

Special edition: Can you trust investment research?

April 28, 2016

by Jeffrey Briskin

Investment strategy research has become a cottage industry, with hundreds of studies published every year claiming that actively managed strategies, from factor-based investing to market timing, have greater potential to generate alpha. Those claims are usually based on backtesting performance over various market timeframes.

Campbell Harvey views most of this research with a high degree of skepticism.

Harvey is the J. Paul Sticht Professor of International Business at Duke University's Fuqua School of Business in Durham, North Carolina, and a research associate of the National Bureau of Economic Research in Cambridge, Massachusetts. He also serves as president of the American Finance Association.



We previously wrote about Harvey's work [here](#). But in an April 5 presentation at The Journal of Portfolio Management Research Summit in Boston, Harvey elaborated on his research and voiced his skepticism of published investment research that is often misleading or based on false assumptions. His presentation summarized some of the ideas that appeared in *Backtesting*, an article he co-wrote with Professor Yan Liu of Texas A&M that was originally published in 2015 in *The Journal of Portfolio Management*.

He began his presentation by demonstrating that researchers – and advisors and investors who act upon their research – often fall victim to committing “type 1” and “type 2” errors.

A type 1 error occurs when someone acts on a belief when the evidence clearly demonstrates that this belief is false. An example is when an antelope flees after hearing what it believes to be the sound of a stalking cheetah when in reality there are no cheetahs for miles around.

Type 2 errors occur when someone fails to act even when his or her belief in a certain reality is proven false. An example would be the same antelope, believing that there are no cheetahs nearby, fails to flee when one suddenly appears in front of him.

Type 1 errors are hard-wired into our DNA. Without them, we wouldn't react quickly to threats – real or imagined. The cost of making a type 1 error is, in general, relatively small. Fleeing from a false alarm costs time and energy, but we live to tell the tale of our mistake. Committing a type 2 error, on the other hand, can lead to death, injury or other catastrophes.

Investors and advisors often commit type 1 errors, such as timing a trade with the mistaken belief that certain market or economic events will result in a large gain or protect against loss. Of course, sometimes those bets pay off, but over the long run, repeated use of these same tactics – especially in short-term trading situations – rarely results in market-beating performance for anyone other than the most experienced traders with the most sophisticated research and technology at their disposal.

Likewise, investors who commit type 2 errors, such as irrationally refusing to bail out of a free-falling stock of a company on the verge of bankruptcy, can lose their entire investment.

Since most investment professionals aren't likely to commit many type 2 errors, Campbell's presentation and paper focused on how advisors can detect type 1 errors in investment research and use this knowledge to determine whether a given strategy merits serious consideration.

False patterns and flawed assumptions

Harvey contends that too much investment research is flawed because researchers base their findings on analysis of limited data sets or commit the errors of *overfitting* and *apophenia*, the human tendency to believe in false cause-and-effect relationships or perceive patterns in nature that aren't really there. The superstition that broken mirrors cause bad luck is an obvious example. But Harvey asserts that a great deal of investment research reflects similar flawed thinking.

From factor-based investing to quantitative analysis, researchers are constantly scouring decades of market and economic data to uncover "secrets" that can give investors an edge. But, according to Harvey, most published research uses faulty methodologies or contains only results that appear to "prove" that their strategy works.

A hypothetical example of flawed research at work

In this very simplified scenario (mine, not Harvey's) researchers wish to see if there is a relationship between weather on Wall Street and closing stock market prices. After comparing years of daily Manhattan weather reports with closing S&P 500 prices, they publish a paper claiming that there is a positive correlation between rainy and snowy days in New York City and declines of 0.25% or more in the S&P 500 when trading closes on these days. Their results claim a statistical significance level of $p \leq .04$. This means that the finding has only a 4% chance or less of being untrue (in the research world, any finding with a p factor of .05 or less is generally considered to be a statistically significant result).

Convinced that this new "factor" will give them a competitive edge, advisors schedule short trades to execute on days during March when stormy weather is predicted for Manhattan. When the month ends, they find that market declined on only six of March's eleven rainy days. As the advisors count their losses, they ask, "What went wrong?"

Advisors inevitably discover that the researchers tested this "weather/market close" hypothesis over 20

different time periods ranging from six months to 50 years and only published the results based on one 10-year period where the correlation was statistically significant. Researchers also excluded days when the S&P 500 rose or declined by less than 0.25%. And during the “significant” 10-year period the average number of bad-weather days in New York City was much higher than those in the other periods they analyzed, creating many more opportunities for “stormy weather/stormy market” correlations to occur.

Weighing the evidence

Should advisors ignore the hundreds of studies advocating various investment strategies published in financial journals each year? Not at all, said Harvey. But advisors should try to evaluate the evidence supporting the results and adjust expectations accordingly. His article offers statistical tools that can aid in this process.

Advisors should start by investigating the methodology of the research. Is it flawed by statistical biases and selective data mining that can exaggerate results and ignore the possibility of empirical abnormalities? Harvey believes these flaws taint most research and are particularly endemic to factor-based research because the limited “pool” of historical data that researchers can use to identify profitable investment strategies creates a greater risk of overfitting and apophenia.

Advisors also should investigate how many data tests the researchers conducted before they identified one with statistically significant results. The more “failed tests,” the higher the likelihood that the results of the “successful” test are less significant than they appear. To calculate this, Harvey recommends using a simplified version of the Bonferri method, which is designed to counteract the effect of multiple comparisons.

While Bonferri formulas can be highly sophisticated, Harvey said that most advisors can use a simple method where the significance level of the “successful” test is multiplied by the total number of tests (or an estimate of how many tests may have been conducted). Using our “weather/market” example above, you would multiply the p-value of 0.04 times the 20 total tests researchers conducted. This results in a p-value of 0.80, which means that there is only an 80% probability that the research results are statistically valid.

Statistical validity and the Sharpe haircut

Most published investment research includes a Sharpe ratio to show how the strategy would have performed on a risk-adjusted basis during the selected timeframe. According to Harvey, many investors and advisors who are considering implementing these strategies typically apply a standard “50% haircut” that discounts the Sharpe ratio by half to account for potential research biases and flaws.

However, Harvey thinks that while the standard 50% haircut appropriately “weeds out” research with low Sharpe ratios, it disproportionately penalizes research generating high Sharpe ratios that may deserve consideration. Instead, he outlines a formula in his article for calculating a more appropriate haircut based on an estimate of the statistical validity of research findings.

In the example he provides in the article, he postulates that a researcher who claims his strategy generates a Sharpe ratio of 0.75 with a p-value of 0.0008 is showing only one single test out of 200 total tests conducted (if the actual number of tests isn't known, a best-guess estimate can be used). Using the Bonferroni method, the significance would be reduced to a much less reliable p-value of 0.15 (200 x .0008). Inputting this result and the original 0.75 Sharpe ratio into Harvey's formula generates an adjusted Sharpe ratio of 0.32. In this case, employing multiple testing assumptions results in a "Sharpe haircut" of about 60%.

This is a vastly simplified summary of one approach Harvey discussed at the presentation. In his article, he discusses three different approaches to calculating appropriate Sharpe haircuts and tests each of these against three investment strategies: earnings-to-price ratio (E/P), momentum (MOM), and betting against beta factor (BAB). You can view his results by downloading the [article](#).

Is there any truly trustworthy research?

While Harvey thinks that some research-based investment strategies may be worth considering, he cautions that most empirical research findings are likely to be false even when the statistics are done properly. This suggests that factor-based products such as smart-beta ETFs and funds, which have become a staple of many retirement accounts, will not deliver the "outperformance" predicted by their underlying research assumptions. Advisors and investors should approach investment products whose performance is predicated on historical results with the same level of skepticism they should bring to evaluating nutritional supplements backed by questionable claims of efficacy: *Caveat emptor*.

Jeffrey Briskin is director of marketing at Advisor Perspectives.