Busting the Myth About Size
December 16, 2014
by Vitali Kalesnik, Noah Beck
of Research Affiliates

Many market participants (including investors, product providers, and analysts alike) assume that, just as value stocks on average outperform growth, small-cap stocks on average outperform large-caps. Unlike value, however, and contrary to popular opinion, there is little solid evidence that stock size affects performance.

A recent Research Affiliates article by Hsu and Kalesnik (2014) concluded that there are at best three factors from which investors can benefit through passive investing: market, value, and low beta. The size premium was conspicuously missing from that short list. In this article we explore empirical evidence behind the size premium in more detail. The summary below offers a preview of our findings. We let the reader examine the evidence and draw his or her own conclusion. In our opinion the preponderance of evidence does not support the existence of a size premium.

□

We are not arguing that investors should stop investing in small stocks. A portfolio of small stocks offers a certain level of diversification in an investment program dominated by large-stock strategies. Moreover, major anomalies are stronger in the universe of small stocks (likely because small stocks are more prone to mispricing). Thus, small stocks have the potential to serve as an alpha pool for skilled active managers and rules-based strategies that primarily target factors other than size. Nonetheless, we are skeptical that investors will earn a higher return simply by preferring small stocks over large.

Updating the Evidence

Banz (1981) reported that small-cap stocks outperformed large-cap stocks. For the subsequent decade the phenomenon Banz observed was considered a curious anomaly. The situation changed in 1993, when Eugene Fama and Kenneth French suggested that small stocks may expose investors to some undiversifiable risk that warrants a higher required rate of return. At that moment, the size factor took its place alongside the market and value factors in the original Fama–French three-factor model. Carhart (1997) then made the case for momentum as a fourth return factor. Today the most standard equity pricing model used in academia includes four factors: market, value, size, and momentum.

But consider this: What if a large company were split, on paper only, into two small companies? Suppose there is no change in operations, and imagine that one of the small companies booked all the cash flows on even-numbered days of the month, and the other one accounted for all the cash on odd days. In this scenario, it would be most surprising if the small companies both delivered higher returns than the original large company. Yet the size premium is precisely based on the expectation that small-cap stocks will outperform large-cap stocks!

For any reasonable economic theory explaining why small-cap stocks are supposed to outperform large-cap stocks, there is an equally plausible theory explaining why the reverse should be true. The source of the specific risk postulated by Fama and French (1993) was unclear 21 years ago, and it is still murky today. Theoretical explanations for the size premium were provided after researchers observed the anomalous regularity in returns—not the other way around. Today investors believe in the size premium on the basis of empirical evidence, not on theoretical arguments. So let's turn to the evidence with updated data.

Following the methodology employed in Fama and French (2012), we grouped stocks in each country by size into two portfolios. The large stock portfolio consists of the top 90% of the market by market capitalization, and the small stock portfolio consists of the bottom 10% of the market. Stocks within the large and small portfolios are weighted by market capitalization. To measure the premium we looked at the arithmetic difference between the small and large stock portfolio returns. We report in Table 1 the average annualized returns, volatilities, and t-statistics in 18 major developed countries from January 1982 to July 2014. Table 1 also displays data for the United States over the longer period from July 1926 to July 2014.

In the 88-year U.S. sample, the size premium is 3.4% per annum. Assuming a normal distribution of premium estimates (we will discuss later why this assumption may not be warranted), the size premium is statistically significant with a t-stat of 2.38, which corresponds to a p-value of 1.7%. After 1981, when Banz's paper appeared, the premium is positive in the United States and positive on average in the international sample, but it is not statistically significant anywhere. The substantial, statistically significant average return observed in the long-term U.S. dataset is the main reason why size is popularly believed to be one of the most important factors.

Examining the U.S. Data

Existence of the size premium in the United States is practically an article of faith in the practice of asset management as well as the academic literature. The empirical evidence, however, does not stand up very well to closer scrutiny. The data are doubtful for several reasons, including overestimated small-cap returns due to missing data on delisted stocks; the absence of transaction costs in the calculation of index returns; biases resulting from data-mining and the publishing process; and misestimated statistical measures based on the assumption of normality. In addition, there proves to be no return advantage on a risk-adjusted basis.

Delisting bias. Shareholders do not necessarily lose the full amount of their investment in a company when it is delisted from a major stock exchange. Often the stock can still be traded in the over-the-counter (OTC) market, and the investor may receive some residual value if the company is liquidated. Nonetheless, returns on stocks after they have been delisted are likely to be very negative. Moreover, all companies are subject to business and financial risks that might result in their stock's falling short of listing requirements, but small stocks by market capitalization are appreciably more likely to be removed from an exchange. Shumway (1997) pointed out that regular performance databases overestimated small-cap stock returns because they did not include returns on delisted stocks. If a database that is used in simulating portfolios omits the strongly negative returns of delisted stocks, the hypothetical results will be better than what actual portfolios can achieve in practice.

To estimate the impact of the delisting bias on the size premium, Shumway and Warther (1999) looked at the smallest and the most distressed stocks for which they could obtain reliable data, namely, stocks listed on the NASDAQ exchange. We represent their findings in Figure 1. The chart shows the average monthly returns for 20 groups of stocks sorted by size before and after correcting for the upward bias in the database. Clearly, the smallest stocks are significantly more affected by the delisting bias. After adjusting for the delisting bias, the statistical significance of the size premium completely disappears. It is unreasonable to suppose that the effect Shumway and Warther quantified for NASDAQ stocks is missing from other exchanges.

Transaction costs. Theoretical simulations ignore an important component of investment performance measurement: trading expenses—the actual costs of buying or selling investments. Small stocks by definition have much lower trading capacity and, correspondingly, much higher transaction costs. Soon after the first articles documenting the size effect appeared, researchers asked how much of the premium remains when trading costs are taken into account. Stoll and Whaley (1983) showed that transaction costs accounted for a significant part of the size premium for stocks listed on the New York Stock Exchange and the American Stock Exchange.

Data-mining and reporting bias. There are literally hundreds of known factors in the existing literature, and many papers documenting new factors are published every year. In our opinion the vast majority of these factors are spurious products of data-mining. We are not alone in taking a skeptical position. Lo and MacKinlay (1990), Black (1993), and MacKinlay (1995), among others, have argued that many factors, notably including size, are likely to be a result of data-mining. And, in finance no less than the physical and biological sciences, striking results—especially new discoveries—tend to win the competition for space in academic journals.

The standard procedure for determining whether a factor is statistically significant is to see if its t-stat crosses a certain threshold. Normally the threshold is set at 1.96 for a 5% confidence level. With a t-stat of 2.38, the U.S. size premium passes this test for the 1926–2014 sample. But Harvey, Liu, and Zhu (2014) rightly observed that if many researchers are looking for statistical irregularities, then the 1.96 criterion is too low; it allows many inherently random outliers to be misidentified as valid factors. They argue that the threshold for the size factor should have been closer to a t-stat of 2.50 in 1993.[1] Size does not pass this test.

Non-normality of returns. Standard statistical testing assumes that the estimate of a variable—in this case, the average of the size premium —quickly converges to a normal distribution.[2] If, however, the underlying data include large outliers, then the assumption of normality is

unfounded. The differences between the small and large stock portfolio returns exhibit just such outliers. Figure 2 is a histogram of the return differences. For comparison, we display on the same chart a normal distribution with the same mean and standard deviation.

We indicate on the chart four extreme outliers of 6 sigma or higher. "Sigma" may be an unfamiliar statistical term, so let us put these outlier returns in perspective. The 23.6% premium registered in January 1934 is a 6-sigma event. If it were drawn from normal distribution, this would be a one-in-67-million-year event, like the one that wiped out the dinosaurs. The 27.2% difference in returns in September 1939 is a 6.9-sigma event; in a normal distribution, it would have about a one-in-five chance of occurring in the 4.5 billion years since the planet earth came into existence. The 33.8% premium in August 1932 is an 8.6-sigma event, and the 51.6% premium in May 1933 is a 13.1-sigma event. If these last two outliers were drawn from a normal distribution, each would have much less than a one-in-a-hundred chance of occurring in the entire 13.8 billion years the universe has existed.

To add to the problem, all four outliers occurred in the 1930s. If they were removed, the estimated size premium in Table 1 would drop from 3.4% to 1.9% and lose statistical significance. (There is a similar outcome in the post-war period: The estimated size premium is about 1.9% premium with a t-stat of 1.52.) We do not argue, however, that truncating or otherwise transforming the sample will give us a better estimate. What happened in the 1930s is very valuable information about the economy and the stock market. The average return from the full sample, including the unadjusted outliers, is the best estimate available as long as the statistical bounds around it are borne in mind. If the size premium is predicated on exceedingly rare events, then we'll have to wait many lifetimes to determine with confidence whether or not it exists.

No risk-adjusted benefit. Academics are interested in the arithmetic average returns in a simulated long/short portfolio, but practitioners are concerned with the actual risk-adjusted returns that they can generate from their investments—and the majority do not engage in short-selling. We display in Table 2 the average geometrically chained cumulative returns of the long-only portfolios of small and large stocks. These results are produced using the same databases we used earlier in this article, so they contain the same biases that we noted above.

Small stocks outperform large stocks in this sample, but, because small stocks are generally more volatile, the Sharpe ratios reveal that small-cap investing provides a miniscule advantage in the risk-adjusted return. If investors are switching from large stocks to small in the hope of a premium, they should realize that they are increasing the volatility, too. The estimates of average returns are very noisy, and are likely overstated due to the biases we described earlier; the estimates of volatility on the other hand are real. (Estimates of the mean are always less certain than estimates of standard deviation.) We suggest that investors seeking higher returns consider boosting their overall equity allocation rather than chasing the illusory size premium in an attempt to add risk on the cheap within the existing allocation. A large-cap stock portfolio would have higher returns than a mix of small-cap stocks and risk-free assets designed to have the same volatility. In other words, the added risk of small-cap stocks is essentially uncompensated. Note that even in the only data set with a statistically significant size premium (i.e., the U.S. full sample from 1926–2014), the Sharpe ratio is actually lower for small stocks.

Concluding Remarks

We placed our inquiry in a historical context, starting with Banz's (1981) paper, because the widespread belief in a size premium is largely a result of its early discovery. Market capitalization data were readily available to early researchers writing doctoral dissertations and journal articles, and, as we have seen, the performance of small stocks was exceptional in the 1930s. Eugene Fama was one of Rolf Banz's professors at the University of Chicago; in fact, as a member of Banz's dissertation committee, he was intimately familiar with Banz's research on the small-cap anomaly.[3] Fama and Kenneth French included the size premium in their influential three-factor model, an analytical advance that opened the gate for empirical research into studying factors previously unexplained by then-existing theories. Riding on the popularity of the Fama–French theory, the size premium was soon entrenched in the pantheon of risk factors.

Berk (1997) argued that the size premium observed in the data is nothing more than a poor way of value investing. Value investing relies on buying cheaply priced companies as measured by a ratio of price to company fundamentals. Investing based on size, measured by company market capitalization, would use only the price side of the valuation measure. Because it would therefore use only a fraction of the relevant information, the strategy is significantly weaker than a value strategy that uses prices as they relate to company fundamentals. In our view, Berk's argument is, to date, the strongest explanation why the size premium is observed.

However, we go one step further. If Berk questioned the size premium as a separate factor, we question the size premium as a

phenomenon. Today, more than 30 years after the initial publication of Banz's paper, the empirical evidence is extremely weak even before adjusting for possible biases. The return premium is not statistically significant in any of the international markets, whether taken alone or in combination. The U.S. long-term size premium is driven by the extreme outliers, which occurred three-quarters of a century ago. These extreme outliers confound the standard techniques of setting confidence bounds around the estimated premium. Finally, adjusting for biases, most notably the delisting bias, makes the size premium vanish. If the size premium were discovered today, rather than in the 1980s, it would be challenging to even publish a paper documenting that small stocks outperform large ones. All this evidence makes us question the existence of the size premium as such.

We are not arguing that investors should completely abandon small stocks. Small stocks are more volatile than large stocks, and they receive considerably less attention from sell-side analysts. Consequently, small stocks are more likely to be mispriced. The major anomalies are, in fact, stronger in the small-cap sector. Small stocks are more attractive as an alpha pool to be fished by skillful active managers and exploited by rules-based value and momentum strategies.

Endnotes

1. The authors argue further that "a newly discovered factor today should have a t-ratio that exceeds 3.0." Page 35.

2. This result relies on the central limit theorem, which says that, as the number of random observations increases, the arithmetic average converges to a normal distribution. If the observations include extreme outliers, the convergence can be either extremely slow or may not occur at all.

3. Fox (2009), page 204.

References

Banz, Rolf W. 1981. "The Relationship Between Return and Market Value of Common Stocks." Journal of Financial Economics, vol. 9, no. 1 (March):3-18.

Berk, Jonathan B. 1997. "Does Size Really Matter?" Financial Analysts Journal, vol. 53, no. 5 (September/October):12–18.

Black, Fischer. 1993. "Beta and Return." Journal of Portfolio Management, vol. 20, no. 1 (Fall):8–18.

Carhart, Mark M. 1997. "On Persistence in Mutual Fund Performance." Journal of Finance, vol. 52, no. 1 (March):57–82.

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

———. 2012. "Size, Value, and Momentum in International Stock Returns." Journal of Financial Economics, vol. 105, no. 3 (September):457–472.

Fox, Justin. 2009. The Myth of the Rational Market: A History of Risk, Reward, and Delusion on Wall Street. HarperCollins e-books.

Harvey, Campbell R., Yan Liu, and Heqing Zhu. 2014. "…And the Cross-Section of Expected Returns." NBER Working

Paper No. 20592. Available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2513152## OR Available at nber.org/papers/w20592.

Hsu, Jason and Vitali Kalesnik. 2014. "Finding Smart Beta in the Factor Zoo." Research Affiliates (July).

Lo, Andrew W., and A. Craig MacKinlay. 1990. "Data-Snooping Biases in Tests of Financial Asset Pricing Models." Review of Financial Studies, vol. 3, no. 3 (Fall):431–467.

MacKinlay, A. Craig. 1995. "Multifactor Models Do Not Explain Deviations from the CAPM." Journal of Financial Economics, vol. 38, no. 1 (May):3–28.

Shumway, Tyler. 1997. "The Delisting Bias in CRSP Data." Journal of Finance, vol. 52, no. 1 (March):327-340.

Shumway, Tyler, and Vincent A. Warther. 1999. "The Delisting Bias in CRSP's Nasdaq Data and Its Implications for the Size Effect." Journal of Finance, vol. 54, no. 6 (December):2361–2379.

Stoll, Hans R. and Robert E. Whaley. 1983. "Transaction Costs and the Small Firm Effect." Journal of Financial Economics, vol. 12, no. 1 (June):57–79.