MIN QI

YANGRU WU

Technical Trading-Rule Profitability, Data Snooping, and Reality Check: Evidence from the Foreign Exchange Market

We report evidence on the profitability and statistical significance among 2,127 technical trading rules. The best rules are found to be significantly profitable based on standard tests. We then employ White's (2000) Reality Check to evaluate these rules and find that data-snooping biases do not change the basic conclusions for the full sample. A sub-sample analysis indicates that the data-snooping problem is more serious in the second half of the sample. Profitability becomes much weaker in the more recent period, suggesting that the foreign exchange market becomes more efficient over time. Evidence from cross exchange rates confirms the basic findings.

SINCE THE BREAKDOWN of the Bretton Woods System in early 1973, many major currencies have floated against the U.S. dollar. Since then, a

large body of research has been devoted to studying the time-series properties of exchange rates and to testing the efficiency of the foreign exchange (FX) market. One strand of such literature examines the profitability of technical trading rules. The central idea underlying this research is that if the FX market is efficient, one should not be able to use publicly available information to predict future changes in exchange rates and hence to make an abnormal (risk-adjusted) profit.[1] In particular, popular technical trading strategies, which use current and past price and volume data and are guided by mechanical algorithms, should not be able to beat the market. This suggests a simple and robust test for the weak form efficient market hypothesis. Cornell and Dietrich (1978) are among the first to document profitability of filter and moving average rules using the post-Bretton Wood data. Sweeney (1986) finds that filter rules are profitable for the dollar–deutsche mark exchange rate and the profitability cannot be explained by risk. Using futures prices for five currencies, Levich and Thomas (1993) demonstrate based on bootstrap re-sampling that some filter and moving average rules may be significantly profitable even after adjusting for risk.[2]

The significant profitability documented in these studies is often interpreted as evidence against market efficiency.[3] However, since a researcher can examine many trading rules for a given data set, there is a high probability that one such rule will yield superior performance for that data set due to pure luck, even if it does not provide any useful economic signal for a new data set. In practice multiple technical trading rules are often explored, but only the best one is reported and statistical inference about the best ignores the full set of alternatives. To put it in a different way, since the best trading rule is only known *ex post* and at the beginning of the sample period one would not know which rule will do well, finding that the best trading rule outperforms a benchmark based on a usual statistical test (say the standard *t*-statistic) does not necessarily imply that the rule is significantly profitable. This is the so-called data-snooping or data-mining problem. While the significance of this problem is widely recognized in empirical studies, data snooping is prevailing and there had been a lack of practical procedures to formally deal with it. White (2000) develops a novel procedure, called Reality Check, to measure and correct for data-snooping biases. The idea is to generate the empirical distribution from the full set of models (or trading strategies in this context) that lead to the best-performing strategy and to draw inference from this distribution for certain performance measures. This is the first rigorously founded and easily applicable method to test whether the best rule encountered during a specification search has predictive ability

1. A complementary literature studies the forecasting accuracy of exchange rates using either fundamental analysis or time-series techniques. See, for example, Baillie and Bollerslev (1994), Mark (1995), and Qi and Wu (2003). In contrast to the literature on technical analysis with high-frequency data, these researchers employ lower-frequency data (monthly or quarterly) and find that exchange rates are in general unforecastable at the 1- to 12-month horizons.

2. Studies in this area also include Kho (1996), LeBaron (1999), Lee and Mathur (1996), Neely, Weller, and Dittmar (1997), Neely and Weller (2001), Neely (2002), and Sapp (2004), among many others.

3. We use the term "market efficiency" in the sense that one cannot make an abnormal (or risk-adjusted) profit using publicly available information.

over a benchmark. This methodology allows data mining to be undertaken with some degree of confidence that one will not mistake results that might be generated by pure chance for genuinely good ones.[4]

This paper studies the profitability of technical trading rules in the FX market and examines the biases due to potential data snooping. Our paper is the first study to formally investigate data-snooping biases of technical rules in the FX market using White's (2000) Reality Check methodology. Previous studies have employed various techniques to try to control for or mitigate potential data-snooping biases. For example, Brock, Lakonishok, and LeBaron (1992) argue that the problem of data snooping can be mitigated by deliberate choice of a simple class of rules that has been in common use for a long period of time. But there still remains a significant amount of latitude in choosing the exact form of the rule. Neely, Weller, and Dittmar (1997) employ genetic programming to identify an optimal trading rule and the performance of the optimal rule obtained from the estimation sample is then examined over a validation sample. However, there still exists an issue of how to select a fitness criterion and how to choose the training, selection, and validation periods, etc., in addition to the complexity of implementing the genetic program in a given application. The Reality Check, on the other hand, generates a useful estimate of the true p-value for the null hypothesis that the best trading rule does not beat a benchmark strategy through simple bootstrap simulation. Comparing the true p-value to the nominal p-value provides an easy yet rigorous way to quantify the data-snooping bias in a given application. The Reality Check thus offers important insights and is a useful complement to the genetic programming and other methods in fending off data-snooping biases. Results in this paper should shed more light on the debate on whether technical analysis is useful in the FX market and whether the FX market is efficient.

We employ four types of popular trading strategies: filter, moving average, trading range break, and channel breakout. These rules have been widely used by FX market practitioners and have been subject to extensive research.[5] For each type, we consider various parameterizations. Overall, we investigate the universe of 2,127 parameterizations of trading rules, which are by far the largest in number and the most comprehensive in scope in FX studies. We employ daily U.S. dollar prices of seven currencies for the period between April 2, 1973, and December 31, 1998. Our evaluation is based on three performance measures: mean excess return, Sharpe ratio, and Jensen alpha. The Sharpe ratio criterion accounts for total (stand-alone) risk, while the Jensen alpha criterion considers the market risk-adjusted excess return.

---

4. The Reality Check methodology has been employed by Sullivan, Timmermann, and White (1999, 2001, 2003) to study equity returns, among others. In particular, using 7,846 rules and applying the Reality Check to 100 years of daily data on S&P500 index returns, Sullivan, Timmermann, and White (1999) find significant profitability of the best trading rule in equity returns but the profitability becomes insignificant in an out-of-sample analysis. Since the FX market is substantially different from the equity market, it is worthy of a careful investigation using the Reality Check methodology.

5. See, for example, LeBaron (1999), Levich and Thomas (1993), Neely (1997), Neely, Weller, and Dittmar (1997), Osler (2003), Rosenberg (1996), Schulmeister (1987), Surajaras and Sweeney (1992), Sweeney (1986), and the Group of Thirty (1985).

We find that for the full sample period certain rules appear to outperform even after adjustments are made for data-snooping biases. In particular, using White's bias-adjusted p-values, the best-performing strategies produce significant mean excess returns, Sharpe ratios, and Jensen alphas at the 1% level for seven U.S. dollar exchange rates. However, a sub-sample analysis indicates that the profitability of trading rules becomes mostly insignificant after accounting for data-snooping biases in the second half of the sample. An out-of-sample analysis further shows that the performance of the best rules identified in sample is substantially weaker out of sample. Results obtained by using the Japanese yen and the deutsche mark as vehicle currencies are qualitatively similar to those for the dollar exchange rates. Our overall results suggest that there is a potential change in the dynamics of exchange rate series and that the FX market becomes increasingly more efficient over time.

The remainder of the paper is organized as follows. Section 1 describes the universe of technical trading rules that we explore in this study. Section 2 explains the bootstrap methodology that we adopt to investigate data-snooping biases in trading-rule profitability in the FX market. The main empirical results are reported in Section 3. Section 4 carries out further checks for the robustness of results. Concluding remarks are contained in the final section.

1. TECHNICAL TRADING RULES

Technical trading rules have been extremely popular tools used by traders in financial markets for many decades and the study on their performance has been a subject of extensive research. For applications in the FX market, in a survey of major FX dealers based in London in 1990, Taylor and Allen (1992) report that at least 90% of the respondents place some weight on technical analysis when forecasting currency prices, particularly for short horizons. In order to obtain meaningful estimate of data-snooping biases, we must provide a comprehensive coverage of the trading rules in use. In this study, we consider four popular types of trading rules: filter, moving average, trading range break (support and resistance), and channel breakout. These rules are among the most popular and are drawn from existing academic and practitioner's literature. Intuitively, choosing too few rules can cause biases in statistical inference due to data mining. On the other hand, Hansen (2003) points out that loading too many irrelevant rules can reduce the test power. We therefore find a balance and select a fairly large variety of reasonable parameters that lie in the ranges used in the literature. The Appendix cites the relevant references and explains the rationale to choose the rules and specific parameter values in our universe for the experiments.

Each type of rule is explained below. We assume that the speculator initially has the necessary initial wealth to be used as margins to trade FX contracts. The speculator can take a long or short position in the FX.[6]

---

6. Short sale restrictions can have a significant impact on the performance of investment strategies, as analyzed by Alexander (2000) for common stocks. For FX, the problem is less important and we assume that the transaction cost is the same whether the investor is taking a long or a short position in FX.

1.1 Filter Rules

A filter rule strategy is specified as follows. If the daily closing price (in U.S. dollars) of a foreign currency goes up by $x\%$ or more from its most recent low, then the speculator borrows the dollar and uses the proceeds to buy the foreign currency. When the closing price of the foreign currency drops by at least $y\%$ from a subsequent high, the speculator short sells the foreign currency and uses the proceeds to buy the dollar. We define the subsequent high as the highest price over the e most recent days and the subsequent low as the lowest price over the e most recent days. We also consider the case where a given long or short position is held for c days during which time all other signals are ignored.

1.2 Moving Average Rules

The moving average of a currency price for a given day is computed as the simple average of prices over the previous n days, including the current day. Under a 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 dollar to buy the foreign currency. Similarly, when the short moving average is below the long moving average by $b\%$, the speculator short sells the FX to buy the dollar. In addition to this fixed percentage band filter, we also implement the moving average rules with a time delay filter, which requires that the long or short signals remain valid for d days before he takes any action. As in the filter rule case, we also consider the case where a given long or short position is held for c days during which time all other signals are ignored.

1.3 Trading Range Break (or Support and Resistance) Rules

Under a trading range break rule, when the price of a foreign currency exceeds the maximum price (resistance level) over the previous n days by $b\%$, the speculator borrows the dollar to buy the foreign currency. When the price goes below the minimum price over the previous n days by $b\%$, the speculator sells short the FX to buy the dollar. We also consider an alternative definition for the resistance level, i.e., the local maximum (minimum), which is the most recent closing price higher (lower) than the e previous closing prices. As with the moving average rules, we implement the rules with a time delay filter, d, and as well we consider the case where a given long or short position is held for c days during which time all other signals are ignored.

1.4 Channel Breakout Rules

A channel is defined to be one that 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, the speculator borrows the dollar to buy FX when the closing price of the foreign currency goes above the channel by $b\%$ and sells short the FX to buy the dollar when the closing price goes below the channel by

$b\%$. Once again, we consider holding a given long or short position for c days during which all other signals are ignored.

2. THE BOOTSTRAP EXPERIMENT TO STUDY DATA-SNOOPING BIASES

White's (2000) Reality Check procedure tests the null hypothesis that the model selected in a specification search has no predictive superiority over a given benchmark model. It allows aggressive model searching to be undertaken with confidence that one will not mistake results that could have been generated by chance for genuinely good results. The procedure was developed based on Diebold and Mariano's (1995) and West's (1996) methods to test whether a given model has predictive superiority over a benchmark model after taking into consideration the effects of data snooping. The test evaluates the distribution of a performance measure, accounting for the full set of models that lead to the best-performing model, and is based on the $L \times 1$ performance statistic

$$\bar{R} = N^{-1} \sum_{t=P}^{T} R_t \tag{1}$$

where L is the number of models, N is the number of prediction periods, $R_t = f(X_{t-1}, \hat{\theta}_{t-1})$ is a performance measure (such as the difference in returns between a trading rule and a benchmark), $\hat{\theta}_{t-1}$ is a vector of estimated parameters, and X is a vector of variables.

Our full sample of daily exchange rates starts on April 2, 1973, and ends on December 31, 1998, with a total of 6,463 observations for each currency. The first trading signal is generated for the 256th observation for all specifications because some rules require 255 days of previous data in order to provide a trading signal. Therefore, P is 256, T is 6,463, and N is 6,208. We investigate 69 filter rules, 858 moving average rules, 560 trading range break rules, and 640 channel breakout rules. Thus, the total number of trading rules in our universe (L) is 2,127.

We consider three performance measures in this study. The first one is the mean excess return, which is most popularly used in the literature. The second one is the Sharpe ratio which measures the average excess return per unit total risk. As for the third, we follow Jensen (1968) by using the alpha parameter from the CAPM, which is the excess return after adjusting for systematic market risk, as a performance measure. We call this criterion the Jensen alpha criterion.

The excess return rate of the kth trading rule at time t is computed as[7]

$$R_{k,t} = (s_t - s_{t-1} + r_t^* - r_t)I_{k,t-1} - abs(I_{k,t-1} - I_{k,t-2})g,$$
$$k = 1,...,L; t = 256,...,T \tag{2}$$

---

7. Sweeney (1986) shows that the excess returns from filter rules are primarily dependent on exchange rate variations and not much on interest-rate differentials. Because of that, most authors ignore interest-rate differentials in calculating returns, see, for example, Sweeney (1986), Surajaras and Sweeney (1992), Lee and Mathur (1996), and Szakmary and Mathur (1997). We conducted our analysis without considering interest-rate differentials and obtained similar results.

where s_t is the natural logarithm of exchange rate (domestic currency price of one unit foreign currency); r_t and r_t^* are, respectively, domestic and foreign interest rates from time t-1 to t; $I_{k,t}$ represents the trading signal generated by the kth trading rule using information available at time t and this dummy variable may take three values: 1 which represents a long position, 0 which represents a neutral position, and -1 which represents a short position; and g is a one-way transaction cost. Note that $R_{k,t}$ defined in Equation (2) is the excess return (zero net investment) for an investor who takes a speculative position in the FX market. A natural benchmark is one where the investor does not take a position in the FX market and therefore earns a zero return.

The Sharpe ratio for the kth trading rule is accordingly defined as:[8]

$$G_k = \frac{E(R_{k,t})}{\sqrt{E(R_{k,t}^2) - [E(R_{k,t})]^2}} \tag{3}$$

Similarly, the Jensen alpha for the kth rule, α_k, is computed after estimating the standard CAPM:

$$R_{k,t} = \alpha_k + \beta_k(r_{m,t} - r_{f,t}) + \varepsilon_{k,t} \tag{4}$$

where $r_{m,t}$ is the market rate of return, $r_{f,t}$ is the risk-free rate, and $\varepsilon_{k,t}$ is the error term that is assumed to satisfy classical assumptions for OLS regression. We use returns on the value-weighted U.S. stock market index as a proxy for the market return, and the three-month U.S. Treasury bill rate as the risk-free rate. The market index is obtained from the CRSP tape, while the Treasury bill rate data is obtained from the Federal Reserve Board's website. For results reported in succeeding sections, the mean return, Sharpe ratio, and Jensen alpha are all annualized.

Based on our first performance measure, to assess whether there exists a superior rule, the null hypothesis to be tested is that the best rule is no better than the benchmark, i.e.,

$$H_0^{\text{return}}: \max_{k=1,\ldots,L} \{E(R_k)\} \leq 0 \tag{5}$$

where the expectation is evaluated with the average $\bar{R}_k = N^{-1}\sum_{t=P}^{T} R_{k,t}$. Rejection of the null will lead to the conclusion that the best trading rule achieves performance superior to the benchmark.

Similarly, in the case of the Sharpe ratio criterion, the null hypothesis becomes

$$H_0^{\text{Sharpe}}: \max_{k=1,\ldots,L} \{G_k\} \leq 0 \tag{6}$$

---

8. We do not subtract the risk-free rate in computing the Sharpe ratio because $R_{k,t}$ is already in excess return terms.

and in the case of the Jensen alpha criterion, the null hypothesis is:

$$H_0^{\text{Jensen}}: \max_{k=1,...,L} \{\alpha_k\} \leq 0 \tag{7}$$

Following White (2000), we test the null hypothesis H_0^{return} by applying the stationary bootstrap method of Politis and Romano (1994) to the observed values of $R_{k,t}$ as follows. Step 1, for each trading rule k, we resample the realized excess return series $R_{k,t}$, one block of observations at a time with replacement, and denote the resulting series by $R_{k,t}^*$. This process is repeated B times. Step 2, for each replication i, we compute the sample average of the bootstrapped returns, denoted by $\bar{R}_{k,i}^* \equiv N^{-1} \sum_{t=P}^{T} R_{k,t}^*$, $i = 1,...,B$. Step 3, we construct the following statistics:

$$\bar{V} = \max_{k=1,...L} \{\sqrt{N}(\bar{R}_k)\}, \tag{7}$$

$$\bar{V}_i^* = \max_{k=1,...,L} \{\sqrt{N}(\bar{R}_{k,i}^* - \bar{R}_k)\}, i = 1,...,B \tag{8}$$

Step 4, White's Reality Check p-value is obtained by comparing $\bar{V}$ to the quantiles of $\bar{V}_i^*$. By employing the maximum value over all L models, the Reality Check p-value incorporates the effects of data snooping from the search over the L trading rules. To keep the amount of computation manageable, we choose $B = 500$ with the smoothing parameter equal to 0.5.[9] Similarly, the Politis and Romano (1994) stationary bootstrap procedure can be applied to compute White's Reality Check p-value under the Sharpe ratio and the Jensen alpha criteria.

3. EMPIRICAL EVIDENCE FROM SEVEN U.S. DOLLAR EXCHANGE RATES

3.1 Data and Summary Statistics

We collect daily data from the Federal Reserve Board's website for the following currency prices relative to the U.S. dollar: the Canadian dollar (CAN), the deutsche mark (GER), the French franc (FRA), the Italian lira (ITA), the Japanese yen (JAP), the Swiss franc (SWI), and the pound sterling (UK). For interest rates, we obtain the overnight Euro-currency interest rates from *Datastream International*. Table 1 reports some summary statistics of the daily returns (changes in the natural logarithm of exchange rates). The mean return rates show that *on average* the dollar appreciates against the Canadian dollar, the French franc, the Italian lira, and the British pound, while depreciates against the other three currencies over the sample period. The mean *absolute* daily changes in exchange rates are quite large, ranging from 0.18% to 0.55%. With the exception of the Canadian dollar, all exchange rates display substantial daily volatility with the standard deviation of daily change in the range

---

9. Our overall results are not sensitive to the choice of the smoothing parameter which determines the block length. Results from a sensitivity analysis with different values of the smoothing parameter were reported in an earlier version of the paper and are available upon request.

TABLE 1

DESCRIPTIVE STATISTICS FOR DAILY CHANGES IN THE LOGARITHM OF EXCHANGE RATES

$s_t - s_{t-1}$	CAN	FRA	GER	ITA	JAP	SWI	UK
Mean*100	−0.0067	−0.0031	0.0083	−0.0161	0.0132	0.0134	−0.0062
Mean Absolute*100	0.1841	0.4505	0.4757	0.4215	0.4388	0.5487	0.4323
Min	−0.0190	−0.0587	−0.0587	−0.0669	−0.0626	−0.0583	−0.0384
Max	0.0186	0.0416	0.0414	0.0404	0.0563	0.0441	0.0459
Standard Deviation	0.0026	0.0065	0.0066	0.0062	0.0065	0.0076	0.0062
Sharpe Ratio	−0.4091	−0.0757	0.1996	−0.4122	0.3224	0.2799	−0.1588
Skewness	−0.1843	−0.1911	−0.0680	−0.5281	0.4251	−0.0184	−0.1421
Kurtosis	4.3654	5.5238	3.5506	8.0005	6.3622	3.4030	3.9643
$\rho(1)$	0.0552	0.0323	0.0333	0.0374	0.0443	0.0353	0.0630
$\rho(2)$	−0.0069	−0.0067	−0.0052	0.0022	0.0122	−0.0089	0.0014
$\rho(3)$	0.0015	0.0120	0.0137	0.0110	0.0061	0.0032	−0.0047
$\rho(4)$	−0.0028	0.0085	0.0002	−0.0133	0.0082	−0.0074	0.0019
$\rho(5)$	0.0128	0.0196	0.0193	0.0134	0.0176	0.0052	0.0331
$\rho(6)$	−0.0055	−0.0206	−0.0004	0.0008	−0.0054	0.0084	−0.0089

NOTES: This table reports summary statistics for daily changes in the logarithm of exchange rates for the period between April 2, 1973, and December 31, 1998, with 6,463 observations for each currency. Exchange rate is defined as the U.S. dollar price of one unit foreign currency; s_t is the logarithm of exchange rate; $\rho(k)$ is the kth order serial correlation of $(s_t - s_{t-1})$. The Sharpe ratios are annualized.

of 0.62%–0.76%. The maximum daily appreciation of the dollar relative to these six currencies lies between 3.8% (the British pound) to 6.69% (the Italian lira), while the maximum depreciation of the dollar ranges between 4.04% (the deutsche mark) and 5.63% (the Japanese yen). The corresponding numbers are much smaller for the Canadian dollar. On the other hand, the Canadian dollar has the smallest Sharpe ratio (average return per unit risk) among all seven currencies. All daily return distributions have excess kurtosis relative to the normal distribution (Equation 3) and all distributions appear to skew to the left with the exception of the Japanese yen. The standard Jarque-Bera test (not reported) overwhelmingly rejects the null hypothesis of normality for all currencies. The first-order serial correlation ranges between 0.0331 and 0.0628. Serial correlation at higher orders is far smaller for all currencies.

3.2 Full Sample Results

Table 2 Panel A reports the performance of the best trading rules under the three alternative criteria in the absence of transaction cost. Interestingly, all three criteria select the same best rule for each currency. Two currencies (the Japanese yen and the Swiss franc) choose channel breakout as the best-performing rule, while the remaining five currencies select single moving average as the best-performing rule. The moving averages chosen are relatively short (between 15 and 25 days). Both channel breakout rules use the shortest period (5 days) to form a channel. The mean excess returns are all positive but vary substantially across currencies, from a low of 4.02% per annum for the Canadian dollar to a high of 12.81% per annum for the Japanese yen. The nominal p-values are all close to zero. When the White's Reality Check procedure is applied to correct for data-snooping biases, the p-values are greater than the nominal p-values as expected. Nevertheless, the evidence

TABLE 2

PERFORMANCE OF THE BEST TRADING RULE: U.S. DOLLAR AS VEHICLE CURRENCY

Currency	Best Trading Rule	Number of Trades	Mean Return Criterion			Sharpe Ratio Criterion			Jensen α Criterion		
			Mean Return	Nominal p-value	White's p-value	Sharpe Ratio	Nominal p-value	White's p-value	α	Nominal p-value	White's p-value
Panel A. Without Transaction Cost											
CAN	MA($n = 20, b = 0$)	1178	4.02	0.000	0.000	0.97	0.000	0.000	4.10	0.000	0.000
FRA	MA($n = 20, b = 0$)	1016	10.67	0.000	0.000	1.06	0.000	0.000	11.13	0.000	0.000
GER	MA($n = 25, b = 0.001$)	780	10.00	0.000	0.000	0.97	0.000	0.000	10.36	0.000	0.002
ITA	MA($n = 15, b = 0.0005$)	1144	9.69	0.000	0.000	0.98	0.000	0.000	10.12	0.000	0.000
JAP	CBO($n = 5, x = 0.01, c = 1$)	842	12.81	0.000	0.000	1.24	0.000	0.000	12.97	0.000	0.000
SWI	CBO($n = 5, x = 0.05, b = 0.001, c = 1$)	1192	10.26	0.000	0.006	0.86	0.000	0.006	10.70	0.000	0.006
UK	MA($n = 15, b = 0.001$)	1026	9.93	0.000	0.000	1.00	0.000	0.000	10.36	0.000	0.000
Panel B. With One-way Transaction Cost 0.025%											
CAN	TRB($n = 5, c = 10, d = 4$)	29	3.38	0.000	0.014	0.81	0.000	0.020	3.54	0.000	0.002
FRA	MA($n = 20, b = 0$)	1016	9.64	0.000	0.000	0.96	0.000	0.000	10.09	0.000	0.000
GER	MA($n = 25, b = 0.001$)	780	9.21	0.000	0.002	0.89	0.000	0.004	9.57	0.000	0.002
ITA	MA($n_1 = 2, n_2 = 20, b = 0.0005$)	652	8.85	0.000	0.004	0.90	0.000	0.002	9.25	0.000	0.000
JAP	CBO($n = 5, x = 0.01, c = 1$)	842	11.95	0.000	0.000	1.16	0.000	0.000	12.11	0.000	0.000
SWI	MA($n = 10, b = 0.01$)	560	9.54	0.000	0.020	0.80	0.000	0.018	9.87	0.000	0.016
UK	MA($n = 15, b = 0.001$)	1026	8.89	0.000	0.004	0.89	0.000	0.004	9.32	0.000	0.000

NOTES: This table reports the performance of the best trading rule from the universe of 2,127 trading rules for each currency under three alternative criteria. Panel A presents the results in the absence of transaction cost while Panel B displays results with a one-way transactions cost equal to 0.025%. The U.S. dollar is used as the vehicle currency. The three criteria select the same best trading rule for each currency. The "White's p-value" is computed by applying the Reality Check methodology to the universe of 2,127 trading rules and incorporates the effects of data-snooping bias. The "nominal p-value" is calculated by applying the Reality Check methodology to the best trading rule only. The mean returns and the Jensen α are annualized and in percentage terms. The Sharpe ratios are annualized.

of profitability is so strong that even after correcting for data snooping, the null that technical analysis does not have genuine merit can be rejected at the 1% level for all seven currencies.

It may be premature, however, to say that the best-performing rules identified above are actually profitable because we have not considered the effects of transaction costs. These effects can indeed be significant when the trading frequencies of the preferred rules are high. The total number of trades for each best rule shows that over the 25-year period with 6,208 trading days, there are trades for roughly between 13% and 19% of the days.[10] While the percentage transaction costs per trade for foreign currencies are in general much lower than those for stocks, frequent trading can accumulate the small cost per trade into a substantial figure.

Let x be the maximum transaction cost per one-way trade for a rule to be breaking even over the 25 trading years, we find that $x = 25*4.02/1178 = 0.085\%$ for the Canadian dollar. Similarly, we find $x = 0.26, 0.32, 0.21, 0.38, 0.22$, and 0.24% for the French franc, the deutsche mark, the Italian lira, the Japanese yen, the Swiss franc, and the British pound, respectively. With the exception of the Canadian dollar, these threshold transaction costs are indeed much larger than the actual transaction costs incurred in major FX markets. For example, Bessembinder (1994) documents that the *round-trip* transaction costs in the inter-bank market, as measured by bid-ask spreads, lie between 0.05% and 0.08% for the deutsche mark, the Japanese yen, the Swiss franc, and the British pound. The bid-ask spreads for currencies less heavily traded and with greater volatility can be larger (Shapiro, 1999, pp. 148–157).[11] Thus the one-way costs are in the range of 0.025%–0.04%, which are not in the same order of magnitude compared to the threshold transaction costs estimated above. Using a more conservative 0.04% one-way cost, we find that the *after-cost* profit for the Canadian dollar is 2.14% (4.02–1178*0.04/25). Similarly, the *after-cost* profits are 9.04, 8.75, 7.86, 11.46, 8.35, and 8.29%, respectively, for the French franc, the deutsche mark, the Italian lira, the Japanese yen, the Swiss franc, and the British pound. Except for the Canadian dollar, these *after-cost* profit figures are clearly economically significant.

Panel B of Table 2 reports results on the best trading rules and their performance with a transaction cost filter $q = 0.025\%$ imposed on Equation (2).[12] Compared to Panel A of Table 2, we find that this cost filter does have some effect on the overall performance of the trading rules. In particular, two currencies—the Canadian dollar and the Swiss franc—select the optimal rules that are of different types from the baseline case without a transaction cost filter. The number of trades under the optimal

---

10. On certain days the number of trades may be 2 when going from short to long or from long to short, thus the actual percentage of days with trades is somewhat lower.

11. Neely, Weller, and Dittmar (1997) also claim that 0.05% is a reasonable cost per round-trip trade. McCormick (1975) reports lower estimates of bid-ask spread, in the range of 0.02%–0.04%. Using surveys of several major markets, Cheung and Chinn (2001) estimate the bid-ask spreads to be around 0.05% for the Swiss franc and the British pound, and 0.03%–0.05% for the deutsche mark and the Japanese yen.

12. Results with $g = 0.05\%$ are similar and are omitted for brevity. They are available upon request.

rule is greatly reduced for these two currencies. Furthermore, after accounting for data snooping biases, the profitability of the optimal trading rule becomes less significant statistically (at the 5% level as measured by the White p-value). The other five currencies continue to choose moving average rules as the optimal, although the Italian lira selects a longer and double moving average rule. The profitability of the best trading rule continues to be statistically significant at the 1% level for these five currencies.

The statistical significance of the best trading rules as well as the nature of data-snooping biases under the Sharpe ratio criterion are quite similar to those under the mean return criterion. The nominal p-values are close to zero. The White p-values are all significant at the 1% level, implying that even after data-snooping biases are properly accounted for, technical trading rules yield superior Sharpe ratios to the benchmark of not participating in the market.

The results for the Jensen alpha criterion reported in Table 2 are obtained by using the value-weighted U.S. stock market index as a proxy for the market return. Results obtained using the S&P500 index as the market return are similar and omitted. For the best-selected rules, the pricing errors α, which are expressed in annualized percentage terms, are somewhat larger than the mean return without market risk adjustment. Furthermore, after accounting for data-snooping biases, the White p-values are all lower than 1%. These results show that the market risk factor is unlikely to explain the excess return.[13]

3.3 Sub-sample Analysis

While the above results seem to suggest that technical analysis is useful in forecasting the direction of exchange rate changes, it is important to note that the dollar exchange rates experienced long swings in the early 1980s and several crises in the 1990s. The exchange rates may well be subject to regime changes during the sample period. A methodology that works in one period may not work in another period. To check for the robustness of results, we conduct a sub-sample analysis. We divide the full sample into two sub-samples with roughly equal lengths. For each currency in each sub-sample, we select the best trading rule from the same universe of rules that is used for the full sample analysis and apply the Reality Check methodology to adjust for data-snooping biases.

Table 3 reports the performance of the best trading rules for the seven U.S. dollar exchange rates in the two sub-samples. As can be seen from Panel A, the results for the first sub-sample (1974–85) are basically consistent with the full sample results reported in Table 2. For each currency, the White p-value is near zero, implying that the null hypothesis can be rejected at the 1% level even after data-snooping biases are accounted for. Furthermore, the mean excess returns are all stronger than the corresponding ones for the full sample.

---

13. We have also used the Japanese stock market index, the German stock market index, and the world stock market index, respectively, as a proxy for the market return to compute the Jensen alpha. The statistical significance of the best trading rules selected based on these alternative Jensen alpha measures is similar. The results are not reported but are available upon request.

TABLE 3

PERFORMANCE OF THE BEST TRADING RULE IN SUB-PERIODS

Currency	Best Trading Rule	Number of Trades	Mean Return	Nominal p-value	White's p-value	Sharpe Ratio	Nominal p-value	White's p-value	Jensen α	Nominal p-value	White's p-value
Panel A. Sub-sample 1974–85											
CAN	TRB($n = 10$, $c = 1$)	270	5.72	0.000	0.000	1.57	0.000	0.000	5.83	0.000	0.000
FRA	MA($n = 2$, $d = 3$)	458	14.42	0.000	0.000	1.47	0.000	0.000	14.67	0.000	0.000
GER	MA($n = 10$, $b = 0.0005$)	684	14.67	0.000	0.000	1.48	0.000	0.000	14.92	0.000	0.000
ITA	MA($n = 5$, $c = 5$)	494	12.88	0.000	0.000	1.41	0.000	0.000	12.97	0.000	0.000
JAP	MA($n_1 = 2$, $n_2 = 10$, $b = 0.0005$)	514	15.43	0.000	0.000	1.71	0.000	0.000	15.39	0.000	0.000
SWI	CBO($n = 5$, $x = 0.05$, $b = 0.001$, $c = 1$)	498	16.66	0.000	0.000	1.39	0.000	0.002	16.86	0.000	0.000
UK	MA($n = 10$, $b = 0$)	800	15.02	0.000	0.000	1.53	0.000	0.000	15.56	0.000	0.000
Panel B. Sub-sample 1986–98											
CAN	TRB($n = 10$, $c = 5$, $d = 4$)	10	4.55	0.000	0.062	1.00	0.000	0.064	4.75	0.000	0.030
FRA	MA($n = 20$, $b = 0$)	562	8.75	0.002	0.242	0.85	0.002	0.256	9.37	0.000	0.146
GER	TRB($e = 3$, $b = 0.005$, $c = 25$)	14	10.31	0.000	0.094	0.96	0.000	0.098	10.58	0.000	0.070
ITA	MA($n = 25$, $b = 0.001$)	436	9.77	0.000	0.110	0.93	0.000	0.104	10.23	0.000	0.074
JAP	MA($n_1 = 10$, $n_2 = 50$, $b = 0.001$)	112	13.15	0.000	0.008	1.16	0.000	0.010	13.65	0.000	0.004
SWI	TRB($e = 2$, $c = 25$)	38	11.50	0.000	0.094	0.96	0.000	0.100	11.61	0.000	0.074
UK	CBO($n = 5$, $x = 0.005$, $c = 10$)	22	8.46	0.004	0.306	0.84	0.002	0.302	9.34	0.000	0.148

NOTES: This table reports the performance of the best trading rule from the universe of 2,127 trading rules for each currency under three alternative criteria for two sub-samples. The U.S. dollar is used as the vehicle currency. The "White's p-value" is computed by applying the Reality Check methodology to the universe of 2,127 trading rules and incorporates the effects of data-snooping bias. The "nominal p-value" is calculated by applying the Reality Check methodology to the best trading rule only. The mean returns and the Jensen α are annualized and in percentage terms. The Sharpe ratios are annualized.

TABLE 4

PERFORMANCE OF THE BEST TRADING RULE IN SUB-SAMPLE 1986–98: COMPARISON OF REALITY CHECK AND TEST FOR SUPERIOR PREDICTIVE ABILITY

Currency	Best Trading Rule	Number of Trades	Mean Return	Nominal p-value	White's p-value	SPA_l p-value	SPA_c p-value	SPA_u p-value
CAN	TRB($n = 10$, $c = 5$, $d = 4$)	10	4.55	0.000	0.062	0.040	0.056	0.062
FRA	MA($n = 20$, $b = 0$)	562	8.75	0.002	0.242	0.078	0.140	0.170
GER	TRB($e = 3$, $b = 0.005$, $c = 25$)	14	10.31	0.000	0.094	0.074	0.086	0.086
ITA	MA($n = 25$, $b = 0.001$)	436	9.77	0.000	0.110	0.076	0.090	0.092
JAP	MA($n_1 = 10$, $n_2 = 50$, $b = 0.001$)	112	13.15	0.000	0.008	0.008	0.008	0.008
SWI	TRB($e = 2$, $c = 25$)	38	11.50	0.000	0.094	0.074	0.094	0.094
UK	CBO($n = 5$, $x = 0.005$, $c = 10$)	22	8.46	0.004	0.306	0.238	0.300	0.304

NOTES: This table reports the performance of the best trading rule from the universe of 2,127 trading rules for each currency for the sub-sample 1986–98. The U.S. dollar is used as the vehicle currency. The "White's p-value" is computed by applying the Reality Check methodology to the universe of 2,127 trading rules and incorporates the effects of data-snooping bias. SPA_c refers to the test for superior predictive ability of Hansen (2005), and SPA_l and SPA_u are the lower and upper bounds of Hansen's test statistic. The mean returns are annualized and in percentage terms.

The evidence for the second sub-sample as shown in Panel B of Table 3 is, however, much weaker. While the nominal p-value is smaller than 1% for each currency, the corresponding White p-value is much larger, indicating that the significance suggested by the nominal p-value is misleading in the second sub-sample. After correcting for data snooping, the null hypothesis can be rejected at the 1% level only for the Japanese yen, and at the 10% level for the Canadian dollar, the deutsche mark, and the Swiss franc. The null hypothesis cannot be rejected even at the 10% level for the remaining three currencies.

Hansen (2003) points out that White's Reality Check may be impacted by the presence of outliers and trading strategies with low profitability, rendering the test with low power. Since this is more likely to be a problem in the second sub-sample, we implement a more powerful test for superior predictive ability suggested by Hansen (2005) for the second sub-sample. Test results are reported in Table 4 where White's p-values are repeated for ease of comparison. We find that the test results based on Hansen's SPA_c statistic are qualitatively the same as White's Reality Check for all currencies except the Italian lira where the null hypothesis can be rejected by SPA_c at the 10% level (p-value $= 0.090$) but cannot be rejected by the Reality Check (p-value $= 0.110$). The fact that the Reality Check and the SPA tests produce similar results suggests that few of the technical trading rules do much worse than the benchmark.[14]

Results for the Sharpe ratio criterion and the Jensen alpha criterion are qualitatively the same as those for the mean return criterion. Also results with transaction cost filters (not reported) remain qualitatively similar. These results indicate that the null hypothesis can be rejected strongly for the first sub-sample, but not for the second sub-sample.

14. Hansen and Lunde (2005) show that sometimes a single poor alternative can have a large impact on the Reality Check.

In summary, we find that for the full sample even after adjustments for data-snooping biases are made, the best-performing rules generate significant excess returns, Sharpe ratios, and Jensen alpha's for all seven currencies. The excess returns cannot be easily attributed to systematic market risk. However, most predictability of exchanger rates comes from the early part of the sample. Trading-rule profitability becomes much weaker in the second part of the sample, suggesting that exchange rates may have undergone structural changes and the market has become more efficient over time.[15] These results are consistent with LeBaron (2002) and Olson (2004), who claim that technical analysis has become less profitable in the recent decade.

4. FURTHER RESULTS ON TRADING-RULE PROFITABILITY

The proceeding section demonstrates that when technical trading rules are applied to the FX market, a potential data-snooping problem arises and can lead to incorrect inference. Our results for seven U.S. dollar exchange rates show that after correcting for such biases, there exists substantial evidence that mechanical strategies yield superior performance for the first half of the sample, but the evidence becomes much weaker in the second part of the sample. In this section, we conduct further analyses to examine the robustness of these results. As can be seen in the preceding section, the statistical results based on the Sharpe ratio and the Jensen alpha criteria are similar to those based on the mean return criterion. We also confirm this finding for all experiments in this section. To economize on space, we only report results based on the mean return criterion in this section.

4.1 Evidence from Cross Exchange Rates

One distinct characteristic of the U.S. dollar exchange rates over the post-Bretton Woods period is that they exhibit very long swings. In particular, the dollar shows persistent appreciation in the early 1980s and deep depreciation starting in the mid 1980s. The simple trading rules work perhaps because of the trend-like behavior of the dollar exchange rates, making it relatively easy for the rules to identify the general pattern. In this subsection, we employ the Japanese yen and the deutsche mark respectively as the vehicle currency and apply the same rules to these cross exchange rates. This experiment is warranted because it allows us to examine whether the profitability of technical rules is purely a dollar event or a phenomenon in the FX market in general. Neely, Weller, and Dittmar (1997) examine the predictability of the Japanese yen–deutsche mark and Swiss franc–British pound cross rates as a guard against data mining and find significant excess returns. On the other hand, Lee and Mathur (1996) report little profitability using six European currency cross rates. The Reality Check test applied to cross exchange rates can therefore shed more light on this issue.

---

15. Previous studies report that exchange rates in many developed countries may be subject to structural breaks. See Perron and Vogelsang (1992) and Wu (1997) among others. Our results provide another piece of evidence supporting the view that exchange rate series may be non-stationary over time.

Panel A of Table 5 reports the results when the Japanese yen is used as the vehicle currency. For the full sample, using the nominal *p*-value, the null hypothesis is rejected at the 1% level for all currencies. After adjusting for data-snooping biases, the null hypothesis is rejected at the 1% level for all currencies except for the Swiss franc where the null is rejected at the 5% level. Furthermore, the mean excess returns are all economically significant. It is interesting to note that the filter rule is selected to be the best trading rule for three currencies (the French franc, the Swiss franc, and the British pound). The filter rules involve much more frequent trading than the moving average and channel breakout rules, and greatly increase transaction costs. Using the same 0.04% transaction cost per one-way trade, the *after-cost* return for the French franc gets reduced by roughly a half to 5.36% per annum. For the Swiss franc, the transaction costs from the 5,122 trades nearly eliminate all profits, leaving the *after-cost* return to a mere 0.55% per annum. Similarly, for the British pound the *after-cost* annual excess return is only 1.57% after accounting for transactions costs of 6,334 trades.

A sub-sample analysis shows, however, that trading rules appear to be useful mostly for the first sub-sample only. For the second sub-sample, after accounting for data-snooping biases, the profitability is statistically significant at the 5% level only for the lira–yen and the pound–yen exchange rates.

Results for five currency prices against the deutsche mark are reported in Panel B of Table 5. For the full sample, the null hypothesis can be rejected at the 1% level for the Canadian dollar, the Swiss franc, and the British pound, and at the 5% level for the French franc. All five currencies select moving averages as the best trading rules. The excess returns are in general economically less significant than those for the dollar exchange rates. Results from sub-sample analyses are in general much weaker. While the null hypothesis can be rejected at the 5% level for four out of five exchange rates for the first sub-sample, it cannot be rejected at the 5% level for any exchange rate for the second sub-sample after correcting for data snooping.

Overall, results for the cross exchange rates are qualitatively similar to those for the U.S. dollar rates. Trading rules appear to work for the first sub-sample for most currencies, but become mostly unprofitable for the second sub-sample. These results are consistent with Lee and Mathur (1996) but are in contrast with Neely, Weller, and Dittmar (1997).

4.2 Out-of-Sample Test

As pointed out by Lo and MacKinlay (1990) and others, out-of-sample experiments are a way of purging the effects of data-snooping biases in a given analysis. Our last check for the robustness of the profitability of technical trading rules involves an out-of-sample test. To this end, we consider the seven U.S. dollar exchange rates and separate again the full sample into two sub-samples with roughly equal sizes. We use the first sub-sample (1974–85) to select the best rule out of our universe of 2,127 rules based on the mean return criterion. The best-performing rule selected

TABLE 5

PERFORMANCE OF THE BEST TRADING RULE: CROSS EXCHANGE RATES

Currency	Best Trading Rule	Number of Trades	Mean Return	Nominal p-value	White's p-value	Best Trading Rule	Number of Trades	Mean Return	Nominal p-value	White's p-value
	Panel A. Japanese Yen as Vehicle Currency					Panel B. Deutsche Mark as Vehicle Currency				
Full Sample 1974–98										
CAN	CBO($n = 5$, $x = 0.01$, $b = 0.0005$, $c = 1$)	790	12.17	0.000	0.000	MA($n = 5$, $b = 0.005$)	1208	11.11	0.000	0.000
FRA	Filter($x = 0.01$, $e = 10$)	2722	9.72	0.000	0.000	MA($n_1 = 2$, $n_2 = 50$, $c = 5$)	390	3.12	0.000	0.020
GER	MA($n = 15$, $b = 0$)	1354	10.50	0.000	0.000					
ITA	MA($n_1 = 5$, $n_2 = 25$, $c = 1$)	460	10.71	0.000	0.000	MA($n_1 = 15$, $n_2 = 20$, $c = 5$)	634	4.69	0.000	0.058
SWI	Filter($x = 0.001$, $c = 1$)	5122	8.75	0.000	0.020	MA($n_1 = 15$, $n_2 = 20$, $d = 3$)	610	4.53	0.000	0.008
UK	Filter($x = 0.001$, $e = 2$)	6334	11.70	0.000	0.000	MA($n_1 = 20$, $n_2 = 50$, $b = 0$)	230	7.15	0.000	0.000
Sub-sample 1974–85										
CAN	TRB($n = 5$, $c = 5$, $d = 2$)	100	14.66	0.000	0.000	MA($n = 25$, $c = 5$)	192	14.78	0.000	0.000
FRA	MA($n = 10$, $c = 10$)	214	12.52	0.000	0.000	MA($n_1 = 5$, $n_2 = 10$, $d = 2$)	564	5.44	0.000	0.042
GER	Filter($x = 0.01$, $e = 10$)	1048	13.64	0.000	0.000					
ITA	CBO($n = 15$, $x = 0.05$, $b = 0.0005$)	78	10.73	0.000	0.020	MA($n_1 = 2$, $n_2 = 5$, $d = 4$)	372	6.17	0.002	0.158
SWI	Filter($x = 0.001$, $c = 1$)	2354	12.84	0.000	0.002	MA($n_1 = 20$, $n_2 = 25$, $b = 0$)	296	7.51	0.000	0.000
UK	MA($n = 2$, $b = 0$)	2058	13.52	0.000	0.004	TRB($e = 4$, $b = 0.001$, $c = 5$)	148	11.19	0.000	0.004
Sub-sample 1986–98										
CAN	CBO($n = 5$, $x = 0.01$, $b = 0.0005$, $c = 1$)	372	12.13	0.000	0.062	MA($n = 5$, $b = 0.005$)	718	10.59	0.000	0.170
FRA	MA($n_1 = 2$, $n_2 = 25$, $d = 4$)	242	9.83	0.002	0.084	MA($n_1 = 5$, $n_2 = 25$, $c = 25$)	124	1.52	0.012	0.784
GER	Filter($x = 0.0005$, $c = 1$)	2844	9.69	0.000	0.106					
ITA	MA($n_1 = 10$, $n_2 = 15$, $d = 2$)	422	14.04	0.000	0.002	TRB($n = 25$, $c = 10$, $d = 2$)	24	6.09	0.002	0.056
SWI	TRB($n = 5$, $b = 0.001$, $c = 10$)	120	8.46	0.000	0.350	TRB($n = 25$, $b = 0.0005$, $c = 5$)	54	3.01	0.002	0.674
UK	Filter($x = 0.001$, $e = 2$)	3400	12.48	0.000	0.028	TRB($n = 50$, $c = 10$)	18	6.91	0.000	0.132

NOTES: This table reports the performance of the best trading rule from the universe of 2,127 trading rules for each currency under the maximum mean return criterion. In Panel A, the Japanese yen is used as the vehicle currency, while in Panel B the deutsche mark is used as the vehicle currency. The "White's p-value" is computed by applying the Reality Check methodology to the universe of 2,127 trading rules and incorporates the effects of data-snooping bias. The "nominal p-value" is calculated by applying the Reality Check methodology to the best trading rule only. The mean returns are annualized and in percentage terms.

TABLE 6

OUT-OF-SAMPLE (1986–98) PERFORMANCE OF THE BEST RULE SELECTED FROM THE
SAMPLE PERIOD (1974-–85)

Currency	Best Trading Rule	In Sample (1974–85)		Out of Sample (1986–98)		
		Number of Trades	Mean Return	Number of Trades	Mean Return	p-value
CAN	TRB($n = 10, c = 1$)	270	5.72	356	1.72	0.072
FRA	MA($n = 2, d = 3$)	458	14.42	578	5.00	0.038
GER	MA($n = 10, b = 0.0005$)	684	14.67	922	4.07	0.090
ITA	MA($n = 5, c = 5$)	494	12.88	648	1.03	0.378
JAP	MA($n_1 = 2, n_2 = 10, b = 0.0005$)	514	15.43	760	4.30	0.074
SWI	CBO($n = 5, x = 0.05, b = 0.001, c = 1$)	498	16.66	694	4.50	0.078
UK	MA($n = 10, b = 0$)	800	15.02	1054	4.81	0.030

NOTES: This table reports the out-of-sample (1986–98) performance of the best trading rule selected from the sample period of 1974–85 based on the mean return criterion. The U.S. dollar is used as the vehicle currency. The p-value for the out-of-sample period (1986–98) is calculated from bootstrap with 500 replications. Since there is only one rule to consider in the out-of-sample period, there is no data-snooping problem as such. The mean returns are annualized.

in the first sub-sample is then evaluated using the out-of-sample data (1986–98). The p-value for the out-of-sample period (1986–98) is calculated from bootstrap with 500 replications. Since there is only one rule to consider in the out-of-sample period, there is no data-snooping problem.

Table 6 reports the mean excess return and the associated p-value for the out-of-sample test. For ease of comparison, the in-sample (1974–85) performance of the best trading rule is also displayed. We find that for each currency the best-selected rule that outperforms the benchmark significantly in the first half of the sample in general becomes much less significant out of sample. The mean excess returns, while positive, all become less important economically. Furthermore, for the out-of-sample period the null hypothesis can be rejected at the 5% level only for the French franc and the British pound. We also conduct the out-of-sample test based on the Sharpe ratio criterion and the Jensen alpha criterion, respectively. Results are similar to those based on the mean return criterion and are omitted for brevity.

5. CONCLUSION

Researchers have long been interested in examining whether technical trading strategies can predict changes in exchange rates and whether the FX market is efficient. While numerous studies document significant profitability of certain rules, it is unclear to what degree these results are subject to data snooping. The fundamental problem arises when more than one trading rule is applied to a given financial time series, only the results of the best rule are reported, and the statistical analysis of the performance of the best rule does not take into consideration the whole universe of the candidate rules. Although the data-snooping problem has been widely recognized, it has not been properly treated due to the lack of commonly accepted and easily implementable procedures.

Using White's (2000) Reality Check, our paper makes a first attempt to formally characterize the data-snooping biases of various popular mechanical algorithms in the FX market and to re-examine the profitability of these rules. Evidence from seven U.S. dollar exchange rates shows that for the full sample the null hypothesis that trading rules do not produce excess returns can be rejected at the 1% significance level even after data-snooping biases are properly accounted for. The excess returns generated by these trading rules are not only statistically significant, but also economically important. The same conclusion can be reached under the Sharpe ratio criterion which measures the excess return per unit risk, and under the Jensen alpha criterion which measures the market risk-adjusted excess return. A sub-sample analysis indicates, however, that the profitability of trading rules is much less pronounced in the second part of the sample than in the first part. Furthermore, an out-of-sample test indicates that the null hypothesis can be rejected at the 5% level for only two currencies. These results suggest that exchange rate series may have undergone structural changes over time, and the FX market has become increasingly more efficient. Our analyses for cross exchange rates produce similar results and confirm the robustness of our basic findings.

Several researchers have attempted to explain why technical rules are profitable in terms of their economic contents. LeBaron (1999) claims that much exchange rate predictability is due to central bank intervention. Neely and Weller (2001) and Neely (2002) show that trading-rule returns precede central bank interventions. Intervention does not create market inefficiency from which technical rules profit. But rather, intervention is intended to halt predictable trends from which trading rules have profited. Sapp (2004) finds that both profitability and market uncertainty increase preceding central bank intervention and remain high during intervention. He demonstrates that a time-varying risk premium may play a possible role around interventions. Kho (1996) also shows that time-varying risk premium and conditional volatility explain a substantial part of trading-rule returns. However, before researchers are ready to consider these and other economic explanations, it is important to know to what extent the existing statistical results are subject to data snooping. Our preliminary analysis using the Reality Check makes such an attempt and produces useful results indicating that data-snooping biases are more serious in the later part of the sample, and that the evidence of trading-rule profitability becomes much weaker due to potential structural changes in exchange rate dynamics. We do not find significant empirical evidence suggesting that the FX market is inefficient.

APPENDIX
TRADING RULES INVESTIGATED

Selection Rationale

In order to obtain meaningful estimate of data-snooping biases, we include a fairly large number of specifications in our universe. The types of rules we consider are the most popularly employed by market participants and are extensively studied by previous researchers. See for example, LeBaron (1999), Levich and Thomas

(1993), Neely (1997), Neely, Weller, and Dittmar (1997), Osler (2003), Rosenberg (1996), Schulmeister (1987), Surajaras and Sweeney (1992), Sweeney (1986), and the Group of Thirty (1985). Some of the existing studies report a selective number of parameterizations. For example, Sweeney (1986) chooses filter sizes equal to 0.005, 0.01, 0.02, 0.03, 0.04, 0.05, and 0.1. Levich and Thomas (1993) use similar filter rules and moving average rules of 1–5, 5–20, and 1–200 days. LeBaron (1999) considers the 150-day moving average rule. Schulmeister (1987) employs double moving average rules of 3–10, 5–10, and 4–16 days. Most technical analysis studies do not report their choice of parameter values. Lo and MacKinlay (1990) point out that many rules that did not generate superior returns do not get published and are therefore filtered out by subsequent studies focusing on successful ones. Because of this problem, it is important not only to include the rules/parameterizations that were actually publicly reported, but also to include other rules/parameterizations that were ever considered. Intuitively, choosing too few rules can cause biases in statistical inference due to data mining. On the other hand, loading too many irrelevant rules can reduce the test power. This latter point is emphasized by Hansen (2003). Monte-Carlo simulations by Sullivan and White (1999) show that the test appears to be conservative. We therefore find a balance and select a fairly large variety of reasonable parameters that lie in the ranges used in the literature. They cover the sorts of rules that FX traders may have considered over time. Because rules with similar parameters are not independent and we want to keep the amount of computation manageable, we choose the grid of parameters naturally and not evenly. Overall, the candidate trading rules included in our universe are by far the largest in number and the most comprehensive in scope in foreign exchange studies.

Notations

S_t: Exchange rate (U.S. dollar price of one unit foreign currency), $t = 1, 2, ..., T$ ($T = 6,463$)

s_t: $\ln S_t$

g: transaction cost adjustment factor, $g = 0, 0.025\%, 0.05\%$

r_t: domestic interest rate

r_t^*: foreign interest rate

I_t: signal = 1: long; = 0: neutral; = −1: short

R_t: rate of return from a trading rule

$R_t = (s_t - s_{t-1} + r_t^* - r_t)I_{t-1} - abs(I_{t-1} - I_{t-2})g$

Benchmark: $I = 0$, no participation in the FX market.

A. Filter Rules

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

y: decrease in the log dollar 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

$x = 0.0005, 0.001, 0.005, 0.01, 0.05, 0.10$ (6 values)

$y = 0.0005, 0.001, 0.005, 0.01, 0.05$ (5 values)

$e = 1, 2, 5, 10, 20$ (5 values)

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

Total number of filter rules $= x{\cdot}c + x{\cdot}e + x{\cdot}y = 24 + 30 + 15 = 69$

B. 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

$n = 2, 5, 10, 15, 20, 25, 50, 100, 150, 200, 250$ (11 values)

$m = \sum_{i=1}^{10} i = 55$

$b = 0, 0.0005, 0.001, 0.005, 0.01, 0.05$ (6 values)

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

$c = 5, 10, 25$ (3 values)

Total number of MA rules $= nb + mb + nd + md + nc + mc$
$$= 66 + 330 + 44 + 220 + 33 + 165 = 858$$

C. Trading Range Break (TRB, or Support and Resistance) Rules

n: number of days in the support and resistance range

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

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

$n = 5, 10, 15, 20, 25, 50, 100$ (7 values)

$e = 2, 3, 4, 5, 10, 25, 50$ (7 values)

$b = 0.0005, 0.001, 0.005, 0.01, 0.05$ (5 values)

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

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

Total number of TRB rules $= nc + ec + nbc + ebc + ncd + ecd = 28 + 28 + 140 + 140 + 112 + 112 = 560$

D. Channel Breakout Rules (CBO)

n: number of days for a channel

x: difference between the high price and the low price ($x \times$ 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

$n = 5, 10, 15, 20, 25, 50, 100, 200$ (8 values)

$x = 0.001, 0.005, 0.01, 0.05, 0.10$ (5 values)

$b = 0.0005, 0.001, 0.005, 0.01, 0.05$ (5 values)

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

Note that b must be less than x. There are 15 xb combinations.

Total number of CBO rules $= n \cdot x \cdot c + n \cdot c \cdot (xb$ combinations$) = 160 + 480 = 640$

Total number of trading rules $= 69 + 858 + 560 + 640 = 2{,}127$

LITERATURE CITED

Alexander, Gordon J. (2000). "On Back-Testing 'Zero-Investment' Strategies." *Journal of Business* 73, 255–278.

Baillie, Richard T., and Tim Bollerslev (1994). "Cointegration, Fractional Cointegration and Exchange Rate Dynamics." *Journal of Finance* 49, 737–745.

Bessembinder, Hendrik (1994). "Bid-Ask Spreads in the Interbank Foreign Exchange Markets." *Journal of Financial Economics* 35, 317–348.

Brock, William, Josef Lakonishok, and Blake LeBaron (1992). "Simple Technical Trading Rules and the Stochastic Properties of Stock Returns." *Journal of Finance* 37, 1731–1764.

Cheung, Yin-Wong, and Menzie D. Chinn (2001). "Currency Traders and Exchange Rate Dynamics: A Survey of the U.S. Market." *Journal of International Money and Finance* 20, 439–471.

Cornell, W. Bradford, and J. Kimball Dietrich (1978). "The Efficiency of the Market for Foreign Exchange under Floating Exchange Rates." *Review of Economics and Statistics* 60, 111–120.

Diebold, Francis X., and Roberto Mariano (1995). "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics* 13, 253–265.

Group of Thirty (1985). *The Foreign Exchange Market in the 1980's*. New York, NY.

Hansen, Peter R. (2003). "Asymptotic Tests of Composite Hypotheses." Mimeo, Brown University.

Hansen, Peter R. (2005). "A Test for Superior Predictive Ability." *Journal of Business and Economic Statistics* 23, 365–380.

Hansen, Peter R., and Asger Lunde (2005). "A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?" *Journal of Applied Econometrics* 20, 873–889.

Jensen, Michael C. (1968). "Problems in Selection of Security Portfolios: The Performance of Mutual Funds in the Period 1945–1964." *Journal of Finance* 23, 389–416.

Kho, Bong-Chan (1996). "Time-Varying Risk Premia, Volatility, and Technical Trading Rule Profits: Evidence from Foreign Currency Futures Markets." *Journal of Financial Economics* 41, 249–290.

LeBaron, Blake (1999). "Technical Trading Rule Profitability and Foreign Exchange Intervention." *Journal of International Economics* 49, 125–143.

LeBaron, Blake (2002). "Technical Trading Profitability in Foreign Exchange Markets in the 1990's." Mimeo, Brandeis University.

Lee, Chun I., and Ike Mathur (1996). "Trading Rule Profits in European Currency Spot Cross-Rates." *Journal of Banking and Finance* 20, 949–962.

Levich, Richard M., and Lee R. Thomas (1993). "The Significance of Technical Trading-Rule Profits in the Foreign Exchange Markets: A Bootstrap Approach." *Journal of International Money and Finance* 12, 451–474.

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

Mark, Nelson (1995). "Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability." *American Economic Review* 85, 201–218.

McCormick, Frank (1975). *Transactions Costs in the Foreign Exchange Markets under the Fixed and Floating Exchange Rates*. Riverside, CA: University of California.

Neely, Christopher (1997). "Technical Analysis in the Foreign Exchange Market: A Layman's Guide." *Federal Reserve Bank of St. Louis Review*, September/October.

Neely, Christopher (2002). "The Temporal Pattern of Trading Rule Returns and Exchange Rate Intervention: Intervention Does Not Generate Technical Trading Profits." *Journal of International Economics* 58, 211–232.

Neely, Christopher, and Paul A. Weller (2001). "Technical Analysis and Central Bank Intervention." *Journal of International Money and Finance* 20, 949–970.

Neely, Christopher, Paul Weller, and Rob Dittmar (1997). "Is Technical Analysis in the Foreign Exchange Market Profitable? A Genetic Programming Approach." *Journal of Financial and Quantitative Analysis* 32, 405–426.

Olson, Dennis (2004). "Have Trading Rule Profits in the Currency Markets Declined over Time?" *Journal of Banking and Finance* 28, 85–106.

Osler, Carol L. (2003). "Currency Orders and Exchange Rate Dynamics: An Explanation for the Predictive Success of Technical Analysis." *Journal of Finance* 58, 1791–1820.

Perron, Pierre, and Timothy J. Vogelsang (1992). "Nonstationarity and Level Shifts with an Application to Purchasing Power Parity." *Journal of Business and Economics Statistics* 10, 301–320.

Politis, Dimitris, and Joseph Romano (1994). "The Stationary Bootstrap." *Journal of the American Statistical Association* 89, 1303–1313.

Qi, Min, and Yangru Wu (2003). "Nonlinear Prediction of Exchange Rates with Monetary Fundamentals." *Journal of Empirical Finance* 10, 623–640.

Rosenberg, Michael R. (1996). *Currency Forecasting, A Guide to Fundamental and Technical Models of Exchange Rate Determination.* Chicago: Irwin.

Sapp, Stephen (2004). "Are All Central Bank Interventions Created Equal? An Empirical Investigation." *Journal of Banking and Finance* 28, 443–474.

Schulmeister, Stephen (1987). "An Essay on Exchange Rate Dynamics." Mimeo, Austrian Institute of Economics Research.

Shapiro, Alan C. (1999). *Multinational Financial Management*, 6th edition. Upper Saddle River, NJ: Prentice-Hall.

Sullivan, Ryan, and Halbert White (1999). "Finite Sample Properties of the Bootstrap Reality Check for Data-Snooping: A Monte-Carlo Assessment." Mimeo, UCSD.

Sullivan, Ryan, Allan Timmermann, and Halbert White (1999). "Data-Snooping, Technical Trading Rule Performance, and the Bootstrap." *Journal of Finance* 54, 1647–1691.

Sullivan, Ryan, Allan Timmermann, and Halbert White (2001). "Dangers of Data Mining: The Case of Calendar Effects in Stock Returns." *Journal of Econometrics* 105, 249–286.

Sullivan, Ryan, Allan Timmermann, and Halbert White (2003). "Forecast Evaluation with Shared Data Sets." *International Journal of Forecasting* 19, 217–227.

Surajaras, Patchara, and Richard Sweeney (1992). *Profit-Making Speculation in Foreign Exchange Markets.* Boulder, CO: Westview Press.

Sweeney, Richard (1986). "Beating the Foreign Exchange Market." *Journal of Finance* 41, 163–182.

Szakmary, Andrew C., and Ike Mathur (1997). "Central Bank Intervention and Trading Rule Profits in Foreign Exchange Markets." *Journal of International Money and Finance* 16, 513–535.

Taylor, Mark P., and Helen Allen (1992). "The Use of Technical Analysis in the Foreign Exchange Market." *Journal of International Money and Finance* 11, 304–314.

West, Kenneth (1996). "Asymptotic Inference about Predictive Ability." *Econometrica* 64, 1067–84.

White, Halbert (2000). "A Reality Check for Data Snooping." *Econometrica* 68, 1097–1126.

Wu, Yangru (1997). "The Trend Behavior of Real Exchange Rates: Evidence from OECD Countries." *Weltwirtschaftlichcs Archiv* 133, 282–296.