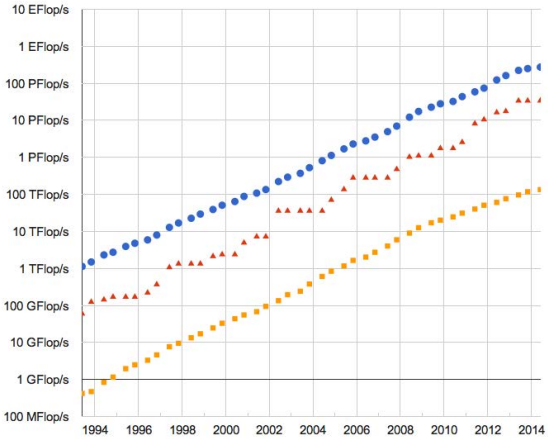
Financial Mathematics and Big Data Computing: Science and Pseudoscience

David H. Bailey
http://www.davidhbailey.com
Lawrence Berkeley National Lab (retired) and Univ. of California, Davis

Colleagues: Jonathan M. Borwein (Univ. of Newcastle, Australia), Marcos Lopez de Prado (Guggenheim), Jim Q. Zhu (Univ. of Western Michigan), Stephanie Ger (Harvard), Alexander Sim (LBNL), Victoria Stodden (Univ. of Illinois), Kesheng Wu (LBNL) and others.

March 25, 2015

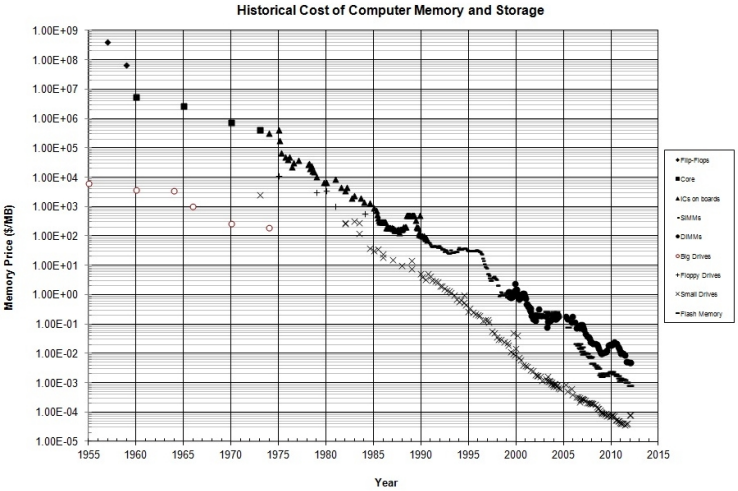
Increasing performance of the top 500 supercomputers (1994 – present)



Red = #1
Orange = #500
Blue = Sum of #1 thru #500

- "Performance development," Top500.org, available at `http://top500.org/statistics/perfdevel`.

Declining cost of data storage (1955 – 2012)

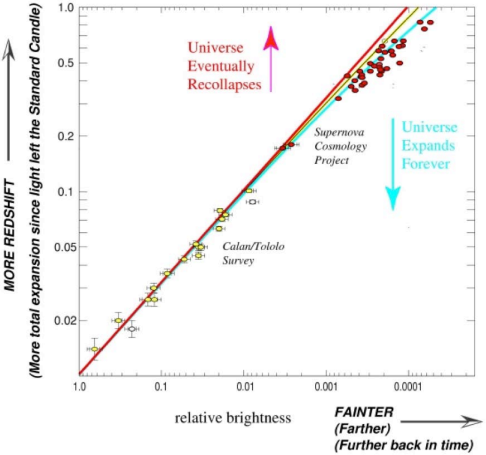


- John C. McCallum, "Disk drive storage price decreasing with time (1955-2012)," available at `http://www.jcmit.com/disk2012.htm`.

Big data application: The accelerating universe

In 1998, two teams of astronomers (one led by Saul Perlmutter of LBNL, and the other led by Brian P. Schmidt of ANU), came to the paradoxical conclusion that the expansion of the universe is accelerating, not slowing down.

Both teams based their results on careful measurements of distant supernovas, which in turn were found by sifting through reams of digital telescope data. The U.S. team in particular relied heavily on large computer systems at the NERSC facility, coupled with a worldwide network of collaborating astronomers.



Implications of big data computing

1. Scientific computing is in the midst of a paradigm shift, from data-poor to data-rich computing — many fields are now *drowning* in data.
2. Nowadays most data being processed on supercomputers is *experimental*, not simulation data.
3. Advanced visualization and analytical facilities are no longer optional — they are absolutely essential.
4. Sophisticated statistical machine learning techniques are now being applied to classify data and to focus on "interesting" items in the data.

DANGER AHEAD

In spite of the successes of big data computing, danger lies ahead:
Supercomputers operating on big data 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 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`

Email from DHB to a financial colleague, 10 June 2013

One thing that has always puzzled me about [the financial world] is the following sort of thing. Look at these two examples: [examples cited from recent news]. 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. ... In addition, ... 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. ...

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. ...

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

Email from financial colleague to DHB, 17 June 2013

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.

Backtest overfitting in mathematical 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, primarily due to *backtest overfitting*.
- In mathematical finance, a **backtest** is the usage of historical market data (e.g., the past ten years of daily S&P 500 or FTSE 100 closing averages) to assess the performance of a proposed trading strategy.
- **Backtest overfitting** (i.e., statistical overfitting of backtest data) means either proposing a model for a dataset that inherently possesses a higher level of complexity than the data, or else trying many variations of a model on that data, and then only presenting results from the variation that best fits the data.
- When a computer can analyze thousands, millions or even billions 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.
- Backtest overfitting appears to be rather widespread in the field, not just in the U.S. but elsewhere as well.

John von Neumann on the danger of statistical overfitting

Enrico Fermi recalled John von Neumann's warning:

> *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.*

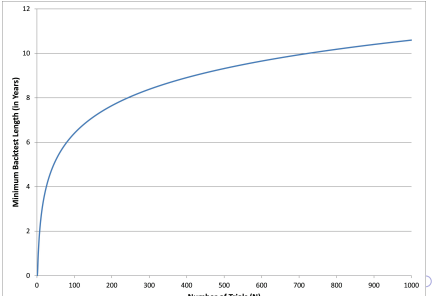
New papers on backtest overfitting

- **Presents formulas relating size of dataset to likelihood of backtest overfitting:**
  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, `http://www.ams.org/notices/201405/rnoti-p458.pdf`.
- **Presents formulas for calculating the probability of backtest overfitting:**
  D. H. Bailey, J. M. Borwein, M. Lopez de Prado and Q. J. Zhu, "The probability of backtest overfitting," *Journal of Computational Finance*, to appear, 27 Feb 2015, `http://www.davidhbailey.com/dhbpapers/backtest-prob.pdf`.
- **Introduces backtest overfitting for a general audience:**
  D. H. Bailey, S. Ger, M. Lopez de Prado, A. Sim and K. Wu, "Statistical overfitting and backtest performance," manuscript, 07 Oct 2014, `http://ssrn.com/abstract=2507040`.
- **Defines a "deflated Sharpe ratio," correcting for some forms of distortion:**
  David H. Bailey and Marcos Lopez de Prado, "The deflated Sharpe ratio: Correcting for selection bias, backtest overfitting and non-normality," *Journal of Portfolio Management*, to appear, 31 Jul 2014, `http://ssrn.com/abstract=2460551`.

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!

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



An absurd investment strategy

- A financial advisor sends letters to $10,240 = 10 \times 2^{10}$ potential clients, with half (5120) 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 half (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.
- Ten investors will have been sent ten consecutive correct predictions! They may be so impressed by the advisor's ten consecutive spot-on predictions that will entrust to him/her all of their assets...

This strategy is absurd, even fraudulent, because the final ten investors have not been told of the thousands of others letters sent with different predictions.

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

A not-so-absurd investment strategy

Suppose an investor believes that there are daily, weekly or monthly patterns in stock market data, and he/she seeks to exploit them for gain.

Sample strategies:

- Basic strategy: Buy a set of stocks each Monday, then sell on Wednesday; buy on the 6th of the month, then sell on the 19th; etc.
- Refinements: Sell the portfolio whenever it drops more than 10% from its initial value; purchase shares only when they increase in value more than 10% from start; etc.

Even with these very simple strategies, there are literally millions of variations (by changing various parameters), which can be quickly explored by computer.

Selecting only the best combination of parameters (and not mentioning the many others that were tried) is a classic **selection bias** statistical error.

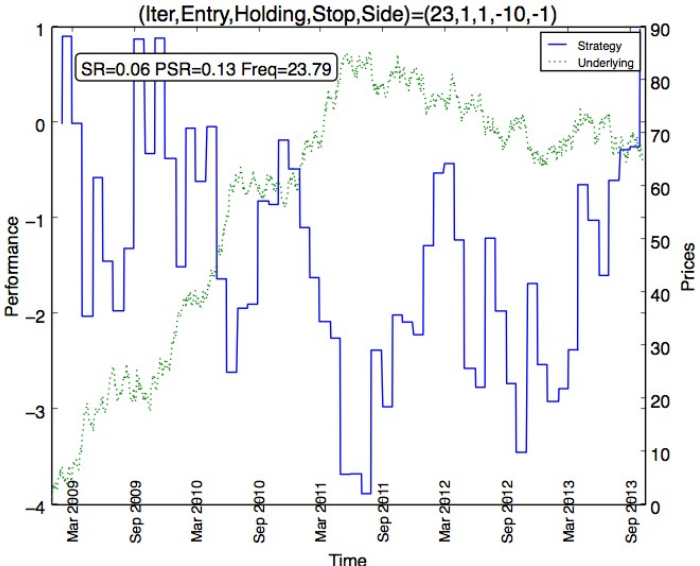
Optimizing an investment strategy to fit pseudorandom time series

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 time series dataset:

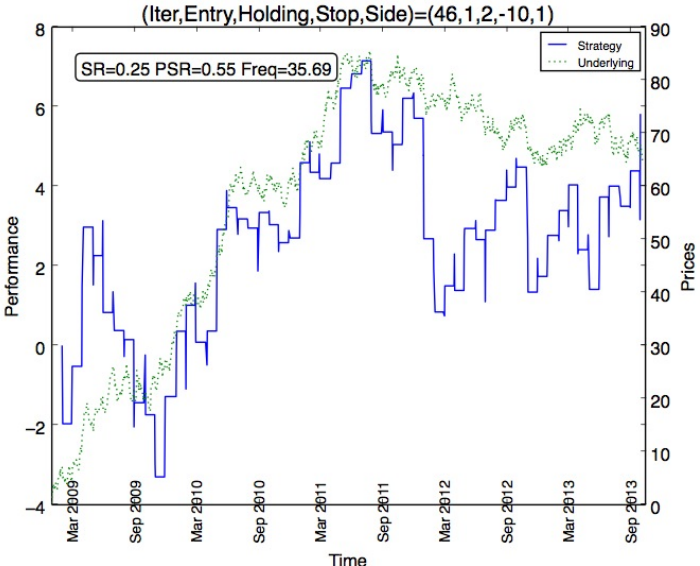
- The underlying dataset was generated by a pseudorandom number generator!
- As you can see, by tweaking some very basic parameters (entry price, sell price, stop-loss, etc), we can fit and "predict" the underlying time series quite well.
- The final (24th) viewgraph presents the results of implementing the resulting strategy on a continuation of the underlying (pseudorandom) dataset.

- The code is due to Stephanie Ger, Harvard University, modified from earlier code by Marcos Lopez de Prado.
- This software is NOW AVAILABLE in an online demo (try it yourself!): `http://datagrid.lbl.gov/backtest`.

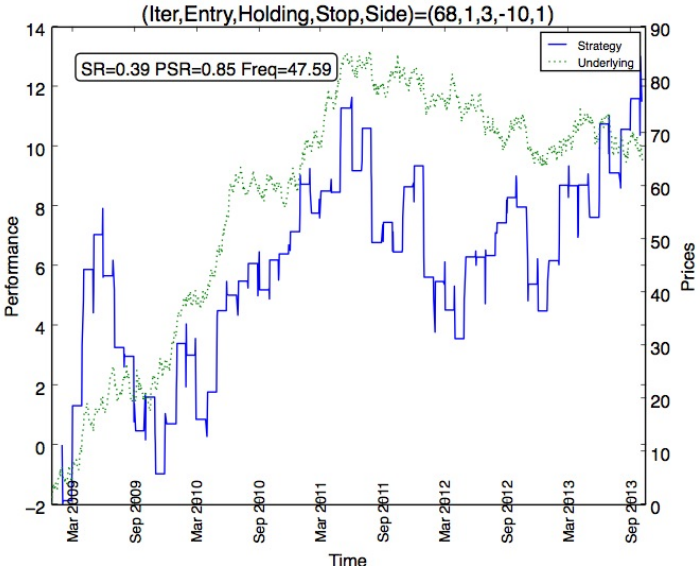
Optimizing an investment strategy to fit input time series, pg. 01



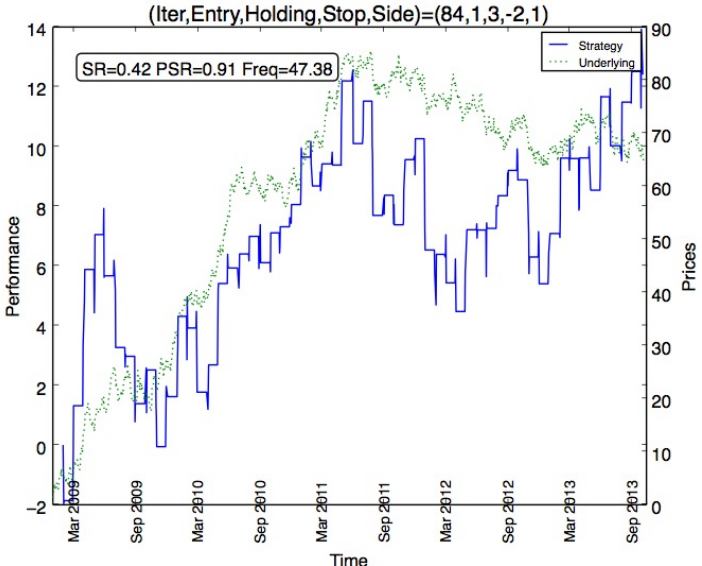
Optimizing an investment strategy to fit input time series, pg. 02



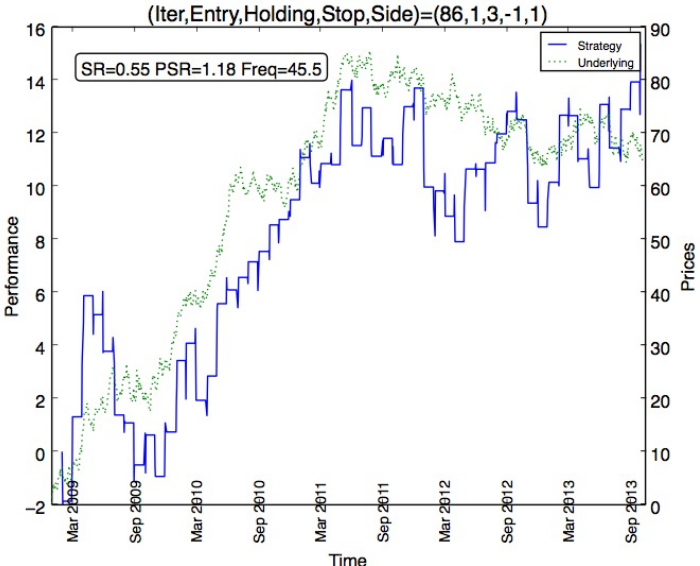
Optimizing an investment strategy to fit input time series, pg. 03



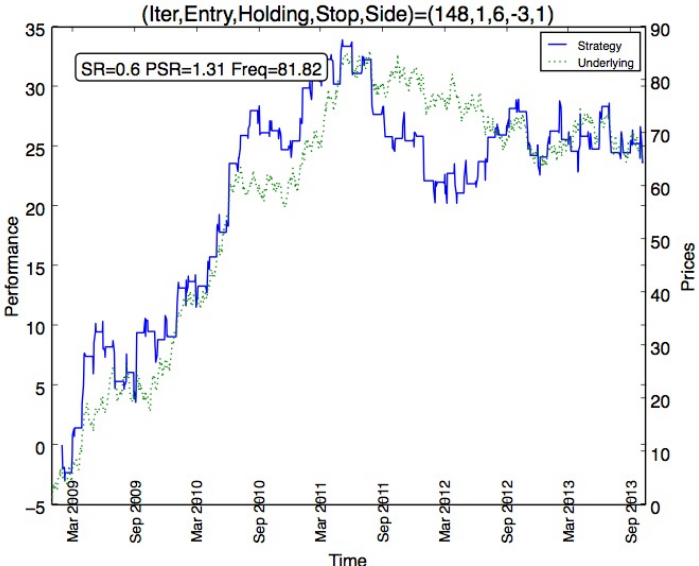
Optimizing an investment strategy to fit input time series, pg. 04



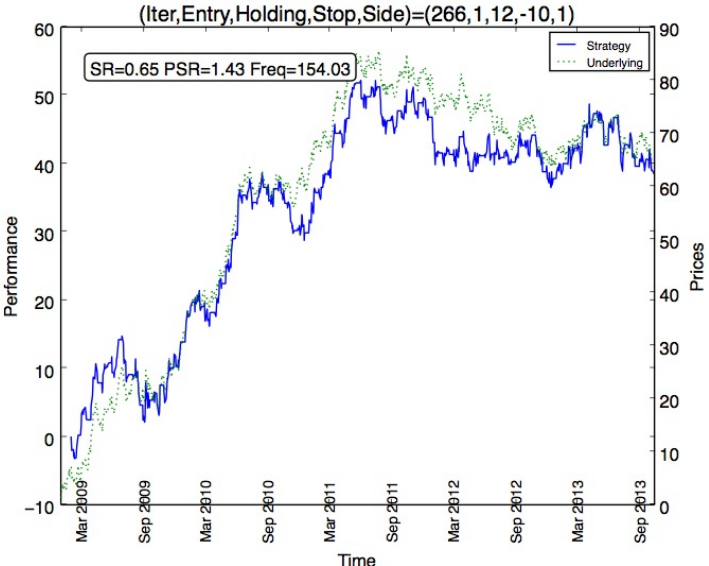
Optimizing an investment strategy to fit input time series, pg. 05



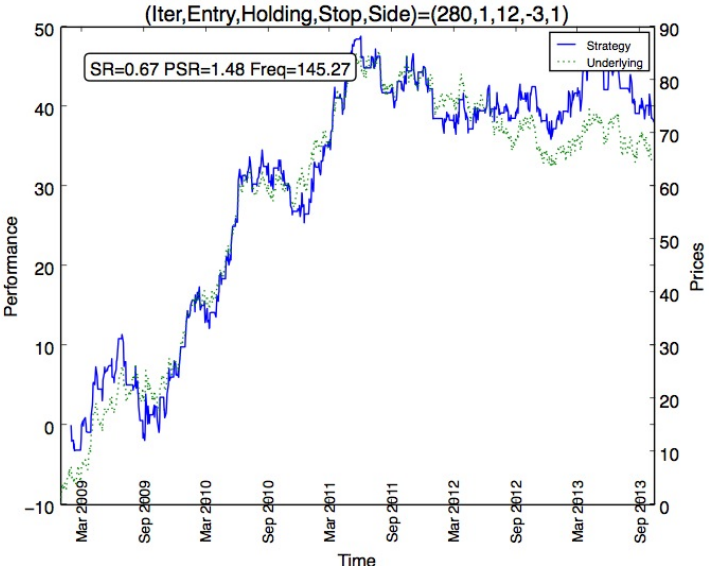
Optimizing an investment strategy to fit input time series, pg. 06



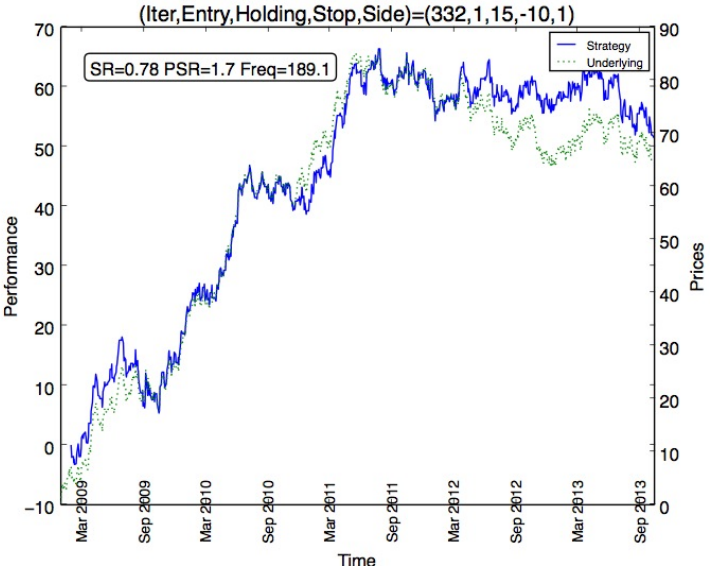
Optimizing an investment strategy to fit input time series, pg. 07



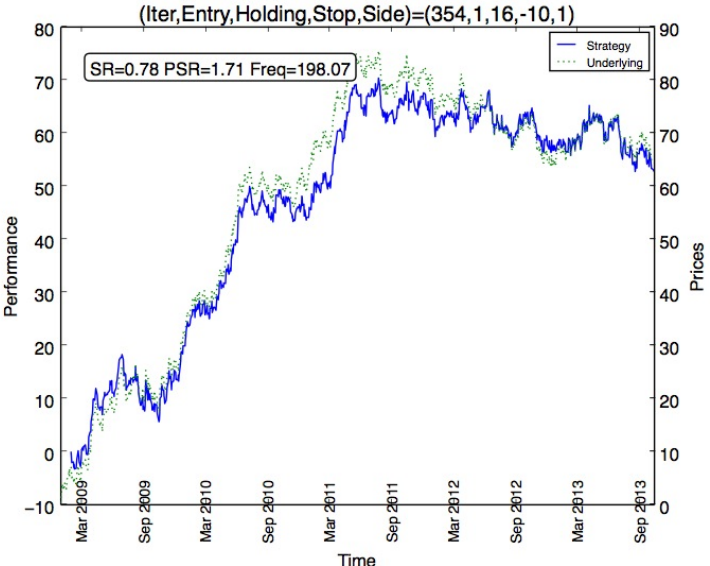
Optimizing an investment strategy to fit input time series, pg. 08



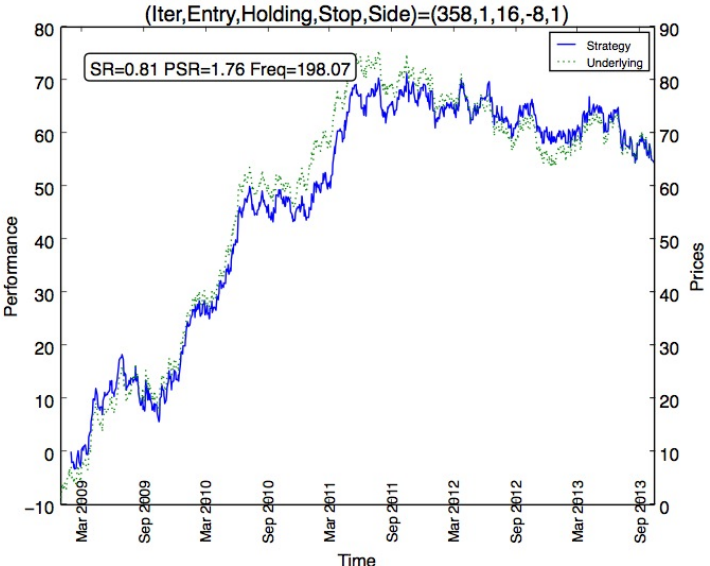
Optimizing an investment strategy to fit input time series, pg. 09



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

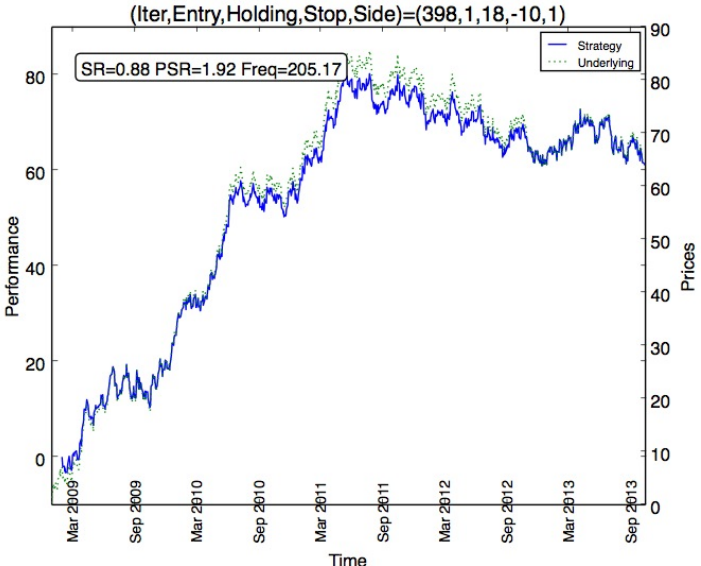


Optimizing an investment strategy to fit input time series, pg. 11



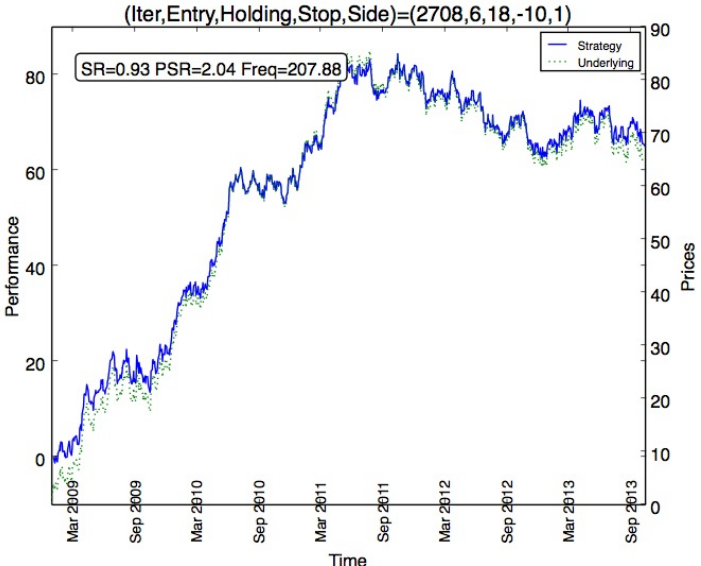
Optimizing an investment strategy to fit input time series, pg. 12

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

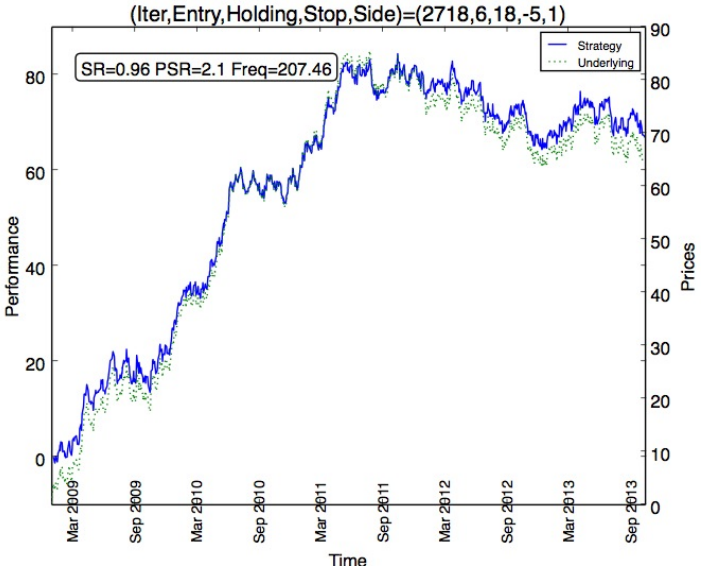


Optimizing an investment strategy to fit input time series, pg. 14

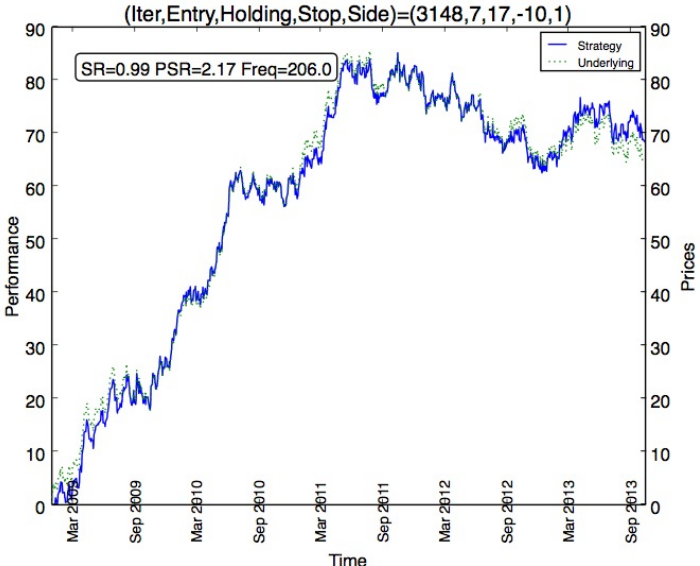
Optimizing an investment strategy to fit input time series, pg. 15



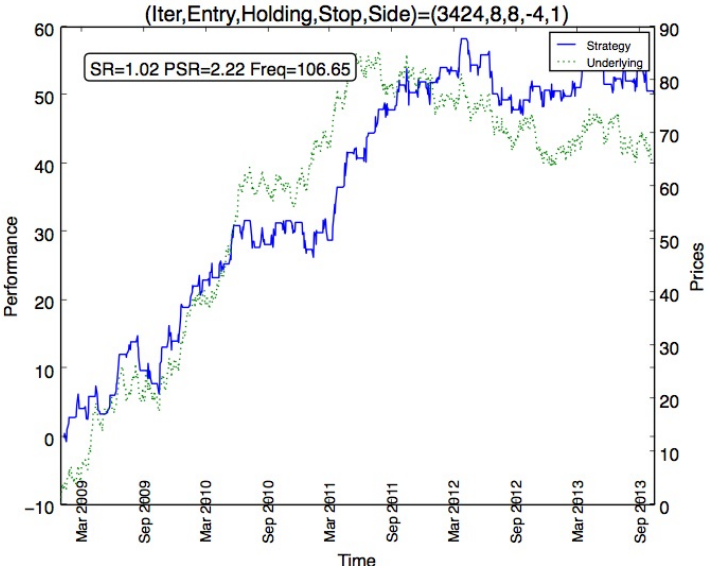
Optimizing an investment strategy to fit input time series, pg. 16



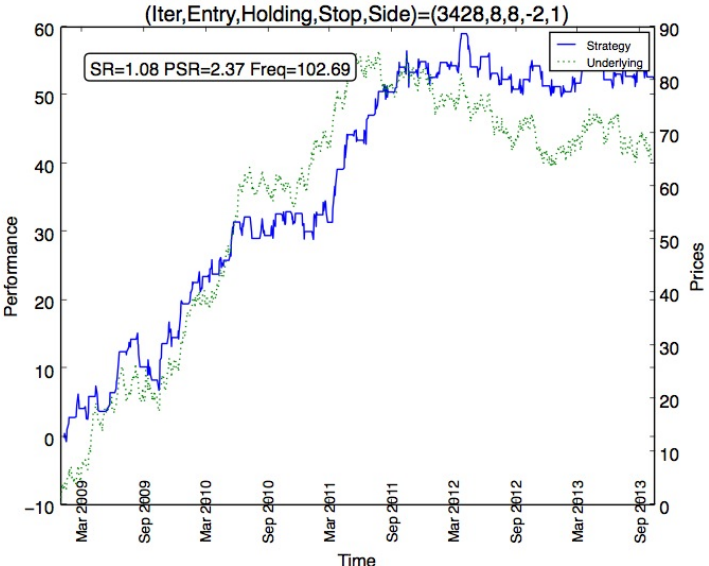
Optimizing an investment strategy to fit input time series, pg. 17



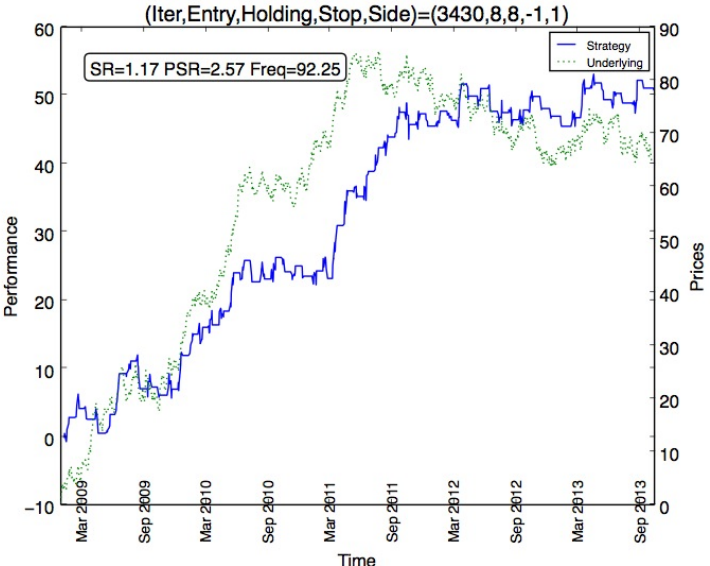
Optimizing an investment strategy to fit input time series, pg. 18



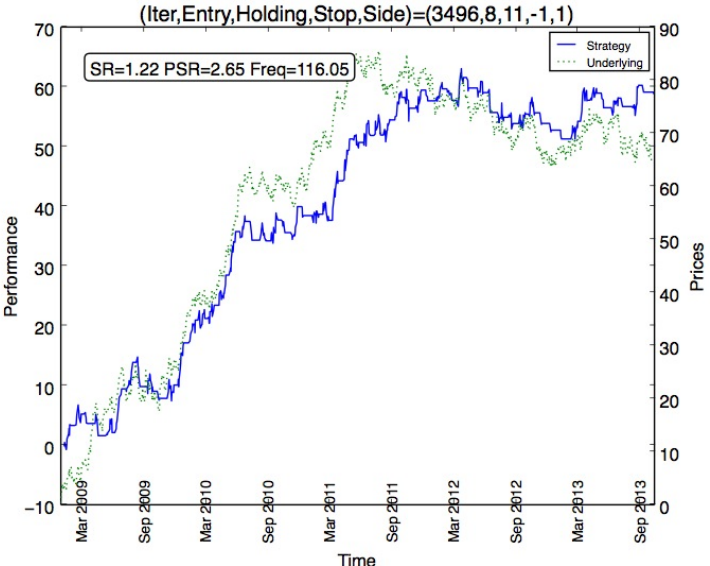
Optimizing an investment strategy to fit input time series, pg. 19



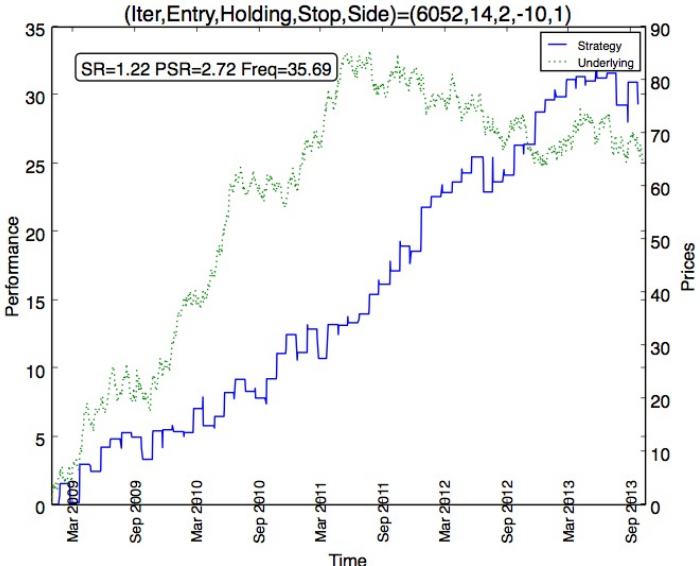
Optimizing an investment strategy to fit input time series, pg. 20



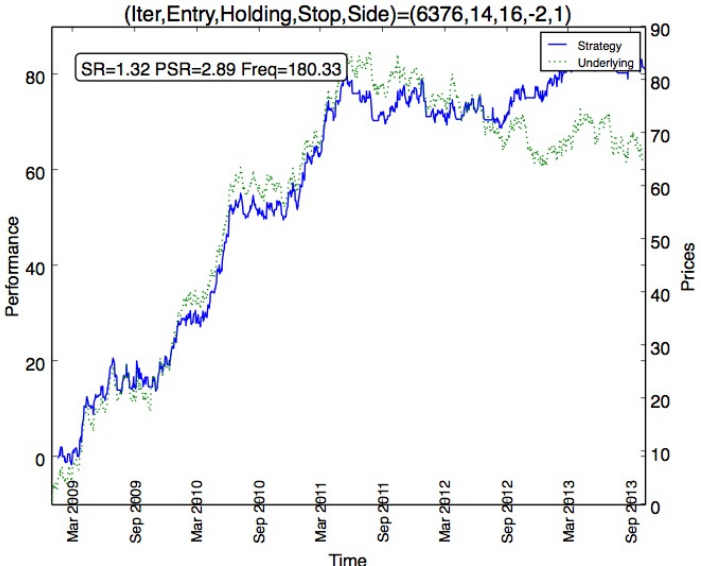
Optimizing an investment strategy to fit input time series, pg. 21



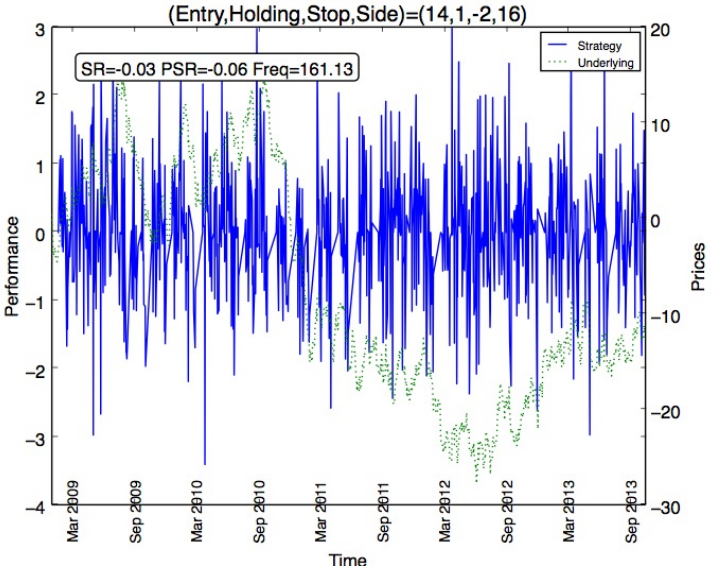
Optimizing an investment strategy to fit input time series, pg. 22



Final (optimal) strategy and its performance on the input time series



Deploying the resulting strategy on a continuation of the time series



Analysis

- After exhaustively exploring the space of strategy variations, the computer program found a strategy that achieved a Sharpe ratio of 1.32 on the input (backtest) pseudorandom time series (i.e., 1.32 standard deviations above zero).
- However, this optimal strategy, when applied to a new (pseudorandom) time series, failed miserably — the Sharpe ratio was -0.03 (i.e., slightly prone to *lose* money).
- In other words, the "optimal" strategy found by the computer search only fit idiosyncrasies of the input (backtest) dataset — it has no fundamental "intelligence" whatsoever.

For additional analysis (aimed at a fairly elementary audience), see:

- D. H. Bailey, S. Ger, M. Lopez de Prado, A. Sim and K. Wu, "Statistical overfitting and backtest performance," manuscript, 07 Oct 2014, `http://ssrn.com/abstract=2507040`.
- The software demo program is NOW AVAILABLE online: `http://datagrid.lbl.gov/backtest`.

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.

Is this just a problem for financial professionals?

The problem of statistical overfitting is hardly restricted to mathematical finance:

- As we noted at the beginning, many fields, long starved for good data, are now *drowning* in data.
- However, the underlying mathematical and statistical expertise of those who deal with this data has not kept up with the growth of this data.
- Example: At the University of California, Berkeley, typical enrollment in rigorous statistical methods courses is less than 200 per year, yet thousands of graduates will likely deal with "big data" during their career.
- At the least, practitioners in many fields should consult those with the appropriate statistical expertise before publishing the results of a study or analysis.

Visit our website and blog

Mathematicians Against Fraudulent Financial and Investment Advice (MAFFIA):
`http://www.financial-math.com`

This talk is available at:
`http://www.davidhbailey.com/dhbtalks/dhb-zurich-finance.pdf`.