University of
St Andrews

Centre for Dynamic Macroeconomic Analysis

Illusory Profitability of Technical Analysis in Emerging Foreign Exchange Markets

Pei Kuang, M. Schröder and Q. Wang

CDMA Working Paper Series No. 1302
1 Aug 2013
JEL Classification: C12, F31, G14, G15
Keywords: foreign exchange, technical trading, data mining, bootstrap test, emerging market

Illusory Profitability of Technical Analysis in Emerging Foreign Exchange Markets

P. Kuang[a], M. Schröder[b], Q. Wang[c,*]

[a]*Department of Economics, University of Birmingham*

[b]*Centre for European Economic Research*

[c]*Bangor Business School*

Abstract

We conduct an extensive examination of profitability of technical analysis in ten emerging foreign exchange markets. Studying 25988 trading strategies for emerging foreign exchange markets, we find that best rules can sometimes generate an annually mean excess return of more than 30%. Based on standard tests, we find hundreds to thousands of seemingly significant profitable strategies. Almost all these profits vanish once the data snooping bias is taken into account. Overall, we show that the profitability of technical analysis is illusory.

Keywords: foreign exchange, technical trading, data mining, bootstrap test, emerging market

JEL: C12, F31, G14, G15

1. Introduction

After decades of debate, the profitability of technical trading rules (hereafter "TTR") remains a conundrum. If the market is (weakly) efficient, rational investors should quickly arbitrage away the profits, implying that TTR is useless. If TTR cannot generate persistent profits, why do at least 90% percent of experienced traders place some weight on it in costly trading activity [Taylor and Allen (1992)]?

We show that the profitability of technical analysis is illusory. Studying 25988 trading strategies for emerging foreign exchange markets, we find that best rules can sometimes generate an annually mean excess return of 30%. Applying standard tests, we find hundreds to thousands of seemingly significant profitable strategies.[1] Almost all these profits become insignificant once the data snooping bias is taken into account.

*Corresponding author. Bangor Business School, Hen Goleg, College Road, Bangor LL57 2DG, United Kingdom.

Email addresses: `p.kuang@bham.ac.uk` (P. Kuang), `schroeder@zew.de` (M. Schröder), `q.wang@bangor.ac.uk` (Q. Wang)

[1] We call these strategies "seemingly significant profitable" because the standard tests ignore the effect of data snooping, and in reality they might not be profitable if the data snooping bias is properly controlled for. In the later section, "seemingly significant profitable" TTRs are those detected by individual rule nominal p-values (see Section 5).

Economists have long acknowledged the data snooping bias in this context. To find the desired trading rules, investors usually need to search intensively among a potentially large universe of trading rules on a single historical data. Trying hard and long enough makes it very likely to find seemingly profitable but in fact wholly spurious trading strategies. Such a search can be done by researchers and investors as a whole. The classical statistical inference typically conducted in the previous literature is biased in this case.

How much can the data snooping bias explain the identified profitability? Due to its statistical difficulty, few existing research is able to answer this question. Our paper quantifies the extent of data snooping bias by applying two recently developed and more powerful methods: the StepM test [Romano and Wolf (2005)] and the SSPA test [Hsu, Hsu and Kuan (2010)]. For a given universe of trading strategies, these tests are data-snooping free since they take the entire search process into account, and hence are able to detect the genuine profitable trading rules from the universe. We compare the number of profitable rules from these two tests to that from the classical tests, which do not control for data snooping bias. The comparison can show how many profits are genuine and how many are spurious. This provides richer information about the extent of data snooping bias than only knowing whether the best performing rule has a true predictive ability or not (as in Reality Check [White (2000)] and the SPA test [Hansen (2005)]). Therefore we are able to make the extent of data snooping bias more transparent.

Our study provides a comprehensive test of trading rule profitability across 25,988 trading rules. The majority of these rules are popular among professional traders, but have not been studied in the literature for emerging FX markets. Hence, our paper depicts a more complete picture regarding the performance of trading rules and market efficiency for these markets. A large universe of trading rules raises a concern about the power of tests. However, we show that our major conclusions continue to hold when testing with smaller universes.

A study of TTR profitability for emerging FX market is interesting in its own right. Factors such as spot exchange rate movements, interest rate differentials and transaction costs in emerging markets can contribute to TTR profitability differently from their developed FX market counterparts. In addition, emerging markets have stricter regulation and capital control, which makes it more difficult for speculation to arbitrage away the profits.

The remainder of this paper is structured as follows. Section 2 provides a short literature review. Section 3 presents the universe of the trading rules. Section 4 discusses the empirical methodology. Section 5 briefly documents our data and empirical findings. Section 6 concludes. A detailed documentation of the trading rules considered can be found in the appendix.

2. Literature

The FX market has substantial supportive evidence for the profitability of TTR [e.g., Sweeney (1986), LeBaron (1999), and Qi and Wu (2006)]. Several other studies show the opposite [see Lee and Mathur (1996), and Neely and Weller (2003)]. Most of these studies confined themselves to the currencies of developed economies. It is unclear whether their findings can be carried over to the emerging markets, which themselves are heterogeneous.

Existing studies on the profitability of TTR for emerging FX markets report mixed results (e.g., Martin (2001), Lee, Gleason and Mathur (2001), Pojarliev (2005)). De Zwart, Markwat, Swinkels and van Dijk (2009) provide evidence that combining technical analysis with fundamental analysis can improve the risk-adjusted performance of the investment strategies. These studies do not formally control for the effect of data snooping bias, which is a critical concern in this line of research. A notable exception is Qi and Wu (2006). They are the first to study the TTR profitability with formal data snooping check for developed countries' currencies. They consider a universe of 2,127 simple technical trading rules and find that data snooping biases do not change the conclusion of profitability of trading rules in full sample. But in the second half of the sample the data snooping bias is more serious.

Data snooping has long been considered by academic researchers [Jensen (1967), Jensen and Bennington (1970), Lo and MacKinlay (1990) and Brock, Lakonishok and LeBaron (1992)]. However, a rigorously founded and generally applicable test remained unavailable until White (2000) introduced the Reality Check. The Reality Check can directly quantify the effects of data snooping when testing the best trading rule from the "full universe" of trading strategies. Since then, a couple of studies [e.g., Sullivan, Timmermann and White (1999), White (2000), Hsu and Kuan (2005) and Qi and Wu (2006)] have applied the Reality Check and found mixed results.

Hansen (2005) points out that the power of the Reality Check can be reduced and even be driven to zero when too many poor and irrelevant rules are included in the set of alternatives. Simply excluding poor performing alternatives does not lead to valid inference in general either. To solve this problem, Hansen (2005) proposes a new test statistic for superior predictive ability (hereafter "SPA"), which invokes a sample-dependent distribution under the null hypothesis. This SPA test is more powerful compared to White's Reality Check and less sensitive to the inclusion of poor and irrelevant alternatives.

The question of interest for both the Reality Check and the SPA test is whether the best trading rule beats the benchmark. An investor might want to know whether a particular trading rule is profitable. A researcher may want to test whether a certain trading rule found profitable in the literature indeed outperforms the market. Furthermore, as pointed out by Timmermann (2006), choosing the forecast with the best track record is often a bad idea, and a combination of forecasts dominates the best individual forecast in out-of-sample forecasting experiments. So we may want to know all or some of the profitable trading strategies and combine them for decision making (this is what our complex trading rules do).[2] We may also want to avoid worse strategies when combining trading rules, since trimming them off often helps to improve forecasting performance as found in the forecasting literature [Timmermann (2006)]. Romano and Wolf (2005) modify the Reality Check and propose a stepwise multiple test (hereafter "StepM"), which can detect as many profitable trading rules as possible for a given significance level. Hsu, Hsu and Kuan (2010) propose a more powerful stepwise SPA (hereafter "SSPA") test to combine the SPA test and the StepM test. These two stepwise tests enable us to separate the genuine profitable rules from spurious ones among all seemingly profitable rules obtained from classical tests, and hence provide a more complete picture of the extent of data snooping bias.

3. Universe of Trading Rules

Defining the universe of trading rules is a key step for obtaining valid inference in a superior predictive ability test. On the one hand, having a too small universe may miss the data snooping bias incurred during the search for profitable trading rules. On the other hand, the power of test may be reduced when too many irrelevant trading rules are included. In balancing these considerations, we expand the trading rules to a large universe while keeping it computationally feasible. In total we have 25,988 trading rules, consisting of simple rules, charting rules with kernel smoothing and complex trading rules. Table 1 provides an overview of the universe of trading rules we considered. Since the class of simple trading rules has well been described in the literature, we leave its details to the appendix. In the following, we focus on describing the other two group of trading. The parameters used in these trading strategies are reported in the appendix.

Charting Rules with Kernel Smoothing

Similarly to Lo, Mamaysky and Wang (2000), we consider 5 pairs of technical patterns, applying each to kernel smoothed spot series. These strategies include Head and Shoulders (HS) and Inverse Head and Shoulders (IHS), Triangle (TA), Rectangle (RA), Double Tops and Bottoms (DTB) and Broadening Tops and Bottoms (BTB). Using non-parametric smoothing is to find which degree of smoothing can mimic the eyeball smoothing adopted by investors. Kernel smoothing is an ideal way to smooth the spot rates. The critical issue in kernel smoothing is the choice of bandwidth. Lo, Mamaysky and Wang (2000) recommend a bandwidth of $0.3\times$*(optimal band width)*. The optimal band width is calculated from a cross validation method [Lo, Mamaysky and Wang (2000)].[3] Lo, Mamaysky and Wang (2000) advocate the use of this

[2] Some evidence of unstable performance of the trading strategy is provided in Sullivan, Timmermann and White (1999). They find that the best trading rule applying to DJIA for the period of 1897-1986 does not outperform the benchmark for the period of 1987-1996.

[3] For a detailed discussion on kernel smoothing and cross validation, see Härdle (1990).

Table 1. The Universe of Trading Rules

Trading Rules	Number of rules
Simple trading rules	
Filter Rules (FR)	497
Moving Averages (MA)	2049
Support and Resistance (SR)	1220
Channel Break-Outs (CB)	2040
Momentum Strategies in Price (MSP)	1040
Chart rules with kernel smoothing	
Head and Shoulders (HS, IHS)	3384
Triangle (TA)	1128
Rectangle (RA)	3384
Double Tops and Bottoms (DTB)	3384
Broadening Tops and Bottoms (BTB)	1128
Complex trading rules	
Learning Strategies (LS)	1872
Voting Strategies (VS)	26
Fraction Position Strategies (FPS)	26
Voting by Learning Strategies (VLS)	4810
Total number of trading rules	25988

under-smoothed stock price series since they are neither too volatile nor too smooth, and the professional technical analysts they interviewed feel that such a smoothing is more acceptable than other versions of smoothed series. They admit, however, that such an approach is ad hoc. Jegadeesh (2000) recommends the use of different choices of bandwidth to assess whether their results are sensitive to these choices. Savin, Weller and Zvingelis (2007) consider the multiples 1, 1.5, 2 and 2.5 of the optimal bandwidth. We use three different bandwidths (scaling the optimal bandwidth with 0.3, 1 and 4), as well as the original series, in order to see the effect from under-smoothing, optimal smoothing, over-smoothing and no smoothing. Therefore, each pair of charting rules in this class is applied to four versions of spot rates, depending on the smoothing parameters used.

3.0.1. Head and Shoulders and Inverted Head and Shoulders

Head and Shoulders occurs when the second (head) of three consecutive peaks exceeds the first (left shoulder) and the third (right shoulder). The minima between left (right) shoulder and the head are called left (right) troughs. We require the two shoulders (troughs) to be approximately equal such that their differences are no more than a differential rate x. The HS pattern is completed when the adjacent local minimum of the right shoulder penetrates the neckline and the band. Short position in foreign currency is taken exactly on the day when the price crosses the neckline and the band. Following Chang and Osler (1999), two exit rules for HS strategy are considered, namely endogenous and exogenous. For the endogenous one, we distinguish two kinds of situations. We define a cutoff as y times the standard deviation of the daily exchange rate change. If the price falls by d percent times difference between head and average trough(the difference is referred to as "measuring objective" or "price objective" in technical manual), we exit on the day when the price has risen above a local minimum by the cutoff percentage, which implies that the price has conclusively stopped moving in the predicted direction. In the second case, if the price does not fall by such an amount, we allow for a possible bounce or interruption, that is, the price may temporarily move back toward the neckline. When the second trough falls below the aforementioned *d* percent line, we are back to the former case. Otherwise, we liquidate the position at the second trough. In both situations, in order to limit the loss, a stop-loss line is incorporated whenever the price goes sufficiently far in the wrong direction. An exogenous exit rule means we close our position after holding for an exogenous specified number of days *f*. As the name tellingly reveals, the inverted Head and Shoulders is simply an inverse version of Head and Shoulders; once the IHS is completed, the speculator expects an upward trend in the future spot exchange rate. As a result he or she will borrow the British Pound to buy the foreign currency.

3.0.2. Triangle

Triangle is one of the reverse patterns which is also based on pricing movements showing five consecutive local extrema. Triangle tops (TTOP) are characterized by three descending local maxima and two ascending local minima. Triangle bottoms (TBOP) are characterized by three ascending local minima and two descending local maxima. Once a triangle is completed, it will constitute a signal for taking a long (short) position in foreign currency if the future closing spot exchange rate exceeds the latest top (or falls below the latest bottom) by a fixed proportion x, known as the "trend filter". We consider similar liquidation methods as for HS.

3.0.3. Rectangle

The rectangle pattern is also characterized by five consecutive local extrema. Rectangle tops (RTOP) require three tops and two bottoms to lie near an upper horizontal lines, that is, within x percent of their average, respectively. Moreover, we require the lowest top to be higher than the highest bottom. Similarly, the rectangle bottoms (RBOT) require two tops and three bottoms lie near the upper horizontal lines. Signals are generated in a similar way as the Triangle rule, so are the liquidation methods.

3.0.4. Double Tops and Bottoms

The double tops (bottoms) are characterized by two tops (bottoms) that lie near an upper horizontal lines, with one bottom (top) lies in between. Following Lo, Mamaysky and Wang (2000), we require the two tops (bottoms) to occur at least a month, or with 22 intermediatory trading days. In addition, the second top (bottom) should be higher (lower) than all the local maxima (minima) in between.

3.0.5. Broadening Tops and Bottoms

Similarly to the Triangle class, the Broadening Tops and Bottoms are characterized by five consecutive local extrema. Broadening Tops (BTOP) requires 3 descending local maxima and two ascending local minima. Broadening Bottoms requires 3 descending local minima and two ascending local maxima. Compared to the Triangle class, this class presupposes a "divergence" shape, while the Triangle class requires a "convergence" shape.

Complex Trading Rules

Single rules can generate false signals when prices fluctuate in a broad sideways pattern. Relying on a single rule can be a dangerous practice, even for the historical best rule. Technical analysts can combine other trading rules to confirm the prediction of price direction. Following Hsu and Kuan (2005), we consider three classes of complex trading rules: the learning strategy (LS), the voting strategy (VS), and the fractional position strategy (FPS). Besides these complex trading rules, we also propose a new class of complex rules, namely the voting by learning strategy (VLS), which combines the voting and learning rules.

3.0.6. Learning Strategy (LS)

A learning strategy assumes that after an investor learns about the strategies' performances from the past m days (memory span) within the certain class, he can switch his position by following the best strategy found during the memory span. After his switch, he waits for r days (review span) and then reevaluates the strategies' performances of the past m days (memory span) to decide whether to switch. For the evaluation of the trading rules, we use the mean return and the Sharpe ratio during the memory span as the performance measure.

3.0.7. Voting Strategy (VS)

Voting strategy is a system of voting by trading rules within each class. Each rule is assigned one vote for recommending a ballot. We consider the three choice ballot, where either a long, a short or no position can be voted. The decision follows the recommendation of majority votes. To avoid the voting result being dominated by a class with a large number of rules, we consider every class of non-complex rules separately but not the one consisting of all rules.

3.0.8. Fractional Position Strategy (FPS)

Note that both the learning strategy and the voting strategy yield the signal as an integer. The fractional position strategy, in contrast, allows to take the position of a non-integer between -1 and 1. The fraction of a position is determined by an "evaluation index". In our case, it is only the fraction of winning votes that recommend the same positions relative to all votes within the same class.

3.0.9. *Voting by Learning Strategy(VLS)*

To implement this strategy, we consider the best n trading rules evaluated from the memory span within each class of non-complex rules, and then assign votes to these rules as in the voting strategy class. The decision follows the recommendation of majority votes from the best n trading rules. In addition, we consider the case in which the top n trading rules are selected from the set of all non-complex trading rules.

4. Test with Data Snooping Check

In this section, we discuss the empirical tests used in this paper. The discussion is kept informal and heuristic, with more technical description provided in Appendix C.

4.1. *Performance Measures*

Suppose that our universe of trading rules includes m rules. Let δ^k_{t-1} be a "signal" function which generates a trading signal by the k^{th} trading rule using information up to $t-1$. This signal function can assume three values that instructs a trader to take a short position ($\delta^k_{t-1} = -1$), a long position ($\delta^k_{t-1} = 1$), or no position ($\delta^k_{t-1} = 0$) in a foreign currency at time $t-1$. The k^{th} trading rule yields the profit as:

$$R^k_t = (s_t - s_{t-1} + r^*_t - r_t)\delta^k_{t-1} - abs(\delta^k_{t-1} - \delta^k_{t-2})g \tag{1}$$

where R^k_t is the excess return from trading on the currency in period t using k^{th} trading rule, s_t is the logarithm of the spot exchange rate (British Pound price of one unit foreign currency) at time t; r_t and r^*_t are domestic and foreign interest rates from time t-1 to t, respectively; g is a one way transaction cost.

We consider two performance measures: the mean excess return ($E(R^k_t)$) and the Sharpe ratio.[4] We use a natural benchmark that the investor does not take position in the foreign exchange market and hence earns a zero excess return (alternatively, zero Sharpe ratio). Therefore our performance measure is in fact a relative performance measure [$d_{k,t}$ in the notation of Hansen (2005)] of k^{th} trading strategy compared to the benchmark.

4.2. *Empirical Tests*

Reality Check

White (2000) tests the null hypothesis that the benchmark is not inferior to any of the alternative trading rules. Suppose we search in 1000 trading rules to test whether the best trading rule from these 1000 trading

[4]The Sharpe ratio used here is defined as $\frac{E(R^k_t)}{\sigma}$, where σ is the standard deviation of excess returns. We calculate σ through the circular block bootstrap to account for serial correlations in returns. We use Algorithm 3.1 in Ledoit and Wolf (2008) to choose the block size. This resampling method can account for unknown dependence structure in excess returns.

rules is significantly better than a benchmark. Instead of comparing one trading rule to the benchmark as in the classical test, we look at a vector of 1000 relative performance measure at the same time. Each element in the vector represents the relative performance of one model in the universe. Testing whether the best rule is profitable is equivalent to ask whether the **maximal** value in this **vector** is significantly larger than zero. The distribution of the maximum in a vector of elements differs from the distribution of the element in the vector. The latter is used for inference in classical tests. Doing this gives consideration to the full set of models underlying the vector that led to the best performing trading rule. Rejecting the null hypothesis implies that at least one trading rule beats the benchmark.

Due to the complication of the true distribution of the test, White (2000) recommends the stationary bootstrap method of Politis and Romano (1994) to first obtain the empirical distribution of the test statistic, and obtain the p-value by comparing it with the quantiles of the empirical distribution from bootstrap re-sampling. If the p-value is smaller than a given significance level, the null hypothesis is rejected.

SPA test

The null hypothesis of SPA test [Hansen (2005)] is the same as in White's Reality Check. Unlike the White's Reality Check, the SPA test uses the studentized test statistic, which will typically improve the power. Hansen (2005) provides a concrete example for highlighting the advantage of studentizing the individual statistics, since it avoids a comparison of the performance measured in different "units of standard deviation". Furthermore, the SPA test invokes a sample-dependent distribution under the null hypothesis, which can discard the poor models asymptotically. Therefore, the new test is more powerful and less sensitive to the inclusion of poor and irrelevant alternatives. The improvement of the power of the SPA test over the Reality Check is further confirmed by the simulation experiment conducted in Hansen (2005). The p-value of SPA test is calculated by bootstrapping its empirical distribution.

StepM test

Both the Reality Check and the SPA test seek to answer whether the best trading strategy beats the benchmark. As discussed in section 1, it is often more interesting to identify all outperforming trading rules, or to know whether a particular trading rule improves upon the benchmark. The Reality Check can be modified easily for identifying potential strategies that beat the benchmark, but Romano and Wolf (2005) show that this is only suboptimal, and only amounts to the first step of StepM test of Romano and Wolf (2005), which can detect more good strategies from the second step on. The StepM test is therefore more powerful than the Reality Check in detecting superior trading rules. The aim of the StepM test is to find as many profitable trading rules as possible. They test whether the individual trading rule is better than the benchmark by checking whether it falls into a joint confidence region constructed from all trading strategies. If not, that trading strategy is detected as profitable. When some trading strategies are detected, the remaining trading strategies can be used to construct a new joint confidence region. Then the first step is repeated to find whether there are remaining profitable trading strategies. One continues this way until no profitable rules can be detected.

Romano and Wolf (2005) also propose the use of studentization to improve the power and level properties of the StepM test.

Stepwise SPA test

The Stepwise SPA test of Hsu, Hsu and Kuan (2010) aims at combining the advantages of both the SPA test and the StepM test, to improve upon the White's Reality Check in two different ways. In this setting, the null hypothesis is similar to the StepM test, though it invokes a sample-dependent distribution under the null hypothesis, such that poor and irrelevant trading rules will be discarded asymptotically. They provide a formal proof and simulation results to demonstrate that the SSPA test is more powerful than the Reality Check, SPA test, and StepM test.

5. Empirical Results

Data

We collect daily spot exchange rate data from Datastream, which covers the period from January 1994 to July 2007. The spot exchange rate is quoted as foreign currency units per British pound (hereafter GBP). We study the countries which have both a spot exchange rate and an overnight interest rate available (if no overnight interest rate is available, we use other daily short rate instead). In total, we include ten currencies from emerging markets, which represent various geographic regions. The data for most of them starts in the 1990s. All considered currencies and their data availability are reported in table 2.

Table 2. Data Used for Testing the Profitability of Trading Rules

All data are from Datastream.

Currency	Symbol	Available data	First Sub-period	Out-of-Sample period
Brazil, Reais	BRL	01/2000-07/2007	01/2000-07/2002	08/2002-07/2007
Czech Republic, Koruny	CZK	01/1998-07/2007	01/1998-07/2002	08/2002-07/2007
Hungary, Forint	HUF	09/1995-07/2007	09/1995-07/2002	08/2002-07/2007
India, Rupees	INR	01/1994-07/2007	01/1994-07/2002	08/2002-07/2007
Indonesia, Rupiahs	IDR	01/1999-07/2007	01/1999-07/2002	08/2002-07/2007
Mexico, Pesos	MXN	01/1995-07/2007	01/1995-07/2002	08/2002-07/2007
Poland, Zlotych	PLN	01/1995-07/2007	01/1995-07/2002	08/2002-07/2007
South Africa, Rand	ZAR	01/1996-07/2007	01/1996-07/2002	08/2002-07/2007
Thailand, Baht	THB	01/1999-07/2007	01/1999-07/2002	08/2002-07/2007
Turkey, New Lira	TRY	08/2002-07/2007	n.a.	n.a.

5.1. Full Sample Results

In this section we report the performance of trading rules using the full sample. We first present the data snooping bias in the best performing rules, as those reported in the literature. We then report the bias in all seemingly profitable rules.

5.1.1. Biases in the most profitable trading rules

Transaction costs are an important concern when testing TTR profitability. Lee, Gleason and Mathur (2001) uses one-way transaction costs of 0.1% for Latin American Countries' currencies. Burnside, Eichenbaum and Rebelo (2007) mention that the emerging countries' bid-ask spreads are usually 2-4 times higher than the developed counties' currencies. Qi and Wu (2006) use transaction costs of 0.04% for developed FX markets. So we choose 0.1% as one-way transaction costs. We have also considered transaction costs of 0%, 0.04%, and 0.3%. Results with these different transaction costs are qualitatively similar.

Table 3. Performance of the Best FX Trading Rules in Emerging Market (with 0.1% one way transaction cost)

Results are based on the whole universe containing 25988 trading rules. The mean excess return and Sharpe ratio are both annualized. The whole sample data are used for detecting the best rules.

Panel A: Mean Return Criteria

Currency	Nr. of Trades	Mean Excess Return	P-value (RC)	P-value (SPA_l)	P-value (SPA_c)	P-value (SPA_u)
BRL	114	0.30	0.43	0.27	0.30	0.37
CZK	75	0.05	1.00	1.00	1.00	1.00
HUF	15	0.05	1.00	0.81	0.89	0.94
INR	309	0.05	1.00	1.00	1.00	1.00
IDR	23	0.14	1.00	1.00	1.00	1.00
MXN	188	0.14	0.21	0.14	0.17	0.21
PLN	79	0.08	1.00	1.00	1.00	1.00
TRY	57	0.19	1.00	1.00	1.00	1.00
THB	3	0.05	1.00	1.00	1.00	1.00
ZAR	235	0.18	0.49	0.89	0.93	0.99

Panel B: Sharpe Ratio Criteria

Currency	Nr. of Trades	Sharpe Ratio	P-value (RC)	P-value (SPA_l)	P-value (SPA_c)	P-value (SPA_u)
BRL	60	1.72	0.12	0.04	0.04	0.12
CZK	140	1.02	1.00	1.00	1.00	1.00
HUF	168	1.12	0.75	0.16	0.16	0.64
INR	4	0.65	1.00	1.00	1.00	1.00
IDR	14	1.17	1.00	0.48	0.49	0.98
MXN	154	1.29	0.07	0.02	0.02	0.09
PLN	79	0.78	1.00	1.00	1.00	1.00
TRY	50	1.51	1.00	0.92	0.92	0.99
THB	148	1.08	1.00	1.00	1.00	1.00
ZAR	235	1.11	0.95	0.70	0.70	0.96

Table 3 documents the performance of the best trading rules according to both the mean excess return (panel A) and Sharpe ratio (panel B). Both the mean excess returns and the Sharpe ratio are positive and large for all currencies. The annual mean excess return ranges from 0.05 (Thai Baht) to 0.30 (Brazilian Real). The annualized Sharpe ratio lies between 0.65 (Indian Rupee) and 1.72 (Brazilian Real).[5]

Panel A and B also report four p-values (one p-value from the Reality Check and three p-values from the SPA test, including the upper bound and lower bound p-values).[6] A close look at p-values of the Reality Check and the SPA tests shows no profitable rules according to the mean excess return criteria. For the Sharpe ratio, we find that two out of ten currencies (Brazilian Real and Mexican Peso) have profitable rules, even after the data-snooping bias is controlled for.[7]

Following Sullivan, Timmermann and White (1999), we also calculate the best rule nominal p-value by applying the Reality Check to each best trading rule. Although the best rule is obtained from a search among all trading rules, the best rule nominal p-value does not account for the search, it is therefore subject to data snooping bias. We find the smallest best rule nominal p-values very close to zero for every currency (not reported in the table). Note that the best performing rule from the original sample is not necessarily the best performing rule from bootstrap samples, which has the smallest best rule nominal p-value calculated. Its past performance may simply stem from pure luck. The reported four p-values can be compared to the smallest best rule nominal p-value, which is an easy but rigorous way of quantifying the effect of data-snooping bias. Not surprisingly, all the p-values with data snooping check are bigger than the smallest best rule nominal p-values, which ignore the data snooping bias.

In general, since we have the smallest best rule nominal p-value close to zero for every currency, we will find profitable rules if the data snooping bias is not considered. Therefore, ignoring the data snooping effect leads to a significant bias.

5.1.2. Bias in all "profitable" trading rules

It is often desirable to know all profitable trading rules, as discussed in the previous sections. We detect all the profitable rules according to the individual rule nominal p-value, the StepM test and the SSPA test. The individual rule nominal p-value is obtained from applying the Reality Check to each individual trading rule. When many trading rules are evaluated using individual rule nominal p-value, some are bound to appear to be profitable by chance alone [Romano and Wolf (2005)]. We report the total number of profitable

[5] Daily Sharpe ratio is annualized as in Lo (2002) with the number of lags selected from Akaike information criterion (AIC).

[6] To calculate the p-values, we apply circular block bootstrap with a block length of 2 and 500 bootstrap replications. Our results hardly change when we use a different block length or other bootstrap procedures such as the stationary bootstrap or the moving block bootstrap.

[7] When zero transaction cost is assumed, we find two (five) currencies have profitable rules according to mean excess return (Sharpe ratio) criteria. In the case of a 0.04% one-way transaction cost, we find no (three) profitable rules according to the mean excess return (Sharpe ratio) criteria. Assuming a transaction cost of 0.3%, we find no profitable rules under either criteria.

trading rules detected by the individual rule nominal p-value and compare them to those found by the StepM or the SSPA test. This comparison can serve as a way of summarizing the danger of data snooping for *all* profitable rules according to the individual rule nominal p-value.

Table 4. Number of Profitable Trading Rules

All profitable trading rules are detected at 5% level. The universe of trading rules contains 25988 rules. The whole sample data are used for detecting the profitable rules.

One-way transaction cost of 0.1%

	Mean Excess Return Criteria			Sharpe Ratio Criteria		
Currency	Nominal P-value	StepM	SSPA	Nominal P-value	StepM	SSPA
BRL	2599	0	0	2744	0	1
CZK	117	0	0	192	0	0
HUF	338	0	0	417	0	0
INR	198	0	0	419	0	0
IDR	396	0	0	657	0	0
MXN	1139	0	0	1316	0	24
PLN	323	0	0	358	0	0
TRY	680	0	0	871	0	0
THB	264	0	0	1224	0	0
ZAR	1166	0	0	1361	0	0

Table 4 reports the number of profitable trading rules with one-way transaction costs of 0.1% under both the mean excess return and the Sharpe ratio criteria. we find hundreds or thousands of trading rules for every currency according to both criteria when we consider the individual rule nominal p-value. When the data snooping bias is taken into account, we find no profitable rules for all ten currencies according to the mean excess return. According to the Sharpe ratio criteria, the StepM test finds no profitable rules. Only the SSPA test finds some profitable rules for two out of ten currencies under the Sharpe ratio criteria. Even in these two cases, the total number of trading rules detected to be profitable is quite small compared to those found by individual rule nominal p-values: 1 profitable rule for the Brazilian Real and 24 profitable rules for the Mexican Peso, which is in stark contrast to 2774 profitable rules for the Brazilian Real and 1316 profitable rules for the Mexican Peso detected by individual rule nominal p-values. These striking results indicate that the data snooping bias is substantial.

Reducing the transaction costs to zero enables us to find much more (typically more than one thousand) profitable trading rules for each criteria for every currency when data snooping bias is ignored.[8] When it is taken into account, we find the total number of trading rules detected to be profitable is usually less than one percent of those found by individual rule nominal p-values.

We are not claiming the optimality of the tests used. Such optimal tests are rare in composite hypothesis

[8]Results are not reported but are available from the authors upon request.

testing. In both StepM tests and SSPA tests, we do not allow for false discovery of unprofitable rules. When we are willing to tolerate some false discoveries (as in the tests of controlling for false discovery proportion (FDP)), we might be able to find more profitable rules.[9] We implement the FDP-StepM method suggested by Romano, Shaikh and Wolf (2008) which provides asymptotic control of FDP. Although the results are not tabulated in the paper, doing this does not change our major conclusions. For example, since we find the StepM test detects no profitable rules with one way transaction cost of 0.1%, our results remain the same even after controlling for FDP. Reducing assumed transaction costs to zero or 0.04% does increase the number of profitable rules, but the number of profitable rules are still small comparing to thousands of profitable rules found when ignoring the data snooping bias.

5.2. Sub-sample Analysis

Trading rule performance is often not stable. Today's profitable rule does not guarantee sure profits for tomorrow. For example, Sullivan, Timmermann and White (1999) find that the best trading rule applying to DJIA for the period of 1897-1986 does not outperform the benchmark for the period of 1987-1996. Qi and Wu (2006) find the profitability has declined in recent periods. For this reason, we document the trading rule profitability in two subperiods of our sample. The second subperiod is between 08/2002 and 07/2007 for each currency. Having the same length of second subperiods facilitates a comparison across currencies. The first subperiod covers the rest of the whole sample period (see Table 2). Since our sample of TRY starts from 08/2002, we exclude it from our sub-sample analysis.

In unreported tables, we find the best rules have almost zero best rule nominal p-values in both sub-samples for all those transaction costs we have considered. After taking data mining bias into account, almost all these rules become insignificant. In addition, we find the best rules vary across different currencies and sub-periods.

5.3. Out-of-sample Test

An out-of-sample test is often considered as an effective way of detecting a data snooping bias. For example, if one trading rule is wrongly identified as profitable by looking at the individual rule nominal p-value, the same rule is unlikely to perform well in a new sample. Investors and researchers sometime conduct such out-of-sample tests to check whether previously identified profitable trading rules are still profitable in the new data (for example, Neely, Weller and Ulrich (2009)).

Still, caution has to be taken, when the same new data is used repeatedly for testing various models in out-of-sample experiments. This can occur if investors search (individually and as a whole) among large

[9] We thank one anonymous referee for pointing out this possibility.

number of rules that are profitable in both historical and new data.[10] In this paper we also investigate this potential bias.

We conduct the out-of-sample experiment as follows. First, we take the first sub-sample as the in-sample data, and detect all profitable trading rules according to the individual rule nominal p-value. That is, we ignore the data snooping bias. The detected profitable rules are then taken as the universe of trading rules for our out-of-sample analysis. The out-of-sample test uses the second sub-sample and detects all profitable trading rules using tests with and without a data snooping check.

Table 5. Number of profitable trading rules in an out-of-sample experiment

All profitable trading rules detected at the 5% level. The TTR universe for sub-sample analysis is the whole universe of 25988 rules. The TTR universe for the out-of-sample test consists of the profitable rules found in the first sample according to individual rule nominal p-values. "NP" refers to the number of profitable rules according to individual rule nominal p-values. "OOS" refers to out-of-sample tests. We assume one-way transaction costs of 0.1%. Sub-sample periods are defined in Table 2.

Panel A: Mean Excess Return Criteria

Currency	$NP_{Sample1}$	$NP_{Sample2}$	NP_{OOS}	$StepM_{OOS}$	$SSPA_{OOS}$
BRL	4665	1903	427	0	0
CZK	131	210	52	0	0
HUF	159	150	97	0	0
INR	219	102	1	0	0
IDR	296	285	12	0	0
MXN	1234	300	72	0	0
PLN	438	296	2	0	0
THB	44	119	1	0	0
ZAR	980	97	10	0	0

Panel B: Sharpe Ratio Criteria

Currency	$NP_{Sample1}$	$NP_{Sample2}$	NP_{OOS}	$StepM_{OOS}$	$SSPA_{OOS}$
BRL	6320	2188	704	0	0
CZK	311	328	75	0	0
HUF	485	205	106	0	0
INR	572	212	9	0	0
IDR	753	1395	22	0	0
MXN	1400	430	83	0	0
PLN	501	372	12	1	1
THB	481	1030	44	0	0
ZAR	1094	232	21	0	0

Table 5 reports the number of trading rules in the out-of-sample tests when we impose a one-way transaction cost of 0.1%. "$NP_{Sample1}$" refers to the number of profitable trading rules found in the first sub-sample when the data snooping bias is ignored. "NP_{OOS}" stands for the number of profitable trading rules found in the out-of-sample test without controlling for a data snooping bias. This implies that all these rules are found to be profitable both in-sample and out-of-sample according to the individual rule nominal

[10] Inoue and Kilian (2004) point out that out-of-sample tests can have as serious data mining bias as in-sample test.

p-value. "StepM$_{OOS}$" and "SSPA$_{OOS}$" report the number of profitable rules detected out-of-sample according to the StepM test and the SSPA test, respectively. The difference between NP$_{OOS}$ and StepM$_{OOS}$ or SSPA$_{OOS}$ provides a quantification of the extent of the out-of-sample data snooping bias. Indeed, we find that ignoring the repeating testing in out-of-sample experiments can bias the results. For example, with a one-way cost as high as 0.1%, there are 704 trading rules incorrectly detected as profitable for Brazilian Real under Sharpe ratio criteria. Both the StepM test and SSPA test only find one rule to be profitable out-of-sample, which is free of data snooping bias. Results with other transaction costs are qualitatively similar.

5.4. Reduced Universe of Trading Rules

Our universe of trading rules is the largest in number in the FX trading rule literature so far. This, however, raises the concern about the power of the test, as the number of trading rules is large and is usually five times higher than the number of the observations.

Table 6. Number of Profitable Trading Rules (Reduced Universe)

All profitable trading rules are detected at the 5% significance level. We assume that there is no transaction cost involved. The whole sample data are used for detecting the profitable rules.

Panel A: Filter Rule class (total number of trading rules: 497)

	Mean Excess Return Criteria			Sharpe Ratio Criteria		
Currency	Nominal P-value	StepM	SSPA	Nominal P-value	StepM	SSPA
BRL	80	0	0	88	0	0
CZK	6	0	0	8	0	0
HUF	17	0	0	20	0	0
INR	19	0	0	20	0	0
IDR	18	0	0	19	0	0
MXN	6	0	0	6	0	0
PLN	32	0	0	33	0	0
TRY	32	0	0	27	0	0
THB	0	0	0	0	0	0
ZAR	25	0	0	24	0	0

Panel B: Moving Average class (total number of trading rules: 2049)

	Mean Excess Return Criteria			Sharpe Ratio Criteria		
Currency	Nominal P-value	StepM	SSPA	Nominal P-value	StepM	SSPA
BRL	476	0	0	438	0	0
CZK	4	0	0	3	0	0
HUF	16	0	0	16	0	0
INR	16	0	0	16	0	0
IDR	107	0	0	114	0	0
MXN	13	0	0	16	0	0
PLN	90	0	0	96	0	0
TRY	11	0	0	14	0	0
THB	0	0	0	0	0	0
ZAR	102	0	0	96	0	0

To address this issue, we consider a reduced universe, which contains the trading rules in each class.[11] The size of these universes ranges from 497 (filter rule) to 3,384 (Head and Shoulders for example). By doing this, we assume that the professional traders and researchers only search for profitable rules within each class. We have also considered a universe which either contains all simple trading rules, or all charting rules with kernel smoothing.

We report the results for the filter rule and the moving average rule, which are the most frequently studied classes in the FX trading rules literature. Results based on other classes of trading rules are similar, and are available from the authors upon request.

Table 6 shows that even in the absence of transaction costs and a small universe of trading rules, we detect no profitable trading rules according to the StepM and the SSPA test for all currencies. Ignoring the data snooping bias involved, however, one can find much more profitable trading rules based on the individual rule nominal p-values. The extent of data snooping bias is substantial. Given the results in these two examples and similar results for unreported universes, we conclude that our major results, namely that the data snooping bias is substantial and that evidence supporting the TTR profitability is rare, are robust to a smaller universe of trading rules.

6. Conclusion

There is substantial empirical evidence documenting the success of the TTR in various markets, especially in FX markets. Critics, however, cite data snooping as one major source of this anomaly.

The purpose of this paper is to quantifying the extent of data snooping bias in technical trading rule profitability. We test trading rule performances across 25,988 trading rules for the emerging foreign exchange markets. We find that nearly all seemingly successes of TTR in our trading rule universe are in fact the result of data mining bias. Our findings continue to hold when we consider different transaction costs, sub-sample analysis and various reduced universes of trading rules.

Our empirical results also provide a first comprehensive evidence of the technical trading rule profitability for 10 emerging countries' exchange rates. The universe of our trading rules includes simple rules, charting rules with kernel smoothing and complex trading rules. Although popular among professional traders, the majority of these trading rules have not been studied in the literature for emerging FX markets. Our paper depicts a more complete picture of the performance of trading rules and market efficiency for emerging foreign exchange markets. Overall we find rare evidence against the efficiency of emerging FX markets.

[11] We do not consider complex rules here, since they are based on simple rules and charting rules with kernel smoothing, which search implicitly in large universes.

Investors have to be careful when applying TTR to these markets, given the substantial data snooping bias that is likely to be involved. Further research, such as the investigation of economic explanations for trading rules' profitability or determinants of its cross country variation, also needs to account for the danger of data snooping.

Appendix A: Description of Simple Trading Rules

Our descriptions of simple trading rules draw heavily on Sullivan, Timmermann and White (1999), Qi and Wu (2006) and Hsu and Kuan (2005), though we have made some modifications to avoid ambiguities.

Filter Rules

The filter rule strategy for generating a trading signal follows Sullivan, Timmermann and White (1999) and Qi and Wu (2006). The basic filter rule could be stated as follows: if the daily closing price (in British Pound) of a foreign currency moves up by $x\%$ or more from its most recent low, the speculator borrows the British Pound and uses the proceeds to buy and hold the foreign currency until its price moves down at least $x\%$ from a subsequent high, at which time the speculator short sells the foreign currency and uses the proceeds to buy the British Pound. Two definitions of the subsequent high (low) are considered. One is the highest (lowest) closing price over the period of holding a particular long (short) position. The alternative high (low) refers to the highest (lowest) price over the e most recent days. We also consider that a given long or short position is held for prespecified c days during which period all other signals are ignored.

Moving Average Rules

Moving averages are among the oldest trading rules used by chartist. The (equally weighted) moving average of a currency price for a given day t over the n days is $\frac{1}{n}\sum_{i=0}^{n-1} s_{t-i}$. Under a simple single moving average rule, when the current price is above the moving average by an amount larger than the band with $b\%$, the speculator borrows the British Pound to buy the foreign currency. Similarly, when the current price is below the moving average by $b\%$, the speculator short sells the foreign currency to buy the British Pound. Under dual moving average rule, when the short moving average of a foreign currency price is above the long moving average by an amount larger than the band with $b\%$, the speculator borrows the British Pound to buy the foreign currency. If the short moving average of a foreign currency price penetrates the long moving average from above, the speculator short sells the foreign currency to buy the British Pound. Following Sullivan, Timmermann and White (1999) and Qi and Wu (2006), we implement the moving average rules with a time delay filter in addition to the fixed percentage band filter as described above. The time delay filter requires that the long or short signals remain valid for d days before action is taken. Similar to the filter rule case, we also consider holding a given long or short position for c days during which period all other signals are ignored.

Trading Range Break (or Support and Resistance) Rules

The support and resistance level refers to certain price levels acting as barriers to prevent traders from pushing the price of an underlying asset in a certain direction. Under a trading range break rule, when the price of a foreign currency moves above the maximum price (resistance level) over the previous n days by b%, the speculator borrows the British Pounds to buy the foreign currency. When the price falls below the minimum price over the previous n days by b%, the speculator sells short the foreign currency and buys the British Pound. Alternatively, we use the local maximum (minimum), which refers to the highest (lowest) price over the e most recent days, as the definition for the resistance level. Here we also allow a time delay filter, d, as well as the holding period of c days to be included, as in the case of moving average rules.

Channel Breakout Rules

A channel occurs when the high price of a foreign currency over the previous n days is within x% of the low over the previous n days. Under a channel breakout rule, when the closing price of the foreign currency exceeds the channel by b%, a signal is generated for the speculator to borrow the British Pound and buy the foreign currency. Likewise, when the closing price of the foreign currency drops below the channel by b%, a signal is generated for the speculator to short sell the foreign currency and buy the British Pound. Again, we consider a holding period of c days.

Momentum strategies

A momentum strategy attempts to predict the strength or weakness of the current market based on an "oscillator" constructed from a momentum measure. We follow Hsu and Kuan (2005) and use the rate of change (ROC) as the momentum measure. The m-day ROC at time t is defined as the change of spot exchange rate divided by the closing spot exchange rate at time $t - m$. Two oscillators are considered: the simple oscillator (which is just m-day ROC), and the moving average oscillator (which is the w-day moving average of m-day ROC with $w \leq m$). An overbought/oversold level k (say 5% or 10%) is needed to determine whether a position should be initiated. When the oscillator penetrates the overbought level from below, the speculator borrows the British Pound to buy the foreign currency. If the oscillator crosses the oversold level from above, the speculator sells short the foreign currency and buys the British Pound. Once again, we consider a holding period of c days.

Appendix B: Documentation of Trading Rules Parameters

Simple Rules

Filter Rules (FR)

x: increase in the log pound value of foreign currency required to generate a "buy" signal

y: decrease in the log pound value of foreign currency required to generate a "sell" signal

e: the number of the most recent days needed to define a low (high) based on which the filters are applied to generate "long" ("short") signal

c: number of days a position is held during which all other signals are ignored

$x = 0.0005, 0.001, 0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, 0.045, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.12, 0.14, 0.16, 0.18, 0.2, 0.25, 0.3$ (24 values)

$y = 0.0005, 0.001, 0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.04, 0.05, 0.075, 0.1$ (12 values)

$e = 1, 2, 3, 4, 5, 10, 15, 20$ (8 values)

$c = 5, 10, 25, 50$ (4 values)

Noting that y must be less than x, there are 185 (x,y) combinations

Number of rules in FR class = $x + x \times e + x \times c +$ ((x,y) combinations) = 24 + 192 + 96 + 185 = 497

Moving Average Rules (MA)

n: number of days in a moving average

m: number of fast-slow combinations of n

b: fixed band multiplicative value

d: number of days for the time delay filter

c: number of days a position is held, ignoring all other signals during that time

$n = 2, 5, 10, 15, 20, 25, 30, 40, 50, 75, 100, 125, 150, 200, 250$ (15 values)

$m = \sum_{i=1}^{n-1} i = 105$

$b = 0.0005, 0.001, 0.005, 0.01, 0.02, 0.03, 0.04, 0.05$ (8 values)

$d = 2, 3, 4, 5$ (4 values)

$c = 5, 10, 25, 50$ (4 values)

Number of rules in MA class: $= n{+}m{+}b{\times}(n{+}m){+}(d{\times}(n{+}m){+}c{\times}(n{+}m){+}9 = 15{+}105{+}960{+}480{+}480{+}9 =$ 2049

Support and Resistance (SR, or Trading Range Break) Rules

n: number of days in the support and resistance range;

e: used for an alternative definition of extrema where a low (high) can be defined as the most recent closing price that is less (greater) than the n previous closing prices;

b: fixed band multiplicative value;

d: number of days for the time delay filter;

c: number of days a position is held, ignoring all other signals during that time

$n = 5, 10, 15, 20, 25, 50, 100, 150, 200, 250$ (10 values);

$e = 2, 3, 4, 5, 10, 20, 25, 50, 100, 200$ (10 values);

$b = 0.0005, 0.001, 0.005, 0.01, 0.02, 0.03, 0.04, 0.05$ (8 values);

$d = 2, 3, 4, 5$ (4 values);

$c = 5, 10, 25, 50$ (4 values);

Number of rules in SR class = $[(1{+}c){\times}(n{+}e)]{+}[b{\times}(n{+}e){\times}(1{+}c)]$ $+[d{\times}c{\times}(n{+}e)]$= 100+800+320 = 1220

Channel Breakout Rules (CBO)

n: number of days for a channel

x: difference between the high price and the low price (x×low price) required to form a channel

b: fixed band multiplicative value ($b < x$)

c: number of days a position is held, ignoring all other signals during that time

$n = 5, 10, 15, 20, 25, 50, 100, 150, 200, 250$ (10 values);

$x = 0.001, 0.005, 0.01, 0.02, 0.03, 0.05, 0.075, 0.10$ (8 values)

$b = 0.0005, 0.001, 0.005, 0.01, 0.02, 0.03, 0.04, 0.05$ (8 values)

$c = 1, 5, 10, 25$ (4 values)

Noting that b must be less than x. There are 43 (x,b) combinations.

Number of rules in CBO class = $n \times x \times c + n \times c \times$((x,b) combinations)= 320 + 1720 = 2040

Momentum Strategies in Price (MSP)

m: number of days rate of change in price;

w: number of days in moving average;

k: overbought/oversold level;

f: fixed holding days;

$m = 2, 5, 10, 20, 30, 40, 50, 60, 125, 250$ (10 values);

$w = 2, 5, 10, 20, 30, 40, 50, 60, 125, 250$ (10 values);

$k = 0.05, 0.10, 0.15, 0.2$ (4 values);

$f = 5, 10, 25, 50$ (4 values);

Noting that w must be less than or equal to m, there are 55 $w - m$ combinations.

Number of rules in MSP class = $m \times k \times f$+ ((m,w) combinations)$\times k \times f$ = 160 + 880 = 1040

Charting Rules with Kernel Smoothing

The Charting rules share common parameters, so we first describe them here.

k: fixed band multiplicative value;

f: fixed holding days;

r: stop loss rate;

d: parameter for fixed liquidation price;

y: multiple of standard deviation of daily exchange-rate changes used to liquidate the position;

b: multiple of optimal bandwidth of kernel regression;

$k = 0, 0.005, 0.01$; (3 values)

$f = 1, 5, 10, 25$ (4 values)

$r = 0.005, 0.0075, 0.01$; (3 values)

$d = 0.25, 0.5, 0.75$; (3 values)

$y = 1, 1.25, 1.5, 1.75, 2.00, 2.5, 3.00, 3.50, 4.00, 4.50$ (10 values)

$b = 0.3, 1, 4$ (3 values)

Head-and-Shoulders (HS)

x: differential rate of shoulders or troughs

$x = 0.015, 0.03, 0.05$; (3 values)

Number of rules in HS class = $(x \times k \times r \times d \times y + x \times k \times f) \times (1 + b)$= $(810 + 36) \times 4 = 3384$

Triangle (TA)

Number of rules in TA class: = $(k \times r \times d \times y + k \times f) \times (1 + b)$= $(270 + 12) \times 4 = 1128$

Rectangle (RA)

Number of rules in RA class = $(x \times k \times r \times d \times y + x \times k \times f) \times (1 + b)$= $(810 + 36) \times 4 = 3384$

Double Tops and Bottoms (DTB)

x: differential rate of shoulders or troughs

n: least day differential between two tops/bottoms;

$x = 0.015, 0.03, 0.05$; (3 values)

$n = 22$ (1 value);

Number of rules in DTB class: = $(x \times k \times r \times d \times y \times n + x \times k \times f \times n) \times (1 + b) = (810 + 36) \times 4$= 3384

Broadening Tops and Bottoms (BTB)

Number of rules in BTB class: = $(k \times r \times d \times y + k \times f) \times (1 + b)$= $(270 + 12) \times 4 = 1128$

Complex Rules

Learning strategies (LS)

m: memory span;

r: review span;

$m = 2, 5, 10, 20, 40, 60, 125, 250$ (8 values);

$r = 1, 5, 10, 20, 40, 60, 125, 250$ (8 values);

Noting that $r \leq m$, there are 36 (m,r) combinations.

We have 2 performance measures: the sum of m daily returns, and the sharpe ratio.

In addition, including class with kernel smoothed series and learning on all non-complex rules, there are $5 + 5 \times 4 + 1 = 26$ classes of trading rules.

Number of rules in LS class = $36 \times 2 \times 26 = 1872$

Voting Strategies (VS)

Number of rules in VS class = 26

Fraction Position Strategies (FPS)

Number of rules in FPS class = 25 + 1 = 26

Voting by learning strategies (VLS)

m: memory span;

r: review span;

n: number of top trading rules chosed within a class;

$m = 1, 2, 5, 10, 20, 40, 60, 125, 250$ (9 values);

$r = 1, 5, 10, 20, 40, 60, 125, 250$ (8 values);

$n = 2, 3, 5, 10, 50$ (5 values)

Number of rules in VLS class: $= 37 \times 26 \times 5 = 4810$

Total number of trading rules = 497 + 2049 + 1220 + 2040 + 1040 + 3384 + 1128 + 3384 + 3384 + 1128 + 1872 + 26 + 26 + 4810 = 25988

Appendix C: Tests without Data Snooping Bias

In this section, we discuss the two stepwise bootstrap tests (StepM and SSPA) mentioned in the previous section. However, we start with the Reality Check and SPA test, since StepM is a stepwise version of the former and the SSPA test is a stepwise version of the latter. Our notations are similar to Hansen (2005).

Reality check

White (2000) tests the null hypothesis that the benchmark is not inferior to any of the m alternative trading rules:

$$\mathbf{H_0} : \mu \leq \mathbf{0}, \tag{.1}$$

where $\mu = E(\mathbf{d_t})$ and is a $m \times 1$ vector ($\mu \in \mathbb{R}^m$), while $\mathbf{d_t} = (d_{1,t}, \cdots, d_{m,t})$ is the $m \times 1$ vector of relative performance measures. Look at the vector of relative performance measure gives consideration to the full set of models underlying the vector that led to the best performing trading rule. Rejecting equation .1 implies that at least one trading rule beats the benchmark. White proceeds to construct the reality check from the test statistics,

$$T_n \equiv max(n^{1/2}\overline{d}_1, \cdots, n^{1/2}\overline{d}_m), \tag{.2}$$

$\overline{d}_m$ is the average performance of model m across n observations.

To calculate the p-value for the null hypothesis, White (2000) recommends the stationary bootstrap method of Politis and Romano (1994) to first obtain the empirical distribution of T_n^*:

$$T_n^*(b) = max[n^{1/2}(\overline{d}_1(b) - \overline{d}_1), \cdots, (n^{1/2}\overline{d}_m(b) - \overline{d}_m)], \qquad b = 1, \cdots, B \tag{.3}$$

The p-value is obtained by comparing T_n with the quantiles of the empirical distribution of $T_n^*(b)$. If the p-value is smaller than a given significance level, the null hypothesis is rejected.

SPA test

The null hypothesis of SPA test [Hansen (2005)] is the same as in White's Reality Check. Unlike the White's Reality Check, the SPA test uses the studentized test statistic, which will typically improve the power. Hansen (2005) provides a concrete example for highlighting the advantage of studentizing the individual statistics, since it avoids a comparison of the performance measured in different "units of standard deviation". Furthermore, the SPA test invokes a sample-dependent distribution under the null hypothesis, which can discard the poor models asymptotically. Therefore, the new test is more powerful and less sensitive to the inclusion of poor and irrelevant alternatives. The test statistic for SPA is given as follows:

$$T_n^{SPA} = max[max_{k=1,\cdots,m} \frac{n^{1/2}\overline{d}_k}{\hat{\omega}_k}, 0], \tag{.4}$$

where $\hat{\omega}_k^2$ is some consistent estimator of $\omega_k^2 = var(n^{1/2}\overline{d}_k)$. Then an re-centered estimator $\hat{\mu}^c$ for μ is chosen such that it conforms with the null hypothesis based on $N_m(\hat{\mu}^c, \hat{\Omega})$. The particular $\hat{\mu}^c$ suggested by Hansen (2005) is $\hat{\mu}_k^c = \overline{d}_k \mathbb{I}_{\{n^{1/2}\overline{d}_k/\hat{\omega}_k \leq -\sqrt{2loglogn}\}}$ for $k = 1, \cdots, m$, where $\mathbb{I}$ is the indicator function. It can be shown that for a poor trading rule $\mu_k < 0$, it has little effect on the distribution, and for a sufficiently large n, it will be discarded eventually. The improvement of the power of the SPA test over the Reality Check is further confirmed by the simulation experiment conducted in Hansen (2005). The p-value of T_n^{SPA} is calculated by bootstrapping its empirical distribution and then comparing the T_n^{SPA} with the quantiles of the empirical distribution of $T_n^{SPA*}(b, n)$. That is:

$$\hat{p}^{SPA} \equiv \sum_{b=1}^{B} \frac{\mathbb{I}_{\{T_n^{SPA*}(b,n) > T_n^{SPA}\}}}{B} \tag{.5}$$

Furthermore, Hansen (2005) defines another two p-values, which are not consistent, but can serve as the upper and lower bound for the consistent p-value. It imposes the null by recentering the bootstrap variables at $\hat{\mu}^l$ or $\hat{\mu}^u$, instead of $\hat{\mu}^c$. That is:

$$Z_{k,b,t}^* \equiv d_{k,b,t}^* - g_i(\overline{d}_k), \tag{.6}$$

where $i = l, c, u$, $g_l(x) = max(0, x)$, $g_u(x) = x$, and $g_c(x) = x \times \mathbb{I}_{\{x \geq -\sqrt{(\hat{\omega}_k^2/n)2loglogn}\}}$. One can show that $E(Z_{k,b,t}^*) = \hat{\mu}^i$ for $i = l, c, u$.

StepM test

Both the Reality Check and the SPA test seek to answer whether the best trading strategy beats the benchmark. As discussed in section 1, it is often more interesting to identify all outperforming trading rules, or to know whether a particular trading rule improves upon the benchmark. The Reality Check can be modified easily for identifying potential strategies that beat the benchmark, but Romano and Wolf (2005) show that this is only suboptimal, and only amounts to the first step of StepM test of Romano and Wolf (2005), which can detect more good strategies from the second step on. The StepM test is therefore more powerful than the Reality Check in detecting superior trading rules. The aim of the StepM test is to find as many profitable trading rules as possible. Their null hypothesis is considered as the individual hypothesis test:

$$H_0^k : \mu_k \leq 0, \tag{.7}$$

The individual decisions are made in a manner that asymptotically controls for the familywise error rate (FWE) at the significance level α. The FWE is defined as the probability of incorrectly identifying at least one strategy as superior. The joint confidence region is constructed to have a nominal joint coverage probability of $1 - \alpha$ in each stage, which is fulfilled by choosing a parameter c_i at the stage i. In the first stage it assumes the form of

$$[d_1 - c_1, \infty) \times \cdots \times [d_m - c_1, \infty) \tag{.8}$$

If $0 \notin [d_k - c_1, \infty)$, then the k^{th} trading rule is detected as profitable. This first step can detect more than one profitable rule. For the second step, all the profitable rules found in the first step are dropped from the universe and only remaining rules are used to form the new universe. Forming a similar confidence interval as in (.8), though replacing c_1 by c_2, one can detect the profitable trading rules, if any, from the remaining rules by checking whether or not the individual confidence interval $[d_k - c_2, \infty)$ contains zero. If no profitable rules are found, one should stop; otherwise, one continues this way until no profitable rules can be detected. Since the individual confidence interval in the joint confidence interval typically shrinks with the increasing number of testing steps, more profitable trading rules can be detected than when only relying on the first step.

Romano and Wolf (2005) also propose the use of studentization to improve the power and level properties of the StepM test.

Stepwise SPA test

The Stepwise SPA test of Hsu, Hsu and Kuan (2010) combines the SPA test and the StepM test to improve the power. It uses the studentized test statistic as in SPA test, and imposes a null to center the bootstrap variables around $\hat{\mu}^c$ as in the SPA test, such that poor and irrelevant trading rules will be discarded asymptotically. They provide formal proof and simulation results to demonstrate that the SSPA test is more powerful than the Reality Check, SPA test, and StepM test.

Acknowledgements

We are grateful to Waldemar Rotfuß, Stefan Ruenzi, Andreas Schrimpf, Ernst Maug, Min Qi, Michael Shill, Yangru Wu, and especially Graham Elliott (the editor) and an anonymous referee for helpful discussions and comments. We are also thankful to seminar participants at Goethe University Frankfurt, University of Mannheim, Centre for European Economic Research (ZEW), Forecasting Financial Markets conference 2008, International Workshop on computational and Financial Econometrics 2008, European Financial Management Association annual conference 2008, International Symposium on Forecasting 2009, for their helpful discussions. We thank Jonas Vogt and Hela Hellerich for their excellent research assistance. Errors and omissions remain the responsibility of the authors.

References

Brock, W., Lakonishok, J., LeBaron, B., 1992. Simple technical trading rules and the stochastic properties of stock returns. Journal of Finance 47, 1731–64.

Burnside, C., Eichenbaum, M., Rebelo, S., 2007. The returns to currency speculation in emerging markets. American Economic Review Papers and Proceedings 97(2), 333–338.

Chang, P.H.K., Osler, C.L., 1999. Methodical madness: Technical analysis and the irrationality of exchange-rate forecasts. Economic Journal 109, 636–61.

De Zwart, G., Markwat, T., Swinkels, L., van Dijk, D., 2009. The economic value of fundamental and technical information in emerging currency markets. Journal of International Money and Finance 28, 581–604.

Hansen, P.R., 2005. A test for superior predictive ability. Journal of Business & Economic Statistics 23, 365–380.

Härdle, W., 1990. Applied Nonparametric Regression. Cambridge University Press, Cambridge, U.K.

Hsu, P., Hsu, Y., Kuan, C., 2010. Testing the predictive ability of technical analysis using a new stepwise test without data snooping bias. Journal of Empirical Finance 17, 471–484.

Hsu, P.H., Kuan, C.M., 2005. Reexamining the profitability of technical analysis with data snooping checks. Journal of Financial Econometrics 3, 606–628.

Inoue, A., Kilian, L., 2004. In-sample or out-of-sample tests of predictability: Which one should we use? Econometric Reviews 23, 371–402.

Jegadeesh, N., 2000. Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation: Discussion. The Journal of Finance 55(4), 1765–1770.

Jensen, M., 1967. Random walks: reality or myth – comment. Financial Analysts Journal 23, 77–85.

Jensen, M.C., Bennington, G.A., 1970. Random walks and technical theories: Some additional evidence. Journal of Finance 25(2), 469–482.

LeBaron, B., 1999. Technical trading rule profitability and foreign exchange intervention. Journal of International Economics 49, 125–143.

Ledoit, O., Wolf, M., 2008. Robust performance hypothesis testing with the sharpe ratio. Journal of Empirical Finance 15, 850–859.

Lee, C.I., Gleason, K.C., Mathur, I., 2001. Trading rule profits in latin american currency spot rates. International Review of Financial Analysis 10, 135–156.

Lee, C.I., Mathur, I., 1996. Trading rule profits in european currency spot cross-rates. Journal of Banking & Finance 20, 949–962.

Lo, A.W., 2002. The statistics of sharpe ratios. Financial Analysts Journal , 36–52.

Lo, A.W., MacKinlay, A.C., 1990. Data-snooping biases in tests of financial asset pricing models. Review of Financial Studies 3, 431–67.

Lo, A.W., Mamaysky, H., Wang, J., 2000. Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. Journal of Finance 55, 1705–1765.

Martin, A.D., 2001. Technical trading rules in the spot foreign exchange markets of developing countries. Journal of Multinational Financial Management 11, 59–68.

Neely, C.J., Weller, P.A., 2003. Intraday technical trading in the foreign exchange market. Journal of International Money and Finance 22, 223–237.

Neely, C.J., Weller, P.A., Ulrich, J.M., 2009. The adaptive markets hypothesis: evidence from the foreign exchange market. Journal of Financial and Quantitative Analysis 44, 467–488.

Pojarliev, M., 2005. Performance of currency trading strategies in developed and emerging markets: Some striking differences. Financial Markets and Portfolio Management 19, 297–311.

Politis, D.N., Romano, J.P., 1994. The stationary bootstrap. Journal of the American Statistical Association 89, 1303–1313.

Qi, M., Wu, Y., 2006. Technical trading-rule profitability, data snooping, and reality check: Evidence from the foreign exchange market. Journal of Money, Credit and Banking 30, 2135–2158.

Romano, J.P., Shaikh, A.M., Wolf, M., 2008. Formalized data snooping based on generalized error rates. Econometric Theory 24, 404.

Romano, J.P., Wolf, M., 2005. Stepwise multiple testing as formalized data snooping. Econometrica 73, 1237–1282.

Savin, G., Weller, P., Zvingelis, J., 2007. The predictive power of "head-and-shoulders" price patterns in the u.s. stock market. Journal of Financial Econometrics 5, 243 – 265.

Sullivan, R., Timmermann, A., White, H., 1999. Data-snooping, technical trading rule performance, and the bootstrap. Journal of Finance 54, 1647–1691.

Sweeney, R.J., 1986. Beating the foreign exchange market. Journal of Finance 41, 163–82.

Taylor, M.P., Allen, H., 1992. The use of technical analysis in the foreign exchange market. Journal of International Money and Finance 11, 304–314.

Timmermann, A., 2006. Handbook of Economic Forecasting. Amsterdam: North-Holland. chapter "Forecast Combinations".

White, H., 2000. A reality check for data snooping. Econometrica 68, 1097–1126.