

# Scientific Integrity in Mathematical Finance

David H. Bailey

<http://www.davidhbailey.com>

Lawrence Berkeley National Lab (retired) and Univ. of California, Davis

11 July 2014

## The dangers of reading too much into data

*Another thing I must point out is that you cannot prove a vague theory wrong. [...] Also, if the process of computing the consequences is indefinite, then with a little skill any experimental result can be made to look like the expected consequences. [Richard Feynman, 1964]*

*I remember my friend Johnny von Neumann used to say, with four parameters I can fit an elephant, and with five I can make him wiggle his trunk. [Enrico Fermi]*

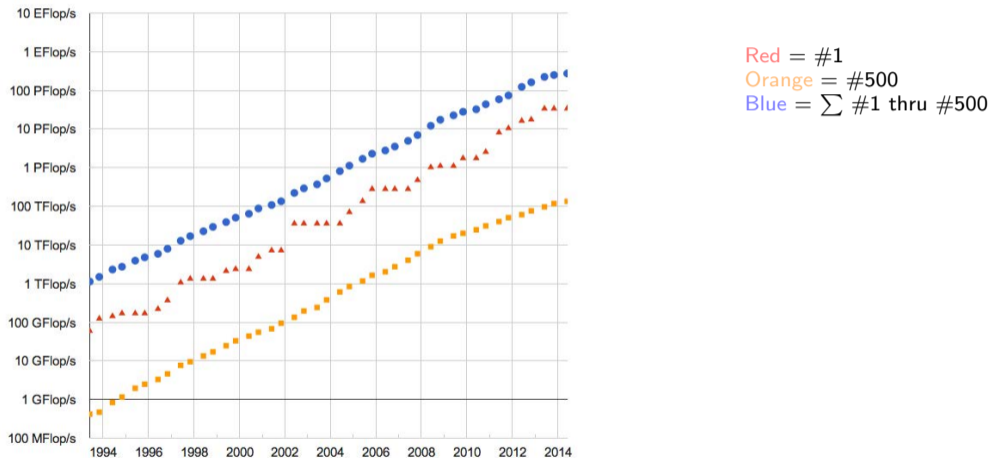
# High-performance computing meets mathematical finance

The advent of highly parallel scientific computing is revolutionizing the world finance, just as it has almost every other field of modern science and engineering.

Researchers and analysts can:

- ▶ Access and analyze petabytes of past stock and bond market data.
- ▶ Analyze hundreds, thousands or millions of different strategies or portfolio options.
- ▶ Apply extremely sophisticated mathematical algorithms to investment and finance data — software is widely available, albeit at a fee.
- ▶ Implement trading algorithms that can perform thousands of transactions on multiple trading venues, in a flash.

# Increasing performance of the top 500 supercomputers (1994 – present)



► “Performance development,” Top500.org, available at <http://top500.org/statistics/perfdevel>.

# DANGER AHEAD

In spite of the successes of computing in the finance field, danger lies ahead, just as in every other field of modern science:

*Supercomputers can generate nonsense faster than ever before!*

Key concerns:

- ▶ Are the algorithms, data sources and processing methods well documented?
- ▶ Are the results reproducible by other researchers, or even by the same team of researchers?
- ▶ Are the results statistically sound?
- ▶ Are the results numerically reliable?
- ▶ Have the results been validated using tests designed by the researchers or others?

## Reproducibility in scientific computing

A December 2012 workshop on reproducibility in computing, held at Brown University in Rhode Island, USA, noted that

*Science is built upon the foundations of theory and experiment validated and improved through open, transparent communication. With the increasingly central role of computation in scientific discovery, this means communicating all details of the computations needed for others to replicate the experiment. ... The “reproducible research” movement recognizes that traditional scientific research and publication practices now fall short of this ideal, and encourages all those involved in the production of computational science ... to facilitate and practice really reproducible research.*

- ▶ V. Stodden, D. H. Bailey, J. Borwein, R. J. LeVeque, W. Rider and W. Stein, “Setting the default to reproducible: Reproducibility in computational and experimental mathematics,” <http://www.davidhbailey.com/dhbpapers/icerm-report.pdf>.

## Reproducibility in scientific computing, continued

Issues identified in the ICERM report and other studies include:

- ▶ The need to carefully document the full context of computational experiments—system environment, input data, code used, computed results, etc.
- ▶ The need to save the code and output data in a permanent repository.
- ▶ The need for reviewers, research institutions and funding agencies to recognize the importance of computing and computing professionals, and to allocate funding for after-the-grant support and repositories.
- ▶ The increasing importance of numerical reproducibility, and the need for tools to ensure and enhance numerical reliability.
- ▶ The need to encourage publication of negative results—other researchers can often learn from them.
- ▶ The re-emergence of the need to ensure responsible reporting of performance.

## Reproducibility in biomedicine

The biomedical field has been stung by numerous cases where pharma products look good based on clinical trials, but later disappoint in real-world usage, or the results cannot be reproduced in separate studies. Examples:

- ▶ In 2004, GlaxoSmithKline acknowledged that while some trials of Paxil found it effective for depression in children, other unpublished studies showed no benefit.
- ▶ In 2011, Bayer researchers reported that they were able to reproduce the results of only 17 of 67 published studies they examined.
- ▶ In 2012, Amgen researchers reported that they were able to reproduce the results of only 6 of 53 published cancer studies.
- ▶ In 2014, a review of Tamiflu found that while it made flu symptoms disappear a bit sooner, it did not stop serious complications or keep people out of the hospital.

These experiences have exposed a fundamental flaw in methodology:

**Only publicizing the results of successful trials introduces a bias into the results.**

The AllTrials movement would require all results to be public: <http://www.alltrials.net>

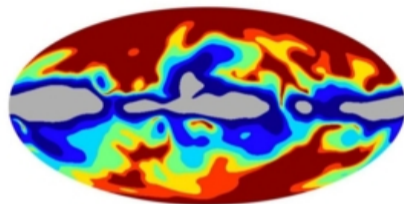


## Reproducibility in physics

In March 2014, a team of researchers from Harvard University made the dramatic announcement that they had discovered an interesting “twisting” pattern in cosmic microwave background data, measured using their BICEP2 experimental system.

This pattern fit very well with the hypothesized pattern of the most commonly assumed model of “inflation” in the first tiny fraction of a second after the big bang, and thus has been trumpeted as the first experimental evidence of the inflationary cosmology.

But other researchers had difficulty reconstructing the claimed results. Finally, two teams challenged the BICEP2 findings, saying that the results could more readily be explained by dust in the Milky Way.



- ▶ Ron Cowen, “Doubt grows about gravitational waves detection,” *Scientific American*, 2 Jun 2014.

## Reproducibility in social science

The “blank slate” paradigm (1920–1990):

- ▶ The human mind at birth is a *tabula rasa* (“blank slate”).
- ▶ Heredity and biology play no significant role in human psychology; all personality and behavioral traits are socially constructed.
- ▶ Pre-modern societies were peaceful, devoid of psychological and social problems.

Current consensus, based on latest research:

- ▶ Humans at birth possess sophisticated facilities for social interaction, language acquisition, pattern recognition, navigation and counting.
- ▶ Heredity, evolution and biology are major factors in human personality.
- ▶ Some personality traits are more than 50% heritable.
- ▶ Pre-modern societies had more crime, war and social problems than today.

How did the 20th century social scientists get it so wrong?

- ▶ Sloppy experimental methodology and analysis.
- ▶ Pervasive wishful thinking and politically correct biases.
- ▶ Ignoring or dismissing data that runs counter to predisposition.

## Email from DHB to a financial colleague, 10 June 2013

One thing that has always puzzled me about financial news and commentary is the following sort of thing. Look at these two examples:

This fellow says that “we are heading over 1700 before the summer swoon”, and makes number of other rather specific predictions: [URL given]. Or this fellow “tells you where he thinks the stock market is headed, and why” [URL given]. ... Excuse me for being “dumb”, but this sort of thing seems to me to be outright nonsense — callous attempts to fleece money out of credulous, ignorant investors.

After all, the stock market, by definition, contains the consensus of all available information, including the tens of thousands of stock market analysts and economists worldwide who scour every morel of information in the business world, and then advise the leading mutual funds and pension funds. Thus the big players have already made their moves long before individual investors hear anything. In addition, as is amply documented even in semi-popular books such as *Dark Pools*, there are thousands more very bright mathematicians using program-trading schemes, plying every trick of time series analysis, machine learning, stealth and anti-stealth that money can buy, to wriggle every conceivable angle out of the market and beat their competitors to the punch with trades.

## Email from DHB to a financial colleague, continued

How can amateurs possibly compete with these outfits? Obviously, they can't — the only rational strategy for an amateur individual investor is simply to regularly buy, say via a 401K program, into one or a handful of broad-market index funds, or, with the help of a qualified professional, buy and hold a balanced, diversified portfolio.

So when people like those above assure their audiences that they “know” the stock market is heading up, or down, or up for a month or two, then down, or that by following their strategies, John Q Public can enjoy reliable, above-market returns, this cannot have any scientific basis. After all, if these pundits really had a substantive, scientifically defensible basis for making such predictions, they wouldn't be selling their advice online, but instead they would be making millions advising institutional clients with many billions of dollars.

So why doesn't somebody blow this whistle on this sort of thing? Am I missing something?

## Email from financial colleague to DHB

It is not a dumb question at all. It is a question I have struggled with and which answer makes me an unhappy man. The truth is, most people in this industry are charlatans. They do not have any particular model or theory to understand the world. They are not scientists. They are market wizards. Some of them made a lot of money and therefore claim to have magic powers. But this is a zero sum game, someone has to make a lot of money, as a matter of probability distribution, not magic.

I completely agree with your assessment. The amount of nonsense in the airways is incredible. The good news is, the quants are silently taking over Wall Street, thanks to high frequency and big data. For the same reason that alchemists and astrologers fought the chemists and astronomers, the market wizards are fighting the quants. So all this media nonsense is in part the tug of that war. An attempt of the wizards to squeeze out a few more dimes.

## Reproducibility in finance

Finance, like the pharmaceutical world, has been stung with numerous instances of investment strategies that look great on paper, but fall flat in practice. A primary cause is *statistical overfitting of backtest (historical market) data*.

When a computer can analyze thousands or millions of variations of a given strategy, it is almost certain that the best such strategy, measured by backtests, will be overfit and thus of dubious value.

## Paper on “Pseudo-mathematics and financial charlatanism”

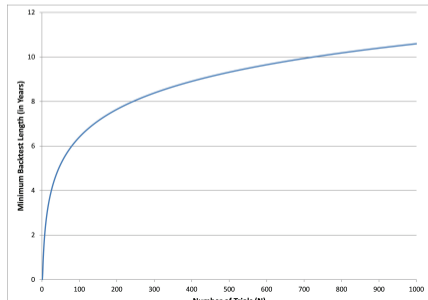
In two 2014 papers by DHB, JMB, Marcos Lopez de Prado and Jim Zhu, we derive (a) a formula relating the number of trials to the minimum backtest length, and (b) a formula for the probability of backtest overfitting. We also show that under the assumption of memory in markets, overfit strategies are actually prone to *lose* money.

- ▶ D. H. Bailey, J. M. Borwein, M. Lopez de Prado and Q. J. Zhu, “Pseudo-mathematics and financial charlatanism: The effects of backtest over fitting on out-of-sample performance,” *Notices of the American Mathematical Society*, May 2014, pg. 458–471.
- ▶ D. H. Bailey, J. M. Borwein, M. Lopez de Prado and Q. J. Zhu, “The probability of backtest overfitting,” manuscript, 12 Feb 2014, <http://ssrn.com/abstract=2326253>.

## How easy is it to overfit a backtest? Answer: Very easy!

- ▶ If only 2 years of daily backtest data are available, then no more than 7 strategy variations should be tried.
- ▶ If only 5 years of daily backtest data are available, then no more than 45 strategy variations should be tried.
- ▶ *A backtest that does not report the number of trials  $N$  makes it impossible to assess the risk of overfitting.*
- ▶ Given any desired performance level, a financial researcher just needs to keep trying alternative parameters for that strategy!

$$\text{MinBTL} \approx \left( \frac{(1 - \gamma)Z^{-1} \left[1 - \frac{1}{N}\right] + \gamma Z^{-1} \left[1 - \frac{1}{N}e^{-1}\right]}{E[\max_N]} \right)^2$$





## An absurd investment strategy

- ▶ A financial advisor sends letters to  $10,240 = 10 \times 2^{10}$  potential clients, with 5120 letters predicting a certain security will go up, and the other half predicting it will go down.
- ▶ One month later, the advisor sends letters only to the 5120 investors who were previously sent the correct prediction, with 2560 letters predicting a certain security will go up, and the other half predicting it will go down.
- ▶ The advisor continues this process for 10 months.
- ▶ The remaining ten investors, so impressed by the advisor's ten consecutive spot-on predictions, will entrust to him/her all of their assets!

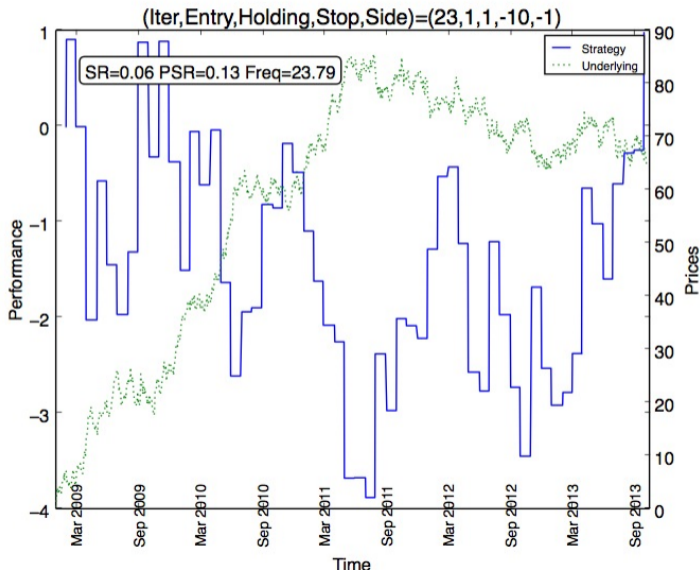
This strategy is absurd; even fraudulent.

But why is marketing a statistically overfit strategy, where potential investors are not informed of the number of trials behind the strategy, any different?

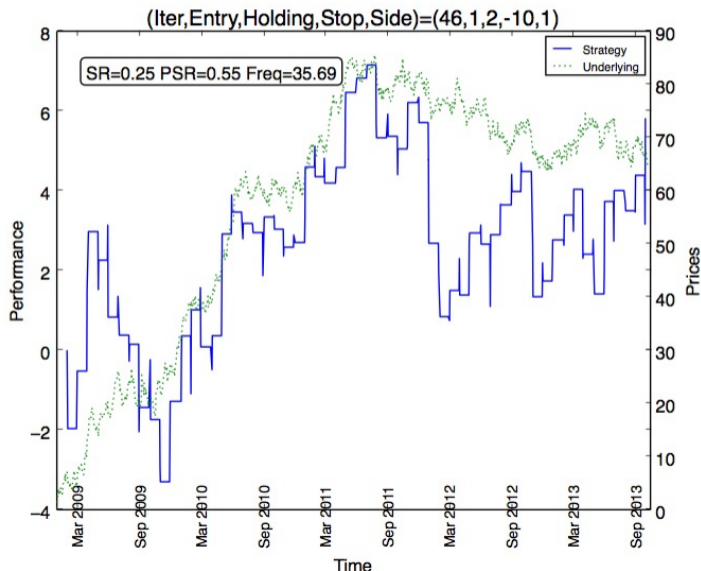
## Optimizing an investment strategy to fit pseudorandom time series

- ▶ The following 23 viewgraphs present the results of different steps in an attempt to find an “optimal” investment strategy, based on a fixed market price dataset.
- ▶ The underlying dataset was generated by a pseudorandom number generator!
- ▶ As you can see, by tweaking some very basic parameters (entry price, sell price, stop-loss, etc), we can fit and “predict” the underlying dataset quite well.
- ▶ The final (24th) viewgraph presents the results of implementing the resulting strategy on a continuation of the underlying (pseudorandom) dataset.

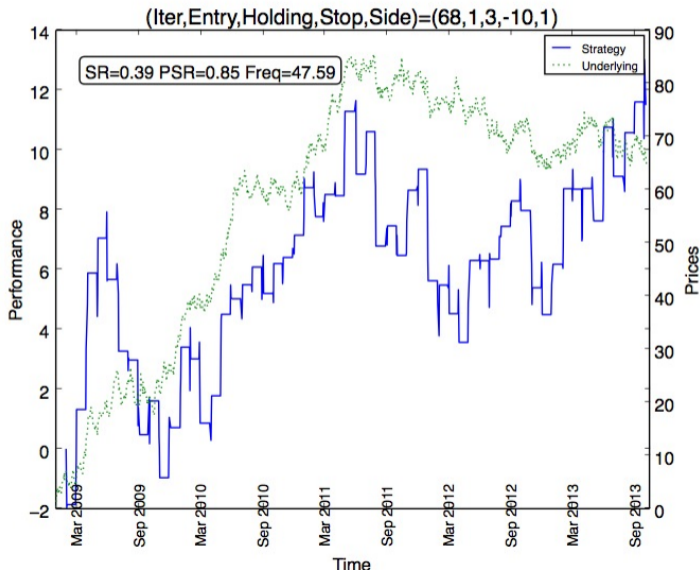
# Optimizing an investment strategy to fit pseudorandom time series, pg. 01



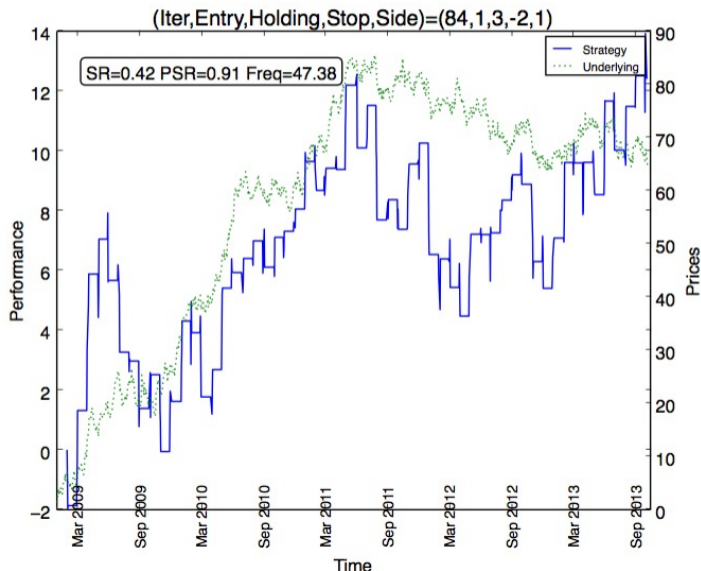
# Optimizing an investment strategy to fit pseudorandom time series, pg. 02



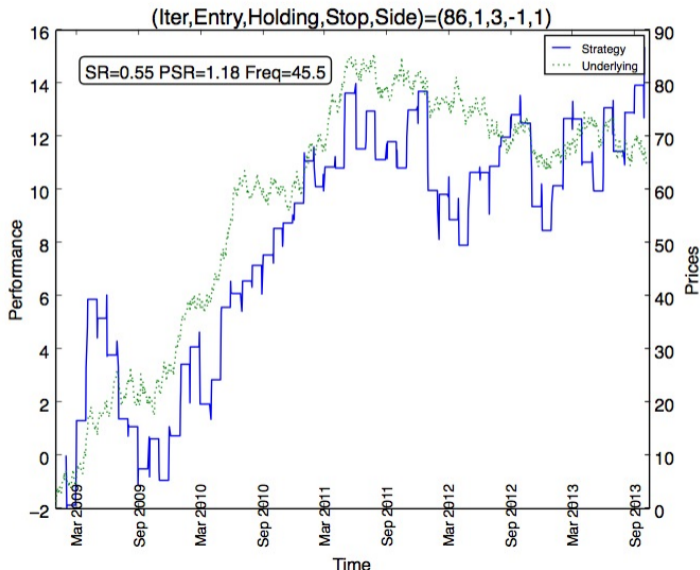
# Optimizing an investment strategy to fit pseudorandom time series, pg. 03



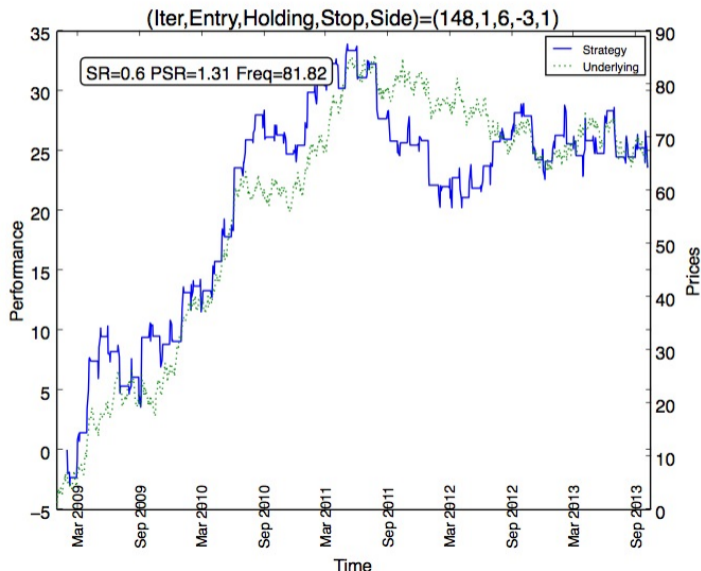
# Optimizing an investment strategy to fit pseudorandom time series, pg. 04



# Optimizing an investment strategy to fit pseudorandom time series, pg. 05

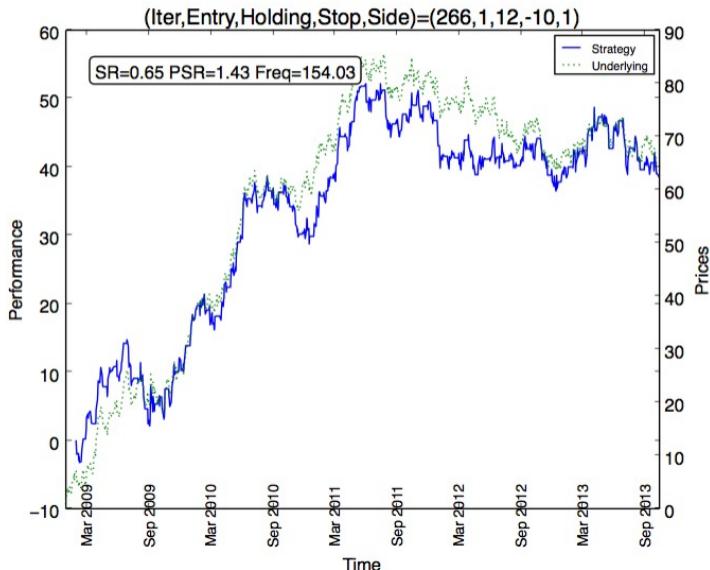


# Optimizing an investment strategy to fit pseudorandom time series, pg. 06

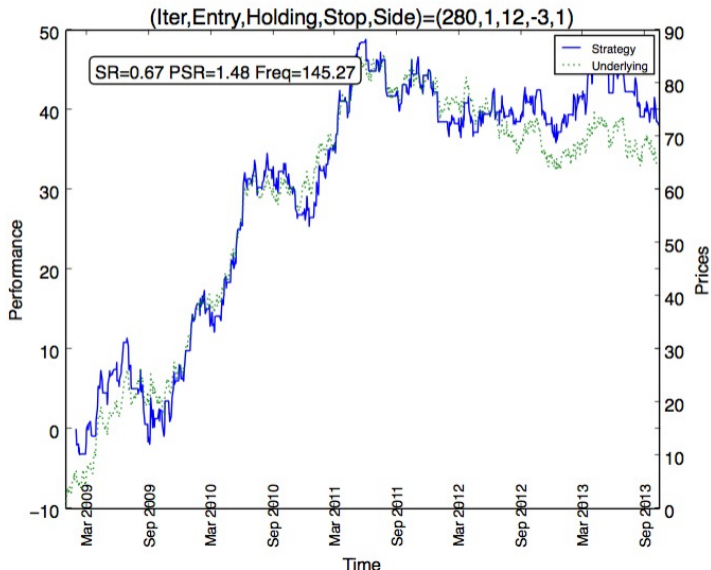




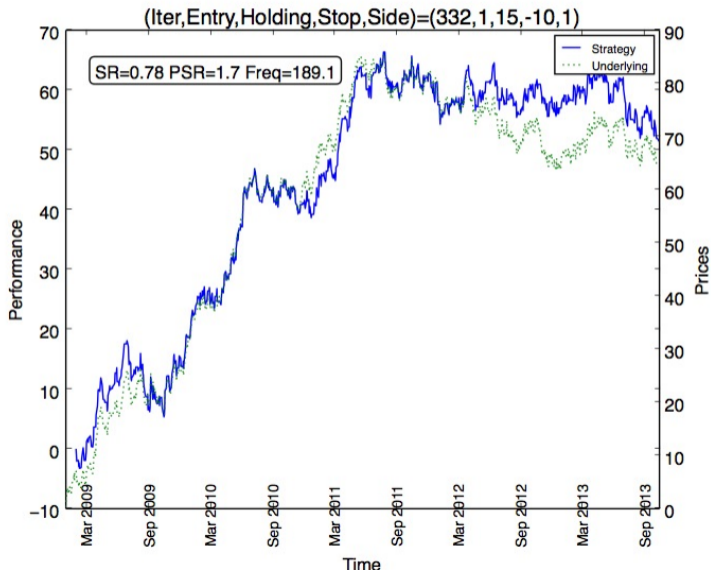
# Optimizing an investment strategy to fit pseudorandom time series, pg. 07



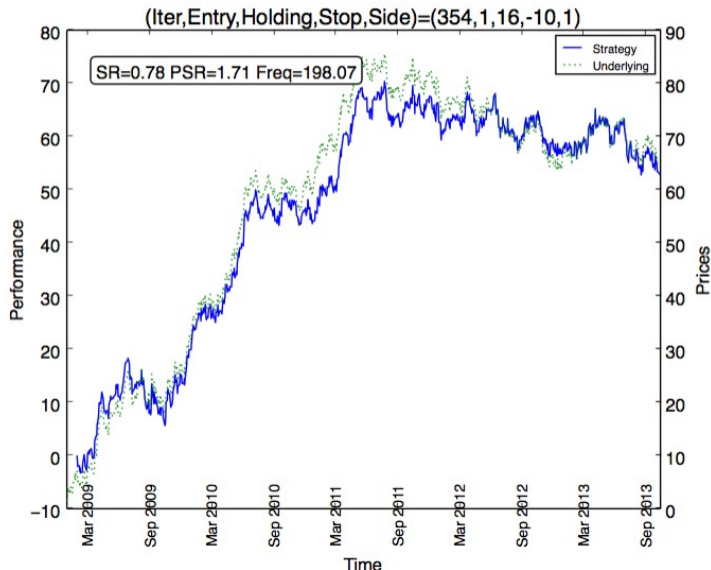
# Optimizing an investment strategy to fit pseudorandom time series, pg. 08



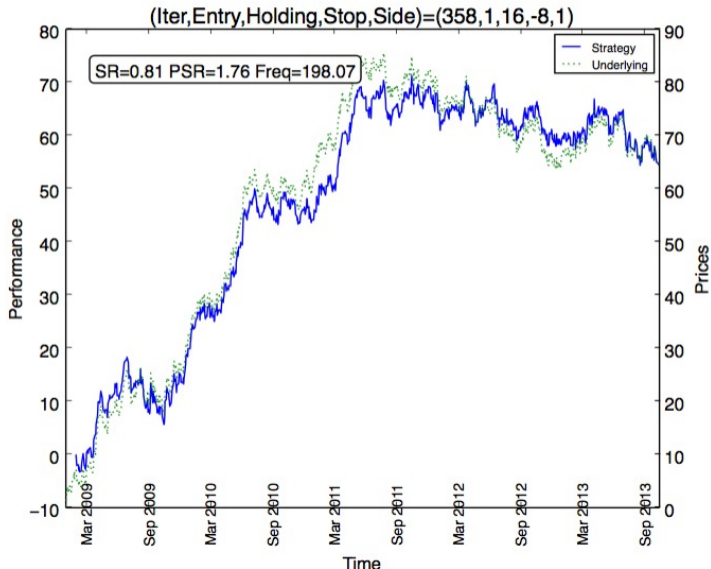
# Optimizing an investment strategy to fit pseudorandom time series, pg. 09



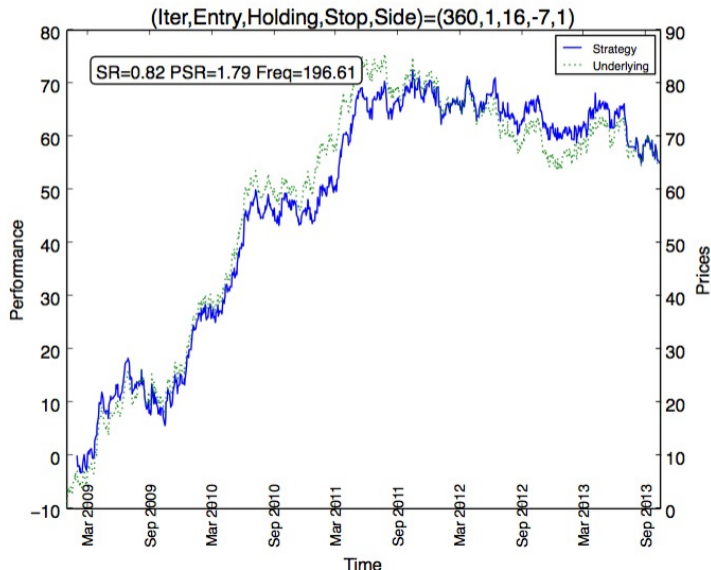
# Optimizing an investment strategy to fit pseudorandom time series, pg. 10



# Optimizing an investment strategy to fit pseudorandom time series, pg. 11



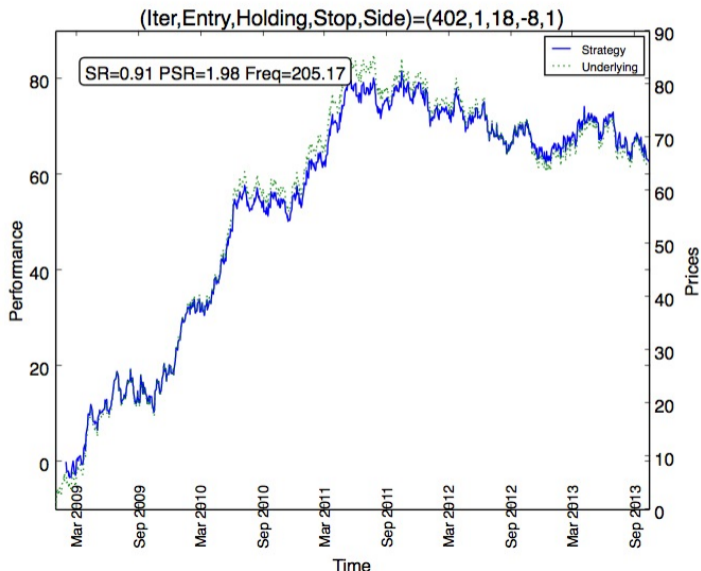
# Optimizing an investment strategy to fit pseudorandom time series, pg. 12



# Optimizing an investment strategy to fit pseudorandom time series, pg. 13

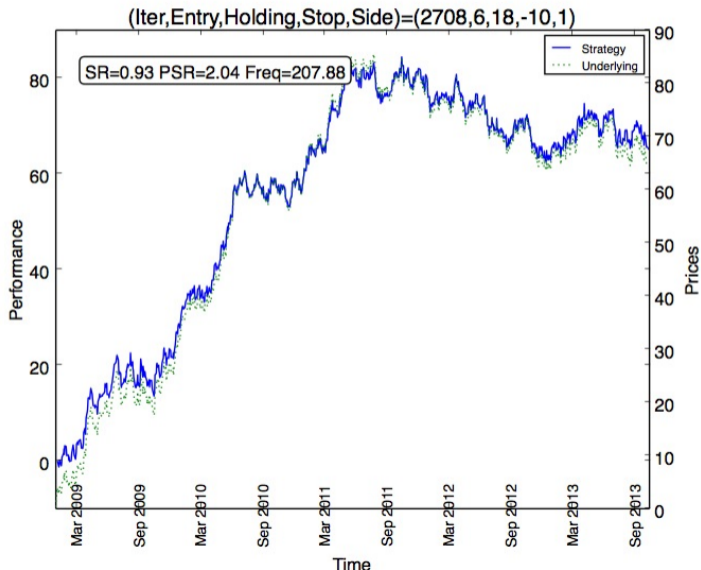


# Optimizing an investment strategy to fit pseudorandom time series, pg. 14

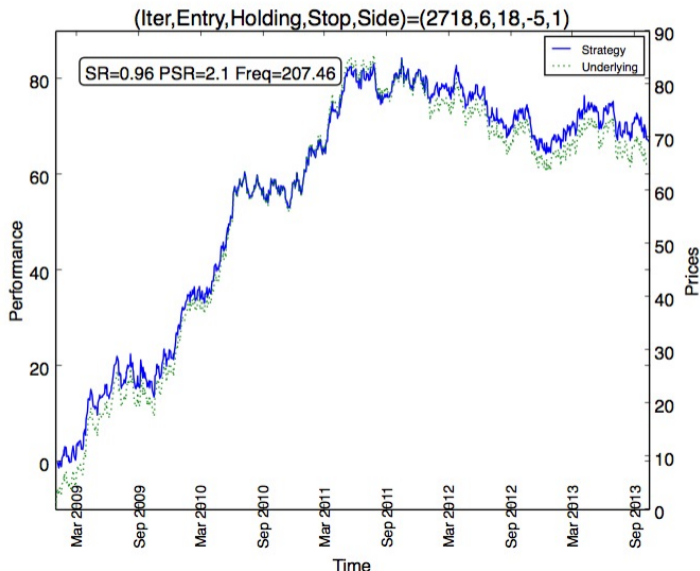




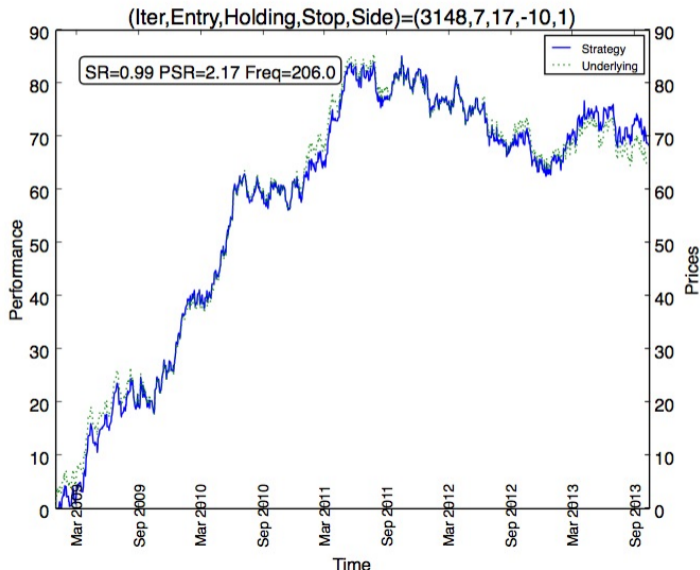
# Optimizing an investment strategy to fit pseudorandom time series, pg. 15



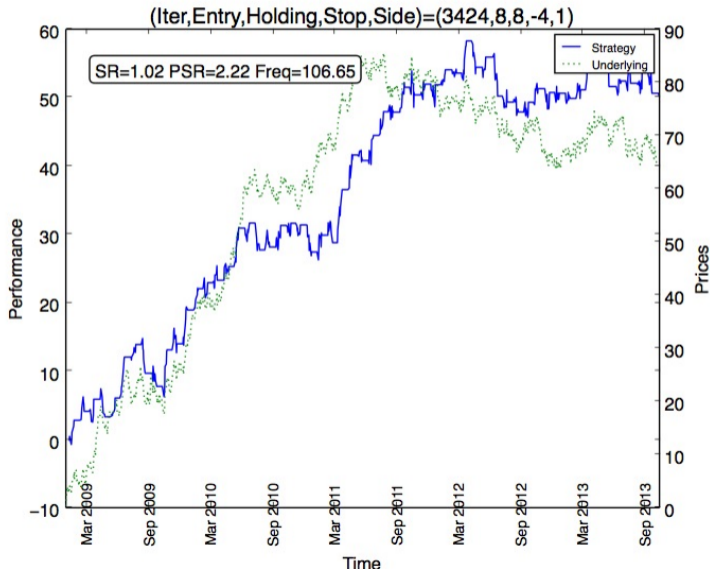
# Optimizing an investment strategy to fit pseudorandom time series, pg. 16



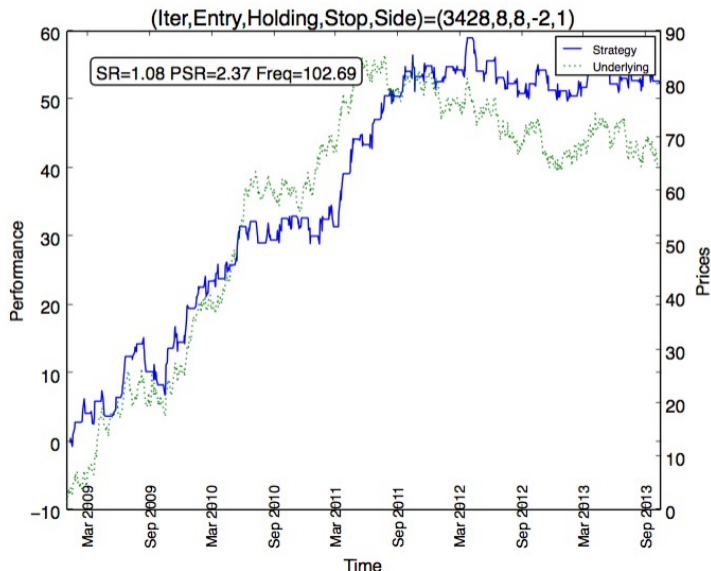
# Optimizing an investment strategy to fit pseudorandom time series, pg. 17



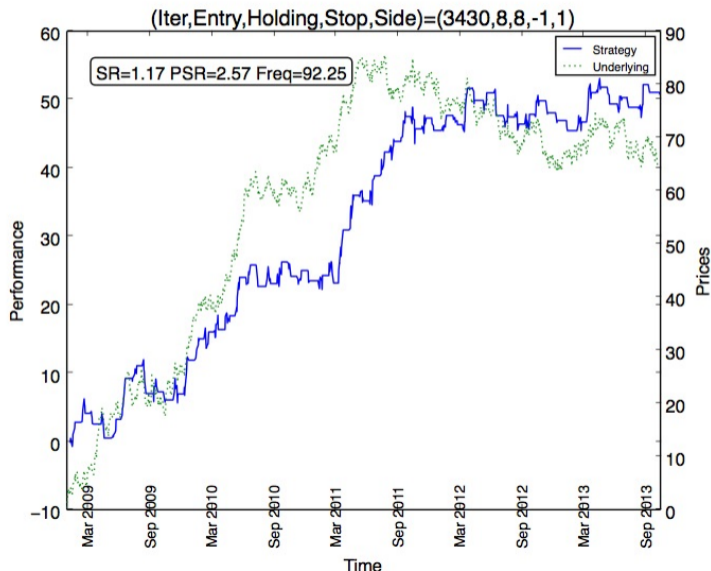
# Optimizing an investment strategy to fit pseudorandom time series, pg. 18



# Optimizing an investment strategy to fit pseudorandom time series, pg. 19



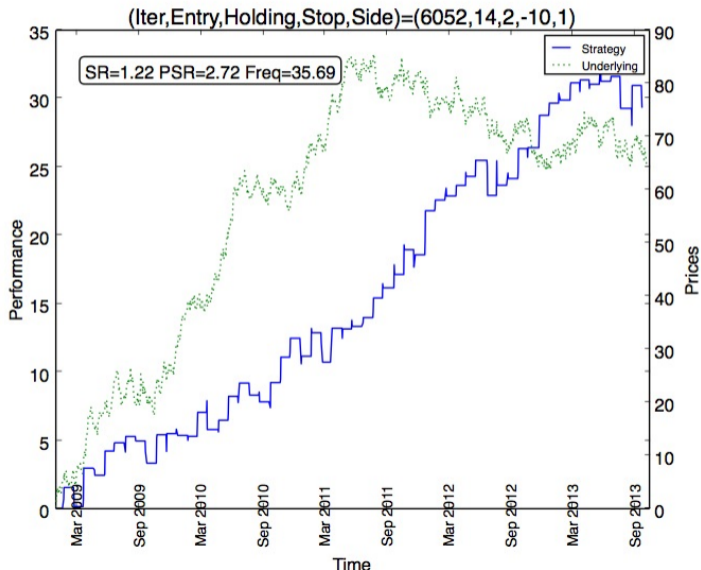
# Optimizing an investment strategy to fit pseudorandom time series, pg. 20



# Optimizing an investment strategy to fit pseudorandom time series, pg. 21

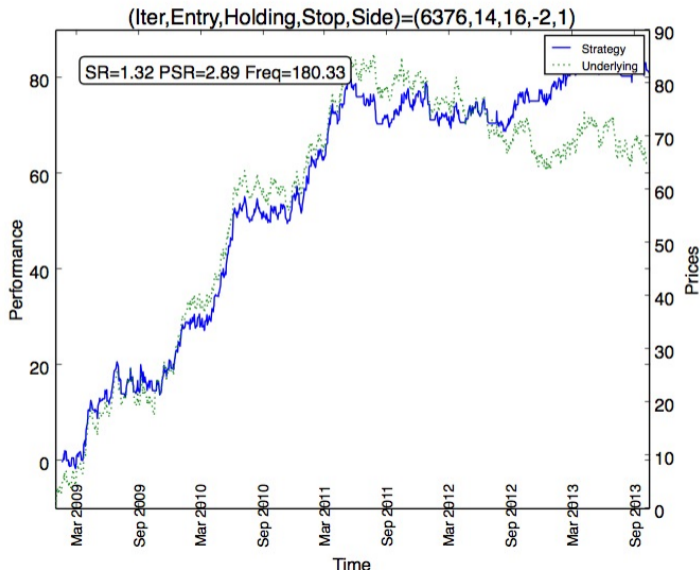


# Optimizing an investment strategy to fit pseudorandom time series, pg. 22

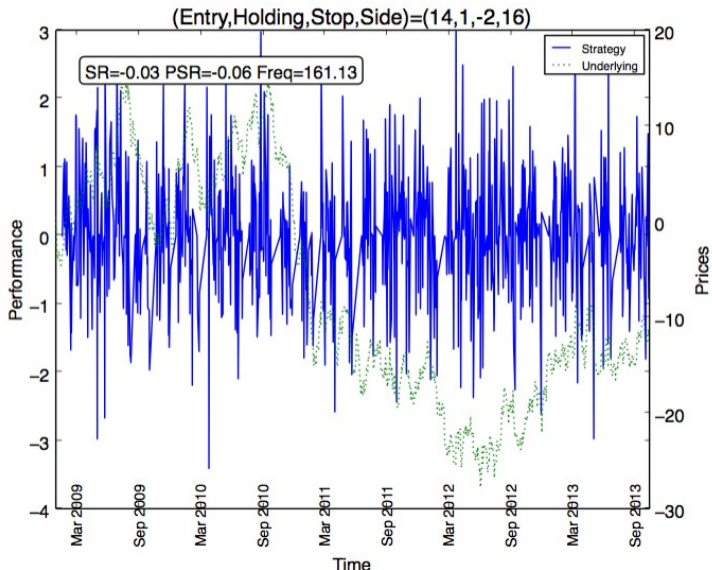




# Optimizing an investment strategy to fit pseudorandom time series, pg. 23



# Deploying the resulting strategy on a continuation of the time series



## Why the silence in the mathematical finance community?

- ▶ Historically scientists have led the way in exposing those who utilize pseudoscience to extract a commercial benefit: e.g., in the 18th century, physicists exposed the nonsense of astrologers.
- ▶ Yet financial mathematicians in the 21st century have remained disappointingly silent with the regards to those in the community who, knowingly or not:
  1. Fail to disclose the number of models or variations that were used to develop an investment strategy.
  2. Make vague predictions that do not permit rigorous testing and falsification.
  3. Misuse probability theory, statistics and stochastic calculus.
  4. Use dubious technical jargon: “stochastic oscillators,” “Fibonacci ratios,” “cycles,” “Elliot wave,” “Golden ratio,” “parabolic SAR,” “pivot point,” “momentum,” etc.

As we recently wrote in our paper “Pseudo-Mathematics and Financial Charlatanism”:  
**“Our silence is consent, making us accomplices in these abuses.”**

- ▶ D. H. Bailey, J. M. Borwein, M. Lopez de Prado and Q. J. Zhu, “Pseudo-mathematics and financial charlatanism: The effects of backtest over fitting on out-of-sample performance,” *Notices of the American Mathematical Society*, May 2014, pg. 458–471.

## Summary

- ▶ Super-powerful, highly parallel computer systems, and correspondingly powerful software, are revolutionizing the field of finance.
- ▶ But dangers lie ahead, particularly in statistical reliability.
- ▶ In mathematical finance, almost any desired performance can be achieved by massaging a model long enough.
- ▶ Backtest overfitting (the statistical overfitting of historical market data) is rampant in the financial world, an unfortunate by-product of powerful new computer systems and software.
- ▶ A financial program that does not disclose the number of variations of the strategy that have been explored to produce the optimal strategy is almost certainly overfit and of no predictive value.
- ▶ There are numerous other abuses, including making vague, untestable predictions, and using impressive-sounding pseudoscientific jargon.

**Why the silence? Why do not more in the finance field speak out?**