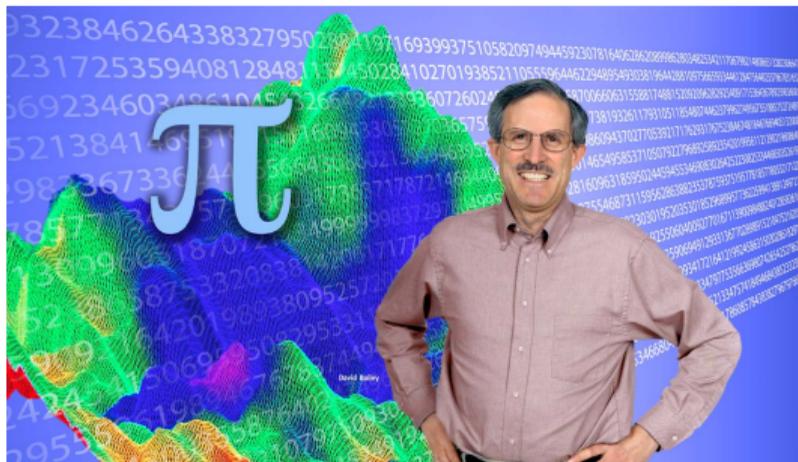


Getting financial predictions right

David H. Bailey

<http://www.davidhbailey.com>

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



DeepMind's AlphaGo Zero teaches itself to play Go

- ▶ In October 2017, DeepMind researchers programmed an AI computer with the rules of Go, then had the program play itself — literally teaching itself with no human input.
- ▶ After just three days of training, the resulting program “AlphaGo Zero” defeated their earlier program (which had defeated Ke Jie, the highest rated human) 100 games to 0.
- ▶ The Elo rating of Ke Jie, the world’s highest rated Go player, is 3661. After 40 days of training, AlphaGo Zero’s Elo rating was over 5000.
- ▶ Thus AlphaGo Zero is as far ahead of Ke as Ke is ahead of a good amateur.
- ▶ “The latest AI can work things out without being taught,” *The Economist*, 21 Oct 2017, <https://www.economist.com/news/science-and-technology/21730391-learning-play-go-only-start-latest-ai-can-work-things-out-without>.



Artificial intelligence and big data in finance

According to a Bloomberg report, Wall Street is entering a new era:

The fraternity of bond jockeys, derivatives mavens and stock pickers who've long personified the industry are giving way to algorithms, and soon, artificial intelligence.

Banks and investment funds have been tinkering for years, prompting anxiety for employees. Now, firms are rolling out machine-learning software to suggest bets, set prices and craft hedges. The tools will relieve staff of routine tasks and offer an edge to those who stay. But one day, machines may not need much help.

- ▶ S. Kishan, H. Son and M. Rojanasakul, "Robots are coming for these Wall Street jobs," <https://www.bloomberg.com/graphics/2017-wall-street-robots/>.

Artificial intelligence and big data in finance, continued

- ▶ **Sell side credit markets:** Automating subjective human decisions.
- ▶ **Sell side foreign exchange:** Anticipating variations in demand and prices.
- ▶ **Sell side equities:** Optimizing order execution.
- ▶ **Buy side equities:** Timing stock purchases and assessing risk.
- ▶ **Buy side credit:** Understanding bond deals, legal documents and court rulings.
- ▶ **Buy side macroeconomics:** Analyzing satellite images of Chinese industrial sites, Walmart parking lots, etc., to spot trends in the economy.

The message to the finance world is clear: **Prepare to adapt, or face irrelevance.**

Challenging times for hedge funds

- ▶ As of 1 July 2017, the HFRI Fund Weighted Composite Index is up 3.28% year-to-date, and 4.79% annualized gain for the previous 5 years.
- ▶ The corresponding figures for the S&P500 Index (including dividends) are 9.34% and 13.6%.
- ▶ In 2014, the California Public Employees' Retirement System (CalPERS) announced that it would liquidate its USD\$4 billion investment in hedge funds.
- ▶ In 2016, the New York City public pension fund announced that it will also liquidate all hedge fund investments.

Yet some hedge funds do consistently beat the market! According to *Institutional Investor's Alpha*:

Five of the six largest firms in this year's ranking rely all or mostly on computers and algorithms to make their investment decisions, a theme that has increasingly played a role in the top-100 ranking over the past few years. And all five posted asset increases last year.

DANGER AHEAD



Artificial intelligence, big data and high-performance computing can generate nonsense faster than ever before!



Details to consider any computer-based research and development activity:

- ▶ Are the theoretical models realistic?
 - ▶ Is the computer implementation debugged?
 - ▶ Are the statistical methods and data analysis techniques appropriate?
 - ▶ Is there sufficient data to draw statistically reliable conclusions?
 - ▶ Can other researchers reproduce the findings?
-
- ▶ V. Stodden, M. McNutt, D. H. Bailey, E. Deelman, Y. Gil, B. Hanson, M. A. Heroux, J. P. A. Ioannidis and M. Tauber, "Enhancing reproducibility for computational methods," *Science*, vol. 356, no. 3617 (3 Dec 2016), 1240–1241.

Reproducibility crises in biomedicine, psychology and economics

- ▶ In 2011, Bayer researchers reported that they were able to reproduce only 17 of 67 pharma studies.
- ▶ In 2012, Amgen researchers reported that they were able to reproduce only 6 of 53 cancer studies.
- ▶ In August 2015, the Reproducibility Project in Virginia reported that they were able to reproduce only 39 of 100 psychology studies.
- ▶ In September 2015, the U.S. Federal Reserve was able to reproduce only 29 of 67 economics studies.



Reproducibility Project staff

Credit: NY Times

A crisis in finance?

- ▶ Many individual investors believe that the financial system is rigged against them.
- ▶ Many are confused about the thousands of investment funds and services available.
- ▶ Many are skeptical of the claims of financial forecasters.
- ▶ Financial news is replete with questionable charts, graphs and pseudomathematical jargon: “Fibonacci ratios,” “cycles,” “waves,” “golden ratios,” “parabolic SARs,” “pivot points,” “symmetrical triangles,” “rising wedges,” etc.



What should responsible finance researchers do?

First and foremost, ensure that their own published research is mathematically and statistically sound.

Email exchange between DHB and a finance colleague

Email from DHB to finance colleague, 10 June 2013:

One thing that has always puzzled me about the financial world is the following sort of thing: [press examples cited]. Excuse me for being “dumb,” but this sort of thing seems to me to be outright nonsense. ... When people like those above say that they “know” where the stock market is heading, this cannot have any scientific basis. ...

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

Response from finance 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. ...

I completely agree with your assessment. The amount of nonsense ... is incredible.

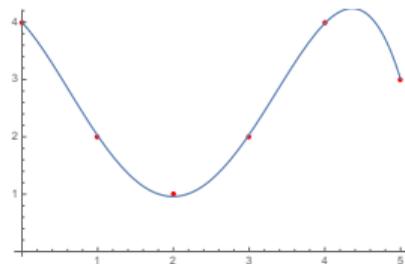
What is backtest overfitting?

Backtest: Testing a model based on historical market data.

Backtest overfitting: Statistical overfitting of historical market data.

Examples:

- ▶ Proposing a theoretical model that inherently possesses a higher level of complexity than the backtest data.
- ▶ Using a computer to try millions or billions of variations of a model or strategy on historical market data, and then only presenting results from the variation that works best.
- ▶ Constructing an investment fund by using a computer to explore millions or billions of weighting factors, then only marketing the one that works best.

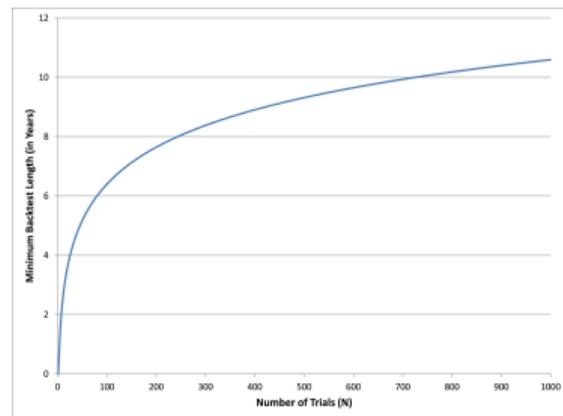


Fitting six data points (almost perfectly!) with a fourth-degree function.

When a computer can analyze many variations of a fund or strategy on a fixed dataset, it is almost certain that the “optimal” selection will be statistically overfit.

How easy is it to overfit a backtest? Very!

- ▶ If only 2 years of daily market data are available, then no more than 7 variations should be tried.
- ▶ If only 5 years of daily market data are available, then no more than 45 variations should be tried.



Backtest overfitting is now thought to be the principal reason why financial funds and strategies that look great on paper often disappoint in practice.

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

Letters to clients: An absurd investment scheme

- ▶ A financial advisor sends letters to $5,120 = 5 \times 2^{10}$ prospective clients, with 2560 predicting a certain stock will go up, and 2560 predicting it will go down.
- ▶ One month later, the advisor sends letters only to the 2560 investors who were previously sent the correct prediction, with 1280 letters predicting a certain stock will go up, and 1280 predicting it will go down.
- ▶ After ten months, the final five investors will have been sent ten consecutive spot-on predictions!



This strategy is absurd, even fraudulent, because the final five investors are not told of the 10,235 other letters with different predictions.

But why is promoting a statistically overfit fund or strategy, where investors are not informed of the millions of failed computer trials, any different?

Proliferation of new stock funds

- ▶ Roughly USD\$2.1 trillion is held in U.S.-listed exchange-traded funds (ETFs), with hundreds of new ETFs minted each year.
- ▶ In a 2012 study, researchers found that the median time between the definition of a new index and the inception of a new ETF based on the index dropped from almost three years in 2000 to only 77 days in 2011.
- ▶ As a result, the report concludes, “most indexes have little live performance history for investors to assess in the context of a new ETF investment.”

Performance of the new funds:

Out of 370 new indexes, 87% of the indexes outperformed the broad U.S. market, based on computer-based tests before inception, but only 51% outperformed after inception.

The funds had an average 12.25% return above the broad market, based on computer-based tests before inception, but -0.26% return after inception.

- ▶ Vanguard Research, “Joined at the hip: ETF and index development,” July 2012, <http://www.vanguard.com/pdf/s319.pdf>.

How difficult is it to design a stock portfolio to achieve a desired performance profile?

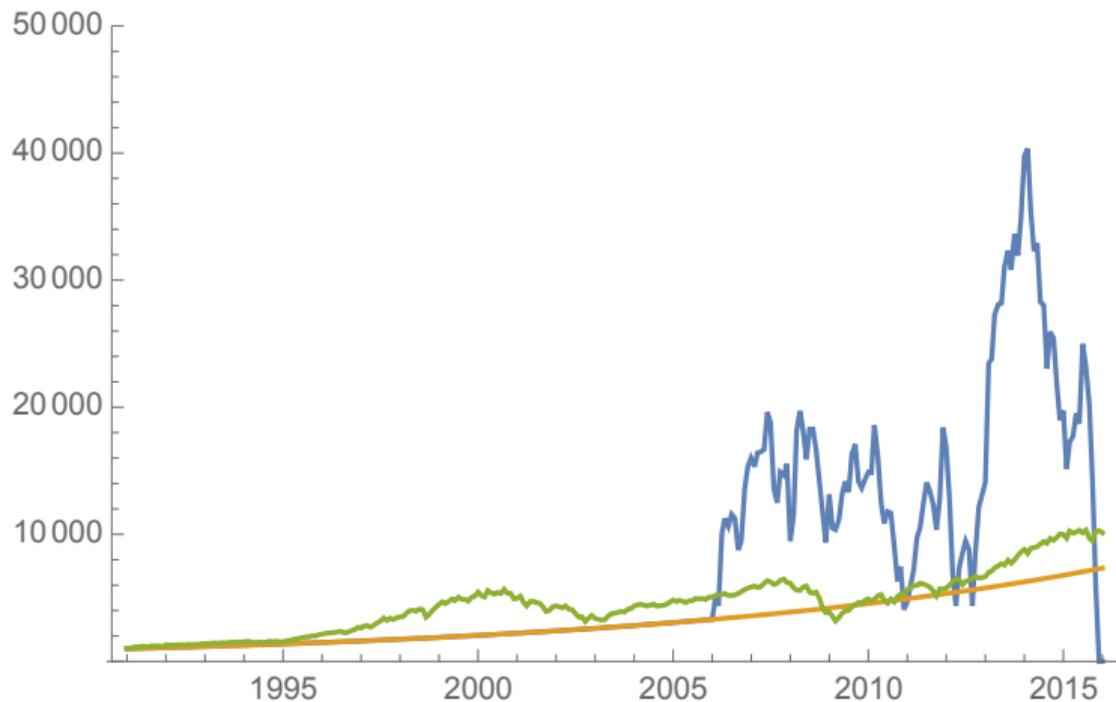
- ▶ Given some desired performance profile (a time series), we devised a computer program that constructs a weighted subset of S&P500 stocks whose performance matches the profile over the specified backtest time period (1991–2005).
- ▶ But most of these portfolios fail miserably when presented with new data (2006–2015).

These erratic and often catastrophic results on new (out-of-sample) data are symptomatic of statistical overfitting.

Technical details are given here:

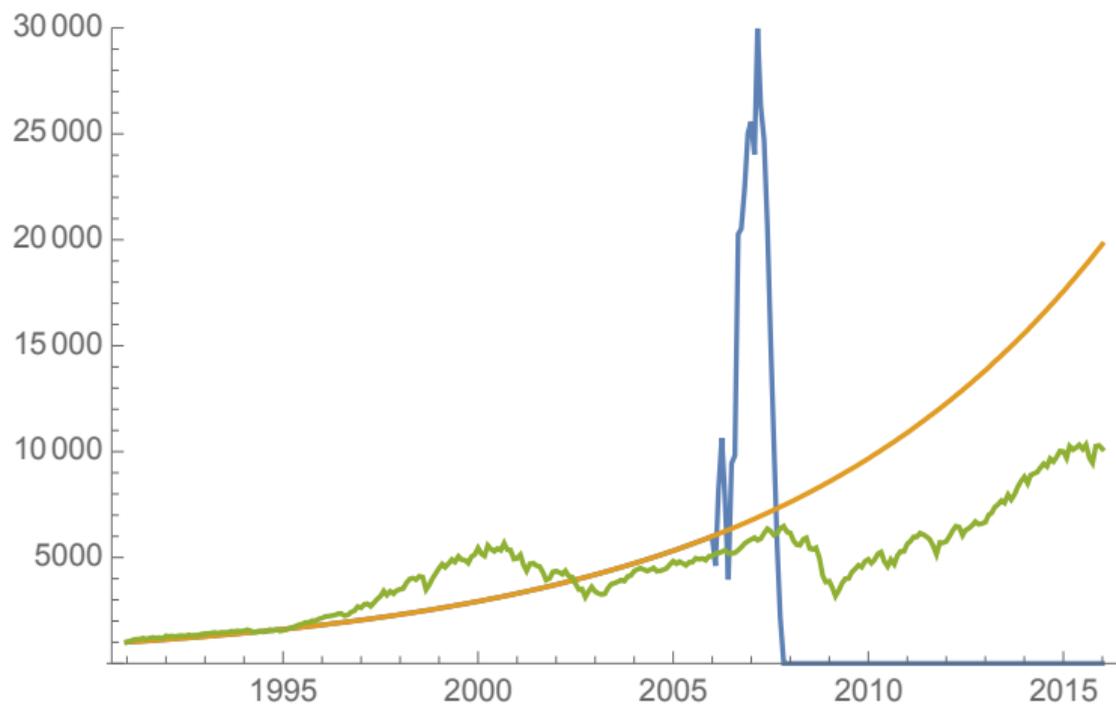
- ▶ D. H. Bailey, J. M. Borwein and M. Lopez de Prado, “Stock portfolio design and backtest overfitting,” *Journal of Investment Management*, vol. 17 (2017), no. 1, preprint available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2739335.

Steady growth profile, APR = 8%



Blue: constructed portfolio; orange: target profile; green: S&P500.

Steady growth profile, APR = 12%



Blue: constructed portfolio; orange: target profile; green: S&P500.

Can stock market “gurus” forecast the market?

Kaissar's analysis of prominent market forecasters (covering 1999 to 2016):

- ▶ The forecasters *overestimated* the S&P 500's year-end price by 26.2% on average during the three recession years 2000 to 2002.
- ▶ They *underestimated* the index's level by 10.6% for the initial recovery year 2003.
- ▶ They *overestimated* the S&P 500's year-end level by a whopping 64.3% in 2008.
- ▶ They *underestimated* the index by 10.9% for the first half of 2009.

Kaissar's conclusion: **The forecasts were least useful when they mattered most.**

- ▶ N. Kaissar, “S&P 500 forecasts: Crystal ball or magic 8?,” *Bloomberg News*, 23 December 2016, <https://www.bloomberg.com/gadfly/articles/2016-12-23/s-p-500-forecasts-mostly-hit-mark-until-they-matter-most>.

Our analysis of market forecasters

In 2012, the CXO Advisory Group ranked 68 forecasters based on their 6,582 forecasts (forecasts made for the S&P 500 index). We extended and advanced this study:

- ▶ Each individual forecast is treated according to two factors: (a) *time frame* and (b) *specificity*.
- ▶ Long-term forecasts are considered more significant than the short-term forecasts, since long-term underlying trends tend to overcome short-term noise.
- ▶ Specific forecasts are treated more important than non-specific ones.

Summary of forecaster study

- ▶ Across all forecasts, the average accuracy was 48%.
- ▶ Two-thirds of forecasters had an accuracy level below 50%.
- ▶ Only about 6% of forecasters had accuracy values between 70% and 79%; the highest accuracy value was still below 80%.

For details:

- ▶ D. H. Bailey, J. M. Borwein, A. Salehipour, and M. Lopez de Prado, "Evaluation and ranking of market forecasters", *Journal of Investment Management*, to appear, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2944853.

More than just a finance problem: Pitfalls in big data science

Artificial intelligence and big data have great promise to accelerate progress in science, engineering, finance and commerce.

But there are numerous potential pitfalls:

- ▶ Using unrealistic theoretical models.
- ▶ Using questionable statistical tests and techniques.
- ▶ Using insufficient data to draw statistically reliable conclusions.
- ▶ Failing to test thoroughly for software errors.
- ▶ Failing to facilitate reproducibility.
- ▶ Becoming so enamored with the science that we fail to see errors and weaknesses.

Let's not be guilty of generating scientific nonsense!

Why the silence?

Historically scientists have exposed pseudoscience — astrology in the 19th century; young-earth creationism and health fads in the 20th century, etc.

Yet 21st century scientific researchers have remained disappointingly silent with regards to those in the community who, knowingly or not:

- ▶ Engage in “p-hacking” to inflate the statistical significance of their results.
- ▶ Fail to disclose the number of models or variations that were tried in their study.
- ▶ Misuse probability theory and statistical methods.
- ▶ Present misleading charts and graphs.
- ▶ Use questionable pseudomathematical jargon.
- ▶ Make vague claims that do not permit rigorous testing and falsification.
- ▶ Withhold key details, so other researchers cannot test or reproduce their results.

As we wrote in a recent paper:

Our silence is consent, making us accomplices in these abuses.

Thanks!

Visit our website at:

<http://www.mathinvestor.org>

This talk is available at:

<http://www.davidhbailey.com/dhbtalks/dhb-mediAx-2017.pdf>

Collaborators:

- ▶ Jonathan M. Borwein (University of Newcastle, Australia, deceased 2016).
- ▶ Marcos Lopez de Prado (Guggenheim Partners and LBNL).
- ▶ Amir Salehipour (University of Newcastle, Australia).
- ▶ Jim Zhu (Western Michigan University).