



# Backtesting with Integrity

**Tools, whether a chainsaw, a backhoe, or a math formula, can be incredibly useful, relevant and powerful if properly used, or destructive, dangerous or ineffective if misused and not respected.**

## Introduction

Backtesting is an important tool utilized by investment model and strategy developers. Investors, advisers, and regulatory authorities are rightfully skeptical of backtests. Admittedly, the backtesting process, which produces hypothetical data or results, is inherently flawed. Backtests are intended to show investment decisions that would have been made had a strategy been utilized in the past. Without adhering to best practices, backtests can lead to misleading results.

A model derived with the retroactive use of backtested performance results are suspect and potentially dangerous. Backtesting is performed with the benefit of hindsight, which makes the results unreliable and potentially not a reflection of a strategy's performance that would have been obtained had live dollars been invested in the same manner as the tested strategy. Many risks, which are outlined and discussed in this paper, cause backtested data to be dangerous. If constructed, utilized and disclosed properly, backtesting can be useful when constructing and evaluating new investment strategies, including ones customized for a specific client.

At Newfound, backtesting is a necessary component in our strategy development process and we constantly strive to evolve our process to more adequately address its inherent issues. However, all too frequently, we see strategies being presented in the industry that succumb to these issues. We approach the use of backtested data first by recognizing and addressing its risks and drawbacks. We spend considerable time ensuring that we are not placing undue reliance upon its results, but considering them as one of many metrics in analyzing a new investment strategy. We also take



proactive steps to inform our clients, both in writing and in discussion, of our use of historical data to construct backtested investment products. We believe transparency, diligence, and open discussions with clients are necessary to ensure that backtesting is used properly.

Newfound also believes that when using backtested data, the audience needs to be considered. Newfound deals primarily with sophisticated institutions and not a retail audience. However, products that we construct may be designed for investment by retail investors. We caution our clients to pay extra attention to the quantitative nature of our investment strategies (and their related risks) and the use of backtested data and performance results in constructing and analyzing the proposed investment. Additional care should be taken with less sophisticated investors.

We believe that by providing investors with more transparency into the potential problems arising from backtesting, they can be equipped to identify those strategies that adhere to backtesting best practices and those that don't. In this paper we categorize risks based on Cognitive, Data, and Process-Driven issues, discuss implications, and provide potential solutions used by Newfound to allow for meaningful and responsible use of the backtest tool.

## **Cognitive Issues**

### **Hindsight Bias**

**Description:** Hindsight bias is a cognitive effect whereby we believe events that occurred in the past are more predictable than they truly were at the time. This bias arises because of our knowledge of what has already transpired, as well as revealing data often becoming more available after the fact.

**The Problem:** As we develop models, we may be influenced by our knowledge of how events transpired and include information or methods that we may not have otherwise had the event not transpired. This may lead to our process or methodology addressing a particular event in history instead of being developed to be robust for future events.

**Solution:** There is a fine balance between an evolution in process derived from new market evidence and a process that is derived to address a specific event in history. To achieve robustness, we should always err on the side of including less information, instead of more, and always be cognizant of why we are including a particular piece of information and how it is used over time. If a piece of information becomes critical only within a single historical environment and for no qualitatively defensible reason, we are likely succumbing to hindsight bias.

### **Confirmation Bias**

**Description:** Confirmation bias is a cognitive effect whereby we bias our search for information, our interpretation of information, or how we remember things based on our beliefs.

**The Problem:** Instead of searching all possible scenarios, we only find supporting evidence for our existing beliefs and are blinded to contrary evidence.

**Solution:** At Newfound, we make sure that there are always parties who are independent to model and strategy development; it is their job to try to identify flaws and problems in the underlying supporting theories, data, and process.



## Data Issues

### Survivorship Bias

**Description:** Survivorship bias is a problem whereby we only look at data from “surviving” sources and mistakenly overlook data from sources that did not survive.

**The Problem:** Often our datasets contain currently listed companies, and therefore backtests are unable to account for how our process would have applied for those businesses that have been acquired or gone bankrupt. Consider, for example, a strategy that was backtested on single name equities but whose dataset did not include Enron. This “survivorship bias” in the dataset could dramatically positively skew results.

**The Solution:** When dealing with a strategy that back tests using single-name securities, it must be ensured that a “point-in-time” database is used.

### Restated Data

**Description:** Companies often restate quarterly financial information, likewise exchanges frequently restate “tick” and pricing information long after the date the event actually transpired. Sometimes the re-statement is more subtle; the CBOE revised their VIX methodology in 2003 and many data-providers restated calculations from the new methodology over those that were actually available from 1990-2003.

**The Problem:** If we are unknowingly using a database with restated data, we may be making investment or trading decisions based on information that was either a) different at the actual time or b) unavailable at the actual time.

**The Solution:** We must ensure that our database of information is “point-in-time,” meaning that it reflects only information that was available for a given date, and that restated information is only made available on the date of restatement.

### Incorrect Data

**Description:** Despite best efforts, information in a dataset is not always accurate. Trade data, including pricing or timing, may be off; financial data may be incorrect; even economic data may be misstated.

**The Problem:** Making decisions based on incorrect data can lead to the misallocation of capital and potential loss.

**The Solution:** When new information is added to the database, its cleanliness should be checked. This can be done statistically by comparing it against related values or by comparing it against another dataset. Sensitivity tests should also be performed to see the potential impact of incorrect data for a given model or strategy.

## Process-Driven Issues

### Look-ahead Bias



**Description:** Look-ahead bias is the effect whereby our model or strategy uses information from the future to make a decision in the past. For example, before 2008, the VIX had never spiked above 45.74%. If in backtesting before 2008 we used a distribution for the VIX that included the high of 80.86%, we would be incorporating future data and succumbing to Look-ahead bias.

**The Problem:** This bias is not one that necessarily comes from a dataset, but like Hindsight Bias, comes from knowing how events transpired. Using information from the future will skew our results positively, making us believe our results are better than they actually should be.

**The Solution:** When running a back test, we must always be cognizant only to use the information that was available at the current point in time that we are testing.

### **Data Snooping Bias**

**Description:** A form of data mining and statistical bias that leads to the discovery of misleading relationships that are found to be statistically significant in the dataset, but ultimately meaningless outside of it.

**The Problem:** If we discover a relationship and do not rule out Data Snooping bias, we could build a model or strategy on a relationship that is guaranteed to not persist into the future.

**The Solution:** At Newfound, we solve this problem philosophically and technically. We solve it philosophically by ensuring that any process that is supported by empirical evidence must also be supported by qualitative insight. We also attempt to solve this problem technically by exploring the discovered relationship over independent datasets (both across assets and time-frames), checking its robustness.

### **Over-optimization**

**Description:** Over-optimization is the problem where a model is too highly calibrated to an in-sample dataset that it is no longer robust or accurate for an out-of-sample dataset.

**The Problem:** Models and strategies are often calibrated to data, but unless steps are taken to test the robustness of the models and strategies on different data sets, they run the risk of failure going forward. Unless future environments perfectly replicate the conditions the model or strategy was calibrated on, the effectiveness of the model or strategy will not persist.

**The Solution:** At Newfound, we take several steps to combat over-optimization. First, we strive to make all of the parameters of our models and strategies adaptive, and analyze the sensitivity of model and strategy performance to parameter changes. We also test our models and strategies on out-of-sample data, testing for robustness on different asset classes and market environments.

### **Selection Bias**

**Description:** Selection bias occurs when the individual elements chosen for a scientific study do not accurately reflect the population we are trying to examine.



**The Problem:** Survivorship bias, which we addressed earlier, is a type of Selection bias known as Sampling bias. Selection bias can also manifest by the time interval we choose to run our analysis. Unless we can be sure that the population of our test data accurately reflects the population of the data we will be using going forward, we run the risk of developing inadequate models.

**The Solution:** At Newfound, we limit the risk of Selection bias by expanding the scope or assets and horizon of time over which our models and strategies must be effective. By only developing more general models, we reduce the risk of Selection bias.

### **Unrealistic Competition Bias**

**Description:** Unrealistic Competition is a bias by which we compare our results against an unfair or unrealistic benchmark, inflating our results.

**The Problem:** Often we compare hypothetical performance in a back test against the performance of a strategy or benchmark that existed at the time. However, unless we are cognizant of the investment constraints due to access, cost, liquidity and investor preference at the time, we can create an unrealistic comparison. For example, allocating more than 10% of a portfolio to Emerging Market equities in the early 2000s would be highly unrealistic for the average investor in that time-period. Therefore, comparing a backtest that allocates more than 10% to Emerging Market equities in that era to a live strategy from that era is an unrealistic comparison.

**The Solution:** In developing models and strategies, we must remain cognizant of investor behavior in the past and either limit our allocations to fall within the behavioral constraints or change our benchmark to be in line with our allocation methodology. Using the same example from above, we may consider changing our benchmark to one with greater structural concentration to Emerging Markets to determine if our performance was due to our investment decisions or just an unrealistically high allocation to an asset class.

### **Conclusion**

In this paper we have discussed multiple risks to consider when performing and evaluating back tests. We explored three types of issues: Cognitive, Data, and Process-Driven. Within each of these categories, we described several biases, explained their impact, and provided the solution that we use at Newfound to combat the issue.



**Newfound Research**  
FINANCIAL INNOVATORS

Newfound Research LLC  
425 Boylston Street, 3rd Floor  
Boston, MA 02116  
Phone: (617) 531 - 9773

**For more information about Newfound Research call us at + 1-617-531-9773 or visit us at [www.thinknewfound.com](http://www.thinknewfound.com) or email us at [info@thinknewfound.com](mailto:info@thinknewfound.com)**

- 
- These materials are proprietary to Newfound Research LLC and may not be reproduced, modified, transmitted, transferred or distributed in any form without the prior written consent of Newfound Research LLC.