The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns

Jeremiah Green
Pennsylvania State University

John R. M. Hand
UNC-Chapel Hill

X. Frank Zhang
Yale University

We take up Cochrane's (2011) challenge to identify the firm characteristics that provide independent information about average U.S. monthly stock returns by simultaneously including 94 characteristics in Fama-MacBeth regressions that avoid overweighting microcaps and adjust for data-snooping bias. We find that while 12 characteristics are reliably independent determinants in non-microcap stocks from 1980 to 2014 as a whole, return predictability sharply fell in 2003 such that just two characteristics have been independent determinants since then. Outside of microcaps, the hedge returns to exploiting characteristics-based predictability also have been insignificantly different from zero since 2003. (*JEL* G12, G14)

Received 28 January 2015; editorial decision 28 November 2016 by Editor Andrew Karolyi.

In his 2011 American Finance Association Presidential address, John H. Cochrane (2011, 1,060) challenges researchers to identify the firm characteristics that provide *independent* information about average U.S. stock returns. Cochrane issues his challenge because of the "veritable zoo" of hundreds of characteristics that have been presented as statistically significant predictors of the cross-section of returns in the anomalies literature since 1970 (Green, Hand, and Zhang 2013; Hou, Xue, and Zhang 2016). The goal of our

---

Our paper has greatly benefitted from the comments of Andrew Karolyi, two anonymous referees, Jeff Abarbanell, Sanjeev Bhojraj, Matt Bloomfield, John Cochrane, Oleg Grudin, Bruce Jacobs, Bryan Kelly, Juhani Linnainmaa, Ed Maydew, Scott Richardson, Jacob Sagi, and Eric Yeung and workshop participants at the University of Chicago, Cornell University, UNC-Chapel Hill, the Fall 2013 Conference of the Society of Quantitative Analysts, the Fall 2013 Chicago Quantitative Alliance Conference, and the 2014 Jacobs Levy Quantitative Financial Research Conference. Our SAS programs are available online at https://sites.google.com/site/jeremiahrgreenacctg/home. Supplementary data can be found on *The Review of Financial Studies* web site. Send correspondence to John R. M. Hand, Kenan-Flagler Business School, UNC-Chapel Hill, Chapel Hill, NC 27599; telephone: 919-962-3173. E-mail: hand@unc.edu.

doi:10.1093/rfs/hhx019

study is to begin to answer Cochrane's call by using 94 characteristics from CRSP, Compustat, and I/B/E/S; one-month-ahead U.S. returns from 1980 to 2014; and empirical methods that avoid overweighting microcap stocks (Fama and French 2008; Hou, Xue, and Zhang 2016) and adjust for data-snooping biases (Benjamini and Yekutieli 2001; Harvey, Liu, and Zhu 2016).

Our general approach follows that of Fama and French (1992) and others in that we estimate a series of Fama-MacBeth regressions. However, we diverge from the anomalies literature in four ways. First and foremost, we focus on simultaneously including as explanatory variables all 94 of the large set of firm characteristics that we study. This enables us to identify those characteristics that provide independent, nonredundant information. Our method of simultaneously evaluating all 94 firm characteristics is feasible because the mean absolute correlation across characteristics is small (0.07) and because we retain all firm-month observations by setting missing characteristic values to the standardized mean of that month's nonmissing values.

Second, because microcap stocks—those with a market cap below the NYSE 20th percentile—make up only 3% of the value of the U.S. stock market, we focus on the cross-section of stocks in a manner that avoids overweighting microcaps (Hou, Xue, and Zhang 2015). We estimate regressions based on two combinations of sample and method: market-value-weighted least squares (VWLS) on all stocks and ordinary least squares (OLS) on all-but-microcaps. The former approach places the most weight on large cap stocks, while the latter approach emphasizes small, but not microcap, stocks. We pool results across these two approaches as a way of arriving at inferences about the broad cross-section of non-microcap stocks. To demonstrate the influence of microcaps and to allow comparisons with prior research, we also provide results using OLS on all stocks.

Third, mindful of the data-snooping concerns that arise from estimating regressions with a large number of regressors and from using characteristics already identified by prior studies, we evaluate whether a characteristic is statistically significant by using two-tailed p-values on Fama-MacBeth estimated slope coefficients adjusted for false detection rates that take into account dependency across hypothesis tests, in our case regression coefficient p-values (Benjamini and Yekutieli 2001). Our inferences are largely similar if we adopt the recommendation made by Harvey, Liu, and Zhu (2016) that the absolute value of a Fama-MacBeth t-statistic be 3.0 or more to be significant.

Fourth, and finally, we assess the stability of the return-generating process over calendar time in light of the substantial changes in technology, information, trading volume, and the types and capacity of long/short investment vehicles that have occurred in U.S. capital markets since 1980, the start of our data window. We find a marked shift in characteristics-based predictability in 2003, with the hedge portfolio returns to non-microcap stocks after 2003 dropping to zero and the number of independent characteristics falling to just two.

For microcaps, the number of independent characteristics also falls, whereas microcap hedge portfolio returns remain reliably positive after 2003.

We start our empirical analyses by establishing a conventional baseline against which to compare the results of our approach. This baseline comprises the results of estimating for the full window from 1980 to 2014 regressions that contain each of the 94 characteristics as a single independent variable, and regressions that add to a given single characteristic the characteristic equivalents of the factors in the prominent benchmark factor models of Carhart (1997), Fama and French (2015), and Hou, Xue, and Zhang (2015) as explanatory variables.

We find that only one of the 94 characteristics is significant (growth in long-term net operating assets) when the Fama-MacBeth regressions contain a single characteristic using VWLS on all stocks, and that 12 characteristics are significant using OLS on all-but-microcap stocks (asset growth, growth in industry-adjusted sales, percentage change in shares outstanding, growth in inventory, earnings announcement return, growth in book equity, growth in CAPEX, growth in long-term net operating assets, growth in PP&E plus inventory, number of consecutive quarters with earnings higher than the same quarter a year ago, growth in sales less growth in inventory, and standardized unexpected quarterly earnings). Pooling across the two approaches yields a total of 12 univariately significant characteristics in the cross-section of non-microcap stocks, as compared to 30 univariately significant characteristics in the cross-section dominated by microcaps. This indicates that the substantial majority of characteristics put forward as anomalies are not robustly present in a univariate manner over the full 35-year period from 1980 to 2014, especially for non-microcap stocks.

Next, we extend the univariate perspective by singly evaluating characteristics after controlling for the 3 or 4 characteristics associated with prominent benchmark factor models. This yields 6, 4, and 1 significant characteristics out of the 90–91 tested on the Carhart, five-factor, and q-factor models using VWLS on all stocks, and 17, 4, and 4 significant characteristics using OLS on all-but-microcaps. This indicates that within prominent factor-inspired benchmark models, Hou, Xue, and Zhang's q-theory model best captures the independent determinants of average returns in that it yields the fewest incrementally significant firm characteristics beyond those specified by the q-theory model itself.

Central to our study, we then relax the approach of evaluating non–benchmark model characteristics one at a time by simultaneously including all 94 characteristics in the Fama-MacBeth regressions. Given the assumptions of OLS and the low average absolute cross-correlations among characteristics, this enables us to identify the *independent* determinants of returns. We estimate that six characteristics are independent determinants using VWLS on all stocks, and that nine characteristics are independent determinants using OLS on all-but-microcaps. Pooling across the two approaches yields 12 independent characteristics in non-microcaps: book-to-market, cash, change in the number of analysts, earnings announcement return, one-month momentum, change in six-month momentum, number of consecutive quarters with earnings higher than the same quarter a year ago, annual R&D to market cap, return volatility, share turnover, volatility of share turnover, and zero trading days. In comparison, there are 23 independent characteristics in the micro-cap-dominated cross-section of OLS on all stocks.

Taken together, our univariate versus multivariate findings regarding the determinants of average monthly U.S. stock returns over the full window from 1980 to 2014 suggest three main takeaways. First, the fact that the small number of multivariately identified independent characteristics is the same as the small number of univariately significant characteristics (12 each time) leads us to conclude that it is not the case that a few independent characteristics are able to absorb the information in a large number of other characteristics that are univariately significant. Rather, there are just intrinsically few characteristics that independently predict average returns in non-microcap stocks after adjusting for the influence of microcap stocks and data-snooping concerns.

Second, the identity and nature of independent characteristics differs from that of univariately significant characteristics. Based on the fundamental and market classification of anomalies proposed by McLean and Pontiff (2016), 10 of the 12 univariately significant characteristics are fundamental-based (asset growth, growth in industry-adjusted sales, growth in inventory, growth in book equity, growth in CAPEX, growth in long-term net operating assets, growth in PP&E plus inventory, number of consecutive quarters with earnings higher than the same quarter a year ago, growth in sales less growth in inventory, and standardized unexpected quarterly earnings), while two are market-based (percentage change in shares outstanding, and earnings announcement return). In contrast, only one of the 12 multivariately identified independent characteristics is fundamental-based (number of consecutive quarters with earnings higher than the same quarter a year ago), while seven are market-based (change in six-month momentum, earnings announcement return, one-month momentum, return volatility, share turnover, volatility of share turnover, zero trading days). These differences suggest that research that needs to control for the determinants of average returns may do so best using the independent, rather than univariately significant, characteristics we identify.

Third, the characteristics that are independent determinants in non-microcap returns are also typically independent determinants in microcap returns, but not vice versa. Of the 12 independent characteristics in non-microcaps, 10 are independent characteristics in microcaps (OLS on all stocks), while 13 of the 23 independent characteristics in microcaps are not independent characteristics in non-microcaps. It is also the case that 11 of the 12 independent characteristics in non-microcaps lie outside the characteristics that equate to the factors in the Carhart, five-factor, and q-factor benchmark models, with book-to-market being the exception. This suggests that future work may benefit from developing models of average returns that are tailored so as to allow for differences across firm size, and that are broader in the characteristics they include than current benchmark models.

The final set of results in our study extend the preceding analyses by showing that there are marked differences in the number and economic importance of independent characteristics over calendar time, and in a manner that varies by firm size. Specifically, we document that in 2003 there is a sharp kink downward in the magnitude of the hedge portfolio returns to characteristics-based predictability, especially in non-microcap stocks. The mean monthly raw hedge return from exploiting predictability using the full set of 94 characteristics falls from 1.9% (t-statistic = 4.4) before 2003 to 0.5% (t-statistic = 1.1) after 2003 in the value-weighted all-stocks hedge portfolio, and from 2.8% (t-statistic = 5.7) before 2003 to 0.1% (t-statistic = 0.2) after 2003 in the equal-weighted all-but-microcap stocks hedge portfolio. For microcaps, although the mean monthly raw hedge return is significantly positive before and after 2003, it is almost two-thirds smaller after 2003. Limiting the characteristics used to create hedge returns to the subset of 12 previously mentioned yields similar conclusions, as does controlling for the factor returns of the Carhart, five-factor, and q-factor benchmark factor models.

Consistent with the marked drop in hedge returns, we also find that after 2003 only two characteristics are independent determinants of non-microcap returns after 2003, as compared to 12 before 2003. For microcaps, the number of independent determinants also drops, falling two-thirds from 25 to 8. Thus, beyond data snooping and the post-publication decay, and despite a low 0.07 average cross-correlation in characteristics, since 2003 it has been the case that almost no characteristics-based anomalies have existed in the non-microcap cross-section of returns, and fewer than ten have been present in the microcap cross-section. This suggests that the 2003 shift in the economic and statistical significance of characteristics to average monthly U.S returns presents a meaningful challenge to both past and future research.

Overall, our findings that only 12 of 94 characteristics provide independent information in non-microcaps from 1980 to 2014 and that just two characteristics matter after 2003 add even more doubt about findings in prior research already raised by Harvey, Liu, and Zhu's (2016) data-snooping critique and McLean and Pontiff's (2016) finding of post-publication decay. While our results are consistent with the view that the return predictability documented in prior studies suffers from substantial data-snooping problems (Harvey, Liu, and Zhu 2016; Linnainmaa and Roberts 2016), the striking differences in predictability that we document pre- versus post-2003 and by firm size indicate that data snooping is not a complete explanation.

While we do not identify the exact reason for the sudden drop in characteristics-based predictability in 2003, one explanation is that characteristics-based predictability reflects mispricing and that mispricing declines as the cost of exploiting it declines. Consistent with characteristics-based predictability reflecting mispricing, Engelberg, McLean, and Pontiff (2016) conclude from a set of 97 characteristics that anomaly returns are due to biased investor expectations, since they are seven times higher on earnings announcement days, are two times higher on corporate news days, and reliably predict analyst forecast errors. Consistent with the falling cost over time of exploiting mispricing, we note that a number of changes occurred in the information and trading environment from July 2002 to June 2003, including the passing of the Sarbanes-Oxley Act, the accelerating of 10-Q and 10-K filing requirements by the SEC, and the introduction of autoquoting by the NYSE. While the temporal confluence of these changes makes it difficult to causally identify one or more of them with the shifts we observe in the monthly return generating process, we propose that the changes made it cheaper and technologically easier to rapidly implement quantitative long/short trading strategies. Thus, consistent with the costly-limits-to-arbitrage arguments of Shleifer and Vishny (1997), Lesmond, Schill, and Zhou (2004), Chordia, Roll, and Subrahmanyam (2008), Li and Zhang (2010), and Lam and Wei (2011), we conjecture that such changes in the information and trading environment increased arbitrage activity, increased the efficiency of the stock market, and to the degree that the significant pre-2003 pricing of characteristics reflected high costly limits to arbitrage, decreased the influence of characteristics in determining average returns after 2003.

1. Existing Literature on Firm Characteristics and the Cross-Section of Stock Returns

Our study is related to three main areas of research. The first is work that models average returns as a function of firm characteristics or exposure to priced factors. Papers in this area have concluded that factor models based on a small number of characteristics are largely able to explain the portfolio returns formed by individually ranking firms on a large number of characteristics (Hou, Xue, and Zhang 2015; Fama and French 2015, 2016). For example, motivated by q-theory, Hou, Xue, and Zhang (2015) find that a factor model consisting of the excess market return, a small-minus-big size factor, a high-minus-low investment factor, and a high-minus-low return on equity factor performs similarly to a model featuring size, book-to-market, and 12-month momentum but also captures many patterns not explained by the three factors. Hou, Xue, and Zhang therefore propose that their four-factor model is a powerful alternative to other factor models and that any new anomaly variable warrants being benchmarked against their q-factor model to determine if the anomaly truly provides any incremental information. In a related approach, Fama and French (2015, 2016) develop a five-factor model that augments the three-factor model of Fama and French (1993) by adding profitability and investment factors (Li, Livdan, and Zhang 2009; Novy-Marx 2013). Taking a different approach entirely, Light, Maslov, and Rytchkov (2016) treat expected returns as latent variables and develop a procedure that distills 13 factors based on specific characteristics into two new factors, one of which they argue summarizes the information from all anomalies.

Our work also connects to studies that have examined the predictability of returns. Prior studies find that return predictability is strongest among stocks with the highest levels of arbitrage frictions and has declined over time as arbitrage frictions have declined and arbitrage activity has risen. Lesmond, Schill, and Zhou (2004), Hou and Moskowitz (2005), Chordia, Roll, and Subrahmanyam (2008), Li and Zhang (2010), and Lam and Wei (2011) all report that firm characteristics predict returns primarily among stocks with high trading costs or arbitrage frictions. Consistent with this view, Schwert (2003), Green, Hand, and Soliman (2011), Hendershott, Jones, and Menkveld (2011), Chordia, Subrahmanyam, and Tong (2014), and McLean and Pontiff (2016) find that returns to various characteristic-based anomalies decline in response to greater arbitrage activity or as anomalies are made public. Most recently, Novy-Marx and Velikov (2016) observe that while returns to low turnover strategies are robust to adjustments for trading costs, many higher turnover strategies are not.

The third area to which our research speaks is studies that have sought to measure the dimensionality of returns by primarily using firm characteristics, either indirectly by cataloging the characteristics that have been found to be significant by using no- or low-dimensional control methods, or directly by placing medium-sized sets of characteristics into multidimensional models. Illustrating the catalog approach, Subrahmanyam (2010) identifies 50 significant characteristics; Green, Hand, and Zhang (2013) list 330 characteristics in the anomalies literature; Harvey, Liu, and Zhu (2016) itemize 315 such characteristics and/or factors; and Hou, Xue, and Zhang (2016) further expand the set of anomaly-based firm characteristics to 430+. Illustrating instead the multidimensional perspective, Jacobs and Levy (1988) analyze 25 characteristics and find that 10 are significant, and work by Haugen and Baker (1996) reports that 11 of the 40 interrelated characteristics they study are significant.[1] From sets chosen based on published research, Fama and French (2008) and Lewellen (2015) find that 7 out of 7, and 10 out of 15 characteristics, respectively, are significant.[2] Contemporaneous with our study, DeMiguel et al. (2016) use an approach that incorporates firm characteristics into investors' portfolio optimization process by modeling portfolio weights as a function of firm characteristics and find that 6 of the set of 50 characteristics they study

[1] Many of the 40 characteristics used by Haugen and Baker (1996) are highly correlated variants of a few constructs, with the likely result that their analysis is based on fewer than 40 independent characteristics.

[2] Fama and French (2008) orient their analysis around whether the significance of characteristics is robust across firm size. Lewellen (2015) focuses his work on the cross-sectional dispersion and out-of-sample predictive ability of the stock return forecasts that he extracts from the 15 characteristics he studies.

are jointly significant predictors. Taking a factor pricing view of anomalies, Stambaugh and Yuan (2016) find for 1967–2013 that a four-factor model that contains two mispricing factors in addition to market and size accommodates a large set of anomalies.

Notwithstanding such prior research, the challenge issued by Cochrane (2011) remains: only a fraction of the 430+ characteristics in the anomalies literature have been studied in a way that identifies which firm characteristics provide *independent* information about average returns. Our paper responds to Cochrane's challenge by simultaneously evaluating a much larger set of characteristics than prior work, and in that larger set estimating not just the number, identity, and economic significance of the independent determinants, but newly highlighting the degree to which results obtained over the full data period from 1980 to 2014 vary over time, particularly before and after 2003.

2. Data Set Construction and Correlations among Firm Characteristics

2.1 Data set aligned in calendar time

As the chief goal of our paper is to empirically identify the independent determinants of average returns by regressing one-month-ahead returns on a large number of characteristics all at the same time, we face design decisions as to which and how many characteristics to include; how to combine characteristics across firms, time periods, and databases; and how to address missing data. To maximize the ability of others to replicate and/or expand our work, we seek to transparently detail the choices we make in selecting, aligning, and coding firm characteristics. In doing so, we recognize that some choices we make distance us from the exact research designs, characteristics definitions, and sample periods used in the papers in which the firm characteristics were originally identified.

We initially selected 102 of the 330 characteristics listed in Green, Hand, and Zhang (2013), requiring that each characteristic be entirely calculable from CRSP, Compustat, and/or I/B/E/S data.[3] Our data cover the 35-year period from January 1980 to December 2014. We begin in 1980 because most characteristics only become robustly available in that year. The 102 characteristics we select are listed in Table 1. The Appendix provides the details of each characteristic, including a description of how it is calculated and the author(s), journal, and year of publication or working paper of the underlying academic study. The characteristics span both highly and sparsely cited papers, published and working papers, and publication dates between 1977 and 2016. On occasion, more than one characteristic comes from a given paper.

We begin our data creation with all firms with common stock on the NYSE, AMEX, or NASDAQ that have a month-end market value on CRSP and a

[3] We also restricted the firm characteristics to main-effect signals. We do not include characteristics that are interactions between other characteristics.

Table 1
Listing of firm characteristics used in the study

The Appendix provides the source and definition of each characteristic.

Acronym	Firm characteristic	Acronym	Firm characteristic
absacc	Absolute accruals	*divo*	Dividend omission
acc	Working capital accruals	*dolvol*	Dollar trading volume
aeavol	Abnormal earnings announcement volume	*dy*	Dividend to price
age	# years since first Compustat coverage	*ear*	Earnings announcement return
agr	Asset growth	*egr*	Growth in common shareholder equity
baspread	Bid-ask spread	*ep*	Earnings to price
beta	Beta	*fgr5yr*	Forecasted growth in 5-year EPS
betasq	Beta squared	*gma*	Gross profitability
bm	Book-to-market	*grCAPX*	Growth in capital expenditures
bm_ia	Industry-adjusted book to market	*grltnoa*	Growth in long-term net operating assets
cash	Cash holdings	*herf*	Industry sales concentration
cashdebt	Cash flow to debt	*hire*	Employee growth rate
cashpr	Cash productivity	*idiovol*	Idiosyncratic return volatility
cfp	Cash-flow-to-price ratio	*ill*	Illiquidity
cfp_ia	Industry-adjusted cash-flow-to-price ratio	*indmom*	Industry momentum
chatoia	Industry-adjusted change in asset turnover	*invest*	Capital expenditures and inventory
chcsho	Change in shares outstanding	*IPO*	New equity issue
chempia	Industry-adjusted change in employees	*lev*	Leverage
chfeps	Change in forecasted EPS	*lgr*	Growth in long-term debt
chinv	Change in inventory	*maxret*	Maximum daily return
chmom	Change in 6-month momentum	*mom12m*	12-month momentum
chnanalyst	Change in number of analysts	*mom1m*	1-month momentum
chpmia	Industry-adjusted change in profit margin	*mom36m*	36-month momentum
chtx	Change in tax expense	*mom6m*	6-month momentum
cinvest	Corporate investment	*ms*	Financial statement score
convind	Convertible debt indicator	*mve*	Size
currat	Current ratio	*mve_ia*	Industry-adjusted size
depr	Depreciation / PP&E	*nanalyst*	Number of analysts covering stock
disp	Dispersion in forecasted EPS	*nincr*	Number of earnings increases
divi	Dividend initiation	*operprof*	Operating profitability
orgcap	Organizational capital	*roeq*	Return on equity
pchcapx_ia	Industry adjusted % change in capital expenditures	*roic*	Return on invested capital
pchcurrat	% change in current ratio	*rsup*	Revenue surprise
pchdepr	% change in depreciation	*salecash*	Sales to cash
pchgm_pchsale	% change in gross margin - % change in sales	*saleinv*	Sales to inventory
pchquick	% change in quick ratio	*salerec*	Sales to receivables
pchsale_pchinvt	% change in sales - % change in inventory	*secured*	Secured debt
pchsale_pchrect	% change in sales - % change in A/R	*securedind*	Secured debt indicator
pchsale_pchxsga	% change in sales - % change in SG&A	*sfe*	Scaled earnings forecast
pchsaleinv	% change sales-to-inventory	*sgr*	Sales growth
pctacc	Percent accruals	*sin*	Sin stocks
pricedelay	Price delay	*SP*	Sales to price
ps	Financial statements score	*std_dolvol*	Volatility of liquidity (dollar trading volume)

(*continued*)

Table 1
Continued

Acronym	Firm characteristic	Acronym	Firm characteristic
quick	Quick ratio	*std_turn*	Volatility of liquidity (share turnover)
rd	R&D increase	*stdacc*	Accrual volatility
rd_mve	R&D to market capitalization	*stdcf*	Cash flow volatility
rd_sale	R&D to sales	*sue*	Unexpected quarterly earnings
realestate	Real estate holdings	*tang*	Debt capacity/firm tangibility
retvol	Return volatility	*tb*	Tax income to book income
roaq	Return on assets	*turn*	Share turnover
roavol	Earnings volatility	*zerotrade*	Zero trading days

nonmissing value for common equity in their annual financial statements. We then integrate data across Compustat, I/B/E/S, and CRSP and compute and align characteristics in calendar time. Since Green, Hand, and Zhang (2013) report that 57% of the 330 characteristics they list are evaluated by the original authors through the lens of monthly returns, we remeasure and realign characteristics every month. For each month t's return we calculate characteristics as they were at the end of month $t-1$, assuming that annual accounting data are available at the end of month $t-1$ if the firm's fiscal year ended at least six months before the end of month $t-1$, and that quarterly accounting data are available at the end of month $t-1$ if the fiscal quarter ended at least four months before the end of month $t-1$. I/B/E/S and CRSP data are aligned in calendar time using the I/B/E/S statistical period date and the CRSP monthly or daily end date.

While monthly updating is consistent with the portfolio rebalancing approach used by many quantitative institutional investors, we recognize that some practitioners update their data as often as every minute, or as infrequently as every 12 months. We choose monthly updating because we view it as a reasonable trade-off between the lower transactions and trading costs of longer frequencies, and the benefit of greater timeliness (Novy-Marx and Velikov 2016). Our choice means that those characteristics in our data set that come from studies with a shorter-than-monthly frequency will be less timely than in the original studies, while those that come from studies using longer frequencies will be more timely. Such slippage may lower our ability to detect the incremental significance of individual characteristics, but it also reduces the chances that we will identify return predictability that does not survive the trading-cost effects of high turnover strategies (Novy-Marx and Velikov 2016).

We take monthly stock returns from CRSP and include delisting returns per Shumway and Warther (1999). We delete 20 observations that have a monthly return less than -100% and set blank values of analyst following, *nanalyst*, to zero. I/B/E/S is the most restrictive of our databases in its coverage of firms and availability over time, so we only use I/B/E/S-based characteristics starting in January 1989, when I/B/E/S's more expansive coverage began.[4]

[4] The firm characteristics that employ I/B/E/S data are *chnanalyst*, *chfeps*, *disp*, *fgr5yr*, *nanalyst*, *sfe*, and *sue*.

Following prior work such as Hou, Xue, and Zhang (2015) and Fama and French (2015), we delineate firm size groupings based on their monthly NYSE-based percentiles. We label stocks with a market cap greater than the median NYSE stock at the end of month $t-1$ as big, stocks below the median and above the 20th percentile as small, stocks with values less than or equal to the 20th percentile as microcap, and stocks excluding microcaps as all-but-microcap.

Since in our key analyses we identify the independent determinants of average returns by regressing one-month-ahead returns on all characteristics simultaneously, in such regressions we avoid discarding a characteristic simply because its value is missing in one or more firm-months. Just 4% of characteristics have full data with no missing observations from January 1980 to December 2014. The approach we take to retaining as much characteristic information as possible is to first winsorize all characteristics at the 1st and 99th percentiles of their monthly distributions and standardize each to have a zero mean and unit standard deviation. We then set missing characteristic values to the characteristic's post-standardized monthly mean of zero.[5] For each firm characteristic, we report in Table 2 the number of our data set of 1,933,898 firm-month observations over the period from January 1980 to December 2014 with nonmissing data, and the percentage missing before being reset to zero. We note that the characteristics with the largest fraction of missing data tend to be those that use I/B/E/S analyst information (*chnanalyst, chfeps*, *disp*, *fgr5yr*, *nanalyst*, *sfe*, and *sue*) or that use sparsely populated Compustat data (*rd_mve*, *rd_sale*, and *realestate*).

2.2 Correlations among firm characteristics

We gauge the potential for multicollinearity concerns in our Fama-MacBeth regressions by measuring the cross-correlations among our initial set of 102 firm characteristics. While multicollinearity does not lead to bias in estimated slope coefficients, it does increase their standard errors. Thus, to the extent we are able to exclude characteristics with very large cross-correlations with other characteristics because they are mechanically or economically related—for example, beta and beta squared—we expect to be able to more powerfully identify the independent determinants of average returns when we estimate regressions that simultaneously contain very large numbers of characteristics, especially if the number of highly collinear characteristics is found to be small.

We therefore calculate the variance inflation factors (VIFs) of each characteristic, since VIFs summarize the extent to which a given characteristic is explained by a linear combination of all other characteristics (Greene 2011). Panel A of Table 3 describes the distribution of VIFs. While the median VIF of 1.8 is not large, VIFs are right skewed in that 10% of characteristics have

---

[5] The approach of replacing missing independent variables with their sample means is known as the zero-order regression method (Wilks 1932). Under multivariate normality of the dependent and independent variables, the zero-order method is generally expected to lead to unbiased slope coefficient estimates (Afifi and Elashoff 1966).

Table 2
Available data and missing observations for firm characteristics

Firm characteristic	# obs.	% miss.	Firm characteristic	# obs.	% miss.	Firm characteristic	# obs.	% miss.
absacc	1,662,391	14%	*egr*	1,801,782	7%	*pchsale_pchxsga*	1,499,205	22%
acc	1,662,391	14%	*ep*	1,933,898	0%	*pchsaleinv*	1,409,289	27%
aeavol	1,710,488	12%	*fgr5yr*	760,194	61%	*pctacc*	1,662,379	14%
age	1,933,898	0%	*gma*	1,797,401	7%	*pricedelay*	1,913,115	1%
agr	1,801,938	7%	*grcapx*	1,614,361	17%	*ps*	1,801,971	7%
baspread	1,933,853	0%	*grltnoa*	1,347,837	30%	*quick*	1,856,862	4%
beta	1,913,144	1%	*herf*	1,933,888	0%	*rd*	1,801,971	7%
betasq	1,913,144	1%	*hire*	1,797,477	7%	*rd_mve*	931,907	52%
bm	1,933,898	0%	*idiovol*	1,913,144	1%	*rd_sale*	918,493	53%
bm_ia	1,933,898	0%	*ill*	1,871,487	3%	*realestate*	800,058	59%
cash	1,712,248	11%	*indmom*	1,933,741	0%	*retvol*	1,933,841	0%
cashdebt	1,864,432	4%	*invest*	1,740,556	10%	*roaq*	1,720,520	11%
cashpr	1,913,829	1%	*ipo*	1,933,898	0%	*roavol*	1,453,242	25%
cfp	1,775,090	8%	*lev*	1,928,357	0%	*roeq*	1,720,115	11%
cfp_ia	1,775,090	8%	*lgr*	1,795,724	7%	*roic*	1,844,919	5%
chatoia	1,657,109	14%	*maxret*	1,933,897	0%	*rsup*	1,709,439	12%
chcsho	1,801,190	7%	*mom12m*	1,791,487	7%	*salecash*	1,917,657	1%
chempia	1,797,477	7%	*mom1m*	1,933,898	0%	*saleinv*	1,526,455	21%
chfeps	1,000,523	48%	*mom36m*	1,502,039	22%	*salerec*	1,865,648	4%
chinv	1,754,017	9%	*mom6m*	1,875,749	3%	*secured*	1,132,250	41%
chmom	1,791,487	7%	*ms*	1,723,698	11%	*securedind*	1,933,898	0%
chnanalyst	1,455,603	25%	*mve*	1,933,898	0%	*sfe*	988,655	49%
chpmia	1,774,849	8%	*mve_ia*	1,933,898	0%	*sgr*	1,778,807	8%
chtx	1,688,004	13%	*nanalyst*	1,480,584	23%	*sin*	1,933,898	0%
cinvest	1,684,236	13%	*nincr*	1,723,698	11%	*sp*	1,928,325	0%
convind	1,933,898	0%	*operprof*	1,797,245	7%	*std_dolvol*	1,868,232	3%
currat	1,866,803	3%	*orgcap*	1,422,289	26%	*std_turn*	1,873,288	3%
depr	1,849,276	4%	*pchcapx_ia*	1,751,021	9%	*stdacc*	1,213,569	37%
disp	825,468	57%	*pchcurrat*	1,732,739	10%	*stdcf*	1,213,569	37%
divi	1,801,971	7%	*pchdepr*	1,714,330	11%	*sue*	1,711,920	11%
divo	1,801,971	7%	*pchgm_pchsal*	1,778,662	8%	*tang*	1,853,308	4%
dolvol	1,860,698	4%	*pchquick*	1,721,978	11%	*tb*	1,703,039	12%
dy	1,928,742	0%	*pchsale_pchinvt*	1,427,946	26%	*turn*	1,861,339	4%
ear	1,722,142	11%	*pchsale_pchrect*	1,727,611	11%	*zerotrade*	1,871,513	3%

This table shows the number of firm-month observations (# obs.) over January 1980–December 2014 with sufficient data to calculate a given firm characteristic, and the percent of firm-months with missing observations (% miss.). The sample is all common stocks on the NYSE, AMEX, and NASDAQ exchanges with available annual and quarterly Compustat accounting data and CRSP stock return data. Analyst data are from I/B/E/S. Firm-month observations are created using monthly stock returns, including delisting returns, for each month t. Stock returns in month t are matched with the accounting data most recently available as of the end of month $t-1$, under the assumption that that annual accounting data are available at the end of month $t-1$ if the firm's fiscal year ended at least six months before the end of month $t-1$, and that quarterly accounting data are available at the end of month $t-1$ if the fiscal quarter ended at least four months before the end of month $t-1$.

a VIF $\geqslant$ 9.8. We therefore seek to mitigate the effects of multicollinearity in our key regressions that simultaneously include all characteristics by removing the eight characteristics that are most strongly related to other characteristics (*betasq*, *dolvol*, *lgr*, *maxret*, *mom6m*, *pchquick*, *quick*, and *stdacc*), each of which has a VIF >7. We then use the remaining 94 characteristics throughout our analyses.[6]

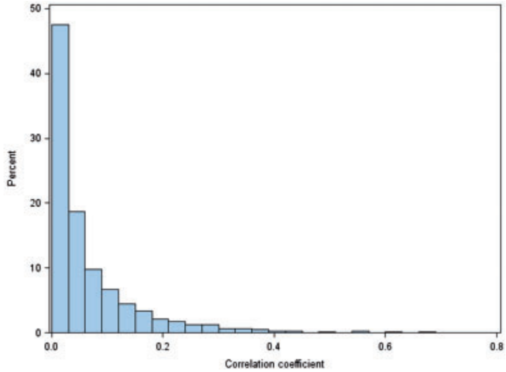
[6] In untabulated analyses, we vary the VIF cutoff at which we exclude a characteristic from 5 to 10 and find results similar to those obtained using our default VIF cutoff of 7. Our results are also similar if we do not exclude any characteristics based on VIF cutoffs.

Table 3
Descriptive statistics on the degree of cross-correlation among firm characteristics

A. VIFs

	# characteristics	Min.	10th pct.	Median	Mean	90th pct.	Max.
Before removing VIF >7	102	1	1.1	1.8	3.8	9.8	22.1
After removing VIF >7	94	1	1.1	1.6	2.1	3.9	6.9
B. Absolute cross-correlations							
Before removing VIF >7	102	0	0.004	0.03	0.07	0.18	0.95
After removing VIF >7	94	0	0.004	0.03	0.07	0.18	0.77

C. Distribution of the absolute cross-correlations among characteristics after removal of characteristics with VIFs >7



This table presents descriptive statistics on the degree of cross-correlation among firm characteristics. VIFs, variance inflation factors. Panel A shows statistics from the distribution of variance inflation factors (VIFs) where the VIF for each characteristic is calculated as $1/(1-R^2)$, with R^2 being that obtained from regressing each characteristic on all the other characteristics in a pooled regression. After examining the VIFs we remove 8 characteristics with high VIFs such that the resulting maximum VIF $\leqslant 7$. The characteristics removed are *betasq, dolvol, lgr, maxret, mom6m, pchquick, quick, stdacc*. Panel B describes the distribution of the absolute value of Pearson correlations between all pairs of characteristics using the pooled sample of firm-month observations. Panel C graphs the distribution of the absolute cross-correlations after the 8 cross-correlated characteristics with large VIFs are removed.

In panel B of Table 3 we report the key percentiles of the absolute cross-correlations among characteristics before and after removing the eight characteristics with VIFs greater than seven. In both cases, the mean and median absolute cross-correlations are quite low at 0.07 and 0.03, respectively.[7] Not surprisingly, after removing characteristics with VIFs >7, the largest cross-correlations also decline. Panel C shows the distribution of the absolute cross-correlations among the 94 nonhighly cross-correlated characteristics, making clear that 90% of absolute cross-correlations are below 0.2.

---

[7] The mean absolute cross-correlation among characteristics is similarly small if cross-correlations are calculated after missing characteristic values are reset to each characteristic's monthly mean. (Panels B and C use characteristics that may have missing values, while the pooled regression to calculate VIFs uses characteristics that have had missing values set to their monthly mean value.)

3. Empirical Methods and Results

We estimate standard Fama and MacBeth (1973) regressions over the period from 1980 to 2014 to determine how many and which of our set of 94 firm characteristics are statistically significant determinants of one-month-ahead returns, most especially when they are all simultaneously included in the regressions. In assessing whether a characteristic is reliably related to average returns, we are mindful of the inferential biases that can arise in our tests from data snooping or from overweighting of microcap stocks, which make up on average only about 3% of the market cap of the NYSE-Amex-NASDAQ universe (Fama and French 2008; Hou, Xue, and Zhang 2016).

We seek to avoid the latter bias by focusing the majority of our analyses on the non-microcap cross-section of firms, implemented in two ways: by applying value-weighted least squares (VWLS) to all stocks and by using OLS on all-but-micro-cap stocks. The VWLS approach places most weight on large cap stocks, while the OLS approach emphasizes smaller (but not microcap) stocks. This approach is similar to that used when creating portfolios that seek to minimize the influence of microcap stocks (Hou, Xue, and Zhang 2016). In arriving at inferences about non-microcaps, we pool results across our two approaches, while for reference purposes we arrive at inferences inclined toward microcaps using OLS on all stocks.

In our setting, we face the risk that we will falsely reject the null for some characteristics purely by chance, given our goal of assessing the significance of individual characteristics rather than the joint significance of all characteristics as a single set. This multiple testing concern is common in other research literatures and recently has been raised in the context of return prediction by Harvey, Liu, and Zhu (2016). While the concern they address is that new return predictors may be falsely deemed statistically significant because tests use the same data sets, our concern is that we risk falsely concluding that a return predictor is significant because we simultaneously include a large number of characteristics in the same regression, and/or because we estimate several regressions using the same data set. We seek to do this by evaluating the statistical significance of an estimated coefficient on a characteristic using two-tailed p-values adjusted for false detection rates in a manner that takes into account the dependency among p-values across a given set of estimated regressions (Benjamini and Yekutieli 2001).[8] We refer to such p-values as DFDR p-values and require that a DFDR p-value be less than or equal to 0.05 for the associated parameter estimate to be considered statistically significant. The DFDR procedure controls for the false discovery rate by assuming there is some expected proportion of test statistics that will be erroneously judged

[8] To adjust tests of significance to take these issues into account, other literatures have developed tools including family-wise error rate (FWER) adjustments and false discovery rates (FDR). Due to the very conservative nature of FWER adjustments, we choose to use FDR adjustments to p-values when making inferences about statistical significance from our regressions.

as significant, which prior work suggests is likely to be the case for the large set of anomalies that we study. As shown by Benjamini and Yekutieli (2001), applying the DFDR procedure yields p-values that result in a false discovery rate less than or equal to the desired confidence level.

The DFDR approach is implemented by sorting the p-values from the hypothesis tests (in our case the t-statistics on the coefficients in the regressions) so that with m test statistics the p-values are ordered $p_{(1)} \leqslant p_{(2)} \leqslant p_{(3)} \leqslant \ldots \leqslant p_{(\mathrm{m})}$. For a desired level of significance q (0.05), define for the set of p-values

$$k = max\left\{ i : p_{(i)} \leqslant \frac{iq}{m\sum_{i=1}^{m}\frac{1}{i}} \right\}. \tag{1}$$

This procedure steps through each p-value, starting with the most significant (lowest p-value), until k is reached such that the next p-value no longer meets the criteria. Hypotheses 1 through k are selected as significant or, more precisely, the null hypotheses are rejected. The DFDR p-values are given by the following equation:

$$DFDR\, p_{(i)} = \begin{cases} \left[\sum_{i=1}^{m}\frac{1}{i}\right] p_m, & i = m \\ \min\left(DFDR\ p_{i+1}, \left[\sum_{i=1}^{m}\frac{1}{i}\right]\frac{m}{i}p_i\right), & i < m \end{cases}. \tag{2}$$

Following the approach of Harvey, Liu, and Zhu (2016) we also report the number of t-statistics that are greater than or equal to 3.0 in absolute value, where t-statistics are computed from the time-series of mean monthly coefficient estimates, given Newey-West (1994) adjustments over 12 monthly lags.

3.1 Baseline models of average returns

We begin by establishing a conventional baseline against which to compare the results of our main approach of simultaneously including all 94 characteristics in the Fama-MacBeth regressions. This baseline comprises the results of estimating regressions that contain each of the 94 characteristics as a single independent variable, followed by regressions that add to the characteristic equivalents of the factors in the prominent benchmark factor models of Carhart (1997), Fama and French (2015), and Hou, Xue, and Zhang (2015) one and only one of the characteristics that are not already in the benchmark models, using data for 1980–2014 as a whole.[9]

[9] Characteristics that require I/B/E/S data are used beginning in January 1989. Therefore, the Fama-MacBeth slope coefficients are the means of the time series of monthly coefficients when available from January 1980 through December 2014. For I/B/E/S characteristics, the number of time series coefficients is smaller than for non-I/B/E/S characteristics and are therefore not directly comparable. In the multivariate regressions, I/B/E/S-based characteristics are omitted from the regressions prior to January 1989, so the possible number of significant characteristics is smaller from January 1980 through December 1989.

In Table 4, Column A, we detail the results of placing each characteristic individually into its own Fama-MacBeth regression, followed in Columns B–D by the results of adding each characteristic singly to the characteristics called for by a given benchmark model. The benchmark characteristics are size, book-to-market, and 12-month momentum for the Carhart model; size, book-to-market, investment, and operating profitability for the five-factor model; and size, investment, and quarterly return-on-equity for the q-factor model. In each column we report results for all stocks using WLS, all-but-microcap stocks using OLS, and all stocks using OLS. To help guide the reader's eye to which coefficients are significant, given the large number of characteristics studied, we bold the t-statistic for coefficient estimates with a DFDR p-value $\leqslant 0.05$. The number of coefficient estimates with a DFDR p-value $\leqslant$ 0.05 and the number of absolute t-statistics $\geqslant 3.0$ are shown in the first and second lines of Table 4.

As judged by coefficient estimates with DFDR p-values $\leqslant 0.05$, in Column A we find that just one of 94 characteristics is significant in univariate regressions for non-microcap firms as measured by VWLS on all stocks (growth in long-term net operating assets), while 12 characteristics are significant using OLS on all-but-microcaps (asset growth, growth in industry-adjusted sales, growth in shares outstanding, growth in inventory, earnings announcement return, growth in book equity, growth in CAPEX, growth in long-term net operating assets, growth in PP&E plus inventory, number of consecutive quarters with earnings higher than the same quarter a year ago, growth in sales less growth in inventory, and standardized unexpected quarterly earnings). Pooling across the two measures of the cross-section of non-microcaps yields a total of 12 univariately significant characteristics, in identity the same as those found by using OLS on all-but-microcaps. Underscoring the observation that the more heavily that smaller stocks are weighted, the more characteristics are reliably univariately significant, we note that there are 30 significant characteristics when microcaps are given the same regression weighting as larger stocks (all stocks, OLS).[10] The results of Column A therefore indicate that the great majority of characteristics put forward as anomalies in the prior literature are not robustly present in a univariate manner over the period from 1980 to 2014 taken as a whole. Moreover, where they are reliably present they are concentrated in microcap stocks.

We note that our assessments about the number and identity of significant coefficients rely on our ability to determine which coefficients are likely to be driven by in-sample data overfitting. Our inference that a large number of coefficients are insignificant differs from prior research because we de-emphasize microcaps and because we adjust statistical significance

[10] We also note that a comparison of the number of significant characteristics based on DFDR p-values = 0.05 versus absolute t-statistics = 3.0 indicates that the former approach yields fewer significant characteristics than the latter, reflecting the goal in DFDR of taking into account dependency across multiple coefficient p-values.

Table 4
Fama-MacBeth regressions of monthly stock returns on each of the 94 firm characteristics studied one at a time

In Column A, no other characteristics are controlled for, and in Columns B–D, the charactertistics version of promiment benchmark factor models are controlled for. Coefficients with a two-tailed DFDR p-value $\leqslant$ 0.05 are shown with a bold t-statistic. The data window is January 1980–December 2014. FM, Fama-MacBeth.

	(A) Single characteristic, No benchmark model controls						(B) Single characteristic, Carhart benchmark model controls						(C) Single characteristic, 5-factor benchmark model controls						(D) Single characteristic, q-factor benchmark model controls					
	All stocks, VWLS		All-but-microcap, OLS		All stocks, OLS		All stocks, VWLS		All-but-microcap, OLS		All stocks, OLS		All stocks, VWLS		All-but-microcap, OLS		All stocks, OLS		All stocks, VWLS		All-but-microcap, OLS		All stocks, OLS	
	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.
# DFDR p $\leqslant$ 0.05		1		12		30		5		15		33		6		4		20		2		3		19
# $\lvert t$-stats$\rvert \geqslant$ 3.0		3		16		34		14		20		36		8		7		25		3		5		26
agr	−0.21	−3.1	−0.26	**−4.6**	−0.42	**−8.1**	−0.19	**−3.5**	−0.24	**−5.1**	−0.35	**−8.3**												
bm	0.17	1.1	0.14	1.8	0.32	**4.8**													0.13	0.9	0.10	1.4	0.19	2.6
mom12m	0.30	2.4	0.20	1.9	0.21	2.1							0.26	2.2	0.20	2.0	0.22	2.5	0.27	2.3	0.20	1.9	0.21	2.4
mve	−0.10	−1.0	−0.05	−1.0	−0.25	−2.3																		
operprof	0.08	1.5	0.06	2.1	0.05	1.4	0.10	2.3	0.08	2.7	0.07	1.9							0.09	1.9	0.07	2.6	0.07	2.4
roeq	0.18	2.5	0.14	2.5	0.20	2.3	0.21	3.2	0.14	2.9	0.23	**3.6**	0.20	**3.4**	0.15	3.0	0.26	**4.0**						
absacc	−0.03	−0.3	−0.08	−1.3	−0.06	−0.8	−0.07	−0.9	−0.08	−1.9	−0.08	−1.3	−0.01	−0.1	−0.04	−1.0	−0.06	−1.1	0.02	0.2	−0.03	−0.6	−0.05	−1.0
acc	−0.20	−2.5	−0.11	−2.5	−0.16	−2.6	−0.12	−2.0	−0.08	−2.2	−0.12	−2.5	−0.17	−2.5	−0.08	−2.2	−0.07	−1.6	−0.18	−2.4	−0.09	−2.3	−0.09	−2.1
aeavol	0.02	0.3	0.00	0.0	0.08	**4.7**	−0.02	−0.4	−0.01	−0.6	0.04	2.4	0.01	0.2	0.00	0.1	0.06	**3.6**	0.01	0.2	0.00	0.1	0.05	3.1
age	0.02	0.2	0.08	1.2	0.13	1.8	0.06	0.9	0.10	1.8	0.21	**4.5**	−0.03	−0.3	0.02	0.3	0.14	2.8	0.00	0.0	0.04	0.5	0.17	3.0
baspread	−0.42	−1.2	−0.20	−1.5	0.10	0.6	−0.61	−2.2	−0.26	−2.2	−0.01	−0.1	−0.40	−1.3	−0.17	−1.3	0.00	0.0	−0.41	−1.3	−0.17	−1.2	0.01	0.1
beta	−0.07	−0.4	−0.10	−0.7	−0.11	−0.8	−0.15	−1.0	−0.14	−1.1	−0.08	−0.6	−0.03	−0.2	−0.06	−0.4	−0.04	−0.3	−0.03	−0.2	−0.06	−0.4	−0.04	−0.3
bm_ia	0.04	0.8	0.00	−0.1	0.01	0.2	0.03	0.7	0.01	0.2	0.02	0.4	0.04	0.9	0.01	0.3	0.02	0.4	0.05	1.1	0.01	0.1	0.00	0.1
cash	0.11	1.1	0.04	0.3	0.10	1.0	0.11	1.4	0.05	0.5	0.14	1.5	0.17	1.9	0.10	1.0	0.17	1.8	0.15	1.6	0.08	0.8	0.14	1.5
cashdebt	0.24	1.9	0.08	2.0	0.05	0.7	0.29	**3.4**	0.11	3.2	0.06	1.1	0.29	3.3	0.10	2.8	0.08	1.5	0.25	2.3	0.07	1.9	0.06	1.2
cashpr	−0.08	−1.7	−0.10	−2.3	−0.14	−3.2	−0.03	−1.0	−0.05	−1.8	−0.04	−1.2	−0.06	−1.8	−0.06	−2.2	−0.03	−1.1	−0.07	−1.7	−0.08	−2.2	−0.09	−1.9
cfp	0.14	1.5	0.11	1.8	0.13	1.7	0.08	1.3	0.09	2.2	0.11	2.0	0.07	1.0	0.07	1.7	0.12	2.1	0.05	0.6	0.06	1.2	0.10	1.6
cfp_ia	−0.01	−0.2	0.00	0.1	0.04	1.0	−0.01	−0.3	0.02	0.5	0.04	1.2	0.00	0.1	0.02	0.6	0.04	1.2	0.00	0.0	0.01	0.3	0.03	0.9
chatoia	0.07	1.6	0.08	**3.9**	0.08	**5.0**	0.06	1.5	0.07	**3.6**	0.07	**4.7**	0.02	0.5	0.04	1.9	0.05	2.9	0.02	0.5	0.04	1.8	0.04	2.6
chcsho	−0.12	−3.0	−0.17	**−3.4**	−0.26	**−6.5**	−0.10	−3.0	−0.13	**−3.4**	−0.18	**−5.8**	−0.07	−2.5	−0.05	−2.1	−0.10	−3.3	−0.09	−2.8	−0.07	−2.6	−0.11	**−3.8**
chempia	−0.02	−0.5	−0.06	−1.3	−0.16	**−4.1**	−0.01	−0.2	−0.05	−1.2	−0.12	−3.3	0.07	1.5	0.05	1.7	0.01	0.2	0.06	1.4	0.06	1.8	0.01	0.2
chfeps	0.06	1.8	0.06	2.1	0.11	**4.8**	0.03	1.0	0.05	2.0	0.10	**4.7**	0.05	1.8	0.05	2.0	0.11	**4.7**	0.06	1.8	0.06	2.1	0.11	**4.9**
chinv	−0.15	−3.2	−0.14	**−4.3**	−0.24	**−6.5**	−0.11	−2.9	−0.11	**−4.3**	−0.19	**−5.8**	−0.10	−2.3	−0.07	−2.8	−0.11	**−3.6**	−0.08	−2.0	−0.07	−2.6	−0.11	**−3.8**
chmom	−0.30	−3.0	−0.12	−1.9	−0.20	**−3.5**	−0.27	−3.3	−0.11	−1.9	−0.17	−3.1	−0.31	**−3.7**	−0.12	−2.4	−0.20	**−4.0**	−0.30	**−3.4**	−0.11	−2.1	−0.18	**−3.6**
chnanalyst	−0.01	−0.6	−0.06	−2.1	−0.05	−2.4	−0.01	−0.5	−0.06	−2.4	−0.04	−2.1	0.00	−0.1	−0.05	−1.8	−0.03	−1.2	0.00	−0.2	−0.05	−1.8	−0.03	−1.5
chpmia	0.01	0.2	0.01	0.3	0.02	0.5	0.00	0.1	0.01	0.2	0.01	0.4	0.00	0.1	0.01	0.2	0.02	0.6	0.01	0.3	0.01	0.2	0.02	0.8
chtx	0.08	1.7	0.06	1.7	0.13	**5.3**	0.05	1.3	0.03	0.9	0.11	**4.8**	0.12	**3.4**	0.09	3.1	0.16	**6.9**	0.08	2.0	0.05	1.5	0.12	**5.6**
cinvest	0.02	0.7	−0.03	−1.2	0.01	0.6	0.01	0.3	−0.03	−1.3	0.01	0.5	0.00	0.1	−0.04	−1.6	0.00	−0.1	−0.01	−0.2	−0.04	−1.9	−0.02	−1.0
convind	−0.17	−1.7	−0.13	−1.4	−0.39	**−4.0**	−0.15	−1.8	−0.10	−1.2	−0.28	−2.9	−0.12	−1.3	−0.08	−0.9	−0.28	−2.6	−0.10	−1.2	−0.04	−0.5	−0.23	−2.3
currat	−0.09	−1.5	−0.10	−2.2	−0.03	−0.8	−0.08	−1.6	−0.08	−2.1	−0.03	−1.1	−0.04	−1.0	−0.05	−1.4	−0.01	−0.2	−0.07	−1.4	−0.07	−1.8	−0.02	−0.9

(*continued*)

Table 4
Continued

	(A) Single characteristic, No benchmark model controls						(B) Single characteristic, Carhart benchmark model controls						(C) Single characteristic, 5-factor benchmark model controls						(D) Single characteristic, *q*-factor benchmark model controls					
	All stocks, VWLS		All-but-microcap, OLS		All stocks, OLS		All stocks, VWLS		All-but-microcap, OLS		All stocks, OLS		All stocks, VWLS		All-but-microcap, OLS		All stocks, OLS		All stocks, VWLS		All-but-microcap, OLS		All stocks, OLS	
	FM coef.	*t*-stat.	FM coef.	*t*-stat.	FM coef.	*t*-stat.	FM coef.	*t*-stat.	FM coef.	*t*-stat.	FM coef.	*t*-stat.	FM coef.	*t*-stat.	FM coef.	*t*-stat.	FM coef.	*t*-stat.	FM coef.	*t*-stat.	FM coef.	*t*-stat.	FM coef.	*t*-stat.
depr	0.02	0.2	0.01	0.1	0.10	1.4	0.01	0.2	0.01	0.3	0.09	1.7	0.03	0.4	0.02	0.6	0.08	1.4	0.05	0.6	0.03	0.6	0.07	1.3
disp	−0.07	−1.1	−0.05	−1.2	−0.08	−2.4	−0.07	−1.4	−0.06	−1.6	−0.10	**−3.7**	−0.05	−1.0	−0.05	−1.2	−0.10	**−3.4**	−0.05	−0.9	−0.04	−1.1	−0.09	**−3.4**
divi	−0.13	−0.7	−0.29	−2.1	−0.29	−2.4	−0.20	−1.2	−0.28	−2.3	−0.36	**−3.6**	−0.14	−0.8	−0.15	−1.3	−0.23	−2.3	−0.11	−0.7	−0.15	−1.2	−0.22	−2.3
divo	0.15	0.7	−0.12	−0.8	0.04	0.4	−0.03	−0.2	−0.16	−1.3	−0.06	−0.7	0.17	1.0	−0.02	−0.2	−0.01	−0.1	0.21	1.2	0.00	0.0	0.05	0.6
dy	0.03	0.2	0.05	0.7	0.03	0.5	0.01	0.1	0.01	0.3	0.03	0.7	−0.04	−0.5	−0.02	−0.3	0.00	−0.1	−0.01	−0.1	0.01	0.1	0.02	0.4
ear	0.12	2.2	0.11	**4.9**	0.18	**6.9**	0.07	1.8	0.08	**5.4**	0.14	**7.0**	0.13	2.4	0.11	**5.1**	0.16	**7.1**	0.12	2.2	0.10	**4.6**	0.14	**6.3**
egr	−0.18	−2.9	−0.19	**−3.9**	−0.21	**−4.8**	−0.17	−3.2	−0.17	**−4.2**	−0.16	**−4.7**	−0.10	−2.2	−0.08	−2.7	−0.02	−1.0	−0.06	−1.7	−0.05	−2.0	0.00	−0.2
ep	0.22	1.0	0.07	0.9	0.01	0.1	0.25	1.5	0.07	1.2	0.07	0.8	0.19	1.1	0.06	1.0	0.13	1.5	0.16	0.8	0.04	0.6	0.08	1.0
fgr5yr	0.00	0.0	−0.06	−0.5	−0.03	−0.4	−0.05	−0.5	−0.08	−0.8	−0.05	−0.9	0.03	0.3	0.00	0.0	0.00	0.1	0.02	0.2	−0.01	−0.1	0.00	−0.1
gma	0.09	1.1	0.05	1.1	0.07	1.7	0.15	2.5	0.10	2.5	0.10	2.4	0.18	2.4	0.12	2.6	0.20	**5.2**	0.12	1.7	0.08	1.9	0.14	3.3
grcapx	−0.15	−2.8	−0.14	**−4.2**	−0.20	**−7.5**	−0.15	−3.2	−0.13	**−4.5**	−0.17	**−7.6**	−0.08	−2.2	−0.07	−3.1	−0.10	**−4.9**	−0.09	−2.2	−0.07	−3.1	−0.11	**−5.3**
grltnoa	−0.14	**−3.8**	−0.16	**−4.4**	−0.27	**−6.8**	−0.12	**−3.9**	−0.14	**−4.4**	−0.22	**−6.4**	−0.03	−0.8	−0.04	−1.6	−0.07	−2.4	−0.02	−0.6	−0.05	−1.6	−0.07	−2.5
herf	0.03	1.0	0.00	0.0	−0.05	−1.4	0.04	1.3	0.00	0.0	−0.08	−2.4	0.04	1.3	0.00	0.1	−0.07	−2.2	0.03	1.2	0.00	−0.1	−0.07	−2.1
hire	−0.13	−2.0	−0.16	−3.2	−0.29	**−7.2**	−0.11	−2.4	−0.14	**−3.4**	−0.24	**−6.7**	0.01	0.3	0.00	−0.1	−0.08	−3.1	0.02	0.4	0.00	0.0	−0.09	**−3.7**
idiovol	−0.12	−0.5	−0.17	−1.2	−0.06	−0.4	−0.34	−1.5	−0.25	−2.0	−0.23	−1.6	−0.11	−0.4	−0.14	−1.0	−0.19	−1.2	−0.11	−0.4	−0.13	−0.9	−0.17	−1.2
ill	−0.12	−1.0	−0.06	−2.7	0.32	**4.4**	−0.21	−2.2	−0.06	−2.9	0.27	**4.6**	−0.31	−3.0	−0.08	**−3.4**	0.23	**4.0**	−0.29	−3.1	−0.08	−3.3	0.26	**4.4**
indmom	0.07	1.1	0.15	1.6	0.35	**3.8**	0.02	0.4	0.10	1.6	0.31	**4.1**	0.06	1.0	0.15	1.8	0.34	**4.0**	0.06	0.9	0.14	1.6	0.34	**3.8**
invest	−0.13	−2.7	−0.19	**−4.1**	−0.34	**−7.0**	−0.12	−3.3	−0.17	**−4.6**	−0.28	**−7.3**	−0.05	−1.0	−0.07	−1.9	−0.14	−3.2	−0.04	−0.8	−0.08	−1.9	−0.15	**−3.3**
ipo	−0.10	−0.5	−0.30	−1.2	−0.73	**−4.5**	−0.03	−0.1	−0.26	−1.1	−0.73	**−5.1**	−0.07	−0.4	−0.24	−1.1	−0.63	**−4.7**	−0.10	−0.5	−0.31	−1.3	−0.72	**−5.1**
lev	0.05	0.4	0.08	1.0	0.05	0.6	0.00	0.0	0.02	0.3	−0.06	−0.8	0.01	0.1	0.03	0.4	−0.06	−0.7	0.03	0.3	0.06	0.7	0.00	0.0
mom1m	−0.12	−1.1	−0.13	−1.9	−0.58	**−6.2**	−0.25	−2.8	−0.20	−3.0	−0.65	**−7.3**	−0.23	−2.3	−0.19	−2.6	−0.64	**−7.1**	−0.20	−2.0	−0.17	−2.3	−0.63	**−7.1**
mom36m	−0.05	−0.7	−0.14	−2.3	−0.19	−2.4	−0.08	−1.1	−0.11	−2.1	−0.10	−1.6	−0.03	−0.5	−0.06	−1.4	−0.04	−0.7	−0.05	−0.8	−0.09	−1.8	−0.08	−1.6
ms	0.06	1.0	0.08	1.7	0.07	1.0	0.14	2.9	0.15	3.3	0.26	**5.7**	0.12	2.3	0.11	2.6	0.25	**5.7**	0.08	1.4	0.08	1.9	0.22	**5.2**
mve_ia	−0.02	−1.1	−0.05	−1.3	−0.04	−0.8	0.00	0.2	0.02	0.7	0.14	**3.9**	−0.03	−1.4	−0.03	−0.8	0.08	2.1	−0.04	−1.6	−0.03	−0.8	0.10	2.7
nanalyst	−0.01	−0.3	−0.03	−0.9	−0.07	−1.4	0.02	0.7	0.02	0.5	0.30	**3.6**	0.03	0.8	0.02	0.4	0.26	2.9	0.03	0.7	0.03	0.5	0.28	3.1
nincr	0.09	2.8	0.12	**3.9**	0.16	**5.8**	0.08	3.2	0.10	**3.9**	0.19	**8.0**	0.10	**3.9**	0.14	**4.8**	0.22	**8.7**	0.08	2.9	0.11	**3.6**	0.18	**7.6**
orgcap	0.13	1.9	0.09	2.1	0.21	3.1	0.15	2.2	0.11	2.6	0.18	**3.4**	0.06	0.9	0.05	1.2	0.12	2.1	0.08	1.2	0.06	1.3	0.13	2.3
pchcapx_ia	0.02	0.4	0.00	−0.1	−0.01	−0.1	0.02	0.5	0.00	0.1	0.00	0.0	0.03	0.8	0.01	0.4	0.02	0.4	0.04	1.1	0.02	0.5	0.02	0.4
pchcurrat	−0.10	−2.7	−0.07	−3.3	−0.07	−2.5	−0.10	−3.1	−0.07	**−3.7**	−0.06	−2.4	−0.07	−2.0	−0.02	−1.2	0.00	0.0	−0.07	−2.3	−0.02	−1.5	−0.01	−0.3
pchdepr	0.03	0.4	0.02	0.6	−0.01	−0.2	0.00	−0.1	0.01	0.4	−0.02	−0.7	0.04	0.9	0.01	0.5	−0.05	−1.7	0.04	0.8	0.02	0.7	−0.05	−1.7
pchgm_pchsale	0.08	2.4	0.07	2.8	0.10	**3.6**	0.07	2.7	0.07	3.2	0.10	**4.3**	0.04	1.4	0.05	2.3	0.07	3.3	0.04	1.4	0.05	2.0	0.07	3.3
pchsale_pchinvt	0.06	1.5	0.08	**3.7**	0.12	**5.3**	0.05	1.4	0.07	**4.2**	0.11	**5.2**	0.03	0.9	0.05	3.0	0.08	**3.7**	0.02	0.7	0.04	2.1	0.07	3.2
pchsale_pchrect	0.02	0.6	0.03	1.7	0.08	**4.6**	0.01	0.4	0.03	1.5	0.08	**5.0**	−0.03	−0.7	−0.03	−1.3	0.02	0.9	−0.03	−0.7	−0.02	−1.1	0.01	0.8
pchsale_pchxsga	−0.09	−1.9	−0.08	−3.1	−0.08	−3.3	−0.10	−2.1	−0.07	−3.3	−0.06	−2.7	−0.03	−0.6	−0.03	−1.6	−0.01	−0.3	−0.05	−1.1	−0.04	−1.8	−0.02	−0.9
pchsaleinv	0.04	0.9	0.05	2.0	0.04	1.6	0.02	0.5	0.03	1.8	0.04	1.5	0.03	0.9	0.04	1.9	0.03	1.5	0.02	0.6	0.03	1.5	0.02	0.9
pctacc	−0.02	−0.6	−0.04	−2.0	−0.10	**−4.0**	−0.01	−0.2	−0.03	−1.6	−0.08	**−4.2**	0.00	0.0	−0.02	−1.0	−0.07	**−3.6**	−0.01	−0.3	−0.02	−1.0	−0.07	−3.2
pricedelay	0.04	0.5	0.01	0.3	0.03	1.3	0.03	0.5	0.01	0.7	−0.01	−0.4	0.05	0.7	0.01	0.6	−0.01	−0.3	0.04	0.6	0.01	0.3	−0.01	−0.5
ps	0.07	1.3	0.12	2.0	0.18	2.8	0.06	1.1	0.12	2.3	0.22	**4.9**	0.04	0.9	0.10	1.8	0.21	**4.6**	0.03	0.6	0.09	1.8	0.20	**4.8**

(*continued*)

Table 4
Continued

	(A) Single characteristic, No benchmark model controls						(B) Single characteristic, Carhart benchmark model controls						(C) Single characteristic, 5-factor benchmark model controls						(D) Single characteristic, q-factor benchmark model controls					
	All stocks, VWLS		All-but-microcap, OLS		All stocks, OLS		All stocks, VWLS		All-but-microcap, OLS		All stocks, OLS		All stocks, VWLS		All-but-microcap, OLS		All stocks, OLS		All stocks, VWLS		All-but-microcap, OLS		All stocks, OLS	
	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.
rd	0.10	1.2	0.19	1.4	0.47	2.3	0.09	1.2	0.19	1.8	0.47	2.8	0.03	0.3	0.12	1.0	0.34	1.9	0.03	0.3	0.14	1.1	0.35	2.0
rd_mve	0.07	0.7	0.14	2.0	0.29	**3.4**	0.02	0.2	0.12	1.8	0.24	3.2	0−0.01	−0.2	0.10	1.5	0.20	2.7	0.02	0.2	0.13	2.0	0.24	**3.6**
rd_sale	−0.15	−1.7	−0.06	−1.2	−0.01	−0.3	−0.13	−1.8	−0.04	−1.0	0.01	0.2	−0.08	−1.2	−0.02	−0.5	0.02	0.5	−0.11	−1.4	−0.02	−0.4	0.02	0.3
realestate	0.06	1.3	0.05	1.2	0.01	0.1	0.07	1.7	0.05	1.4	0.01	0.3	0.05	1.2	0.03	1.0	0.00	0.0	0.05	1.4	0.04	1.0	0.00	0.1
retvol	−0.26	−1.1	−0.25	−2.0	−0.17	−1.3	−0.46	−2.2	−0.31	−2.9	−0.32	−2.8	−0.28	−1.2	−0.23	−2.0	−0.31	−2.6	−0.29	−1.3	−0.23	−1.9	−0.29	−2.6
roaq	0.27	2.5	0.15	2.1	0.19	1.8	0.33	**3.9**	0.16	2.8	0.24	2.8	0.33	**4.0**	0.17	2.9	0.28	3.2	0.24	2.3	0.08	1.2	0.18	2.0
roavol	−0.05	−0.5	−0.07	−1.0	−0.05	−0.5	−0.10	−1.3	−0.07	−1.3	−0.08	−1.0	−0.01	−0.1	−0.03	−0.5	−0.06	−0.7	0.00	0.0	−0.02	−0.3	−0.04	−0.5
roic	0.30	2.6	0.12	1.9	0.04	0.4	0.36	**4.1**	0.13	2.5	0.05	0.6	0.33	**4.1**	0.11	1.9	0.07	0.9	0.30	**3.5**	0.08	1.4	0.05	0.7
rsup	0.04	0.5	0.04	0.9	−0.01	−0.2	0.01	0.2	0.02	0.6	0.01	0.3	0.07	0.9	0.07	1.9	0.08	2.0	0.07	1.0	0.06	1.8	0.05	1.3
salecash	0.05	0.9	0.02	0.7	0.01	0.4	0.02	0.4	0.02	0.7	−0.01	−0.4	0.00	0.0	0.00	−0.1	−0.04	−1.2	0.01	0.2	0.00	0.2	−0.02	−0.6
saleinv	−0.01	−0.3	0.02	0.8	0.03	1.3	−0.02	−0.7	0.02	0.6	0.03	1.8	−0.01	−0.3	0.02	0.8	0.04	1.9	−0.01	−0.3	0.02	0.7	0.03	1.7
salerec	0.08	1.7	0.06	1.4	0.03	0.8	0.07	1.8	0.06	1.5	0.03	0.7	0.05	1.2	0.03	0.8	0.00	0.0	0.07	1.6	0.04	1.1	0.02	0.5
secured	−0.02	−0.5	−0.04	−0.8	0.02	0.4	−0.04	−1.0	−0.05	−1.6	−0.05	−2.1	0.00	−0.1	−0.01	−0.4	−0.03	−1.0	−0.01	−0.3	−0.02	−0.6	−0.05	−1.5
securedind	0.11	0.9	0.05	0.4	0.06	0.4	0.15	1.0	0.08	0.7	−0.01	−0.1	0.10	0.9	0.07	0.7	0.01	0.1	0.15	1.1	0.10	0.8	0.04	0.3
sfe	−0.14	−1.3	−0.03	−0.5	0.05	0.7	−0.06	−0.7	−0.03	−0.6	0.05	1.0	−0.12	−1.2	−0.04	−1.0	0.06	1.2	−0.13	−1.1	−0.05	−1.2	0.05	0.9
sgr	−0.24	−2.8	−0.16	−3.0	−0.24	**−6.6**	−0.23	−3.3	−0.15	**−3.6**	−0.20	−6.4	−0.08	−1.3	−0.03	−0.7	−0.06	−2.1	−0.09	−1.5	−0.03	−0.7	−0.08	−2.4
sin	0.42	1.8	0.32	1.7	0.38	2.0	0.41	2.1	0.28	1.7	0.47	2.5	0.32	1.6	0.32	1.8	0.48	2.5	0.38	1.9	0.29	1.6	0.43	2.2
sp	0.26	1.7	0.15	2.2	0.23	3.3	0.09	1.0	0.09	1.7	0.05	0.8	0.00	0.0	0.04	0.7	−0.01	−0.2	0.15	1.2	0.11	1.7	0.11	1.5
std_dolvol	0.06	0.5	0.06	1.8	0.12	2.3	−0.08	−0.9	0.05	1.5	−0.06	−0.9	−0.08	−0.9	0.03	1.0	−0.09	−1.1	−0.07	−0.7	0.03	1.0	−0.09	−1.1
std_turn	−0.01	−0.1	−0.04	−0.5	−0.05	−0.8	−0.10	−1.2	−0.07	−1.1	−0.09	−1.5	0.01	0.1	0.00	0.0	−0.02	−0.3	0.02	0.2	0.00	0.0	−0.02	−0.4
stdcf	−0.06	−1.6	−0.06	−1.8	−0.07	−1.7	−0.09	−2.9	−0.07	−2.6	−0.07	−2.1	−0.06	−1.8	−0.04	−1.4	−0.06	−1.7	−0.05	−1.5	−0.03	−1.2	−0.05	−1.5
sue	0.20	2.7	0.13	**5.0**	0.25	**7.5**	0.14	1.9	0.10	**4.0**	0.24	**8.3**	0.20	2.9	0.13	**5.2**	0.26	**8.8**	0.15	2.3	0.10	**4.2**	0.18	**6.1**
tang	0.02	0.4	0.01	0.1	0.09	1.2	0.02	0.4	0.01	0.2	0.09	1.4	0.03	0.6	0.03	0.4	0.10	1.5	0.02	0.4	0.02	0.2	0.08	1.1
tb	0.06	1.4	0.06	1.8	0.06	1.6	0.07	1.9	0.07	2.3	0.09	2.8	0.05	1.3	0.05	1.5	0.08	2.5	0.06	1.5	0.04	1.4	0.08	2.5
turn	−0.03	−0.3	−0.15	−1.4	−0.31	**−3.7**	−0.11	−1.2	−0.19	−2.2	−0.30	**−3.6**	0.00	0.0	−0.09	−0.9	−0.20	−2.1	0.00	0.0	−0.09	−0.9	−0.20	−2.2
zerotrade	−0.02	−0.2	−0.01	−0.7	0.08	1.6	−0.12	−1.6	−0.02	−1.2	−0.04	−0.6	−0.16	−2.1	−0.02	−1.4	−0.06	−0.9	−0.15	−2.0	−0.02	−1.3	−0.05	−0.7

This table presents the results of cross-sectional Fama-MacBeth (1973) regressions of monthly stock returns on each of the 94 firm characteristics one at a time, after controlling for no other characteristics (Column A), or for the characteristics of notable benchmark models (Columns B–D). The data window is January 1980–December 2014. Missing characteristic values are replaced by zeros, the mean of the pre-replacement values of each standardized characteristic in each month. The benchmark models are those of Carhart (1997), Fama and French (2015), and Hou, Xue, and Zhang (2015), and the first six lines of the table list the six firm characteristics associated in total with these models (*agr*, *bm*, *mom12m*, *mve*, operprof, and *roeq*). Three sets of regressions are shown for each column: Regressions using all stocks and weighted least squares where the weight is the market value of *equity* for stock i at time $t-1$ (all stocks, VWLS); regressions using all-but-microcap stocks and OLS (all-but-microcap, OLS); and regressions using all stocks and OLS (all stocks, OLS). Microcaps are defined as stocks in month $t-1$ that have a market value of equity less than the 20th percentile of stocks on the NYSE stock exchange in month $t-1$. Intercepts are estimated but not reported. FM coefficients are the means of the monthly estimated coefficients*100; $t-$statistics are taken from the time series of monthly coefficient estimates and employ Newey-West adjustments of 12 lags. The sample is all common stocks on the NYSE, AMEX, and NASDAQ exchanges with sufficient annual and quarterly Compustat accounting data and CRSP stock returns. Analyst data are from I/B/E/S. Firm-month observations are created using monthly stock returns, including delisting returns, for each month t. Stock returns in month t are matched with the accounting information most recently available as of the beginning of month t, assuming that the most recently available annual accounting information is for fiscal years ending at least six months prior to the beginning of month t and that the most recently available quarterly accounting information is for fiscal quarters ending at least four months prior to the beginning of month t. Estimated coefficients with a two-tailed p-value $\leqslant$ 0.05 adjusted using the dependent failure discovery rate method (DFDR; Benjamini and Yekutieli 2001) are shown with a bold t-statistic.

to take multiple testing into account using DFDR p-values. In Figure 1, we visually present the way in which the number of coefficients deemed significant in Table 4 using DFDR p-values varies with DFDR p-values that are higher or lower than our 5% cutoff value. The number of statistically significant coefficients varies most substantially for very low levels of statistical significance (p-values <.01). Around our 5% p-value cutoff, the curves in Figure 1 are relatively flat, indicating that within any given column in Table 4 the number of significant coefficient estimates is relatively insensitive to whether the p-value cutoff is 1%, 5%, or 10%. More visually striking is the difference in the number of significant characteristics across our approaches that emphasize microcaps (OLS on all stocks) or deemphasize microcaps (VWLS on all stocks, and OLS on all-but-microcaps). Overall, Figure 1 shows the value of accounting for multiple testing and the disproportionate effect of microcaps in inferring the degree of characteristics-based predictability in average monthly returns.

Columns B–D of Table 4 show that when added singly to the Carhart, five-factor, and q-factor models, it is the case that 6, 4, and 1 characteristics are incrementally significant using VWLS on all stocks, and 17, 4, and 4 characteristics are incrementally significant using OLS on all-but-microcaps. Since Columns B–D also show that the q-factor model leads to the fewest number of significant characteristics outside of those in the benchmark model itself, we infer that between the three purposefully low-dimension benchmark models, the q-factor model best captures the cross-sectional variation in average returns. Our conclusion echoes that of Hou, Xue, and Zhang (2015, 2016), who conduct similar tests but using the factor versions of the Carhart, five-factor, and q-factor models. Table 4 also highlights one reason for the performance of the q-factor and five-factor models relative to the Carhart model: These models include a measure of growth important when considering return predictability with characteristics one at a time.

3.2 Identifying the independent determinants of the cross-section of average U.S. monthly returns by simultaneously including all 94 characteristics in Fama-MacBeth regressions

Table 5 presents the results of relaxing the univariate approach in Table 4 of evaluating nonbenchmark model characteristics singly by simultaneously including all 94 characteristics in Fama-MacBeth regressions, using the full period from 1980 to 2014. Given the assumptions of OLS, we argue that this method enables us to most powerfully identify the independent determinants of average returns, in that we expect that only those characteristics that are truly independent determinants of returns will have significant estimated coefficients when the effects of all other characteristics are controlled for.

Based on DFDR p-values, Column A of Table 5 shows that for the period from 1980 to 2014, a total of six characteristics are reliably independent determinants when using VWLS applied to all stocks (book-to-market, cash, one-month momentum, change in six-month momentum, return volatility,

Table 5
Results of Fama-MacBeth regressions of monthly stock returns on all 94 firm characteristics simultaneously

	Set of stocks, regression method					
	(A) All stocks, VWLS		(B) All-but-microcap stocks, OLS		(C) All stocks, OLS	
	FM coef.	*t*-stat.	FM coef.	*t*-stat.	FM coef.	*t*-stat.
# DFDR ⩽ 0.05		6		9		23
# |*t*-stats|⩾ 3.0		8		12		27
agr	−0.02	−0.5	−0.10	−3.0	−0.14	**−4.7**
bm	0.21	**3.6**	0.07	2.1	0.10	3.2
mom12m	0.16	1.7	0.13	1.6	0.10	1.4
mve	−0.25	−2.5	−0.17	−3.0	−0.67	**−5.9**
operprof	0.01	0.4	0.01	0.5	0.01	0.0
roeq	0.08	2.1	0.05	2.1	0.09	2.8
absacc	−0.01	−0.2	−0.04	−1.5	−0.03	−0.8
acc	−0.15	−2.9	−0.06	−2.2	−0.04	−1.5
aeavol	−0.00	−0.1	−0.01	−0.8	0.01	1.1
age	−0.05	−2.0	−0.04	−2.1	0.03	1.1
baspread	−0.07	−0.5	0.04	0.5	0.25	2.8
beta	0.03	0.2	0.00	0.0	0.07	1.1
bm_ia	−0.24	−1.1	−0.04	−0.5	−0.10	−0.9
cash	0.21	**3.8**	0.17	**3.4**	0.20	**4.9**
cashdebt	0.02	0.4	0.03	1.1	−0.01	−0.5
cashpr	−0.04	−2.1	−0.03	−1.7	−0.02	−1.4
cfp	−0.05	−1.1	0.02	0.8	0.08	2.7
cfp_ia	0.23	1.1	0.05	0.7	0.12	1.2
chatoia	0.07	2.6	0.04	2.2	0.06	3.1
chcsho	−0.03	−2.1	0.01	0.1	−0.05	−3.1
chempia	0.09	1.8	0.05	1.4	0.04	1.1
chfeps	0.05	2.1	0.07	3.2	0.14	**6.2**
chinv	−0.01	−0.2	−0.04	−1.5	−0.06	−2.3
chmom	−0.16	**−3.5**	−0.03	−1.1	0.02	0.7
chnanalyst	−0.02	−1.5	−0.07	**−3.3**	−0.08	**−4.3**
chpmia	0.03	0.9	0.02	0.8	0.03	1.2
chtx	0.00	0.1	−0.01	−0.5	0.08	**5.1**
cinvest	−0.04	−1.6	−0.04	−2.3	−0.02	−1.1
convind	−0.07	−1.7	−0.04	−0.8	−0.20	**−3.8**
currat	−0.04	−1.4	−0.04	−2.1	−0.02	−0.9
depr	0.01	0.4	0.04	1.3	0.05	1.9
disp	0.00	0.2	−0.02	−1.2	−0.08	**−4.6**
divi	−0.23	−1.9	−0.14	−1.7	−0.23	−2.9
divo	0.08	0.7	0.01	0.0	0.03	0.4
dy	−0.05	−1.4	−0.08	−2.5	−0.05	−2.1
ear	0.08	3.0	0.07	**5.7**	0.10	**6.6**
egr	−0.05	−1.5	−0.03	−1.4	−0.01	−0.7
ep	0.15	2.3	0.02	0.6	0.14	**3.9**
fgr5yr	0.03	0.4	−0.01	−0.1	0.03	0.8
gma	0.04	1.0	0.09	2.1	0.08	1.9
grcapx	−0.06	−2.9	−0.05	−2.9	−0.06	**−4.0**
grltnoa	−0.06	−2.0	−0.04	−1.5	−0.02	−0.7
herf	0.04	1.7	0.01	0.3	−0.05	−2.5
hire	−0.04	−0.8	−0.01	−0.3	−0.06	−1.3
idiovol	−0.08	−0.8	−0.04	−0.9	−0.15	−2.5
ill	0.18	1.9	−0.07	−2.1	0.45	**8.4**
indmom	0.02	0.5	0.11	2.6	0.34	**6.5**
invest	0.02	0.4	−0.00	−0.1	−0.02	−0.7
ipo	0.05	0.4	−0.04	−0.3	−0.32	−2.5
lev	0.02	0.2	0.05	1.1	0.01	0.1
mom1m	−0.50	**−6.9**	−0.37	**−6.5**	−0.81	**−9.4**
mom36m	−0.01	−0.3	−0.02	−1.0	−0.02	−0.8

(*continued*)

Table 5
Continued

	Set of stocks, regression method					
	(A) All stocks, VWLS		(B) All-but-microcap stocks, OLS		(C) All stocks, OLS	
	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.
ms	0.04	1.5	0.06	2.8	0.09	2.7
mve_ia	−0.01	−0.4	0.00	0.1	0.03	1.3
nanalyst	0.04	1.1	0.04	1.1	0.34	**5.7**
nincr	0.05	3.3	0.08	**4.8**	0.12	**6.9**
orgcap	−0.00	−0.1	−0.03	−0.8	0.01	0.3
pchcapx_ia	0.03	1.4	0.04	1.9	0.05	1.5
pchcurrat	−0.06	−2.0	−0.01	−0.9	−0.00	−0.2
pchdepr	0.02	0.7	0.02	1.0	−0.03	−1.3
pchgm_pchsale	−0.00	−0.2	0.02	1.4	0.03	1.6
pchsale_pchinvt	−0.01	−0.3	0.03	1.9	0.03	1.9
pchsale_pchrect	−0.01	−0.6	−0.04	−2.1	0.02	1.2
pchsale_pchxsga	−0.02	−0.6	−0.02	−1.0	−0.00	0.0
pchsaleinv	0.03	1.0	−0.01	−0.5	−0.02	−0.7
pctacc	0.02	0.9	0.02	1.4	−0.00	−0.2
pricedelay	−0.00	0.0	0.02	1.7	−0.00	−0.1
ps	0.03	1.2	0.03	1.0	0.05	2.1
rd	0.04	0.7	0.07	1.4	0.12	2.2
rd_mve	0.13	2.4	0.14	**3.3**	0.24	**6.5**
rd_sale	−0.02	−0.4	0.01	0.3	0.04	1.6
realestate	0.03	1.3	0.04	2.0	0.02	1.2
retvol	−0.51	**−3.8**	−0.26	**−4.9**	−0.44	**−6.3**
roaq	0.00	0.1	0.02	0.5	0.09	2.0
roavol	−0.01	−0.3	0.04	1.2	0.01	0.3
roic	0.12	2.0	0.03	1.0	0.01	0.2
rsup	0.04	1.0	0.05	2.2	0.05	2.1
salecash	0.01	0.4	0.01	1.1	−0.03	−1.4
saleinv	−0.03	−2.1	0.01	0.4	0.03	2.4
salerec	0.02	0.9	0.02	0.8	0.00	0.1
secured	−0.02	−0.7	−0.02	−1.3	−0.03	−2.0
securedind	0.13	1.4	0.18	1.4	0.14	1.1
sfe	−0.17	−2.3	−0.09	−1.9	0.04	0.9
sgr	−0.11	−2.5	−0.00	−0.1	−0.04	−1.2
sin	0.30	2.0	0.22	1.5	0.50	3.2
sp	0.00	0.0	0.03	0.9	0.02	0.4
std_dolvol	−0.06	−0.8	−0.04	−1.5	−0.17	**−4.2**
std_turn	0.12	2.6	0.23	**5.0**	0.42	**7.3**
stdcf	−0.04	−1.9	−0.03	−1.4	−0.02	−1.2
sue	0.08	1.7	0.07	3.2	0.12	**5.4**
tang	−0.03	−1.0	−0.01	−0.4	0.02	0.5
tb	0.03	1.3	0.03	1.5	0.04	2.6
turn	−0.11	−2.0	−0.28	**−5.6**	−0.55	**−10.2**
zerotrade	−0.20	**−4.6**	0.00	0.0	−0.34	**−6.9**
Mean # obs.	4,605		1,835		4,605	
Mean adj. R^2	28.6%		15.5%		7.9%	

This table shows the results of monthly Fama-MacBeth (1973) regressions of stock returns in month t when all 94 firm characteristics are simultaneously included as independent variables. The data window is January 1980–December 2014. Missing characteristic observations are set to the zero mean of the nonmissing values of the characteristic in that month after the nonmissing values have been winsorized at the 1st and 99th percentiles and standardized to have a zero mean and unit standard deviation. Three regressions are shown: using all stocks and WLS where the weight is the market value of equity for stock i at time $t-1$ (all stocks, VWLS), using all-but-microcap stocks and OLS (all-but-microcap stocks, OLS), and using all stocks and OLS (all stocks, OLS). Microcaps are stocks in month $t-1$ with a market value of equity less than the 20th percentile of stocks on the NYSE stock exchange in month $t-1$. FM coefficients are the means of the monthly estimated coefficients*100; t-statistics employ Newey-West adjustments of 12 lags. The sample is all common stocks on the NYSE, AMEX, and NASDAQ exchanges with available annual and quarterly Compustat accounting data and CRSP stock return data. Analyst data are from I/B/E/S. Estimated coefficients with a two-tailed p-value $\leqslant 0.05$ adjusted using the dependent failure discovery rate method (DFDR; Benjamini and Yekutieli 2001) are shown with a bold t-statistic. FM, Fama-MacBeth.

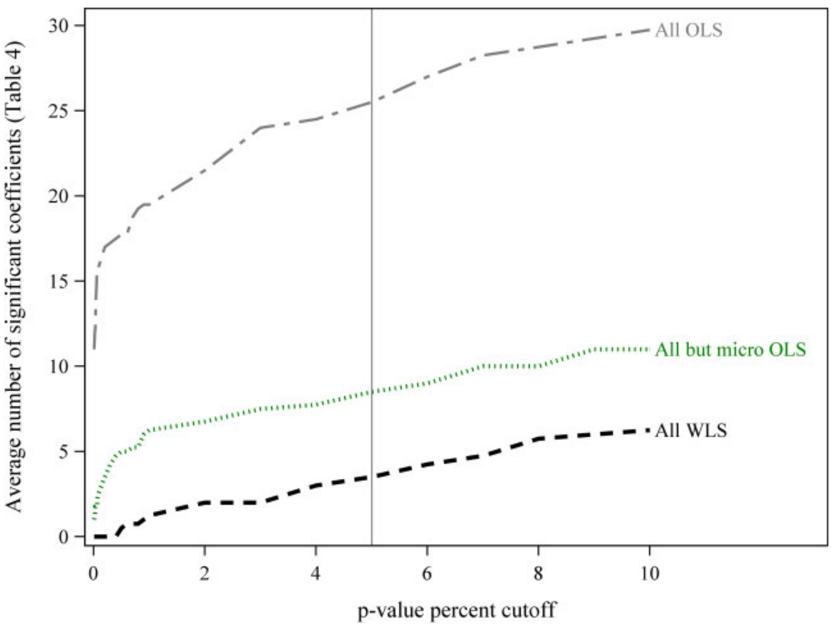


Figure 1
The number of coefficient estimates in each column of Table 4 that are statistically significant at dependent false discovery rate *p*-values ranging from 0% to 10%
This figure plots the average number of coefficient estimates in each column of Table 4 that are statistically significant (*y*-axis) using dependent false discovery rate (DFDR) *p*-value cutoffs ranging from 0% to 10% (*x*-axis). False discovery rate (FDR) *p*-values take into account the risk that we falsely conclude that a firm characteristic is a significant predictor of returns due to the fact that we simultaneously include a large number of characteristics in the same regression, and/or because we estimate several regressions using the same data set. We use the method of Benjamini and Yekutieli (2001) that extends the basic FDR approach of Benjamini and Hochberg (1995) in order to also adjust for the dependency among *p*-values that arises because our tests are performed with the same sets of dependent and independent variables across the various samples and/or methodologies. The figure shows that as the *p*-value cutoff increases, the number of significant coefficients increases, but at a decreasing rate. The vertical lines at 5% show the DFDR cutoff value that we employ. The DFDR *p*-values on individual coefficients are calculated taking into account the dependency in the hypotheses about coeffiicent estimates across columns. DFDR is an extension of the FDR method that adjusts *p*-values to take into account the risk that we falsely conclude that a firm characteristic is a significant predictor of returns due to the fact that we simultaneously include a large number of characteristics in the same regression, and/or because we estimate several regressions using the same data set.

and the number of zero trading days). This increases to nine characteristics if OLS is applied to all-but-microcaps (cash, change in the number of analysts, earnings announcement return, one-month momentum, the number of consecutive quarters with an increase in earnings over the same quarter a year ago, annual R&D to market cap, return volatility, share turnover, and volatility of share turnover). Pooling across the two approaches yields 12 multivariately determined independent characteristics, namely book-to-market, cash, change in the number of analysts, earnings announcement return, one-month momentum, change in six-month momentum, number of consecutive quarters with earnings higher than the same quarter a year ago, annual R&D

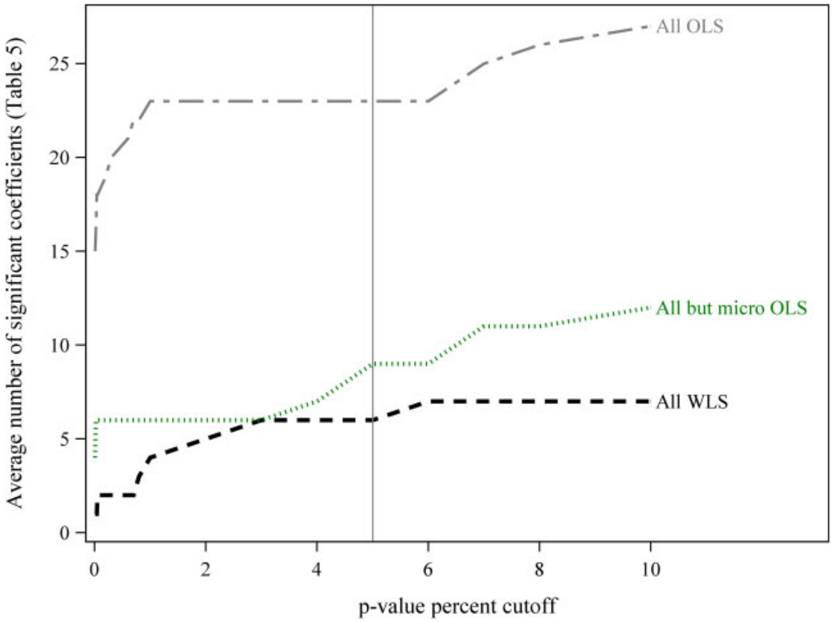


Figure 2
The number of coefficient estimates in each column of Table 5 that are statistically significant at dependent false discovery rate (DFDR) p-values ranging from 0% to 10%
This figure plots the number of coefficient estimates in each column of Table 5 that are statistically significant (y-axis) using dependent false discovery rate (DFDR) p-value cutoffs ranging from 0% to 10% (x-axis). False discovery rate (FDR) p-values take into account the risk that we falsely conclude that a firm characteristic is a significant predictor of returns due to the fact that we simultaneously include a large number of characteristics in the same regression, and/or because we estimate several regressions using the same data set. We use the method of Benjamini and Yekutieli (2001) that extends the basic FDR approach of Benjamini and Hochberg (1995) in order to also adjust for the dependency among p-values that arises because our tests are performed with the same sets of dependent and independent variables across the various samples and/or methodologies. The figure shows that as the p-value cutoff increases, the number of significant coefficients increases, but at a decreasing rate. The vertical lines at 5% show the DFDR cutoff value that we employ. The DFDR p-values on individual coefficients are calculated taking into account the dependency in the hypotheses about coeffiicent estimates across columns. DFDR is an extension of the FDR method that adjusts p-values to take into account the risk that we falsely conclude that a firm characteristic is a significant predictor of returns due to the fact that we simultaneously include a large number of characteristics in the same regression, and/or because we estimate several regressions using the same data set.

to market cap, return volatility, share turnover, volatility of share turnover, and zero trading days. As in Table 4, the disproportionate influence of microcap stocks can be seen in Column C of Table 5 by noting that the application of OLS to all stocks yields 23 independent characteristics. In Figure 2, we visually present the number of significant coefficients in Table 5 as a function with DFDR p-values different than our 5% cutoff value, noting as in Figure 1 that the curves are relatively flat, particularly for non-microcaps.

Taken together, the results in Tables 4 and 5 suggest several takeaways regarding the number and identity of the independent determinants of average monthly stock returns over the period from 1980 to 2014. First, of our sample of 94 characteristics that prior work has presented as predictive of average returns, 12 provide robustly independent information for non-microcap stocks, whereas the remaining 82 do not. This contrast paints an unfavorable picture of the reliability of the inferences made in prior research.

Second, the fact that the number of multivariately identified independent characteristics is the same as the number of univariately significant characteristics (12) indicates that it is not the case that a few independent characteristics are able to powerfully absorb the information contained in a large number of univariately significant characteristics. Instead, there are simply intrinsically few characteristics that independently predict average returns in non-microcap stocks.

Third, the identity and nature of independent characteristics differs substantially from the identity and nature of univariately significant characteristics. By using the fundamental and market approach to classifying anomalies proposed by McLean and Pontiff (2016), we categorize 10 of the 12 univariately significant characteristics as fundamental (asset growth, growth in industry-adjusted sales, growth in inventory, growth in book equity, growth in CAPEX, growth in long-term net operating assets, growth in PP&E plus inventory, number of consecutive quarters with earnings higher than the same quarter a year ago, growth in sales less growth in inventory, and standardized unexpected quarterly earnings) and two as market-based (percentage change in shares outstanding, and earnings announcement return). In contrast, only one of the 12 multivariately identified independent characteristics is fundamental (number of consecutive quarters with earnings higher than the same quarter a year ago), while seven are market-based (change in six-month momentum, earnings announcement return, one-month momentum, return volatility, share turnover, volatility of share turnover, and zero trading days).[11] These differences suggest that research needing to control in its design for the determinants of average returns may do so most powerfully by using the independent characteristics that we identify.

Fourth, the independent characteristics in non-microcaps tend to be independent characteristics in microcaps, but not vice versa. Of the 12 independent characteristics in non-microcaps, 10 are independent characteristics in OLS regressions on all stocks, and 13 of the 23 independent characteristics in microcaps (OLS on all stocks) are not independent characteristics in non-microcaps. It is also the case that 11 of the 12 independent characteristics differ from the characteristics in the Carhart, five-factor, and q-factor benchmark models, with book-to-market being the notable exception. This suggests that past work that has used the characteristics versions of the Carhart, five-factor, and q-factor models to control for cross-sectional variation in average

[11] The average absolute cross-correlation between the univariately significant fundamental characteristics is 0.19, suggesting that the lack of fundamental independent characteristics is not due to high correlation, and that only one characteristic emerges as significant after controlling for all others.

Table 6
Fama-MacBeth regressions of monthly stock returns on simultaneously including as independent variables the subset of twelve firm characteristics in Table 5 that are significant in the All stocks, VWLS, or All-but-microcap OLS regressions

	(A) All stocks, VWLS		(B) All-but-microcap stocks, OLS		(C) All stocks, OLS	
	FM coef.	***t*-stat.**	**FM coef.**	***t*-stat.**	**FM coef.**	***t*-stat.**
# DFDR ⩽ 0.05		5		7		7
# |*t*-stats|⩾ 3.0		2		6		7
bm	0.17	1.4	0.08	1.5	0.25	**4.7**
cash	0.21	**3.3**	0.13	1.8	0.20	2.7
chmom	−0.21	**−2.8**	−0.07	−1.7	−0.05	−1.3
chnanalyst	−0.03	−1.4	−0.08	**−3.0**	−0.06	−2.3
ear	0.12	**2.9**	0.10	**5.7**	0.16	**8.2**
mom1m	−0.26	**−2.8**	−0.24	**−4.2**	−0.74	**−8.3**
nincr	0.08	**3.4**	0.13	**5.1**	0.20	**7.6**
rd_mve	0.00	0.0	0.14	2.6	0.25	**4.1**
retvol	−0.46	−2.3	−0.26	−2.6	−0.19	−1.5
std_turn	0.10	1.5	0.23	**4.0**	0.35	**5.8**
turn	−0.05	−0.6	−0.25	**−3.4**	−0.50	**−7.2**
zerotrade	−0.07	−1.0	−0.04	**−3.0**	−0.06	−1.7
Mean # obs.	4,605		1,835		4,605	
Mean adj. R^2	9.0%		5.9%		3.4%	

This table shows the results of monthly Fama-MacBeth (1973) regressions of stock returns in month t when the firm characteristics that are included as independent variables are restricted to the the subset of 12 characteristics that are significant in table 5 for union of the All-stocks VWLS and the All-but-microcap stocks OLS regressions. The data window is January 1980–December 2014. Missing characteristic values are replaced by zeros, the mean of the pre-replacement values of each standardized characteristic in each month. Estimated coefficients with a two-tailed p-value $\leqslant 0.05$ adjusted using the dependent failure discovery rate method (DFDR; Benjamini and Yekutieli 2001) are shown with a bold t-statistic. Three regressions are shown: using all stocks and WLS where the weight is the market value of equity for stock i at time $t-1$ (all stocks, VWLS), using all-but-microcap stocks and OLS (all-but-microcap stocks, OLS), and using all stocks and OLS (all stocks, OLS). Microcaps are stocks in month $t-1$ with a market value of equity less than the 20th percentile of stocks on the NYSE stock exchange in month $t-1$. FM coefficients are the means of the monthly estimated coefficients*100; t-statistics employ Newey-West adjustments of 12 lags. The sample is all common stocks on the NYSE, AMEX, and NASDAQ exchanges with available annual and quarterly Compustat accounting data and CRSP stock return data. Analyst data are from I/B/E/S. FM, Fama-MacBeth.

returns outside of the particular independent variable that was the focus of the work may have failed to achieve the degree of control that was expected, particularly to the extent that the set of firms being studied comprised large cap stocks.

In Table 6 we assess the robustness of our conclusion from Table 5 that over the period from 1980 to 2014 there are 12 independent determinants of average monthly returns in non-microcap stocks. Specifically, Table 6 reports the results of estimating the regressions reported in Table 5 when the characteristics included as independent variables are restricted to the 12 that were multivariately identified as being independent determinants in Table 5. An inspection of Columns A and B indicates that, as compared with the 6 and 9 independent determinants identified by DFDR p-values in Table 5, similar numbers of 5 and 7 are seen in Table 6. Likewise, there are a total of 9 characteristics across Columns A and B in Table 6, as compared to 12 characteristics in Table 5. We note that in comparing Tables 5 and 6, the sole representative of the characteristics equating to the factors in the Carhart, five-factor, and q-factor models, namely book-to-market, does not remain significant in Table 6.

3.3 Hedge portfolio returns from predicting the cross-section of returns using characteristics

Since statistical significance does not necessarily translate into economic importance, in this section we report the results of hedge portfolio tests aimed at measuring the magnitude of the economic benefits of exploiting the full set of characteristics when predicting the cross-section of one-month-ahead returns. Specifically, following the method of Lewellen (2015), in panel A of Table 7 we report the magnitude and significance of three mean one-month holding period out-of-sample hedge portfolio raw returns that are calculated as follows.

In panel A, for every month $t-1$ starting in January 1990, we use data from months $t-120$ through $t-1$ to estimate the same three sets of regressions as those reported in Columns A, B, and C of Table 5. As in Tables 4–6, we assume that annual accounting data are available at the end of month $t-1$ if the firm's fiscal year ended at least six months before the end of month $t-1$ for annual data and that quarterly accounting data are available at the end of month $t-1$ if the firm's fiscal quarter ended at least four months before the end of month $t-1$. Separately for the set of stocks defined in each column, on a firm-by-firm basis we then apply the resulting mean coefficient estimates to the values of the corresponding 94 characteristics as of the end of month $t-1$. This yields three predicted returns for month t for each firm, one for each of the Columns A, B, and C. Then we calculate the realized return to each of three hedge portfolios for month t by using two types of breakpoints for identifying firms in the top and bottom deciles of predicted returns and two methods of weighting the individual realized returns within each decile.

For the Column A portfolio in panel A of Table 7, labeled portfolio A, decile breakpoints are based on only NYSE firms (so there are an equal number of NYSE stocks in each decile, but not necessarily an equal number of stocks across the NYSE, AMEX, and NASDAQ combined) and the realized returns in the top/long and bottom/short deciles are weighted by firms' market caps at the end of month $t-1$. For the Column B portfolio, referred to as portfolio B, the decile breakpoints are based on all-but-microcap stocks, and the realized returns in the top/long and bottom/short deciles are equally weighted. Lastly, for the Column C portfolio, labeled portfolio C, the decile breakpoints are based on NYSE stocks and the realized returns in the top/long and bottom/short deciles are equally weighted. Since predicted returns are available for each firm at the end of each month, our approach is in theory implementable in real time using only historically available data.[12] In total, applying this method yields a time

[12] We confirm that our approach is only quasi out-of-sample, since we use characteristics that were not first reported in the academic literature until after month t-1. However, potentially counterbalancing some the upward bias

Table 7
Statistics on raw monthly hedge portfolio returns for January 1990–December 2014

A. Descriptive statistics for raw monthly hedge portfolio returns, where predicted returns are based on using all characteristics in Table 5 (except for those that require I/B/E/S data) and coefficients from rolling 120-month Fama-MacBeth regressions of month $t-119$ returns on $t-120$ characteristics through month $t-1$ returns on $t-2$ characteristics

Long top/short bottom decile hedge portfolio	Mean return	Std. dev.	t-stat.
Portfolio A: VW all-stocks hedge portfolio with NYSE decile breakpoints	1.2%	5.3%	3.8
Portfolio B: EW all-but-microcaps hedge portfolio with all-but-microcaps decile breakpoints	1.4%	5.5%	4.5
Portfolio C: EW all-stocks hedge portfolio with NYSE decile breakpoints	3.1%	4.7%	11.3

B. Descriptive statistics for raw monthly hedge portfolio returns, where predicted returns are based on limiting the firm characteristics to the subset of the 12 characteristics shown in table 6: bm, cash, chmom, chnanalyst, ear, mom1m, nincr, rd_mve, retvol, std_turn, turn, zerotrade. From these we omit chnanalyst, because analyst data are not available early in the sample data window.

Portfolio A: VW all-stocks hedge portfolio with NYSE decile breakpoints	0.4%	5.7%	1.2
Portfolio B: EW all-but-microcaps hedge portfolio with all-but-microcaps decile breakpoints	1.0%	5.2%	3.4
Portfolio C: EW all-stocks hedge portfolio with NYSE decile breakpoints	2.4%	4.2%	9.6

C. Differences in the mean hedge returns reported in panel A versus panel B

Long top/short bottom decile hedge portfolio	Mean return diff.	Std. dev. of diff.	t-stat. on diff.
Portfolio A: VW all-stocks hedge portfolio with NYSE decile breakpoints	0.8%	4.5%	3.0
Portfolio B: EW all-but-microcaps hedge portfolio with all-but-microcaps decile breakpoints	0.4%	3.1%	2.2
Portfolio C: EW all-stocks hedge portfolio with NYSE decile breakpoints	0.7%	3.1%	4.0

This table presents statistics for the monthly hedge portfolio returns calculated as the value-weighted (VW) or equally-weighted (EW) mean return in month t for stocks in the top decile of stocks minus the VW or EW return for stocks in the bottom decile of stocks for the monthly cross-section of predicted returns. Missing observations for a given characteristic in a given month are set to the zero mean of the nonmissing values of the characteristic in that month after the nonmissing values have been winsorized at the 1st and 99th percentiles and standardized to have a zero mean and unit standard deviation. VW uses the equity market value for month $t-1$. Decile cutoffs for each month are created from NYSE stocks or all-but-microcap stocks. Microcap stocks are stocks with equity market values less than the 20th percentile of NYSE stocks. Predicted returns in panel A are calculated using all characteristics from Table 5 (except for those that require I/B/E/S data) available as of the end of month $t-1$, and coefficients that are the mean estimated coefficients from rolling 120-month Fama-MacBeth (1973) regressions of month $t-119$ returns on $t-120$ characteristics through month $t-1$ returns on $t-2$ characteristics. Predicted returns in panel B are calculated using the twelve characteristics from table 6 (except *chnanalyst,* which requires I/B/E/S data). For panels A and B, the first estimation window begins January 1, 1980, and hedge portfolio returns are calculated from January 1990 through December 2014.

series of 274 realized raw monthly hedge returns over the period from January 1990 to December 2014 for each of the three hedge portfolios A, B, and C.

Panel A reports descriptive statistics for the realized raw monthly hedge portfolio returns. Inspection shows that the mean raw returns for all three types of hedge portfolios are always statistically and economically large, with the

in implementable hedge returns that such data snooping creates, we never discard a characteristic, even when an estimated coefficient would indicate that the characteristic is either no longer incrementally significant or is significant, but with a sign opposite to that expected based on the anomalies literature.

largest (smallest) in the smallest (largest) cap firms. For the cross-section of stocks measured by portfolio A's VW all-stocks long top/short bottom decile hedge portfolio with NYSE decile breakpoints, the mean raw monthly return over 1990–2014 is 1.2% (t-statistic = 3.8), while for the cross-section of stocks measured by portfolio B's EW all-but-microcaps hedge portfolio using all-but-microcaps decile breakpoints the mean monthly raw return is 1.4% (t-statistic = 4.5). These results show that there are substantial economic benefits to using the predictability in the full set of all 94 firm characteristics. At the same time, we note (once more highlighting the disproportionate influence that microcaps can exert) the mean monthly raw return for portfolio C's EW all-stocks hedge portfolio with NYSE decile breakpoints is 3.1% (t-statistic = 11.3).[13]

Panel B reports the same statistics as in panel A, but using predicted returns to construct top/long and bottom/short deciles based on limiting the firm characteristics to the subset of the 12 shown in Table 6, less the change in the number of analysts because analyst data are not available early in the sample data window. Panel C tabulates the differences in the mean hedge returns reported in panels A and B. It can be seen that limiting the construction of hedge portfolios designed to exploit characteristics-based predictability to the 12 characteristics identified in Columns A and B of Table 5 does reduce mean hedge returns in a material and statistically significant manner, especially for large cap firms. The mean hedge return falls by two-thirds for portfolio A, from 1.2% to 0.4% (t-statistic on difference = 3.0), but by only one-third for portfolio B, from 1.4% to 1.0% (t-statistic on difference = 2.2). We note in passing that although Table 5 identifies 23 different independent characteristics in microcap stocks (Column C), restricting the construction of hedge return portfolios to the 12 characteristics identified in Columns A and B of Table 5 only reduces the mean hedge return to the microcap portfolio by one-quarter, from 3.1% to 2.4% (t-statistic on difference = 4.0). The finding that hedge portfolio returns are higher when the full set of firm characteristics is used to predict returns would seem to contradict the regression results that point to there being only a small set of independent firm characteristics. We note, however, that the statistical adjustments used to determine the significance of individual regression coefficients may be too conservative from a portfolio perspective. In addition, since the hedge portfolio returns are based on anomalies that may have been discovered by chance from multiple tests of the same data, the inclusion of large numbers of characteristics when creating the hedge portfolio returns may accentuate multiple testing concerns.

---

[13] Since raw hedge portfolio returns in the anomalies literature are often orthogonalized against key factor returns, in our Internet Appendix, we report the estimated alpha intercepts and associated t-statistics from regressing the raw hedge portfolio returns described in panel A on the factor returns relevant to each of the Carhart, five-factor, and q-factor models. We find that the estimated alphas are similar in magnitude and significance to the mean raw hedge returns found in panel A of Table 7. Moreover, the q-factor model always has the smallest alpha (eliminating more of the mean raw hedge return) than either the Carhart or five-factor model, supporting the inference made from Table 4 that of the benchmark models, Hou, Xue, and Zhang's q-factor model best captures the cross-sectional variation in returns due to the truly independent determinants of average returns.

3.4 The 2003 change in predictability and the pre- versus post-2003 differences in the number and economic importance of the independent determinants of average monthly U.S. stock returns

Consistent with almost all research in the anomaly and asset pricing literatures, the results described in Tables 4–7 treat the years 1980–2014 as a uniform block of calendar time during which the return generating process is presumed to remain constant. We now relax this assumption in light of the substantial changes in the volume, nature, and costs of trading in stocks that occurred from 1980 to 2014, including Reg. FD, the decimalization of trading quotes, Sarbanes-Oxley, accelerated SEC filing requirements, autoquoting, and computerized long/short quantitative investment (Chordia, Roll, and Subrahmanyam 2001; Jones 2002; Schwert 2003; French 2008; Green, Hand, and Soliman 2011; Hendershott, Jones, and Menkveld 2011).

The first approach we take to assessing the constancy of the return-generating process in calendar time is shown in Figure 3, where we plot the natural log of one plus the cumulative mean raw hedge portfolio returns calculated in a manner equivalent to those in panel A of Table 7, but where the seven characteristics that require I/B/E/S data have been excluded because I/B/E/S data are only robustly available starting in 1990. Visual inspection of Figure 3 indicates that the mean hedge returns for each of the three portfolios displayed sharply and persistently fell in late 2002/early 2003. Indeed, the declines are so marked for the non-microcap cross-section of stocks that the characteristics-based predictability of one-month-ahead U.S. returns has been, on average, zero since early 2003, leaving economically meaningful predictability only in microcaps, and there to a much reduced degree.

We provide statistical evidence consistent with these visual assessments in Table 8, where we report the mean raw hedge returns, together with their associated t-statistics, for each of the three hedge portfolios pre-2003, post-2003, and post- versus pre-2003. We define pre-2003 as the period ending December 31, 2002, and the post-2003 period as starting January 1, 2004. The mean raw monthly hedge return in the VW all-stocks hedge portfolio with NYSE decile breakpoints declines from 1.9% before 2003 (t-statistic = 4.4) to 0.5% after 2003 (t-statistic = 1.1), with the fall of −1.4% per month being reliably negative (t-statistic = −2.3). Similarly, the mean raw monthly hedge return for the EW all-but-microcap hedge portfolio with all-microcap decile breakpoints drops from 2.8% before 2003 (t-statistic = 5.7) to 0.1% after 2003 (t-statistic = 0.2), with the post- versus pre-2003 decline of −2.7% being reliably negative (t-statistic = −4.4). In contrast, while the mean raw monthly return for the EW all-stocks hedge portfolio with NYSE decile breakpoints also declines by a significant −2.8% post- versus pre-2003 (t-statistic = −5.2), it remains reliably positive in the post-2003 period with a mean of 1.7% per month (t-statistic = 4.4). All told, the economic hedge returns measure of the degree of predictability in the set of 94 characteristics we study plunges across all sizes of firms, and that the drop is so great in non-microcap

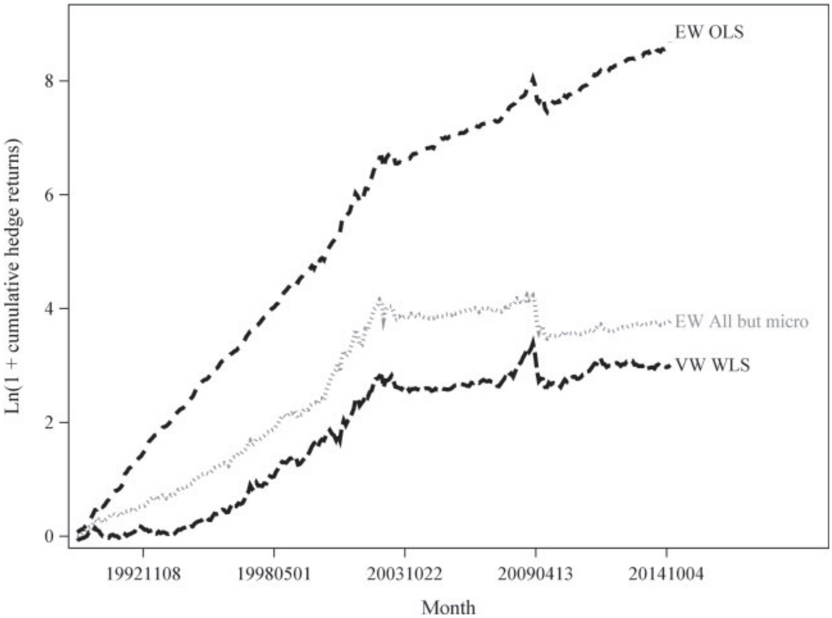


Figure 3
Plot of ln(1 + cumulative mean monthly raw hedge portfolio returns to predicting the cross-section of returns using all 94 characteristics, except those based on I/B/E/S data), displayed separately for (1) value-weighted returns for all stocks with NYSE decile breakpoints (VW WLS), (2) equally weighted returns for all-but-microcap stocks with all-but-microcap decile breakpoints (EW all-but-microcap), and (3) equally weighted returns for all stocks with NYSE decile breakpoints (EW OLS)
This figure plots the natural log of 1 + the cumulative mean monthly raw hedge portfolio returns to predicting the cross-section of returns using all 94 characteristics, less the 8 characteristics that are based on I/B/E/S data. Mean monthly raw hedge portfolio returns calculated as the value-weighted (VW) or equally weighted (EW) mean return in month t for stocks in the top decile of stocks minus the VW or EW return for stocks in the bottom decile of stocks for the monthly cross-section of predicted returns. Missing observations for a given characteristic in a given month are set to the zero mean of the nonmissing values of the characteristic in that month after the nonmissing values have been winsorized at the 1st and 99th percentiles and standardized to have a zero mean and unit standard deviation. VW uses the equity market value for month $t-1$. Decile cutoffs for each month are created from NYSE stocks or all-but-microcap stocks. Micro stocks are stocks with equity market values less than the 20th percentile of NYSE stocks. Predicted returns are calculated using all characteristics from Table 6 (except for characteristics that require I/B/E/S data) available as of the end of month $t-1$, and coefficients that are the mean estimated coefficients from rolling 120-month Fama-MacBeth (1973) regressions of month $t-119$ returns on $t-120$ characteristics through month $t-1$ returns on $t-2$ characteristics. The first estimation window begins January 1, 1980. The hedge portfolio returns are calculated from January 1990 through December 2014.

stocks that since 2003 the mean hedge return to exploiting characteristics-based predictability is insignificantly different from zero. Only in microcap stocks does characteristics-based predictability survive, and for such firms the economic magnitude of the predictability after 2003 is two-thirds less than it was before 2003.

Table 9 confirms the inferences from mean hedge returns in Table 8 by reporting the results of estimating the same Fama-MacBeth regressions as in Table 5, but separately for the subperiods 1980–2002 and 2004–2014. We focus first on Columns A and B in each subperiod, which report results for our two

Table 8
Statistics on raw hedge portfolio returns pre- versus post-2003

	Pre-2003			Post-2003			Post- minus pre-2003	
Portfolio	**Mean return**	**Std. dev.**	***t*-stat.**	**Mean return**	**Std. dev.**	***t*-stat.**	**Diff.**	***t*-stat.**
Portfolio A: VW all-stocks hedge portfolio with NYSE decile breakpoints	1.9%	5.5%	4.4	0.5%	4.9%	1.1	−1.4%	−2.3
Portfolio B: EW all-but-microcaps hedge portfolio with all-but-microcaps breakpoints	2.8%	6.0%	5.7	0.1%	4.4%	0.2	−2.7%	−4.4
Portfolio C: EW all-stocks hedge portfolio with NYSE decile breakpoints	4.4%	4.4%	12.5	1.7%	4.4%	4.4	−2.8%	−5.2

This table presents statistics for the monthly hedge portfolio returns calculated as the value-weighted (VW) or equally-weighted (EW) mean return in month t for stocks in the top decile of stocks minus the VW or EW return for stocks in the bottom decile of stocks for the monthly cross-section of predicted returns. Missing observations for a given characteristic in a given month are set to the zero mean of the nonmissing values of the characteristic in that month after the nonmissing values have been winsorized at the 1st and 99th percentiles and standardized to have a zero mean and unit standard deviation. VW uses the equity market value for month $t-1$. Decile cutoffs each month are created from NYSE stocks or all-but-microcap stocks. Microcap stocks are stocks with equity market values less than the 20th percentile of NYSE stocks. Predicted returns in panel A are calculated using all characteristics from table 5 (except for those that require I/B/E/S data) available as of the end of month $t-1$, and coefficients that are the mean estimated coefficients from rolling 120-month Fama-MacBeth (1973) regressions of month $t-119$ returns on $t-120$ characteristics through month $t-1$ returns on $t-2$ characteristics. The first estimation window begins January 1, 1980, and hedge portfolio returns are calculated from January 1990–December 2002 (pre-2003) and January 2004–December 2014 (post-2003).

alternative definitions of non-microcap firms, and as in our analysis of Table 5, we measure the number of independent determinants of average returns by the number of different characteristics across Columns A and B combined that have DFDR p-values $\leqslant 0.05$. An inspection of Columns A and B reveals that 12 characteristics are independent determinants of non-microcap stocks' average returns from 1980 to 2002, as compared to two characteristics from 2004 to 2014. We note that of the latter, one is an independent determinant both pre- and post-2003 (the number of consecutive quarters with earnings higher than the same quarter a year ago) while one is not (the industry-adjusted change in the number of employees). We also note from the two Column Cs that the number of independent determinants using OLS on all stocks falls from 25 characteristics before 2003 to 8 characteristics after 2003.

We interpret the sharp shift in late 2002/early 2003 in the number and economic importance of the independent determinants of the monthly return generating process, especially across firm size, as being consistent with the costly-limits-to-arbitrage arguments of Shleifer and Vishny (1997), Lesmond, Schill, and Zhou (2004), Chordia, Roll, and Subrahmanyam (2008), Li and Zhang (2010), and Lam and Wei (2011). We note that during the period from July 2002 to June 2003, a number of changes occurred in the information and trading environment that made it cheaper and technologically more feasible to rapidly implement quantitative long/short trading strategies.

Table 9
Results of Fama-MacBeth regressions of monthly stock returns on all 94 firm characteristics simultaneously

	Table 5 Fama-MacBeth regressions, pre-2003 Set of stocks, regression method						Table 5 Fama-MacBeth regressions, post-2003 Set of stocks, regression method					
	(A) pre: All stocks, VWLS		(B) pre: All-but-microcaps, OLS		(C) pre: All stocks, OLS		(A) post: All stocks, VWLS		(B) post: All-but-microcaps, OLS		(C) post: All stocks, OLS	
	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.
# DFDR $\leqslant$ 0.05		6		11		25		1		1		8
# $\|t$-stats$\|\geqslant$ 3.0		11		16		25		2		1		13
agr	−0.03	−0.4	−0.09	−2.4	−0.16	**−4.3**	−0.02	−0.3	−0.10	−1.9	−0.10	−2.1
bm	0.31	**4.2**	0.11	2.8	0.14	**3.5**	0.03	0.3	−0.02	−0.6	0.03	0.6
mom12m	0.31	3.3	0.28	**3.8**	0.23	**4.6**	−0.12	−0.7	−0.15	−0.9	−0.13	−0.8
mve	−0.27	−2.1	−0.24	−3.4	−0.75	**−5.7**	−0.21	−1.3	−0.04	−0.5	−0.52	−2.6
operprof	0.02	0.4	0.00	−0.1	0.00	0.1	0.00	0.0	0.03	1.6	0.00	−0.1
roeq	0.13	2.4	0.10	3.0	0.12	2.7	−0.01	−0.2	−0.03	−1.0	0.03	0.9
absacc	−0.04	−0.6	−0.04	−1.1	−0.04	−1.3	0.05	0.7	−0.05	−1.0	0.00	0.0
acc	−0.20	−3.0	−0.10	−3.0	−0.06	−1.8	−0.05	−0.7	0.01	0.3	−0.02	−0.3
aeavol	0.02	0.6	0.00	−0.1	0.01	0.9	−0.04	−0.9	−0.03	−1.4	0.01	0.6
age	−0.07	−1.7	−0.04	−1.6	0.05	1.3	−0.03	−1.1	−0.04	−1.6	0.00	−0.2
baspread	0.02	0.2	0.04	0.6	0.41	**4.3**	−0.26	−1.0	0.02	0.2	−0.05	−0.3
beta	0.06	0.4	0.02	0.2	0.06	0.8	−0.05	−0.4	−0.03	−0.3	0.09	0.9
bm_ia	−0.35	−1.0	−0.06	−0.4	−0.15	−0.9	−0.02	−0.4	−0.02	−0.3	−0.02	−0.3
cash	0.28	**4.1**	0.26	**4.0**	0.20	**3.8**	0.08	0.9	0.02	0.3	0.20	3.2
cashdebt	−0.04	−0.6	0.05	1.0	−0.02	−0.6	0.13	1.6	0.01	0.3	0.00	−0.1
cashpr	−0.03	−1.1	−0.02	−0.8	−0.04	−1.8	−0.07	−2.4	−0.05	−1.7	0.01	0.4
cfp	−0.04	−0.7	0.02	0.5	0.07	1.6	−0.08	−0.8	0.03	0.8	0.12	2.5
cfp_ia	0.37	1.1	0.04	0.4	0.17	1.1	−0.03	−0.5	0.07	1.3	0.03	0.5
chatoia	0.09	3.0	0.06	2.5	0.04	1.5	0.03	0.6	0.02	0.5	0.09	**4.0**
chcsho	−0.02	−1.3	0.01	0.2	−0.06	−2.7	−0.04	−1.9	−0.01	−0.4	−0.04	−1.6
chempia	0.04	0.6	−0.01	−0.3	0.00	0.0	0.17	2.9	0.16	**3.7**	0.11	2.2
chfeps	0.08	2.2	0.11	**3.7**	0.19	**7.3**	0.02	0.8	0.03	1.2	0.09	3.0
chinv	0.03	0.6	−0.02	−0.6	−0.06	−1.7	−0.09	−1.8	−0.07	−2.0	−0.06	−1.7
chmom	−0.19	−3.2	−0.03	−0.8	0.06	1.4	−0.11	−1.6	−0.03	−0.7	−0.04	−0.8
chnanalyst	−0.02	−1.1	−0.09	−2.5	−0.09	−2.8	−0.02	−1.0	−0.05	−2.8	−0.08	−3.1
chpmia	0.01	0.2	0.03	1.2	0.03	0.9	0.06	1.5	0.00	−0.1	0.04	0.9
chtx	−0.04	−1.0	−0.01	−0.4	0.08	**4.2**	0.08	2.8	−0.01	−0.4	0.06	3.0
cinvest	−0.06	−2.1	−0.05	−3.1	−0.01	−0.7	0.00	0.0	−0.02	−0.5	−0.03	−0.8
convind	−0.10	−2.1	−0.07	−1.5	−0.26	**−4.2**	−0.01	−0.2	0.04	0.4	−0.08	−0.9
currat	−0.04	−1.0	−0.04	−1.6	0.00	0.1	−0.05	−1.3	−0.04	−1.4	−0.06	−2.0
depr	0.03	0.7	0.05	1.4	0.07	2.0	−0.02	−0.5	0.01	0.2	0.01	0.3
disp	0.02	0.4	−0.03	−1.0	−0.10	**−4.6**	−0.01	−0.3	−0.02	−0.6	−0.05	−2.4

(*continued*)

Table 9
Continued

	Table 5 Fama-MacBeth regressions, pre-2003 Set of stocks, regression method						Table 5 Fama-MacBeth regressions, post-2003 Set of stocks, regression method					
	(A) pre: All stocks, VWLS		(B) pre: All-but-microcaps, OLS		(C) pre: All stocks, OLS		(A) post: All stocks, VWLS		(B) post: All-but-microcaps, OLS		(C) post: All stocks, OLS	
	FM coef.	*t*-stat.	FM coef.	*t*-stat.	FM coef.	*t*-stat.	FM coef.	*t*-stat.	FM coef.	*t*-stat.	FM coef.	*t*-stat.
divi	−0.40	−2.6	−0.20	−1.8	−0.30	−3.0	0.08	0.4	−0.03	−0.3	−0.11	−0.9
divo	0.10	0.6	0.06	0.4	0.14	1.4	0.05	0.4	−0.12	−0.8	−0.18	−1.6
dy	−0.05	−1.0	−0.10	−2.3	−0.05	−1.6	−0.06	−1.1	−0.03	−1.2	−0.06	−1.4
ear	0.09	2.5	0.07	**5.2**	0.09	**5.4**	0.06	1.8	0.07	2.7	0.11	**4.2**
egr	−0.08	−1.7	−0.06	−2.1	−0.04	−1.5	0.00	0.1	0.03	0.9	0.04	1.3
ep	0.19	2.3	0.05	1.3	0.16	**3.8**	0.08	0.7	−0.04	−0.5	0.12	1.6
fgr5yr	0.05	0.4	0.01	0.1	0.08	1.4	0.00	0.1	−0.03	−0.7	−0.03	−1.1
gma	0.08	1.4	0.10	1.7	0.10	1.8	−0.02	−0.3	0.07	1.3	0.04	0.9
grcapx	−0.08	−3.1	−0.04	−2.1	−0.05	−2.4	−0.03	−0.8	−0.05	−2.0	−0.08	**−3.9**
grltnoa	−0.06	−1.7	−0.05	−2.1	−0.04	−1.3	−0.05	−1.0	−0.01	−0.2	0.02	0.6
herf	0.08	2.2	0.02	0.6	−0.05	−1.8	−0.02	−0.8	−0.02	−0.9	−0.04	−2.1
hire	−0.04	−0.6	0.03	0.6	−0.02	−0.4	−0.05	−0.7	−0.08	−1.7	−0.12	−2.1
idiovol	−0.22	−1.7	−0.12	−2.0	−0.23	−2.8	0.19	1.6	0.10	1.5	0.00	0.0
ill	0.36	3.1	−0.03	−1.2	0.56	**8.4**	−0.16	−1.6	−0.14	−1.8	0.25	**4.8**
indmom	0.01	0.2	0.11	1.8	0.41	**5.6**	0.03	0.6	0.10	2.4	0.22	**4.9**
invest	−0.03	−0.5	−0.04	−1.1	−0.09	−2.3	0.10	2.2	0.07	1.7	0.10	2.1
ipo	−0.05	−0.3	−0.17	−1.0	−0.36	−2.9	0.23	1.1	0.20	0.8	−0.25	−0.8
lev	0.13	2.0	0.13	2.1	0.11	2.0	−0.20	−2.0	−0.08	−1.2	−0.18	−1.4
mom1m	−0.65	**−8.4**	−0.49	**−7.6**	−1.07	**−14.9**	−0.23	−2.1	−0.16	−1.9	−0.31	−3.0
mom36m	−0.07	−1.5	−0.05	−1.7	−0.05	−1.2	0.11	2.6	0.02	0.7	0.02	0.4
ms	0.09	2.3	0.09	3.1	0.16	**4.4**	−0.04	−1.1	0.00	0.2	−0.06	−1.8
mve_ia	0.00	0.0	0.02	0.6	0.04	1.4	−0.02	−0.7	−0.02	−0.6	0.01	0.3
nanalyst	0.01	0.2	0.13	2.8	0.50	**7.5**	0.07	1.4	−0.05	−1.1	0.18	2.5
nincr	0.04	1.9	0.09	**4.3**	0.14	**6.4**	0.08	**3.6**	0.06	2.3	0.09	3.1
orgcap	0.00	−0.1	−0.01	−0.3	0.02	0.7	−0.01	−0.1	−0.05	−1.3	−0.02	−0.5
pchcapx_ia	0.05	1.6	0.02	0.9	0.00	0.0	0.01	0.2	0.10	1.8	0.14	1.7
pchcurrat	−0.07	−1.9	−0.02	−1.0	−0.01	−0.3	−0.04	−0.9	−0.01	−0.2	0.00	0.1
pchdepr	0.02	0.7	0.00	−0.2	−0.05	−1.8	0.01	0.1	0.06	1.9	0.02	0.6
pchgm_pchsale	−0.02	−0.7	0.02	0.7	0.02	0.9	0.03	1.0	0.04	1.4	0.04	1.7
pchsale_pchinvt	0.00	0.1	0.03	1.3	0.02	0.9	−0.02	−0.6	0.03	1.7	0.06	2.4
pchsale_pchrect	−0.01	−0.2	−0.05	−1.8	0.01	0.5	−0.03	−0.9	−0.03	−1.0	0.04	1.9
pchsale_pchxsga	−0.05	−1.1	−0.03	−1.2	−0.01	−0.2	0.04	1.0	0.00	0.1	0.01	0.3
pchsaleinv	0.05	1.3	0.01	0.6	0.02	0.8	−0.02	−0.7	−0.06	−1.8	−0.09	−2.3
pctacc	0.00	0.0	0.03	1.5	−0.01	−0.3	0.06	1.7	0.01	0.3	0.00	0.1
pricedelay	0.03	0.6	0.00	0.2	−0.01	−0.4	−0.05	−0.9	0.06	2.2	0.01	0.7
ps	0.04	1.0	0.05	1.6	0.07	2.0	0.02	0.6	−0.02	−0.5	0.02	0.6

(*continued*)

Table 9
Continued

	Table 5 Fama-MacBeth regressions, pre-2003 Set of stocks, regression method						Table 5 Fama-MacBeth regressions, post-2003 Set of stocks, regression method					
	pre: All stocks, VWLS		pre: All-but-microcaps, OLS		(C) pre: All stocks, OLS		(A) post: All stocks, VWLS		(B) post: All-but-microcaps, OLS		(C) post: All stocks, OLS	
	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.
rd	0.00	0.0	0.08	1.3	0.16	2.2	0.12	1.1	0.06	0.6	0.04	0.5
rd_mve	0.10	1.8	0.17	2.9	0.25	**5.4**	0.18	1.6	0.08	2.0	0.23	3.6
rd_sale	0.01	0.1	0.02	0.5	0.06	1.9	−0.06	−1.0	0.00	0.0	0.01	0.1
realestate	0.03	1.1	0.04	1.5	0.02	0.5	0.02	0.7	0.04	1.5	0.04	1.7
retvol	−0.78	**−5.5**	−0.39	**−7.5**	−0.49	**−5.9**	0.00	0.0	0.00	0.0	−0.36	−2.9
roaq	−0.01	−0.2	0.01	0.1	0.08	1.5	0.04	0.4	0.04	0.8	0.09	1.5
roavol	0.00	0.0	0.05	1.2	0.03	0.8	−0.03	−0.4	0.01	0.3	−0.03	−0.8
roic	0.19	2.9	0.06	1.5	0.06	1.6	−0.02	−0.2	−0.02	−0.4	−0.10	−2.2
rsup	0.02	0.3	0.05	2.0	0.06	1.9	0.08	1.4	0.04	1.1	0.04	1.1
salecash	0.00	0.2	0.01	0.5	−0.02	−0.8	0.01	0.4	0.03	1.0	−0.04	−1.3
saleinv	−0.03	−1.5	0.00	−0.3	0.03	2.2	−0.04	−1.6	0.02	1.0	0.04	1.3
salerec	0.03	0.8	0.02	0.6	0.01	0.3	0.02	0.4	0.02	0.5	−0.01	−0.3
secured	−0.01	−0.2	−0.01	−0.6	−0.03	−1.2	−0.04	−0.9	−0.04	−1.4	−0.05	−1.9
securedind	0.18	1.2	0.27	1.4	0.17	0.8	0.05	1.4	0.01	0.2	0.08	1.5
sfe	−0.26	−2.5	−0.18	**−3.8**	−0.08	−2.2	−0.08	−0.8	0.01	0.1	0.16	2.4
sgr	−0.11	−1.9	0.00	0.0	−0.01	−0.4	−0.11	−1.7	−0.01	−0.2	−0.08	−1.6
sin	0.19	0.9	0.16	0.9	0.54	2.6	0.49	3.3	0.31	1.2	0.45	1.8
sp	−0.07	−0.8	0.03	0.6	−0.06	−1.1	0.14	1.9	0.05	1.0	0.18	2.3
std_dolvol	−0.13	−1.6	−0.07	−2.3	−0.17	**−3.5**	0.07	0.5	0.02	0.7	−0.17	−2.2
std_turn	0.21	**4.0**	0.37	**10.1**	0.58	**13.8**	−0.05	−0.7	−0.04	−0.8	0.12	1.2
stdcf	−0.03	−1.2	−0.01	−0.6	−0.02	−1.0	−0.07	−1.5	−0.06	−1.5	−0.03	−0.7
sue	0.04	1.1	0.05	2.2	0.10	**3.5**	0.16	1.4	0.10	2.5	0.16	**5.2**
tang	−0.06	−1.5	−0.03	−0.9	0.05	1.3	0.02	0.5	0.02	0.5	−0.05	−0.8
tb	0.02	0.7	0.03	1.0	0.03	1.7	0.05	1.5	0.04	1.2	0.06	2.1
turn	−0.17	−2.5	−0.40	**−7.8**	−0.66	**−12.3**	0.01	0.1	−0.06	−1.0	−0.32	**−4.0**
zerotrade	−0.24	**−4.4**	−0.07	**−3.7**	−0.45	**−8.3**	−0.12	−1.9	0.14	1.8	−0.12	−2.2
Mean # obs.	4,909		1,896		4,909		4,027		1,720		4,027	
Mean adj. R^2	28.0%		15.4%		7.7%		29.8%		1 5.7%		8.3%	

This table presents the results of estimating the regressions reported in Table 6 separately pre-2003 versus post-2003, where post-2003 is defined as beginning in January 2004. All 94 firm characteristics are simultaneously included as independent variables. Missing characteristic observations are set to the zero mean of the nonmissing values of the characteristic in that month after the nonmissing values have been winsorized at the 1st and 99th percentiles and standardized to have a zero mean and unit standard deviation. Three regressions are shown: using all stocks and WLS where the weight is the market value of equity for stock i at time $t-1$ (all stocks, VWLS), using all-but-microcap stocks and OLS (all-but-microcap stocks, OLS), and using all stocks and OLS (all stocks, OLS). Microcaps are stocks in month $t-1$ with a market value of equity less than the 20th percentile of stocks on the NYSE stock exchange in month $t-1$. FM coefficients are the means of the monthly estimated coefficients*100; t-statistics employ Newey-West adjustments of 12 lags. The sample is all common stocks on the NYSE, AMEX, and NASDAQ exchanges with available annual and quarterly Compustat accounting data and CRSP stock return data. Analyst data are from I/B/E/S. Estimated coefficients with a two-tailed p-value $\leqslant 0.05$ adjusted using the dependent failure discovery rate method (DFDR; Benjamini and Yekutieli 2001) are shown with a bold t-statistic. FM, Fama-MacBeth.

Following the adoption of Reg. FD in October 2000 and the decimalization of quotes in January 2001, both of which reduced effective spreads, price impact, and trading costs (Bessembinder 2003; Eleswarapu, Thompson, and Venkataraman 2004), the information environment changed in two ways. July 2002 saw the passing of the Sarbanes-Oxley Act, which increased auditing quality requirements and imposed managerial responsibility on the quality of firms' internal controls for financial statements, beginning with those for fiscal 2002 that started to be reported in early 2003. Then, beginning in November 2002, the deadlines for filing annual and quarterly reports with the SEC following a fiscal year-end or quarter-end were accelerated, to make 10-Q and 10-K filings more timely.

Importantly from a technological perspective, between January and May 2003 the NYSE introduced its autoquoting software, a change that Hendershott, Jones, and Menkveld (2011) argue led to dramatic reductions in trading frictions and costs and an equally dramatic increase in the algorithmic trading that long/short equity hedge funds use to implement long/short quantitative trading strategies. While the temporal confluence of these changes makes it difficult to causally identify one or more of them as explaining the shifts seen in the monthly return generating process, we propose that in total the changes made it much cheaper and technologically easier to rapidly implement high-volume quantitative long/short trading strategies, thereby increasing arbitrage activity, increasing the efficiency of the stock market, and to the degree that the statistically and economically significant pre-2003 pricing of characteristics reflected high costly limits to arbitrage, decreasing the number and influence of characteristics in determining average returns after 2003.[14]

3.5 Robustness tests

In this section, we summarize the results of additional analyses that are detailed in the Internet Appendix and that assess the robustness of our main findings. First, because our multivariate approach to identifying the independent determinants of average returns is made feasible by our method of replacing missing characteristic values with the characteristic's standardized mean for that month (zero), we re-estimate Table 4 without replacing missing characteristic values.[15] We find very few differences in the identities or numbers of significant Fama-MacBeth mean slope coefficients, providing reassurance that at least at the univariate level, and also for any given characteristic at the level of controlling for only the characteristics versions of notable benchmark

[14] We do not ascribe the shift in late 2002/early 2003 to the post-publication reduction in the univariate mean hedge return documented by McLean and Pontiff (2016). This is because we do not observe a spike in the number of anomalies either discovered or published in late 2002/early 2003 (see Green, Hand, and Zhang 2013, figure 1; Harvey, Liu, and Zhang 2016, figure 1). We also differentiate our study from that of Chordia, Subrahmanyam, and Tong (2014, CST) because while CST document an attenuation over calendar time in the returns to anomalies, they analyze only six anomalies and test only for a linear or exponential decay in anomaly hedge returns.

[15] Internet Appendix, Table IA-1.

factor models, the replacement of missing values with standardized means does not materially affect or drive our results.

Second, given the sharp reduction in characteristics-based return predictability in 2003 that we document in Tables 8 and 9, we separately re-estimate the univariate analyses in Column A of Table 4 for pre- and post-2003.[16] In doing so, we confirm that drop in characteristics-based return predictability in 2003 happens at both the univariate and multivariate levels. We also confirm our finding based on the full window from 1980 to 2014, namely the fact that the characteristics that matter switch from being primarily fundamental at the univariate level to primarily market-based at the multivariate level does not arise simply because the full window 1980–2014 pools data from two very different pre- versus post-2003 subperiods. We observe that in a manner similar to that seen over 1980–2014 as a whole, prior to 2003 10 of 12 univariately significant characteristics are fundamental, as compared to 2 of 12 independent characteristics. After 2003 the contrast is moot, in that there are no univariately significant characteristics and just two independent characteristics.

Third, since we use hedge returns to assess the economic importance of characteristics-based return predictability, we confirm that the magnitude and significance of the hedge returns we report in Tables 7 and 8 are not driven by a failure to control for common factor exposures. We do so by regressing the hedge portfolio returns that exploit the full set of 94 characteristics on the monthly factor returns from the Carhart, Fama-French five-factor, and Hou-Xue-Zhang q-factor models.[17] We also repeat these tests (unreported) for the periods 1990–2002 and 2004–2014 and find results very similar to those reported in Table 8.

Fourth, we re-estimate many of our regressions separately for big, small, and microcap stocks by using both VWLS and OLS.[18] The findings and inferences from these mutually exclusive size-based analyses are similar to those we present in the main results of the paper. In our Internet Appendix, we provide visual depictions of the overlaps in which characteristics are and are not significant across alternative regression specifications.[19]

3.6 Limitations

We recognize several caveats to and limitations of our research. While our study is the first to assess the simultaneous predictive power of a very large number of individual firm characteristics, we examine only one-quarter of the 430+ characteristics reported in the anomalies literature to date. Thus, we likely do not identify all the truly independent determinants of average monthly

16 Internet Appendix, Table IA-2.

17 Internet Appendix, Table IA-3.

18 Internet Appendix, Table IA-4.

19 Internet Appendix, Figures IA-1 and IA-2.

returns. It may also not be appropriate to extrapolate our findings to the full population of 430+ characteristics, because our approach has been to focus on those that can be calculated from CRSP, Compustat, or I/B/E/S data, meaning that we have not sampled firm characteristics randomly from the population of characteristics. We may also have introduced measurement error through our approach to treating missing data and by aligning characteristics in calendar month time rather than in the daily or weekly time used in the original research. In addition, we note that while in our quasi out-of-sample tests we have sought to avoid using data that were unavailable in real time, implementing the positions dictated by our hedge portfolios would expose investors to high trading costs, especially in microcap stocks, such that the resultant net-of-trading-costs hedge returns might not be positive (Novy-Marx and Velikov 2016).

4. Conclusions and Implications

In this paper we have sought to respond to the challenge posed by Cochrane (2011, 1060) that researchers begin to identify which of the "veritable zoo" of hundreds of firm characteristics reported in prior studies are statistically significant predictors of the cross-section of average stock returns. We use the same approach as Fama and French's seminal 1992 paper but employ a much larger set of 94 firm characteristics and simultaneously include all 94 of these as explanatory variables in Fama-MacBeth regressions that seek to avoid overweighting microcaps and adjust for data-snooping biases. Our approach leads us to estimate that 12 characteristics provide significant independent information about average U.S. monthly returns in non-microcap stocks over the full period from 1980 to 2014, with the remaining 82 characteristics providing no independent information. We show that the reason that few characteristics provide independent information is because the number of independent determinants of returns is intrinsically small, rather than because a small number of characteristics are able to absorb the information in many other characteristics that are univariately important.

We also document that the number of independent characteristics and their ability to yield positive hedge returns fell sharply in 2003, especially in non-microcap stocks. We estimate that since 2003 just two characteristics have been independent determinants in non-microcap returns, as compared to 12 characteristics before 2003, and that the mean hedge return to non-micro-cap stocks has been insignificantly different from zero after 2003. Characteristics-based predictability currently exists only in microcap stocks. We interpret the decline in the number and economic importance of the firm characteristics over calendar time and across firm size as being most consistent with the costly limits-to-arbitrage market efficiency arguments of Shleifer and Vishny (1997) and others.

In conclusion, by identifying the independent determinants of average monthly returns, and by relaxing the constraints that the independent determinants be the same across firm size and over time, we provide additional evidence beyond that of Harvey, Liu, and Zhu (2016) and McLean and Pontiff (2016) that the inferences that have been made from hundreds of return anomaly studies warrant substantial skepticism. At the same time, we also surface new facts and puzzles to be digested, the most prominent of which is the strong, sudden, and seemingly permanent decline in the characteristics-based predictability of returns in 2003, especially among non-microcap stocks. Our results suggest that future empirical models of average returns may benefit from weighting post-2003 data more strongly than pre-2003 data, as well as from conditioning the return generating process on firm size and from using as controls the characteristics that we identify (pre- versus post-2003) as being independent.

Appendix

Acronym	Author(s)	Date, Journal	Definition of the characteristic-based anomaly variable
absacc	Bandyopadhyay, Huang, and Wirjanto	2010, WP	Absolute value of *acc*
acc	Sloan	1996, TAR	Annual income before extraordinary items (*ib*) minus operating cash flows (*oancf*) divided by average total assets (*at*); if *oancf* is missing then set to change in *act* - change in *che* - change in *lct* + change in *dlc* + change in *txp-dp*
aeavol	Lerman, Livnat, and Mendenhall	2008, WP	Average daily trading volume (*vol*) for 3 days around earnings announcement minus average daily volume for 1-month ending 2 weeks before earnings announcement divided by 1-month average daily volume. Earnings announcement day from Compustat quarterly (*rdq*)
age	Jiang, Lee, and Zhang	2005, RAS	Number of years since first Compustat coverage
agr	Cooper, Gulen, and Schill	2008, JF	Annual percent change in total assets (*at*)
baspread	Amihud and Mendelson	1989, JF	Monthly average of daily bid-ask spread divided by average of daily spread
beta	Fama and MacBeth	1973, JPE	Estimated market beta from weekly returns and equal weighted market returns for 3 years ending month *t*-1 with at least 52 weeks of returns
betasq	Fama and MacBeth	1973, JPE	Market beta squared
bm	Rosenberg, Reid, and Lanstein	1985, JPM	Book value of equity (*ceq*) divided by end of fiscal year-end market capitalization
bm_ia	Asness, Porter, and Stevens	2000, WP	Industry adjusted book-to-market ratio
cash	Palazzo	2012, JFE	Cash and cash equivalents divided by average total assets
cashdebt	Ou and Penman	1989, JAE	Earnings before depreciation and extraordinary items (*ib*+*dp*) divided by avg. total liabilities (*lt*)
cashpr	Chandrashekar and Rao	2009, WP	Fiscal year-end market capitalization plus long-term debt (*dltt*) minus total assets (*at*) divided by cash and equivalents (*che*)
cfp	Desai, Rajgopal, and Venkatachalam	2004, TAR	Operating cash flows divided by fiscal-year-end market capitalization

(*continued*)

Acronym	Author(s)	Date, Journal	Definition of the characteristic-based anomaly variable
cfp_ia	Asness, Porter and Stevens	2000, WP	Industry adjusted *cfp*
chatoia	Soliman	2008, TAR	2-digit SIC - fiscal-year mean-adjusted change in sales (*sale*) divided by average total assets (*at*)
chcsho	Pontiff and Woodgate	2008, JF	Annual percent change in shares outstanding (*csho*)
chempia	Asness, Porter, and Stevens	1994, WP	Industry-adjusted change in number of employees
chfeps	Hawkins, Chamberlin, and Daniel	1984, FAJ	Mean analyst forecast in month prior to fiscal period end date from I/B/E/S summary file minus same mean forecast for prior fiscal period using annual earnings forecasts
chinv	Thomas and Zhang	2002, RAS	Change in inventory (*inv*) scaled by average total assets (*at*)
chmom	Gettleman and Marks	2006, WP	Cumulative returns from months t-6 to t-1 minus months t-12 to t-7
chnanalyst	Scherbina	2008 RF	Change in *nanalyst* from month t-3 to month t
chpmia	Soliman	2008, TAR	2-digit SIC - fiscal-year mean adjusted change in income before extraordinary items (*ib*) divided by sales (*sale*)
chtx	Thomas and Zhang	2011, JAR	Percent change in total taxes (*txtq*) from quarter$t$-4 to t
cinvest	Titman, Wei, and Xie	2004, JFQA	Change over one quarter in net PP&E (*ppentq*) divided by sales (*saleq*) - average of this variable for prior 3 quarters; if *saleq* = 0, then scale by 0.01
convind	Valta	2016, JFQA	An indicator equal to 1 if company has convertible debt obligations
currat	Ou and Penman	1989, JAE	Current assets / current liabilities
depr	Holthausen and Larcker	1992, JAE	Depreciation divided by PP&E
disp	Diether, Malloy, and Scherbina	2002, JF	Standard deviation of analyst forecasts in month prior to fiscal period end date divided by the absolute value of the mean forecast; if *meanest* = 0, then scalar set to 1. Forecast data from I/B/E/S summary files
divi	Michaely, Thaler, and Womack	1995, JF	An indicator variable equal to 1 if company pays dividends but did not in prior year
divo	Michaely, Thaler, and Womack	1995, JF	An indicator variable equal to 1 if company does not pay dividend but did in prior year
dolvol	Chordia, Subrahmanyam, and Anshuman	2001, JFE	Natural log of trading volume times price per share from month t-2
dy	Litzenberger and Ramaswamy	1982, JF	Total dividends (*dvt*) divided by market capitalization at fiscal year-end
ear	Kishore et al.	2008, WP	Sum of daily returns in three days around earnings announcement. Earnings announcement from Compustat quarterly file (*rdq*)
egr	Richardson et al.	2005, JAE	Annual percent change in book value of equity (*ceq*)
ep	Basu	1977, JF	Annual income before extraordinary items (*ib*) divided by end of fiscal year market cap
fgr5yr	Bauman and Dowen	1988, FAJ	Most recently available analyst forecasted 5-year growth
gma	Novy-Marx	2013, JFE	Revenues (*revt*) minus cost of goods sold (*cogs*) divided by lagged total assets (at)
grCAPX	Anderson and Garcia-Feijoo	2006, JF	Percent change in capital expenditures from year$t$-2 to year t
grltnoa	Fairfield, Whisenant, and Yohn	2003, TAR	Growth in long-term net operating assets
herf	Hou and Robinson	2006, JF	2-digit SIC - fiscal-year sales concentration (sum of squared percent of sales in industry for each company).

(*continued*)

Acronym	Author(s)	Date, Journal	Definition of the characteristic-based anomaly variable
hire	Bazdresch, Belo, and Lin	2014, JPE	Percent change in number of employees (*emp*)
idiovol	Ali, Hwang, and Trombley	2003, JFE	Standard deviation of residuals of weekly returns on weekly equal weighted market returns for 3 years prior to month end
ill	Amihud	2002, JFM	Average of daily (absolute return / dollar volume).
indmom	Moskowitz and Grinblatt	1999, JF	Equal weighted average industry 12-month returns
invest	Chen and Zhang	2010, JF	Annual change in gross property, plant, and equipment (*ppegt*) + annual change in inventories (*invt*) all scaled by lagged total assets (*at*)
IPO	Loughran and Ritter	1995, JF	An indicator variable equal to 1 if first year available on CRSP monthly stock file
lev	Bhandari	1988, JF	Total liabilities (lt) divided by fiscal year-end market capitalization
lgr	Richardson et al.	2005, JAE	Annual percent change in total liabilities (*lt*)
maxret	Bali, Cakici, and Whitelaw	2011, JFE	Maximum daily return from returns during calendar month$t$-1
mom12m	Jegadeesh	1990, JF	11-month cumulative returns ending one month before month end
mom1m	Jegadeesh and Titman	1993, JF	1-month cumulative return
mom36m	Jegadeesh and Titman	1993, JF	Cumulative returns from months$t$-36 to t-13
mom6m	Jegadeesh and Titman	1993, JF	5-month cumulative returns ending one month before month end
ms	Mohanram	2005, RAS	Sum of 8 indicator variables for fundamental performance
mve	Banz	1981, JFE	Natural log of market capitalization at end of month t-1
mve_ia	Asness, Porter, and Stevens	2000, WP	2-digit SIC industry-adjusted fiscal year-end market capitalization
nanalyst	Elgers, Lo, and Pfeiffer	2001, TAR	Number of analyst forecasts from most recently available I/B/E/S summary files in month prior to month of portfolio formation. *nanalyst* set to zero if not covered in I/B/E/S summary file
nincr	Barth, Elliott, and Finn	1999, JAR	Number of consecutive quarters (up to eight quarters) with an increase in earnings (*ibq*) over same quarter in the prior year
operprof	Fama and French	2015, JFE	Revenue minus cost of goods sold - SG&A expense - interest expense divided by lagged common shareholders' equity
orgcap	Eisfeldt and Papanikolaou	2013, JF	Capitalized SG&A expenses
pchcapx_ia	Abarbanell and Bushee	1998, TAR	2-digit SIC - fiscal-year mean-adjusted percent change in capital expenditures (*capx*)
pchcurrat	Ou and Penman	1989, JAE	Percent change in *currat.*
pchdepr	Holthausen and Larcker	1992, JAE	Percent change in *depr*
pchgm_pchsale	Abarbanell and Bushee	1998, TAR	Percent change in gross margin (*sale-cogs*) minus percent change in sales (*sale*)
pchquick	Ou and Penman	1989, JAE	Percent change in *quick*
pchsale_pchinvt	Abarbanell and Bushee	1998, TAR	Annual percent change in sales (*sale*) minus annual percent change in inventory (*invt*).
pchsale_pchrect	Abarbanell and Bushee	1998, TAR	Annual percent change in sales (*sale*) minus annual percent change in receivables (*rect*)
pchsale_pchxsga	Abarbanell and Bushee	1998, TAR	Annual percent change in sales (*sale*) minus annual percent change in SG&A (*xsga*)
pchsaleinv	Ou and Penman	1989, JAE	Percent change in *saleinv*
pctacc	Hafzalla, Lundholm, and Van Winkle	2011, TAR	Same as *acc* except that the numerator is divided by the absolute value of *ib*; if *ib* = 0 then *ib* set to 0.01 for denominator

(*continued*)

Acronym	Author(s)	Date, Journal	Definition of the characteristic-based anomaly variable
pricedelay	Hou & Moskowitz	2005, RFS	The proportion of variation in weekly returns for 36 months ending in month*t* explained by 4 lags of weekly market returns incremental to contemporaneous market return
ps	Piotroski	2000, JAR	Sum of 9 indicator variables to form fundamental health score
quick	Ou and Penman	1989, JAE	(current assets - inventory) / current liabilities
rd	Eberhart, Maxwell, and Siddique	2004, JF	An indicator variable equal to 1 if R&D expense as a percentage of total assets has an increase greater than 5%.
rd_mve	Guo, Lev, and Shi	2006, JBFA	R&D expense divided by end-of-fiscal-year market capitalization
rd_sale	Guo, Lev, and Shi	2006, JBFA	R&D expense divided by sales (*xrd*/*sale*)
realestate	Tuzel	2010, RFS	Buildings and capitalized leases divided by gross PP&E
retvol	Ang et al.	2006, JF	Standard deviation of daily returns from month t-1
roaq	Balakrishnan, Bartov, and Faurel	2010, JAE	Income before extraordinary items (*ibq*) divided by one quarter lagged total assets (*atq*)
roavol	Francis et al.	2004, TAR	Standard deviation for 16 quarters of income before extraordinary items (*ibq*) divided by average total assets (*atq*)
roeq	Hou, Xue, and Zhang	2015 RFS	Earnings before extraordinary items divided by lagged common shareholders' equity
roic	Brown and Rowe	2007, WP	Annual earnings before interest and taxes (*ebit*) minus nonoperating income (*nopi*) divided by non-cash enterprise value (*ceq*+*lt*-*che*)
rsup	Kama	2009, JBFA	Sales from quarter t minus sales from quarter t-4 (*saleq*) divided by fiscal-quarter-end market capitalization (*cshoq* * *prccq*)
salecash	Ou and Penman	1989, JAE	Annual sales divided by cash and cash equivalents
saleinv	Ou and Penman	1989, JAE	Annual sales divided by total inventory
salerec	Ou and Penman	1989, JAE	Annual sales divided by accounts receivable
secured	Valta	2016, JFQA	Total liability scaled secured debt
securedind	Valta	2016, JFQA	An indicator equal to 1 if company has secured debt obligations
sfe	Elgers, Lo, and Pfeiffer	2001, TAR	Analysts mean annual earnings forecast for nearest upcoming fiscal year from most recent month available prior to month of portfolio formation from I/B/E/S summary files scaled by price per share at fiscal quarter end
sgr	Lakonishok, Shleifer, and Vishny	1994, JF	Annual percent change in sales (*sale*)
sin	Hong & Kacperczyk	2009, JFE	An indicator variable equal to 1 if a company's primary industry classification is in smoke or tobacco, beer or alcohol, or gaming
SP	Barbee, Mukherji, and Raines	1996, FAJ	Annual revenue (*sale*) divided by fiscal year-end market capitalization
std_dolvol	Chordia, Subrahmanyam, and Anshuman	2001, JFE	Monthly standard deviation of daily dollar trading volume
std_turn	Chordia, Subrahmanyam, and Anshuman	2001, JFE	Monthly standard deviation of daily share turnover
stdacc	Bandyopadhyay, Huang, and Wirjanto	2010, WP	Standard deviation for 16 quarters of accruals (*acc* measured with quarterly Compustat) scaled by sales; if *saleq* = 0, then scale by 0.01

(*continued*)

Acronym	Author(s)	Date, Journal	Definition of the characteristic-based anomaly variable
stdcf	Huang	2009, JEF	Standard deviation for 16 quarters of cash flows divided by sales (*saleq*); if *saleq* = 0, then scale by 0.01. Cash flows defined as *ibq* minus quarterly accruals
sue	Rendelman, Jones, and Latane	1982, JFE	Unexpected quarterly earnings divided by fiscal-quarter-end market cap. Unexpected earnings is I/B/E/S actual earnings minus median forecasted earnings if available, else it is the seasonally differenced quarterly earnings before extraordinary items from Compustat quarterly file
tang	Almeida and Campello	2007, RFS	Cash holdings + 0.715 × receivables +0.547 × inventory + 0.535 × PPE/ totl assets
tb	Lev and Nissim	2004, TAR	Tax income, calculated from current tax expense divided by maximum federal tax rate, divided by income before extraordinary items
turn	Datar, Naik, and Radcliffe	1998, JFM	Average monthly trading volume for most recent 3 months scaled by number of shares outstanding in current month
zerotrade	Liu	2006, JFE	Turnover weighted number of zero trading days for most recent 1 month

References

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

Afifi, A. A., and R. M. Elashoff. 1966. Missing observations in multivariate statistics: I. Review of the literature. *Journal of the American Statistical Association* 61:595–604.

Ali, A., L.-S. Hwang, and M. A. Trombley. 2003. Arbitrage risk and the book-to-market anomaly. *Journal of Financial Economics* 69:355–73.

Almeida, H., and M. Campello. 2007. Financial constraints, asset intangibility, and corporate investment. *Review of Financial Studies* 20:1429–60.

Amihud, Y. 2002. Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets* 5:31–56.

Amihud, Y., and H. Mendelson. 1989. The effects of beta, bid-ask spread, residual risk, and size on stock returns. *Journal of Finance* 44:479–86.

Anderson, C. W., and L. Garcia-Feijoo. 2006. Empirical evidence on capital investment, growth options, and security returns. *Journal of Finance* 61:171–94.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang. 2006. The cross section of volatility and expected returns. *Journal of Finance* 61:259–99.

Asness, C. S., R. B. Porter, and R. L. Stevens. 2000. Predicting stock returns using industry-relative firm characteristics. Working Paper, AQR Capital Management.

Balakrishnan, K., E. Bartov, and L. Faurel. 2010. Post loss/profit announcement drift. *Journal of Accounting & Economics* 50:20–41.

Bali, T., N. Cakici, and R. Whitelaw. 2011. Maxing out: Stocks as lotteries and the cross section of expected returns. *Journal of Financial Economics* 99:427–46.

Bandyopadhyay, S. P., A. G. Huang, and T. S. Wirjanto. 2010. The accrual volatility anomaly. Working Paper, University of Waterloo.

Banz, R. W. 1981. The relationship between return and market value of common stocks. *Journal of Financial Economics* 9:3–18.

Barbee, W. C., S. Mukherji, and G. A. Raines. 1996. Do the sales-price and debt-equity ratios explain stock returns better than the book-to-market value of equity ratio and firm size? *Financial Analysts Journal* 52:56–60.

Barth, M., J. Elliott, and M. Finn. 1999. Market rewards associated with patterns of increasing earnings. *Journal of Accounting Research* 37:387–413.

Basu, S. 1977. Investment performance of common stocks in relation to their price-earnings ratios: A test of market efficiency. *Journal of Finance* 32:663–82.

Bauman, W. S., and R. Dowen. 1988. Growth projections and common stock returns. *Financial Analysts Journal* 44:79–80.

Bazdresch, S., F. Belo, and X. Lin. 2014. Labor hiring, investment, and stock return predictability in the cross section. *Journal of Political Economy* 122:129–77.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)* 57:289–300.

Benjamini, Y., and D. Yekutieli. 2001. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* 29:1165–88.

Bessembinder, H. 2003. Trade execution costs and market quality after decimalization. *Journal of Financial and Quantitative Analysis* 38:747–77.

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

Brown, D. P., and B. Rowe. 2007. The productivity premium in equity returns. Working Paper, University of Wisconsin.

Carhart, M. M. 1997. On persistence in mutual fund performance. *Journal of Finance* 59:3–32.

Chandrashekar, S., and R. K. S. Rao. 2009. The productivity of corporate cash holdings and the cross section of expected stock returns. Working Paper, University of Texas at Austin.

Chen, L., and L. Zhang. 2010. A better three-factor model that explains more anomalies. *Journal of Finance* 65:563–95.

Chordia, T., R. Roll, and A. Subrahmanyam. 2001. Market liquidity and trading activity. *Journal of Finance* 56:501–30.

———. 2008. Liquidity and market efficiency. *Journal of Financial Economics* 87:249–68.

———. 2011. Recent trends in trading activity and market quality. *Journal of Financial Economics* 101:243–63.

Chordia, T., A. Subrahmanyam, and R. Anshuman. 2001. Trading activity and expected stock returns. *Journal of Financial Economics* 59:3–32.

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 & Economics* 58:41–58.

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

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

Datar, V. T., N. Y. Naik, and R. Radcliffe. 1998. Liquidity and stock returns: An alternative test. *Journal of Financial Markets* 1:203–19.

DeMiguel, V., A. Martín-Utrera, F. J. Nogales, and R. Uppal. 2016. Firm characteristics and stock returns: An investment perspective. Working Paper, London Business School.

Desai, H., S. Rajgopal, and M. Venkatachalam. 2004. Value-glamour and accruals mispricing: One anomaly or two? *Accounting Review* 79:355–85.

Diether, K., C. Malloy, and A. Scherbina. 2002. Differences of opinion and the cross section of stock returns. *Journal of Finance* 57:2113–41.

Eberhart, A. C., W. F. Maxwell, and A. R. Siddique. 2004. An examination of long-term abnormal stock returns and operating performance following R&D increases. *Journal of Finance* 59:623–50.

Eisfeldt, A. L., and D. Papanikolaou. 2013. Organization capital and the cross section of expected returns. *Journal of Finance* 68:1365–406.

Eleswarapu, V., R. Thompson, and K. Venkataraman. 2004. The impact of regulation fair disclosure: Trading costs and information asymmetry. *Journal of Financial and Quantitative Analysis* 39:209–25.

Elgers, P. T., M. H. Lo, and R. J. Pfeiffer, Jr. 2001. Delayed security price adjustments to financial analysts' forecasts of annual earnings. *Accounting Review* 76:613–32.

Engelberg, J., R. D. McLean, and J. Pontiff. 2016. Anomalies and news. Working Paper, University of California at San Diego.

Fairfield, P. M., J. S. Whisenant, and T. L. Yohn. 2003. Accrued earnings and growth: Implications for future earnings performance and market mispricing. *Accounting Review* 78:353–71.

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

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

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

———. 2008. Dissecting anomalies. *Journal of Finance* 63:1653–78.

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

———. 2016. Dissecting anomalies with a five-factor model. *Review of Financial Studies* 29:69–103.

Fama, E., and J. MacBeth. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81:607–36.

Francis, J., R. LaFond, P. M. Olsson, and K. Schipper. 2004. Costs of equity and earnings attributes. *Accounting Review* 79:967–1010.

French, K. 2008. The costs of active trading. *Journal of Finance* 63:1537–73.

Gettleman, E., and J. M. Marks. 2006. Acceleration strategies. Working Paper, University of Illinois at Urbana-Champaign.

Green, J., J. Hand, and M. Soliman. 2011. Going, going, gone? The apparent demise of the accruals anomaly. *Management Science* 55:797–816.

Green, J., J. R. M. Hand, and X. F. Zhang. 2013. The supraview of return predictive signals. *Review of Accounting Studies* 18:692–730.

Greene, W. H. 2011. *Econometric analysis*, 7th ed. Upper Saddle River, NJ: Prentice Hall.

Guo, R., B. Lev, and C. Shi. 2006. Explaining the short- and long-term IPO anomalies in the US by R&D. *Journal of Business, Finance & Accounting* 33:550–79.

Hafzalla, N., R. Lundholm, and M. Van Winkle. 2011. Percent accruals. *Accounting Review* 86:209–36.

Harvey, C. R., Y. Liu, and H. Zhu. 2016. …And the cross section of expected returns. *Review of Financial Studies* 29:5–68.

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

Hawkins, E. H., S. C. Chamberlin, and W. E. Daniel. 1984. Earnings expectations and security prices. *Financial Analysts Journal* 40:24–27, 30–38, 74.

Hendershott, T., C. M. Jones, and A. J. Menkveld. 2011. Does algorithmic trading improve liquidity? *Journal of Finance* 66:1–33.

Holthausen, R. W., and D. F. Larcker. 1992. The prediction of stock returns using financial statement information. *Journal of Accounting & Economics* 15:373–412.

Hong, M., and M. Kacperczyk. 2009. The price of sin: The effects of social norms on markets. *Journal of Financial Economics* 93:15–36.

Hou, K., and T. J. Moskowitz. 2005. Market frictions, price delay, and the cross section of expected returns. *Review of Financial Studies* 18:981–1020.

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.

———. . A comparison of new factor models. Working Paper, The Ohio State University.

Huang, A. G. 2009. The cross section of cash flow volatility and expected stock returns. *Journal of Empirical Finance* 16:409–29.

Jacobs, B. I., and K. N. Levy. 1988. Disentangling equity return regularities: New insights and investment opportunities. *Financial Analysts Journal* 44:18–43.

Jegadeesh, N. 1990. Evidence of predictable behavior of security returns. *Journal of Finance* 45:881–98.

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

Jiang, G., C. M. C. Lee, and G. Y. Zhang. 2005. Information uncertainty and expected returns. *Review of Accounting Studies* 10:185–221.

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

Kama, I. 2009. On the market reaction to revenue and earnings surprises. *Journal of Business Finance & Accounting* 36:31–50.

Kishore, R., M. W. Brandt, P. Santa-Clara, and M. Venkatachalam. 2008. Earnings announcements are full of surprises. Working Paper, Duke University.

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

Lam, F. Y. E. C., and K. C. J. Wei. 2011. Limits-to-arbitrage, investment frictions, and the asset growth anomaly. *Journal of Financial Economics* 102:127–49.

Lerman, A., J. Livnat, and R. R. Mendenhall. 2008. The high-volume return premium and post-earnings announcement drift. Working Paper, New York University.

Lesmond, D. A., M. J. Schill, and C. Zhou. 2004. The illusory nature of momentum profits. *Journal of Financial Economics* 71:349–80.

Lev, B., and D. Nissim. 2004. Taxable income, future earnings, and equity values. *Accounting Review* 79:1039–74.

Lewellen, J. 2015. The cross section of expected stock returns. *Critical Finance Review* 4:1–44.

Li, E.X.N., D. Livdan, and L. Zhang. 2009. Anomalies. *Review of Financial Studies* 22:4301–34.

Li, D., and L. Zhang. 2010. Does q-theory with investment frictions explain anomalies in the cross section of returns? *Journal of Financial Economics* 98:297–314.

Light, N., D. Maslov, and O. Rytchkov. 2016. Aggregation of information about the cross section of stock returns: A latent variable approach. *Review of Financial Studies*. Advance Access published December 26, 2016, 10.1093/rfs/hhw102.

Linnainmaa, J. T., and M. Roberts. 2016. The history of the cross section of stock returns. Working Paper, University of Chicago.

Litzenberger, R. H., and K. Ramaswamy. 1982. The effects of dividends on common stock prices: Tax effects or information effects? *Journal of Finance* 37:429–43.

Liu, W. 2006. A liquidity-augmented capital asset pricing model. *Journal of Financial Economics* 82:631–71

Loughran, T., and J. R. Ritter. 1995. The new issues puzzle. *Journal of Finance* 50:23–51.

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

Michaely, R., R. H. Thaler, and K. L. Womack. 1995. Price reactions to dividend initiations and omissions: Overreaction or drift? *Journal of Finance* 50:573–608.

Mohanram, P. S. 2005. Separating winners from losers among low book-to-market stocks using financial statement analysis. *Review of Accounting Studies* 10:133–70.

Moskowitz, T. J., and M. Grinblatt. 1999. Do industries explain momentum? *Journal of Finance* 54:1249–90.

Newey, W. K., and W. K. West. 1994. Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61:631–54.

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

Novy-Marx, R., and M. Velikov. 2016. A taxonomy of anomalies and their trading costs. *Review of Financial Studies* 29:104–47.

Ou, J. A., and S. H. Penman. 1989. Financial statement analysis and the prediction of stock returns. *Journal of Accounting & Economics* 11:295–329.

Palazzo, B. 2012. Cash holdings, risk, and expected returns. *Journal of Financial Economics* 104:162–85.

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

Pontiff, J., and A. Woodgate. 2008. Share issuance and cross-sectional returns. *Journal of Finance* 63:921–45.

Rendleman, R. J., C. P. Jones, and H. A. Latané. 1982. Empirical anomalies based on unexpected earnings and the importance of risk adjustments. *Journal of Financial Economics* 10:269–87.

Richardson, S. A., R. G. Sloan, M. T. Soliman, and I. Tuna. 2005. Accrual reliability, earnings persistence and stock prices. *Journal of Accounting & Economics* 39:437–85.

Rosenberg, B., K. Reid, and R. Lanstein. 1985. Persuasive evidence of market inefficiency. *Journal of Portfolio Management* 11:9–16.

Scherbina, A. 2008. Suppressed negative information and future underperformance. *Review of Finance* 12:533–65.

Schwert, G. W. 2003. Anomalies and market efficiency. In *Handbook of the Economics of Finance*, eds. G. Constantinides, M. Harris, and R. M. Stulz, vol. 1, 939–74. Amsterdam: Elsevier.

Shiller, R. J. 1984. Stock prices and social dynamics. *Brookings Papers on Economic Activity* 2:457–510.

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

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. 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.

Stambaugh, R. F., and Y. Yuan. 2016. Mispricing factors. *Review of Financial Studies*. Advance Access published December 31, 2016, 10.1093/rfs/hhw107.

Subrahmanyam, A. 2010. The cross section of expected stock returns: What have we learned in the past twenty-five years of research? *European Financial Management* 16:27–42.

Thomas, J., and F. X. Zhang. 2011. Tax expense momentum. *Journal of Accounting Research* 49:791–821.

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

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

Tuzel, S. 2010. Corporate real estate holdings and the cross section of stock returns. *Review of Financial Studies* 23:2268–302.

Valta, P. 2016. Strategic default, debt structure, and stock returns. *Journal of Financial and Quantitative Analysis* 51:197–229.

Wilks, S. S. 1932. Moments and distributions of estimates of population parameters from fragmentary samples. *Annals of Mathematical Statistics* 3:163–95.