Lecture Notes in Management Science (2013) Vol. **5**: 176–180
5th International Conference on Applied Operational Research, Proceedings

ISSN 2008-0050 (Print), ISSN 1927-0097 (Online)

Performance analysis of investment strategies – pitfalls and surprises

Peter Scholz [1] and Ursula Walther [2]

[1] *Hamburg School of Business Administration, Hamburg, Germany*

[2] *Berlin School of Economics and Law, Berlin, Germany*
ursula.walther@hwr-berlin.de

Abstract. Active investment strategies are a subject of endless debates. Myriads of studies have been conducted to proof performance potential - or to reject previous studies due to flaws or misinterpretations. The presentation will address three specific aspects which often are disregarded when performance is measured. Firstly, we will discuss the role of backtests and show that this instrument – even when used carefully and skilled – may lead to biased and misleading results. Secondly, we give an example that the concepts of performance and forecast power must be strictly distinguished. Finally, we demonstrate that implementation details, while largely neglected, may strongly impact and bias a strategy's performance.

Keywords: backtest; forecast power; implementation; market efficiency; performance analysis; simple moving average; simulations; technical timing strategies

Introduction

A rigorous performance analysis of a technical trading strategy is a surprisingly complex task. Many researchers are not fully aware of the many pitfalls and potential biases established methodology may cause. In this short note we address three aspects: The validity of standard backtestings, the distinction between performance and forecast power and the impact of implementation.

Validity of backtestings

Classic backtests

A standard tool often used by practitioners to test an investment strategy's performance is a classic backtest. When well done such a backtest is performed as in the study by Faber (2007/2009). He tests a market timing strategy which is applied to a stock market index. The thesis is that the strategy improves the overall performance. Compared to a direct investment in the index the investor would bear less risk by avoiding bad market times. The data sample used for the backtest is a very long S&P500 return time series from 1900 to 2008. The specific timing strategy signals when to enter the market (i.e. open a long position) and when to close it again. To follow this rule results in the strategy's result time series. Now the pure index returns (underlying) and the strategy returns are to be compared. The strategy is "better" than the underlying when the result time series offers a higher return and/or lower risk. The study by Faber (2007/2009) reports the following data to do the comparison:

	S&P 500	*Timing results*
Annualized Return	9.21%	10.45%
Volatility	17.87%	12.01%
Sharpe ratio	0.29	0.54
Maximum Drawdown	(83.66%)	(50.31%)
Best Year	52.88%	52.40%
Worst Year	(43.86%)	(26.87)

We see that the timing strategy slightly increases the average return (10.45% vs. 9.21%) and significantly improves a couple of risk figures, namely volatility, maximum drawdown, and the worst year return. So the conclusion is that indeed the timing strategy offers better performance than the pure underlying. What is wrong with this conclusion?

Pathwise versus terminal distribution

A critical insight for the understanding of performance analysis is the distinction between a pathwise and the terminal distribution. A backtest, like in the Faber (2007/2009) study above, analyses a pathwise distribution. This is the distribution of monthly (or daily) returns as generated by the strategy when applied to the one observed historical index development. However, an investor who has to decide whether or not to apply a strategy faces a very different situation. She does not know beforehand how the underlying will evolve in the future. If we assume the underlying to follow any stochastic process the potential future paths show a variety of different developments. The strategy only is attractive if it improves the performance on all or at least on a majority of the potential paths. Essentially, the investor is interested in what we call the terminal distribution. This distribution consists of the returns of all potential paths at the investment horizon (see figure 1). The terminal distribution reflects the investor's true chance-risk-position.

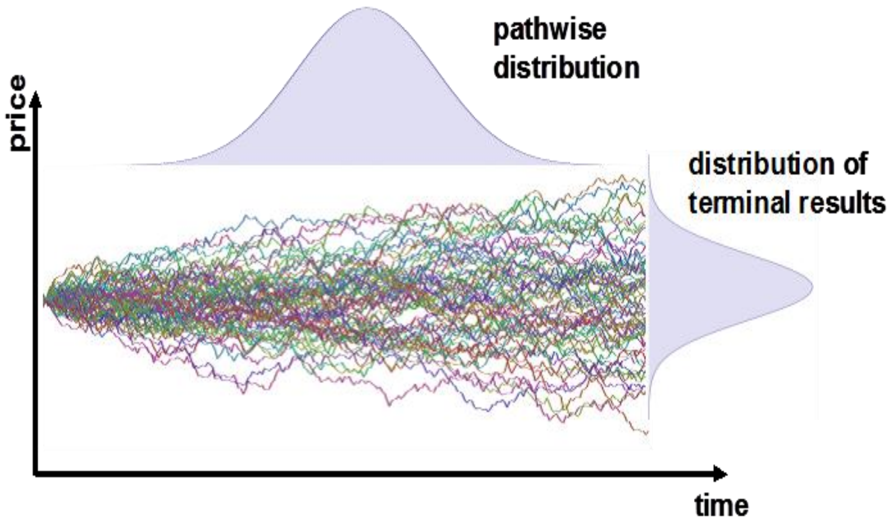

Fig. 1. The terminal distribution.

The relevant effect of a trading strategy is its impact on the terminal distribution which we call "return shaping". Performance in this sense means that the terminal distribution after applying a strategy (i.e. the "shaped" distribution) is more attractive than the original terminal distribution of the underlying.

Potential biases between pathwise and terminal distribution

To better understand the relation between pathwise and terminal distribution we ran a simulation study. As underlying we generated return paths by a bootstrapping procedure based on the German stock market index DAX. The data was DAX returns from 2000 to 2009. As trading strategy we used a simple moving average with 200 days averaging period (SMA 200) which is rather similar to the strategy used by Faber. A central finding is that the strategy performs extremely different on different paths. While indicating an attractive outperformance potential on some path (see table path 1) and at least a risk reduction on others (see table path 2) it proved completely worthless on others (see table path 3). A general conclusion from a single backtest path therefore is potentially hasty.

	average return	*volatility*	*VaR (1%)*	*kurtosis*	*max drawdown*
Path 1 underlying	- 3.02%	27%	0.054	11.54	1.13
Path1 SMA200	1.83%	24%	0.07	29.96	0.59
Path 2 underlying	-5.4%	27%	0.049	8.9	1.03
Path 2 SMA200	-4.91%	17%	0.047	33.8	0.72
Path 3 underlying	-0.28	26%	4.89	7.29	0.59
Path 3 SMA200	-12.9	30%	9.86	29.6	1.54

Nevertheless, on average the strategy might still work. This cannot be refuted by selected paths with bad performance. To check this we ran another simulation. This time we used a Brownian motion with drift as underlying in order to strictly control the statistics of the terminal distribution. Price paths were generated by the formula $r_t = \mu \cdot \Delta t + \sigma \cdot \sqrt{\Delta t} \cdot \varepsilon_t$ with drift $\mu = 8.6\%$, volatility $\sigma = 26\%$ and a standard normally distributed stochastic term ε. The following table compares the strategy's average performance measured from the different paths with its performance measured by its return shaping effect on the terminal distribution. The difference is striking. While the average pathwise performance indicates a slight outperformance but no increase in volatility the true terminal distribution reveals a significant increase in risk but no return. Hence, the pathwise distributions are biased. They overestimate success and underestimate risk.

timing results	*average return*	*volatility*	*skew*	*kurtosis*
pathwise distributions average of paths	0.02	25.9%	-0.267	15.6
terminal distribution	0.0054	59.9%	-4.81	30.1

Distinction between performance and forecast power

In further simulation studies we found additional pitfalls in the interpretation of performance found in empirical studies. It is common to interpret a finding of empirical performance as an indication of market inefficiency. Very often a subsequent conclusion is to ascribe forecast power to the strategy. Indeed, sustainable performance of technical trading strategies is inconsistent with efficient markets. However, from a pure backstest study such conclusion cannot be drawn. In Scholz and Walther (2011) we found that performance of SMA-timing strategies is systematically related to the drift and other process parameters of the underlying: The higher the drift the worse the performance. On market data we could confirm the finding. SMA timing outperformed the underlying only if the market data had a negative drift during the analysed time period. As long as the drift is not predictable such performance neither indicates market inefficiency nor forecast potential. For both conclusions more sophisticated methodology is necessary (see e.g. Brock et al (1992), Fifield et al (2005)).

Impact of implementation

Another finding from our simulation studies (Scholz (2012)) is that implementation details heavily influence the performance of trading strategies. Surprisingly, we found that many academic studies do not even fully report their detailed implementation. Hence, different studies are hardly comparable. Overall, the finding shows that performance is not an attribute of a signal generating strategy alone but an attribute of the combination of a strategy and an appropriate implementation.

Summary

By running a couple of simulation studies we found that results of empirical performance studies, especially when done as a classic backtesting, must be interpreted with great care. Firstly, pathwise distributions as generated by backtestings may be systematically biased compared to the terminal distribution which contains the relevant information from an investor's point of view. Secondly, performance potential found by a classic backtest neither indicates market inefficiency nor forecasting ability. Finally, implementation details do influence performance. Therefore, a trading strategy cannot be soundly assessed without a clear description of the implementation.

References

Brock W, Lakonishok J, LeBaron B (1992). Simple technical trading rules and the stochastic properties of stock returns. Journal of Finance 47(5): 1731-1764

Faber MT (2007): A quantitative approach to tactical asset allocation. The Journal of Wealth Management. 9(4): 69-79. Updated version February 2009.

Fifield SG, Power DM, Sinclair CD (2005): An analysis of trading strategies in eleven European stock markets. The European Journal of Finance 11(6): 531-584

Park CH, Irwin SH (2007): What do we know about the profitability of technical analysis? Journal of Economic Surveys 21(4): 786-826

Scholz P, Walther U (2011). The Trend is not Your Friend! – A Demystifcation of Empirical Timing Success based on Asset Price Characteristics and its Compatibility with Market Efficiency. working paper. Frankfurt School of Finance and Management.

Scholz P (2012). Size Matters! – How position sizing determines risk and return from technical timing strategies. working paper Frankfurt School of Finance and Management.