Journal of Empirical Finance 17 (2010) 471–484

Contents lists available at ScienceDirect

Journal of Empirical Finance

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

Testing the predictive ability of technical analysis using a new stepwise test without data snooping bias

Po-Hsuan Hsu [a,*], Yu-Chin Hsu [b], Chung-Ming Kuan [c]

[a] Department of Finance, University of Connecticut, United States
[b] Department of Economics, University of Texas at Austin, United States
[c] Department of Finance, National Taiwan University, Taiwan

ARTICLE INFO

Article history:
Received 25 October 2008
Received in revised form 19 July 2009
Accepted 5 January 2010
Available online 18 January 2010

JEL classification:
C12
C32
C52
G11

Keywords:
Data snooping
Exchange traded funds
Reality check
SPA test
Stepwise test
Technical trading rules

ABSTRACT

In the finance literature, statistical inferences for large-scale testing problems usually suffer from data snooping bias. In this paper we extend the "superior predictive ability" (SPA) test of Hansen (2005, *JBES*) to a stepwise SPA test that can identify predictive models without potential data snooping bias. It is shown analytically and by simulations that the stepwise SPA test is more powerful than the stepwise Reality Check test of Romano and Wolf (2005, *Econometrica*). We then apply the proposed test to examine the predictive ability of technical trading rules based on the data of growth and emerging market indices and their exchange traded funds (ETFs). It is found that technical trading rules have significant predictive power for these markets, yet such evidence weakens after the ETFs are introduced.

Published by Elsevier B.V.

1. Introduction

Technical analysis has been widely applied in stock markets since W.P. Hamilton wrote a series of articles in *The Wall Street Journal* in 1902. Its predictive power (or profitability), however, remains a long-debated issue in both industry and academia. A recent article in *The Wall Street Journal* observes: "Some brokerage firms have eliminated their technical research departments altogether. Still, when the markets begin to sag, investors rediscover technical analysis" (Browning, July 30, 2007). Indeed, the same article reports that some technical analysts did foresee and warn their clients right before the stock market plunge on July 23, 2007. There are also numerous empirical results in the literature that support technical analysis, such as Sweeney (1988), Blume et al. (1994), Brown et al. (1998), Gencay (1998), Lo et al. (2000), and Savin et al. (2007). Such evidences, however, may be criticized for their data snooping bias; see, e.g., Lo and MacKinlay (1990) and Brock et al. (1992).

Data snooping is common in the finance and economics literature. In practice, only a few financial data sets are available for empirical examination. Data snooping arises when researchers rely on the same data set to test the significance of different models (technical trading rules) individually. As these individual statistics are generated from the same data set and hence related to each other, it is difficult to construct a proper joint test, especially when the number of models (rules) being tested is large. White (2000)

* Corresponding author.
E-mail address: Paul.Hsu@business.uconn.edu (P.-H. Hsu).

0927-5398/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.jempfin.2010.01.001

proposes a large-scale joint testing method for data snooping, also known as Reality Check (RC), which takes into account the dependence of individual statistics. Sullivan et al. (1999) apply the RC test and find that technical trading rules lose their predictive power for major U.S. stock indices after the mid 80s.

White's RC test suffers from two drawbacks. First, Hansen (2005) points out that the RC test is conservative because its null distribution is obtained under the least favorable configuration, i.e., the configuration that is least favorable to the alternative. In fact, the RC test may lose power dramatically when many poor models are included in the same test. To improve on the power property of the RC test, Hansen (2005) proposes the "superior predictive ability" (SPA) test that avoids the least favorable configuration. Empirical studies, such as Hansen and Lunde (2005) and Hsu and Kuan (2005), also show that the SPA test is more powerful than the RC test. Second, the RC test checks whether there is any significant model but does not identify all such models. Note that Hansen's SPA test shares the same limitation. Romano and Wolf (2005) introduce an RC-based stepwise test, henceforth Step-RC test, that is capable of identifying as many significant models as possible. Nonetheless, Romano and Wolf's Step-RC test is conservative because its stepwise critical values are still determined by the least favorable configuration, as in the original RC test.

In this paper, the SPA test is further extended to a stepwise SPA (Step-SPA) test that can identify predictive models in large-scale, multiple testing problems without data snooping bias. This is analogous to the extension of White's RC test to Romano and Wolf's Step-RC test. It is shown that the Step-SPA test is consistent, in the sense that it can identify the violated null hypotheses (models or rules) with probability approaching one, and its familywise error (FWE) rate can be asymptotically controlled at any pre-specified level, where FWE rate is defined as the probability of rejecting at least one correct null hypothesis. This paper makes additional contribution by showing analytically and by simulations that the Step-SPA test is more powerful than the Step-RC test, under any power criterion defined in Romano and Wolf (2005).

In our empirical study, the proposed Step-SPA test is applied to evaluate the predictive power of 9120 moving average rules and 7260 filter rules in several growth and emerging markets. Unlike many existing studies on technical analysis, we examine not only market indices but also their corresponding Exchange Traded Funds (ETFs). Considering ETFs is practically relevant because ETFs have been important investment vehicles since their inception in late 90s. Moreover, due to the tradability and low transaction costs, ETFs help to increase market liquidity and hence may improve market efficiency (e.g. Hegde and McDermott, 2004). Our empirical study thus enables us to assess whether the predictive power of technical rules, if any, is affected after ETFs are introduced.

Our empirical results provide strong evidence that technical rules have significant predictive ability in pre-ETF periods, yet such evidence weakens in post-ETF periods. In particular, we find many technical rules with significant predictive power prior to the inception of ETFs in U.S. growth markets but *none* when the ETFs that track these market indices become available. For emerging markets, we find technical rules have predictive ability for 4 (out of 6) index returns but for only 2 ETF returns. For these two predictable ETFs, far fewer rules with significant predictive power can be identified by the proposed stepwise test. The high break-even transaction costs associated with the top rules in those predictable ETFs further suggest that some technical rules may be exploited to make profit in certain emerging markets. Our findings therefore indicate a negative impact of the inception of ETFs on the predictive ability of technical trading rules. This is compatible with the intuition that ETFs allow arbitrageurs to trade away most potential profits in young markets.

To summarize, this paper makes the following contributions to the literature. First, we develop a new test for empirical testing problems in finance that require correction of data snooping bias. Second, we provide new evidence of technical predictability (and potential profitability) of growth and emerging stock markets based on recently available data of ETFs. Last, but not least, this study supports the adaptive market efficiency hypothesis of Lo (2004). Using technical predictability as a barometer of market efficiency, our results suggest that the existence of ETFs effectively improves market efficiency.[1]

This paper proceeds as follows. We summarize the existing tests and introduce the Step-SPA test in Section 2. The simulation results for the Step-SPA test are reported in Section 3. The data and performance measures are discussed in Section 4. The empirical results are presented in Section 5. Section 6 concludes the paper. The proofs and some details of the technical rules considered in the paper are deferred to Appendices.

2. Tests without data snooping bias

Given m models for some variable, let $d_{k,t}$ ($k=1,2,...,m$ and $t=1,2,...,n$) denote their performance measures (relative to a benchmark model) over time. Suppose that for each k, $IE(d_{k,t})=\mu_k$ for all t, and for each t, $d_{k,t}$ may be dependent across k. We wish to determine whether these models can outperform the benchmark and would like to test the following inequality constraints:

$$H_0^k : \mu_k \leq 0, \quad \mathbf{k} = 1, \ldots, \mathbf{m}. \tag{1}$$

For example, we may test if there is any technical trading rule that can generate positive return for an asset. Let r_t be the return of this asset at time t and $\delta_{k,t-1}$ be the trading signal generated by the k-th trading rule at time $t-1$, which takes the values of 1, 0, or -1, corresponding to a long position, no position, and a short position, respectively. Then, $d_{k,t}=\delta_{k,t-1}r_t$ is the realized return of the k-th trading rule, and Eq. (1) is the hypothesis that *no* trading rule can generate positive mean return. Note that $d_{k,t}$ depend on each other because they are based on the same return r_t.

[1] Neely et al. (2009) also suggest that the weakening technical predictability in foreign exchange markets can be explained by the adaptive market efficiency hypothesis.