Review of Financial Economics 23 (2014) 30–45

Contents lists available at ScienceDirect

Review of Financial Economics

journal homepage: www.elsevier.com/locate/rfe

Predictability of the simple technical trading rules: An out-of-sample test

Jiali Fang *, Ben Jacobsen, Yafeng Qin

Massey University, Private Bag 102904, North Shore, Auckland 0745, New Zealand

ARTICLE INFO

Article history:
Received 21 January 2013
Received in revised form 10 May 2013
Accepted 15 May 2013
Available online 25 May 2013

JEL classification:
G11
G14

Keywords:
Technical analysis
Market efficiency
Out-of-sample tests
Return predictability

ABSTRACT

In a true out-of-sample test based on fresh data we find no evidence that several well-known technical trading strategies predict stock markets over the period of 1987 to 2011. Our test safeguards against sample selection bias, data mining, hindsight bias, and other usual biases that may affect results in our field. We use the exact same technical trading rules that Brock, Lakonishok, and LeBaron (1992) showed to work best in their historical sample. Further analysis shows that this poor out-of-sample performance most likely is not due to the market becoming more efficient – instantaneously or gradually over time – but probably a result of bias.

1. Introduction

Technical analysis studies patterns in historical stock market series generated by day-to-day market activities, with the aim to predict future market movements. The key information technical analysts use is volume and price. We evaluate the profitability of 26 classic technical trading strategies that are formed by using the underlying price on the Dow Jones Industrial Average (DJIA) during the period from 1987 to 2011. These trading rules were first tested extensively by Brock et al. (1992) which allows us to perform a comprehensive out-of-sample test by using exactly the same trading rules on a fresh new data set that minimises the effect of any possible statistical biases. With the benefit of a fresh data set, we find little predictability of the 26 technical trading strategies out-of-sample, which is in strong contrast with their in-sample findings. Further analysis of these out-of-sample results shows that the profitability of these strategies does not gradually disappear suggesting that the market becomes more efficient over time, but trading strategies based on these rules underperform the market from the beginning of our out-of-sample period. While it is possible that all investors started using these technical rules and made the market instantaneously more efficient, it seems more likely that the earlier results are caused by some sort of statistical bias. Particularly because we also find no evidence of success for these trading rules in another 12 year out-of-sample period from 1885 to 1896. Moreover, the in-sample success of the technical trading strategies does not alter in several robustness tests. It does not change when we use OLS robust regressions to limit the impact of outliers, nor when we use rolling window regressions to check if any particular period drives these results in the original sample. Also the technical trading rules remain successful when we consider the S&P 500 rather than the Dow Jones index. Similarly, the failure of the technical results out-of-sample is equally robust. The technical trading rules also do not generate profits when we correct for outliers, or consider specific sample periods using rolling windows. Additionally the 2008 financial crisis period does not appear to drive the out-of-sample results as the profitability of the 26 technical trading rules also does not persist out-of-sample when we remove the crisis period from our sample. No other alternative hypothesis seems to explain the difference between in-sample and out-of-sample results, but the statistical biases. Last but not least, the inclusion of transaction cost that further eliminates the profitability of technical trading strategies may cast even stronger doubts on the efficiency of the technical trading strategies. Our study shows the importance of studying new data to safeguard against the danger of possible statistical biases.

The possible danger of biases of all sorts is well known. Jensen and Bennington (1970) indicate that superior trading rule performance is often a consequence of survivorship bias. Merton (1985) points out the danger of selection bias and cognitive bias that could affect results, while studying the behaviour of stock market returns; Lo and MacKinlay (1990) state that the degree of data snooping bias in a particular field increases with the number of studies published on the topic. Others like Denton (1985), Black (1993), and Ferson, Sarkissian, and Simin (2003) also emphasize the threats of statistical biases. In the

* Corresponding author at: School of Economics and Finance, Massey University, Private Bag 102904, North Shore, Auckland 0745, New Zealand. Tel.: +64 9 441 8176x9242.
E-mail addresses: j.fang@massey.ac.nz (J. Fang), b.jacobsen@massey.ac.nz (B. Jacobsen), y.qin@massey.ac.nz (Y. Qin).

1058-3300/$ – see front matter
http://dx.doi.org/10.1016/j.rfe.2013.05.004

field of technical analysis, Sullivan, Timmermann, and White (1999) utilise the White's Reality Check technique to check for any data snooping bias in particular, and Bajgrowic and Scaillet (2012) employ a false discovery rate strategy to deal with the same problem. However, it is difficult to guard against other statistical biases that could affect the results. Fama (1991) and Lakonishok and Smidt (1988) both provide us with the best solution for these statistical biases: The use of new data. Fama (1991, p. 1587) states that: "We should also keep in mind that the CRSP data… are mined on a regular basis by many researchers. Spurious regularities are a sure consequence. Apparent anomalies in returns thus warrant out-of-sample tests before being accepted as regularities likely to be present in future returns". Lakonishok and Smidt (1988) prescribe long and new data series as the best remedy against data snooping, noise and 'boredom' (selection bias). Fortunately, with the passage of time many earlier studies can now be replicated with fresh data. Our study is, therefore, primarily motivated to perform such an out-of-sample test, by having access to another 25 years of out-of-sample data other than that used in Brock et al. (1992).

The study of Brock et al. (1992) is an important milestone in the field of technical analysis. Not only because they tested a large number of popular technical trading rules but also because it marks a turning point in the academic view on technical analysis. Before the publication of their work, technical analysis was largely dismissed by academics in the 1960s and 1970s. Although Alexander (1964) provides supportive evidence for the profitability of technical analysis on stock markets by utilising the filter rules, Fama (1965) and Samuelson (1965) both question the value of technical analysis by providing evidence in favour of random walk models. The debate on the usefulness of technical analysis has continued since these studies. But it suffered a relatively quiet period until the beginning of the 1990s. Modern studies in the field of technical analysis are boosted from the beginning of the 1990s, which coincides with the publication of Brock et al. (1992). According to Park and Irwin (2004, p 17): "The number of technical trading studies over the 1995–2004 period amounts to about half of all empirical studies conducted since 1960". Following the strength of their findings, many studies further confirm the predictive power of their set of technical trading rules in many different economic circumstances. These trading strategies are found to beat the buy-and-hold strategy in different stock markets across the world. For example, Raj and Thurston (1996), Parisi and Vasquez (2000) and Vasiliou, Eriotis, and Papathanasiou (2008) provide supportive evidence from the Hong Kong, Chile and Greek markets, respectively. Bessembinder and Chan (1995) take transaction costs into account on six Asian stock markets (Hong Kong, Japan, Korea, Malaysia, Thailand and Taiwan) during the period of 1975 to 1991, with these trading rules again found to significantly beat the buy-and-hold strategy across all markets and all trading rules. Previous literature also confirms the predictive ability of the technical trading strategies when different forecasting techniques are employed. For instance, Fernandez-Rodrıguez, Gonzalez-Martel, and Sosvilla-Rivero (2000) use artificial neutral networks and they discover predictability for the Madrid stock market from 1966 to 1997. Gencay (1996) and Gencay and Stengos (1998) both use feed-forward networks and report positive results on the DJIA during the period 1963 to 1988. Using the same data, Gencay and Stengos (1997) reach a similar conclusion when they apply the nearest neighbour regression technique. For a longer sample period from 1897 to 1988, Gencay (1998) also provides supportive evidence by using the same feedforward networks method on the DJIA. Lastly, not just in the stock markets, Gencay, Dacorogna, Olsen, and Pictet (2003), Gençay, Ballocchi, Dacorogna, Olsen, and Pictet (2002) and Gencay (1999) further report the merit of the technical trading strategies in the forex markets.

The concern of data snooping arises with the increasing supportive evidence reported in the field of technical analysis. Sullivan et al. (1999) find that the results of Brock et al. (1992) are not altered after taking into account the quantified data snooping effects. They also show that the same significant profitability is not realised in shorter out-of-sample tests on either the DJIA 1987 to 1996 data, or the S&P 500 futures data. They state at the end of their study that: "…it is possible that, historically, the best technical trading rule did indeed produce superior performance, but that, more recently, the markets have become more efficient and hence such opportunities have disappeared" (Sullivan et al., 1999, p. 1684). Bajgrowic and Scaillet (2012) also show that technical trading rules do not outperform after 1986. Their study uses a different method to account for the data snooping effects. These two studies focus on examining the data snooping adjusted predictability of a large number of technical trading rules (in both cases, they use the same universe of 7846 technical trading rules). Our study differs as we do not consider a large universe of trading rules but focus on what would have happened to an investor had he or she implemented the 26 trading rules that seemed to perform so well in the past. Our paper uses a substantially longer sample of fresh data available over the last 25 years, which safeguards against any possible biases with respect to the Brock Lakonishok and LeBaron set of trading rules. Last but not least we investigate why these specific technical trading rules might not work. Is that caused by bias or the market becoming (gradually) more efficient with respect to these trading rules over time?

2. Out of sample tests

Fresh out of sample data are generally considered to offer the strongest safeguard against possible statistical biases. For instance, Neely and Weller (2012) consider fresh data based out-of-sample study as the most certain solution against data snooping, data mining and publication bias; Cooper and Gulen (2006) report that many features of a researcher's *out-of-sample* experiment such as the choice of assets, predictive variables, length of the in-sample window used to obtain forecast parameters, and model selection methods are typically exogenously determined by the researcher after having obtained familiarity with the entire data, whereas it does not induce a bias when out-of-sample tests are performed on new data. Additionally, Andrikopoulos, Daynes, Latimer, and Pagas (2008), Davis (1994), Foster, Smith, and Whaley (1997), Rapach and Wohar (2006), Hand, Mannila, and Smyth (2001), McQueen and Thorley (1999), Ilmanen (2011), DeFusco, McLeavey, Pinto, and Runkle (2007), and Cortes, Mohri, Riley, and Rostamizadeh (2008) all claim the cleanness of the results that the *true* out-of-sample studies could provide.

True out-of-sample test is opposed to sample splitting which is also common in the academic literature. For instance, researchers sometimes validate the in-sample results by using split samples — using one part of the sample for calibration and the other for verification. Faraway (1992), Camstra and Boomsma (1992) and Inoue and Kilian (2005) question the efficiency of such method, and Chatfield (1995) considers the use of new data as irreplaceable. As Chatfield puts it: "Statisticians sometimes think that they can overcome the need for new data by splitting a sample into two parts… this is a poor substitute for true replication and the same sentiment also applies to techniques like cross-validation. 'The only real validation of a statistical analysis, or of any statistical enquiry, is confirmation by independent observations' (Anscombe, 1967, p. 6) and so model validation needs to be carried out on a completely new set of data" (Chatfield, 1995, p. 439). We should also distinguish between using completely fresh new data from those using the appended *new* data set. In the latter case, only small amount of new data is added to the original data set, and the resulting longer data set is used for the *out-of-sample* confirmation. Conrad, Cooper, and Kaul (2003) argue that such *out-of-sample* experiment is likely to be affected by any snooping bias that is present in the original results. Besides, while some in-sample tests provide remedies for a particular type of statistical bias (for instance, Sullivan et al., 1999 and Bajgrowic and Scaillet, 2012 for data snooping), the use of fresh sample helps to avoid many