Developing & Backtesting Systematic Trading Strategies

Brian G. Peterson

19 November 2014

Backtesting, art or science?

Back-testing. I hate it — it's just optimizing over history. You never see a bad back-test. Ever. In any strategy. - Josh Diedesch (2014) CalSTRS

Every trading system is in some form an optimization. - Emilio Tomasini (2009)

Moving Beyond Assumptions

Many system developers consider "*I hypothesize that this strategy idea will make money*" to be adequate.

- understand your business constraints and objectives
- build a hypothesis for the system
- build the system in pieces
- test the system in pieces
- measure how likely it is that you've overfit

Constraints and Objectives

Constraints

- capital available
- products you can trade
- execution platform

Benchmarks

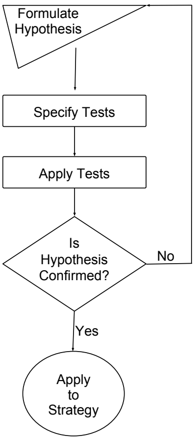
- published or synthetic?
- what are the limitations?
- are you held to it, or just measured against it?

Objectives

- formulate objectives for testability
- make sure they reflect your real business goals

Building a Hypothesis

Essentially, all models are wrong, but some are useful. - George Box (1987)

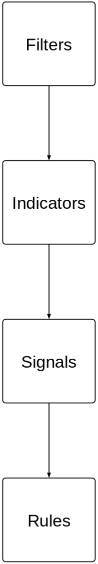


To create a testable idea (a hypothesis), we need to:

- formulate a declarative conjecture
- make sure the conjecture is predictive
- define the expected outcome
- describe means of verifying (testing) the outcome

Building Blocks

Separating the strategy into components aids testing, and increases productivity.



Definitions

Filters

- select the instruments to trade
- may be part of the hypothesis
- categorize market characteristics that are favorable to the strategy

Indicators

- quantitative values derived from market data

Signals

- describe the interaction between filters, market data, and indicators
- can be viewed as a prediction at a point in time

Rules

- make path-dependent actionable decisions

Test the System in Pieces, or, How to Screw Up Less

Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise. - John Tukey (1962) p. 13

Fail quickly, think deeply, or both?

No matter how beautiful your theory, no matter how clever you are or what your name is, if it disagrees with experiment, it's wrong. - Richard P. Feynman (1965)

Things to Watch Out For, or, Types of Overfitting

Look Ahead Bias

- directly using knowledge of future events

Data Mining Bias

- caused by testing multiple configurations and parameters over multiple runs, with adjustments between backtest runs
- exhaustive searches may or may not introduce biases

Data Snooping

- knowledge of the data set can contaminate your choices
- making changes after failures without having strong experimental design

Measuring Indicators

A good indicator is describing some measurable aspect of reality: a theoretical “fair value” price, or the impact of a factor on that price, or turning points of the series, or slope.

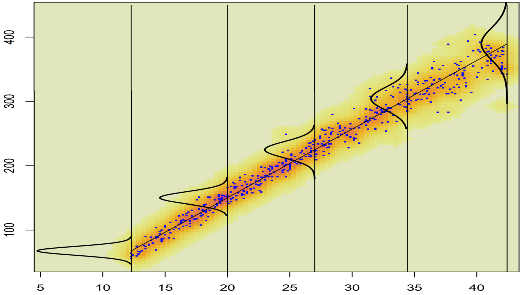
- hypothesis and tests for the indicator
- custom ‘perfect foresight’ models
- lessons from signal processing: symmetric filters

If your indicator doesn’t have testable information content, throw it out and start over.

Measuring Signals

Signals make predictions; all the literature on forecasting is applicable:

- mean squared forecast error, BIC, etc.
- box plots or additive models for forward expectations
- “revealed performance” approach of Racine and Parmeter (2009)
- re-evaluate assumptions about the method of action of the strategy
- detect information bias or luck before moving on

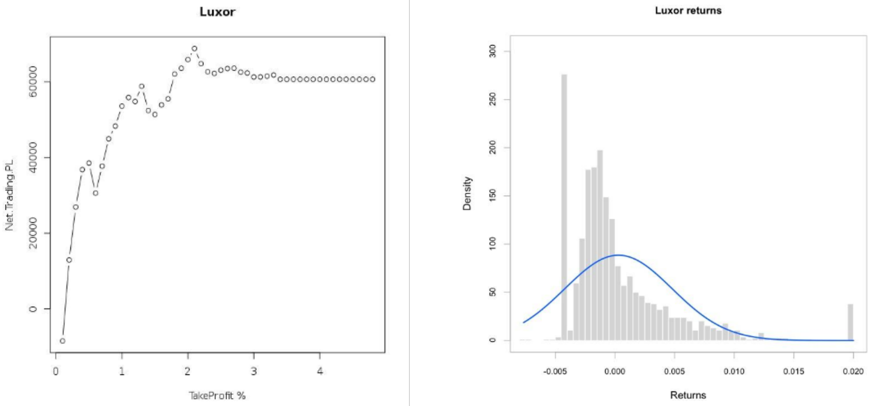


Measuring Rules

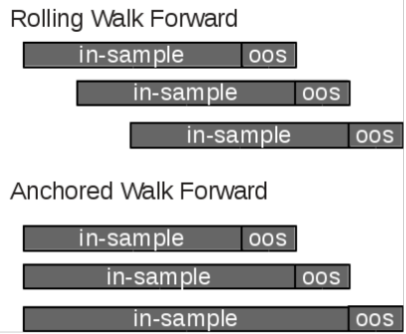
If your signal process doesn't have predictive power, stop now.

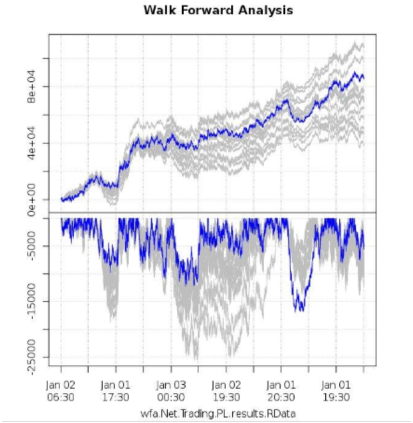
- rules should refine the way the strategy 'listens' to signals
- entries may be passive or aggressive, or may level or pyramid into a position
- exits may have their own signal process, or may be empirical
- risk rules should be added near the end, for empirical 'stops' or to meet business constraints

Parameter Optimization



Walk Forward





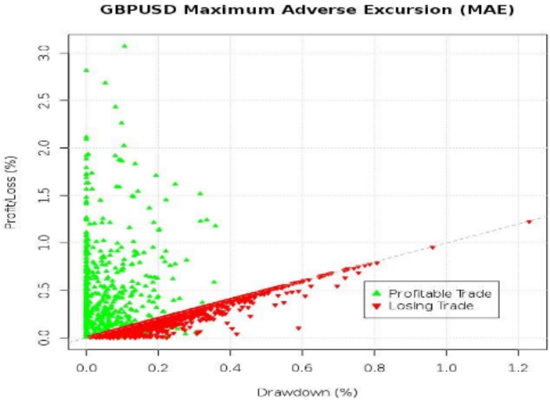
Proper formulation of your business objective is critical to results.

Beware of Rule Burden

- having too many rules is an invitation to overfitting
- adding rules after being disappointed in backtest results is almost certainly an exercise in overfitting (data snooping)
- strategies with fewer rules are more likely to be robust out of sample

Measuring the Whole System

Net profit as a sole evaluation method ignores many of the characteristics important to this decision. - Robert Pardo (2008)



Using Trade Statistics

All trading and backtesting platforms (should) provide trade statistics:

- number of trades w/ gross and net P&L
- mean/median, standard deviation of trading P&L per trade
- percent of positive/negative trades
- Profit Factor : absolute value ratio of gross profits over gross losses
- Drawdown statistics
- start-trade drawdown (Fitschen 2013, 185)
- win/loss ratios of winning over losing trade P&L (total/mean/median)

Dangers of aggregate statistics:

- hiding the most common outcomes
- focusing on extremes
- not enough trades or history for validity
- colinearities of overlapping "trades"

Using Returns

- Returns create a *standard* mechanism for comparing multiple strategies or managers
- Choice of the denominator matters

Sample Analyses:

- tail risk measures
- volatility analysis
- factor analysis / factor model monte carlo
- style analysis
- comparing strategies in return space
- applicability to asset allocation

Asset Allocation

- we tend to do asset allocation studies only after strategies are in production.
- backtests are most often done on 1-lots, and initial scaling is done ad-hoc.
- strategy daily returns become returns of a synthetic asset (the strategy) as inputs to optimization
- optimizer should use your business objectives as the portfolio objective

Did we over do it?

A big computer, a complex algorithm and a long time does not equal science. - Robert Gentleman

Detecting Backtest Overfitting

- **White's Reality Check** : from White (2000) and Hansen (2005)
- **k-fold cross validation** : improves single hold-out model by randomly dividing the sample of size T into sequential sub-samples of size T/k.(Hastie, Tibshirani, and Friedman 2009)
- **CSCV sampling** (combinatorially symmetric cross validation): "generate $S/2$ testing sets of size $T/2$ by recombining the S slices of the overall sample of size T". (Bailey et al. 2014, p.17)
- **Multiple Hypothesis Testing** looks at Type I vs Type II error in evaluating backtests and look at appropriate haircuts based on this. (Harvey and Liu 2013a 2013b 2014)

Conclusion & Questions

- understand the business context you operate in
  - constraints
  - benchmarks
  - objectives
- separate the components of the strategy
- construct testable hypotheses at each step of the process
- evaluate the components separately
- test yourself often

Thanks

Thank You for Your Attention

Thanks to my team, and my family, who make it possible.

brian@braverock.com

Code to apply the techniques discussed here may be found in the **R** *quantstrat* and *PerformanceAnalytics* packages. (Peterson et al. 2014; Peterson and Carl 2014)

Resources

Bailey, David H, Jonathan M Borwein, Marcos López de Prado, and Qiji Jim Zhu. 2014. "The Probability of Backtest Overfitting." http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2326253.

Box, George E.P., and Norman R. Draper. 1987. *Empirical Model-Building and Response Surfaces.* John Wiley & Sons.

Diedesch, Josh. 2014. "2014 Forty Under Forty." *Chief Investment Officer.* California State Teachers' Retirement System. http://www.ai-cio.com/Forty_Under_Forty_2014.aspx?page=9.

Feynman, Richard P, Robert B Leighton, Matthew Sands, and EM Hafner. 1965. *The Feynman Lectures on Physics*. Vols. 1-3.

Fitschen, Keith. 2013. *Building Reliable Trading Systems: Tradable Strategies That Perform as They Backtest and Meet Your Risk-Reward Goals*. John Wiley & Sons, Inc.

Hansen, Peter R. 2005. "A Test for Superior Predictive Ability." *Journal of Business and Economic Statistics*.

Harvey, Campbell R., and Yan Liu. 2013a. "Backtesting." *SSRN*. http://ssrn.com/abstract=2345489.

———. 2013b. "Multiple Testing in Economics." *SSRN*. http://ssrn.com/abstract=2358214.

———. 2014. "Evaluating Trading Strategies." *SSRN*. http://ssrn.com/abstract=2474755.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition*. Springer.

Pardo, Robert. 2008. *The Evaluation and Optimization of Trading Strategies, Second Edition*. John Wiley & Sons.

Peterson, Brian G., and Peter Carl. 2014. *PerformanceAnalytics: Econometric Tools for Performance and Risk Analysis: R Package Version 1.4.3541*. http://CRAN.R-project.org/package=PerformanceAnalytics.

Peterson, Brian G., Joshua Ulrich, Jan Humme, and Peter Carl. 2014. *quantstrat: Quantitative Strategy Model Framework: R Package Version 0.9.1632*. http://r-forge.r-project.org/projects/blotter/.

Racine, Jeffrey S, and Christopher F Parmeter. 2009. "Data-Driven Model Evaluation: a Test for Revealed Performance." *Mac Master University*. https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=FEMES09&paper_id=152.

Tomasini, Emilio, and Urban Jaekle. 2009. "Trading Systems: A New Approach to System Development and Portfolio Optimisation."

Tukey, John W. 1962. "The Future of Data Analysis." *The Annals of Mathematical Statistics*. JSTOR, 1–67. http://projecteuclid.org/euclid.aoms/1177704711.

White, Halbert L. 2000. "System and Method for Testing Prediction Models and/or Entities." Google Patents. http://www.google.com/patents/US6088676.