New financial analysis tools at CARMA

Amir Salehipour

CARMA, The University of Newcastle

Joint work with Jonathan M. Borwein, David H. Bailey and Marcos López de Prado

November 13, 2015

Table of Contents

Purope and motivation

Motivation

1. Extracting the optimal investment strategy from the historical data may not be an optimal strategy from now on / in the future;
2. Situations exist when a model targets a specific behavior than a general one.

What should we do?

1. We illustrate two online tools in order to understand the problem of 'overfitting':
   BODT: Backtest Overfitting Demonstration Tool, and
   TMST: Tenure Maker Simulation Tool
2. We introduce advanced statistics to be used;
3. We develop an advanced iterative algorithm that minimizes the impact of overfitting.

Major sources of overfitting

Two sources of error in evaluating strategies:

1. **Multiple testing**: If we are analyzing and evaluating multiple strategies (which is a typical case in choosing the best investment strategy), the probability of choosing at least one poor strategy grows.

2. **Selection bias**: If we are conducting multiple tests and we are ignoring negative outcomes, we are exposed to a biased sample of outcomes.

These two sources of error 'may lead to high estimated values for Sharpe Ratio (SR), where the true SR may even be null'.

Example 1*: Tossing a fair coin

- A fair coin is tossed for 10 times. Assume the following output: $\{+,+,+,+,+,-,-,-,-,-\}$.

*Example from: Bailey and López de Prado (2014)

Example 1*: Tossing a fair coin

- A fair coin is tossed for 10 times. Assume the following output: $\{+,+,+,+,+,-,-,-,-,-\}$.
- One may claim that the best strategy for betting on the outcome is to expect '$+$' on the first five tosses and '$-$' on the last five tosses.

*Example from: Bailey and López de Prado (2014)

Example 1*: Tossing a fair coin

- A fair coin is tossed for 10 times. Assume the following output: $\{+,+,+,+,+,-,-,-,-,-\}$.
- One may claim that the best strategy for betting on the outcome is to expect '$+$' on the first five tosses and '$-$' on the last five tosses.
- **How probable is to have this outcome repeated in the future?** Repeating the experiment may result in $\{-,-,+,-,+,+,-,-,+,-\}$; still we win five times and we lose five times; but what about our bet?

*Example from: Bailey and López de Prado (2014)

Example 1*: Tossing a fair coin

- A fair coin is tossed for 10 times. Assume the following output: $\{+,+,+,+,+,-,-,-,-,-\}$.
- One may claim that the best strategy for betting on the outcome is to expect '$+$' on the first five tosses and '$-$' on the last five tosses.
- **How probable is to have this outcome repeated in the future?** Repeating the experiment may result in $\{-,-,+,-,+,+,-,-,+,-\}$; still we win five times and we lose five times; but what about our bet?
- It is clear that the claimed betting rule was overfit.

*Example from: Bailey and López de Prado (2014)

Backtest overfitting

Definition 1: Backtest

Backtest is an evaluation / a simulation of a new strategy by using historical data.

Backtest overfitting

Definition 1: Backtest

Backtest is an evaluation / a simulation of a new strategy by using historical data.

Definition 2: Backtest overfitting

Typically a backtest is realistic when the training performance (on In-Sample or IS data) is consistent with the testing performance (on Out-Of-Sample or OOS data); however, this may not happen in practice, which causes overfitting concept.

Backtest overfitting

Definition 1: Backtest

Backtest is an evaluation / a simulation of a new strategy by using historical data.

Definition 2: Backtest overfitting

Typically a backtest is realistic when the training performance (on In-Sample or IS data) is consistent with the testing performance (on Out-Of-Sample or OOS data); however, this may not happen in practice, which causes overfitting concept.

Backtest overfitting is now thought to be a primary reason why quantitative investment models and strategies that look good on paper often disappoint in practice.

Backtest overfitting

Definition 1: Backtest

Backtest is an evaluation / a simulation of a new strategy by using historical data.

Definition 2: Backtest overfitting

Typically a backtest is realistic when the training performance (on In-Sample or IS data) is consistent with the testing performance (on Out-Of-Sample or OOS data); however, this may not happen in practice, which causes overfitting concept.

Backtest overfitting is now thought to be a primary reason why quantitative investment models and strategies that look good on paper often disappoint in practice.

The two online tools

1. show the impact of backtest overfitting on investment strategies, and
2. propose correction tests.

Types of investment strategies

1. General trading rules: very popular among investors, and are marketed every day in TV shows, business publications and academic journals. Example is seasonal strategies.
   $\longrightarrow$ Backtest Overfitting Demonstration Tool (BODT) illustrates how easy is to overfit a backtest involving a seasonal strategy.

Types of investment strategies

1. General trading rules: very popular among investors, and are marketed every day in TV shows, business publications and academic journals. Example is seasonal strategies.
   $\longrightarrow$ Backtest Overfitting Demonstration Tool (BODT) illustrates how easy is to overfit a backtest involving a seasonal strategy.

2. The rest of strategies (those based on econometric or statistical (forecasting) methods): the strategies often published in highly respected academic journals.
   $\longrightarrow$ Tenure Maker Simulation Tool (TMST) shows these investments strategies are even easier to overfit than in the 'seasonal' counterpart.

   (We call it the 'Tenure Maker' because many academic research results in finance are subject to the criticism that they have been produced in this way.)

Backtest Overfitting Demonstration Tool (BODT)

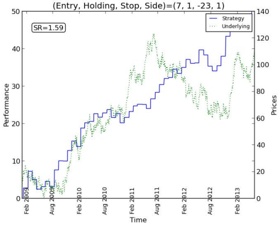
- The BODT is an online interface available to the public (https://carma.newcastle.edu.au/backtest/)
- The tool finds optimal strategies on random (unpredictable) data as well as on real-world stock market data (S&P500);
- The tool demonstrates that high **Sharpe Ratios (SR)** are meaningless unless investors control for the number of trials.

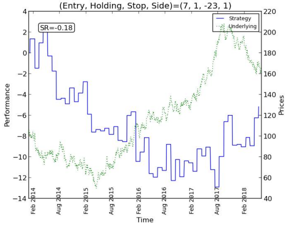
How BODT functions?

1. **Data generation/import**: Generates a pseudorandom time series which reflects the history of stock market prices/imports real-world S&P500 stock market data (historical data);
2. **Combination generation**: Generates all combinations of the parameters *entry date*, *holding period*, *stop loss* and *side* (by successively adjusting values);
3. **Evaluation/training**: Evaluates all generated combinations of the parameters; if an improved combination (improvement over Sharpe Ratio) than the best recorded is found, a new strategy is then recorded;
4. **Optimal strategy**: Reports a trading strategy which maximizes the Sharpe Ratio statistic (optimal values of the parameters);
5. **Evaluation/testing**: Evaluates the performance of the optimal strategy on OOS data (by using the optimal values for the parameters);
6. **Reporting**: Generated numerical and visualized performance reports.

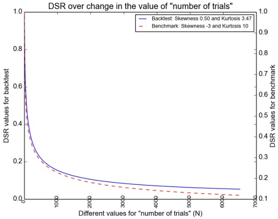
Report and analysis

1. Performance of the optimal strategy on IS (left) and on OOS (right)





2. Performance analysis: normality vs non-normality.



The tutorial is available at https://carma.newcastle.edu.au/backtest/

Tenure Maker Simulation Tool (TMST)

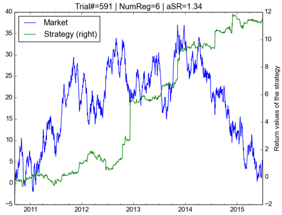
- The TMST is an online interface available to the public (https://carma.newcastle.edu.au/tenuremaker/)
- The TMST looks for econometric specifications that maximize the predictive power (in-sample) of a random (unpredictable) time series;
- The resulting Sharpe Ratio tends to be even higher than in the ‘seasonal’ (illustrated by BODT) counterpart.

How TMST functions?

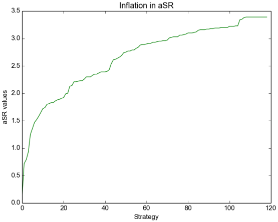
1. **Data generation**: A series of IID (independent, identically distributed) Normal returns is generated;
2. **Combination generation**: A large number of time series models is automatically generated, where the series is forecasted as a fraction of past realizations of that same series. The time series models include:
   1. Moving (rolling) sums of the past series;
      (similar to the moving average with the difference that sum, instead of average, is used over the last past series)
   2. Polynomials of the past series;
      ($y = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n + \epsilon$, where, y is the dependent (response) variable, x is the explanatory variable (regressor), a is regression coefficient (a parameter) and ϵ is the error)
   3. Lags of the past series;
      (the dependent variable y is predicted based on both the current values of an explanatory variable, x, as well as the lagged (past period) values; e.g. $y = a_0 + a_1 x_t + a_1 x_{t-1} + a_2 x_{t-2} + \epsilon$)
   4. Cross-products of the above.
3. **Optimal strategy**: A forward-selection algorithms is applied on these alternative specifications, in order to derive an optimal forecasting equation. As the selection algorithm progresses, it publishes the improved model;
4. **Reporting**: Generated visualized performance reports.

Report and analysis

1. Progress movie (the left plot)



2. Inflation in Sharpe Ratio (the right plot)



The tutorial is available at
https://carma.newcastle.edu.au/tenuremaker/

Sharpe Ratio (SR)

The most widely used performance statistic.

How much additional return we can receive for the additional deviation when holding the risky asset over a risk-free asset (e.g. governmental bond). The greater the SR, the better.

Definition 3: Sharpe Ratio (SR)

$$SR = \frac{E(R_a - R_b)}{\sigma_a}$$

Ratio of the expected values of the excess return over a benchmark asset return to the standard deviation of the excess return or risky asset (Sharpe, 1994).

where

- R_a is the asset return;
- R_b is the benchmark (risk free) asset return;
- $E(R_a - R_b)$ is the expected value of the excess of the asset return over the benchmark return;
- σ_a is the standard deviation of the excess return.

Weakness: the returns must be normally distributed; because the behavior and effectiveness of the standard deviation changes in case of non-normally distributed data.

Probabilistic Sharpe Ratio (PSR)

Overcomes normality requirement of SR.

PSR incorporates information regarding the non-normality of the returns. (Bailey and López de Prado, 2012).

Definition 4: Probabilistic Sharpe Ratio (PSR)

PSR is the probability that an estimated SR exceeds a given threshold when non-normally distributed returns exist.

$$\widehat{PSR} = P(\widehat{SR} > SR^*) = Z[\frac{(\widehat{SR} - \widehat{SR^*})\sqrt{T-1}}{\sqrt{1 - \widehat{\gamma_3}\widehat{SR} + \frac{\widehat{\gamma_4}-1}{4}\widehat{SR}^2}}]$$

where

- N is the number of independent trials;
- T is the sample length/track record/returns; for instance, three years of stock market data;
- $\widehat{\gamma_3}$ is the skewness of the returns distribution;
- $\widehat{\gamma_4}$ is the kurtosis of the returns distribution;
- Z is the cumulative standard normal distribution.

Deflated Sharpe Ratio (DSR)

Definition 5: Deflated Sharpe Ratio (DSR)

DSR is a PSR where the threshold is set so that the impact of all tried strategies is captured as well as the non-normality of the returns' distribution (Bailey and López de Prado, 2014).

$$\widehat{DSR} = P(\widehat{SR} > SR^*)$$

and

$$\widehat{SR^*} = \sqrt{Var[\{\widehat{SR_n}\}]} \times ((1-\gamma)Z^{-1}(1-\frac{1}{N}) + \gamma Z^{-1}(1-\frac{1}{N}e^{-1}))$$

where

- $\gamma \approx 0.5772$ (the Euler-Mascheroni constant);
- $e \approx 2.7182$ (the Euler's number);
- $\widehat{SR_n}$ is an SR estimate associated with each trial (out of N trials) with variance $Var[\{\widehat{SR_n}\}]$.

Example 2: How DSR can be helpful?

- Assume combining different parameters yields the optimal investment strategy with annualized $SR = 2.5$, and the sample length $T = 1250$ (5 years).

  Suppose $N = 100$ independent trials;
  $\widehat{\gamma_3} = -3$ and $\widehat{\gamma_4} = 10$ (the skewness and the kurtosis of the returns distribution);
  $Var[\{\widehat{SR_n}\}] = 0.5$

Example 2: How DSR can be helpful?

- Assume combining different parameters yields the optimal investment strategy with annualized $SR = 2.5$, and the sample length $T = 1250$ (5 years).

  Suppose $N = 100$ independent trials;
  $\widehat{\gamma}_3 = -3$ and $\widehat{\gamma}_4 = 10$ (the skewness and the kurtosis of the returns distribution);
  $Var[\{\widehat{SR_n}\}] = 0.5$

- How does the strategy sound to an investor? **Very good**;

Example 2: How DSR can be helpful?

- Assume combining different parameters yields the optimal investment strategy with annualized $SR = 2.5$, and the sample length $T = 1250$ (5 years).

  Suppose $N = 100$ independent trials;
  $\widehat{\gamma}_3 = -3$ and $\widehat{\gamma}_4 = 10$ (the skewness and the kurtosis of the returns distribution);
  $Var[\{\widehat{SR_n}\}] = 0.5$

- How does the strategy sound to an investor? **Very good**;

Example 2: How DSR can be helpful?

- Assume combining different parameters yields the optimal investment strategy with annualized $SR = 2.5$, and the sample length $T = 1250$ (5 years).

  Suppose $N = 100$ independent trials;
  $\widehat{\gamma}_3 = -3$ and $\widehat{\gamma}_4 = 10$ (the skewness and the kurtosis of the returns distribution);
  $Var[\{\widehat{SR}_n\}] = 0.5$

- How does the strategy sound to an investor? **Very good**;

- Does it really work? **Not really sure!**

Example 2: How DSR can be helpful?

- Assume combining different parameters yields the optimal investment strategy with annualized $SR = 2.5$, and the sample length $T = 1250$ (5 years).

  Suppose $N = 100$ independent trials;
  $\widehat{\gamma}_3 = -3$ and $\widehat{\gamma}_4 = 10$ (the skewness and the kurtosis of the returns distribution);
  $Var[\{\widehat{SR_n}\}] = 0.5$

- How does the strategy sound to an investor? **Very good**;

- Does it really work? **Not really sure!**

- How to ensure? **Compute the DSR**;

Example 2: How DSR can be helpful?

- Assume combining different parameters yields the optimal investment strategy with annualized $SR = 2.5$, and the sample length $T = 1250$ (5 years).

  Suppose $N = 100$ independent trials;
  $\widehat{\gamma}_3 = -3$ and $\widehat{\gamma}_4 = 10$ (the skewness and the kurtosis of the returns distribution);
  $Var[\{\widehat{SR_n}\}] = 0.5$

- How does the strategy sound to an investor? **Very good**;

- Does it really work? **Not really sure!**

- How to ensure? **Compute the DSR**;

- At a 95% confidence level $\widehat{DSR} = 0.9004 < 0.95$; **thus, not a good investment strategy**.

Documentation

We have documented the BODT and the TMST tools as two tutorials (in `.pdf`) which may be downloaded from the tools' websites, as well as our manuscript entitled:

Online tools for demonstration of backtest overfitting
By David H. Bailey, Jonathan M. Borwein, Amir Salehipour, Marcos L. de Prado, Qiji J. Zhu (to be submitted to ANZIAM Journal)

which may be downloaded at SSRN (`ssrn.com`), ID 2597421 (Bailey et al., 2015) (`http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2597421`)

An iterative PSR-based algorithm

Motivation: How one may benefit PSR when selecting investment strategies in order to minimize the impact of overfitting?

An iterative optimization algorithm: We developed an iterative algorithm which benefits PSR in making decisions; more precisely, the algorithm

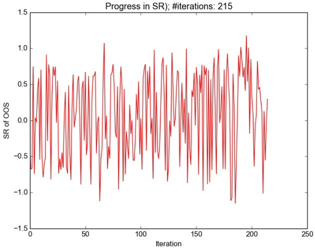
1. Computes PSR for every strategy;
2. Keeps positive SR;
3. Sorts SR in decreasing order of PSR;
4. Selects the first `item_max` (a user defined parameter) strategies out of the list;
5. Applies all strategies (of the list) on OOS;
6. Chooses the best strategy.

Example 3: Outcome of the iterative PSR-based algorithm

Preliminary results highlight:

Parameters	Without algorithm	With algorithm
entry day	15	20
holding period	5	7
stop loss	-1	-4
side	-1	-1
SR (OOS)	-0.1034	1.1755

- With algorithm: the plot shows all strategies; the best strategy is selected where SR on OOS is 1.1755;
- Without algorithm: BODT yields the optimal strategy with IS SR of 1.5627 and the true SR (on OOS) of -0.1034!



Bibliography

D. H. Bailey and M. López de Prado. The sharpe ratio efficient frontier. *The Journal of Risk*, 15 (2):3–44, 2012.

D. H. Bailey and M. López de Prado. The deflated sharpe ratio: Correcting for selection bias, backtest overfitting, and non-normality. *Journal of Portfolio Management*, 40(5):94–107, 2014.

D. H. Bailey, J. M. Borwein, M. López de Prado, and Q. J. Zhu. Pseudomathematics and financial charlatanism: The effects of backtest over fitting on out-of-sample performance. *Notices of the AMS*, 61(5):458–471, 2014.

D. H. Bailey, J. M. Borwein, A. Salehipour, M. López de Prado, and Q. J. Zhu. Backtest overfitting demonstration tool: An online interface. *Available at SSRN: http://ssrn.com/abstract=2597421 or http://dx.doi.org/10.2139/ssrn.2597421*, 2015.

W. F. Sharpe. The sharpe ratio. *The Journal of Portfolio Management*, 21(1):49–58, 1994.

Question?