Fact, Fiction, and the Size Effect

RON ALQUIST, RONEN ISRAEL, AND TOBIAS MOSKOWITZ

This draft: May 2018

Not for quotation without permission

Abstract

In the earliest days of empirical work in academic finance, the size effect was the first market anomaly to challenge the standard asset pricing model and prompt debates about market efficiency. The notion that small stocks have higher average returns than large stocks, even after risk-adjustment, was a pathbreaking discovery, one that for decades has been taken as an unwavering fact of financial markets. In practice, the discovery of the size effect fueled a crowd of small cap indices and active funds to a point where the investment landscape is now segmented into large and small stock universes. Despite its long and illustrious history in academia and its commonplace acceptance in practice, there is still confusion and debate about the size effect. We examine many claims about the size effect and aim to clarify some of the misunderstanding surrounding it by performing simple tests using publicly available data.

RON ALQUIST is a Vice President at AQR Capital Management in Greenwich, CT.
Ron.Alquist@aqr.com

RONEN ISRAEL is a Principal at AQR Capital Management in Greenwich, CT.
Ronen.Israel@aqr.com

TOBIAS MOSKOWITZ is the Dean Takahashi Professor of Finance and Economics at Yale University, Research Associate at the NBER, and a Principal at AQR.
Tobias.Moskowitz@yale.edu; tobias.moskowitz@aqr.com

Acknowledgments: We thank Cliff Asness, Jeremy Getson, Antti Ilmanen, Sarah Jiang, John Liew, Lasse Pedersen, Scott Richardson, Rodney Sullivan, and Ryan Wei for useful comments and suggestions.

Introduction

After confronting the myths surrounding momentum investing[1] and value investing[2] we realized two things: 1) we had passed over the first anomaly discovered in academic finance and the one that had been around the longest – size, and 2) that despite its longevity and the attention it has received, there is still much confusion and debate surrounding the size anomaly.

The size effect is the phenomenon that "small" stocks (i.e., those with lower market capitalizations) on average outperform "large" stocks (i.e., those with higher market caps) over time, on average. The size premium is the return achieved by buying (being long in an absolute sense or overweight relative to a benchmark) small stocks and selling (shorting or underweighting) large ones. The size effect was first documented by several academic papers in the early 1980s,[3] and it quickly became the first real challenge to the field's preeminent asset pricing framework, the Capital Asset Pricing Model or CAPM.

Broadly speaking, researchers responded to its discovery in two ways. On the one hand, proponents of market efficiency argued that this evidence simply indicated that the CAPM was misspecified and that size was related to a second source of priced risk beyond the market. According to this view, a stock's market beta and its size were now required in order to understand its expected returns. As long as size was correlated with a fundamental source of risk, rational investors needed to be compensated for holding assets more exposed to that risk. Other scholars interpreted the evidence of a size premium as a fundamental conceptual challenge to market efficiency, where small stocks relative to large stocks were simply mispriced, having nothing to do with compensation for risk. For example, the evidence available at the time showed that while small stocks had higher market betas, the difference in risk was not large enough to account for the difference in average returns. The size premium, therefore, represented the first true "market anomaly."

Yet, despite size's legacy and its subsequent prominence in the field, there remains much debate about the size effect, including its reliability. The very existence of a size premium, for example, turns out to be a less well-established empirical fact than its younger cousins value and momentum (and defensive and quality premia as well) – something we will investigate in depth in this article.

The paper is organized around a number of facts and fictions about the size effect that warrant clarification. The facts we present include: that the size effect diminished shortly after its discovery and publication; that it is dominated by a January seasonal effect; that it is not applicable or does not work for other asset classes outside of individual equities; that it can be made much stronger when looked at in conjunction with other factors (namely, quality or

[1] "Fact, Fiction, and Momentum Investing," Journal of Portfolio Management [2014], 40th Anniversary edition.
[2] "Fact, Fiction, and Value Investing," Journal of Portfolio Management [2015].
[3] See Banz [1981], Keim [1981], Roll [1983].

defensive factors); that the size premium mostly comes from microcap stocks and is difficult to implement in practice; and, finally, that the size effect continues to receive a disproportionate amount of attention relative to other factors with similar or stronger evidence behind them. The fictions we attempt to clarify include: that the size effect is one of the strongest anomalies; that other factors performing better among small stocks is evidence of a size effect; that the size effect is robust to how you measure it; that it works in other markets and settings; and that it seems to be more than just an illiquidity premium.

Finally, we will address theories that propose an economic story, unrelated to liquidity, where small stocks should deserve a marginal premium over their other risk characteristics, and that a size premium is consistent with a risk-based efficient markets view of the world. While a size premium can certainly occur in a world of efficient or inefficient markets, we find that economic stories, other than as a proxy for illiquidity, for why the size of a firm should matter for pricing, to be puzzling.

As done in our prior papers, we address the facts and fictions of the size effect using published and peer-reviewed academic papers and conduct tests using the most well-known and straightforward publicly available data.[4]

Finally, the topics we address include both positive and negative attributes of size-based investing. Our intention is not meant to completely denigrate a strategy that many believe is a cornerstone of good investing. Rather, our goal is to see the evidence and theoretical arguments behind the size effect more clearly.

Based on the facts we uncover, size does not appear to be on equal footing with other prominent factors, such as value, momentum, and defensive/quality investing. The returns to size are far less stable, less persistent, and less robust than these other factors. Although we do not completely deny the existence of a size effect or advocate actively betting against or shorting it, we also do not believe size is a key factor for constructing portfolios. We believe the size effect captures part of a broader effect – an illiquidity premium – that can add value at the margin in conjunction with other factors, but where it is also (by definition) more difficult and expensive to trade. On its own, a size factor is not a particularly strong source of expected returns in practice, despite its prominence in the literature and the attention it has received from the investment world.

[4] Kenneth French's data library (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html) provides returns for market (RMRF), small (SMB), value (HML), momentum (UMD), and profitability (RMW) factors, including returns for the long and short sides separately and for both large and small capitalization securities separately, all of which we use in this article. AQR's data library (https://www.aqr.com/library/data-sets) provides returns for a betting against beta, BAB, factor from Frazzini and Pedersen [2014], which we use in this article.

Fiction: The Size effect is one of the strongest documented anomalies/factors.

Despite its longevity and prominence, the size anomaly is not one of the strongest anomalies commonly studied in the literature. In fact, it is one of the weakest. It is significantly weaker than other well-known anomalies such as value, momentum, profitability, and defensive or low volatility.

Size has never been a very strong effect. Let's start by examining the original study on the size effect. Banz [1981] documented that small stocks outperformed large stocks over his sample period, which spanned January 1936 to December 1975. The table below attempts to replicate his results using the same sample period. It reports the annualized mean, volatility, t-statistic of the mean, Sharpe ratio, annual alpha, t-statistic of the alpha, and the information ratio (alpha divided by residual standard deviation) from a regression of the size factor's returns on the market portfolio (CAPM). All statistics are computed using monthly returns, but the numbers reported are annualized. We use two specifications of long-short portfolios that seek to capture the size premium: the first is Fama and French's small minus big factor, SMB, which is long the smallest half of stocks (based on NYSE breakpoints) and short the largest half, and the second is a portfolio that is long the smallest decile of stocks (based on NYSE breakpoints) and short the largest. Banz's [1981] original study used Fama-MacBeth [1973] regressions to show a size premium, which is probably closer to the decile portfolio returns approach.[5]

As the table shows, over the 1936 to 1975 period, the evidence in favor of a strong size premium is weak. The first four columns of the table report the annual return, volatility (standard deviation), *t*-statistic of the mean, and Sharpe ratio of the two size strategies. SMB has a 1.9% annualized mean return with almost 10% annual volatility, translating into a 0.19 Sharpe ratio. The mean return of SMB over the 1936 to 1975 period, however, is not statistically significant with a *t*-statistic of only 1.21. The 1-10 decile portfolio has a much higher mean return of 7.1%, but with more than twice the volatility at 25.3% per year for a Sharpe ratio of 0.28. Here, the *t*-statistic of 1.78 barely meets the 10% significance threshold, but not the commonly used 5% threshold. In fact, if Banz's paper had been written today and using Harvey, Liu, and Zhu's [2016] and Harvey's [2017] suggested threshold value of 3.0 for the datamining robust t-statistic, the statistical evidence for a size effect would be even weaker. These results indicate that the size effect is not particularly strong, even over the original sample period in which it was discovered.

The next three columns of the table report the alpha of the size strategies versus the market portfolio (CAPM alpha). Since small stocks typically have larger market betas than large stocks, part of the size premium may simply be the equity market risk premium in disguise. The CAPM alphas account for these beta differences. As the table shows, SMB has zero (in fact slightly negative) alpha with respect to the market once the betas are controlled for, and the decile spread portfolio has a positive alpha (2.5%) that is indistinguishable from zero (*t*-statistic of 0.66).

[5] As Fama [1976] shows, Fama and MacBeth [1973] regressions tend to place more weight on the smallest, most volatile stocks. Hence, a decile sort, which emphasizes the most extreme stocks, will match these results better.

These results suggest that the size premium in its original sample is not only weak, but seems to be captured by general market exposure.

Original Size Anomaly Sample Period

	Annual Return	Annual Vol	Raw T-Stat	Sharpe	Annual Alpha vs CAPM	Alpha T-Stat vs CAPM	IR vs CAPM	Start Date	End Date
SMB	1.9%	9.8%	1.21	0.19	-0.3%	-0.22	-0.03	1/31/1936	12/31/1975
Decile 1-10	7.1%	25.3%	1.78	0.28	2.5%	0.66	0.11	1/31/1936	12/31/1975

The poor showing of the size effect in its original sample begs the question as to how it received so much initial attention and was considered a challenge to the CAPM, when it appears that the CAPM captures it well. We will return to this question at the end, but one issue that may have weakened the size effect since the original studies is that errors in our historical databases of stock prices have been discovered and fixed. The most commonly used database for stock returns is the Center for Research in Security Prices (CRSP) at the University of Chicago, who continually fix data errors they encounter going back in time. One such data error that plagued early studies was a delisting bias. Stocks delisted from the exchanges simply had no return information available to them and were therefore dropped from the analysis. Shumway [1997] painstakingly backfills these delisting returns by hand collecting delisting events and recording the delisted prices, which on average were for negative events.[6] Since these negative delisting returns were omitted from the original data sources of the original studies on size, and since delisting events are more likely to occur for smaller firms, this bias made the performance of small stocks look better than it actually was relative to large stocks. Hence, part of the size premium originally discovered by researchers in the late 1970s and early 1980s may have been driven by these data errors that have since been fixed. Thus, a researcher in 1980 might find no return information for a delisted stock in say January 1965, but a researcher in 2018 looking at that same stock in January 1965 would find (on average) a -30% return. Hence, even if one goes back to the original sample periods of the early studies, the returns during those original sample periods contain fewer errors today than observed at the time researchers were investigating them. Thus, replication of the size anomaly appears weaker than in the original studies, even when the exact same sample period is being examined (Asness, Frazzini, Israel, Moskowitz, and Pedersen [2017] discuss this as well).

We can also look at the size effect over the much longer period for which we have data, including going back to 1926, and of course going forward until 2017. Over the full sample period over the last 91 years, the size premium looks to be about the same magnitude, but is statistically a bit stronger due to the larger sample size – SMB has 2.5% annual return with a *t*-statistic of 2.13 and the decile spread portfolio has a 6.1% return with a *t*-statistic of 2.29. However, as the next three columns show, the CAPM still prices these portfolios nicely, as both alphas of SMB and the Decile 1-10 portfolios are statistically indistinguishable from zero.

[6] Shumway [1997] and Shumway and Warther [1999] find that the delisting return is -55% on average for Nasdaq firms and -30% on average for NYSE/Amex firms when delisting is for performance-related reasons.

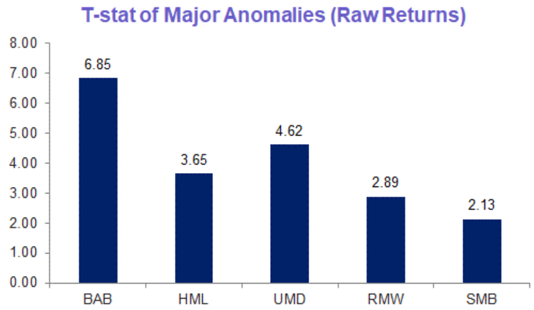
Full Sample Period	Annual Return	Annual Vol	Raw T-Stat	Sharpe	Annual Alpha vs CAPM	Alpha T-Stat vs CAPM	IR vs CAPM	Start Date	End Date
SMB	2.5%	11.1%	2.13	0.22	0.9%	0.86	0.09	7/31/1926	12/31/2017
Decile 1-10	6.1%	25.6%	2.29	0.24	2.2%	0.89	0.09	7/31/1926	12/31/2017

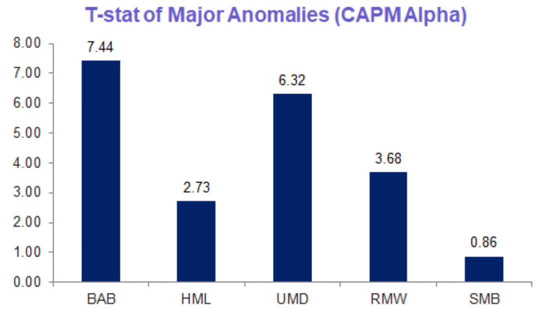
So, the size premium on its own is significant, but adjusting for market beta renders it insignificant. How do these results compare to other well-known factors from the literature? The table below reports the performance of five popular academic factors based off of the five most prominent asset pricing anomalies found in the academic literature, over the longest sample of data available. The factors include the "betting against beta" factor, BAB, of Frazzini and Pedersen [2014] that is long low beta stocks and short high beta stocks, levered to have the same beta and taken from AQR's data library; the "high minus low" value factor, HML, of Fama and French [1993], which is a portfolio long the top 30% of stocks based on high ratios of book-to-market equity (BE/ME) and short the lowest 30% of BE/ME stocks, taken from Ken French's website; the "up minus down" momentum factor, UMD, which is long high momentum stocks (top 30%) and short low ones (bottom 30%), taken from Ken French's website; the "robust minus weak" profitability factor, RMW, which is long high profitable firms based on the top 30% profits-to-assets ratio and short the bottom 30%, following Fama and French [2015] and taken from Ken French's website; and finally the "small minus big" size factor, SMB.

As the table shows, no matter what metric of performance used – mean, Sharpe ratio, *t*-statistic, alpha, or IR – the size factor, SMB, has the *worst* performance among the five factors, and often by a decent margin. For example, SMB's full sample Sharpe ratio is 0.22 per year, while the next lowest Sharpe ratio is that of HML at 0.38, and UMD and BAB have Sharpe ratios almost two to three times larger (0.48 and 0.73). The CAPM alphas are significant for all of the factors except SMB, indicating that the factors other than size add a return premium above and beyond traditional equity market risk. This evidence also shows that the size effect is in fact not a market anomaly, unlike the other factors. Furthermore, as we will show later, the other four factors have a wealth of out of sample evidence showing their efficacy in other time periods, other equity markets, and even other asset classes. The size factor fails to yield significantly positive effects out of sample in all of these settings.

Full Sample	Annual Return	Annual Vol	Raw T-Stat	Sharpe	Annual Alpha vs CAPM	Alpha T-Stat vs CAPM	IR vs CAPM	Start Date	End Date
BAB	8.1%	11.0%	6.85	0.73	8.7%	7.44	0.80	12/31/1930	12/31/2017
HML	4.6%	12.1%	3.65	0.38	3.4%	2.73	0.29	7/31/1926	12/31/2017
UMD	7.9%	16.3%	4.62	0.48	10.2%	6.32	0.67	1/31/1927	12/31/2017
RMW	3.0%	7.7%	2.89	0.39	3.7%	3.68	0.50	7/31/1963	12/31/2017
SMB	2.5%	11.1%	2.13	0.22	0.9%	0.86	0.09	7/31/1926	12/31/2017

The figures below show clearly that on either a raw or risk-adjusted return perspective, the size effect is the weakest of the anomalies.





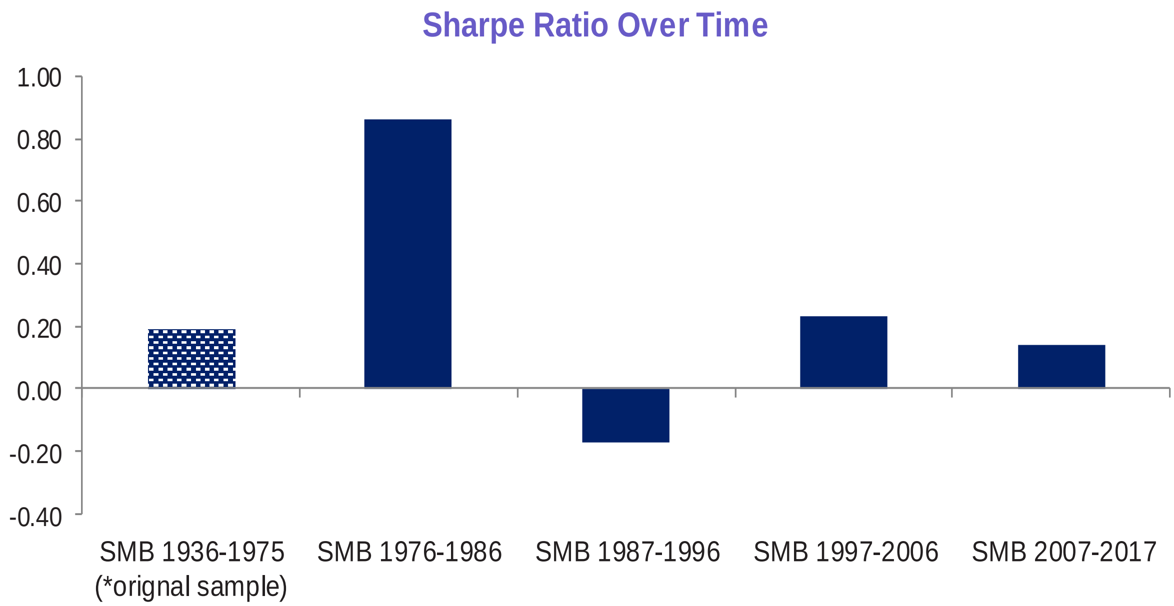
So, while we can debate whether there is in fact a significant size premium at all (and whether there ever was), there is little debate about whether size is one of the strongest anomalies – it is not. It is one of the weakest.

Fact: The size effect has disappeared or weakened since its discovery

As noted above, there is some debate as to whether a size effect ever existed at all. But, even among those that believe there was a healthy size premium, many more believe it has significantly weakened over time since its discovery, to the point that it is no longer there.[7]

The figure below plots the Sharpe ratio of the small minus big factor, SMB, over its original sample when it was discovered (1936 to 1975), as well as decade-by-decade over the four decades following the original size discovery: 1976 to 1986, 1987 to 1996, 1997 to 2006, and 2007 to 2017.

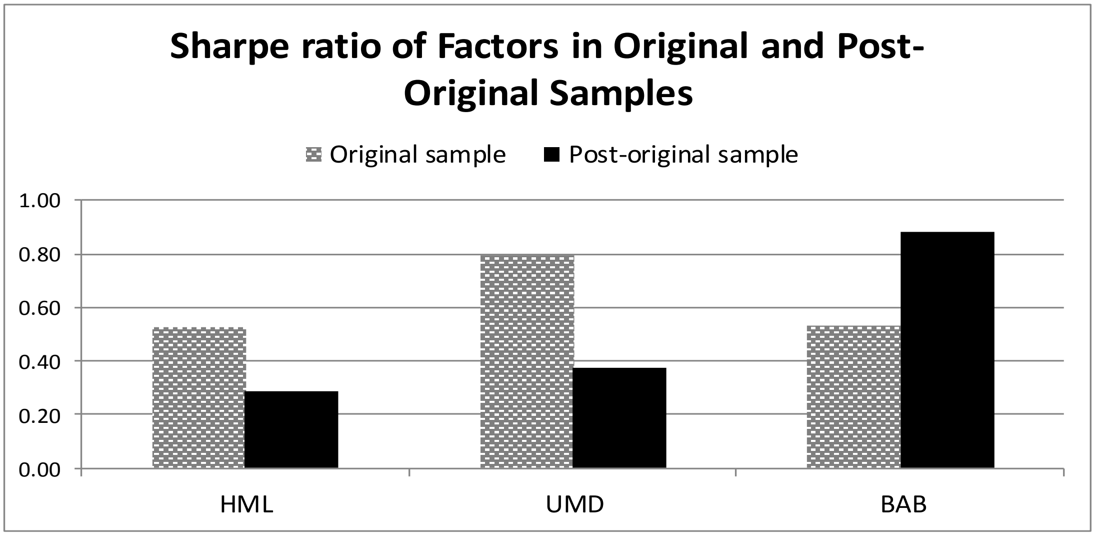
[7] See Dichev [1998], Chan, Karceski, and Lakonishok [2000], Horowitz, Loughran, and Savin [2000], Gompers and Metrick [2001], Van Dijk [2013], Israel and Moskowitz [2013], McLean and Pontiff [2015], and Chordia, Subrahmanyam, and Tong [2015].



One of the most interesting findings is the very strong performance of SMB in the decade immediately following the paper by Banz [1981], when (perhaps not coincidentally) most of the original follow-on studies of the size effect were published. From 1976 to 1986, the size effect experienced returns almost four times larger than its original sample. Following Banz's [1981] seminal study, and perhaps spurred on by the immediate rise in the strength of the size effect following that study, a suite of papers by Reinganum [1981], Keim [1983], Reinganum [1983], Schwert [1983], and Roll [1983] all dissected the size anomaly during a decade when its returns were particularly strong and being noticed by practitioners. However, following the publication of these papers in the 1980's the returns to size fell precipitously and were actually negative over the subsequent decade, turning slightly positive over the next two decades, but essentially remaining flat. Since the slew of publications on the size effect, there has been no significant positive premium associated with small cap strategies.

Scholars have offered various explanations for the disappearance of the size effect. For instance, Schwert [2003] suggests that the small-firm anomaly disappeared shortly after the initial publication of the papers that discovered it because of an explosion of small cap funds and indices that may have priced it away. Gompers and Metrick [2001] argue that institutional investors' continued demand for large stocks in the 1980s and 1990s increased the prices of large companies relative to small companies, which may account for a large part of the size premium's disappearance over this period. Finally, Hou and Van Dijk [2014] argue that small firms experienced a series of negative profitability shocks in the 1980s and 1990s and that these shocks help to explain the disappearance of the size premium during that period.

There is also the specter of data mining having exaggerated the original results and explaining why the out of sample evidence looks poorer. McLean and Pontiff [2015] argue that many anomalies from the finance literature suffer from poorer out of sample performance due to both data mining and arbitrage activity that lowers their returns.[8] We can compare how the other prominent anomalies – value, momentum, and defensive – fare in the out of sample periods since their discovery. We use HML, UMD, and BAB factors to represent each of the other anomalies and define their original sample periods following McLean and Pontiff [2015], who use the seminal papers of Fama and French [1992] for value, Jegadeesh and Titman [1993] for momentum, and Fama and MacBeth [1973] for market beta. The corresponding original sample periods from those studies are 1963 to 1990, 1964 to 1989, and 1926 to 1968, respectively. We therefore report the out of sample performance of HML, UMD, and BAB from 1991 to 2017, 1990 to 2017, and 1969 to 2017, respectively.



As the figure shows, the out of sample performances of the three factors are mixed compared with their performances in sample. The Sharpe ratios of the HML and UMD factors decline in the sample period after their discovery, while the BAB factor's Sharpe ratio improves. Declining out of sample performance is thus not unique to the size factor. More importantly, however, the size premium remains insignificant in the out of sample period, while the other three factors continue to exhibit significant return premia in their respective out of sample periods. So, although the HML and UMD premia fall out of sample, their returns remain significantly positive.[9]

[8] Interestingly, McLean and Pontiff [2015] show a bias in published papers where the last few years of a paper's data sample tends to exhibit returns that are much stronger than the first few years of out of sample data following the original sample. This bias could result from sample-specific data mining, or more indirectly selective updating of data, where the authors only update samples when the added few years improve their results, but do not bother if the results are unchanged or weaker. It could also be the case that papers are written *because* the recent sample is so strong, where strong recent performance may make the effect more hotly debated, salient, and interesting. This may also partly explain the proliferation of size-related papers in the immediate years following the original study.

[9] In another out of sample test, Ilmanen, Israel, Moskowitz, Thapar, and Wang [2018] examine a century of evidence on these factor premia across many asset classes and test their out of sample performance both before and after the

Regardless of the reason, and data mining, arbitrage, and shifting demand for small stocks may all have partly contributed to its demise, the size premium has weakened over time and is absent since its original discovery from a congregation of papers published in the early 1980s.

Fiction: The size effect is robust to how you measure size.

A single measure of anything seems unlikely to be optimal, given estimation error, data mining concerns, and absent any strong theory. Indeed, we showed for both value and momentum,[10] multiple measures of each tend to provide better and more stable performance, providing robustness driven by diversification benefits from different measures that diminish data errors, noise, and the influence of missing data that can otherwise limit samples.

For size, we also put this statement to the test. The predominant (in academia and practice) way to measure size is to use the firm's market capitalization, which is the share price of the equity in the firm multiplied by the number of outstanding shares of the stock. However, the size of the firm could be captured in many ways. How robust is the size effect to different measures of size?

Academia has considered this question. Berk [1995a] for instance, using an argument from Ball [1978], argues that when size is measured by market capitalization, which contains market prices, it can mechanically lead to a negative relation between size and average returns. The idea is simple: returns equal today's price plus dividends, divided by yesterday's price, which will have a statistical negative relationship with market cap (which equals shares outstanding times yesterday's price) by construction if prices move. So, if running the following regression,

$$\underbrace{\frac{P_t + D_t}{\boldsymbol{P_{t-1}}}}_{Return_t} = \alpha + \beta(\underbrace{\boldsymbol{P_{t-1}} \times S_{t-1}}_{market\ cap_{t-1}}) + \delta' X_{t-1} + \varepsilon_t$$

where P_t, D_t, and S_t, are price, dividends, and shares outstanding, respectively, at time t and X_{t-1} is a set of control variables, if the controls do not completely account for all price movements, then mechanically there will be a negative relation between returns and market cap, since the price at time $t-1$, highlighted in **red**, shows up on both sides of the regression.

To address this potential bias, Berk [1995b, 1997] suggests using non-price based measures of size as a better way to test the true relation between size and average returns. He finds, however, that using non-price based size measures (such book equity or number of employees) results in

---

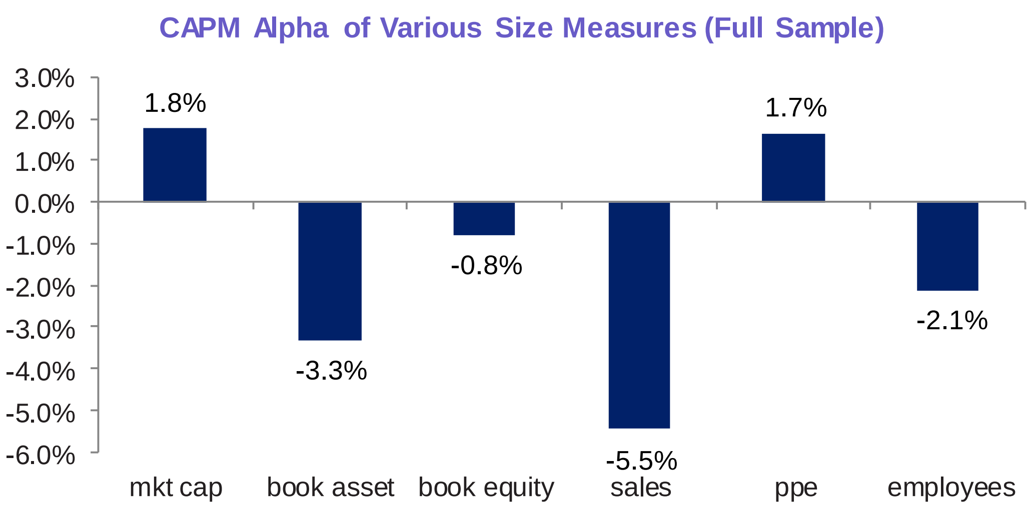
original sample periods in which they were discovered. The before-discovery sample should be immune from arbitrage trading effects since the anomalies were not yet known. Hence, contrasting the two out-of-sample periods provides a distinguishing test of data mining versus arbitrage-driven return degradation. They find that the out of sample evidence in both periods similar, but worse than the original sample period, suggesting that data mining rather than arbitrage may be contributing to the weaker out of sample performance.

[10] Asness, Frazzini, Israel, and Moskowitz [2014, 2015].

no reliable size premium. Hence, the size effect does not appear robust to these other measures of size that do not contain market prices.

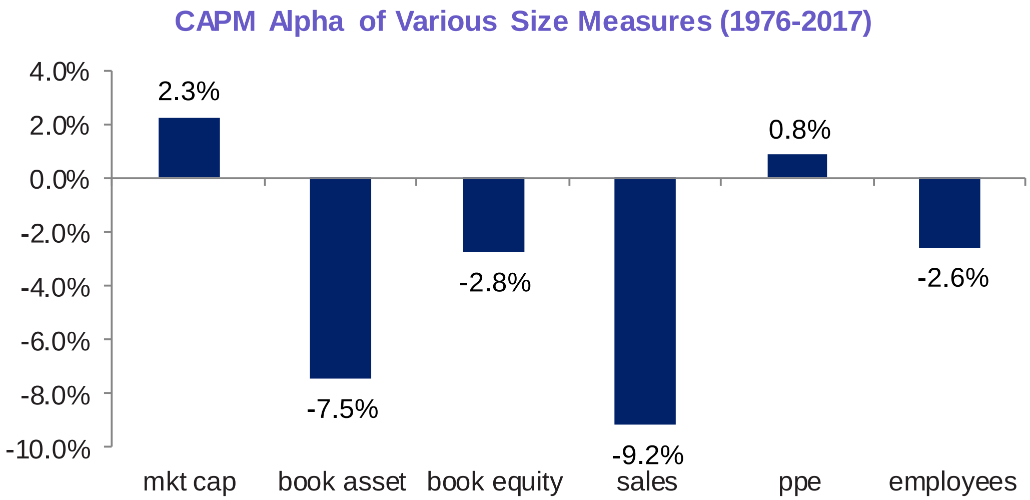
We examine the robustness of different measures of size for predicting returns by using non-price based size measures. Specifically, we use the book value of assets, book value of equity, sales, property, plant, and equipment (PP&E), and the number of employees as alternative non-price based measures of the size of a firm. For each non-price size measure, we form an SMB portfolio in the exact same manner as above (e.g., Fama and French [1993]) and use each non-price size measure to rank and sort stocks.

The figure below shows the alphas with respect to the market (CAPM) of these SMB portfolios based on the different measures of size over the full sample or longest period for which we have available data (January 1951 to December 2017, where accounting numbers are available). The first bar shows the results for market cap as the measure of size and the remaining bars show the results for the non-price size measures.



As the figure shows, the market cap measure of size (which uses prices) delivers the strongest size premium, while the non-price based measures of size are weaker, with four out of the five measures producing a negative result.

The next figure reproduces the graph above for the out-of-sample period from 1976 to 2017 after the original study by Banz [1981]. Here, the performance of the non-price size measures is even worse, and the only substantial return premium exhibited is for the market cap measure of size.



These results are broadly consistent with Berk [1995a, 1995b, and 1997] and suggest a much weaker relation between non-price based size measures and average returns. What little size premium might be present when using market cap to measure size disappears entirely (and switches sign) when using non-price based measures of size. These results are even starker out of sample following the original research. Hence, the size effect seems to vary considerably with different measures of size and does not appear very robust.

This result is counterintuitive since any individual measure has error to it (due to mismeasurement, missing data for some firms, and random errors), so an average of similar measures should help reduce noise and be more robust. Frazzini et al. [2013] and Israel and Moskowitz [2013] show that multiple measures of value produce more stable value portfolios that deliver higher Sharpe ratios, higher information ratios, and more robust returns. The same is true for momentum (Frazzini et al. [2015]), and for quality/defensive (Asness, Frazzini, and Pedersen [2016]). As with any systematic process, unless theory dictates preferring one metric to all others, an average of sensible measures is generally the best and most robust approach. While this is true for all of the other commonly used factors, it does not appear true for size.

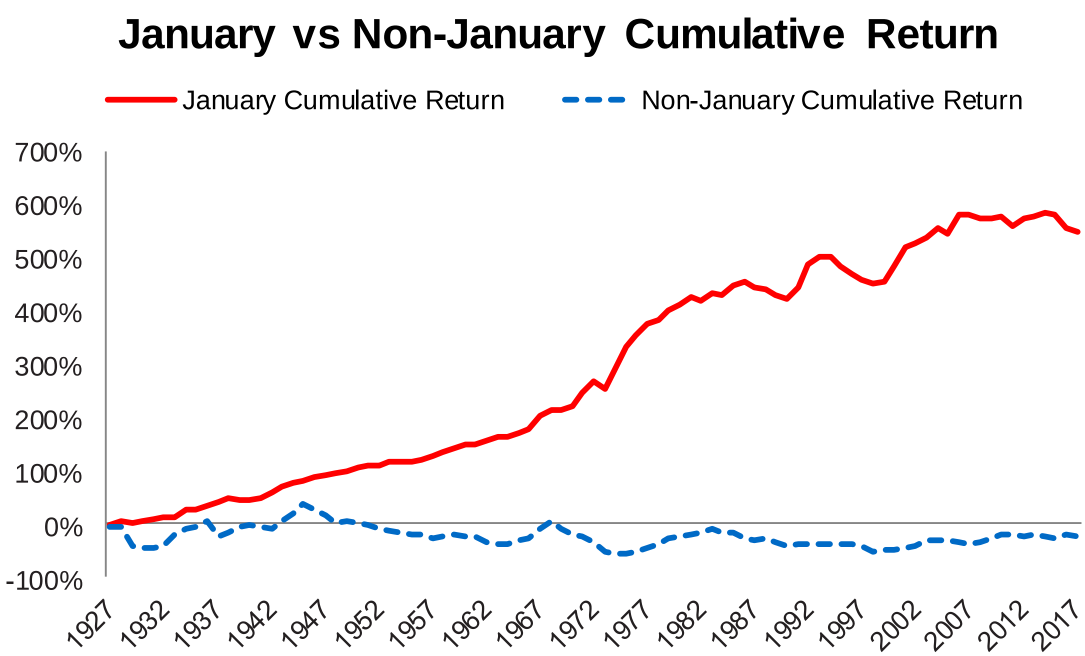
In addition, using multiple measures to reduce errors generally improves the out-of-sample performance of a strategy. As with any specific sample of data, you will always find some measures that work particularly well in sample and some that do not. However, without theory telling you why one measure should outperform another, this is usually due to chance. Using multiple measures can therefore guard against the dangers of data mining – picking one particular measure over others that happened to work well in one particular sample, and that one is often overfitted to that sample.

As only the market cap measure of size seemed to deliver any sort of premium, and all other measures produced a negligible or opposite signed premium, the robustness of the size effect is questionable. Moreover, the significantly worse performance of the market cap based measure of size in the out-of-sample period following the original studies is also troubling from a data mining perspective. Combining these results, the size effect does not appear robust. Unlike other factors (e.g., value, momentum, quality) the size premium is quite sensitive to changes in how you measure it and over what sample you look at it.

Fact: The size effect is dominated by a January effect.

One of the earliest findings of the size effect was that it mostly resided in January (see Keim [1983], Roll [1983], and Reinganum [1983a], as well as recent work by Asness, Frazzini, Israel, Moskowitz, and Pedersen [2017]). This strong seasonal component to the size effect has long been a focal point – for both advocates and critics – of the size factor.

Again, let's start with the full sample evidence from 1926 to 2017. The graph below simply plots the cumulative returns to the size factor, SMB, for the months of January only versus all other months. For the January cumulative returns, we invest in SMB in January of each year then put the returns in cash for the remaining months (February to December). For the non-January cumulative returns, we invest in SMB for all months except January (putting the money in cash for January). A plot of the time series of the two cumulative returns is reported below.



As the graph clearly shows, there is a substantial return to the size factor in January, but absolutely no evidence of any size premium outside of January. The returns to size are completely flat throughout most of the year. Whatever premium the size factor has seems to be generated almost exclusively in January.[11]

A more formal test of the size effect in and outside of January is contained in the table below, where we report the average monthly return, volatility, *t*-statistic, Sharpe ratio, and CAPM alphas to SMB in January and non-January months. SMB delivers an impressive 2.1% return in the month of January alone, with a *t*-statistic of 6.25 that is not captured at all by the CAPM (alpha equals 1.9% with a *t*-statistic of 5.64). These results are dramatically stronger than what we obtained for SMB over all months over the same sample period. The non-January months exhibit literally zero size premium (average return of 0.0% from 1926 to 2017) and an alpha of -0.1%. All of the returns to size are concentrated in January exclusively, with no evidence of any size effect – economically or statistically – outside of January.

SMB Performance in January and Non-January Months (full sample)

	Monthly Return	Stdev	Raw T-Stat	Sharpe	Monthly Alpha vs CAPM	Alpha T-Stat vs CAPM	IR vs CAPM	Start Date	End Date
January	2.1%	3.2%	6.25	0.65	1.9%	5.64	0.61	1/31/1927	1/31/2017
Non January	0.0%	3.1%	0.32	0.04	-0.1%	-0.88	-0.10	7/31/1926	12/31/2017

Because of these results, the size effect and the January effect have been inextricably linked. Since its discovery, many researchers have argued that the January effect has weakened over time, driven possibly by increased arbitrage trading that exploited it, less price impact in the market from turn-of-the year trades due to improved market liquidity, more passive index investing, etc. The weaker January effect may in turn have contributed to the weaker size effect over time.

The table below reports the same statistics on SMB in January and non-January months for the more recent sample from 1976 to 2017, following the original size studies. As the table shows, the January effect is indeed much weaker in the more recent sample, but it still dominates what is left of the size effect in this sample. SMB in January averages only 1.0% per month in this period compared to the 2.1% return it exhibited in January over the longer sample dating back to 1926, with a *t*-statistic of 2.39. The CAPM once again cannot explain these returns. Outside of January, there is no SMB premium in the recent period – CAPM alpha of 0.0% with a *t*-statistic of 0.27.

[11] Moreover, early researchers (Keim [1983] and Roll [1983]) showed that it was in fact the first few trading days of the year that generated nearly all of the January premium and hence all of the size premium as well. This empirical fact has been attributed to year-end tax loss selling, rebalancing, and cash infusion at the beginning of the year from investors as well as window dressing by mutual fund managers at the turn of the year.

SMB Performance in January and Non-January Months (1976-2017)

	Monthly Return	Stdev	Raw T-Stat	Sharpe	Monthly Alpha vs CAPM	Alpha T-Stat vs CAPM	IR vs CAPM	Start Date	End Date
January	1.0%	2.7%	2.39	0.37	1.0%	2.24	0.35	1/31/1976	1/31/2017
Non January	0.2%	3.0%	1.09	0.18	0.0%	0.27	0.04	2/29/1976	12/31/2017

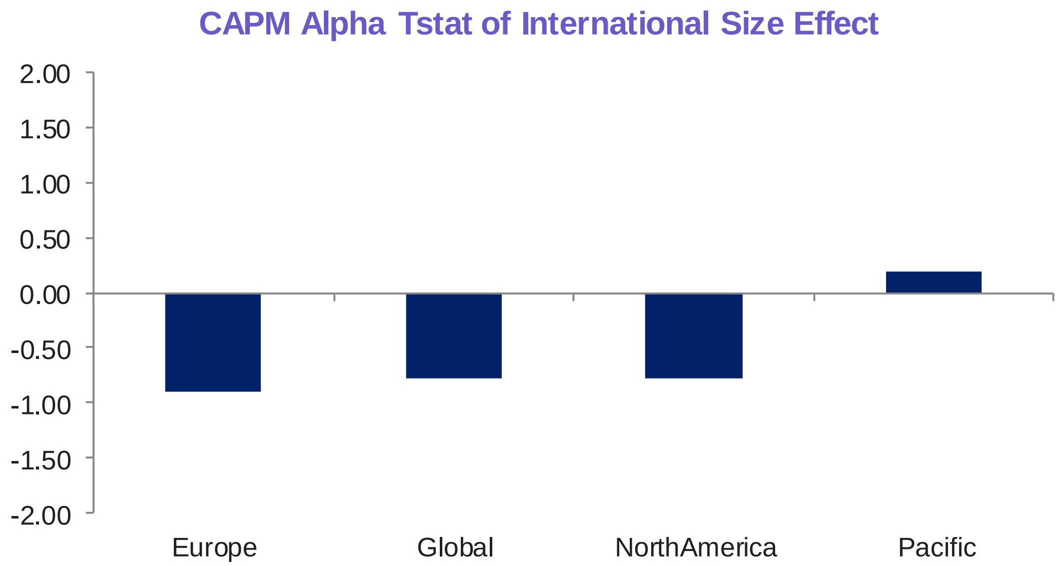
The bottom line is that other than January there is, and never was, a size premium. All of the returns to size seem to come from January alone, and the fact that the January effect has diminished over time has contributed to the demise of the size effect.

Fiction: The size effect works in other equity markets.

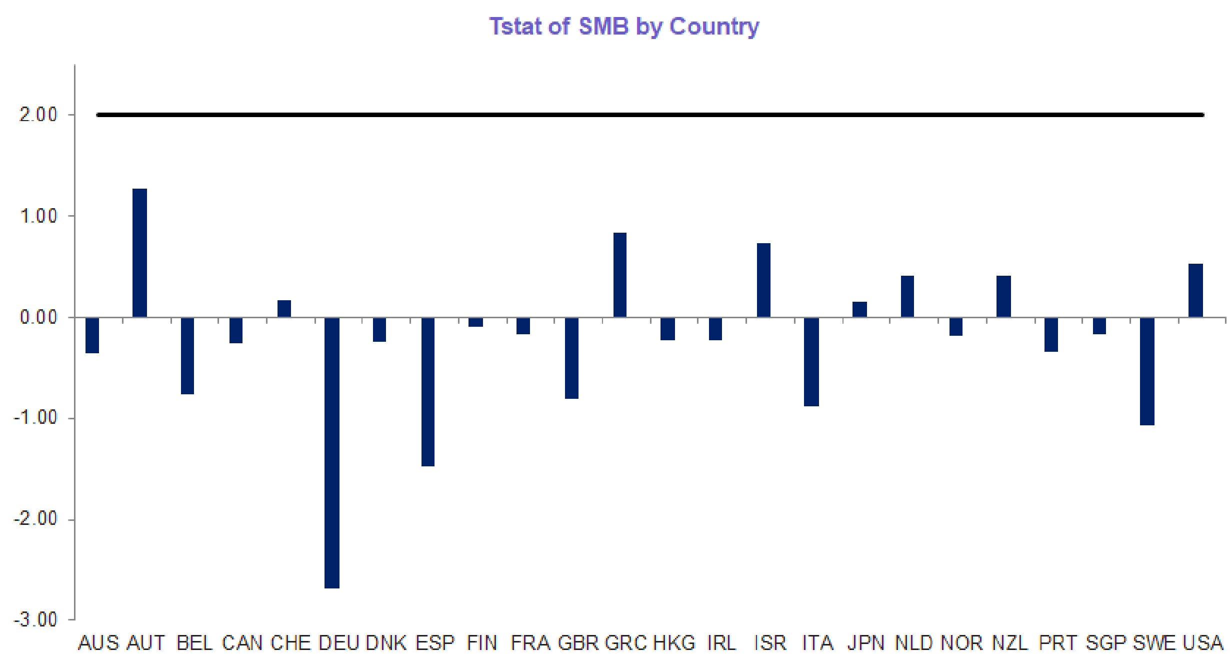
Another way to assess the robustness of any factor is to examine its efficacy in other markets. Other equity markets provide a set of out of sample tests for any factor and help to guard against data mining. They also can help build a better diversified global factor that offers a more stable return premium, since diversification benefits often exist across international equity markets. Much research has shown that factors such as value, momentum, and quality/defensive work extremely well in other markets (Fama and French [1998], Rouwenhorst [1998], Liew and Vassalou [2000], Griffin, Ji, and Martin [2003], Chui, Wei, and Titman [2010], Asness, Moskowitz, and Pedersen [2013], Fama and French [2014], Frazzini and Pedersen [2013], and Asness, Frazzini, and Pedersen [2016]). How well does size fare in other equity markets?

We examine 24 international equity markets and compute an SMB portfolio in each market following the same procedure above, which matches that of Fama and French [1993]. The universe of stocks in each country is the MSCI universe and data are obtained from World Scope and cover the period January 1984 to December 2017. The graph below reports the average SMB returns across countries grouped into regions: Europe (Austria, Belgium, Switzerland, Germany, Denmark, Spain, Finland, France, UK, Greece, Ireland, Israel, Italy, Netherlands Norway, Portugal, Sweden), North America (Canada, USA), Pacific (Australia, Honk Kong, Japan, New Zealand, Singapore), and Global portfolios. The region portfolios are equal-weighted averages of the country-specific SMB portfolios in each region.

The figure below reports the *t*-statistic of the CAPM alphas of these regional SMB portfolios. As the figure shows, none of the *t*-statistics are even close to being reliably positive, and in fact most are negative. Thus, we see no evidence at all of a positive size premium in these other markets.



We can also look country-by-country at how size has fared. The figure below plots *t*-statistics of the SMB returns for each of the 23 countries we examine outside of the U.S.



As the figure shows, none of the countries exhibits a significant SMB premium (the standard threshold for significance of a *t*-statistic of 2.0 is highlighted on the graph). The highest positive *t*-statistic is for Austria and it is only 1.28. Moreover, 16 out of 23 countries exhibit a *negative t*-statistic, where the average return to SMB is actually negative, not positive (though we note these are largely statistically insignificant). Hence, there's more evidence to support a negative size premium than a positive one, though the evidence is most consistent with there being no size premium at all.

We can also look at emerging markets. Here, the historical time series of returns is more limiting (beginning in 1994). The table below reports results for emerging markets and for the U.S. over the same time period for comparison.

	Annual Return	Annual Vol	Raw T-Stat	Sharpe	Annual Alpha vs CAPM	Alpha T-Stat vs CAPM	IR vs CAPM	Start Date	End Date
Emerging mkts	3.8%	14.0%	1.31	0.27	3.7%	1.27	0.27	7/31/1994	1/31/2018
USA	2.6%	16.6%	0.75	0.16	1.6%	0.47	0.10	7/31/1994	12/31/2017

The size premium in emerging markets is positive, and more so than in the U.S., but it still remains insignificant (*t*-statistic of only 1.31 for raw returns and 1.27 for CAPM alpha).

Finally, we note that the international samples cover a period over which the U.S. size premium is weak (1984 to 2017). Hence, these are not completely independent tests. Nevertheless, nearly every country fails to deliver a size effect in this sample, so the poor performance of size over this period is robust in every country.

Overall, there is little evidence of a size premium in other equity markets globally. This finding highlights another robustness test the size effect seems to fail.

Fact: The size effect is either not applicable or does not work for other asset classes.

Another virtue of some of the leading asset pricing factors is that they can be applied more broadly to other asset classes. For example, value, momentum, carry, and defensive factors have all been shown to work well in explaining returns in other assets classes, such as fixed income, credit, currencies, commodities, equity index futures and options (see Asness, Moskowitz, and Pedersen [2013], Koijen, Moskowitz, Pedersen, and Vrugt [2016], Frazzini and Pedersen [2013], Asness, Ilmanen, Israel, and Moskowitz [2015], and Israel, Palhares, and Richardson [2017]). The application of a factor to other assets is appealing because general theories of asset pricing are not asset-specific; they should apply to any financial claim or asset. Also, using the same characteristic to describe returns in many asset classes provides a unifying framework tying those asset classes together. Finally, looking at other asset classes also provides yet another out of sample test to guard against data mining biases.

Does size also help as a unifying concept across asset classes? No. First, the concept of "size" is a more difficult concept to apply outside of equities. What is the "size" of a currency or government bond or a commodity? So, right away the concept of size is ill-suited to describe returns in many asset classes.

But, perhaps we can think a bit more creatively about size in some other asset classes to test for a size premium in those asset classes. We can start by looking at country equity indices, where size can be defined as the aggregate sum of market caps of all stocks that comprise the index in each country. Examining country index portfolios, we can rank countries by their total stock market capitalization and form an SMB portfolio from among the countries. We examine two universes of country index portfolios: 1) developed markets (containing the 24 country index portfolios from January 1975 to December 2017) and 2) emerging markets (containing 25 emerging country index portfolios from January 1988 to December 2017). We go long the smallest half of countries and short the largest, equal weighting the countries in each leg of the strategy. The table below reports the performance of these size-based portfolios among country indices.

SMB Country Index Portfolios

	Annual Return	Annual Vol	Raw T-Stat	Sharpe	Annual Alpha vs CAPM	Alpha T-Stat vs CAPM	IR vs CAPM	Start Date	End Date
EQ	1.3%	7.8%	1.06	0.16	2.0%	1.68	0.26	1/31/1975	12/31/2017
EQE	0.6%	3.7%	0.88	0.16	0.5%	0.81	0.15	1/31/1988	12/31/2017

There is no evidence of a size premium among developed equity market or among emerging market country indices. The returns to size are positive for both, but statistically insignificant. This contrasts with what researchers have found for value, momentum, carry, and quality/defensive across these same country indices (see Asness, Moskowitz, and Pedersen [2013], Asness, Ilmanen, Israel, and Moskowitz [2015], and Koijen, Moskowitz, Pedersen, and Vrugt [2017]), where significantly positive premia are present.

Since size is a firm attribute that researchers have applied to equities, another natural place to examine size is the other side of a firm's balance sheet – its corporate debt. Using the market capitalization of the firm's equity, we sort firms into deciles based on size, but instead of going long the equities of the smallest 10% of firms and short the stocks of the largest 10% of firms, we instead go long and short their respective corporate bonds. The table below details the results, separating the universe of US corporate bonds into high yield and investment grade separately (with the investment grade universe containing about 500 bonds on average and the high yield universe about 450 bonds on average). The sample period is January 1997 to December 2017.

SMB Credit Portfolios

	Annual Return	Annual Vol	Raw T-Stat	Sharpe	Annual Alpha vs CAPM	Alpha T-Stat vs CAPM	IR vs CAPM	Start Date	End Date
Credit HY	-5.2%	13.7%	-1.73	-0.38	-8.0%	-2.91	-0.64	1/31/1997	12/31/2017
Credit IG	0.5%	3.2%	0.76	0.17	-0.2%	-0.25	-0.06	1/31/1997	12/31/2017

As the table shows, there is no size premium at all among corporate bonds in the US. A portfolio long small firm credit and short large firm credit produces a negative average return among high yield bonds of -5.2%, with a CAPM alpha of -8.0% (*t*-statistic = -2.91). This is the opposite sign of the size premium claimed in equities. It is a size discount. Among investment grade bonds, we find nothing – an insignificant size premium of 0.5% in raw returns with a -0.2% CAPM alpha. The evidence for a positive size premium in credit is simply not there.

For other asset classes, the notion of size is less direct. For example, we can examine currencies by looking at the size of various countries, using their GDP as a measure of economic size (see Hassan [2013]). We could do the same for government bonds. However, for commodities it is unclear what (if any) measure of size would make sense. In fact, we could not come up with one!

To test one of these markets, we apply a size-based strategy to currencies, where we use the GDP of each country to rank currencies relative to the US dollar. Specifically, we go long the currencies (relative to the US dollar) of the smallest half of countries and short the currencies of the largest half of countries, where we equal weight countries in the long and short legs. We do this for both developed markets (24 currencies from January 1980 to December 2017) and emerging markets (23 currencies from January 1997 to December 2017).

SMB FX Portfolios

	Annual Return	Annual Vol	Raw T-Stat	Sharpe	Annual Alpha vs CAPM	Alpha T-Stat vs CAPM	IR vs CAPM	Start Date	End Date
FX	0.1%	2.3%	0.34	0.06	0.2%	0.41	0.07	1/31/1980	12/31/2017
FXE	-1.6%	4.8%	-1.52	-0.33	-1.0%	-0.95	-0.21	1/31/1997	12/31/2017

The results show that there is no size premium among currencies either.

Despite the fact that many studies document significant value, momentum, carry, and defensive return premia in bonds, country equity index futures, commodities, currencies, and equities globally, we find that size fails to deliver a consistent premium in other asset classes and is less intuitive in other asset classes.

Some might argue that size is really a proxy for liquidity, and that if we had looked at liquidity in these other asset classes we might have found a premium. The relation between size and liquidity is an issue we will discuss later, and the broader concept of liquidity does indeed make more sense and apply more generally to other asset classes. However, the characteristic of size per se does not seem to capture returns in other asset classes and is often ill-suited as a characteristic for differentiating securities in other asset classes.

Fact: The size effect mostly comes from microcap stocks.

One criticism of the size effect is that whatever size premium is present is concentrated in microcap stocks that are extremely small and so are difficult to trade.[12] We will discuss trading costs and other implementation issues later, but for now we test the conjecture that the returns to size are all concentrated in extremely small stocks. We report SMB returns for various subsamples of data where we remove the smallest *n*% of firms, and let *n* range from 0.1% to 30.0%. The table below reports the results.

SMB Returns Removing *n* % of Small Stocks

Exclude	Annual Return	Annual Vol	Raw T-Stat	Sharpe	Annual Alpha vs CAPM	Alpha T-Stat vs CAPM	IR vs CAPM	Start Date	End Date
0.00%	6.3%	26.3%	2.28	0.24	3.6%	1.33	0.14	7/31/1926	12/31/2017
0.10%	6.3%	26.3%	2.28	0.24	3.6%	1.33	0.14	7/31/1926	12/31/2017
0.25%	6.3%	26.0%	2.30	0.24	3.6%	1.35	0.14	7/31/1926	12/31/2017
0.50%	6.2%	25.8%	2.30	0.24	3.5%	1.34	0.14	7/31/1926	12/31/2017
1.00%	5.8%	25.6%	2.15	0.23	3.0%	1.16	0.12	7/31/1926	12/31/2017
5.00%	3.5%	23.5%	1.44	0.15	0.7%	0.30	0.03	7/31/1926	12/31/2017
10.00%	2.4%	21.8%	1.04	0.11	-0.4%	-0.21	-0.02	7/31/1926	12/31/2017
20.00%	0.5%	18.4%	0.25	0.03	-2.0%	-1.06	-0.11	7/31/1926	12/31/2017
30.00%	-0.4%	17.0%	-0.23	-0.02	-2.8%	-1.65	-0.17	7/31/1926	12/31/2017

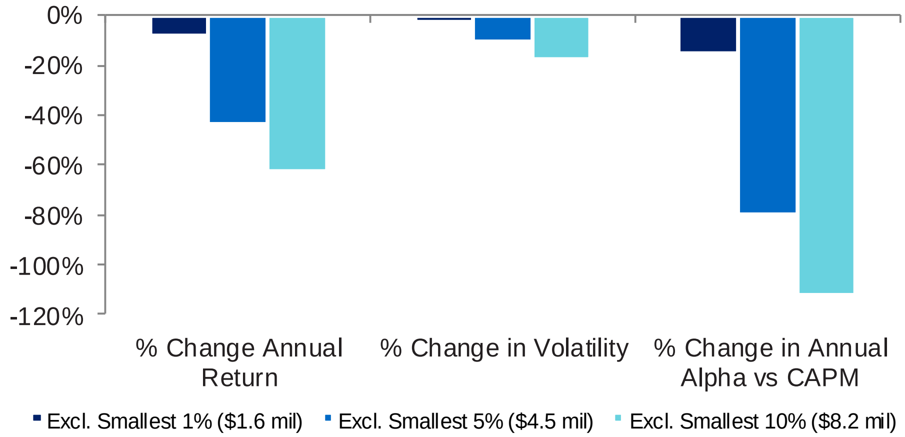
The first row reports the standard result that removes no firms for comparison, where we see a 6.3% return (3.6% CAPM alpha) with a *t*-statistic of 2.28 (1.33).[13] As we remove increasing fractions of the smallest stocks, naturally the SMB premium declines. Once we remove the smallest 5% of stocks, the SMB premium is no longer significant.

On a risk-adjusted basis (relative to the market), there is no size effect, as all of the alphas are indistinguishable from zero. However, we note that the point estimate of the alpha declines rapidly when we remove the smallest 5% of firms as well, which corresponds to firms with an average market cap of only $4.5 million; a size that is well below the Russell 3000 minimum, for instance. Hence, to the extent there is a premium for small stocks, it does indeed appear to be concentrated among the tiniest 5% of firms.

To see the influence of these tiny firms more clearly, the graph below plots the percentage change in average returns, volatility, and alpha of the size factor when the smallest 1%, 5%, and 10% of firms are removed from the portfolio.

[12] Horowitz, Loughran, and Savin [2000] find that removing stocks with less than $5 million in market cap eliminates the small firm premium. Crain [2011] and Bryan [2014] find that the small stock effect is concentrated among the smallest 5% of firms.

[13] The entries in the first row differ slightly from those reported in the table with the full-sample statistics for the 1-10 decile portfolio from Ken French's website. For this table, we computed a 1-10 decile portfolio using the sample of US equities from AQR's data library and Fama-French-style construction methods. This series and the one from French's website are 0.95+ correlated.



Avg cap breakpoints are in parentheses.

The rapid decrease in returns as the smallest fraction of firms is removed is evident from the figure. This fact has been used to argue that the small firm effect is difficult to trade, because the smallest percentage of firms is highly illiquid, volatile, and expensive to trade. We take this issue up next, which is also related to the extreme price impact experienced by such small stocks in January.

Fact: The size effect is difficult to implement in practice.

The fact that the bulk of the size effect seems to come from microcap stocks (see above), indicates that implementing a trading strategy to exploit the size effect might be difficult and costly in practice. One way to assess how costly it is to trade these stocks and how those costs vary with firm size, is to examine various measures of trading costs and liquidity from the literature across different size portfolios.

The table below examines average cost measures from the literature and practice across size decile portfolios, ranging from the smallest 10 percent of stocks (decile 1) to the largest 10 percent of stocks (decile 10). The first column reports the trading cost measure of Frazzini, Israel, and Moskowitz [2018], which is a measure of market impact from a calibrated model estimated from live executed trades from AQR Capital Management. As the table clearly shows, price impact costs are monotonically decreasing in firm size: the smallest decile of stocks has an average 95 basis points of price impact, while the largest decile of stocks experiences only 2 basis points of price impact. These significant differences in cost would severely impact the return differences between small and large cap stocks, which were slight anyway.

The remaining columns of the table report results for other cost and liquidity measures, including Amihud [2002], who uses the daily absolute price change divided by daily share turnover as a measure of illiquidity, the effective bid-ask spread from the Trade and Quote (TAQ) data from the exchanges, a measure of price impact from the TAQ data suggested by the Kyle [1985] model (see Frazzini, Israel, and Moskowitz [2018] and Hasbrouck [2009] for details on how to calculate this measure), the proportion of zero return days (as suggested by Hasbrouck [2009] and Goyenko, Holden, and Trzcinka [2009]), and a modified version of the Roll [1984] illiquidity measure, which is the square root of the negative of the autocovariance of daily log prices over the last month and is designed to capture temporary price movements from liquidity trading. For details on the computation of these measures and how they relate to actual transactions costs, see Frazzini, Israel, and Moskowitz [2018].

As can be seen from the table, all of the measures of costs and illiquidity decline steadily as you move from the smallest decile to the largest.[14] This evidence suggests that a portfolio tilted toward smaller stocks and away from larger stocks will suffer from poorer liquidity and larger transactions costs.

	Market Cap Decile	Frazzini, Israel, and Moskowitz (2018) tcost (MI in bps)	Amihud (2002)	TAQ spread	TAQ price impact (Kyle (1985) lambda)	Proportion of zero return days	Roll (1984)
Small	1	95	132771	225	1.82	37%	137
	2	55	38472	138	1.05	27%	96
	3	29	17851	96	0.68	23%	74
	4	20	8776	72	0.46	20%	57
	5	15	4777	56	0.32	18%	45
	6	12	2544	43	0.21	16%	31
	7	9	1350	33	0.14	14%	21
	8	7	656	24	0.08	12%	12
	9	5	272	16	0.04	10%	9
Large	10	2	83	10	0.02	7%	9

Furthermore, the more weight that is given to the smallest stocks, at the extreme end of the size spectrum, the higher the costs and the worse the liquidity. As a size-based strategy, like those proposed in the literature, requires investing in microcap stocks in deciles 1 and 2, the returns to these strategies are significantly affected by transactions costs that tend to eliminate what little premium might exist. For example, taking the most optimistic of our results on the size premium using a portfolio long decile 1 and short decile 10 that generated a 2.0-2.5% alpha over the market, the trading costs associated with that long-short portfolio would wipe out most of the return premium. Focusing purely on January, where all of the size returns seem to occur, would similarly be hampered significantly by trading costs (and constrained by liquidity as well, especially at larger portfolio sizes). Hence, a size-based strategy is hindered by liquidity and transactions costs that make it difficult to implement in practice.

However, as Frazzini, Israel, and Moskowitz [2018] show, trading costs can be reduced by combining multiple factors that are not perfectly correlated to each other. Much like the diversification benefits a portfolio can achieve with less correlated return sources, a portfolio can

[14] As another sign of illiquidity, we find that smaller deciles load more positively on lagged market returns, consistent with non-synchronous trading for small, illiquid stocks.

also benefit from diversification in trading costs. Combining size with other factors can lower trading costs at the margin that may make a size factor valuable in combination with other factors. Given our other research on size's interaction with quality and how the two factors are negatively correlated and can enhance each other's returns, combining size with quality can also mitigate some of the transactions costs. Taken together, a portfolio of size and quality has both a higher return premium and lower trading costs than a stand-alone size strategy, and hence may be more implementable.[15]

Fiction: The size effect is likely more than just a liquidity effect.

As the previous facts and fictions allude to, size is closely related to measures of liquidity. In fact, many scholars have argued that size is really just a proxy for a liquidity effect and that better and more direct measures of liquidity can explain the size effect. In short, these papers argue that there is no size effect per se, but that it is instead a poor proxy for a stronger liquidity effect. The idea being that size may just be a proxy for illiquidity and liquidity risk, and investors generally require compensation for holding such securities.[16]

We can test this idea using factors that attempt to capture liquidity return premia more directly and see whether the size premium is related to these liquidity factors. Using the same liquidity measures above, we form long-short portfolios that invest in the 10 percent least liquid securities (based in turn on the measures from Frazzini, Israel, and Moskowitz [2018], Amihud [2002], TAQ effective spread and lambda, proportion of zero return days, and Roll [1984]) and short the 10 percent most liquid and compute their returns over time. We then regress the returns to a long-short size decile strategy (decile 1 minus decile 10) on these various liquidity factors and report the results in the table below.

[15] Asness, Frazzini, Israel, Moskowitz, and Pedersen [2017] study the link between liquidity and size when controlling for quality, and find that while there is a tight relation between size and liquidity there is little relation between liquidity and quality measures – high quality small stocks face similar liquidity to junky small stocks. They argue that these results are consistent with liquidity-based theories for the size premium, where size is also correlated with a quality factor that is unrelated to liquidity, and so the size-liquidity relation may be partly obscured by quality. Hence, size seems to be related to both illiquidity (positively) and quality (negatively), but where liquidity and quality are not strongly related.

[16] A large literature argues that the returns to size are captured by measures of illiquidity. See Brennan and Subrahmanyam [1996], Amihud [2002], Hou and Moskowitz [2005], Sadka [2006], and Ibbotson, Chen, Kim, and Hu [2013] and measures of liquidity risk such as those of Pastor and Stambaugh [2003] and Acharya and Pedersen [2005]. Crain [2011] summarizes this evidence.

	SMB Alpha vs Liquidity	SMB Alpha t-Stat vs Liquidity	Return Correlation with Size strategy	Measure Correlation with Size strategy	Start Date	End Date
Frazzini, Israel, and Moskowitz [2018] tcost	-1.3%	-0.74	0.51	0.92	7/31/1988	6/30/2017
Amihud [2002]	-1.8%	-1.81	0.74	0.92	7/31/1963	6/30/2017
TAQ effective spread	-0.2%	-0.12	0.69	0.89	7/31/1993	6/30/2017
TAQ price impact (lambda)	1.7%	0.82	0.47	0.64	7/31/1993	6/30/2017
Proportion of zero trading days	0.6%	0.47	0.40	0.64	7/31/1963	6/30/2017
Modified Roll [1984]	1.6%	1.62	0.72	0.35	7/31/1963	6/30/2017

As the table shows, the returns to the size strategy are highly correlated with the returns to the liquidity strategies. This is not surprising given that the liquidity measures are so highly correlated with size, as indicated in the table as well and consistent with the results from the previous section. All of the alphas on the size factor are insignificant from zero and half of them are negative, after adjusting for the liquidity factor. These results are largely consistent with the literature claiming that size is just a proxy for liquidity and that any detectable size premium is captured by a liquidity premium.

Fiction: There is a strong economic story, ex liquidity, where small stocks should deserve a marginal premium over their other risk characteristics.

There are other scholars, however, who claim that size is distinctly related to expected returns beyond just liquidity. Here, size plays a special role that is linked to a return premium above and beyond those compensating investors for liquidity. While, as discussed earlier, the data seem to confirm that size has trouble predicting returns beyond liquidity measures, we can also evaluate the statement above on theoretical grounds.

If size per se carries a return premium, then there must be an economic story for why size, separate from liquidity, should be related to expected returns. Aside from liquidity, why would size be a characteristic that could drive returns? We can appeal to the two leading paradigms for thinking about return predictability – risk-based explanations consistent with efficient markets and behavioral mispricing explanations consistent with less than perfectly efficient markets.

Among risk-based stories, the size of an asset has to be related to the covariance of that asset's return to some underlying economic source of risk for which investors require compensation. Small stocks may simply have higher betas on those sources of risk – like the market, macro variables, etc. However, this explanation is not about size per se but about size being a proxy for beta with respect to some macroeconomic factor like the market. In fact, we already showed that the CAPM does a good job explaining size's returns, hence there is not a size effect per se, and size is just picking up higher beta stocks. This, therefore, is not a story about *size* carrying a premium.

Similarly, on the behavioral side, scholars have suggested that small stocks are harder to arbitrage and trade (indeed, we found they have much higher trading costs and illiquidity) and hence there will be more mispricing associated with them. But, for such stocks to carry a return premium they must be systematically underpriced. In principle, mispricing should be equally likely to cause overpricing as underpricing, but if anything, behavioral-finance theory following Miller [1977] predicts that small stocks are more likely to be overpriced rather than underpriced. The idea is that it is harder to short sell small stocks, so that their prices primarily reflect the views of optimists and are therefore overvalued. This implies a *negative* size premium. To explain why size itself is compensated it must be that people demand a larger return (lower price) to trade in small, illiquid, and costly to trade (and short) stocks – but this sounds exactly like an illiquidity premium story. The case for size itself to matter seems harder to make.

Finally, size as a characteristic that drives returns is a strange notion compared with other characteristics such as value (book-to-price), momentum (past year returns), and quality (profits-to-assets). For example, if the cost of capital were a function of size, then this by itself would be a reason to merge, and we would observe more mergers, even across very different industries and types of businesses, than we actually do. So, size is a characteristic that would be an unusual return predictor in an economic model. On the other hand, when two firms merge their value, momentum, and quality characteristics are averaged because they are ratios. Hence, the cost of capital predicted by these characteristics following a merger would be the average cost of capital of the two firms. This makes intuitive sense.

So, while the data do not seem to yield a large size premium above and beyond any illiquidity premium, theory, too, struggles with why size per se would provide a return premium separate from market risks and liquidity.

Fiction: Many anomalies being stronger among small stocks is evidence of a size effect.

The first part of this statement is true, but the latter part is false. The size effect – that small stocks outperform large stocks – is often confused with other factors, such as value, being stronger among small stocks than among large stocks. Many anomalies (though not all) are indeed stronger among small stocks, but this has nothing to do with the "size effect" or more precisely a return premium for size per se. This statement is about other return premia being stronger (at least gross of trading costs) among smaller cap stocks. This could be due to illiquidity, more limited arbitrage, higher volatility, or more retail investors associated with small stocks, all of which may exacerbate any return premium associated with other factors, but none of which necessarily have anything to do with a premium associated with small firms themselves.

For example, when looking at other factors among small versus large cap stocks, the factors are neutral to size. When we look at another factor, like value, that is long value stocks (high BE/ME) and short growth stocks (low BE/ME) within the small cap universe, the value stocks are on average the same size stocks as the growth stocks. So, a value factor that is long value and short growth among small cap stocks would net out any small cap exposure. The only return premium being picked up here is a value premium *among* small stocks. Any size premium is effectively hedged away.

The table below reports the performance of various factors – HML, BAB, UMD, RMW – formed among small cap and large cap stocks separately as described above, over the longest sample period for which we have data. Each factor is neutral to size as the longs and the shorts of each leg of the factor have equivalent size characteristics.

	Annual Return	Annual Vol	Raw T-Stat	Sharpe	Annual Alpha vs CAPM	Alpha T-Stat vs CAPM	IR vs CAPM	Start Date	End Date
BAB Large	4.2%	11.5%	3.45	0.37	6.0%	5.12	0.55	1/31/1931	12/31/2017
BAB Small	10.5%	12.3%	7.96	0.85	10.0%	7.54	0.82	1/31/1931	12/31/2017
HML Large	3.4%	13.9%	2.32	0.24	1.4%	1.03	0.11	7/31/1926	12/31/2017
HML Small	5.9%	12.6%	4.46	0.47	5.3%	4.04	0.43	7/31/1926	12/31/2017
UMD Large	6.5%	17.9%	3.45	0.36	8.9%	4.98	0.53	1/31/1927	12/31/2017
UMD Small	9.3%	16.3%	5.46	0.57	11.6%	7.12	0.75	1/31/1927	12/31/2017
RMW Large	2.2%	8.7%	1.86	0.25	3.1%	2.73	0.37	7/31/1963	12/31/2017
RMW Small	3.8%	9.1%	3.12	0.42	4.4%	3.58	0.49	7/31/1963	12/31/2017

As the table shows, for every factor the returns, Sharpe ratio, alpha, and information ratio, are a bit higher when formed among small cap stocks than among large cap stocks, especially for BAB and HML.[17] Thus, other factor returns do indeed seem to be a bit stronger (gross of transactions costs) among small cap stocks. But, none of the additional returns to these factors is driven by a size premium since each factor is neutral to size. The other factors may exhibit slightly higher returns among small stocks simply because those stocks are less liquid, more difficult to trade, have more idiosyncratic volatility, may have more retail (less sophisticated) investors present, and simply face more limited arbitrage capital, all of which could contribute to enhancing the return premium associated with other factors. However, this does not indicate there is a premium for size per se, only that other premia are larger when implemented among small cap stocks.[18]

Finally, as with most of our analysis here, we are looking at gross of t-cost returns. In practice, the net of transactions costs returns often ameliorate any factor performance differences across different size universes, since small cap stocks are more expensive and more difficult to trade. On a net of trading cost basis, the performance of many of these factors is not very different

[17] Note the lack of any value premium among large cap stocks, a notable fact we discussed in "Fact, Fiction, and Value Investing," Journal of Portfolio Management [2015].

[18] As another example, researchers often find that factor return premia are stronger when applied to emerging markets for similar reasons. Again, this indicates factor return premia are stronger within emerging markets (gross of transactions costs), but that does not mean there is an emerging market premium per se.

among small versus large cap stocks because as our earlier evidence showed, small stocks are more difficult and more costly to trade.

Fact and Fiction: The size effect is much stronger when controlling for other factors.

This one depends critically on what other factors are controlled for. We have already shown that controlling for the market portfolio (CAPM) diminishes the size effect, rendering it insignificant in most cases. The CAPM explains the returns to size even better if in addition to including the contemporaneous market portfolio return on the right hand side of the regression, we also put lagged returns of the market to account for possible nonsynchronous trading effects that may affect small, illiquid stocks. Small stocks, especially early in the historical sample period when markets were less liquid, do not trade as actively as large stocks and may not trade at all for several days (or longer). As a result, when the market moves, stocks that did not trade or traded slowly will lag the market movements and hence their betas with respect to the market will be understated (biased toward zero). This does not happen as often for large stocks that are more liquid and trade more continuously. Hence, a small minus big portfolio will have a beta biased toward zero due to potential nonsynchronous trading.

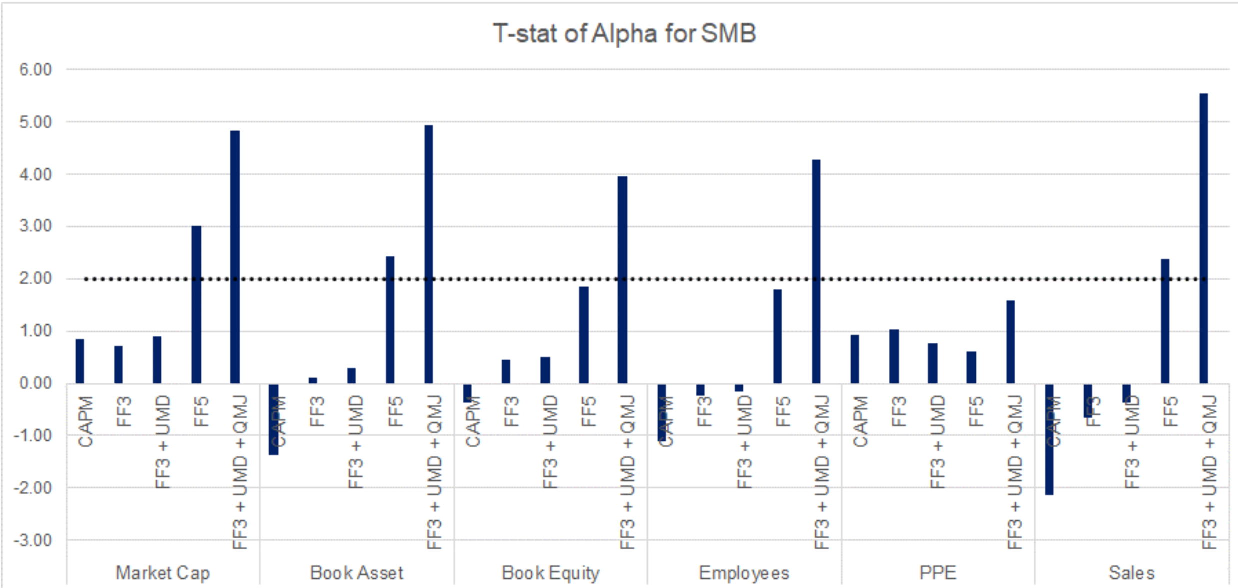
One way to account for this is to include lagged returns on the market, where small firms will exhibit a positive beta to the lagged market returns, while large firms will not. The total market exposure of firms is then simply the sum of these two betas (Dimson [1981]). Asness, Frazzini, Israel, Moskowitz, and Pedersen [2017] show that including a lagged market return increases the market beta exposure of long-short small minus big stock portfolios, which in turn reduces their alphas with respect to the market. Thus, the CAPM works even better at explaining the size effect after taking this lagged market exposure into account.

So, at least with respect to the market factor, the size effect gets weaker, not stronger. But, what about other factors? We can also run a regression of SMB's returns on the Fama and French factor HML. Here, we also include the market portfolio and its lag to see how the size effect fares when both the market and the value effect are controlled for. We refer to this model as FF3, containing the MKT, the MKT lagged a month, and HML.[19]

These regressions are run for SMB formed from market capitalization – the classic measure of size – as well as SMB portfolios formed from the non-price based measures of size we used earlier (book assets, book equity, employees, PP&E, and sales).

We then repeat these regressions by adding the momentum factor – UMD. We refer to this model as FF3+UMD. The figure below reports the alphas from all of these regressions, along with the CAPM alphas for comparison.

[19] The real Fama and French [1993] three factor model contains the market portfolio, SMB, and HML, but since we are using SMB in our analysis as the dependent variable, we obviously cannot include it on the right hand side of the regression as a control. So, instead we add the lagged market return and refer to this model as FF3, which is not to be confused with the Fama and French [1993] three factor model.



As the figure shows, none of the CAPM alphas are significant, with some of the alphas negative and none having a *t*-statistic anywhere close to +2.0. Adding HML does little to change that conclusion, and adding UMD also has a negligible effect on SMB's alpha. This suggests that while controlling for the market makes the size effect weaker, it is relatively immune to controls for value and momentum.

We next run a regression of the same SMB portfolios on Fama and French's new five factor model (FF5), which contains the MKT, its lag, HML, and two new factors from Fama and French [2014] – RMW, a profitability factor that is long "robust" or profitable firms (high operating profits-to-assets) minus "weak" unprofitable firms and CMA, an investment factor that is long "conservative" firms with low investments-to-assets and short "aggressive" firms with high investment-to-assets.[20] Here, the story changes considerably. Suddenly, the alpha of SMB is positive and strongly significant with a *t*-statistic of 3.02.

Moreover, and even more interestingly, the SMB alphas from the non-price based measures of size are also now significantly positive, with *t*-statistics ranging from just under 2.0 to 2.4. In other words, the size effect seems to have been made substantially stronger by including the two new Fama and French factors RMW and CMA. Digging into this result, it is the relation between SMB and RMW in particular that is driving this positive result for size.

So, why does the size effect become significantly stronger when controlling for the profitability factor? Because, as Asness, Frazzini, Israel, Moskowitz, and Pedersen [2017] argue and show, the size effect is confounded by a very powerful quality versus "junk" effect. They investigate the relation between size and quality and find that size, controlling for quality, not only

[20] Again, the Fama and French [1993] five factor model contains the market portfolio, SMB, HML, RMW, and CMA, but since we are using SMB in our analysis as the dependent variable, we replace SMB as a control with the lagged market return and refer to this model as FF5.

resurrects the size premium and elevates it significantly, but also helps resolve some of the patterns associated with size mentioned above.

The interaction between size and quality is especially interesting for several reasons. First, quality can be defined as a characteristic of an asset that all else equal commands a higher price. As such, size, which is based on market values, should have a strong connection to quality, where size's relation to average returns is made much clearer once controlling for quality. We can measure firm quality in a variety of ways, and Asness, Frazzini, and Pedersen [2016] and Asness, Frazzini, Israel, Moskowitz, and Pedersen [2017] measure it by using profitability, payout, growth, and safety, taking an average of these measures to form a quality factor that is long high quality stocks and short low quality/junk stocks they call "quality minus junk," QMJ.

The figure adds QMJ as a factor to the Fama and French three factors and the UMD factor (FF3+UMD+QMJ). As the graph indicates, the size premium is substantially increased after controlling for QMJ. The *t*-statistic of the alpha for size jumps from 0.91 when controlling for the FF3+UMD factors to 4.84 when adding the QMJ factor. The same substantial increase in the size premium also occurs for the non-price based measures of size, as the alpha of each SMB portfolio associated with the non-price size measures jumps from statistically no different from zero (*t*-statistics all well below 1.0) under the FF3+UMD factors to highly significantly positive (*t*-statistics close to and above 5.0).

The interaction between size and quality/junk is far stronger than size's interaction with other factors (beta, value, momentum) and accounting for it produces a more significant size premium. Regardless of the quality metric used, metrics that vary substantially both qualitatively and in terms of measured correlation, we find a much stronger and more stable size effect when controlling for a firm's quality. This is why the Fama and French five factor model also helps resurrect size, as the RMW factor based on profitability is one measure of quality.

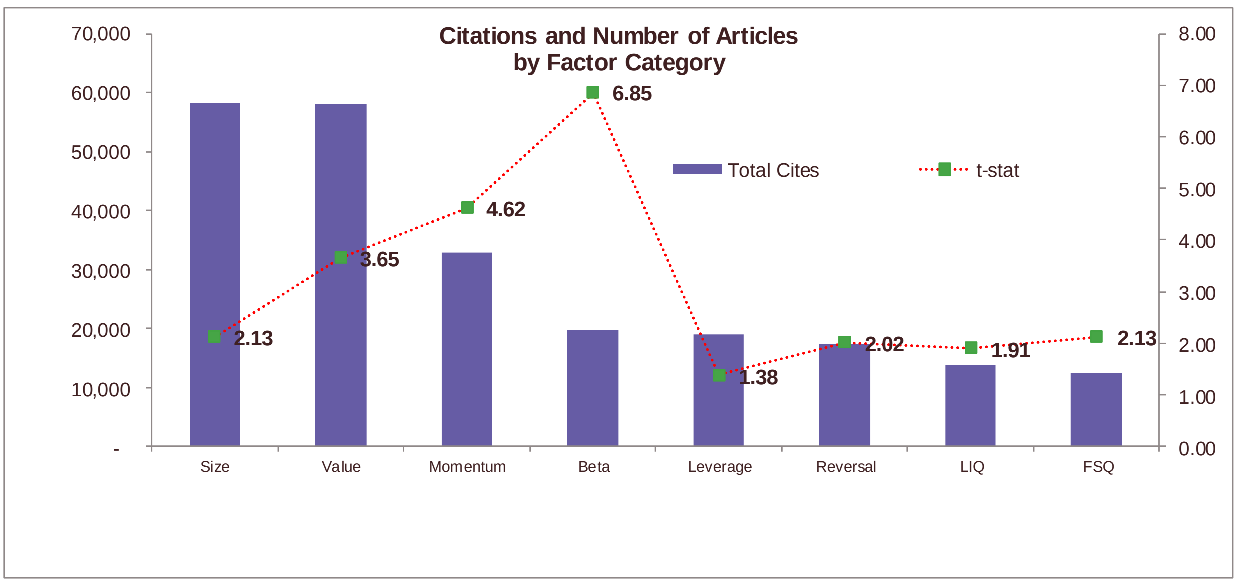
Firm size is highly confounded with firm quality, which distorts the relation between size and expected returns. Large firms tend to be high quality firms, while small firms tend to be "junky." Since high quality stocks outperform junk stocks on average, the basic size effect is fighting a strong quality effect. Going long small stocks and short large stocks, a size-based strategy is long a potential size premium but also short a quality premium, which both understates the actual size effect and introduces additional variation from the quality factor.

In addition to resurrecting the size premium, controlling for quality also reconciles many of the empirical irregularities associated with the size effect that we (and the literature) have documented. For instance, controlling for quality resurrects the size effect after the 1980s and explains its time variation, restores a linear relationship between size and average returns that is no longer concentrated among the tiniest firms, revives the returns to size outside of January and simultaneously diminishes the returns to size in January – making it more uniform across months of the year, and uncovers a larger size effect in almost two dozen international equity markets, where size has been notably weak. These results are robust to using non-market based size measures, making the size premium a much stronger and more reliable effect after controlling for quality (Asness, Frazzini, Israel, Moskowitz, and Pedersen [2017]).

Fact: The size effect receives disproportionately more attention than other factors with similar or much stronger evidence.

Finally, despite its pretty mediocre evidence and lack of theory, the size effect has received disproportionately more attention than other factors with much stronger evidence and theory behind them. For instance, as we showed earlier, value, momentum, and quality/defensive provide much stronger return evidence than size. Also, liquidity seems to provide a stronger empirical premium than size. Yet, the size effect has received a lot more attention in the literature than some of these factors.

Using Google Scholar, we added up the number of papers that have explicitly focused on the size effect (excluding this one) and added up all of the citations to those articles in the academic finance, accounting, and economic literatures. We then did the same for several other prominent factors in the literature – value, momentum, beta, leverage, reversals, liquidity, and quality (broadly defined as financial statement quality or FSQ). The results are plotted below in the figure for each factor, along with *t*-statistics of the raw returns associated with each factor over the longest possible sample period, which begins around 1926-1931 for size, value, beta, reversals, and momentum and begins around 1951 for liquidity, FSQ, and leverage.



As the figure shows, size has received much more attention than just about every factor, except value. Comparing the citations versus historical performance of each factor, it is arguable that size has received much more attention than it deserves. The evidence behind size is far more meager than many other factors that receive much less attention, and other factors that have similar strength of evidence behind them receive a lot less rumination.

The undue prominence of the size effect in the literature and practice is likely due to it being the first anomaly to challenge standard asset pricing theory (namely, the CAPM), and focus in science can often be path-dependent. But, the truth is that the premium associated with size is not very strong, not very persistent, not very robust, and may never have existed in the first place (if not for data errors and improper risk measurement).

Conclusion

As we have stated before in our other "fact, fiction" articles, if one wants to challenge our evidence and conclusions that is fine. If someone discovers something challenging or enlightening versus what we have shown, we welcome it and wish to understand it.

Bottom line for this article: addressing the facts and fictions of the size effect, we find neither strong empirical evidence nor robust theoretical support for a prominent size-based factor, despite the attention it has received in the literature and among practitioners.

References

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

Ang, A., Hodrick, R., Xing, Y., Zhang, X., 2006. The cross-section of volatility and expected returns. Journal of Finance 61, 259–299.

Ang, A., Hodrick, R., Xing, Y., Zhang, X., 2009. High idiosyncratic volatility and low returns: international and further U.S. evidence. Journal of Financial Economics 91, 1–23.

Asness, C., A. Frazzini, R. Israel and T. Moskowitz [2014]. "Fact, Fiction, and Momentum Investing," Journal of Portfolio Management, 40th Anniversary edition.

Asness, C., A. Frazzini and L. Pedersen [2014]. "Quality Minus Junk." Working Paper, AQR Capital Management.

Asness, C., A. Ilmanen, R. Israel and T. Moskowitz [2015]. "Investing with Style." Journal of Investment Management Volume 13, No. 1, 27–63.

Asness, C., T. Moskowitz, and L. Pedersen [2013]. ''Value and Momentum Everywhere.'' Journal of Finance 68(3), 929-985.

Asness, C., A. Frazzini R. Israel, T. Moskowitz and L. Pedersen [2017]. "Size Matters, If You Control Your Junk." Journal of Financial Economics.

Ball, R., 1978. Anomalies in relations between securities' yields and yield-surrogates. Journal of Financial Economics, 6, 103–126.

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

Berk, J., 1995a. A critique of size-related anomalies. Review of Financial Studies 8, 275–286.

Berk, J., 1995b. An empirical reexamination of the relation between firm size and return. Unpublished working paper, University of Washington.

Berk, J., 1997. Does size really matter? Financial Analysts Journal 53, 12–18.

Black, F., 1972. Capital market equilibrium with restricted borrowing. Journal of Business 45, 444–455.

Black, F., 1992. Beta and return. Journal of Portfolio Management 20, 8–18.

Black, F., Jensen, M.C., Scholes, M., 1972. The capital asset pricing model: some empirical tests. In Jensen, M.C., ed., Studies in the Theory of Capital Markets. New York, 79–121.

Brennan, M., Subrahmanyam, A., 1996. Market microstructure and asset pricing: on the compensation for illiquidity in stock returns. Journal of Financial Economics 49, 441–464.

Bryan, A., 2014. Does the small-cap premium exist? Morningstar Inc., Chicago, IL.

Chan, L.K.C., Karceski, J., Lakonishok, J., 2000. New paradigm or same old hype in equity investing? Financial Analysts Journal 56, 23–36.

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

Crain, M.A., 2011. A literature review of the size effect. Unpublished working paper. Financial Valuation Group, Tampa, FL.

Easterday, K., Sen, P., Stephan, J., 2009. The persistence of the small firm/January effect: is it consistent with investors' learning and arbitrage efforts? Quarterly Review of Economics and Finance 49, 1172–1193.

Fama, E.F., and K.R. French [1992]. "The cross-section of expected stock returns." Journal of Finance 47, 427–465.

Fama, E.F. and K.R. French [1993]. "Common Risk Factors in the Returns on Stocks and Bonds." Journal of Financial Economics 33, 3-56.

Fama, E.F. and K.R. French [2012]. "Size, Value, and Momentum in International Stock Returns." Journal of Financial Economics 105 p. 457-472.

Fama, E.F. and K.R. French [2014]. "A Five-Factor Asset Pricing Model." Working paper, University of Chicago.

Frazzini, A., R. Israel, and T.J. Moskowitz [2015]. "Trading Costs of Asset Pricing Anomalies." Working paper, Yale University.

Frazzini, A., R. Israel, and T.J. Moskowitz [2018]. "Trading Costs." Working paper, Yale University.

Frazzini, A., Pedersen, L.H., 2014. Betting against beta. Journal of Financial Economics 111, 1–25.

Gompers, P., Metrick, A., 2001. Institutional investors and equity prices. Quarterly Journal of Economics 116, 229–259.

Harvey, C., Y. Liu and H. Zhu [2014]. "…and the Cross-Section of Expected Returns", Working paper, Duke University Fuqua School of Business.

Horowitz, J.L., Loughran, T., Savin, N.E., 2000. Three analyses of the firm size premium. Journal of Empirical Finance 7, 143–153.

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

Hou, K. and M. A. van Dijk [2014]. Resurrecting the Size Effect: Firm Size, Profitability Shocks, and Expected Stock Returns. Working paper, Ohio State University.

Ibbotson, R., Chen, Z., Kim, D.Y.J, Hu, W., 2013. Liquidity as an investment style. Financial Analysts Journal 69, 30–44.

Israel, R. and T. Moskowitz [2013]. “The Role of Shorting, Firm Size, and Time on Market Anomalies.” Journal of Financial Economics 108(2), 275-301.

Jegadeesh, N. and S. Titman [1993]. “Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency.” Journal of Finance, vol. 48, pp. 65-91.

Keim, D.B., 1983. Size-related anomalies and stock return seasonality: further empirical evidence. Journal of Financial Economics 12, 13–32.

Keim, D.B. [1999], “An analysis of mutual fund design: the case of investing in small-cap stocks,” Journal of Financial Economics, vol. 51, pp. 173-194.

McLean, D., Pontiff, J., Watanabe, A., 2009. Share issuance and cross-sectional returns: international evidence. Journal of Financial Economics 94, 1–17.

Miller, Edward M., 1977. Risk, uncertainty, and divergence of opinion. Journal of Finance 32, no. 4, 1151–1168.

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

Penman, S., F. Reggiani, S. Richardson, A. Irem Tuna [2013]. “An Accounting-Based Characteristic Model for Asset Pricing.” Working paper, Columbia University.

Reinganum, M., 1983a. Portfolio strategies based on market capitalization. Journal of Portfolio Management 9, 29–36.

Reinganum, M., 1983b. The anomalous stock market behavior of small firms in January: empirical tests for tax-loss selling effects. Journal of Financial Economics 12, 89–104.

Roll, R., 1983. The turn-of-the-year effect and the return premia of small firms. Journal of Portfolio Management 9, 18–28.

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

Van Dijk, M.A., 2013. Is size dead? A review of the size effect in equity returns. Journal of Banking and Finance 35, 3263–3274.