Alice's Adventures in Factorland: *Three Blunders That Plague Factor Investing*

Rob Arnott, Campbell R. Harvey, Vitali Kalesnik, and Juhani Linnainmaa

Rob Arnott
is a partner and chairman of the board at Research Affiliates, LLC, in Newport Beach, CA.
arnott@rallc.com

Campbell R. Harvey
is a professor of finance at the Fuqua School of Business at Duke University in Durham, NC, and a partner and senior advisor at Research Affiliates, LLC.
cam.harvey@duke.edu

Vitali Kalesnik
is a partner and director of research for Europe at Research Affiliates Global Advisors (Europe), Limited, in London, UK.
kalesnik@rallc.com

Juhani Linnainmaa
is a professor of finance and business economics at the USC Marshall School of Business in Los Angeles, CA, and a partner and advisor at Research Affiliates, LLC.
juhani.linnainmaa@marshall.usc.edu

Factor investing involves allocating portfolio weights to certain known factors. Investors are typically led to develop their factor return expectations with little more than an extrapolation of the factor's past paper-portfolio returns. But past is not prologue. We argue that both the expected returns and the risks are rarely fully understood by investors, whether retail or institutional.

Factor investing is no longer a niche investment product. This investment style has been embraced by some of the world's largest institutional investors and by retail investors who now have access to hundreds of factor products. Given the widespread adoption of factor investing, it is worthwhile to take a step back and reexamine the process of factor development and investment. Our article details three important misperceptions in factor investing: returns falling far short of expectations because of overfitting and/or crowding, drawdowns that far exceed expectations, and failed diversification as correlations unexpectedly soar.

First, a factor can produce disappointing returns for multiple reasons. Recent research by Harvey, Liu, and Zhu (2016) documented 314 factors published in top academic journals (with many more "newly discovered factors" not making it into those top journals). How many of these added value in historical testing on paper? Almost all of them, of course. How many were statistically significant? Almost all of them. How many investors asked these simple questions: Does this factor make sense, or is it likely data mined? How crowded is the factor now? What are the likely trading costs to implement this factor? Is the historical performance of the candidate factor explained by its exposure to another factor (e.g., rising from low valuation to high valuation during the sample period)? Very few of the papers ask any of these questions, yet all of these issues lead to investors developing exaggerated expectations of factor performance.

Second, investors often have a naive view of the tail behavior of factor strategies. Most of these factor returns stray very far from a normal distribution. As such, it is a mistake to use simple risk management tools that ignore the tail behavior. Too many investors believe that creating a portfolio of factors will eliminate the extreme tail behavior. This is a dangerous misperception.

Third, investors need to understand correlations. Many investors mistakenly believe they can diversify away most of the risks in factor investing by creating a portfolio of several factors. In periods of market stress, however, most diversification benefits can vanish as the factors begin moving in unison. An understanding of how factors behave in different environments (e.g., high or low market volatility, high or low inflation, high or low real bond-market yields, economic expansion or recession), and of how correlations change through time, is essential.

In the second section, we detail the impact of data mining on both factor selection and disappointing out-of-sample factor performance. The third section demonstrates the extent of the nonnormalities in factor returns. The fourth section explores how the degree of diversification is time varying. The fifth section of the article quantifies the underperformance risks of factors exacerbated by nonnormality, serial correlation, and autocorrelation of factor return realizations. The sixth section of the article examines the performance of factors over the last 15 years using a bootstrapping technique that takes nonnormalities into account and tries to answer the following question: Is factor investing broken? Some concluding remarks are offered in the final section.

WHY FACTOR PREMIUMS VANISH

Of the thousands of factors tested, some will look good in the backtest purely by luck—that is, as a consequence of data mining and backtest overfitting (a point made by Harvey and Liu (2015), among others). Importantly, many of these lucky factors have little or no economic foundation, which is emphasized by Harvey (2017). Some of these factors may look good as a result of coding mistakes by researchers or because of problems with the data. For example, McLean and Pontiff (2016) failed to replicate the in-sample performance of 12 out of 97 factors in their examination of the many published anomalies.[1]

Even if a factor has a true structural risk premium, real-world returns can disappoint once the factor becomes crowded.[2] The backtest results do not reflect the market impact of investors pouring capital into the strategy. As the factor becomes crowded, too many investors seek to make approximately the same trades, and the mispricing disappears. The anticipated arbitrage leads to disappointing returns. McLean and Pontiff (2016) and Arnott, Beck, and Kalesnik (2016) showed that factor performance degrades after publication. We will quantify some of the return degradation later in this section.

When paper portfolios move to live trading, transaction costs start to play a very important role. Novy-Marx and Velikov (2016) showed that almost no factor, constructed as a long–short portfolio, with turnover exceeding 50%, has any return left after accounting for transaction costs.[3] Hou et al. (2018) reviewed 447 factors and showed that 64% of them fail to deliver statistically significant alpha when the backtest excludes the illiquid microcap names (defined as the bottom 2% of the market by market cap). This finding differs from overfitting but has a similar consequence: If the strategy cannot be implemented in the manner the backtest assumes, the live experience will likely fall far short of the backtest results, even if an uninvestable paper portfolio concurrently matches the backtest results.

Factors are not independent, a finding that is well documented. Yet investors often mistakenly attribute the backtest result to a particular factor when the observed performance is due to exposure (its loading or factor beta) to another factor. Furthermore, the backtest of the candidate factor might look impressive if it begins when the factor has low valuation levels and ends when it has high valuation levels, a point made by both Fama and French (2002) and Arnott and Bernstein (2002). This surely affects the forward-looking premium, as Arnott et al. (2016) argued.

In Exhibit 1, we report the return characteristics of the 15 factors most closely followed by investors over the period July 1963 through June 2018. Beyond the market excess return factor, we divide the factors into two groups. The first group comprises the six factors used in the most popular academic multifactor models (value, size, operating profitability, investment, momentum, and low beta), and the second group includes another

[1] A host of issues complicate replication, such as data revisions, bid–ask issues, methods used to merge databases, and assumptions on delisting returns (see, e.g., Roll [1984] and Shumway and Warther [1999]).

[2] Crowding in financial markets has been a topic of many studies, including those by Lou and Polk [2013], Asness [2015], Bhansali and Harris [2018], Pedersen [2013], Hong et al. [2016], Khandani and Lo [2011], Wermers [1999], and many others.

[3] Prior to Novy-Marx and Velikov's work, Korajczyk and Sadka (2004) raised the trading cost concerns for the momentum factor. Specifically, they showed that it is difficult to benefit from the momentum factor after accounting for transaction costs. Consistent with these estimates, Arnott, Kalesnik, and Wu (2018) and Arnott et al. (2017) found that many active funds fail to capture the momentum premium. This does not mean, however, that momentum is useless in the investment process. For example, Ross et al. (2017) demonstrated that, with careful implementation, an active manager does not have to incur high transaction costs. In addition, Asness and Frazzini (2013) noted that it is possible to improve the efficacy of the value signal by paying attention to when the market value of equity is measured and how this measurement affects the signal's correlation with momentum.

EXHIBIT 1

Average Factor Returns, United States, July 1963–June 2018

Factor	Year of Discovery	Average Return	Average Volatility	*t*-Stat	CAPM Alpha	CAPM Alpha *t*-Stat	CAPM Beta	CAPM Beta *t*-Stat
Market	1964	4.20	10.00	3.11				
Value	1990	4.15	10.00	3.07	5.24	3.99	–0.26	–6.90
Size	1975	2.53	10.00	1.87	1.32	1.02	0.29	7.69
Operating Profitability	2013	3.70	10.00	2.74	4.74	3.60	–0.25	–6.57
Investment	2003	4.32	10.00	3.20	5.80	4.56	–0.35	–9.69
Momentum	1989	5.48	10.00	4.06	6.06	4.50	–0.14	–3.57
Low Beta	1966	0.16	10.00	0.12	3.18	3.36	–0.72	–26.46
Idiosyncratic Volatility	2000	1.62	10.00	1.20	4.34	4.18	–0.65	–21.73
Short-term Reversals	1987	5.34	10.00	3.96	4.13	3.17	0.29	7.70
Illiquidity	1997	3.01	10.00	2.23	3.96	2.99	–0.23	–5.99
Accruals	1991	4.31	10.00	3.20	4.93	3.67	–0.15	–3.81
Cash Flow to Price	1990	4.82	10.00	3.57	6.34	5.01	–0.36	–9.98
Earnings to Price	1971	3.76	10.00	2.78	5.49	4.44	–0.41	–11.65
Long-Term Reversals	1982	3.43	10.00	2.55	3.50	2.57	–0.02	–0.39
Net Share Issues	2003	5.28	10.00	3.91	6.97	5.60	–0.40	–11.26
Average of Factors 1–6		3.39	10.00	2.51	4.39	3.50	–0.24	–7.58
Average of Factors 7–14		3.95	10.00	2.92	4.96	3.95	–0.24	–7.14
Average of Other Factors		3.58	10.00	2.64	4.44	3.48	–0.20	–5.89
Portfolios of Factors: No Rebalancing, No Vol. Scaling, 1–6		3.61	5.17	5.18	4.44	6.82	–0.20	–10.52
Portfolios of Factors: No Rebalancing, No Vol. Scaling, 7–14		3.91	5.20	5.57	4.67	7.04	–0.18	–9.49
Portfolios of Factors: No Rebalancing, No Vol. Scaling, Other		5.14	4.79	7.94	6.14	10.89	–0.24	–14.82
Portfolios of Factors: Monthly Rebalancing, No Vol. Scaling, 1–6		3.39	4.60	5.47	4.39	8.22	–0.24	–15.56
Portfolios of Factors: Monthly Rebalancing, No Vol. Scaling, 7–14		3.95	5.42	5.40	4.96	7.51	–0.24	–12.73
Portfolios of Factors: Monthly Rebalancing, No Vol. Scaling, Other		3.57	3.47	7.62	4.42	11.53	–0.20	–18.41
Portfolios of Factors: Monthly Rebalancing, 10% Vol. Scaled, 1–6		7.38	10.00	5.47	9.55	8.22	–0.52	–15.56
Portfolios of Factors: Monthly Rebalancing, 10% Vol. Scaled, 7–14		7.28	10.00	5.40	9.15	7.51	–0.44	–12.73
Portfolios of Factors: Monthly Rebalancing, 10% Vol. Scaled, Other		10.28	10.00	7.62	12.73	11.53	–0.58	–18.41

Notes: All factors are scaled to 10% volatility. Factor portfolios are equally weighted and also scaled to 10% volatility unless otherwise indicated. Averages for the CAPM beta should be treated with particular caution because these betas are time varying. Market performance is in excess of the 30-day T-bill rate.

Source: Research Affiliates, LLC, using data from CRSP/Compustat.

eight popular factors (idiosyncratic volatility, short-term reversals, illiquidity, accruals, cash flow to price, earnings to price, long-term reversals, and net share issues). We also provide summary information for an additional 33 factors frequently examined in the literature.

We measure the performance of long–short portfolios, which we build following Fama and French (1993). We first sort the universe of stocks based on market capitalization into large and small using the size of the median NYSE firm as the breakpoint. We then select the 30% of stocks with the highest and lowest factor characteristics for the long and the short legs of the portfolios, respectively, using NYSE breakpoints. We compute value-weighted returns for each of the resulting six portfolios. Each factor is long the two high portfolios and short the two low portfolios with equal weights. We further standardize factor portfolios to have annualized volatilities of 10% a year.[4] Of course,

[4] This scaling not only puts factors on equal risk levels, it also facilitates the computation of the Sharpe ratios. For example, if the value factor has a 4.15% average return at 10% volatility, its Sharpe ratio is $0.0415 \times 10 = 0.415$. We apply the same risk scaling method to construct the portfolios of factors. We first scale each factor to have an annualized volatility of 10% a year. We then create an equally weighted portfolio of these factors and, finally, scale the volatility of the resulting portfolio back up to 10%.

investors should never rely on historical performance characteristics alone for the factor selection. An economic rationale—defined before researching a factor and not as an ex post rationale for performance observed in the data—should be equally important.[5] We provide the economic rationale, as well as the arguments on why factors may not work, in the online supplement to this article.

The capital asset pricing model (CAPM) suggested the market return minus a risk-free rate as a risk premium in the early 1960s.[6] Thus, the market factor's performance results, supported by a strong predefined economic rationale for the equity risk premium, are now largely out of sample after over a half century of live experience since first publication. For the 15 most popular factors we study in this article, the median year of discovery is around 1990; this means that about half of the reported performance is in sample and half is out of sample. For the 33 other factors we study, the median year of discovery is 2000; therefore, these factors' reported performance characteristics are largely in sample.

Unsurprisingly, all popular factors earn positive historical average returns, and for most factors these average returns are statistically significant at least at the 5% confidence level. When we combine the groups of factors into portfolios and when we do not rescale the portfolio volatility, we observe average returns roughly on par with the factor averages, volatility quite a bit lower than 10%, and *t*-statistics on the returns and CAPM alphas quite a bit higher. This reduction in volatility and increase in *t*-statistics is the benefit of diversification. Many factors are not highly correlated with each other; consequently, when we scale the portfolios to match the volatility of 10%, we observe a much higher average return.

Two factors on the list—low beta and idiosyncratic volatility—have average returns that are not statistically significant but have negative betas, −0.72 and −0.65, respectively; because of their negative betas, their CAPM alphas are statistically significant, with *t*-statistics of 3.36 and 4.18, respectively. The combination of statistically insignificant returns and statistically significant CAPM alphas means that investors must use leverage and/or derivatives to capture the premium associated with the low beta and idiosyncratic volatility factors.[7]

A negative beta implies potential diversification benefits for investors. In Exhibit 2, we display the average returns for the factors in down (more than one standard deviation below the mean), neutral, and up months (more than one standard deviation above the mean). When the market does great, most factors do not; when the market moves sideways, most factors deliver positive returns; and when the market falls, most factors are at their best. Similarly, on average, the factors have higher premiums in recessions. In Exhibit A2 in the online supplement, we report performance conditional on yield curve slope and inflation regime, showing that the premiums are higher when the slope of the yield curve is flat or inverted and when inflation is high.[8]

Investors should be careful not to be fooled by the long-term factor return averages because the market betas of factors vary widely over time. The value factor, for example, typically correlates negatively with the market. During the global financial crisis, however, the value factor correlated positively and significantly with the market, performing poorly as the markets tumbled and soaring as the stock markets rebounded.[9] Exhibit 3 displays the five-year rolling market betas for the two sets of the most popular factors. An investor who anticipates factor diversification benefits may be disappointed, as value investors assuredly were during the global financial crisis.

These two portfolios of factors correlated negatively with the market, increasingly so beginning in the late 1990s. Toward the beginning of the 2000s, both factor portfolios' betas were lower than −1.0. This threshold is important: It means that an index fund paired with the long–short factor portfolios would have

[5] See Fabozzi and López de Prado (2018) and Arnott, Harvey, and Markowitz (2019) for a far more thorough review of the ways in which an investor can avoid dangerous forms of data mining.

[6] See Treynor (1961, 1962), Sharpe (1964), Lintner (1965a, 1965b), and Mossin (1966).

[7] For example, the betting-against-beta factor constructed by Frazzini and Pedersen (2014) adjusts the leverage of the long leg up and the leverage of the short leg down to make the factor beta neutral.

[8] Note that NBER expansions generally last much longer than recessions, which results in more factors having stronger statistical significance in expansions than in recessions.

[9] Specifically, in the five-year period ending June 2003, the high-minus-low (HML) and market factors had a correlation of −0.71. In the five-year period ending June 2009—the period containing the global financial crisis—the correlation was positive at 0.29.

EXHIBIT 2

Average Monthly Factor Returns Conditional on Market Performance and Economic Cycles, United States, July 1963–June 2018

	Market Return			NBER Business Cycles	
Factor	**Up**	**Neutral**	**Down**	**Expansions**	**Recessions**
Market	4.3%	0.3%	–4.9%	0.5%	–0.5%
Value	–0.8%	0.4%	1.8%	0.3%	0.6%
Size	0.8%	0.4%	–1.6%	0.2%	0.0%
Operating Profitability	–0.5%	0.2%	1.9%	0.3%	0.5%
Investment	–0.8%	0.3%	2.5%	0.2%	1.1%
Momentum	–0.6%	0.7%	0.7%	0.5%	0.3%
Low Beta	–3.0%	0.1%	3.8%	0.0%	0.3%
Idiosyncratic Volatility	–2.4%	0.1%	3.8%	0.1%	0.6%
Short-Term Reversals	1.9%	0.4%	–1.1%	0.4%	0.5%
Illiquidity	–0.9%	0.4%	1.3%	0.3%	0.2%
Accruals	–0.2%	0.4%	0.7%	0.3%	0.5%
Cash Flow to Price	–1.1%	0.4%	2.4%	0.3%	0.8%
Earnings to Price	–1.3%	0.3%	2.7%	0.2%	0.7%
Long-Term Reversals	0.5%	0.2%	0.8%	0.2%	1.0%
Net Share Issues	–1.1%	0.4%	2.8%	0.4%	0.8%
Average of Factors 1–6	–0.8%	0.3%	1.5%	0.3%	0.4%
Average of Factors 7–14	–0.6%	0.3%	1.7%	0.3%	0.6%
Average of Other Factors	–0.5%	0.3%	1.5%	0.2%	0.6%
Portfolio of Factors 1–6	–1.8%	0.7%	3.3%	0.6%	1.0%
Portfolio of Factors 7–14	–1.1%	0.6%	3.1%	0.5%	1.2%
Portfolio of Other Factors	–1.3%	0.8%	4.3%	0.7%	1.7%

Notes: We highlight the values whose corresponding t-statistics exceed 3.0 in absolute value. All factors are scaled to 10% volatility. Factor portfolios are equally weighted and also scaled to 10% volatility.

Source: Research Affiliates, LLC, using data from CRSP/Compustat.

a beta indistinguishable from zero—it would appear to be market neutral—even though the portfolio is exactly 100% long US stocks in its net equity exposure. Investors in these portfolios would be disappointed to discover that what they believe to be a zero-beta market-neutral strategy had a beta indistinguishable from 1.0 during the 2008–2009 market crash, as the negative factor betas evaporated.[10] More recently, the betas of these factor portfolios have increased toward zero. The amount of time variation evident in factor betas implies that investors should probably not have great confidence in predicting the diversification benefits these factors are likely to provide in the future.

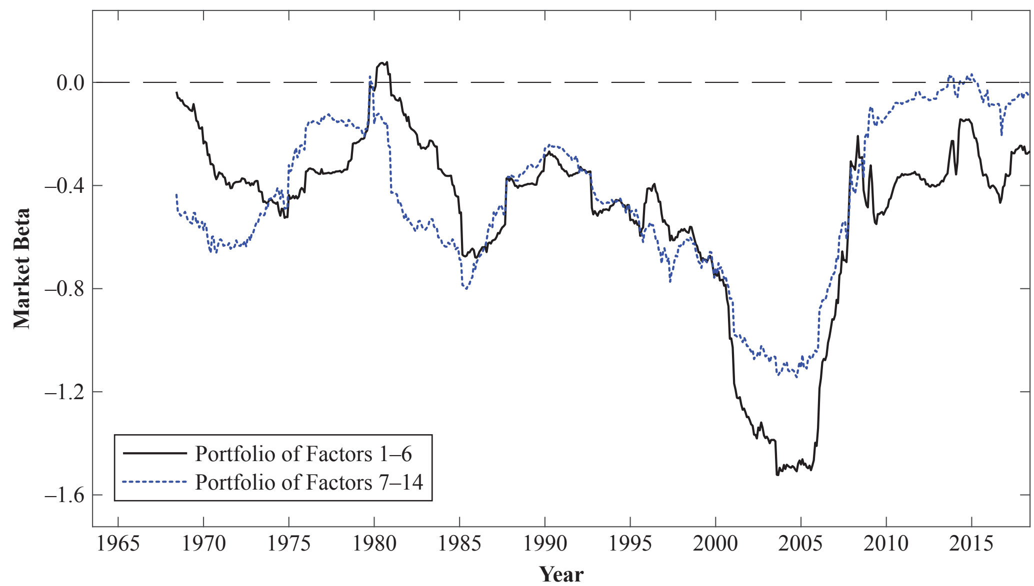
Is this phenomenon of time-varying factor betas an important contributing factor to the poor showing of much of the hedge fund community in 2008–2009? We think so. Is it coincidence that these very different portfolios of factors exhibit their highest and lowest market betas at more or less the same times? We think not.

It is even more important for investors not to be fooled by the largely backtested long-run average returns. Exhibit 4 updates the analysis of Harvey, Liu, and Zhu (2016) and shows the number of factors documented in the top-tier academic journals. Through year-end 2018, over 400 factors have been "discovered," just considering those published in the top academic journals. This number continues to grow each year. The sheer number of factors suggests that many are likely not true factors, having structural drivers of expected returns, but have been found only because numerous

[10] We can see from Exhibit 3 that, as the market beta of the long–short portfolios drifted up toward zero around the global financial crisis and the index fund has market beta exposure of 1.0, the combined portfolio will have a market beta of 1.0.

Exhibit 3

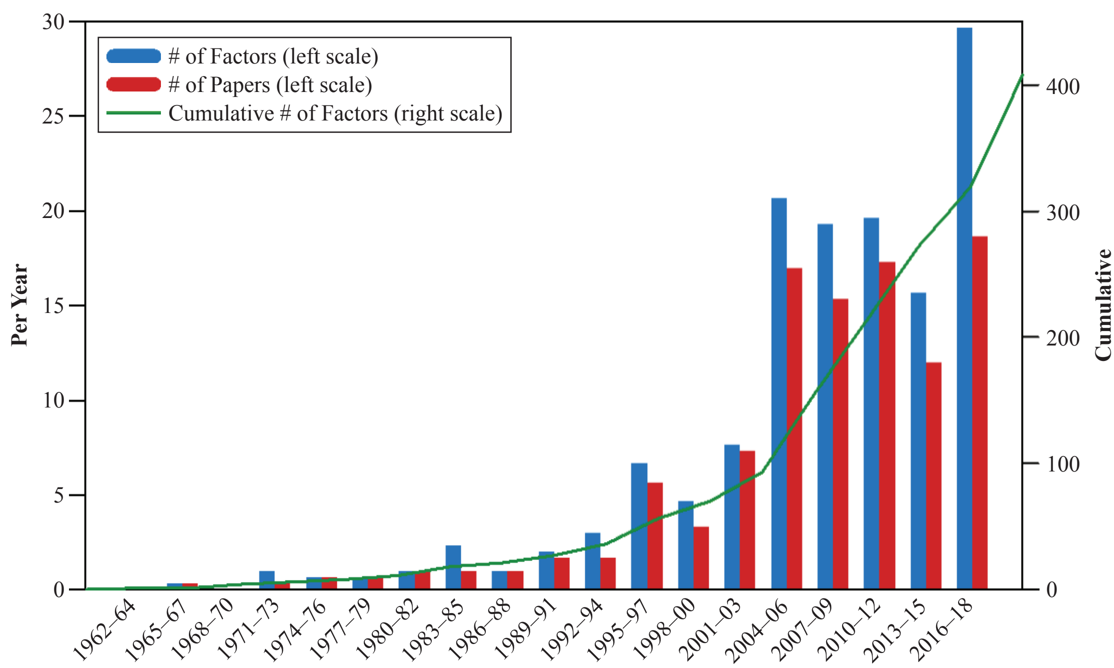
Five-Year Factor Return Market Betas, United States, July 1963–June 2018



Notes: All factors are scaled to 10% volatility. Factor portfolios are equally weighted and also scaled to 10% volatility.
Source: Research Affiliates, LLC, using data from CRSP/Compustat.

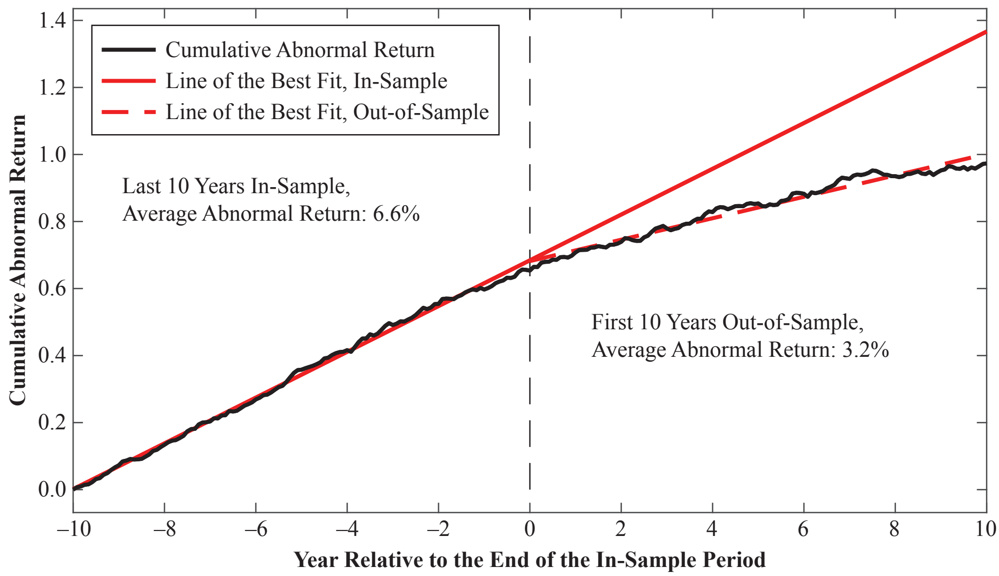
Exhibit 4

Factors and Publication Years, Limited to Top-Tier Academic Journals



Source: Harvey and Liu (2019).

EXHIBIT 5
Cumulative Factor Performance before and after Publication



Notes: We estimate the full-sample CAPM regression for each of the 47 factors using monthly returns. A factor's abnormal return in month t is the estimated alpha plus month-t residual from the CAPM regression.

Source: Research Affiliates, LLC, using data from CRSP/Compustat.

researchers, combing through the large but finite sample of historical data, have found spurious predictors.

As Arnott, Harvey, and Markowitz (2019) discussed, when data are limited, the economic foundations of a factor gain importance. Chordia, Goyal, and Saretto (CGS) (2017) examined 2.1 million equity-based trading strategies that use different combinations of indicators based on data from Compustat. CGS carefully took data mining into account by penalizing each discovery (i.e., by increasing the hurdle for significance). They identified 17 strategies that "survive the statistical and economic thresholds." The statistical significance of the 17 survivors is impressive, nearly matching the high hurdle established by researchers at CERN when combing through quintillions of observations to discover the elusive Higgs boson (ATLAS Collaboration 2012; CMS Collaboration 2012). One of these strategies, for example, has long-term debt issuance minus preferred/preference stock redeemable as its numerator and rental commitments—four years into the future!—as its denominator. This peculiar construction, as with all 17 of the best strategies CGS identified, has no meaningful economic rationale, which—in our view and in the view of the authors of the paper—invalidates the pursuit of these factors.

In Exhibit 5, we measure factor performance before and after the end of the sample period used in the original study that discovered each factor. The figure uses data on all 46 factors that we analyze in our study. We construct this history as follows. We first remove the market component from each factor's return by estimating CAPM regressions using the full sample. We then compute the average residual return (net of market) for each factor, using the average residual as the return for each factor—the growing value of $1 invested in the average of the 46 factors, from 10 years before the end of the in-sample period until 10 years after its end.

Factor performance displays an apparent breakpoint at the end of the in-sample period. The average return in the 10 years after the end of the original sample is less than half the average return over the 10 years prior to this date. At least three drivers likely contribute to this return deterioration: (1) If researchers try many different definitions of predictors, many factors may show high in-sample returns purely by luck or may materially overstate the true factor potential (some may even be entirely spurious); (2) after discovery, many investors try to exploit the anomaly so that the returns weaken and trading costs soar as the trade becomes crowded;

and (3) the backtest returns could simply be a result of correlation with other factor exposures (e.g., a factor might have been cheap at the beginning of the backtest and expensive at the end in the context of value).

Consistent with the results of McLean and Pontiff (2016), the diminishing performance of a factor after its publication is remarkable. An even more striking fact, one that we think has garnered far too little attention, is that factor performance in the most recent 15 years has largely vanished for the most popular factors. Other than the market factor, not a single one has delivered a statistically significant excess return since 2003.[11] We find three—profitability and two of the blended portfolios of factors—have statistically significant CAPM alpha, but only because of a substantial negative beta. Worse, these uninspired results are before trading costs and fees.

In Exhibit 6, we report statistics similar to those we displayed in Exhibit 1, but now we focus only on the last 15 years of the sample. We also illustrate the factors' performance over the full sample (Exhibit 7, Panel A) and the more recent 15 years (Exhibit 7, Panel B). We do not scale portfolio volatilities to 10% in this figure. Instead, to compare the performance of the factors with that of the unlevered market portfolio, we scale the volatilities of the factor portfolios to match the volatility of the market.

For most of the factors we tested, the recent 15-year period was largely out of sample. During this span, four of the six most popular factors had close to zero or negative performance; two of the six factors, size and profitability, had quite attractive positive average returns. A cautionary note applies, however: Profitability is the only one of the six most popular factors whose performance is largely in sample, given its recent publication date of 2013. Indeed, the operating profitability factor's average annualized return is 6.1% from July 2003 through December 2013 and is 2.4% from January 2014 through June 2018. Of course, not all the premiums are gone.[12] The portfolios of the most popular factors (1–6), the next eight (7–14), and the other 33 factors earn average returns of 3.10%, 2.89%, and 5.29%, respectively. These averages are substantially lower than the same portfolios' average returns over the full sample period from July 1963 through June 2018: 7.38%, 7.28%, and 10.28%, respectively. Because these full-span results *include* the latest 15-year span, it is unsurprising that some 75% of the 1963–2002 alphas, on average, have evaporated. Did alpha disappear? Were these strong earlier results a function of cherry-picking and data mining? If the factors delivered genuine alpha in the first place, which has largely disappeared, will the alpha come back?

There are no easy answers here. Our message is that investors need to be skeptical of claims of any new factor's performance and to carefully consider whether the current set of factors is broken (which we address later). Before tackling this difficult question, we explore another investor misperception about the tail behavior of factor returns.

FACTORS ARE FAR FROM NORMAL

Factors are generally prone to big drawdowns. To illustrate this point, we examine each factor's worst monthly return and compare it to the return of the worst month that would be expected with a normal return distribution and the same in-sample volatility. We report in Exhibit 8 the skewness and excess kurtosis of the monthly factor returns.[13] Annual skewness and kurtosis for most of the factors are larger than the monthly results.

The two factors with the most negative skewness are momentum and illiquidity. These results are not surprising: Momentum is known to be prone to crashes,

[11] Caution should even be exercised in interpreting the significance of the market return. Our research uses the last 15 years of data, which happens to begin just after the bear market low of 2002 and ends with the second-longest bull market in US history. Alternative sample selections would lead to different levels of significance.

[12] We may drive a wedge between factor premiums and portfolio return by forming and rebalancing portfolios of factors. Each factor is an equally weighted combination of value-weighted portfolios; HML, for example, is defined as HML = ½ × ($r_{\text{small-value}}$ + $r_{\text{big-value}}$) – ½ × ($r_{\text{small-growth}}$ + $r_{\text{big-growth}}$). A portfolio of factors, in turn, gives the same weight to each factor. We rebalance the portfolio toward equal weights each month; a strategy that rebalances toward equal weights monthly can be viewed as an active strategy. Erb and Harvey (2006, p. 85) discussed this rebalancing issue in the context of commodities trading (also see Booth and Fama 1992). Novy-Marx and Velikov (2018) attributed much of the premium of the betting-against-beta factor to a similar equal-weighting mechanism.

[13] Skewness measures the asymmetry in return distributions. A negatively skewed distribution has more large negative outliers than large positive outliers of a similar magnitude. Excess kurtosis measures the extent to which we observe extreme realizations in both directions. Under normality, skewness and excess kurtosis equal zero.

Exhibit 6

Recent (and largely out-of-sample) Average Annual Factor Returns, United States, July 2003–June 2018

Factor	Year of Discovery	Average Return	Average Volatility	*t*-Stat	CAPM Alpha	CAPM Alpha *t*-Stat	CAPM Beta	CAPM Beta *t*-Stat
Market	1964	6.76	10.00	2.61				
Value	1990	0.15	10.00	0.06	–1.61	–0.63	0.26	3.60
Size	1975	2.66	10.00	1.03	0.33	0.13	0.34	4.90
Operating Profitability	2013	4.97	10.00	1.92	8.17	3.51	–0.47	–7.15
Investment	2003	0.34	10.00	0.13	–0.14	–0.05	0.07	0.94
Momentum	1989	0.77	10.00	0.30	2.92	1.17	–0.32	–4.47
Low Beta	1966	–1.71	10.00	–0.66	3.08	1.65	–0.71	–13.38
Idiosyncratic Volatility	2000	–0.48	10.00	–0.19	4.09	2.10	–0.68	–12.24
Short-term Reversals	1987	0.69	10.00	0.27	–2.12	–0.88	0.42	6.09
Illiquidity	1997	4.23	10.00	1.63	4.47	1.70	–0.04	–0.49
Accruals	1991	2.17	10.00	0.84	0.86	0.33	0.19	2.64
Cash Flow to Price	1990	3.45	10.00	1.33	3.85	1.46	–0.06	–0.80
Earnings to Price	1971	0.69	10.00	0.27	2.75	1.10	–0.31	–4.28
Long-Term Reversals	1982	–2.01	10.00	–0.78	–4.14	–1.65	0.31	4.43
Net Share Issues	2003	0.65	10.00	0.25	3.34	1.38	–0.40	–5.76
Average of Factors 1–6		1.20	10.00	0.463	2.12	0.96	–0.137	–2.59
Average of Factors 7–14		1.17	10.00	0.453	1.64	0.69	–0.069	–1.30
Average of Other Factors		1.64	10.00	0.63	2.86	1.23	–0.18	–2.86
Portfolios of Factors: No Rebalancing, No Vol. Scaling, 1–6		1.12	3.97	1.09	1.98	1.99	–0.13	–4.48
Portfolios of Factors: No Rebalancing, No Vol. Scaling, 7–14		1.04	4.31	0.93	1.49	1.32	–0.07	–2.07
Portfolios of Factors: No Rebalancing, No Vol. Scaling, Other		1.70	3.34	1.96	3.02	4.23	–0.20	–9.65
Portfolios of Factors: Monthly Rebalancing, No Vol. Scaling, 1–6		1.20	3.86	1.20	2.12	2.23	–0.14	–5.07
Portfolios of Factors: Monthly Rebalancing, No Vol. Scaling, 7–14		1.17	4.07	1.11	1.64	1.55	–0.07	–2.29
Portfolios of Factors: Monthly Rebalancing, No Vol. Scaling, Other		1.64	3.11	2.04	2.86	4.27	–0.18	–9.46
Portfolios of Factors: Monthly Rebalancing, 10% Vol. Scaled, 1–6		3.10	10.00	1.20	5.50	2.23	–0.36	–5.07
Portfolios of Factors: Monthly Rebalancing, 10% Vol. Scaled, 7–14		2.89	10.00	1.11	4.03	1.55	–0.17	–2.29
Portfolios of Factors: Monthly Rebalancing, 10% Vol. Scaled, Other		5.29	10.00	2.04	9.20	4.27	–0.58	–9.46

Notes: All factors are scaled to 10% volatility. Factor portfolios are equally weighted and also scaled to 10% volatility unless otherwise indicated. Averages for the CAPM beta should be treated with particular caution because these betas are time varying. Market performance is in excess of the 30-day T-bill.

Source: Research Affiliates, LLC, using data from CRSP/Compustat.

and the prices of illiquid companies tend to plummet when liquidity dries up.[14] Notably, excess kurtosis for all factors is positive, and considerably positive for some factors, suggesting that extreme realizations are not infrequent.

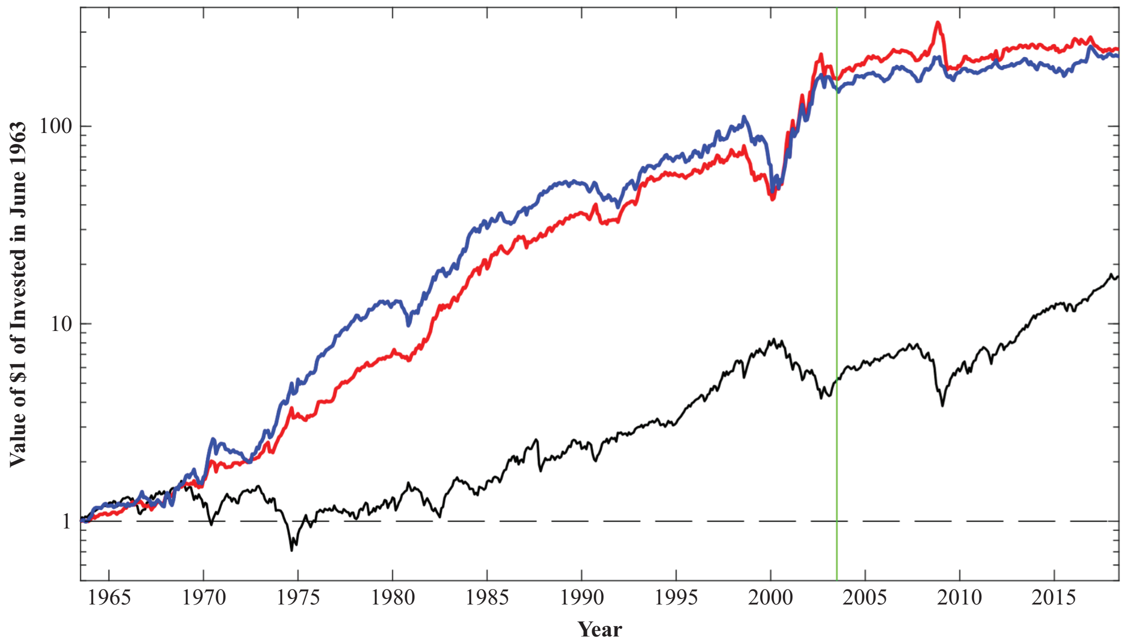
We also report in Exhibit 8 the worst monthly factor returns over the last 55 years together with estimates of how often we would expect to experience losses of these magnitudes if the factor returns were normally distributed with the same volatility. Note that many of these one-month drawdowns were part of a longer drawdown that was, in many cases, considerably worse. (We conduct an analysis of the worst drawdowns in the subsequent sections.) The worst month for 11 of the 14 individual factors shown in the exhibit

[14] Momentum crashes were pointed out by Jegadeesh and Titman (1993) and Daniel and Moskowitz (2016), among others. The nonnormality of momentum factor returns is exacerbated by the time-varying factor volatility. Careful management of risk exposure (which is undoubtedly practiced by many sophisticated asset managers not just for momentum but for most factor exposures) can lead to a significant reduction of outlier risk (a point made by Barroso and Santa-Clara 2015, among others).

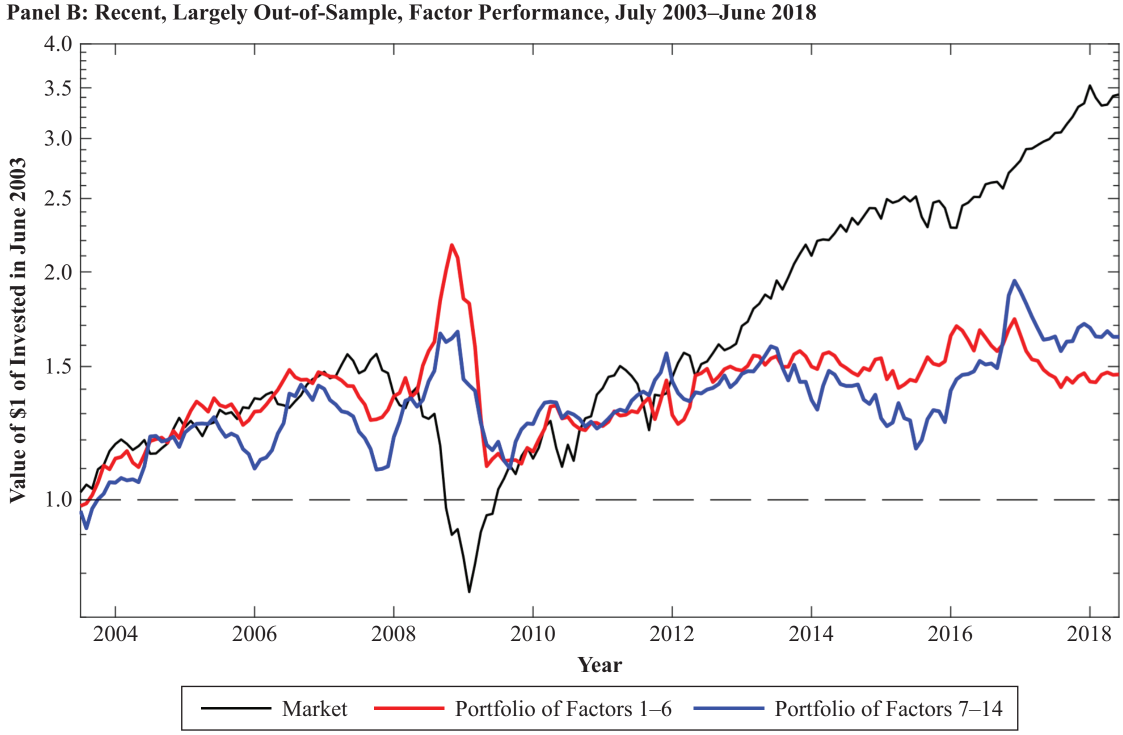
Exhibit 7

Full-Period and Recent Factor Performance, United States, July 1963–June 2018

Panel A: Full-Period Factor Performance, July 1963–June 2018



Panel B: Recent, Largely Out-of-Sample, Factor Performance, July 2003–June 2018



Notes: This exhibit shows the value of $1 invested in the market (financed at the 30-day T-bill rate) for one of the factor portfolios. The factor portfolios are equal weighted, and they are levered to match the volatility of the market portfolio over the July 1963–June 2018 sample period.

Source: Research Affiliates, LLC, using data from CRSP/Compustat.

Exhibit 8

Nonnormality of Monthly Factor Returns, United States, July 1963–June 2018

Factor	Average Annualized Return	Skewness	Excess Kurtosis	Worst Monthly Return	Frequency (in years) of Expected Worst Realized Monthly Return, Assuming Normal Return Distribution
Market	4.20%	–0.54	2.03	–15.3%	1 in 1.6 million
Value	4.15%	0.11	2.05	–11.5%	1 in 2,522
Size	2.53%	0.46	5.54	–16.2%	1 in 8.9 million
Operating Profitability	3.70%	–0.25	12.64	–24.3%	1 in 4.7 quadrillion (10^{15})
Investment	4.32%	0.20	0.79	–9.2%	1 in 109
Momentum	5.48%	–1.41	11.38	–24.3%	1 in 4.1 quadrillion (10^{15})
Low Beta	0.16%	–0.39	3.33	–17.1%	1 in 49.8 million
Idiosyncratic Volatility	1.62%	–0.32	4.39	–16.4%	1 in 12.3 million
Short-Term Reversals	5.34%	0.37	5.82	–13.5%	1 in 58,097
Illiquidity	3.01%	–0.58	5.11	–17.1%	1 in 55.8 million
Accruals	4.31%	–0.11	0.97	–10.0%	1 in 320
Cash Flow to Price	4.82%	–0.35	5.81	–18.9%	1 in 2.5 billion (10^9)
Earnings to Price	3.76%	–0.32	5.94	–19.5%	1 in 10.5 billion (10^9)
Long-Term Reversals	3.43%	0.65	2.62	–9.1%	1 in 106
Net Share Issues	5.28%	–0.40	9.85	–23.4%	1 in 357.5 trillion (10^{12})
Average of Other Factors	3.58%	0.03	4.19	–14.4%	
Portfolio of Factors 1–6	7.38%	0.46	8.12	–16.1%	1 in 6.5 million
Portfolio of Factors 7–14	7.28%	0.12	5.04	–17.6%	1 in 147.6 million
Portfolio of Other Factors	10.28%	0.63	4.93	–13.3%	1 in 38,989

Notes: All factors are scaled to 10% volatility. Factor portfolios are equally weighted and also scaled to 10% volatility.
Source: Research Affiliates, LLC, using data from CRSP/Compustat.

should have occurred less than once in the past 2,000 years; for nine of the factors, it should have occurred less than once during the span that biologically modern humans have roamed the earth; and for three, it should have occurred less than once since the birth of our universe, about 13.8 billion years ago—yet many of us have lived through these extreme realizations over just the last 15 years.[15]

Returns are fat-tailed and asymmetric to the downside. Investors should thus expect and be prepared for extreme factor drawdowns. Take, for example, a momentum strategy. Everything looks great until the sharp drawdown occurs, and then it takes years to recover the losses. In Exhibit 9, we contrast the realized performance of a long–short momentum portfolio with the counterfactual return for which we calibrate returns to be normally distributed.

[15] Kalesnik and Linnainmaa (2018) provided a similar analysis for six factor strategies and an equally weighted six-factor portfolio.

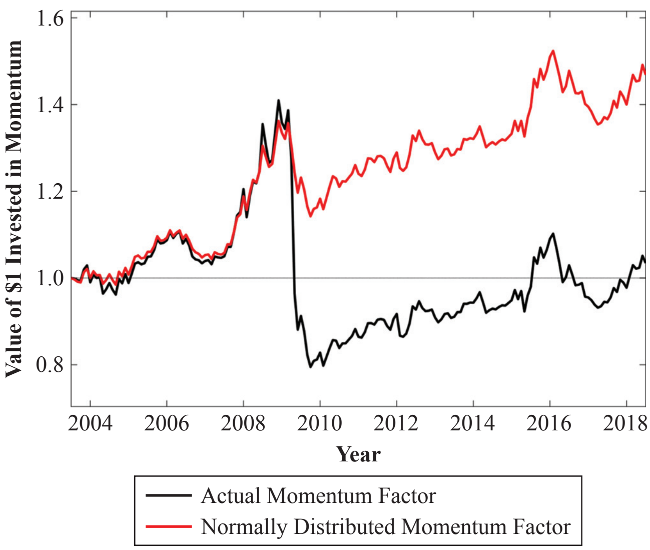
The solid black line in the figure shows the value of $1 invested in the actual momentum strategy, standardized to an annualized volatility of 10%, from July 2003 through June 2018. The momentum strategy crashed in 2009, falling 44% from its peak of $1.41, with a 24% drop in the single month of April 2009. The April 2009 return was momentum's worst month dating back to July 1963; it was an *eight-sigma* event.[16]

The red line in Exhibit 9 represents the counterfactual strategy, which assumes that momentum returns are normally distributed, with the same mean and standard deviation. Using the monthly returns from the momentum factor, as shown on the black line, we

[16] The backtested momentum factor exhibited even more severe crashes in the pre-war period, notably during the Great Depression (Arnott et al. [2016]). In the after-war period, momentum factor performance was quite steady until the early 2000s, when momentum began experiencing crashes again. Investors who were unaware of the pre-war sample would have found the crashes that started in the early 2000s quite surprising.

EXHIBIT 9

Momentum Crash: Actual Momentum Strategy versus Hypothetical Momentum Strategy That Imposes Normality on Realized Returns, July 2003–July 2018



Source: Research Affiliates, LLC, using data from CRSP/Compustat.

then compute the rank of the current month's return, relative to the historical distribution. For example, the worst month ranks 660th in our 660-month history. The midpoint of the bottom 1/660 of the distribution is equivalent to $p = 0.00076$. This turns out to be an 8% loss, if only the distribution were normal!

An investor who assumes that returns are normally distributed could have never foreseen that the factor's returns could be as extremely negative as they sometimes turn out to be.

DIVERSIFICATION FALLS SHORT OF EXPECTATIONS

Whereas any individual factor might have extreme tails, a portfolio of factors might look a lot more like a normal distribution. Indeed, the central limit theorem might encourage us to expect this. This is easy to test and crucially depends on the cross correlations among the factors. The portfolios of factors reported in Exhibit 1 do show material improvements in Sharpe ratios relative to the individual factors as a result of diversification. In Exhibit 6, however, we show that the excess kurtosis for the portfolios of factors is similar to the excess kurtosis of the individual factors. Therefore, the worst monthly drawdowns for the portfolios of factors were not much better than the average of the worst single-factor drawdowns.

To quantify the diversification benefits, consider, for example, the portfolio of factors 1–6. The average return of these factors, as reported in Exhibit 1, is 3.4%. If all six factors were uncorrelated, the annual volatility of this six-factor portfolio would be just 4.1%. Our portfolio of factors is levered up to the same 10% volatility as each of the individual factors, which should increase the return to 8.5%. Exhibit 1, however, shows that the portfolio's return is 7.4%. That is an impressive boost from 3.4% but not as good as 8.5%.[17] Interestingly, the worst single month for the portfolio of factors 1–6 is at −16.0% and is almost identical to the average worst month for the six constituent factors, which is −17.0%. Forming portfolios of factors does not mitigate the risk of large drawdowns to the extent we might expect because the large drawdowns of individual factors often happen at the same time.

Exhibit 10 shows that between mid-2003 and late 2008—before the factor portfolio crash—the returns of the actual and hypothetical portfolios closely track each other. This similarity indicates that, in normal times, returns on portfolios of factors are well approximated by a normal distribution. Note that the factors did perform well during the beginning of the global financial crisis, as global stock markets crashed in the months of September and October 2008. The factor crash happened after this, with the worst occurring after global stock markets had turned sharply higher in early March 2009.

Just before the factor crash began in November 2008, both portfolios were up by approximately 55% from July 2003.[18] As the crash unfolded, the actual

[17] The average return reported in Exhibit 1 for the six-factor portfolio is 7.38% because the portfolio is levered to have a 10% standard deviation of returns. If the factors were uncorrelated, the average return of the factor portfolio would be leveraged back up to 10% risk, leading to an average return of 8.47%: Leverage (10%/4.1%) × Average factor return (3.39%) = 8.47%.

[18] We use the term *factor crash* to refer to the losses accrued by the portfolios of factors starting in November 2008. This event is distinct from the *quant crash* that occurred in August 2007 (see, for example, Khandani and Lo [2011]). The quant crash occurred and the prices recovered within a one-week period, and so this crash does not show up at the monthly resolution.

returns were far more negative than they would have been if the factor return distributions were normal. Normality evaporates in times of market distress. The behavior of portfolios of factors in normal times can therefore be a poor indicator of how these portfolios perform during times of market disruptions. Because many factors—and our portfolios of factors—have large (but demonstrably unstable) negative betas, the largest drawdowns for many factors happened as the stock market rebounded sharply in March and April 2009.

At the bottom of the factor portfolio crash, when the actual portfolio had lost 30%, ceding four years of past gains in a handful of months, the hypothetical portfolio was down by only 15%, giving up less than one year's gains. How much worse are the realized portfolio drawdowns versus what we would expect if we assume a factor is well behaved? To answer this question, we take the following simple approach: We use historical data to calibrate our expectations about factor returns and compare realized factor performance to these expectations. A portfolio of factors may significantly underperform our expectations for three reasons:

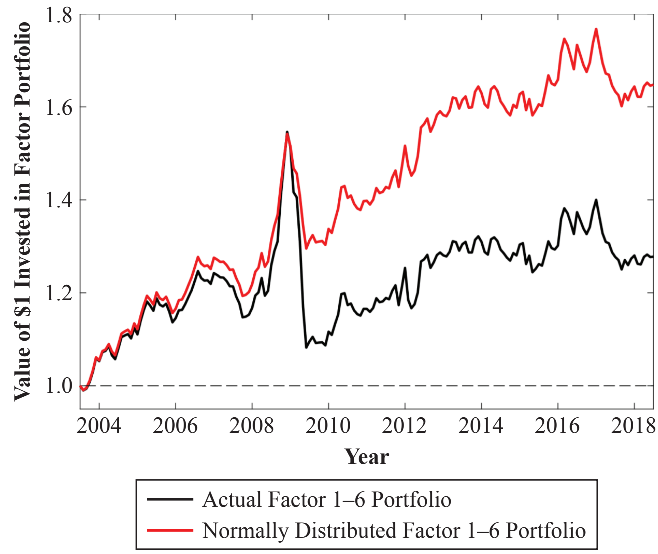
1. Individual factor performance is nonnormal, with a high likelihood of big drawdowns.
2. Cross-factor correlations are time varying, with spikes in correlation around periods of factor underperformance, causing the benefits of diversification to disappear during big drawdowns.
3. Serial correlation of returns exacerbates and prolongs the periods of underperformance.

In Exhibit 11, we examine the largest drawdowns experienced by the portfolios of factors, measured from peak to trough and often spanning multiple months. For example, for factors 1–6, the worst drawdown is 33% (shown in Exhibit 10), which compares to the worst single month of −24.3% for both momentum and profitability, as shown in Exhibit 8. We assess the likelihood of observing a drawdown of this magnitude under different assumptions about the behavior of factor returns and report our findings in Exhibit 11:

- *Simulation under normality* approximates factor returns by normal distributions that we calibrate using the sample means and standard deviations from the full sample. The first row of Exhibit 11 shows that, in the average simulation, the largest drawdown is 21.5%. Just 2.7% of the 10,000 simulations delivered a drawdown worse than the real-world worst-case drawdown of 33.4%.
- The *independent bootstrap* resamples factor returns 10,000 times, resampling each factor's return independently with replacement (i.e., allowing observations to be chosen multiple times) from every other factor. This bootstrap scheme preserves the empirical distributions of factor returns (i.e., it accounts for deviations from normality) but does not account for serial correlations, cross correlations, and cross-serial correlations in factor returns. As a consequence, the empirical probability of 3.7% in this independent bootstrap analysis is close to the probability from the normal simulation—even though we observed before that the individual factor returns are far from normally distributed, when we force factors to be uncorrelated in our independent bootstrap, combining factors into portfolios brings in a lot of the forced diversification.
- The *one-month block bootstrap* simulation resamples factor returns 10,000 times by taking the entire panel of factor returns and drawing months (e.g., June 1997, followed by April 1972) with replacement (Fama and French 2010; Harvey and

EXHIBIT 10

Factor Portfolio Crash: Actual Factor 1–6 Portfolio Returns versus Returns on a Normally Distributed Factor Portfolio, July 2003–July 2018



Source: Research Affiliates, LLC, using data from CRSP/Compustat.

EXHIBIT 11

Estimated Probabilities of Underperformance in Portfolios of Factors, United States, July 1963–June 2018

Factors	Worst Realized Drawdown		Worst Simulated Drawdown and (% of samples that were worse than actual)			
				Bootstrap (resampling returns from realized history)		
			Simulation Assuming Normality	Independent	1-m Blocks, Preserving x- Factor Correlation	12-m Blocks, Preserving x-Factor and Time-Series Correlation
1–6	–33.4%	Return	–21.5%	–21.9%	–25.8%	–30.4%
		[prob]	[2.7%]	[3.7%]	[12.4%]	[29.1%]
7–14	–42.9%	Return	–13.3%	–13.5%	–26.4%	–32.5%
		[prob]	[0.0%]	[0.0%]	[2.2%]	[11.6%]
Other Factors	–22.8%	Return	–6.9%	–6.5%	–20.3%	–23.2%
		[prob]	[0.0%]	[0.0%]	[26.1%]	[44.3%]

Notes: All factors are scaled to 10% volatility. Factor portfolios are equally weighted and also scaled to 10% volatility.
Source: Research Affiliates, LLC, using data from CRSP/Compustat.

Liu 2018). This method accounts for both deviations from normality and cross-factor correlations but does not account for serial or cross-serial correlations. Exhibit 11 shows that the largest drawdown becomes much less of an outlier: The largest drawdown in the average simulation—reported in the 1-m blocks column—is 25.8%; even so, just 12.4% of the simulated largest drawdowns are bigger than the real-world worst-case drawdown of 33.4%.

- The *12-month block bootstrap* follows exactly the same protocol as the one-month block bootstrap, but it draws 12-month blocks, which preserves the nonnormality, cross-factor correlations, and serial and cross-serial correlations. Under these conditions, large drawdowns again become even more probable. The largest drawdown in the average simulation is 30.4%, and in over a quarter of the simulations, the largest drawdown exceeds the actual largest drawdown.

The main message of Exhibit 11 is twofold. First, it shows that portfolios of factors also suffer significant drawdowns because, as we have stated before, diversification across multiple factors cannot fully eliminate the risk of large drawdowns. Moreover, if we falsely viewed returns as being well approximated by independent normal distributions, we would severely underestimate the magnitude of the worst likely drawdowns. Second, Exhibit 11 shows that, if we account for the cross, auto-, and cross-serial correlation structures in factor returns, these large drawdowns are not that surprising.[19] An investor holding a portfolio of factors 7–14, for example, would estimate that the probability of experiencing a drawdown of at least 42.9% is 11.6%.

In the online supplement, we show that the conclusions drawn from Exhibit 11 continue to apply when, instead of studying worst drawdowns, we study the worst one-, three-, and five-year periods for the same portfolios of factors. We show that, just as for the drawdowns reported in Exhibit 11, these prolonged periods of losses would surprise an investor who, in effect, overestimates the extent to which factors are well behaved.

ARE FACTORS BROKEN? OR FAR RISKIER THAN WE THINK?

Exhibit 6 and Panel B of Exhibit 7 show that most factors did not deliver much return in the last 15 years. Are factors broken or just experiencing a period of bad luck? Either way, are they only temporarily impaired and poised to regain the simulated vigor of the past? To attempt to answer these questions, we would first have to know which factors were robust to start with.

In Exhibit 12, we examine the role of bad luck in explaining the recent poor performance of factors. We measure the performance of each factor over the last 15

[19] Arnott et al. (2018) document the momentum effect in factor performance, which implies positive autocorrelation in individual factor performance.

EXHIBIT 12

Estimated Probabilities of Recent Poor Factor Performance, United States, July 1963–June 2018

	Average Return		% of Samples That Were Worse Than Actual		
				Bootstrap (resampling returns from realized history)	
Factor	1963:7–2003:6	2003:7–2018:6	Simulation Assuming Normality	1-m Blocks	12-m Blocks
Market	3.4%	6.8%	98.3%	98.3%	98.8%
Value	5.4%	0.2%	2.3%	2.2%	4.5%
Size	2.5%	2.7%	48.2%	48.0%	48.2%
Operating Profitability	3.4%	5.0%	47.3%	47.1%	44.2%
Investment	5.4%	0.3%	2.8%	2.8%	3.7%
Momentum	7.4%	0.8%	0.3%	0.4%	0.0%
Low Beta	0.8%	–1.7%	5.6%	5.6%	4.6%
Idiosyncratic Volatility	2.3%	–0.5%	8.0%	8.3%	7.4%
Short-Term Reversals	6.9%	0.7%	0.7%	0.7%	0.4%
Illiquidity	2.5%	4.2%	92.6%	92.6%	91.7%
Accruals	5.0%	2.2%	8.3%	8.4%	12.3%
Cash Flow to Price	5.2%	3.4%	15.9%	16.2%	18.0%
Earnings to Price	4.6%	0.7%	6.2%	6.5%	9.5%
Long-Term Reversals	5.4%	–2.0%	0.4%	0.3%	1.1%
Net Share Issues	6.6%	0.7%	1.2%	1.5%	1.1%
Joint Probability			0.0%	1.7%	2.6%

Note: All factors are scaled to 10% volatility in both the 1963:7–2003:6 and the 2003:7–2018:6 subsamples.
Source: Research Affiliates, LLC, using data from CRSP/Compustat.

years ending June 2018 and use simulations to assess the likelihood of observing this level of performance, given how the factors performed before these recent years. We compare factor performance from July 2003 through June 2018 to performance from July 1963 through June 2003. Similar to Exhibit 12, we then simulate returns under different assumptions to assess which features of factor returns, if any, can account for their poor recent performance.

Consider the investment factor as an example. Exhibit 12 shows that the investment factor earned an annualized average return of 5.4% from July 1963 through June 2003. After June 2003, however, the average annual return is a scant 0.3% (before trading costs). Value performed similarly poorly. In the first simulation, reported in the column Simulation Assuming Normality, we generate 10,000 15-year samples by drawing returns from a normal distribution that we calibrate using the investment factor's historical returns from July 1963 through June 2003. In just 2.3% of these simulated paths, the investment factor's average return was 0.3% or worse. This proportion is what we report in Exhibit 12.

We scale each factor to 10% volatility separately in the before- and after-June 2003 samples in the first two columns of Exhibit 12. The simulations themselves, however, use unscaled returns. This difference is responsible for the incongruity between the averages and probabilities reported for (e.g., the size factor). The average returns give the appearance that this factor earned a higher mean after June 2003; in fact, its *unscaled* mean was lower.

The other simulations in Exhibit 12, the same as those described in Exhibit 11, resample historical returns with replacement using different block sizes to account for deviations from normality and for both cross and serial correlation in factor returns. A scheme that resamples individual months—the 1-m Blocks column—assumes no serial correlation in factor returns. In this case, the estimated probability for the investment factor's post-2003 performance is 2.8%, which is the same as the estimate obtained by drawing returns from the normal distribution. This result indicates that deviations from normality, in the case of the investment

EXHIBIT 13

Estimated Probabilities of Recent Poor Performance in Portfolios of Factors Based on Factor Returns (United States, July 1963–June 2003), Actual and Simulated Returns (July 2003–June 2018)

			Worst Simulated Drawdown and (% of samples that were worse than actual)			
				Bootstrap (resampling returns from realized history)		
Factors	Realized Average Return Per Year		Simulation Assuming Normality	Independent	1-m Blocks, Preserving x-Factor Correlation	12-m Blocks, Preserving x-Factor and Time-Series Correlation
1–6	1.9%	Return	9.3%	9.4%	9.4%	9.8%
		[prob]	[0.0%]	[0.1%]	[0.3%]	[0.3%]
7–14	1.9%	Return	9.2%	9.3%	9.3%	9.5%
		[prob]	[0.0%]	[0.0%]	[0.5%]	[1.1%]
Other Factors	4.2%	Return	12.5%	12.6%	12.5%	12.9%
		[prob]	[0.0%]	[0.0%]	[0.1%]	[0.1%]

Notes: All factors are scaled to 10% volatility. Factor portfolios are equally weighted and also scaled to 10% volatility.
Source: Research Affiliates, LLC, using data from CRSP/Compustat.

factor, should not meaningfully change our expectations of the likelihood of observing this previous level of performance in the post-June 2003 sample. The estimated likelihood of the investment factor's poorer performance increases as we resample returns in 12-month blocks, but only modestly so, with the probability increasing from 2.8% to 3.7%. Exhibit 12 therefore suggests that the investment factor's performance was unexpectedly poor over the last 15 years, given the properties of its previous returns.

Our inferences about the likelihood of poor performance vary across factors. Some factors, such as size, operating profitability, and illiquidity, performed well during the last 15 years; of these three factors, the operating profitability factor was discovered only in 2013, so much of its last 15-year performance is not truly out of sample. These factors' recent performance is in line with their earlier performance, and their probabilities are around or above 50%. Most other factors did not come anywhere near their previous performance. Importantly, the entire US equity market performed exceptionally well over the past 15 years.[20]

In Exhibit 13, we use the same framework to revisit the issue of recent poor performance for portfolios of factors. The first row, for example, shows that the annualized average return over the 15-year period from July 2003 through June 2018 was 1.9% for both sets of factors, 1–6 and 7–14.

Exhibit 13 excludes the 15-year period ending June 2018 from the analysis in creating the 10,000 simulated histories of factor returns. If an investor uses these data to calibrate expectations about performance over the next 15 years, how probable would the actual annual return of 1.9% seem? Because we exclude the recent 15-year period of low factor returns, the average returns for the period from mid-1963 to mid-2003 are higher than we report in Exhibit 1. Our average simulation delivers 9.2%–9.8% for both the portfolio of factors 1–6 and the portfolio of factors 7–14. Even when block bootstrapping historical factor returns in 12-month blocks, the probability associated with the actual average return of 1.9% is just 0.3% for factors 1–6 and 1.1% for factors 7–14. Relative to the powerful returns from before mid-2003, the recent performance of these portfolios of factors over the past 15 years is unexpectedly low. The results for the portfolio of the other 33 factors in our study are much the same. Actual returns since mid-2003 are roughly one-third of the pre-2003 returns; random draws from the pre-2003 span deliver a result this weak in just 0.1% of the samples.

CONCLUSION

It is no secret that factor returns have recently fallen far short of investor expectations. Is this a case of

[20] For example, the US equity market's post-2003 Sharpe ratio is 0.68, which is twice as high as its pre-2003 Sharpe ratio of 0.34. Because we normalize the volatility to 10% in both subsamples, Sharpe ratios easily follow from the reported performance.

the factors being broken, or have they just been unlucky over the last 15 years? The answer is probably a combination of both. Recent factor performance has been uncharacteristically bad given pre-2003 performance. Factors' pre-2003 returns were likely inflated by data mining and selection bias, and their post-2003 returns were likely depressed by crowding as they gained widespread adoption.

The role of luck, however, should not be ignored. Just as the factors underperformed the lofty expectations investors might have had back in 2003, the US equity market outperformed these very same expectations to almost the same degree, earning twice its historical Sharpe ratio in the 15 years after mid-2003. The same analysis to which we subjected the factors places the probability of observing this level of performance through luck at less than 2%. Furthermore, the most recent 15-year sample includes the largest drawdown experienced by the portfolio of factors in 2009.

This 15-year period was therefore unusual, not only for the factors but also for the market. What does this mean? If we can accept the possibility that we, as investors, might have been exceptionally lucky with the trajectory of the market, we should also be able to entertain the possibility that the factors might have been exceptionally unlucky. Advocates for factor investing do a disservice to themselves and those they advise by dismissing the recent disappointing performance of factors, but we would also suggest caution before giving up on factor investing based on the same poor performance.

It is important to note that we are not dismissing factor investing. We believe that the factor literature is rich with insights, many of which can be used to deliver superior returns. We also believe that shaping our forward expectations by extrapolating simulated past results (or even live results) is very dangerous. Given that factor investing has not lived up to the expectations of its advocates, do we discard it, or do we rein in our expectations and pay close attention to implementation shortfall as we carefully embrace some use of factor investing? We favor the latter.

REFERENCES

Arnott, R., N. Beck, and V. Kalesnik. "Timing 'Smart Beta' Strategies? Of Course! Buy Low, Sell High!" Research Affiliates Publications, September 2016.

Arnott, R., N. Beck, V. Kalesnik, and J. West. "How Can 'Smart Beta' Go Horribly Wrong?" Research Affiliates Publications, February 2016.

Arnott, R. D., and P. L. Bernstein. 2002. "What Risk Premium Is 'Normal'?" *Financial Analysts Journal* 58, no. 2 (March/April): 64–85.

Arnott, R. D., M. Clements, V. Kalesnik, and J. T. Linnainmaa. "Factor Momentum." Working paper, SSRN, January 31, 2018.

Arnott, R. D., C. R. Harvey, and H. Markowitz. 2019. "A Backtesting Protocol in the Era of Machine Learning." *Journal of Financial Data Science* 1, no. 1 (Winter): 64–74.

Arnott, R., V. Kalesnik, E. Kose, and L. Wu. "Can Momentum Investing Be Saved?" Research Affiliates Publications, October 2017.

Arnott, R., V. Kalesnik, and L. Wu. 2018. "The Folly of Hiring Winners and Firing Losers." *The Journal of Portfolio Management* 45 (1): 71–84.

Asness, C. "How Can a Strategy Everyone Knows about Still Work?" *Cliff's Perspective*, 2015, aqr.com.

Asness, C., and A. Frazzini. 2013. "The Devil in HML's Details." *The Journal of Portfolio Management* 39, no. 4 (Summer): 49–68.

ATLAS Collaboration. 2012. "Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC." *Physics Letters B* 1: 1–29.

Barroso, P., and P. Santa-Clara. 2015. "Momentum Has Its Moments." *Journal of Financial Economics* 116, no. 1 (November): 111–120.

Bhansali, V., and L. Harris. 2018. "Everybody's Doing It: Short Volatility Strategies and Shadow Financial Insurers." *Financial Analysts Journal* 74, no. 2 (Second Quarter): 12–23.

Booth, D. G., and E. F. Fama. 1992. "Diversification Returns and Asset Contributions." *Financial Analysts Journal* 48, no. 3 (May/June): 26–32.

Chordia, T., A. Goyal, and A. Saretto. "p-Hacking: Evidence from Two Million Trading Strategies." Working paper, SSRN, August 12, 2017.

CMS Collaboration. 2012. "Observation of a New Boson at Mass of 125 GeV with the CMS Experiment at the LHC." *Physics Letters B* 1: 30–61.

Daniel, K., and T. Moskowitz. 2016. "Momentum Crashes." *Journal of Financial Economics* 122, no. 2 (November): 221–247.

Erb, C., and C. Harvey. 2006. "The Strategic and Tactical Value of Commodity Futures." *Financial Analysts Journal* 62, no. 2 (March/April): 69–97.

Fabozzi, F. J., and M. López de Prado. 2018. "Being Honest in Backtest Reporting: A Template for Disclosing Multiple Tests." *The Journal of Portfolio Management* 45, no. 1 (Fall): 141–147.

Fama, E. F., and K. R. French. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics* 33, no. 1 (February): 3–56.

——. 2002. "The Equity Premium." *The Journal of Finance* 57, no. 2. (April): 637–659.

——. 2010. "Luck versus Skill in the Cross-Section of Mutual Fund Returns." *The Journal of Finance* 65, no. 5 (October): 1915–1947.

Frazzini, A., and L. H. Pedersen. 2014. "Betting against Beta." *Journal of Financial Economics* 111, no. 1 (January): 1–25.

Harvey, C. R. 2017. "Presidential Address: The Scientific Outlook in Financial Economics." *The Journal of Finance* 72 (4): 1399–1440.

Harvey, C. R., and Y. Liu. 2015. "Backtesting." *The Journal of Portfolio Management* 42 (1): 13–28.

——. "Lucky Factors." Working paper, SSRN, January 15, 2018.

——. "A Census of the Factor Zoo." Working paper, SSRN, February 2019.

Harvey, C., Y. Liu, and H. Zhu. 2016. "… and the Cross-Section of Expected Returns." *Review of Financial Studies* 29, no. 1 (January): 5–68.

Hong, H. G., F. W. Li, S. X. Ni, J. A. Scheinkman, and P. Yan. "Days to Cover and Stock Returns." SSRN, August 2016.

Hou, K., H. Mo, C. Xue, and L. Zhang. "Which Factors." Working paper, SSRN, July 2018.

Jegadeesh, N., and S. Titman. 1993. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *The Journal of Finance* 48, no. 1 (March): 65–91.

Kalesnik, V., and J. Linnainmaa. "Ignored Risks of Factor Investing." Research Affiliates Publications, October 2018.

Khandani, A. E., and A. W. Lo. 2011. "What Happened to the Quants in August 2007? Evidence from Factors and Transactions Data." *Journal of Financial* Markets 14, no. 1 (February): 1–46.

Korajczyk, R. A., and R. Sadka. 2004. "Are Momentum Profits Robust to Trading Costs?" *The Journal of Finance* 59, no. 3 (June): 1039–1082.

Lintner, J. 1965a. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *Review of Economics and Statistics* 47, no. 1 (February): 13–37.

——. 1965b. "Security Prices, Risk, and Maximal Gains from Diversification." *The Journal of Finance* 20, no. 4 (December): 587–615.

Lou, D., and C. Polk. "Comomentum: Inferring Arbitrage Activity from Return Correlations." Paul Woolley Centre working paper No. 36, Financial Markets Group discussion paper no. 721, 2013.

McLean, R. D., and J. Pontiff. 2016. "Does Academic Research Destroy Stock Return Predictability?" *The Journal of Finance* 71, no. 1 (February): 5–32.

Mossin, J. 1966. "Equilibrium in a Capital Asset Market." *Econometrica* 34, no. 4 (October): 768–783.

Novy-Marx, R., and M. Velikov. 2016. "A Taxonomy of Anomalies and Their Trading Costs." *Review of Financial* Studies 29, no. 1 (January): 104–147.

——. "Betting against Betting against Beta." Working paper, SSRN, November 2018.

Pedersen, L. H. 2013. "When Everyone Runs for the Exit." NBER working paper no. 15297, December 2013.

Roll, R. 1984. "A Simple Implicit Measure of the Effective Bid–Ask Spread in an Efficient Market." *The Journal of Finance* 39, no. 4 (September): 1127–1139.

Ross, A., T. J. Moskowitz, R. Israel, and L. Serban. "Implementing Momentum: What Have We Learned?" SSRN, December 1, 2017.

Sharpe, W. F. 1964. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *The Journal of Finance* 19, no. 3 (September): 425–442.

Shumway, T., and V. Warther. 1999. "The Delisting Bias in CRSP's NASDAQ Data and Its Implications for the Size Effect." *The Journal of Finance* 54, no. 6 (December): 2361–2379.

Treynor, J. "Market Value, Time, and Risk." Working paper, SSRN, August 8, 1961.

——. "Toward a Theory of Market Value of Risky Assets." SSRN, Fall 1962.

Wermers, R. 1999. "Mutual Fund Herding and the Impact on Stock Prices." *The Journal of Finance* 54, no. 2 (April): 581–622.

To order reprints of this article, please contact David Rowe at d.rowe@pageantmedia.com or 646-891-2157.

Disclaimer

Since the article's original online publication date of April 1, 2019, text on pages 22, 25, 29, 32, and 35 has been edited as of April 12, 2019.