Biases in Evaluating Trading Strategies

RAYMOND KAN and GEORGE KIRIKOS*

Comments are welcome

First Draft: February, 1995
This version: November, 1995

*Raymond Kan is an assistant professor at the Faculty of Management of the University of Toronto and George Kirikos is the president of Leap of Faith Research Inc., a financial analytics and consulting company in Toronto. We would like to thank Paul Halpern and G. Andrew Karolyi for helpful comments and discussions, and Teresa Chan for research assistance. All remaining errors are ours.

In order to achieve superior returns, investors in financial markets invent new trading strategies on a regular basis. Before putting a new trading strategy into action, a careful investor will test how well this strategy would have performed in the past using historical data. The rationale for back-testing is if a trading strategy did not do well in the past, the chance that it will work in the future is slim. However, to the frustration of many traders and investors, a trading strategy that seems to work well in the past very often ceases to perform once it is put into actual use. It appears like the financial market is baiting the investor with a substantial "would-be profit," but once the investment is made, the financial market acts as a magician and says "now you see it, and now you don't." The phenomenal returns the investor expects mysteriously disappear.

In this paper, we discuss some of the biases that are very often introduced in the process of evaluating historical performance of a trading strategy. These biases tend to overstate the performance of the potential trading strategy in the historical testing period. Without correction, the past performance of a trading strategy is not always a reliable indicator of its future performance. In addition, the magnitude of these biases can be very large at times. As an illustration, we discuss a trading strategy which was recently proposed by Foerster, Prihar and Schmitz (hereafter FPS)[1] and show how its past performance could be contaminated by a survivorship bias. Although FPS reported an extraordinary historical performance of this trading strategy, we show that after controlling for the survivorship bias, their strategy would not have outperformed the market in the historical testing period.

Biases in Evaluating Historical Performance

The success of a trading strategy derives solely from its ability to predict future returns of some financial assets.[2] In evaluating trading strategies, the most important and relevant piece of information is its historical performance. However, the process of evaluating historical performance of a trading strategy is often contaminated by various biases that could make its outcome totally unreliable. Broadly speaking, we can classify these biases into three categories: (1) pre-test bias, (2) trading cost bias, and (3) look-ahead bias. We discuss the nature of these three types of biases and ways to mitigate their influence.

Pre-test Bias

Trading strategies do not come out of thin air. They typically come from personal experience or research on the past data. In either case, the formulation of the trading strategy is heavily influenced by the historical record of the financial assets. If one derives a trading strategy based on one's experience with the financial market over a particular historical period, and then back-tests the performance of the trading strategy using the very same historical period, the trading strategy is likely to perform well in that period. However, traditional statistical testing would not be appropriate in this case because the strategy was developed only after the data was observed. It is always a possibility that an investor uncovers a pattern that occurred simply by coincidence. In the most extreme form, an investor could start off with thousands of useless trading strategies, but by testing all of them using historical data, some would appear to work simply because of chance. Trading strategies that come out of this "data mining" exercise will always do well in back-testing because back-testing is the very process that was used to filter the trading strategies. In a more subtle form, pre-test bias is manifested in the form of "data-snooping" bias as suggested by Lo and MacKinlay.[3] If an investor forms a stock portfolio based on some firm-specific characteristic that is empirically motivated (either due to some casual observation of the data or observed experience from a foreign market), then to the extent that there is even a very small spurious correlation between stock returns and this characteristic, the resulting portfolio could have significantly higher returns in the back-testing period. Therefore, the investor could wrongly believe that a trading strategy based on this firm-specific characteristic is profitable.

Pre-test bias is unavoidable because the development of a successful trading strategy requires a trial-and-error process. In addition, it is unreasonable to expect an investor to come up with a trading strategy without first looking at some historical data. The important thing to remember is that the formal procedure of standard statistical testing is violated and the resulting statistics should not be taken at face value. Adjusting for this bias is difficult but an investor could have more confidence in the profitability of a trading strategy if it has some economic underpinning, as opposed to a strategy like the "Superbowl" theory[4] which is an obvious product of "data mining." It would also be useful to look at the robustness of the performance of the trading strategy over different subperiods. If the superior performance of a trading strategy is entirely driven by one subperiod, then the profit is probably spurious and unlikely to continue in the future.

Trading Cost Bias

Very often in the back-testing of a trading strategy, investors assume they could buy and sell at the closing prices. In reality, there are implicit and explicit trading costs that one has to pay to execute a trade. Explicit trading costs include bid/ask spread and commission. In the case of short-selling, it also includes the opportunity cost of the margin deposit. Implicit trading cost is the cost of market impact, that is the unfavourable price movement that may be induced by a trade. Failure to account for these trading costs will overstate the performance of a trading strategy, especially for one that requires frequent trading and involves small and illiquid stocks. For example, Keim[5] finds that a substantial portion of the turn-of-the-year effect in the U.S. stock market can be accounted for by a systematic movement within the bid-ask spread for low-priced stocks. The measured returns (using closing prices) for the two days bracketing the end of the year are overstated by 3.3%. Therefore, trading strategies based on the turn-of-the-year effect in those two days are not necessarily profitable. Since trading costs differ across investors, it is entirely possible that a trading strategy is profitable to one group of investors but not to another. The adjustments would need to be made on an individual basis and should take into account the potential size of the trades.

Look-ahead Bias

A trading strategy is often designed to pick a portfolio of investments based on specific rules. However, in back-testing the trading strategy, an investor might introduce some information into the selection process that only becomes available subsequent to the selection. A trading strategy which requires future information is not implementable. To the extent that the future information is valuable, the profitability of the trading strategy in the back-testing period could be entirely due to the future information. For example, an investor could trade based on the earnings of a firm but he has to be aware that such information is not publicly available until a few months after the fiscal year-end. Failure to exclude the future information in the back-testing period could significantly overstate the historical performance of a trading strategy.

Out of all the look-ahead biases, the most notorious one is the "survivorship bias." A survivorship bias occurs when an investor tests the trading strategy on a subset of stocks that exist at the end of the testing period. Stocks that are delisted or went into bankruptcy over the period are excluded from the sample. By excluding these stocks, the remaining sample would have an above average performance and therefore trading strategies tested on this sample would tend to find above normal rates of return. Such bias sometimes is unavoidable in the back-testing period because of data availability. Data on firms that went into bankruptcy are more difficult to obtain. However, ignoring this bias could lead an investor to draw an incorrect conclusion on the profitability of a trading strategy. For example, Brown, Goetzmann, Ibbotson and Ross[6] pointed out that in theory survivorship bias could induce a spurious positive autocorrelation in returns and therefore a trading strategy that purchases past winners would appear to be profitable. The magnitude of the bias depends on the trading strategy and the data sample. For example, in the case of mutual fund studies, Hendricks, Patel, and

Zeckhauser[7] find that the effect of survivorship bias on their studies is quite minimal. On the other hand, using a different sample, Malkiel[8] finds that survivorship bias would lead to an overstatement of average fund returns by 150 basis points per annum for the period 1982–91. Therefore, without actually assessing the magnitude of the survivorship bias, one could not in general discount its importance. In the following section, we show that controlling for the survivorship bias can make dramatic differences in the evaluation of the performance of a trading strategy.

Momentum Based Trading Strategy

Motivated by some US studies,[9] FPS suggested a momentum based trading strategy by computing a price momentum indicator based on a weighted average quarterly return over the previous four quarters (with the most recent quarter receiving twice the weight of the others). Based on this momentum index, FPS ranked a sample of stocks on a quarterly basis. The "best outlook" and "worst outlook" portfolios were formed using the top 10 and bottom 10 ranked equities in their sample respectively, and the two portfolios were rebalanced quarterly. The model was tested over the 1978 to 1992 period. During that period, the annual return of the "best outlook" portfolio was an incredible 41.2% (before transaction costs), while the TSE 300 annual total return was 13.1%. The "worst outlook" portfolio's annual return was only 10.9%. Of particular note is the sample universe of stocks which were eligible to be ranked and selected by the trading system. FPS restricted themselves to stocks which subsequently became members of the TSE 100 index at the end of 1993. Such information is obviously not available at the time of the portfolio selection. In their endnote, they recognized that this creates a potential survivorship bias, but argued that "*given the strength of the results, it is safe to conclude that any other reasonable portfolio formation techniques would result in qualitatively similar conclusions.*"

While FPS's sample selection scheme uses future information and renders such a strategy unimplementable, the direction of the bias is not clear in theory. The TSE 100 Index at the end of 1993 can be divided into two sets. The first set consists of stocks which always had large capitalizations relative to the rest of the market, and would include the banks, BCE, Canadian Pacific, Imperial Oil, and other large companies with long histories. Within this subset, the stocks cannot have had continuous bad performance in the sample period. Otherwise, they would probably have fallen out of the TSE 100 at the end of the sample period. Therefore, even though there is momentum in stock prices, it is less likely to be uncovered in this subset of stocks. The second set of stocks are the "big winners" of the past 15 years, which were once tiny companies (such as Cott, Magna, and Newbridge Networks), but due to a sequence of outstanding returns, rose to become members of the TSE 100 at the end of 1993. By having these *ex post* big winners in the sample universe for every period, especially when they were small, a bias is created. Within this subset, the momentum based trading strategy definitely works but the problem is the knowledge of the identity of the second set of stocks would not be available to the investors until the end of 1993. In his commentary on the FPS article, Keith Ambachtsheer[10] wondered aloud as to whether FPS had slipped in a few penny stocks to generate such unbelievably high returns for the "best outlook" portfolio. In essence, this is exactly what FPS did. Stocks like Cott, Magna, and Newbridge Networks which were in the TSE 100 Index at the end of 1993 were not always large cap stocks. In fact, these were amongst the biggest winners in the history of the Toronto Stock Exchange, and they were included in FPS's sample universe for each and every period of their existence, even when they were little more than penny stocks. Although there is reason to believe that the bias from the second set of stocks is bigger than the bias from the first set of stocks, the exact magnitude of the bias is not clear. One needs to control for this survivorship bias to give a fair assessment of the profitability of FPS's trading strategy.

To actually implement FPS's trading strategy, one needs to choose stocks from the TSE 100 at the time of the trade. However, the Toronto Stock Exchange did not introduce the TSE 100 Index until the end of 1993. Fortunately, they have used exactly the same inclusion criteria to create the would-be TSE 100 list at the end of every year from 1981 to 1992.[11] Therefore, our first sample consists of this universe of stocks, which is what FPS would have to use if they tried to actually implement their strategy. As the sample period is shorter than that used by FPS, we created a second sample using the top 100 Canadian equities (ranked by market capitalization). The dataset consisted of the TSE/Western database from January 1955 to May 1991, although we focused on the post-1975 period (due to a potential sample selection bias of the database in the pre-1975 period).[12] At the end of each quarter, all of the Canadian equities listed on the TSE were sorted according to market capitalization (number of shares outstanding multiplied by the closing price), and all but the top 100 were dropped from consideration for investment in the next quarter. Thus, this second sample universe was limited to the largest and most liquid Canadian stocks each quarter. This second sample overlaps with the TSE 100 stocks to a great degree. For example, at the end of 1990, 76 of the stocks in the second sample were members of the TSE 100 index.[13] The two samples differ from that of FPS in only one aspect, in that the investor does not need to know the composition of the TSE 100 far into the future, and therefore our samples are free of survivorship bias. As stocks rise and fall over time, the composition of the TSE 100 would change, and this is reflected in our methodology.[14]

After determining the sample universe for the quarter, we proceed exactly using the methodology in FPS. We ranked the 100 stocks in each sample using the FPS momentum indicator. Based on the indicator value, two portfolios are formed. The ten stocks with the highest indicator values are assigned to the "best outlook" portfolio, while the ten stocks with the lowest indicator values are assigned to the "worst outlook" portfolio. The ten stocks in each portfolio are given equal weights and held for the next quarter. At the end of the next quarter, we repeat the above procedures.

A summary of the results for both samples is presented in Table I. As in FPS, we use the TSE 300 total return as our benchmark. Over the period 1982 Q1 to 1991 Q1, the "best outlook" portfolio had an annual return of 15.28% for the first sample and 13.36% for the second sample. As for the "worst outlook" portfolio, it had an annual return of −0.36% for the first sample and 6.84% for the second sample.[15] Over this period, the TSE 300 annual total return was 11.76%. Over the longer period of 1976 Q1 to 1991 Q1, we can only use the second sample. In this case, the "best outlook" portfolio had an annual return of 16.76% and the "worst outlook" portfolio had an annual return of 7.96%. Over this longer period, the TSE 300 annual total return was 14.2%.

In either case, the "best outlook" portfolio outperformed the market by less than 4% per annum. This outperformance cannot be explained by the systematic risk (beta) of the "best outlook" portfolio since the beta of the "best outlook" portfolio is not much higher than that of the TSE 300.[16] However, subtracting transaction costs would more than offset any remaining excess returns of the "best outlook" portfolio. The quarterly rebalancing of the momentum based trading system would have much higher transaction costs compared to a buy and hold strategy.[17] FPS subtracted 8% annually from their raw returns to account for transaction costs. Even if half of that was subtracted from our raw returns, the "best outlook" portfolio would not outperform the market. The excess returns achieved by FPS of more than 28% annually (before transaction costs) are nowhere to be found after controlling for the survivorship bias.

More interesting results were to be found in the "worst outlook" portfolio, which underperformed the TSE 300 by 4.92%–12.12% per annum, depending on the sample and the data period. Since both portfolios had similar levels of systematic risk, that is not an explanation for the size of the inferior returns. Transaction costs associated with the short selling of these stocks would likely eliminate any trading opportunities, except possibly for market makers. For portfolio managers who already have long positions in these "worst outlook" stocks, one might argue that it is advisable to sell these stocks, in order to avoid the inferior returns in the subsequent quarter. For long-term investors, the round-turn costs of selling the stock and repurchasing it a quarter later, in an attempt to avoid the inferior return, would likely make the trade inadvisable, unless the stock was expected to remain on the list of "worst outlook" stocks for several consecutive quarters. For more nimble traders, who are not married to the stocks, dumping these "poor outlook" equities to avoid further inferior performances might well be advisable. This is consistent with the popular notion amongst traders of taking losses quickly.

Not shown in Table 1 are the results of doing further back-testing back to 1956. (Prior to 1975, there is a potential selection bias in the TSE/Western database, as only a subset of all listed stocks were included in the dataset.) These results are consistent with those shown. From 1956 Q2 (the first year of the TSE 300 index) to 1975 Q4, the "best outlook" portfolio had an annual return of 10.40%, compared to 2.12% for the "worst outlook" portfolio and 7.12% for the TSE 300 (all before transaction costs).

Conclusion

In this paper, we discuss some of the biases in evaluating trading strategies based on their past performance. These biases can cause spurious profits in the back-testing period and as a result past performance does not always foretell future performance. The magnitude of these biases can be very large and should not be overlooked.

As an illustration, we re-examine the profitability of the FPS momentum based strategy after controlling for the survivorship bias. The results show that survivorship bias makes a substantial difference in the evaluation of their trading strategy. Our results show that their reported returns of the "best outlook" portfolio and "worst outlook" portfolio are both overstated by the survivorship bias. Although this paper found that after controlling for the survivorship bias, the profitability of buying the "best outlook" portfolio is limited, there may be some value to using the momentum based indicator proposed by FPS for avoiding stocks in the "worst outlook" portfolio. In addition, the authors have anecdotal evidence that momentum based indicators can yield excess returns and can be very useful as filters. We would not be surprised if momentum based indicators different from the one proposed by FPS are useful and further research into other momentum indicators is warranted.

FOOTNOTES

1. Stephen Foerster, Anoop Prihar, and John Schmitz, "Back to the Future," *Canadian Investment Review*, Winter 1994/95, pp. 9–13.

2. Predicting future returns of financial assets is not equivalent to identifying underpriced or overpriced financial assets. For example, underpriced stocks can continue to be underpriced for a long period of time and buying such stocks does not necessarily give an investor superior returns.

3. Andrew W. Lo, and A. Craig MacKinlay, "Data-Snooping Biases in Tests of Financial Asset Pricing Models," *Review of Financial Studies* 3, 1990, pp. 431–467.

4. The "Superbowl" theory holds that the Dow Jones Industrial Average will be up on the year if the NFC team wins the Superbowl, and down otherwise.

5. Donald B. Keim, "Trading Patterns, Bid-ask Spreads, and Estimated Security Returns," *Journal of Financial Economics* 25, 1989, pp. 75–97.

6. Stephen Brown, William Goetzmann, Roger Ibbotson, and Stephen Ross, "Survivorship Bias in Performance Studies," *Review of Financial Studies* 5, 1992, pp. 553–80.

7. Darryll Hendricks, Jayendu Patel, and Richard Zeckhauser, "Hot Hands in Mutual Funds: Short-Run Persistence of Relative Performance, 1974–1988," *Journal of Finance* 48, 1993, pp. 93–130.

8. Burton G. Malkiel, "Returns from Investing in Equity Mutual Funds 1971 to 1991," *Journal of Finance* 50, 1995, pp. 549–572.

9. For example, Narasimhan Jegadeesh and Sheridan Titman, "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency," *Journal of Finance* 48, 1993, pp. 65–91.

10. Keith Ambachtsheer, "The Price is Right," *Canadian Investment Review*, Winter 1994/95, p. 42.

11. See TSE 100 Index Reference Guide issued by the Toronto Stock Exchange.

12. The choice of the end of the sample period is dictated by data availability. The version of the TSE/Western database that we have access to ends at May 1991.

13. It is not surprising that the two samples have such a high degree of overlap, as one of the main inclusion criteria of the TSE 100 stocks is that they have to rank, by Quoted Market Value, not lower than 120th out of the TSE 300 companies at the end of every year.

14. For example, only 46 of the TSE 100 stocks at the end of 1981 were still in the index at the end of 1993, thus showing that the composition can change dramatically over time.

15. We checked for outliers in our sample and we corrected for a few data errors in the TSE/Western tape. For example, on October 10, 1958, the tape omitted a stock split of the Molson Companies Ltd., Class A, and resulted in a −49.5% return for the month of October 1958 and a 104.6% return for the month of November 1958. On April 11, 1988, the tape mistakenly recorded a dividend as a stock split for Polysar Energy and Chemical Corp., and resulted in a 752.2% return for the month of April 1988.

16. We assume the betas of the two portfolios are constant over the sample period and estimate the betas using monthly returns. If we allow the beta or the market risk premium to be time-varying, we could potentially explain some of the difference. See G. Andrew Karolyi and Bong-Chan Kho, "Time-Varying Risk Premia and the Returns to Buying Winners and Selling Losers: Caveat Emptor et Venditor," working paper, Ohio State University, 1994.

17. On average, the "best outlook" portfolio has 6.5 new stocks every quarter and the "worst outlook" portfolio has 6.1 new stocks every quarter.

Table I
Returns of Momentum Based Trading Strategy

This table reports the returns on the "best outlook" portfolio, the "worst outlook" portfolio and the TSE 300 total return index for two different sample periods. The first sample period is from 1976 Q1 to 1991 Q1 and the momentum based portfolios are selected from the top 100 stocks (in market capitalization) of the Toronto Stock Exchange. The second sample period is from 1982 Q1 to 1991 Q1 where we also include momentum based portfolios selected from the stocks in the TSE 100 index. The "best outlook" and "worst outlook" portfolios are equally weighted and rebalanced every quarter.

	Best Outlook Portfolio		Worst Outlook Portfolio		TSE 300 Total Return
	Top 100	TSE 100	Top 100	TSE 100	
Period: 1976 Q1 – 1991 Q1					
Quarter Mean	4.19%	–	1.99%	–	3.55%
Quarter Standard Deviation	10.93%	–	10.48%	–	8.63%
Annual Mean	16.76%	–	7.96%	–	14.20%
Annual Standard Deviation	21.86%	–	20.96%	–	17.26%
Portfolio Beta	1.02	–	1.00	–	1.00
Percentage of Quarters with Returns Greater than 0	63.93%	–	55.74%	–	75.41%
Period: 1982 Q1 – 1991 Q1					
Quarter Mean	3.34%	3.82%	1.71%	−0.09%	2.94%
Quarter Standard Deviation	11.09%	11.66%	12.45%	14.00%	9.42%
Annual Mean	13.36%	15.28%	6.84%	−0.36%	11.76%
Annual Standard Deviation	22.18%	23.32%	24.90%	28.00%	18.84%
Portfolio Beta	0.90	1.10	1.11	1.28	1.00
Percentage of Quarters with Returns Greater than 0	56.76%	70.27%	54.05%	59.46%	70.27%