Understanding the 2014 Wagner Award Winning Paper in Simple Terms

Introduction

This year's Wagner Award was presented to Mr. Dave Walton, co-founder of StatisTrade, for the paper entitled, "Know Your System! – Turning Data Mining from Bias to Benefit through System Parameter Permutation." While the title may seem daunting, the main concepts and how they can help you are easy to comprehend.

A key to understanding the contribution of Mr. Walton's paper is recognizing an often overlooked and misunderstood reason why systems fail to perform like they did during a back-test: Data Mining Bias (DMB).

A Beautiful Equity Curve Isn't Always Pretty When You Look Inside

Figure 1 shows thousands of trading system equity curves. Each curve was generated by taking a baseline trading system and varying system parameters across multiple back-tests (e.g., # of days for a simple moving average of prices). Notice that some equity curves show returns of up to 400%, while others show returns of -70%, a huge range.

Figure 1. Equity Curves for System Variations

To understand how this relates to DMB and why systems fail to perform, let's walk through a generalized example of how developers often create trading systems.
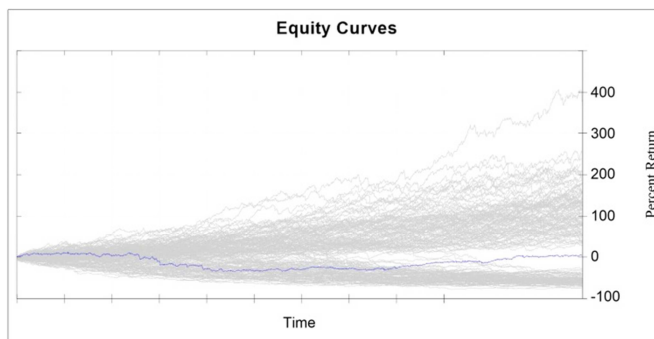
Step 1: Concept. The system developer reads about a trading concept with some historically good results. For example, using a moving average to determine buy and sell signals, or a rotational strategy that buys instruments with high recent momentum.

Step 2: Development. The developer builds the system rules using one of many inexpensive back-testing software tools, adds his own rules, and tests thousands, hundreds of thousands, or even millions of possible variations to get the rules and parameter values "just right." For sake of this example, we'll assume this developer tests an average of 5,000 system variations a day, easily do-able with today's software. The developer works diligently for 3 months exploring system variations. At the end of the period, the developer has evaluated about 5,000 variations a day x 90 days = 450,000 combinations. Look again at the results in **Figure 1**, and imagine each equity curve shows the results from one of these 450,000 system variations.

Step 3: Validation. The developer takes only the best system from Step 2 and validates it using one or several of the methods discussed in the section below. If a given system doesn't hold up well during this validation, the developer returns to Step 2, tweaks the system rules and parameters, and repeats the cycle until he finds a system does successfully pass the validation.

Figure 2. Focus on the "One in a Million" Equity Curve

What has happened here is Data Mining Bias in action.

The system developer has ignored 449,999 systems and focused exclusively on the remaining 1 system showing fantastic results. In other words, the developer 'mined' the data, ignoring the results that failed to produce the results he wanted, and fixated on the single result he was looking for. Why is that a problem? Because it **exponentially increases the chances that the results are due to luck** and not a market edge. Stated another way,
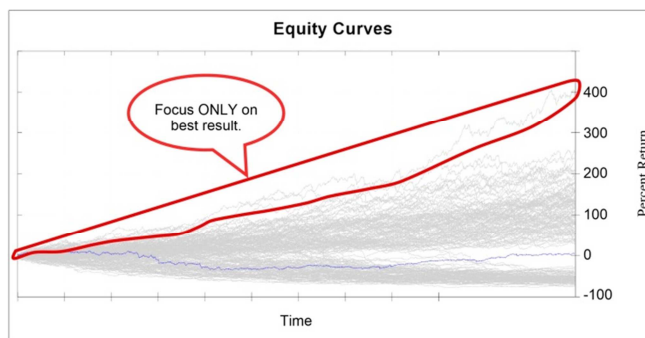
believing the one "best case" result is due to a real market edge is like assuming the single grand-prize winner of a random lottery with 450,000 participants won due to skill.

Monte Carlo and Other Common Approaches Don't Address the Problem

People often turn to Monte Carlo simulation or Out Of Sample (OOS) testing to validate trading systems. Both of these methods can be useful when used under the right conditions. As shown below, though, neither is well suited to address problems caused by DMB.

For example, let's take the "best result" system highlighted in **Figure 2** and run it through a Monte Carlo simulation. After randomizing the data in 10,000 trials, the results look something like **Figure 3** at the right. Judging by the outcome, you might think you have even more evidence to support the belief the system is worth trading. After all, the results at a "95% level of confidence" are great.

Figure 3. Misuse of Monte Carlo Simulation

So why do the results look so good? Monte Carlo is only effective when the input data – in this case, our "best result" trading system – is representative of the entire population, and it is very possible the population is better represented by the 449,999 systems the developer discarded. In other words, the root cause of DMB means Monte Carlo is ill suited to address it.

Out Of Sample testing is another popular means of validating a trading system. While the method is useful, it comes at a steep price – making crucial data unavailable and requiring rigorous adherence to a policy of only using OOS data once, which is harder to follow than it sounds. As a result of these complexities, developers often fail to use OOS effectively and the resulting "validation" is often just as misleading as the Monte Carlo results above.

SPP to the Rescue

The solution proposed in Mr. Walton's paper, called System Parameter Permutation or SPP, is valuable for five primary reasons: (1) it eliminates DMB, meaning all of the problems mentioned above; (2) it overcomes limitations of other system validation techniques like the inability to model real-world system interactions, which can be a significant issue with Monte Carlo simulation; (3) it estimates the system's "real edge" versus random market noise; (4) it provides both an expected range of system performance as well as a statistically sound means of determining a "worst-case" point to stop trading; and (5) it enables these advantages to be realized without a major investment in custom software (i.e., the process can largely be performed using commercially available back-testing software with relatively minor changes).

Space constraints prevent a detailed explanation of the potentially revolutionary advantages these benefits provide to those who trade and those who develop systems. On our web site, www.statistrade.com, we plan to publish a series of articles explaining each benefit in an easy-to-understand format.

How Do I Take Advantage of the Process?

Here are a few ways to put the SPP process to work for you.

1. **Read the Paper and Do It Yourself.** The SPP process is designed to allow people who are familiar with commercial back-testing software to implement it. The required format of the NAAIM paper submission was academic, and the implementation process will vary across software packages, so some expertise with coding and knowledge of basic statistics will be very helpful.
2. **Seek Out Experts.** One of the reasons Mr. Walton co-founded StatisTrade (www.statistrade.com) was to provide fund managers and sophisticated investors access to expertise often only available to hedge funds with full-time quants on staff. To that end, one of the services we offer is executing the process for our customers and providing an easy-to-understand SPP Results Report Card.

System Validation Survey

StatisTrade, co-founded by two NAAIM Wagner Award winners, is a start-up think tank focused on trading system evaluation. Our mission is to provide our clients with unique and critical insights into their trading systems to improve their performance and meet their specific goals. To that end, we employ objective system analyses based on proprietary, statistically-sound tools and processes. **Your feedback helps us provide services of value to you!**

1. What markets do you trade?

   % of allocated capital (totals to 100%)
   a) _____% equities   b) _____% futures   c) _____% options
   d) _____% other: _____

2. What kind of systems do you trade?

   % of all systems traded (totals to 100%)
   a) _____% signal based (capital allocation is dependent on signals)
   b) _____% rotational (portfolio rebalancing; basically always invested)
   c) _____% other: _____

3. How much discretion is involved in your trading?

   % of all systems (totals to 100%)
   a) _____% fully mechanical, no discretion   b) _____% very little discretion
   c) _____% strong discretionary elements

4. Annual spending on trading systems per year?

   a) external suppliers: $_____   If unsure of internal cost
   b) internal resources: $_____   rates, use $100/hr.

5. What factors go into your decision to buy/apply a system or signal service?

   Identify all that apply [scale: 1 (never) - 5 (always)]
   a) _____ backtest results   b) _____ forward test results/track record
   c) _____ information on the design process
   d) _____ statistical evaluations
   e) _____ other: _____

6. How important are these factors for your decision to buy/apply a system or signal service?

   Identify all that apply [scale: 1 (irrelevant) - 5 (critical)]
   a) _____ backtest results
   b) _____ forward test results/track record
   c) _____ information on the design process
   d) _____ statistical evaluations   e) _____ other: _____

7. Once a system or signal service is used, do you validate its ongoing performance (based on your actual results)?

   ❑ I'd love to, but don't know how.
   ❑ I make a discretionary decision to stop trading when results are bad
   ❑ I have a clearly defined numerical thresholds based on statistical methods to determine when to stop trading a system
   ❑ other: _____

8. What products or services are not effectively offered today but would provide significant value to your trading?

   _____
   _____
   _____

   How may we contact you (optional)?

   name: _____   email: _____

   phone: _____

We appreciate your feedback!
Please return this sheet to the NAAIM conference registration desk.
Survey is also available on-line: http://www.statistrade.com/market-survey.html