**Hypothesis Based Strategy Development**

**Introduction**

Developing a robust trading system is extremely difficult. There are many traps that lie in wait to trip us up. The chief of which is overfitting.

Overfitting typically comes in the form of the following cycle:

* create a system you can backtest (typically with your selection of favorite indicators, each with several parameters)
* Run a backtest
* Get results
* Change the parameters
* Repeat until you find a parameter set that produces a desired result.

This was and is a major problem. The biggest issue is you most likely have not discovered a new alpha, or even robustly took advantage of a known alpha. You have optimized your system on the past which will never repeat itself in that manner again.

One simple way to test how overfit your strategy is, is to simply reverse the sequence the of your price data, as suggested by [Marcos Lopez de Prado](mailto:mldp@quantresearch.org). Most strategies break down during this simple robustness check.

Don’t feel bad if this is you, unfortunately we all fall prey to this at times, even me.

In this lecture I will demonstrate a workflow that will help you reduce the odds that your strategy is overfit.

**Start**

The number one thing you can do to reduce the likelihood of overfitting is to start with a hypothesis about the market that you believe can be exploited.

Examples:

1. I think that weekly moving average crossovers are good entries for stocks I wish to hold for several months to years.
2. I think that stocks revert to the 30 day average of OHLC price.
3. I think that I should always be long the market when VIX is below it’s 252 day rolling average.
4. I think intraday If a stock is above VWAP and retraces back to VWAP that’s a good time to enter for a day trade.

What do these examples have in common?

They are all falsifiable. This means you can set up an experiment and test the validity of these beliefs.

For example, using #1, one approach is to first get data for about 30 stocks in different sectors with different performance characteristics over the years. Then test a range of moving average crossovers such that after the crossover you calculate the 1, 3, 6, 9, and 12 month returns. You could then compare those returns to the market benchmark on an absolute and risk adjusted basis to determine if your hypothesis is false. An additional method to test your hypothesis is to compare it against random entries in those same stocks for the same holding periods and make sure your strategy is not simply benefitting from positive drift in those stocks.

Assuming these tests did not invalidate your hypothesis, now you can move forward in the testing process. Now you might want to confirm the robustness of your parameter set, and make sure there is not one single crossover that accounts for the strategy’s success.

There are many tests and slight variations on the workflow but they are typically based on the following steps.

1. Hypothesis
2. Gather preliminary data
3. Set up experiments or tests to falsify your hypothesis
   1. Does the strategy make money on an absolute basis?
   2. Does it work for a broad parameter set?
   3. How does it compare to a benchmark on absolute and risk adjusted?
   4. Is it better than random?
   5. If it passes, move forward to backtest design, if it fails start over with a new hypothesis.
4. Backtest design
   1. This is to decide how the strategy is going to handle:
      1. Rebalancing
      2. Allocation weights or dollar amounts
      3. Exit criteria
   2. You also want to design the strategy in a way that allows you to replicate your results first.
5. Backtest
   1. Simplest backtest possible to replicate your results.
   2. You may test without commissions/slippage models first to see if your results are the same.
   3. You must test with transaction costs also.
   4. Test on a broader set of assets (where possible) over a longer period of time (where possible).
6. Evaluate
7. Refine - this is a possible point of overfitting
8. Repeat
   1. If the strategy has moved to production you will start over again.
   2. Most commonly your strategy will have a major weakness and/or fail robustness tests at which point you start the process over again.

**HYPOTHESIS**

In this lecture I will be exploring the following hypothesis:

It is my belief that stocks mean revert after trending in the same direction after a certain number of days and that a strategy can be designed to exploit that mean reversion.

**EXPERIMENT**

The way we will test this hypothesis is first by gathering data of several stocks and ETFs. We will compute the up and down streaks, and through analysis we will choose the streaks that occur 5% of the time or less to investigate this hypothesis. The threshold is arbitrary at this point but I wanted to test the mean reversion of rare events (this could also be investigated further).

Next we will setup an event based analysis. Each time a streak that meets our requirements is triggered we will log that event for further aggregate analysis. So a streak of positive days will be classified as short positions and a streak of negative days will be classified as long positions.

We will look at the individual and collective cumulative returns of the long and short sides. Then we will look at the relative returns to identify if a long-short strategy is more appropriate.

We will also examine the returns relative to the holding period.

If a strategy cannot be constructed then the hypothesis will be rejected.

**BACKTEST DESIGN**

Based on the results of our experimental analysis we will make the strategy extremely simple to execute. The adjustable parameters will be the streak length, the holding period and the number of symbols to be returned from the coarse filter.

The coarse filter will return stocks with the highest dollar volume.

We will use the same parameters as our analysis.

**EVALUATE**

We analyze the results of our experiment and backtest and discuss future works.

**REFINE**

Are there logical adjustments that can be made to improve the strategy while minimizing overfitting to past data?

Are there additional components we can analyze in the experiment phase, reviewing the hypothesis relative to industry/sector classifications or SIMPLE volatility filters?

**REPEAT**

If it moves to production you will likely have to keep researching and refining the edge. If it fails you will have to start over again.