



Algorithmic Trading
Session 4
Trade Signal Generation II
Backtesting

Oliver Steinki, CFA, FRM

Outline

- ▣ **Introduction**
- ▣ **Backtesting**
- ▣ **Common Pitfalls of Backtesting**
- ▣ **Statistical Significance of Backtesting**
- ▣ **Summary and Questions**
- ▣ **Sources**

Introduction

Where Do We Stand in the Algo Prop Trading Framework?



- ❑ As we have seen, algorithmic proprietary trading strategies can be broken down into three subsequent steps: Signal Generation, Trade Implementation and Performance Analysis
- ❑ The first step, **Signal Generation**, defines when and how to trade. For example, in a moving average strategy, the crossing of the shorter running moving average over the longer running moving average triggers when to trade. Next to long and short, the signal can also be neutral (do nothing). Using moving averages to generate long/short trading signals is an example choice of how to trade
- ❑ Sessions 3 – 6 deal with the question of deciding when and how to trade
 - **Session 3**: Finding Suitable Trading Strategies and Avoiding Common Pitfalls
 - **Today's Session 4**: Backtesting
 - **Session 5**: Mean Reversion Strategies
 - **Session 6**: Momentum Strategies

Introduction

Backtesting

- ❑ Signal Generation describes the process of **deciding when and how to trade**. Backtesting is the process of feeding historical data to your trading strategy to see how it would have performed. A key difference between a traditional investment management process and an algorithmic trading process is the possibility to do so
- ❑ However, if one backtests a strategy without taking care to avoid common backtesting pitfalls, the whole backtesting procedure will be useless. Or worse - it might be misleading and may cause significant financial losses
- ❑ Since backtesting typically involves the computation of an expected return and other statistical measures of the performance of a strategy, it is reasonable to question the statistical significance of these numbers. We will discuss the general way to estimate statistical significance using the methodologies of hypothesis testing and Monte Carlo simulations. In general, the more round trip trades there are in the backtest, the higher will be the statistical significance
- ❑ But even if a backtest is done correctly without pitfalls and with high statistical significance, it doesn't necessarily mean that it is predictive of future returns. Regime shifts can spoil everything, and a few important historical examples will be highlighted

Backtesting

Why is Backtesting Important?

- ❑ Backtesting is the process of feeding historical data to your trading strategy to see how it would have performed. The idea is that the backtested performance of a strategy tells us what to expect as future performance
- ❑ Whether you have developed a strategy from scratch or you read about a strategy and are sure that the published results are true, it is still imperative that you independently backtest the strategy. There are several reasons to do so:
 - ❑ The profitability of a strategy often depends sensitively on the details of implementation, e.g. which prices (bid, ask, last traded) to use for signal generation (trigger) and as entry /exit points (execution)
 - ❑ Only if we have implemented the backtest ourselves, we can analyze every little detail and weakness of the strategy. Hence, by backtesting a strategy ourselves, we can find ways to refine and improve the strategy, hence to improve its risk/reward ratio
 - ❑ Backtesting a published strategy allows you to conduct true out-of-sample testing in the period following publication. If that out-of-sample performance proves poor, then one has to be concerned that the strategy may have worked only on a limited data set
 - ❑ The full list of potential backtesting pitfalls is quite long, but we will look at a few common mistakes on the next pages

Common Pitfalls of Backtesting

Look-Ahead Bias

- ❑ As the name suggests, look-ahead bias describes a strategy which uses data at time t_1 to determine a trading signal at t_0 . A common example of look-ahead bias is a strategy that uses the high or low of a trading day as a trigger. This assumption is not realistic as we only know the high / low after market close
- ❑ Look-ahead bias is essentially a programming error and can infect only a backtest program but not a live trading program because there is no way a live trading program can obtain future information
- ❑ Hence, If your backtesting and live trading programs are one and the same, and the only difference between backtesting versus live trading is what kind of data you are feeding into the program, you're pretty safe to avoid look-ahead bias

Common Pitfalls of Backtesting

Data Snooping Bias – In Sample Over-Optimization

- ❑ If you build a trading strategy that has 100 parameters, it is very likely that you can optimize those parameters in a way that the historical performance looks amazing. It is also very likely that the future performance of this strategy will not at all look like this over-optimized historical performance
- ❑ In general, the more rules a strategy has, and the more parameters the model has to optimize, the more likely it is to suffer from data-snooping bias
- ❑ The way to detect data-snooping bias is easy: We should test the model on out-of-sample data and reject a model that doesn't pass the out-of sample test
- ❑ Cross-validation is probably the best way to avoid data snooping. That is, you should select a number of different subsets of the data for training and tweaking your model and, more important, making sure that the model performs well on these different subsets. One reason why one prefers models with high risk/return ratios and short maximum drawdown durations is that this is an indirect way to ensure that the model will pass the cross-validation test: the only subsets where the model will fail the test are those rare drawdown periods

Common Pitfalls of Backtesting

Survivorship Bias

- ❑ A historical database of asset prices such as stocks that does not include stocks that have disappeared due to bankruptcies, delistings, mergers, or acquisitions suffer from the so-called survivorship bias, because only “survivors” of those often unpleasant events remain in the database
- ❑ Same problem applies to mutual fund or hedge fund databases that do not include funds that went out of business, usually due to negative performance
- ❑ Survivorship bias is especially applicable to value strategies, e.g. investment concepts that buy stocks that seem to be cheap. Some stocks were cheap because the companies were going bankrupt shortly. So if your strategy includes only those cases when the stocks were very cheap but eventually survived (and maybe prospered) and neglects those cases where the stocks finally did get delisted, the backtest performance will be much better than what a trader would actually have suffered at that time

Common Pitfalls of Backtesting

Stock Splits and Dividend Adjustments

- ❑ Whenever a company's stock has an N -to-1 split, the stock price will be divided by N times. However, if you own a number of shares of that company's stock before the split, you will own N times as many shares after the split, so there is in fact no change in the total market value
- ❑ However, in a backtesting environment, we often consider only at the price series to determine our trading signals, not the market-value series of some hypothetical account. So unless we back-adjust the prices before the ex-date of the split by dividing them by N , we will see a sudden drop in price on the ex-date, and that might trigger some erroneous trading signals
- ❑ Similarly, when a company pays a cash (or stock) dividend of $\$d$ per share, the stock price will also go down by $\$d$ (absent other market movements). That is because if you own that stock before the dividend ex-date, you will get cash (or stock) distributions in your brokerage account, so again there should be no change in the total market value
- ❑ If you do not back-adjust the historical price series prior to the ex-date, the sudden drop in price may also trigger an erroneous trading signal

Common Pitfalls of Backtesting

Trading Venue Dependency and Short Sale Constraints

- ❑ Most larger stocks are listed on multiple exchanges, electronic communication networks (ECNs) and dark pools. The historical last daily price might have occurred on any of those trading venues. However, if you enter a market on open (MOO) or market on close order (MOC), this order will be routed to the primary exchange only. Hence, your backtested performance based on open or close might be different to a live trading performance based on MOO/MOC
- ❑ Foreign Exchange (FX) markets are even more fragmented and there is no rule that says a trade executed at one venue has to be at the best bid or ask across all the different FX venues
- ❑ A stock-trading model that involves shorting stocks assumes that those stocks can be shorted, but often there are difficulties in shorting some stocks. This might either be due to limited availability of your broker to locate such stocks or due to regulatory reasons. For example, many European countries and the USA prohibited the short sale of financial stocks during the financial crisis

Common Pitfalls of Backtesting

Futures Continuous Contracts and Futures Close vs. Settlement Prices

- Futures contracts have expiry dates, so a trading strategy on, say, volatility futures, is really a trading strategy on many different contracts. Usually, the strategy applies to front-month contracts. Which contract is the “front month” depends on exactly when you plan to “roll over” to the next month; that is, when you plan to sell the current front contract and buy the contract with the next nearest expiration date. Hence, when choosing a data vendor for historical futures prices, you must understand exactly how they have dealt with the back-adjustment issue, as it certainly impacts your backtest



- The daily closing price of a futures contract provided by a data vendor is usually the settlement price, not the last traded price of the contract during that day. Note that a futures contract will have a settlement price each day (determined by the exchange), even if the contract has not traded at all that day. And if the contract has traded, the settlement price is in general different from the last traded price. Most historical data vendors provide the settlement price as the daily closing price, which can not be replicated via MOC orders in a live strategy environment

Statistical Significance of Backtesting

Hypothesis Testing

- In any backtest, we face the problem of finite sample size: Whatever statistical measures we compute, such as average returns or maximum drawdowns, are subject to randomness. In other words, we may just be lucky that our strategy happened to be profitable in a small data sample. Statisticians have developed a general methodology called *hypothesis testing* to address this issue. The hypothesis testing framework applied to backtesting follows these steps:
 1. Based on a backtest on some finite sample of data, we compute a certain statistical measure called the *test statistic*. For concreteness, let's say the test statistic is the average daily return of a trading strategy in that period
 2. We suppose that the true average daily return based on an infinite data set is actually zero. This supposition is called the *null hypothesis*
 3. We suppose that the probability distribution of daily returns is known. This probability distribution has a zero mean, based on the null hypothesis. We describe later how we determine this probability distribution
 4. Based on this null hypothesis probability distribution, we compute the probability p that the average daily returns will be at least as large as the observed value in the backtest (or, for a general test statistic, as extreme, allowing for the possibility of a negative test statistic). This probability p is called the *p-value*, and if it is very small (let's say smaller than 0.01), that means we can “reject the null hypothesis,” and conclude that the backtested average daily return is statistically significant and not equal to 0

Statistical Significance of Backtesting

Three Ways to Determine the Probability Distribution

- Step 3 of the described Hypothesis Testing Framework requires the most thought. How do we determine the probability distribution under the null hypothesis? There are three ways to do so:
 1. We assume the daily returns follow a standard parametric distribution such as the Gaussian one. If we do this, it is clear that if the backtest has a high Sharpe ratio, it would be very easy for us to reject the null hypothesis. This is because the standard test statistic for a Gaussian distribution is none other than the average divided by the standard deviation and multiplied by the square root of the number of data points
 2. Another way to estimate the probability distribution of the null hypothesis is to use Monte Carlo methods to generate simulated historical price data and feed these simulated data into our strategy to determine the empirical probability distribution of profits. If we do so with the same first moments and the same length as the actual price data, and run the trading strategy over all these simulated price series, we can find out in what fraction p of these price series are the average returns greater than or equal to the backtest return. Ideally, p will be small, which allows us to reject the null hypothesis.
 3. Andrew Lo suggested a third way to estimate the probability distribution: instead of generating simulated price data, we generate sets of simulated trades, with the constraint that the number of long and short *entry* trades is the same as in the backtest, and with the same average holding period for the trades. These trades are distributed randomly over the actual historical price series. We then measure what fraction of such sets of trades has average return greater than or equal to the backtest average return

Statistical Significance of Backtesting

Will a Backtest Be Predictive of Future Returns?

- ❑ Even if we manage to avoid all the common backtesting pitfalls outlined earlier and there are enough trades to ensure statistical significance of the backtest, the predictive power of any backtest rests on the central assumption that the statistical properties of the price series are unchanging, so that the trading rules that were profitable in the past will be profitable in the future. This assumption has often been invalidated in the past:
 - ❑ Decimalization of U.S. stock quotes on April 9, 2001
 - ❑ The 2008 financial crisis that induced a subsequent 50 percent collapse of average daily trading volumes. Retail trading and ownership of common stock was particularly reduced. This has led to decreasing average volatility of the markets, but with increasing frequency of sudden outbursts such as that which occurred during the flash crash in May 2010 and the U.S. federal debt credit rating downgrade in August 2011
 - ❑ The same 2008 financial crisis, which also initiated a multiyear bear market in momentum strategies
 - ❑ The removal of the old uptick rule for short sales in June 2007 and the reinstatement of the new Alternative Uptick Rule in 2010

Summary and Questions

- ❑ Backtesting is useless if it is not predictive of future performance of a strategy, but pitfalls in backtesting will decrease its predictive power
 - ❑ Make sure to avoid the following backtesting pitfalls: look-ahead bias, data-snooping and survivorship bias. Make sure your data is adjusted for stock splits and dividends. It should also take into account trading venue dependency and short sale constraints. Furthermore, roll returns and differences between closing and settlement prices need to be incorporated in your backtesting framework
 - ❑ Make sure your backtested results are statistically significant. Use one of three described ways to determine the probability distribution under the null hypothesis. If possible, use data-validation
 - ❑ Even if you avoid all common pitfalls and ensure that your results are statistically significant, regime shifts could still make your strategy unprofitable in a live trading environment.
-
- ❑ Questions?

Sources

- ▣ Quantitative Trading: How to Build Your Own Algorithmic Trading Business by Ernest Chan
- ▣ Algorithmic Trading: Winning Strategies and Their Rationale by Ernest Chan
- ▣ The Mathematics of Money Management: Risk Analysis Techniques for Traders by Ralph Vince