Simpson's Paradox

Nastja Bethke[1], Ed Tricker[2]

Abstract
For a given level of risk, the average return is a defining statistic for any investment strategy. Often practitioners of finance as well as investors look beyond overall measures such as the mean, and condition the data before calculating the desired statistic. A frequently used condition is the performance during periods when a given reference asset, or the investment strategy itself, posts positive or negative returns. While this is a meaningful component of assessing a strategy's performance, it should not be looked at in isolation. As we show in this note, it can be quite misleading.

[1] *Senior Quantitative Research Analyst*
[2] *CIO - Quantitative Strategies*

1. The Thoughtful Investor

While we often model the returns of financial markets using normal (or Gaussian) distributions, in reality market dynamics often deviate from such simple models. Characteristics such as the presence of fat tails (increasing the likelihood of extreme returns) are frequently observed. Similarly, the PnL of trading strategies can deviate from 'Gaussianity', leading investors to look at various statistical measures of the PnL distribution beyond the mean and volatility. Natural questions to ask are: *"On days when the strategy makes/loses money, what is its average return?"* Mathematically this translates into a *conditional mean* of the returns, only selecting the subset of returns with the sign of (or in the direction of) interest, before averaging the numbers.

A curious investor asks precisely these questions when presented with the opportunity to invest in either of two strategies, Strategy 1 or Strategy 2. On bad days, Strategy 2 has an average (risk-adjusted) return of -0.71, while on good days it has an average (risk-adjusted) return of 1.08. The figures for Strategy 1 are -1.00 and 1.00, respectively. It seems like an obvious choice: the investor decides to allocate money to Strategy 2.

2. Aggregating Statistics

For completeness, the investor also asks to see the overall return statistics, although they feel that the conditional returns already tell a convincing story. When presented with the aggregate figures, see Table 1, it looks like there must be a mistake. The numbers don't make sense. While Strategy 2 outperforms Strategy 1 when looking at the conditional means, highlighted in bold, the reverse is true when looking at the overall average return. How can Strategy 1 beat Strategy 2 overall, when it gains less on positive return days, and loses more on negative return days?

3. A Closer Look

The reason behind this unintuitive result is revealed when looking at the fraction of positive and negative returns for both strategies. While Strategy 1 has a balanced proportion of positive to negative returns, Strategy 2 has a much larger fraction of losing days. (As a reference point, the S&P 500's returns over this period are positive 53% of days.) This drives the overall return, and is graphically illustrated in Figure 1, borrowed from Bassett (2017). While in each subcategory Strategy 2 dominates Strategy 1, overall, due to the different relative numbers of up and down days, the result is reversed. Looking at the cumulative PnL chart in Figure 2 confirms the aggregate result, with the cumulative PnL for Strategy 1 ending above that for Strategy 2 (all returns are risk-adjusted).

Asset	Avg. Return	Avg. Return +	Avg. Return -
Strategy 1	**0.05**	1.00	-1.00
Strategy 2	0.04	**1.08**	**-0.71**

Table 1. Average daily risk-adjusted unconditional and conditional returns for two trading strategies.

Asset	Fraction of Returns +	Fraction of Returns -
Strategy 1	0.53	0.47
Strategy 2	0.42	0.58

Table 2. Fraction of positive and negative returns for two trading strategies.

This phenomenon of a reversal in effect when subpopulations or partitions of the data are analyzed, rather than the data as a whole (this 'whole' being a quite vague concept), is termed *Simpson's Paradox*, after the British statistician Edward H. Simpson.

4. Conclusion

Conditional statistics can give meaningful insight into the behavior of trading strategies. They should however be interpreted in conjunction with other measures of performance.

References

G. W. Bassett. Simpson's paradox and investment management. 2017.

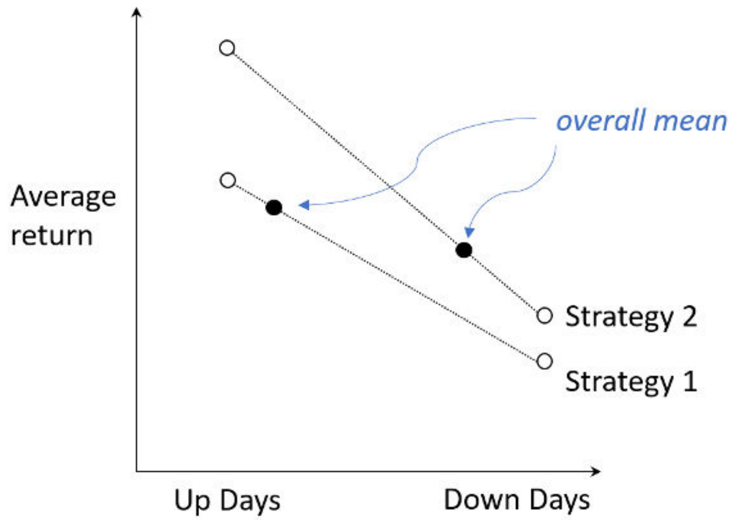

Figure 1. A graphical illustration of Simpson's Paradox for conditional and unconditional average returns. Strategy 2 has higher returns than Strategy 1 both on Up Days (when it gains money) and on Down Days (when it loses money). The overall mean for each strategy is bounded by its respective conditional average returns. The ratio of Up to Down Days determines where along the line the overall mean lies.

Figure 2. The performance of two hypothetical trading strategies over two decades. Both strategies have the same volatility, the end points of the cumulative return curves are therefore indicative of the magnitude of the overall mean.

Legal Disclaimer