JACOBS LEVY EQUITY MANAGEMENT CENTER
FOR QUANTITATIVE FINANCIAL RESEARCH

ILLEGITIMATE SCIENCE:
Why Most *Empirical* Discoveries in Finance Are Likely Wrong, and What Can Be Done About It

Marcos López de Prado
Lawrence Berkeley National Laboratory
Computational Research Division

Friday, May 1, 2015

Key Points

- The proliferation of false discoveries is a pressing issue in Financial research.
- For a large enough number of trials on a given dataset, it is guaranteed that a model specification will be found to deliver sufficiently low *p-values*, even if the dataset is random.
- Most academic papers and investment proposals do not report the number trials involved in an empirical discovery.
- The implication is that *most published empirical discoveries in Finance are likely false*.
- This has severe implications, specially with regards to:
  - Peer review process.
  - Backtesting of investment proposals.
- **Question: What constitutes a *legitimate* discovery in Finance?**

Decision under uncertainty

Suppose that we test whether a factor explains performance:

- Negative condition: The factor does not provide a *poset* of returns.
- Positive condition: The factor provides a *poset* of returns.

Total population	Condition positive	Condition negative	Prevalence = Σ Condition positive / Σ Total population	
Test outcome positive	**True positive**	**False positive** (Type I error)	Positive predictive value (PPV, Precision) = Σ True positive / Σ Test outcome positive	False discovery rate (FDR) = Σ False positive / Σ Test outcome positive
Test outcome negative	**False negative** (Type II error)	**True negative**	False omission rate (FOR) = Σ False negative / Σ Test outcome negative	Negative predictive value (NPV) = Σ True negative / Σ Test outcome negative

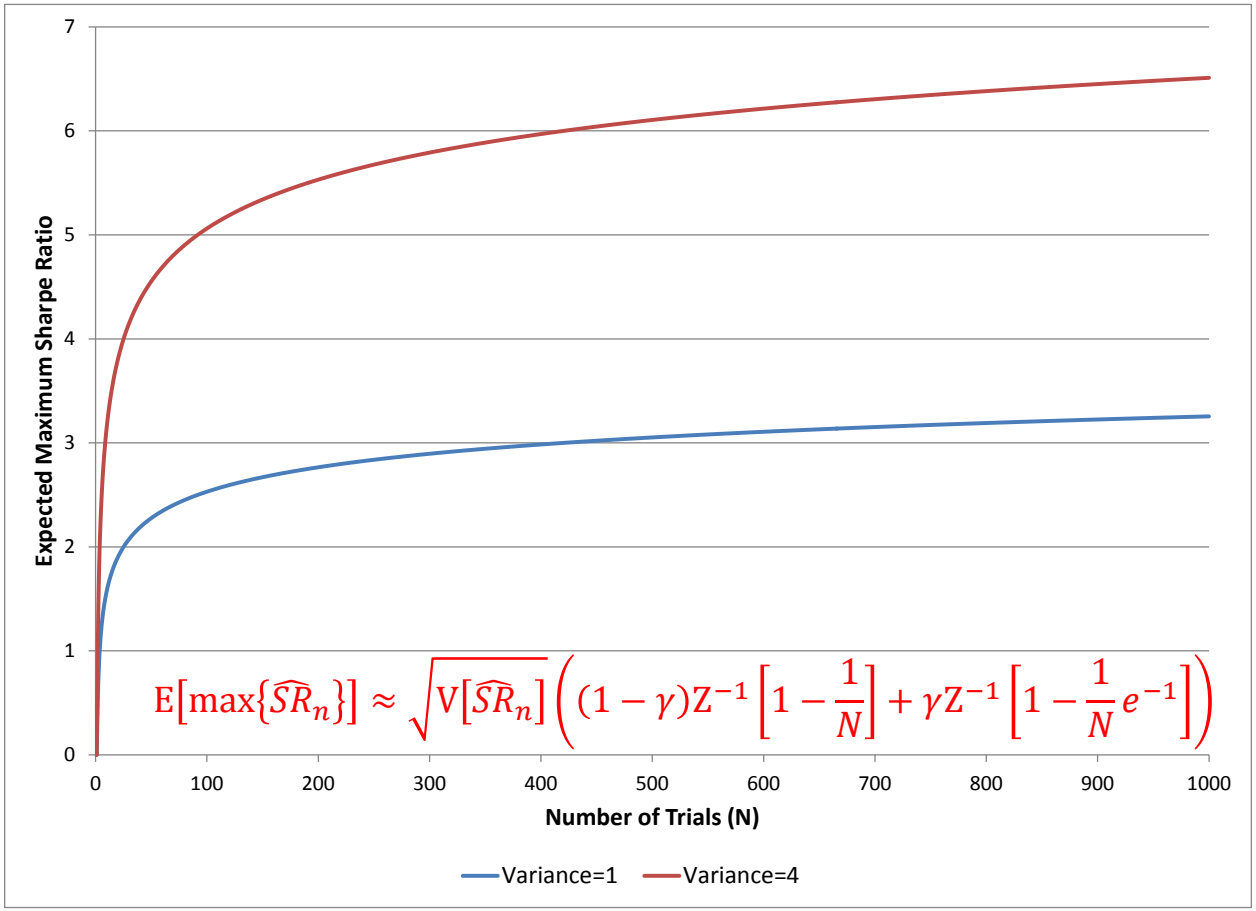
The Neyman-Pearson testing framework made sense 80 years ago, when data and calculations were costly… Not anymore.

Applying the test multiple times

- Suppose now that we are interested in analyzing multiple factors.
- A curious problem then emerges: As we test more and more factors on the same dataset, each at the same significance level α, **the overall probability of choosing at least one false factor grows**.
- This is called the **multiple testing problem**, and it is so pervasive and notorious that the American Statistical Society explicitly warns against it in its Ethical Guidelines (guideline #8):

> *"Running multiple tests on the same data set at the same stage of an analysis increases the chance of obtaining at least one invalid result. Selecting the one "significant" result from a multiplicity of parallel tests poses a grave risk of an incorrect conclusion.* ***Failure to disclose the full extent of tests and their results in such a case would be highly misleading****."*

Backtest Overfitting

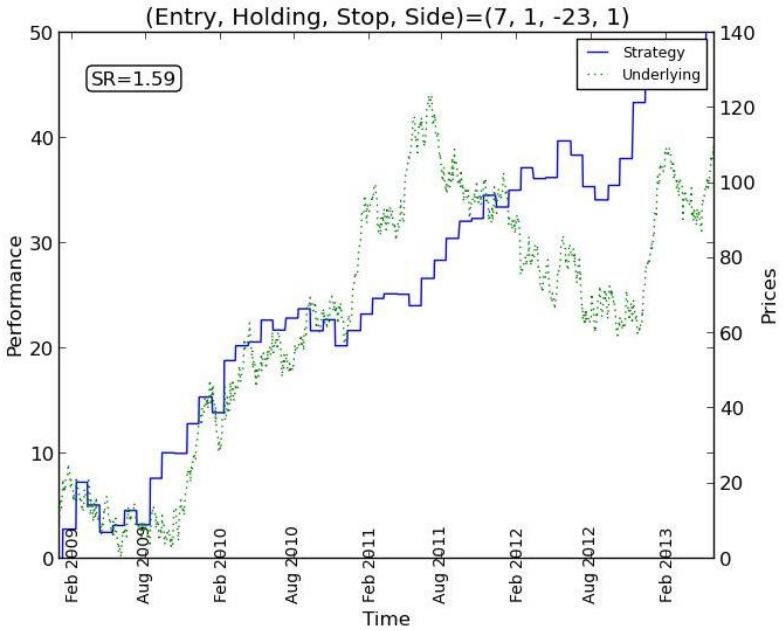


Expected Maximum Sharpe Ratio as the number of independent trials N grows, for $E[\widehat{SR}_n] = 0$ and $V[\widehat{SR}_n] \in \{1,4\}$.

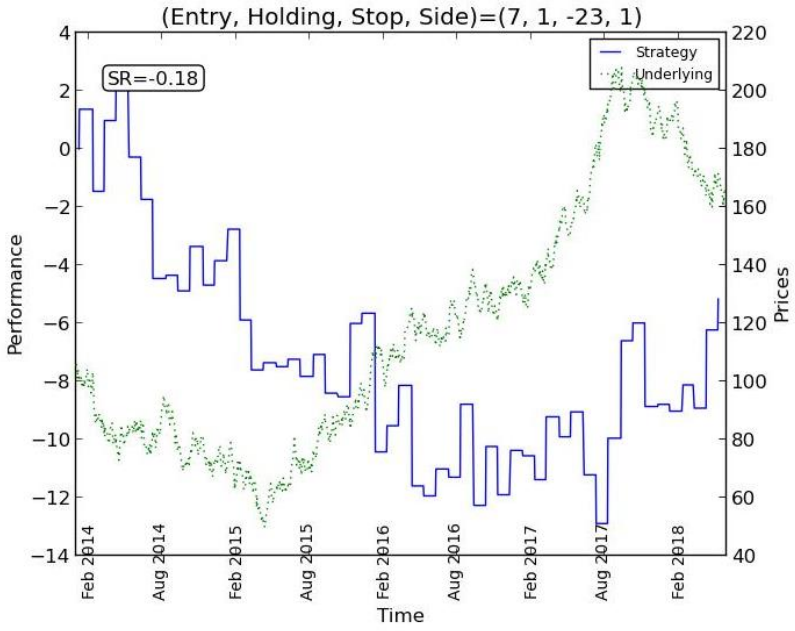
Searching for empirical findings regardless of their theoretical basis is likely to magnify the problem, as $\mathrm{V}[\widehat{SR}_n]$ will increase when unrestrained by theory.

This is a consequence of pure random behavior. We will observe better candidates *even if there is no investment skill associated with this strategy class* ($E[\widehat{SR}_n] = 0$).

The Backtest Overfitting Simulation Tool



An “optimized” investment strategy (in blue) making steady profit while the underlying trading instrument (in green) gyrates in price. It is trivial to make Financial “discoveries” if enough variations are tried.



The same investment strategy performs poorly on a different sample of the same trading instrument.

http://datagrid.lbl.gov/backtest/index.php

A Solution: The Deflated Sharpe Ratio

- The Deflated Sharpe Ratio (DSR) corrects the inflationary effect of multiple trials, non-normal returns and shorter sample lengths:

$$\widehat{DSR} \equiv \widehat{PSR}\left(\widehat{SR}_0\right) = Z\left[\frac{\left(\widehat{SR} - \widehat{SR}_0\right)\sqrt{T-1}}{\sqrt{1 - \hat{\gamma}_3\widehat{SR} + \frac{\hat{\gamma}_4 - 1}{4}\widehat{SR}^2}}\right]$$

where

$$\widehat{SR}_0 = \sqrt{V[\widehat{SR}_n]}\left((1-\gamma)Z^{-1}\left[1 - \frac{1}{N}\right] + \gamma Z^{-1}\left[1 - \frac{1}{N}e^{-1}\right]\right)$$

DSR is a Probabilistic Sharpe Ratio where the rejection threshold is adjusted to reflect the multiplicity of trials.

Some Myths on Prevention of False Discoveries

- **Myth #1**: *p-values* give the probability that a finding is the result of random chance.
  - FALSE: p-values are a statement about hypothetical study replications using imaginary "well-behaved" data.
- **Myth #2**: Holding-out part of the sample for cross-validation prevents false discoveries.
  - FALSE: Hold-out or cross-validation does not control for the number of trials, thus it is equally exposed to selection bias.
- **Myth #3**: Simpler models on longer series are more likely to be correct.
  - FALSE: The Backtest Overfitting Simulation Tool is an extremely simple model, and yet it will deliver a "winning" strategy on random series, of any length, every time.

Why is this a problem, particularly in Finance?

- Most other fields have access to laboratories, where experiments can be reproduced under controlled conditions.
  - Ever wondered why retraction rates are so low in Finance?
  - We will never know what caused the Flash Crash, because we cannot replay that day in absence of Mr. Sarao's spoofing.
- Competition among investment managers means that the ratio of signal-to-noise in financial series is low.
  - Investment professional arbitrage away most significant factors.
  - This increases the probability that academics "discover" a chance configuration, rather than an actual signal.

This "competition" argument is specific to Finance: Researchers from most other fields study immutable laws.

What can be done?

- *p-values* should no longer be the preeminent publication hurdle.
  - Be skeptic of Econometric analysis without strong theoretical basis.
  - We need a new widely accepted statistical method to replace the *p-values*. Harvey et al., Bailey et al. propose such frameworks.
- Journals need to control for the number of trials:
  - Researchers should declare their testing plan in advance.
  - Researchers should make all results available to referees.
  - Editors should require researchers to test their theory on a new dataset provided by the journal.
  - Journal must control for the probability of false positives *across publications*. In other words, at some point journals will cease to accept papers on a particular subject, from any author, until new unpolluted datasets become available.
- If Financial firms do not use a particular discovery, it is likely false.

THANKS FOR YOUR ATTENTION!

Bibliography

- Bailey, D. and M. López de Prado (2012): "The Sharpe Ratio Efficient Frontier," **Journal of Risk**, 15(2), pp. 3-44. Available at http://ssrn.com/abstract=1821643.
- Bailey, D., J. Borwein, M. López de Prado and J. Zhu (2014): "Pseudo-Mathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-Of-Sample Performance," **Notices of the American Mathematical Society**, 61(5), pp. 458-471. Available at http://ssrn.com/abstract=2308659.
- Bailey, D., J. Borwein, M. López de Prado and J. Zhu (2014): "The Probability of Backtest Overfitting." **Journal of Computational Finance**, forthcoming. Available at http://ssrn.com/abstract=2326253.
- Carr, P. and M. López de Prado (2014): "Determining Optimal Trading Rules Without Backtesting," working Paper. Available at http://arxiv.org/abs/1408.1159.
- Harvey, C. and Y. Liu (2014): "Lucky Factors," working paper. Available in SSRN: http://ssrn.com/abstract=2528780.
- Harvey, C. and Y. Liu (2014): "Backtesting," working paper. Available in SSRN: http://ssrn.com/abstract=2345489.
- Harvey, C., Y. Liu and H. Zhu (2015): "…And the Cross-Section of Expected Returns," working paper. Available in SSRN: http://ssrn.com/abstract=2249314.

Bio

Marcos López de Prado is Senior Managing Director at *Guggenheim Partners*. He is also a Research Fellow at *Lawrence Berkeley National Laboratory*'s Computational Research Division (U.S. Department of Energy's Office of Science), where he conducts unclassified research in the mathematics of large-scale financial problems and supercomputing.

Before that, Marcos was Head of Quantitative Trading & Research at Hess Energy Trading Company (the trading arm of *Hess Corporation*, a Fortune 100 company) and Head of Global Quantitative Research at *Tudor Investment Corporation*. In addition to his 17 years of trading and investment management experience at some of the largest corporations, he has received several academic appointments, including Postdoctoral Research Fellow of RCC at *Harvard University* and Visiting Scholar at *Cornell University*. Marcos earned a Ph.D. in Financial Economics (2003), a second Ph.D. in Mathematical Finance (2011) from *Complutense University*, is a recipient of the National Award for Excellence in Academic Performance by the Government of Spain (National Valedictorian, 1998) among other awards, and was admitted into *American Mensa* with a perfect test score.

Marcos serves on the Editorial Board of the *Journal of Portfolio Management* (IIJ) and the *Journal of Investment Strategies* (Risk). He has collaborated with ~30 leading academics, resulting in some of the most read papers in Finance (SSRN), four international patent applications on High Frequency Trading, three textbooks, numerous publications in the top Mathematical Finance journals, etc. Marcos has an Erdös #2 and an Einstein #4 according to the *American Mathematical Society*.

Disclaimer

- The views expressed in this document are the authors' and do not necessarily reflect those of the organizations he is affiliated with.
- No investment decision or particular course of action is recommended by this presentation.
-

Notice:

The research contained in this presentation is the result of a continuing collaboration with

Prof. David H. Bailey, Berkeley Lab
Prof. Jon Borwein, FRS-Canada, AAAS
Prof. Jim Zhu, WMU

For additional details, please visit:
http://ssrn.com/abstract=2308659
www.QuantResearch.org