Recent paper

http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2745220

All that Glitters Is Not Gold: Comparing Backtest and Out-of-Sample Performance on a Large Cohort of Trading Algorithms

Thomas Wiecki
Quantopian Inc

Andrew Campbell
Quantopian Inc.

Justin Lent
Quantopian Inc

Jessica Stauth
Quantopian Inc

March 9, 2016

Abstract:
When automated trading strategies are developed and evaluated using backtests on historical pricing data, there exists a tendency to overfit to the past. Using a unique dataset of 888 algorithmic trading strategies developed and backtested on the Quantopian platform with at least 6 months of out-of-sample performance, we study the prevalence and impact of backtest overfitting. Specifically, we find that commonly reported backtest evaluation metrics like the Sharpe ratio offer little value in predicting out of sample performance ($R^2 < 0.025$). In contrast, higher order moments, like volatility and maximum drawdown, as well as portfolio construction features, like hedging, show significant predictive value of relevance to quantitative finance practitioners. Moreover, in line with prior theoretical considerations, we find empirical evidence of overfitting – the more backtesting a quant has done for a strategy, the larger the discrepancy between backtest and out-of-sample performance. Finally, we show that by training non-linear machine learning classifiers on a variety of features that describe backtest behavior, out-of-sample performance can be predicted at a much higher accuracy ($R^2 = 0.17$) on hold-out data compared to using linear, univariate features. A portfolio constructed on predictions on hold-out data performed significantly better out-of-sample than one constructed from algorithms with the highest backtest Sharpe ratios.

Number of Pages in PDF File: 19