Practical Applications of

Backtesting

PA Practical Applications
Institutional Investor Journals

★ *Bernstein Fabozzi/Jacobs Levy Awards Winner*

Authors: Campbell R. Harvey and Yan Liu
Source: *The Journal of Portfolio Management,* Vol. 42, No. 1.
Report Written By: Howard Moore
Keywords: Backtesting, T-Statistic, Sharpe Ratio, Duke University, Texas A&M University, Man Group

Overview

When a new trading strategy is presented, the results often look too good to be true. In many cases, the good performance is a result of data mining. In order to put the strategy into real-world practice, it's necessary to make some corrections to the backtests. Routinely, practitioners simply haircut the Sharpe ratio by 50%.

Cam Harvey of the **Fuqua School of Business** and **Yan Liu** of **Texas A&M University** always thought it strange that this arbitrary adjustment is so widely used without being seriously questioned. In an interview with **Institutional Investor Journals**, Harvey and Liu discuss a proposed statistical framework that systematically accounts for multiple tests and yields the appropriate haircut for any given Sharpe ratio. In their article, they suggest a profit hurdle for any strategy to be deemed significant. Their method permits investors make decisions in real time on a proposed strategy's viability

Practical Applications

- **Recognize the problem.** Despite its widespread use, the Sharpe ratio for a particular investment strategy can be misleading, because of the finance profession's inevitable data mining.
- **Pay attention to the formula.** Asset management firms developing trading strategies and institutional investors trying to evaluate these strategies should use Harvey and Liu's formula.
- **Don't apply the 50% haircut.** It is always a mistake to use the 50% adjustment. The Harvey and Liu method will lead to fewer false positives (i.e., investing in something that looks good on paper but turns out to be a fluke).

Practical Applications Report

In ***Backtesting*** in ***The Journal of Portfolio Management***, Harvey and Liu present a method to adjust statistics to account for the number of strategies that were tried. "We presented the machinery to incorporate the number of elements tried," says Liu. The intuition is the following: If there is a one-shot test, there is no adjustment. The more strategies you try, however, the higher the chance that some strategy looks good by luck. "So, the more strategies you try, the bigger the haircut," he explains.

★*This article was named Best Article in the 17th Annual* Bernstein Fabozzi/ Jacobs Levy Awards.

Key Definitions

Back testing

Testing a trading strategy on prior time periods via simulation. The process relies on recent past data to gauge the efficacy of a proposed strategy.

Backtest overfitting

A backtest is considered fit when the tested strategy is profiting from a signal. When the tested strategy is profiting from noise, it is considered overfit. Backtest overfitting occurs when a researcher makes a false discovery as a result of overfitting a backtest.

Data mining

The practice of looking for patterns and trends in large amounts of data.

Sharpe ratio

A common way to measure the historical risk-adjusted return on an investment by taking the average historical return minus the risk-free return divided by the standard deviation.

T-statistic

A hypothesis test used to determine whether a population mean is equal to, less than or more than a hypothesized value.

For example, suppose that 100 strategies were tried before settling on the final one, and suppose the selected strategy had a low Sharpe ratio, say, 0.5. The usual approach simply would be to haircut the strategy to 0.25. Harvey and Liu believe there is a more effective way to assess this.

Fifty percent is the wrong adjustment, Harvey cautions. If a large number of strategies have been tested and a modest Sharpe ratio resulted, one should haircut the result to zero. It works the other way also. If you are presented with a strategy that is truly outstanding, even after allowing for the data mining, why would you decrease the Sharpe ratio by a full 50%? "In that case, it seems more reasonable to take just a little off the top," he says.

Harvey and Liu's latest research in JPM follows up on their award-winning article, ***Evaluating Trading Strategies***, in JPM's ***40th Anniversary Issue***. It was awarded Best Article in the **16th Annual Bernstein Fabozzi/Jacobs Levy Awards**.

REVERSE-ENGINEER THE SHARPE RATIO

Harvey and Liu's adjustment to Sharpe ratios is non-linear. At some point, the haircut takes the Sharpe ratio to zero. "We try to achieve 95% confidence in our results," says Harvey. This means there is only a 5% chance of a false positive. "The usual rule is to achieve a return that is two standard deviations from zero," he explains. This is the t-statistic, and it turns out that there is a simple way to translate a Sharpe ratio to a t-statistic. "Practitioners routinely make the mistake of comparing Sharpe ratios based on different sample lengths," Harvey notes. "They should not do that. They should work with t-statistics."

Suppose you are presented with a trading strategy with a Sharpe ratio. You translate this into a t-statistic and suppose it is 2.0. This implies 95% confidence—but only if this is a one-shot test. Now suppose that you learn that 10 different strategies were tried. Following Harvey and Liu's method, instead of having only a 5% chance that the finding is a fluke, there is a 50% chance (10 x 5%) that the finding is a fluke. The t-statistic for a 50% chance of an error is 0.67. "We then reverse-engineer the Sharpe ratio," Harvey says. "Voilà, this is the way to haircut it."

This method is a novel approach. "There are two very important inputs," says Liu. First, one needs to take a view on how many strategies have actually been tried. "The number does not have to be exact—just a ball park." Second, their method explicitly allows for correlation of strategies. "It makes a big difference if you are trying 100 momentum strategies versus 100 relatively uncorrelated global macro strategies," he adds.

Harvey and Liu's method is ideally suited to determine minimum profitability hurdles for proposed strategies. They provide open access code where the inputs are the desired significance level, the number of observations, the strategy volatility and the assumed number of tests. The output is the minimum average monthly return that the proposed strategy must exceed. The program, which can be copied to run the adjustment procedure, is available on Harvey's website (http://bit.ly/1o4qT5E).

True out-of-sample tests are a cleaner way to evaluate a strategy's viability.

HARVEY & LIU: Colleagues and collaborators

CAVEATS

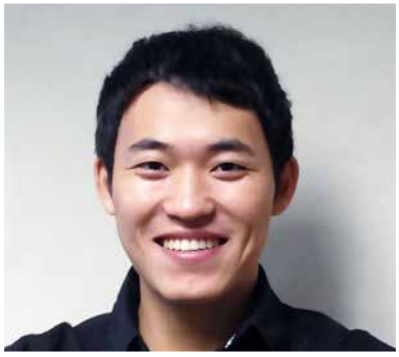
The authors point out that there are a number of caveats to their methodology. They do not observe the entire history of tests and, therefore, need to use judgment as to the number of tests to use as their input. In addition, they use Sharpe ratios as their starting point. "Our method is not applicable insofar as the Sharpe ratio is not the appropriate measure (e.g., non-linearities in the trading strategy or the variance not being a complete measure of risk)," they explain in the article.

True out-of-sample tests of a particular strategy (not a holdout sample of historical data) are a cleaner way to evaluate a strategy's viability, they add. "For some strategies, models can be tested on previously unpublished data or even on uncorrelated markets." However, they note, true out-of-sample tests are not available for the majority of trading strategies.

CREATIVE COLLABORATORS

Harvey is a Finance Professor at **Duke University** and for the last 10 years has served as the investment strategy adviser for **Man Group**, the largest listed hedge fund group in the world. "Part of the genesis for this paper is a result of practical problems that we were trying to deal with at the Man Group," he says.

Liu is an Assistant Professor at Texas A&M. The two met at Duke, where Liu was a PhD student. They started working on a project together, and shortly thereafter, Harvey agreed to chair Liu's dissertation committee. They found that they worked well together as colleagues and collaborators. "We have a very exciting agenda," says Harvey. Neither has a problem proposing ideas that might turn out to be duds, and they are able to keep their thinking creative. "We're about to release a new working paper called ***Rethinking Performance Evaluation*** that will disrupt the usual approaches to performance evaluation," he says.



Campbell R. Harvey

cam.harvey@duke.edu

Cam is the J. Paul Sticht Professor of International Business at **Duke University's Fuqua School of Business** in Durham, NC, and a Research Associate of the **National Bureau of Economic Research** in Cambridge, MA. He also serves as President of the **American Finance Association**. An internationally recognized expert in portfolio management, asset allocation, the cost of capital and global risk management, Cam also serves as an investment strategy advisor to **Man Group plc**, the world's largest, publicly listed, global hedge fund. Cam is a prolific author whose scholarly work has been honored many times. In 2015, he and co-author, **Yan Liu**, won the **16th Annual Bernstein Fabozzi/ Jacobs Levy Award for Best Paper** in ***The Journal of Portfolio Management***. He has received seven Graham and Dodd Awards/Scrolls for excellence in financial writing from the CFA Institute. He served as editor of ***The Journal of Finance*** from 2006 to 2012. A content pioneer on the Internet, Cam created a hypertextual financial glossary, the most comprehensive in the world. The glossary is available as an iPhone app, and the print version, **The New York Times Dictionary of Money and Investing**, was co-authored by **Gretchen Morgenson**. He holds a doctorate in business finance from the University of Chicago.

Yan Liu

y-liu@mays.tamu.edu

Yan is an Assistant Professor at **Texas A&M University** in College Station, TX. His research focuses on asset pricing, particularly, the interaction of statistical tests and asset pricing theories. His article on the cross-section of expected returns, written jointly with **Campbell R. Harvey** and **Heqing Zhu**, won the **NASDAQ OMX Award** for the best paper in asset pricing at the 2014 **Western Finance Association Meeting** and the best paper award at the **2014 INQUIRE Europe-UK Joint Conference**. His research on entropy bounds was accepted by the **2015 American Finance Association Meeting** in Boston. Yan's current work with Harvey is about providing robust inference on regression models in finance. He holds a BS in mathematics from Tsinghua University in China, an MS in statistics from the University of Minnesota and a PhD in finance from Duke University.