Can Statistics Actually Determine if Managers Have No Skill?

By Elisabetta Basilio, PhD, CFA | June 22nd, 2020 | Research Insights, Basilio and Johnsen, Academic Research Insight, Active and Passive Investing

- Campbell Harvey and Yan Liu
- Journal of Finance, 2020
- A version of this paper can be found here
- Want to read our summaries of academic finance papers? Check out our Academic Research Insight category

What are the Research Questions?

Whether they are selecting a manager, a factor, or a strategy, investors can make two types of mistakes.

- We thought we were hiring Peter Lynch, not this loser!
  - A Type I error (or false discovery) when selecting a manager who turns out to be unskilled.
  - That Warren Buffett deal is working up his ass.

This website uses cookies and third party services. Settings OK
The authors of this paper seek to exhibit ways in which type I and type II errors can be measured. Additionally, these types of errors present varying levels of economic importance and cost not fully taken into account by current statistical analysis.

In the context of multiple testing, it is difficult to evaluate the Type II error rate for at least two reasons: large numbers of alternative hypotheses and the multidimensional nature of the data.

In this paper, the authors propose a different approach, to consider the importance of test power. In other words, how well can statistical tests actually identify skill if it genuinely exists. Turns out that the power of common “manager performance” tests is pretty poor. To include the often-cited paper by Fama and French, “Luck versus Skill in the Cross-Section of Mutual Fund Returns.” The TLDR is that the Fama French tests would never find much evidence for manager skill, even if the reality was that there were lots of managers with investing skills.

The authors propose the following:

1. A simple metric to summarize the information contained in the parameters of interest and to evaluate Type I and Type II error rates. In essence, this metric reduces the dimensionality of the parameters of interest and allows us to evaluate error rates around what we consider a reasonable set of parameter values.

2. An error rates evaluation tool using a bootstrap method, which allows investors to capture cross-sectional dependence nonparametrically. Because this method is quite flexible in terms of how it defines the severity of false positives and false negatives, it is possible to evaluate error rate definitions that are appropriate for a diverse set of finance applications.

**What are the Academic Insights?**

The paper proceeds with two practical applications of this framework: to select outperforming strategies (by studying two datasets, the Standard and Poor’s CAPIQ database on a broad set of 484 long-short alpha strategies, and the 18,113 anomalies studied in Yan and Zheng, 2017), and to analyze mutual fund performance (by focusing on the joint test approach used in Fama and French (2010), which treats the mutual fund population as a whole and tests whether the entire mutual fund population has zero alpha (the null hypothesis) versus at least one fund has a positive alpha).

The authors show that when the threshold t-statistic increases, the Type I error rate (the rate of false discoveries among all discoveries) declines while the Type
With regards to the first application, the authors find that neither 2.0 (the usual cutoff for 5% significance) nor 3.0 (based on Harvey, Liu, and Zhu (2016)) is optimal from the perspective of the investor. The preferred choices lie between 2.0 and 3.0 for the examples considered and depends on both p0 and the data that we study. This bootstrap-based framework generates t-statistic thresholds that are calibrated to the particular decision being considered and the particular data under analysis.

With regards to the second application, Fama and French's (2010) joint test suggests that very few (if any) funds exhibit skill on a net return basis. The authors confirm that the Fama and French (2010) approach performs well in terms of the Type I error rate for the mutual fund data. However, they find that the Type II error rates\(^1\) are very high. Even when 2% of funds are truly outperforming and are endowed with an average annualized alpha of 10.66%, there is still an 86.9% chance (at the 5% significance level) of the Fama and French (2010) approach falsely declaring a zero alpha for all funds.

Given the above finding, the authors explore whether there are indeed outperforming funds. Their conclusion is as follows:

> "There is some evidence for the existence of outperforming funds but consistent with the long literature in mutual fund evaluation, it is only modest evidence of fund outperformance."

**Why does it matter?**

Current research on multiple testing focuses on controlling the Type I error rate. This study shows that it is also important to consider the Type II error rate. For the selection of investment strategies, a weighted average of the Type I error rate and the Type II error rate is likely more consistent with the investor’s objective function. With the advent of big data and higher computing power, it is important to try and correct data mining as much as possible. This is an attempt made by the authors of the paper.

It's an intense paper but worth the read!

**The Most Important Chart from the Paper:**
The results are hypothetical results and are NOT an indicator of future results and do NOT represent returns that any investor actually attained. Indexes are unmanaged and do not reflect management or trading fees, and one cannot invest directly in an index.

Abstract

*Multiple testing plagues many important questions in finance such as fund and factor selection. We propose a new way to calibrate both Type I and Type II errors. Next, using a double-bootstrap method, we establish a t-statistic hurdle that is associated with a specific false discovery rate (e.g., 5%). We also establish a hurdle that is associated with a certain acceptable ratio of misses to false discoveries (Type II error scaled by Type I error), which effectively allows for differential costs of the two types of mistakes. Evaluating current methods, we find that they lack power to detect outperforming managers.*

Notes:

1. i.e., the probability of failing to reject the null hypothesis
About the Author: Elisabetta Basilico, PhD, CFA

Dr. Elisabetta Basilico is a seasoned investment professional with an expertise in "turning academic insights into investment strategies." Research is her life's work and by combing her scientific grounding in quantitative investment management with a pragmatic approach to business challenges, she's helped several institutional investors achieve stable returns from their global wealth portfolios. Her expertise spans from asset allocation to active quantitative investment strategies. Holder of the Charter Financial Analyst since 2007 and a PhD from the University of St. Gallen in Switzerland, she has experience in teaching and research at various international universities and co-author of articles published in peer-reviewed journals. She and co-author Tommi Johnsen published a book on research-backed investment ideas, titled Smart(e)r Investing. How Academic Insights Propel the Savvy Investor. You can find additional information at Academic Insights on Investing.

Related Posts

GET UPDATES NOW

FOLLOW US AND LEARN

Recent

DIY Asset Allocation

IMPORTANT DISCLOSURES

Performance figures contained herein...
The General is leading #MFTHarrisburg. Map is here: alphachitect.com/wp-content/upl...
Date is September 12th. NOT 26th like...
twitter.com/i/web/status/1...
4 hours ago

@TorieHappe oh boy. Good times!
3 hours ago

Effects of Portfolio Construction on the Perform of Style Factor ETFs
August 31st, 2020

How to Measure and Understand Portfolio Tail Risk Events
August 27th, 2020

Architect, LLC; hypothetical results are intended for illustrative purposes only. Past performance is not indicative of future results, which may vary. There is a risk of substantial loss associated with trading stocks, commodities, futures, options and other financial instruments. Full disclosures here.