previously thought of as particularly robust, such as investment, are not once one moves back in time, and that this finding is common among the majority of anomalies that we are able to examine. Further, one need not move that far back in time to erode the significance of most anomalies, whose average returns decline between 50% and 75% when we extend the data back just a few years. Thus, data-snooping has had a significant impact on the discovery of most accounting-based return anomalies.

Our results may be good news for asset pricing models. The data-snooping problem is so severe that we would expect to reject even the true asset pricing model in in-sample data. Asset pricing models are evaluated by studying the maximum Sharpe ratios of the factors, the test assets, or their combination (Barillas and Shanken 2017; Gibbons et al. 1989). Therefore, to illustrate the economic significance of data-mining from the perspective of model testing, let us consider the three-factor model information ratios of Table 6. The average anomaly's information ratio is 0.60 in-sample; in the pre- and post-sample, they are 0.28 and 0.25, respectively. The standard errors of these estimates are 0.04 (pre-sample), 0.06 (in-sample), and 0.09 (post-sample). These estimates are therefore all statistically significantly different from zero, leading us to reject the three-factor model.

Now, consider the role of data-mining under arguably conservative assumptions. First, assume that half of the "extra" in-sample information ratio is due to data-mining. This data-snooped information ratio would be $\frac{1}{2}(0.60-0.28)=0.16$. Second, assume that we can measure information ratios almost as precisely as in the post-sample, that is, assume that SE = 0.08. The data-snooped information ratio would therefore be two standard errors away from zero. That is, we would reject the correct asset pricing model for its inability to explain away the data-snooped part of returns.

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 or, if unavailable, a hold-out sample.

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 data.

Appendix: Anomalies

In this appendix, we define the anomalies examined in Section 3. When applicable, we state the formulas using the Compustat item names. The numbering of the anomalies below corresponds to that in Table 4. For those anomalies computed through a process involving multiple steps, we refer to the studies that describe the implementation in detail. We also indicate (1) the first study that used each variable to explain the cross-section of stock returns and (2) the sample period used in that study. When applicable, we use McLean and Pontiff (2016) to identify the first study. We state both the year and the month when the months are provided in the original study; if not given, 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. In the Internet Appendix, we describe these approximations and compare the average returns and the CAPM and three-factor model alphas of the original definitions and the approximations.

A Profitability

1. **Gross profitability** is defined as the revenue minus cost of goods sold, all divided by total assets: $\text{gross profitability}_t=(\text{revt}_t-\text{cogs}_t)/\text{at}_t$. Novy-Marx (2013) examines the predictive power of gross profitability using return data from July 1963 through December 2010.
2. **Operating profitability** is defined as the revenue minus cost of goods sold, SG&A, and interest, all divided by book value of equity: $\text{operating profitability}_t=(\text{revt}_t-\text{cogs}_t-\text{xsga}_t-\text{xint}_t)/\text{be}_t$. Fama and French (2015) construct a profitability factor based on operating profitability using return data from July 1963 through December 2013.

3. **Return on assets** is defined as the earnings before extraordinary items, divided by total assets: return on $\text{assets}_t = \text{ib}_t/\text{at}_t$. 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: return on $\text{equity}_t = \text{ib}_t/\text{be}_t$. Haugen and Baker (1996) use return on equity to predict returns between 1979 and 1993.
5. **Profit margin** is defined as the earnings before interest and taxes, divided by sales: profit $\text{margin}_t = \text{oiadp}_t/\text{revt}_t$. Soliman (2008) uses profit margin to predict returns using return data from 1984 to 2002.
6. **Change in asset turnover** is defined as the annual change in asset turnover, where asset turnover is revenue divided by total assets: change in asset $\text{turnover}_t = \Delta(\text{revt}_t/\text{at}_t)$. Soliman (2008) uses the change in asset turnover to predict returns between 1984 and 2002.

B Earnings Quality

7. **Accruals** is the noncash 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. **Net operating assets** represent the cumulative difference between operating income and free cash flow, scaled by lagged total assets, net operating $\text{assets}_t = ([\text{at}_t - \text{che}_t] - [\text{at}_t - \text{dlc}_t - \text{dltt}_t - \text{be}_t])/\text{at}_{t-1}$. Hirshleifer et al. (2004) form trading strategies based on net operating assets using data from July 1964 through December 2002.
9. **Net working capital changes** is another measure of accruals: net working capital $\text{changes}_t = (\Delta[\text{act}_t - \text{che}_t] - \Delta[\text{lct}_t - \text{dlc}_t])/\text{at}_t$. Soliman (2008) uses net working capital changes to predict stock returns using return data from 1984 to 2002.

C Valuation

10. **Book-to-market ratio** is defined as the book value of equity divided by the December market value of equity: book-to-market $\text{ratio}_t = \text{be}_t/\text{mv}_t$. Fama and French (1992) use book-to-market ratio to predict returns using return data from July 1963 through December 1990.[12]
11. **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 $\text{ratio}_t = (\text{ib}_t + \text{dp}_t)/\text{mv}_t$. Lakonishok et al. (1994) use the cash flow-to-price ratio in tests that use return data from May 1968 through April 1990.
12. **Earnings-to-price ratio** is defined as the income before extraordinary items divided by the December market value of equity: earnings-to-price $\text{ratio}_t = \text{ib}_t/\text{mv}_t$. Basu (1977) measures the predictive power of earnings-to-price ratio using data from April 1957 through March 1971.

[12] The book value of equity is computed as follows. First, we set the book value of equity equal to stockholders' equity (SEQ) if this data item exists. This is also the data item collected by Davis et al. (2000) for the pre-1963 data. Second, if SEQ is missing but both common equity (CEQ) and the par value of preferred stock (PSTK) exist, then we set the book value of equity equal to PSTK + CEQ. Third, if the above definitions cannot be used, but the book values of total assets (AT) and total liabilities (LT) exist, then we set the book value of equity equal to AT − LT. If the book value of equity is now nonmissing, 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.

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

D Growth and Investment

15. **Asset growth** is defined as the percentage change in total assets: asset $\text{growth}_t = \text{at}_t/\text{at}_{t-1} - 1$. Cooper et al. (2008) examine the predictive power of asset growth using return data from July 1968 to June 2003.
16. **Growth in inventory** is defined as the change in inventory divided by the average total assets: growth in $\text{inventory}_t = \Delta\text{invt}_t/([\text{at}_t + \text{at}_{t-1}]/2)$. Thomas and Zhang (2002) use growth in inventory to predict stock returns using return data from 1970 to 1997.
17. **Sales growth** is constructed by ranking firms each year by sales rank and by computing the weighted average sales growth rank over the previous five years: sales $\text{growth}_t = 5 * \text{rank}_t + 4 * \text{rank}_{t-1} + 3 * \text{rank}_{t-2} + 2 * \text{rank}_{t-3} + 1 * \text{rank}_{t-4}$. The ranks are computed using data on firms with six years of sales data. Lakonishok et al. (1994) measure the predictive power of sales growth using return data from May 1968 through April 1990.
18. **Sustainable growth** is defined as the percentage change in the book value of equity: sustainable $\text{growth}_t = (\text{be}_t - \text{be}_{t-1})/\text{be}_{t-1}$. Lockwood and Prombutr (2010) use return data from July 1964 through June 2007 to measure the predictive power of sustainable growth.
19. **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 CAPX growth is computed as CAPX $\text{growth}_t = (\text{capx}_t - [\text{capx}_{t-1} + \text{capx}_{t-2}]/2)/([\text{capx}_{t-1} + \text{capx}_{t-2}]/2)$. The industry adjustment subtracts the average CAPX growth of the firms with the same two-digit SIC code. Abarbanell and Bushee (1998) use industry-adjusted CAPX growth as a return predictor using return data from 1974 through 1993.
20. **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 $\text{inventory}_t = (\text{revt}_t - [\text{revt}_{t-1} + \text{revt}_{t-2}]/2)/([\text{revt}_{t-1} + \text{revt}_{t-2}]/2) - (\text{invt}_t - [\text{invt}_{t-1} + \text{invt}_{t-2}]/2)/([\text{invt}_{t-1} + \text{invt}_{t-2}]/2)$. Abarbanell and Bushee (1998) use growth in sales minus inventory to predict returns from 1974 through 1993.
21. **Investment growth rate** is the percentage change in capital expenditures: investment growth $\text{rate}_t = \text{capx}_t/\text{capx}_{t-1} - 1$. Xing (2008) uses investment growth rate to construct an investment factor using return data from 1964 to 2003.
22. **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 $\text{investment}_t = \frac{\text{capx}_t}{\text{revt}_t} / \left(\frac{1}{3}\sum_{j=1}^{3} \frac{\text{capx}_{t-j}}{\text{revt}_{t-j}}\right)$. Titman et al. (2004) measure the predictive power of abnormal capital investment using return data from July 1973 through June 1996.
23. **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 $\text{ratio}_t = \text{capx}_t/\text{ppent}_{t-1}$.

Xing (2008) uses the investment-to-capital ratio to predict stock returns using return data from 1964 to 2003.

24. **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 $\text{ratio}_t = (\Delta \text{ppent}_t + \Delta \text{invt}_t)/\text{at}_{t-1}$. Lyandres et al. (2008) use the investment-to-assets ratio to predict returns from January 1970 through December 2005.

E Financing

25. **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 $\text{issuance}_t = \mathbf{1}_{\text{dlc}_t+\text{dltt}_t > \text{dlc}_{t-1}+\text{dltt}_{t-1}}$. 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 nonissuers using return data from February 1975 through December 1994. McLean and Pontiff (2016) use Compustat variables dltis_t to measure debt issuances, and we compare this definition to our approximation in the Internet Appendix.
26. **Leverage** is defined as the ratio of long-term debt and the December book value of equity: $\text{leverage}_t = \text{dltt}_t/\text{me}_t$. Bhandari (1988) uses leverage (defined as total assets minus book value of equity, all divided by the market value of equity) as a return predictor using return data from 1948 to 1979.
27. **One-year share issuance** is the log-change in the split-adjusted number of shares outstanding from fiscal year $t-1$ to t, one-year share $\text{issuance}_t =$ adjusted shrout_t/adjusted shrout_{t-1}, where adjusted $\text{shrout}_t = \text{shrout}_t * \text{cfacshr}_t$ from CRSP or, if missing or zero, $1{,}000 * \text{csho}_t * \text{ajex}_t$ from Compustat. The number of shares from CRSP are measured at the fiscal-year ends. The share issuance factor takes long positions in firms that repurchase shares (one-year share $\text{issuance}_t < 0$) and short positions in firms that issue shares (one-year share $\text{issuance}_t > 0$). Pontiff and Woodgate (2008) measure the predictive power of one-year share issuance using return data from 1932 through 2003. Because of the corrections to the historical CRSP data (see footnote 10), we set the beginning of the in-sample period as July 1968, the same as that in Daniel and Titman (2006).
28. **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.
29. **Total external financing** is the sum of net share issuance and net debt issuance minus cash dividends, all scaled by total assets. We compute this measure from the balance sheet as total external $\text{financing}_t = ([\text{adjusted shrout}_t/\text{adjusted shrout}_{t-1} - 1] * \text{me}_t + [\Delta\text{dlc}_t + \Delta\text{dltt}_t] - \text{dvc}_t)/\text{at}_t$, where the first term approximates share issuance and me_t is the fiscal year-end market value of equity. Bradshaw et al. (2006) use a measure of total external financing computed from the statement of cash flows to predict returns using return data from 1971 through 2000. Their measure is defined as total external $\text{financing}_t = (\text{sstk}_t - \text{prstkc}_t + \text{dltis}_t - \text{dltr}_t + \text{dlcch}_t - \text{dv}_t)/\text{at}_t$, with dlcch_t set to zero when missing.

F Distress

30. **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.[13] Dichev (1998) uses Ohlson O-score to predict stock returns using return data from January 1981 through December 1995.

31. **Altman's z-score** is another measure of distress. It is the fitted value from a discriminant function that Altman (1968) estimates to predict bankruptcies among 66 companies from 1946 through 1965: Altman's $\text{z-score}_t = 1.2 * (\text{act}_t - \text{lct}_t)/\text{at}_t + 1.4 * \text{re}_t/\text{at}_t + 3.3 * (\text{ni}_t + \text{xint}_t + \text{txp}_t)/\text{at}_t + 0.6 * \text{me}_t/\text{lt}_t + 1.0 * \text{revt}_t/\text{at}_t$, where me_t is the December market value of equity. Dichev (1998) predicts returns using Altman's z-score from January 1981 through December 1995.
32. **Distress risk** is yet another measure of distress. It is the fitted value from a logistic regression that Campbell et al. (2008) estimate using data for the period from 1963 through 2003 to predict failures. We use the logit-regression estimates for the 12-month horizon reported in Campbell et al. (2008, table IV). The original study details the variable construction rules in Section I and in the appendix. Campbell et al. (2008) measure the relation between distress and stock returns using data from 1981 through 2003. We consider the full 1963–2003 period to be the in-sample period.

G Other

33. **Industry concentration** is the Herfindahl index computed using the 3-digit SIC codes and with sales as the measure of market share. We exclude the same regulated industries over the same time periods used in Hou and Robinson (2006).

H 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 in 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 et al. (2013b). 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 resultant six z-scores. Asness et al. (2013b) examine the predictive power of this composite measure using return data from 1956 to 2012.

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

References

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

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

Ang, A., and J. Chen. 2007. CAPM over the long run: 1926–2001. *Journal of Empirical Finance* 14:1–40.

Asness, C., T. J. Moskowitz, and L. H. Pedersen. 2013a. Value and momentum everywhere. *Journal of Finance* 68:929–85.

Asness, C. S., A. Frazzini, and L. H. Pedersen. 2013b. Quality minus junk. Working Paper, AQR Capital Management.

Ball, R., and R. Watts. 1977. Reply to Salamon and Smith. *Journal of Finance* 32:1802–8.

Ball, R., J. Gerakos, J. T. Linnainmaa, and V. Nikolaev. 2016. Accruals, cash flows, and operating profitability in the cross section of stock returns. *Journal of Financial Economics* 121:28–45.

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* 52:56–60.

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* 52:875–83.

Barberis, N., and A. Shleifer. 2003. Style investing. *Journal of Financial Economics* 68:161–99.

Barberis, N., A. Shleifer, and J. Wurgler. 2005. Comovement. *Journal of Financial Economics* 75:283–317.

Barillas, F., and J. Shanken. 2017. Which alpha? *Review of Financial Studies* 30:1316–38.

Bartov, E., and M. Kim. 2004. Risk, mispricing, and value investing. *Review of Quantitative Finance and Accounting* 23:353–76.

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* 32:663–82.

Bhandari, L. C. 1988. Debt/equity ratio and expected common stock returns: Empirical evidence. *Journal of Finance* 43:507–28.

Bickel, P. J., and K. A. Doksum. 1977. *Mathematical statistics*. Upper Saddle River, NJ: Prentice Hall.

Bond, S. R., J. G. Cummins, J. Eberly, and R. J. Shiller. 2000. The stock market and investment in the new economy: Some tangible facts and intangible fictions. *Brookings Papers on Economic Activity* 2000:61–124.

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* 62:877–915.

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* 42:53–85.

Camerer, C. 1989. Bubbles and fads in asset prices. *Journal of Economic Surveys* 3:3–41.

Campbell, J. Y., J. Hilscher, and J. Szilagyi. 2008. In search of distress risk. *Journal of Finance* 63:2899–939.

Campbell, J. Y., M. Lettau, B. G. Malkiel, and Y. Xu. 2001. Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. *Journal of Finance* 56:1–43.

Chordia, T., A. Subrahmanyam, and Q. Tong. 2014. Have capital market anomalies attenuated in the recent era of high liquidity and trading activity? *Journal of Accounting and Economics* 58:41–58.

Cochrane, J. H. 2011. Presidential address: Discount rates. *Journal of Finance* 66:1047–108.

Cohen, R. B., C. Polk, and T. Vuolteenaho. 2003. The value spread. *Journal of Finance* 58:609–41.

Cooper, M. J., H. Gulen, and M. J. Schill. 2008. Asset growth and the cross-section of stock returns. *Journal of Finance* 63:1609–51.

Daniel, K., and S. Titman. 2006. Market reactions to tangible and intangible information. *Journal of Finance* 61:1605–43.

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

Dichev, I. A. 1998. Is the risk of bankruptcy a systematic risk? *Journal of Finance* 53:1131–47.

Fama, E. F., and K. R. French. 1992. The cross-section of expected stock returns. *Journal of Finance* 47:427–65.

———. 1993. Common risk factors in the returns of stocks and bonds. *Journal of Financial Economics* 33:3–56.

———. 1996. Multifactor explanations of asset pricing anomalies. *Journal of Finance* 51:55–84.

———. 1998. Value versus growth: The international evidence. *Journal of Finance* 53:1975–99.

———. 2008. Average returns, B/M, and share issues. *Journal of Finance* 63:2971–95.

———. 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116:1–22.

———. 2017. International tests of a five-factor asset pricing model. *Journal of Financial Economics* 123:441–63.

French, K. R. 2008. Presidential address: The cost of active investing. *Journal of Finance* 63:1537–73.

Gibbons, M., S. A. Ross, and J. Shanken. 1989. A test of the efficiency of a given portfolio. *Econometrica* 57:1121–52.

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* 118:658–83.

———. 2016. How does government borrowing affect corporate financial and investment policies? Working Paper, Wharton School.

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

Hasbrouck, J. 2009. Trading costs and returns for U.S. equities: Estimating effective costs from daily data. *Journal of Finance* 64:1445–77.

Haugen, R. A., and N. L. Baker. 1996. Commonality in the determinants of expected stock returns. *Journal of Financial Economics* 41:401–39.

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., and D. T. Robinson. 2006. Industry concentration and average stock returns. *Journal of Finance* 61:1927–56.

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

Jaffe, J., D. B. Keim, and R. Westerfield. 1989. Earnings yields, market values, and stock returns. *Journal of Finance* 44:135–48.

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

Jones, C. M. 2002. A century of stock market liquidity and trading costs. Working Paper.

Lakonishok, J., A. Shleifer, and R. Vishny. 1994. Contrarian investment, extrapolation and investment risk. *Journal of Finance* 49:1541–78.

Lee, C. M. C., A. Shleifer, and R. M. Thaler. 1991. Investor sentiment and the closed-end fund puzzle. *Journal of Finance* 46:75–109.

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* 47:13–37.

Lo, A. W., and A. C. MacKinlay. 1990. Data-snooping biases in tests of financial asset pricing models. *Review of Financial Studies* 3:431–67.

Lockwood, L., and W. Prombutr. 2010. Sustainable growth and stock returns. *Journal of Financial Research* 33:519–38.

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* 46:1629–50.

Lyandres, E., L. Sun, and L. Zhang. 2008. The new issues puzzle: Testing the investment-based explanation. *Review of Financial Studies* 21:2825–55.

McLean, R. D., and J. Pontiff. 2016. Does academic research destroy stock return predictability? *Journal of Finance* 71:5–32.

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

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

Ohlson, J. A. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research* 18:109–31.

———. 1995. Earnings, book values, and dividends in equity valuation. *Contemporary Accounting Research* 11:661–87.

Peters, R. H., and L. A. Taylor. 2017. Intangible capital and the investment-q relation. *Journal of Financial Economics* 123:251–72.

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* 63:921–45.

Schwert, G. W. 2003. Anomalies and market efficiency. In Handbook of the economics of finance, volume 1, chapter 15, pages 939–74. Eds. G. M. Constantinides, M. Harris, and R. Stulz, editors. Amsterdam: Elsevier.

Sharpe, W. F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19:425–42.

Shiller, R. J. 1984. Stock prices and social dynamics. *Brookings Papers on Economic Activity* 1984:457–510.

Shleifer, A., and R. Vishny. 1997. Limits of arbitrage. *Journal of Finance* 52:35–55.

Shumway, T. 1997. The delisting bias in CRSP data. *Journal of Finance* 52:327–40.

Shumway, T., and V. A. Warther. 1999. The delisting bias in CRSP's Nasdaq data and its implications for the size effect. *Journal of Finance* 54:2361–79.

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

Soliman, M. T. 2008. The use of DuPont analysis by market participants. *Accounting Review* 83:823–53.

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

Sullivan, R., A. Timmermann, and H. White. 1999. Data-snooping, technical trading rule performance, and the bootstrap. *Journal of Finance* 54:1647–91.

———. 2001. Dangers of data mining: The case of calendar effects in stock returns. *Journal of Econometrics* 105:249–86.

Thomas, J. K., and H. Zhang. 2002. Inventory changes and future returns. *Review of Accounting Studies* 7:163–87.

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

Vuolteenaho, T. 2002. What drives firm-level stock returns? *Journal of Finance* 57:233–64.

Wahal, S. 2016. The profitability and investment premium: Pre-1963 evidence. Working Paper.

White, H. 2000. A reality check for data snooping. *Econometrica* 68:1097–126.

Xing, Y. 2008. Interpreting the value effect through the q-theory: An empirical investigation. *Review of Financial Studies* 21:1767–95.

Yan, X., and L. Zheng. 2017. Fundamental analysis and the cross-section of stock returns: A data-mining approach. *Review of Financial Studies* 30:1382–423.