Essays in international finance

by

Nima Zarrabi

This thesis is submitted for the degree of

Doctor of Philosophy in Finance

Essex Business School

University of Essex

December 2016

To my wonderful parents and brother who have always loved me unconditionally and have taught me to work hard for the things that I aspire to achieve.

Acknowledgements

First and foremost, I offer my sincerest gratitude to my supervisors, Stuart Snaith and Jerry Coakley for their continuous support, patience and immense knowledge. Their guidance helped me in all the time of research and writing of this thesis. I could not have imagined having better supervisors.

Besides my supervisors, I would like to thank the examiners of my thesis, Raphael Markellos and Neil Kellard, for their insightful comments and helpful suggestions. Furthermore, my sincere thanks also goes to Invesco for offering me the internship opportunity in their investment risk team.

Last but not the least, I would like to express my deepest gratitude to my parents and my brother for their endless love and support they provided me through my entire life. I am also sincerely thankful to my beautiful girlfriend, Milica, for her never-failing patience and unwavering love. I could not have completed this effort without her by my side.

Abstract

This thesis investigates three issues related to foreign exchange market. The fist issue is whether commodity prices can beat random walk benchmark by generating more accurate out-of-sample forecasts. The empirical results show that contemporaneous prices outperform the random walk at the daily frequency, however, this predictive ability disappears for monthly data. Using lagged commodity prices, we show that integrating the whole set of commodities into one large model or equally combining forecasts generated by each commodity individually improves the accuracy of the forecasts, implying outperforming the driftless random walk benchmark.

The second issue is the profitability of technical trading rules in foreign exchange market and whether it is consistent with the efficient market hypothesis. The results support profitability of trading rules for different currencies. However, to determine whether one could consistently speculate in the market, we perform a persistence analysis. We construct a portfolio of outperforming rules for each currency at the end of each month and use the selected rules in the following month. These results indicate that profitability of technical trading rules are purely due to luck.

The final issue is the performance of technical analysis and fundamental analysis in forecasting exchange rates. Due to parameter instability, the focus is on local forecasting performance of technical and economic models. We select models with the best performance based on three different criteria on a monthly basis and use them to generate forecasts for the next period. Our results show that if forecasts generated by selected technical and economic models are combined with equal weight, the random walk is beaten by all three criteria. These results underline the importance of considering both fundamental and technical factors in forecasting exchange rates.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The foreign exchange market is the largest financial market in the world, with a daily trading volume of about five trillion U.S. dollars (BIS, 2016). This market connects international institutions participating in currency trading all around the world and encompasses all economic factors, political conditions and market psychology. Due to the large volume, it is virtually impossible to control the market from outside. The foreign exchange market is open 24 hours a day, 5 days a week, due to time differences, with a maximum volume between 13:00 and 16:00 GMT, when both American and European markets are open (Drożdż et al.'s (2010)). All these features make this market an exceptionally complex network.

Attempts to explain exchange rate behavior began with the collapse of the Bretton-Woods system and the introduction of flexible exchange rate regimes. While initial studies found that exchange rates were predictable (see Dornbusch, 1976; Frankel, 1979), the seminal study by Meese and Rogoff (1983) claimed that empirical exchange rate models could not generate more accurate out-of-sample forecasts than a random walk model. In the literature, their finding is known as Meese-Rogoff puzzle. The subsequent literature provided rather mixed and controversial findings. In other words, while the majority of studies support the Meese-Rogoff puzzle, contributions to the recent literature have claimed to be able to forecast exchange rates. Alongside this line of research, many questions have been raised. For instance, which predictors does one use, which model does one estimate, which data frequency does one use, and which evaluation method does one use?

This thesis addresses some of the above questions by focusing on the most important exchange rates, and applying techniques that are novel in the foreign exchange market literature. The goal of this thesis is threefold. First, it revisits the predictive ability of daily and monthly commodity prices in forecasting exchange rates. Second, it analyses the profitability of technical trading rules in the foreign exchange market, while controlling for data snooping and transaction costs. Finally, it investigates the out-of-sample performance of a combination of technical analysis and fundamental analysis in forecasting exchange rates.

1.1 Literature

There has been an ongoing debate on the forecastability of exchange rates following the Meese and Rogoff (1983) study. A wide variety of predictors, models, estimation methods and evaluation tests at different frequencies have been used in the literature to address the Meese-Rogoff puzzle. Since different researchers choose different inputs, their findings are contradictory which causes the literature on predictability of exchange rates to be controversial. Rossi (2013) explains that the answer to the question "are exchange rates predictable?" is that it depends on the several choices about the inputs made by researchers. Therefore, there are studies that find economic models with outperforming forecasts (see Clark and West, 2006; Engle, 1994; Molodtsova and Papell, 2012; Rime et al., 2010; Rossi, 2006; West et al., 1993). while others support the Meese and Rogoff (1983a,b) findings (see Bacchetta et al., 2010; Berkowitz and Giorgianni, 2001; Chinn, 1991; Kilian, 1999; Qi and Wu, 2003; Rossi and Sekhposyan, 2011).

There is an extensive literature on the relationship between commodities and exchange rates. Chen et al. (2010) find that while exchange rates forecast commodity prices at the quarterly frequency, commodity prices are unable to forecast exchange rates. Chen and Rogoff (2003) find that commodity prices indices can generate accurate forecasts for developing currencies at quarterly frequency. Their findings are consistent with those of Cashin et al. (2004) and Coudert et al. (2008) for monthly and annually data, respectively. In a more recent study, Ferraro et al. (2015) find that there is a robust predictive ability for contemporaneous oil prices in forecasting daily Canadian dollar prices. However, predictive ability is very short-lived when lagged prices are used and absent in monthly data. They explain that this is due to ephemeral impact of changes in oil prices on exchange rates.

Technical trading rules use past price behaviour to guide trading decisions in asset markets. The subject of an ongoing debate is whether technical analysis can consistently produce profit or is just a matter of pure luck. The broad use of technical analysis in the foreign exchange market is highlighted in different surveys. For instance, Allen and Taylor (1990) and Taylor and Allen (1992) show that almost all chief foreign exchange dealers in London use technical trading rules to some extent and also combine them with fundamental analysis. The Cheung and Chinn (2001) survey shows that 30% of U.S. FX traders could be considered as technical traders due to the increasing use oftechnical analysis. The broad literature on the performance of trading rules in the FX market finds supportive evidence for their profitability (see Cornell and Dietrich, 1978; Hsu and Taylor, 2013; Neely et al., 1997; Neely and Weller, 2011; Qi and Wu, 2006; Sweeney, 1986). However, these findings may be subject to data snooping since the early literature is mainly focused on the popular trading rules and the risk involved with technical analysis is often disregarded.

The extensive use of technical analysis by practitioners in the foreign exchange market has been given little attention in the literature of forecasting exchange rate. The surveys by Lui and Mole (1998), Cheung and Wong (2000) and Gehrig and Menkhoff (2004) find technical analysis as a better tool in forecasting trends than fundamental analysis, especially at horizons up to six months. Menkhoff and Taylor (2007) explain that the continued use of technical analysis is due to its ability in informing traders about non-fundamental price determinants. They suggest that academic researchers must understand technical analysis and integrate it into economic reasoning at both the macroeconomic and the microstructural levels.

1.2 Contributions

One of the main findings in the current thesis (Chapter 2) is that predictive ability of lagged commodity prices can generate more accurate forecasts than those of a random walk at the daily frequency when their forecasts are combined with equal weight or commodities are integrated into one large model. Our empirical results are consistent with Cashin et al. (2004) and Coudert et al. (2008) who find evidence for outperforming predictive ability for commodity prices. This is in contrast to Ferraro et al. (2015) who find that the predictive ability of lagged oil prices is ephemeral.

The other contribution of this thesis is that it revisits the profitability of technical trading rules by using a new methodology to control for data snooping (Chapter 3). To the best of my knowledge, this is the first study to use the false discovery rate approach of Barras et al. (2010) and perform a comprehensive persistence analysis to evaluate the performance of technical analysis in the foreign exchange market. The empirical results show that this methodology provides a large number of outperforming trading rules which enables investors to combine signals in multiple strategies instead of considering the signal of one trading rule at a time. It finds that, while there are periods with positive Sharpe ratios throughout the sample, these profit opportunities are short-lived, which implies that an investor could not make consistent profit by using previously outperforming trading rules. Our findings support adaptive market hypothesis by Lo (2004) and depart from studies that claim the profitability of technical analysis in the foreign exchange market (see Hsu and Taylor, 2013; Neely and Weller, 2011; Qi and Wu, 2006). Our results caution investors about websites that advertise trading strategies as obvious money-making tools.

The final contribution of this thesis is that it finds support for the predictability of exchange rates by combing technical analysis and fundamental analysis (Chapter 4). Due to the instability in models' forecasting performance, it focuses on local predictive ability, highlighted by Giacomini and Rossi (2010). It shows that an investor could choose outperforming models ex-ante on a monthly basis, on the basis of three different criteria. The empirical results propose that an equally-weighted combination of best technical models and macroeconomic models generates more accurate forecasts than the random walk. It checks robustness of the results by applying the dynamic model averaging method by Raftery et al. (2010). This improves the initial results by finding more support for predictive ability of technical analysis.

1.2.1 Chapter preview

The remainder of this thesis is structured as follows. Chapter 2 considers Australian dollar, Brazilian real, Canadian dollar and eight commodities which are the dominant exports of these countries and have significant shares in the world commodity market. The Clark and West (2007) results suggest that realized commodity prices can forecast all three exchange rates at the daily and monthly frequencies. The superior predictive ability by Hansen (2005) and stepwise SPA by Hsu et al. (2010) suggest that predictive ability of commodities is robust at the daily frequency while it disappears at the monthly frequency after taking into account data snooping. Although the results show that there is no predictive ability for lagged data when the performance of lagged models are evaluated locally, some commodities show significant predictive ability at some points in time. This is the motivation for integrating the whole set of information by constructing a large model consisting of all commodities, combining forecasts obtained from each commodity and applying principal component analysis. The results show that both information combination and forecast combination outperform the random walk benchmark at the daily frequency.

Chapter 3 investigates ten types of well-known and less-studied rules giving us a universe of 7,650 rules. The first contribution of this chapter is to use the false discovery rate of Barras et al. (2010) to check whether the profitability of technical analysis is robust to data snooping. Benjamini and Hochberg (1995) explain that the idea of this approach is that, by allowing a specific proportion of false discoveries, it significantly improves the power of detecting the outperforming rules. The second contribution is that it performs a comprehensive persistence analysis to test whether it would have been possible to select outperforming rules for a month using data prior to that month. It constructs a portfolio of outperforming rules at the end of each month, and uses these selected rules in the following month. It carries out the persistent test for daily U.S. dollar prices of six currencies and evaluates their performance based on Sharpe ratio. The results show that after taking into account data snooping and transaction costs, the profitability of trading rules is not consistent over time. In other words, despite the short-lived profit opportunities, an investor could not consistently make profit by using previously outperforming rules. The results support the adaptive market hypothesis of Lo (2004).

Chapter 4 highlights the importance of local performance evaluation. To do so, it compares two complimentary tests: Giacomini and White (2006) and fluctuation test by Giacomini and Rossi (2010). The former evaluates the relative performance of models globally while the latter studies the entire path of competing model's relative performance. The empirical results show that a few economic models and technical models beat the random walk benchmark at least once over the selected period. This motivates constructing three portfolios of selected models based on three criteria. The selected models are used to generate forecast in the following month, while the portfolios are updated on a monthly basis. This chapter finds that economic models selected by $\bar{R}^2$ outperform the benchmark, however, there is no evidence of predictive ability in the case of technical models. The main finding is that when forecasts generated by selected economic and technical models based on each criterion are combined, the random walk benchmark is beaten in all cases. The second methodology is applied to check for robustness of results. It applies dynamic model averaging approach by Raftery et al. (2010), that allows for recursive estimation and supports the initial results by finding predictive ability for economic models, as well as technical models.

Finally, Chapter 5 briefly summarizes the key findings of this thesis.

Chapter 2

Can commodity prices forecast exchange rates?

The difficulty in forecasting exchange rates has been a longstanding problem in international finance. The pioneering paper in this literature by Meese and Rogoff (1983) finds that a random walk benchmark cannot be beaten by a variety of linear structural models for both nominal and real exchange rates. Many studies have focused on the relationship between commodities, in particular oil, and exchange rates. Some studies support the predictive ability of commodity prices (see Aloui et al., 2013; Bénassy-Quéré et al., 2007; Narayan et al., 2008), while others contradict their findings (see Chen and Rogoff, 2003; Ferraro et al., 2015; Habib and Kalamova, 2007). The controversial and contradictory results are explained by Rossi (2013). She explains that the predictability depends on the choice of predictor, forecast horizon, sample period, model and forecast evaluation method. Although there is no consensus on predictability of exchange rates, recent literature has identified some methodologies that claim to have resolved the Meese and Rogoff puzzle.

This chapter builds on and extends the study by Ferraro et al. (2015) in terms of the number of commodities and currencies considered. It studies the predictive ability of commodity prices in forecasting three exchange rates in countries that have significant commodity exports. It focuses on the Australian dollar (AUD), the Brazilian real (BRL) and the Canadian dollar (CAD). It considers eight commodities that are the dominant exports of these countries and have significant shares in the world commodity market. It evaluates the out-of-sample forecasting performance of realized and lagged prices of the commodities in forecasting exchange rates at the daily and monthly frequencies. The Clark and West (2007) test statistics show that there is significant predictive ability for almost all contemporaneous commodity prices at both frequencies.

It applies the superior predictive ability test of Hansen (2005) with 1,000 bootstrap to control for data snooping. The results support the initial findings and show that predictive ability of contemporaneous commodity prices at the daily frequency is robust for all three currencies. Due to inability of the Hansen (2005) test in detecting the models with significant predictive ability, we apply the Hsu et al. (2010) test which is a stepwise procedure of superior predictive ability. The results show that six, four and three models beat the benchmark for the AUD, BRL and CAD, respectively. These findings are in line with the results of the Hansen et al. (2011) model confidence set.

Given the instability in the relationship between commodities and exchange rates, highlighted by Maier and DePratto (2008), this chapter applies the fluctuation test by Giacomini and Rossi (2010) to evaluate the performance of the models locally. It finds evidence for predictive ability of lagged commodity prices at least once throughout the sample. It is important to note that this information would be ignored if one would only evaluate the performance of the models globally. This finding is the motivation to incorporate the whole information set. To do so, it considers a large model which consists of all commodities (information combination), an equally-weighted average of forecasts generated by each commodity (forecast combination), and a principal component model.

The main contribution of this chapter is that it finds strong evidence for predictive ability of lagged commodity prices in forecasting exchange rates at the daily frequency, implying that both information combination and forecast combination outperform the benchmark. This highlights the importance of integrating the entire information set by either constructing a model consisting of all variables or combining the forecasts generated by them individually. The empirical results also show that, while at the daily frequency information combination generates more accurate forecasts than forecast combination, at the monthly frequency there is evidence of better performance by forecast combination for lagged data.

This paper proceeds as follows. Section 2.1 briefly reviews the literature on predictive ability of commodities in forecasting exchange rates. Section 2.2 describes the data set and discusses the framework used to generate forecasts. Section 2.3 presents the empirical results. Section 2.4 concludes.

2.1 Related literature

The evolution of exchange rates has always been debated by academics and policy-makers. An extensive literature has focused on forecasting exchange rates following the seminal paper by Meese and Rogoff (1983) finds that the random walk provides better forecasts for exchange rates than economic models. While several studies support the Meese-Rogoff puzzle (see Bacchetta et al., 2010; Berkowitz and Giorgianni, 2001; Chinn, 1991; Kilian, 1999; Qi and Wu, 2003; Rossi and Sekhposyan, 2011), many other papers claim that competing models generate more accurate forecasts than those of the random walk benchmark (see Clark and West, 2006; Engle, 1994; Molodtsova and Papell, 2012; Rime et al., 2010; Rossi, 2006; West et al., 1993).

The relationship between commodity prices, particularly oil prices, and exchange rates has been discussed in the academic literature for decades. Previously, the impact of exchange rates on commodity prices was studied. Several papers consider the use of exchange rates to reduce the effects of Dutch disease which describes a decline in the manufacturing sector caused by a commodity price boom, and in particular refers to the natural gas discoveries and the subsequent real exchange rate appreciation in the Netherlands after 1959. For instance, Frankel (2010) considers the 2003-2008 boom in world commodity markets and finds that it affected the currency of several Latin American commodity exporting countries and caused those countries to intervene in foreign exchange market to adjust for currency appreciation. Chen et al. (2010) investigate how the exchange rates of commodity currencies can be considered as in-sample and out-of-sample commodity prices determinants. The reverse relationship, using the commodity price index to predict exchange rates has produced less robust forecasts at quarterly frequency. The studies explain that their findings are due to the forward-looking characteristics of exchange rates and that it is not easy to capture those fluctuations by simple time series models.

Many studies have focused on the in-sample and out-of-sample predictive ability of oil prices in forecasting exchange rates. Issa et al. (2008) and Cayen et al. (2010) focus on the in-sample relationship between real oil prices and real exchange rates in quarterly data. More generally, studies by Chen and Rogoff (2003) and Cashin et al. (2004) consider commodity price indices in developed countries. The former performs an out-of-sample forecasting exercise on quarterly data and finds that a commodity price augmented model cannot beat the random walk benchmark. However, the latter considers monthly data and, using Engle-Granger approach, finds a long-run relationship between the real exchange rates and real commodity prices for about one-third of the countries in their sample. These findings are supported by Coudert et al. (2008) where a larger group of countries, including oil exporters is considered at the annual frequency.

Koranchelian (2005) studies the long-term impact of commodity prices, in particular oil prices, on the real exchange rate of Algeria. He finds the half-life to be 9 months, which is the time required for adjusting the deviations of real exchange rate towards its equilibrium. Habib and Kalamova (2007) examine the same relationship for Norway, Russia and Saudi Arabia at the quarterly frequency and find contradicting results. They find a positive relationship between the real oil price and Russian real exchange rate against U.S. dollar but this relationship does not apply to the Norwegian and Saudi Arabian currencies. They conclude that these currencies do not necessarily explain if there is a relationship between real oil prices and real exchange rates since exchange rate reactions to oil price changes might be explained by other governmental policies, or particular characteristics of the countries in the sample.

Aloui et al. (2013) find a positive relationship between oil prices and U.S. dollar exchange rates where an increase in oil prices leads to U.S. dollar depreciation. Some studies support their findings, while others establish an inverse relationship between oil prices and U.S. dollar in the post-crisis period. For instance, Narayan et al. (2008) examine the relationship between oil prices and the Fijian dollar currency using the generalised autoregressive conditional heteroskedasticity (GARCH) and exponential GARCH (EGARCH) for daily data, and find that a 10% increase in oil prices causes a 0.2% appreciation of the Fijian dollar. In a similar study, Akram (2009) estimates that an increase in commodity prices is caused by a fall in interest rate and the dollar real exchange rate depreciation. Conversely, Bénassy-Quéré et al. (2007) consider the 1974-2004 period and find that a 10% rise in the oil price causes a 4.3% appreciation of the dollar real effective exchange rate in the long run.

A country's dependency on importing or exporting oil is a key factor in discussing the relationship between oil prices and exchange rates. Lizardo and Mollick (2010) add oil prices to the monetary model and find a positive relationship between oil prices and net oil-exporting currencies such as the Canadian dollar, Mexican peso and Russian ruble over the 1970-2008 period. Aloui et al. (2013) find similar results where they show that a rise in real oil prices leads to an appreciation of the dollar in the case of net oil-importing countries such as Japan. Bodart et al. (2012) consider developing countries whose leading exports have a share of at least 20% in the total merchandise exports of the country. Their results show that the price of the leading commodity has a significant long-term impact on the real exchange rate at the monthly frequency where a larger share of the commodity leads to a larger impact on the real exchange rate.

Ferraro et al. (2015) study the contemporaneous and lagged relationship between the oil price and Canadian dollar. Their results show that using realised data, there is a significant relationship at the daily frequency, but that using lagged commodity price changes makes predictive ability very short-lived. While they find robust predictive ability at the daily frequency, at monthly and quarterly frequencies the random walk benchmark cannot be beaten. They explain that it is due to ephemeral impact of changes in oil prices on exchange rates.

Contrary results were obtained in an earlier study by Akram (2004) where a strong nonlinear relationship between the Norwegian krone/ECU and crude oil price was detected. The out-of-sample predictive ability of oil prices in forecasting the exchange rate was also examined using a nonlinear model. The findings showed that the forecasts outperformed those of the random walk model at the monthly and quarterly frequencies. Both nominal and real exchange rates were found to be consistent with the purchasing power parity theory where the half-life was about 1.5 years. Akram's (2004) findings are novel, since previous studies either rejected the purchasing power parity theory or found weak support for it with a half-life of 3-6 years. He noted that the contrary results may be due to the longer sample from the post-Bretton Woods period considered in the study. A particularly interesting conclusion drawn by Akram (2004) is how the effect on exchange rates is less pronounced when oil prices are higher. However, when prices are trending down and, in particular, when they drop below 14 dollars, the effect on the exchange rate is more significant. This may explain discrepancies in the literature generally that the effect of oil prices on the exchange rate is affected not only by directional movement in oil prices but also whether prices are initially higher or lower, and if they fall outside the long-standing price range.

2.2 Models

Our dataset contains daily and monthly prices of eight commodities and three U.S. dollar (USD) nominal exchange rates for the February 27, 1995 to February 27, 2013 period. We consider three exchange rates based on the amount of each per unit of USD in international trade: the Australian Dollar (AUD), the Brazilian Real (BRL) and the Canadian Dollar (CAD), obtained from WM Reuters. The summary statistics for daily and monthly changes in the logarithm of exchange rates are reported in Panels A and B of Table 2.1, respectively. The mean return rates show that on average the U.S. dollar appreciates against the BRL and depreciates against the AUD and CAD. The standard deviation of daily changes lies between 0.5436% and 0.8032% which implies substantial daily volatility. The volatility is remarkably lower for monthly data, lying between 0.0258% and 0.0525%. With the exception of the daily CAD, all distributions appear to be right skewed, while all daily and monthly return distributions have excess kurtosis compared to the normal distribution.

< Table 2.1 around here >

In this paper, we use the price of commodities with significant shares in the world commodity market. In other words, we consider ten commodity groups defined by the UN Comtrade database and calculate their relative average shares in the world export over the 1995-2013 period. We then use the commodity groups with the highest shares.[1] Table 2.2 reports the average share of natural resource commodity groups for the 1995-2013 period. The commodities included in our study are aluminum, coffee, copper, gas, gold, oil, sugar and wheat which have significant shares in the countries' exports. The countries included in our sample are price-takers in the world commodity market since their sizes in that market are relatively small. Table 2.3 represents the average share of commodity export for each country in total world commodity export throughout the sample. We consider 120 observations as the estimation period and one-step-ahead forecasts are calculated using a rolling specification for both frequencies. Since all the variables have stochastic trends according to the Augmented Dickey Fuller (ADF) unit root test, we consider the rate of growth of the variables by taking the first difference of logarithms.

< Tables 2.2 & 2.3 around here >

[1]Note that only natural resources are included in our sample and manufactured commodities are beyond the scope of this paper.

Thus, our models can be expressed as follows:

$$\Delta s_{t+1} = \alpha + \beta_i \Delta X_{i,t+1} + \epsilon_{t+1,i} \tag{2.1}$$

$$\Delta s_{t+1} = \alpha + \beta_i \Delta X_{i,t} + \epsilon_{t+1,i} \tag{2.2}$$

where $\Delta s_{t+1} = s_{t+1} - s_t$, α is constant, β is a vector of parameters, $X_{i,t+1}$ ($X_{i,t}$) is a $t \times 1$ matrix of regressors available at time $t+1$ (t), and $t = 1, \ldots, T$. Note that while Model 2.1 uses the contemporaneous value of commodity prices, Model 2.2 considers lagged commodity prices. Ferraro et al. (2015) explain that the latter is an actual out-of-sample forecast exercise, while the former is an out-of-sample fit.[2]

2.3 Empirical results

2.3.1 Out-of-sample forecasting

This section reports out-of-sample forecasting statistics for all the models defined in section 2.2. The out-of-sample forecast is given by

$$\Delta s_{t+1} = \hat{\beta}_{i,t} X_{i,t} \tag{2.3}$$

where $\hat{\beta}_{i,t}$ is the OLS estimate from regressing $\{\Delta s_k\}_{k-1}^{t}$ on $\{X_{i,k}\}_{k-2}^{t-1}$. The benchmark is the random walk without drift which is defined as the toughest benchmark by Rossi (2013).

We apply Clark and West (2006) (henceforth CW) to evaluate the forecasting performance of our models. The CW is one of the most popular out-of-sample test statistics for nested models and tests the null of equal predictive ability of a linear econometric model and a martingale difference model. The CW test statistic is the adjusted mean squared forecast error (henceforth MSFE-adjusted) statistic which is

[2]It is important to note that since the focus of this chapter is on out-of-sample evaluation rather than in-sample comparison of the competing models, model assessment (such as cointegration, omitted variable bias and nonlinearity) is beyond the scope of this chapter (see Sarno and Valente (2009)).

the modified version of the Diebold and Mariano (1995) and West (1996) statistic. The advantage of the CW over Diebold and Mariano (1995) and West (1996) is that it has an approximately standard normal asymptotic distribution when models are nested. Assume Model 1 is the parsimonious model and Model 2 is the larger model that nests model 1, and $\hat{y}_{1,t+1}$ and $\hat{y}_{2,t+1}$ are the forecasts of y_{t+1} obtained from Models 1 and 2, respectively. We have

$$\hat{\delta}_1^2 = P^{-1}\sum(y_{t+1} - \hat{y}_{1,t+1})^2 \tag{2.4}$$

$$\hat{\delta}_2^2 = P^{-1}\sum(y_{t+1} - \hat{y}_{2,t+1})^2 \tag{2.5}$$

$$\hat{\delta}_2^2 - adj. = P^{-1}\sum(y_{t+1} - \hat{y}_{2,t+1})^2 - P^{-1}\Sigma(\hat{y}_{1,t+1} - \hat{y}_{2,t+1})^2 \tag{2.6}$$

where $adj.$ is the sample average of $(\hat{y}_{1,t+1}-\hat{y}_{2,t+1})^2$, and P is the number of predictions used in calculating these averages. As suggested by Clark and West (2007), the most convenient way to generate the CW test statistic is

$$\hat{f}_{t+1} = (y_{t+1} - \hat{y}_{1,t+1})^2 - (y_{t+1} - \hat{y}_{2,t+1})^2 - (\hat{y}_{1,t+1} - \hat{y}_{2,t+1})^2. \tag{2.7}$$

Clark and West (2006) propose that the CW statistic should be interpreted as the minimum MSFE test statistic. Panel A in Table 2.4 shows that there is strong evidence for predictability of exchange rates at the daily frequency using contemporaneous commodity prices. In other words, all competing models outperform the random walk benchmark for all three currencies. The predictive ability is short-lived when lagged commodity prices are used to forecast future exchange rates which is due to time-variant relationship between commodity prices and exchange rates. This implies that while our results favor the contemporaneous models, the forecasts generated by lagged models are not more accurate than those of the benchmark, with the exception of sugar, aluminum and copper for the AUD, and Oil for the CAD. The implication is that changes in commodity prices are immediately translated into changes in exchange rates and, therefore, we find predictive ability for contemporaneous but not for lagged commodity prices.

At the monthly frequency, the results show that gold, aluminum, coffee and copper significantly outperform the random walk benchmark for all three currencies (see Table 2.5). We find strong evidence for predictive ability for realized oil prices in predicting the monthly BRL. Wheat is also found to outperform the random walk for the AUD and CAD. Comparing Panel A in Table 2.4 and Panel A in Table 2.5 highlights that predictive ability of commodities is short-lived and the frequency of the data is crucial in capturing it. Our results show that lagged commodity prices at monthly frequency have no predictive ability for forecast exchange rates. In other words, all the competing models are outperformed by the random walk benchmark.

< Tables 2.4 & 2.5 around here >

Rogoff and Stavrakeva (2008) demonstrate that one of the main drawbacks of using the CW test for nested models is that it cannot be always interpreted as the minimum MSFE test such as in Diebold and Mariano (1995) and West (1996).[3] They explain that this is more likely where scale bias exists and the CW tests whether the exchange rate is a random walk.[4] They suggest that, since using the asymptotic CW might seem appealing, it is important that one checks the robustness of the results.

One could question whether the previous findings are driven by data snooping. Data snooping occurs when the same data set is used to test the significance of different models. The original joint testing method for data snooping is proposed by White (2000) which is known as Reality Check (henceforth RC). The RC determines whether the best model in the sample has genuine predictive power after controlling for the effects of data snooping. The null hypothesis of the RC is that the performance of the best model does not beat the performance of the benchmark, however, RC is not able to identify further models that generate significant performance. Hansen (2005) points out that power of the RC can be reduced, and even be driven to zero

[3]However, Clark and West (2007) suggest that the CW statistic should be interpreted as the minimum MSFE test statistic.

[4]Holden and Peel (1989) explain that scale bias occurs when a forecaster, on average, undershoots or overshoots the forecast by some percentage.

when too many irrelevant models are included in the set of alternatives. Therefore, he suggests some improvements over the RC and proposes the superior predictive ability (henceforth SPA) test, which invokes a sample dependent distribution under the null. Hansen's (2005) method is more powerful because it is being less sensitive to the influence of poor and irrelevant models. On the other hand, the null hypothesis of the RC test is based on the least favourable configuration which may cause RC test to lose power when many poor models are considered, while the SPA test avoids it.[5]

To check for robustness of our results while controlling for data snooping, we perform the SPA test which evaluates models based on a loss function and defines the best model as the model with the smallest expected loss. Assume $\hat{Y}_{k,t}$ to be the forecast made by model k and realized value to be Y_t. The performance of model k is compared to that of the benchmark as follows:

$$d_{k,t} = L(Y_t, \hat{Y}_{0,t}) - L(Y_t - \hat{Y}_{k,t}) \qquad k = 1, \ldots, m \qquad t = 1, \ldots, T \tag{2.8}$$

The question addressed by the SPA is that whether any of the models $k = 1, \ldots, m$ outperforms the benchmark. The null hypothesis implies that the benchmark produces better forecasts:

$$H_0 : \mu_k = E[d_{k,t}] \leq 0 \quad \text{for all} \quad k = 1, \ldots, m$$

More compactly, it can be shown as an m-dimensional vector as follows:

$$\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix} = E \begin{pmatrix} d_{1,t} \\ \vdots \\ d_{m,t} \end{pmatrix}$$

Therefore, the null can be written as $H_0 : \mu \leq 0$ and a low p-value of 0.05 indicates that the benchmark is outperformed, while a higher p-value implies that the

[5]The least favourable configuration is known as the point least favourable to the alternative.

benchmark cannot be beaten. The SPA results find evidence of predictive ability for contemporaneous models at the daily frequency.

The drawback of both the RC and SPA models is that the null hypothesis is that the best model outperforms the benchmark and, therefore, it cannot detect further models with good performance. Romano and Wolf (2005) extend the RC and propose a stepwise multiple test (henceforth StepM), which allows an investor to identify as many significant models as possible for a given level of significance. Hsu et al. (2010) propose a stepwise SPA (henceforth SSPA) test which is a combination of the SPA test and StepM test, and is a more powerful in detecting all good performing models.[6]

The stepwise procedure enables us to identify significant models where the null distribution does not depend on the LFC. The SSPA test statistics are $\sqrt{n}\bar{d}_1, \ldots, \sqrt{n}\bar{d}_m$ and the test adopts the stationary bootstrap of Politis and Romano (1994). The bootstrapped SPA critical value is determined as $\hat{q}^*_{\alpha_0} = max(\hat{q}_{\alpha_0}, 0)$, where

$$\hat{q}_0 = inf\{q|P^*[\sqrt{n}max_{k=1,\ldots,m}(\bar{d}^*_k - \bar{d}_k + \hat{\mu}_k) \leq q] \geq 1 - \alpha_0\} \tag{2.9}$$

and $(1 - \alpha_0)$-th quantile of the re-centered empirical distribution, P^* is the bootstrapped probability measure, $\bar{d}^*(b) = \sum_{t=1}^{n} d^*_t(b)/n$ is the sample average of the re-sample obtained from the bootstrap, $d^*_t \equiv d^*_{n_{b,t}}$ is the b-th re-sample of d_t and $d_{k,t} = L(Y_t, \hat{Y}_{0,t}) - L(Y_t, \hat{Y}_{k,t})$ is the performance of model k, relative to the benchmark model at time t. The SSPA then proceeds as follows:

- Sort $\bar{d}_k$ in descending order.
- Reject the top model k if $\sqrt{n}\bar{d}_k > \hat{q}^*_{\alpha_0}$. If no model can be rejected, the procedure stops, otherwise, we go to the next step.

[6] The above procedure differs from the original StepM test of Romano and Wolf (2005) in a few ways. First, StepM use a circular block bootstrap instead of stationary bootstrap used in the SSPA. Second, in order to determine the block size of bootstrap, StepM relies on a data-dependent algorithm, while SSPA uses an ex-ante fixed value. Third, SSPA uses heteroskedasticity and autocorrelation consistent (HAC) estimator rather than bootstrapped standard error used by StepM. Hsu et al. (2010) suggest that these differences may affect the finite sample performance of the StepM test.

- Remove d_k of the rejected models from the data. Reject the top model i in the sub-sample of remaining observations if $\sqrt{n}\bar{d}_i > \hat{q}^*_{\alpha_0}$. If no model can be rejected, the procedure stops, otherwise, repeat this step until no model can be rejected.

The SSPA results are reported in Table 2.6. It shows that there is strong evidence for predictive ability of oil, aluminum and copper realized prices for all three currencies at the daily frequency. Our results also favours gold prices for the AUD and CAD, and sugar and wheat for the AUD.[7]

< Table 2.6 around here >

2.3.2 Local performance vs global performance

The relationship between commodities and exchange rates changes over time. For instance, Maier and DePratto (2008) detect in-sample parameter instabilities in the relationship between the Canadian dollar and commodity prices. This instability may affect the forecast performance of our models. This implies that tracking the model relative performance over the entire sample may yield useful information that is ignored when focusing on the overall performance. For instance, a model may perform best on average over a specific period, however, the better local performance of the competing model may be ignored as for example, when considering only the recent past. We perform the fluctuation test by Giacomini and Rossi (2010) to evaluate the performance of models that are unable to beat the benchmark.

The fluctuation test statistic is

$$F_{t,1,M} = \hat{\sigma}^{-1} M^{-1/2} \sum_{j=t-\frac{M}{2}}^{t+\frac{M}{2}-1} \Delta L_{1,j} \tag{2.10}$$

where $\hat{\sigma}^2$ is the HAC estimator of σ^2 calculated to capture the variance of the squared loss function, $\Delta L_{1,t}$. M is the size of the rolling window used to construct the test

[7] In Appendix A we show that the model confidence set of Hansen et al. (2011) provides consistent results.

statistics. We choose the value of M such that $M/P \approx 0.3$, where P is the number of observations in the out-of-sample period. The reason is that, as shown by Giacomini and Rossi (2010), when a rolling window size is a small fraction of the out-of-sample size, the fluctuation test has good size and power properties.

We find that there is evidence for predictive ability of lagged oil, sugar and wheat prices in forecasting daily AUD changes. Figure 2.1 shows that while aluminum beats the random walk from 2013 to mid-2004, oil prices show predictive ability in 2013. Our results show that lagged copper prices generate accurate forecasts for the 2003-2004 period as well as a short period in 2008. The fluctuation test result for BRL is represented by Figure 2.2. It shows that oil, copper and wheat lagged prices are able to outperform the random walk benchmark over last eight months of our sample, while copper shows predictive ability for the 2003-2005 period. Figure 2.3 indicates that the lagged oil model generate better forecasts for CAD that random walk at the daily frequency in March 2007 and February 2013. On the other hand, we find evidence for predictive ability of gas and aluminum prices at some points in 2004 and 2007. It is important to notice that this information about the local predictive ability of commodities is ignored when their performance is evaluated globally.

Our results show that although there is strong evidence for predictive ability of realized commodity prices at the daily frequency, lagged prices may also show predictive ability for a short period of time. The former implies that using accurate forecasts of commodity prices, one could forecast future exchange rates accurately. The latter, however, highlights short-lived predictive ability of lagged commodity prices as well as the importance of assessing the performance of the models locally rather than globally.

$<$ Figures 2.1, 2.2 & 2.3 around here $>$

2.3.3 Information combination vs. forecast combination

Our results so far indicate that commodity prices incorporate predictive ability in forecasting exchange rates movements. Although the individual predictive ability of commodity lagged prices may not result in more accurate forecasts than those of the benchmark, one could integrate the predictive ability of different commodities. There are two directions in the literature to do so: forecast combination (henceforth FC) and information combination (henceforth IC). Engle et al. (1984) argue for the superiority of IC. They propose that if two models are known, one should combine the information used by both models rather than the forecasts obtained from the separate models. Several studies support their findings (see Diebold, 1989; Diebold and Pauly, 1990; Granger, 1989; Hendry and Clements, 2004). On the other hand, a growing literature finds evidence for outperformance of forecast combination (see Giacomini and Komunjer, 2005; Stock and Watson, 2004).

In this section, we compare the out-of-sample forecasting performance of FC and IC in forecasting daily and monthly changes in the exchange rates. IC considers the whole information set as one large model as follows:

$$\Delta s_{t+1} = \alpha + \beta_i \Delta X_{i,t+1} + \epsilon_{t+1,i} \quad (2.11)$$

$$\Delta s_{t+1} = \alpha + \beta_i \Delta X_{i,t} + \epsilon_{t+1,i} \quad (2.12)$$

where $\Delta s_{t+1} = s_{t+1} - s_t$, α is constant, β is a vector of parameters, $X_{i,t+1}$ ($X_{i,t}$) is a $t \times n$ matrix of regressors available at time $t+1$ (t), and $t = 1, \ldots, T$. Similar to Models 2.1-2.2, Model 2.11 uses contemporaneous value of commodities, while Model 2.12 considers commodity lagged prices. FC combines forecasts obtained from simple models (Model 2.1-2.2) where each model incorporates a part of the entire information set. There are different approaches to combining forecasts generated by the different models. In the literature many papers find that an equally-weighted forecast combination often outperforms more complicated weighting methods (see

Stock and Watson, 2004).

We also perform principal component analysis (henceforth PC) to identify patterns in data and express the data in such a way as to highlight their similarities and differences. In other words, PCA is incorporated by estimating a predictive regression based on the principal component. The principal component regressions are given by

$$\Delta s_{t+1} = \alpha + \sum_{k=1}^{K} \beta_k \hat{F}_{k,t+1} + \epsilon_{t+1,i} \tag{2.13}$$

$$\Delta s_{t+1} = \alpha + \sum_{k=1}^{K} \beta_k \hat{F}_{k,t} + \epsilon_{t+1,i} \tag{2.14}$$

where $\hat{F}_t = (\hat{F}_{1,t}, \ldots, \hat{F}_{K,t})$, including the first K principal components. Following Neely et al. (2014), we choose the number of principal components included in a regression model, based on adjusted R^2.[8]

To evaluate the out-of-sample performance of these models, we perform the CW and SSPA tests. While the latter examines if any of the competing models can beat the random walk benchmark, the former compares the predictive ability of the models with each other. Contrary to our previous results, we find that both daily IC and FC models for lagged data outperform the random walk benchmark. This is a striking feature of our results since the bivariate models could not beat random walk (see 2.3.1), however when all the commodities are brought into one model or their forecasts are combined with equal weight, this results in more accurate forecasts than those of the benchmark at daily frequency.

< Table 2.7 around here >

The CW results show that the IC and FC models perform better than the PC for both realized and lagged data at both frequencies (see Table 2.7). At the daily frequency, while the IC model generates more accurate forecasts than FC, the IC and FC models both outperform PC. These findings hold for both contemporaneous and lagged data and all three currencies. At the monthly frequency, while the FC model

[8] Since there are eight principal components, we regress PC_1, $PC_{1,2}$, $PC_{1,2,3}$, and so on, on the exchange rate and choose the model with the minimum adjusted R^2.

beats the random walk for contemporaneous data in the case of AUD and CAD, the CW test results in equal predictive ability of IC and FC. When lagged data are used, the FC model outperforms the IC for the AUD and BRL.

2.4 Conclusion

We analyze the predictability of three commodity currencies. To do so, we use realized and lagged prices of commodities that have significant shares in the world commodity market. To evaluate the performance of eight competing models which consist of a commodity and a constant, we perform the Clark and West (2007) test. The results show that all the competing models outperform the random walk benchmark at the daily frequency when realized data are used. At the monthly frequency, contemporaneous models mostly beat the benchmark, but lagged models are outperformed by random walk. These results hold for all three currencies.

To check for robustness of the results and also to control for data snooping, different metrics are applied such as the SPA test by Hansen (2005) and the SSPA test by Hsu et al. (2010). The results show strong evidence for predictive ability of realized prices of six, four and three commodities in predicting daily changes in AUD, CAD and BRL, respectively. This sheds light on the importance of generating accurate forecasts for commodities which enables us to predict future exchange rates movements.

This chapter investigates whether the instability in the relationship between commodities and currencies proposed by Maier and DePratto (2008) impacts the performance of the models. In other words, this may cause the relative performance of the models to be time-varying, so that averaging this evolution over time will result in a loss of information. This motivates us to apply the fluctuation test by Giacomini and Rossi (2010) which evaluates the relative performance of the models locally. The results show that for lagged daily data, three models for each currency show predictive ability at least once over the selected period while this predictive ability is ignored when their performance is evaluated globally. This finding is our motivation for integrating the predictive ability of all commodities.

First, we construct a model that consists of all the commodities (IC). Second, we combine forecasts generated by individual models with equal weight (FC). Finally, we perform a principal component analysis (PC). The forecasts obtained from these approaches show that both IC and FC approaches outperform the random walk for lagged data at the daily frequency. This is the main contribution of this chapter and implies that while lagged prices of commodities are unable to beat random walk individually, integrating all variables into one model or equally-weighted combinations of their forecasts improves the accuracy of the forecasts, resulting in outperforming the benchmark. The results show that while the PC is outperformed by IC and FC for all currencies and frequencies, information combination performs better than forecast combination at the daily frequency. Our results for monthly data show that the contemporaneous FC outperforms IC in the case of AUD and BRL, but there is no evidence for beating the random walk.

Table 2.1 Descriptive statistics

This table summarizes statistics for daily (monthly) changes in the logarithm of exchange rates for the 27/02/1995-27/02/2013 period, with 4,698 (217) observations for three currencies: Australian dollar (AUD), Brazilian real (BRL) and Canadian dollar (CAD). The reported mean, standard deviation, minimum and maximum are reported in percent. Skewness is a measure of symmetry of the data. The last column represents the sample excess kurtosis relative to the normal distribution.

Currency	Mean	Std.	Min	Max	Skewness	Kurtosis
			Panel A. Daily			
AUD	-0.0069	0.8032	-6.70	8.83	0.75	15.94
BRL	0.0179	0.9703	-11.78	10.80	0.55	23.79
CAD	-0.0066	0.5436	-5.05	4.34	-0.09	9.76
			Panel B. Monthly			
AUD	-0.1425	0.0258	-8.48	22.26	2.85	28.07
BRL	0.3903	0.0525	-16.04	44.80	3.31	27.93
CAD	-0.1493	0.0395	-9.52	31.21	2.37	20.34

Table 2.2 Average share in the world export

This table presents the average share of commodity groups in the world export over the 1995-2013 period. These categories are defined by the UN Comtrade database. It is important to note that since the focus of this paper is on natural resource commodities, the remaining groups which contain manufactured commodities are ignored.

Commodity group	Average share
Group 3 including oil and gas	9.226%
Group 0 including coffee, sugar and wheat	6.346%
Group 2 including copper and aluminum	3.730%
Group 9 including gold	3.504%

Table 2.3 Average export share of Countries

This tables reports the average share of annual commodity export value of the countries included in our samples for the 1995-2013 period. Data is obtained from the UN Comtrade.

Country	Average share of export
Australia	1.17%
Brazil	1.13%
Canada	3.62%
US	10.62%

Table 2.4 CW-statistics for daily data

This table reports the results for the Clark and West (2007) test. The null hypothesis is equal predictive ability. If the p-value is less than 5%, there is evidence for predictive ability of the competing model against the driftless random walk benchmark at the daily frequency. Panel A reports results for realized data when commodity price at time t is used to forecast exchange rate at time t. Panel B reports results for lagged data where commodity price at time $t-1$ is used to predict exchange rate at time t.

	AUD		BRL		CAD	
Commodity	CW-stat	p-val	CW-stat	p-val	CW-stat	p-val
			Panel A. Contemporaneous			
Oil	9.8161	0.0000	9.8221	0.0000	10.1164	0.0000
Gas	4.3196	0.0000	2.1444	0.0160	3.9046	0.0000
Gold	11.9913	0.0000	4.6894	0.0000	9.3920	0.0000
Sugar	6.6055	0.0000	6.0845	0.0000	4.6932	0.0000
Coffee	1.9702	0.0244	3.1795	0.0007	2.6824	0.0037
Aluminum	13.0590	0.0000	12.2752	0.0000	13.1328	0.0000
Wheat	5.9850	0.0000	3.1764	0.0007	3.9567	0.0000
Copper	11.9978	0.0000	11.6718	0.0000	12.0689	0.0000
			Panel B. Lagged			
Oil	0.7758	0.2189	1.0001	0.1586	2.3368	0.0097
Gas	-0.8003	0.7882	0.2649	0.3955	-0.2664	0.6050
Gold	-1.1496	0.8749	0.5444	0.2931	-0.6101	0.7291
Sugar	1.7830	0.0373	1.1298	0.1293	0.6923	0.2444
Coffee	0.3830	0.3509	0.2502	0.4012	-0.2809	0.6106
Aluminum	2.7699	0.0028	0.7560	0.2248	0.1538	0.4389
Wheat	-0.5050	0.6932	1.4390	0.0751	0.5589	0.2881
Copper	2.8185	0.0024	0.1573	0.4375	0.9196	0.1789

Table 2.5 CW-statistics for monthly data

This table reports the results for the Clark and West (2007) test. The null hypothesis is equal predictive ability. If the p-value is less than 5%, there is evidence for predictive ability of the competing model against the driftless random walk benchmark at the monthly frequency. Panel A reports results for realized data when commodity price at time t is used to forecast exchange rate at time t. Panel B reports results for lagged data where commodity price at time $t-1$ is used to predict exchange rate at time t.

	AUD		BRL		CAD	
Commodity	CW-stat	p-val	CW-stat	p-val	CW-stat	p-val
			Panel A. Contemporaneous			
Oil	1.3041	0.0961	2.5943	0.0047	1.5497	0.0606
Gas	1.3488	0.0887	0.6690	0.2517	0.9999	0.1587
Gold	2.1373	0.0163	1.7700	0.0384	2.1661	0.0152
Sugar	1.9346	0.0265	2.4508	0.0071	0.8806	0.1893
Coffee	2.4840	0.0065	3.5639	0.0002	1.9471	0.0258
Aluminum	2.8418	0.0022	3.0255	0.0012	2.4352	0.0074
Wheat	2.2916	0.0110	1.2643	0.1031	2.2455	0.0124
Copper	2.2976	0.0108	3.0567	0.0011	2.1969	0.0140
			Panel B. Lagged			
Oil	-0.7498	0.7733	0.4769	0.3167	-0.4776	0.6835
Gas	1.1417	0.1268	0.1712	0.4320	0.3448	0.3651
Gold	1.3747	0.0846	-0.3530	0.6380	1.4811	0.0693
Sugar	-0.8432	0.8005	0.5652	0.2860	-0.1214	0.5483
Coffee	-0.8199	0.7939	0.5766	0.2821	0.1580	0.4372
Aluminum	-0.8252	0.7954	0.4770	0.3167	-0.5374	0.7045
Wheat	-0.8102	0.7911	0.1670	0.4337	-0.9080	0.8181
Copper	-0.8575	0.8044	0.1176	0.4532	-0.6505	0.7423

Table 2.6 SSPA results

This table reports the results of the SSPA test by Hsu et al. (2010) for contemporaneous models at the daily frequency. The null hypothesis is equal predictive ability. In the application of this test, we set the number of stationary bootstrap B = 1,000 and the parameter of the geometric distribution Q = 0.9, as in Sullivan et al. (1999) and Hsu and Kuan (2005).

Currency	Significant predictors				
AUD	Oil	Aluminum	Copper	Gold	Wheat
BRL	Oil	Aluminum	Copper		
CAD	Oil	Aluminum	Copper	Gold	

Table 2.7 CW results for the IC, FC and PC models

This table reports the Clark and West (2007) results for the information combination (IC), forecast combination (FC) and principal component (PC) models. Note that since the null hypotheiss is equal predictive ability, here we compare the forecasting performance of these three models in pairs.

Currency	Frequency	Model	The CW Test Results
AUD	Daily	Contemp.	The IC and FC models outperform the PC model. The IC model performs better than the FC model.
		Lagged	The IC and FC models outperform the PC model. The IC model performs better than the FC model.
	Monthly	Contemp.	The IC and FC models outperform the PC model. The null of equal predictive ability of the IC and FC cannot be rejected.
		Lagged	The IC and FC models outperform the PC model. The FC model performs better than the IC model.
BRL	Daily	Contemp.	The IC and FC models outperform the PC model. The IC model performs better than the FC model.
		Lagged	The IC and FC models outperform the PC model. The IC model performs better than the FC model.
	Monthly	Contemp.	The IC and FC models outperform the PC model. The null of equal predictive ability of the IC and FC cannot be rejected.
		Lagged	The IC and FC models outperform the PC model. The FC model performs better than the IC model.
CAD	Daily	Contemp.	The IC and FC models outperform the PC model. The IC model performs better than the FC model.
		Lagged	The IC and FC models outperform the PC model. The IC model performs better than the FC model.
	Monthly	Contemp.	The IC and FC models outperform the PC model. The null of equal predictive ability of the IC and FC cannot be rejected.
		Lagged	The IC and FC models outperform the PC model. The null of equal predictive ability of the IC and FC cannot be rejected.

Fig. 2.1 Fluctuation test results - AUD

Figure 2.1 displays the fluctuation test results for commodity lagged prices in the case of AUD at daily frequency. We choose 1,017 forecasts as the size of the rolling window to construct the test statistics. This is 30% of the number of observations in the out-of-sample period. The blue solid line represents the critical value and therefore when it is exceeded, it indicates significant predictive ability. Note that these results are ignored when the performance of the commodities are evaluated globally.

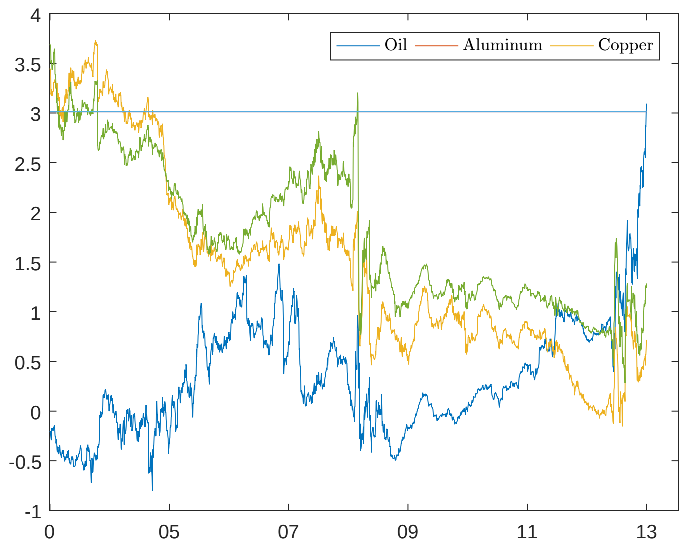


Fig. 2.2 Fluctuation test results - BRL

Figure 2.2 displays the fluctuation test results for commodity lagged prices in the case of BRL at daily frequency. We choose 1,017 forecasts as the size of the rolling window to construct the test statistics. This is 30% of the number of observations in the out-of-sample period. The blue solid line represents the critical value and therefore when it is exceeded, it indicates significant predictive ability. Note that these results are ignored when the performance of the commodities are evaluated globally.

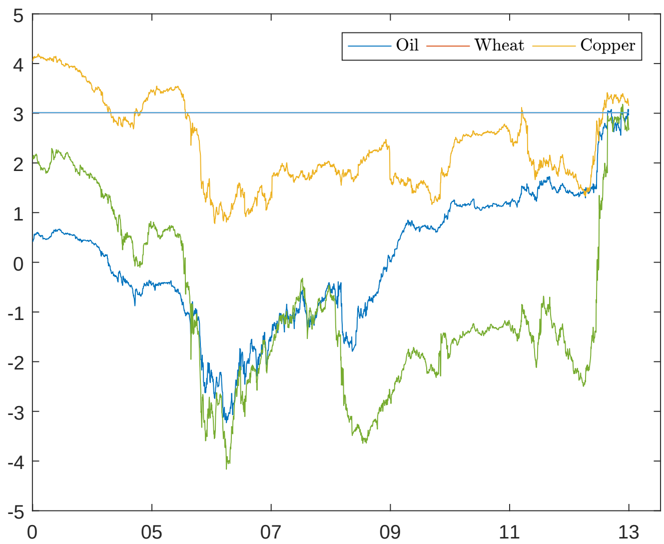
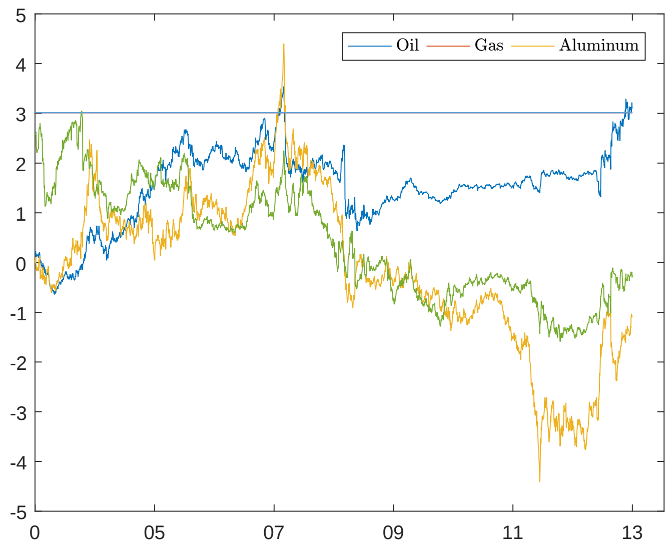


Fig. 2.3 Fluctuation test results - CAD

Figure 2.3 displays the fluctuation test results for commodity lagged prices in the case of CAD at daily frequency. We choose 1,017 forecasts as the size of the rolling window to construct the test statistics. This is 30% of the number of observations in the out-of-sample period. The blue solid line represents the critical value and therefore when it is exceeded, it indicates significant predictive ability. Note that these results are ignored when the performance of the commodities are evaluated globally.



Chapter 3

Are trading rules profitable in the FX?

The efficient market hypothesis (henceforth, EMH) assumes that asset prices fully reflect all available information and that new information will be instantly reflected in security prices. The weak form efficiency implies that technical trading rules (henceforth, TTRs) based on historical data should not be profitable. There is an ongoing debate about whether TTRs can consistently produce profit or just reflect luck. Menkhoff and Taylor (2007) show that technical analysis is very popular among currency market participants, and an extensive literature has developed on this topic. This literature is initially based on surveys of technicians. All these surveys, from primary one by Taylor and Allen (1992) to more recent ones by Cheung and Chinn (2001) and Cheung et al. (2005) highlight the extensive use of technical analysis in the foreign exchange (henceforth, FX) market.

This paper contributes to the literature in two respects. Several studies relate TTR profitability to data snooping bias. Therefore, the first contribution of this paper is to evaluates whether the profitability of TTRs is robust to data snooping. We use the false discovery rate (FDR) by Barras et al. (2010) to check the statistical robustness of our results. To the best of our knowledge, this paper is the first to employ the FDR methodology to control for data snooping in the FX market. Previous studies have employed a variety of methods to control for data snooping biases. For instance, Brock et al. (1992) claim that the deliberate choice of a simple class of rules that has been commonly used for a long period can allay the problem of data snooping. Qi and Wu (2006) and Hsu et al. (2010) use reality check and stepwise-SPA tests, respectively, to address this issue. The advantage of FDR over other methodologies is that it allows for a small (and specific) proportion of false discovery, while detecting almost all outperforming rules. We use the Sharpe ratio criterion to evaluate the long-run in-sample performance of TTRs and implement the FDR methodology. The results indicate that an important proportion of TTRs are profitable.

The second contribution of the paper is that, given the evidence on TTR in-sample predictive ability, it performs a comprehensive persistence analysis to test whether it would have been possible to select the outperforming rules for a month using data prior to that month. The motivation is that as noted by Bajgrowicz and Scaillet (2012), the economic value of the trading strategies can be assessed only by considering performance persistence, transaction costs, and data snooping. We are outperforming models selected at the end of each month in the presence of transaction costs and a portfolio is constructed. This is then adjusted on a monthly basis and used in the following month.Persistence tests are employed for daily U.S. dollar prices for six currencies and the out-of-sample performance of TTRs is evaluated based on Sharpe ratio. The results show that, for all exchange rates, not only does the FDR methodology provide a large number of outperforming trading rules, but also it generates profit for some periods. We find that Sharpe ratio fluctuates over time, implying that a positive Sharpe ratio is often offset by a negative one. The results show that, despite the short-lived profit opportunities, an investor could not consistently make profit by using previously outperforming TTRs, supporting Lo's (2004) Adaptive Market Hypothesis.

This paper proceeds as follows. Section 3.1 reviews the literature on profitability of TTRS. Section 4.1 presents our data and describes the universe of 7,650 technical trading rules. Section 5.1 reviews existing methods to account for data snooping and presents the FDR based approach.Section 3.4 presents the persistence analysis. Section 3.5 concludes.

3.1 Related Literature

Two approaches are extensively adopted by investors to guide their trading decisions. One approach is fundamental analysis which takes into account government policies, domestic and foreign events, political and economic news and corporates' annual reports to examine their effects on the supply and demand of a particular market. By studying fundamentals for a specific market, it is thus believed that one could predict changes in market condition. The second approach is technical analysis which uses past price behavior to guide trading decisions in asset markets.

The philosophy of technical analysis is explained in the seminal textbook of Murphy (1986) in which he points out three principles about the behavior of technical analysts. The first is that an asset's price history incorporates all relevant information which makes the use of asset fundamentals valueless. The second is that asset prices move in trends, implying predictability and also profitability by buying (selling) assets when the price is rising (falling). This is why technicians believe that 'the trend is your friend.' The third principle is that asset price patterns tend to repeat themselves. In other words, traders tend to react in a similar way when confronted by similar conditions.

Pring (1991) explains that the technical method is basically a reflection of the idea that price movements occur in trends. He explains that such trends are determined by the changing attitudes of traders toward a variety of economic, monetary, political and psychological variables. It is, thus, no surprise that an extensive literature has been developed on the performance of trading rules (see Menkhoff and Taylor (2007) for a review). However, regardless of the popularity of TTRs among practitioners, academics have long been skeptical about the advantages of technical analysis. They claim that the profitability of technical trading rules is contrary to the EMH which assumes that no trading rules should be able to generate profits on publicly available information without bearing any risk.

3.1.1 Technical analysis in stock markets

Technical analysis has been extensively employed in stock markets since Dow and Hamilton wrote a series of articles in The Wall Street Journal in the late nineteenth and early twentieth centuries. The early empirical literature finds that technical analysis as filter rules (see Alexander, 1961; Fama and Blume, 1966; Sweeney, 1986), relative strength rules (see Brush and Boles, 1983; Jacobs and Levy, 1988; Jensen and Benington, 1970), and moving average (see Dale and Workman, 1981; Van Horne and Parker, 1967) do not outperform a buy-and-hold strategy. However, more recent studies have shown that simple trading rules are useful for predicting stock market returns.

The seminal paper on predictive ability of trading rules in stock markets is by Brock et al. (1992) (henceforth, BLL) where moving average trading rules and trading range breakout rules are evaluated. After controlling for data snooping with the bootstrap methodology inspired by Efron (1992), Freedman and Peters (1984), and Efron and Tibshirani (1986), they find that these two simple rules generate significant predictions for the U.S. equity index returns. Bessembinder and Chan (1995) study the performance of trading rules in emerging Asia-Pacific stock markets. Their results demonstrate that the rules have significant predictive ability in some stock markets, but they have less power to predict stock price movements in others. In 1998, the authors extend the BLL study by taking into account transaction costs and non-synchronous trading for the U.S. equity market. They confirm BLL's findings on the predictability of trading rules, but accounting for transaction costs and non-synchronous eliminates their profitability.

Sullivan et al. (1999) question the validity of technical analysis due to the fact that the effects of data snooping are not fully accounted for. They perform the bootstrap reality check (henceforth, RC) by White (2000) to control for data snooping bias in their study. They expand the BLL study by considering 26 TTRs and 100 years of daily data on the Dow Jones Industrial Average (DJIA). These in-sample results support the BLL findings even after controlling for data snooping, but their out-of-sample results show that superior performance disappears when they choose the best performing rules with respect to the mean return criterion on a daily basis and using them for the following day. Ito (1999) tests the same TTRs as BLL for six Pacific-Basin countries and finds that the rules generate significant predictions for Indonesia, Japan, Canada, Mexico and Taiwan, but in contrast to BLL, he rejects the predictive power of trading rules for the U.S. index.

Skouras (2001) considers Dow Jones Industrial Average (DJIA) and applies moving average rules with different bands. His results show that time-varying estimated rules outperform the buy-and-hold strategy as well as various fixed moving average rules considered by BLL. Accounting for transaction costs, these findings stay significant only after transaction costs of less than 0.06%. In a similar study,Fang and Xu (2003) study the performance of TTRs in the case of DJIA over a hundred years and suggest that TTRs perform better than trading strategies based on time series models when the market is bullish.

Hsu and Kuan (2005) study a more complete universe of trading techniques compared to previous studies. They apply the RC by White (2000) and superior predictive ability (henceforth, SPA) test by Hansen (2005) to control for the data snooping bias. They study simple TTRs as well as complex ones (e.g. learning strategies and fractional position strategies), and find that the significant TTRs profitability can be found for NASDAQ Composite and Russell 2000 but not for DJIA and S&P-500. They take into account transaction costs and show that in-sample and out-of-sample outperformance of the best rules for NASDAQ Composite and Russell 2000 is robust to transaction costs.

In a recent study, Hsu et al. (2010) examine the performance of a universe of 16,380 TTRs in growth and emerging market indices by using a stepwise extension of the Hansen (2005) SPA test (henceforth stepwise-SPA test).[1] Their results support

[1]Their pool of models encompasses moving average and filter rules used in Brock et al. (1992) and Sullivan et al. (1999).

the predictive ability and profitability of TTRs in such markets. They explain that the predictive ability is due to market maturity. i.e. there are more arbitrage opportunities in younger markets than mature ones.[2] They explain that the profitability can be attributed to tail risk and market frictions.

Bajgrowicz and Scaillet (2012) use the false discovery rate as a new approach to account for data snooping and revisit the historical success of TTRs in previous studies. They evaluate the perfoamnce of 7,846 TTRs on daily prices of the DJIA over more than a hundred years. They demonstrate that their new approach is stronger than the existing methods such as stepwise-SPA test and is able to select more outperforming rules since it allows for a specific (small) proportion of false discoveries. They perform a persistence analysis of TTRs to find whether an investor could reasonably has predicted which rules would generate superior returns after taking into account transaction costs. There results show that the out-of-sample performance is negative throughout the recent period which implies that there is no hot hands phenomenon.

3.1.2 Technical analysis in the foreign exchange market

Menkhoff and Taylor (2007) explain that the FX market differs from equity markets in some aspects. First, total turnover in the global FX market is much greater than the turnover of the largest stock exchanges. Second, as mentioned by Sager and Taylor (2006), due to existence of professional traders in FX markets, the impact of individual investors may be neglected without loss of generality. Third, as highlighted by Lyons (2001), the FX markets have a much higher share of short term interdealer trading. Finally, there is less confidence among traders in models of fair value in the FX market compared to equity markets (see Frankel and Rose, 1996; Taylor, 1994).

Neely et al. (2014) explain that technical analysis dates back at least to 1700 and although modern technical analysis was originally developed in the context of the stock market, after the beginning of floating exchange rates, foreign currency

[2]This explanation is related to Lo's (2004) adaptive market efficiency hypothesis.

traders have broadly employed this approach for trading. The studies by Goodman (1979), Frankel and Froot (1990), and Goodhart (1988) are the first to bring the broad use of technical analysis by FX professionals to the attention of academic researchers. The first survey on how and to what extent technical analysis is used in FX markets is by Taylor and Allen (1992). They show that almost all chief foreign exchange dealers in London use technical analysis to some extent and also tend to combine it with fundamental analysis. They find that at short horizons (less than a week) traders use technical analysis more frequently than fundamental analysis. The Cheung and Chinn (2001) survey finds that 30% of U.S. FX traders could be considered as technical traders due to the increasing use of technical analysis. The Gehrig and Menkhoff (2004) survey is carried out among FX dealers and fund managers located in Germany and reveals that more than 90% of the respondents use technical analysis. Using survey evidence from 400 North American foreign exchange traders, Oberlechner and Osler (2008) find that respondents underestimate uncertainty and overestimate their own skills. They believe that their results explain the high volatility of floating exchange rates, the profitability of trend following strategies and the apparent irrationality of exchange rate forecasts.

The extensive literature on the performance of TTRs in the FX market finds support for their profitability. Poole (1967), Dooley and Shafer (1984) find evidence for the profitability of filter rules over relatively short horizons. Using the post-Bretton Wood data, Cornell and Dietrich (1978) find filter and moving average rules to be profitable. Sweeney (1986) confirms previous findings on the effectiveness of filter rules on many dollar exchange rates while he considers both transaction costs and risk. One could question the findings of such studies due to several drawbacks. First, their results may be subject to the data snooping problem, since the early literature is mainly focused on popular rules such as filter or moving average. It is worth noting that although Jensen (1967) and Jensen and Benington (1970) point out the danger of data snooping in technical analysis, none of the early studies accept their argument. Second, the riskiness of TTRs is often ignored. In other words, TTR returns do not necessarily reject market efficiency since returns may reflect risk premiums. Finally, it is hard to interpret the results of early studies since the performance of the rules are often reported as averages across all the rules or all the assets.

Neely et al. (1997) assess the out-of-sample performance of TTRs. They use genetic programming techniques and find that not only out-of-sample excess returns of the selected rules are economically significant, but also these returns are not compensations for bearing systematic risk. They account for data snooping bias by using the data snooping technique which was first used by BLL, to support their findings on the profitability of technical analysis in FX market. They explain that while central bank intervention generates a source of speculative profit in the market, it is less likely to be the reason for their findings. Instead, they highlight the role of market efficiency as a plausible explanation for their results.

The Qi and Wu (2006) study is the first to use the RC in the FX market to control for data snooping bias. A large number of TTRs is applied to seven daily dollar exchange rates over the 1973-1998 period. They find that the profitability of all seven currencies is significant at the one percent level of significance even after taking into account both data snooping bias and transaction costs. However, to purge the effects of data snooping bias, they carry out an out-of-sample experiment by separating the full sample into two equal sized sub-samples. The best performing rules in the first sub-sample are evaluated using the second sub-sample data. They find that for each currency the significant in-sample outperformance becomes much less significant out-of-sample.

Neely and Weller (2011) analyse the intertemporal stability of excess returns to trading rules in the FX market obtained from out-of-sample tests on previously studied rules. Their results show that the excess returns of the 1970s and 1980s were genuine and not just due to data mining. They demonstrate that, although these profit opportunities disappeared by the early 1990s for well-known rules such as filter and moving average rules, less-studied rules have remained profitable. These findings are consistent with the adaptive market hypothesis by Lo (2004) in which TTRs performance is short-lived with prolonged periods of success and failure.

In a recent study, Hsu and Taylor (2013) carry out a large scale investigation of trading rules in the FX market. They use daily data over a maximum of forty years for thirty developed and emerging market currencies, and also over 21,000 TTRs. They apply the stepwise-SPA test to control for data snooping and find strong evidence for predictability of TTRs in the 1970s and 1980s in developed currencies and in the 2000s in emerging currencies. They show an equally weighted portfolio of TTRs yields an average compound return of 9% over the last thirty years in the case of emerging currencies. They explain that the predictability and profitability of TTRs are due to short lived not-fully-rational behavior and immaturity, instead of autocorrelation, risk permia and central bank intervention.

3.2 Technical trading rules

Technical analysis can be divided into two broad categories: charting and mechanical methods. The former is the older method, graphs the history of prices over a specific period and uses patterns to forecast future movement. Charting is subjective and requires the analyst to be able to find and interpret the patterns. Charting is beyond the scope of this paper and we focus on mechanical methods. Mechanical rules require the analyst to apply rules based on mathematical functions of past and present data.

This paper investigates ten types of trading rules among which there are both well-known and less-studied rules: filter, moving average, linearly weighted moving average, exponentially weighted moving average, moving average convergence-divergence, moving average oscillator, stochastic oscillator, channel breakout, relative strength indicator and bollinger bands. Qi and Wu (2006) point out that choosing too few rules is likely to cause biases in statistical inference due to data mining. On the other hand, Hansen (2003) points out that considering too many irrelevant rules can reduce test power. We study a universe of 7,650 rules which is similar to the number of trading rules used in both the Sullivan et al. (1999) Bajgrowicz and Scaillet (2012) papers.[3] We follow Qi and Wu (2006) in parameters for filter, moving average and channel breakout and for the other rules we select reasonable parameters that lie in the ranges used in the literature. Table 4.1 provides a summary of the parameters used in the calibration of trading rules.

Filter rules (FR)

A filter rule generates a buy signal if the exchange rate rises by $x\%$ or more from its most recent low. The investor borrows the dollar and uses the proceeds to buy the foreign currency. On the other hand, when the exchange rate falls by $y\%$ or more from a subsequent high, the investor shorts the foreign currency and uses the proceeds to purchase the dollar. We define the subsequent high and low as a local maximum and minimum, respectively.

Moving average (MA)

Moving average is a simple average of prices over the previous n days, including the current day. When the short MA of a foreign currency is above (below) the long MA by an amount larger than the band with $b\%$, the investor borrows (short sells) the dollar (foreign currency) to purchase the foreign currency (dollar).

Linearly weighted moving average (LWMA)

This is a type of moving average where a higher weighting is assigned to recent price data than in the case of the simple MA. The LWMA is calculated by multiplying each one of the prices within the selected series, by a particular weight which is determined by dividing the position of time periods selected by the sum of the number of time periods. It is worth noting that weights in LWMA decrease in an arithmetic

[3] Sullivan et al. (1999) study 7,846 trading rules. Bajgrowicz and Scaillet (2012) study the same trading rules and sample but employ a different bootstrap method.

progression. The signal producing process is the same as the simple MA, implying that the rule generates buy (sell) signal when the short MA is above (below) the long MA.

Exponentially weighted moving average (EWMA)

Similar to the LWMA, the EWMA allocates a higher weighting to recent price data than does the simple MA. However, the weights are calculated differently. We define α as the smoothing factor,

$$\alpha = \frac{2}{n+1} \tag{3.1}$$

where n is the number of observations. To calculate the EWMA at each point in time, we multiply the last price by α and sum it with the product of the last day EWMA and $1-\alpha$. Here, weights decrease exponentially and the signal producing process is the same as the simple MA, implying that the rule generates buy (sell) signal when short MA is above (below) the long MA.

Moving average convergence-divergence (MACD)

The MACD measures the difference between two EWMAs. The representation of the MACD includes another EWMA which acts as a trigger indicator, known as the signal line. In other words, MACD (l, s, m) is an indicator where the MACD series is the difference of a long (l) and short (s) EWMAs, and the signal line is an EWMA of the MACD series with parameter m,

$$MACD_t = (1-\lambda_l)\sum_{i=0}^{\infty}\lambda_l^i P_{t-i} - (1-\lambda_s)\sum_{i=0}^{\infty}\lambda_s^i P_{t-i} \tag{3.2}$$

$$Signal\,Line_t = (1-\lambda_m)\sum_{i=0}^{\infty}\lambda_m^i \delta_t \tag{3.3}$$

where $\lambda_l = 1-\frac{2}{l+1}$, $\lambda_s = 1-\frac{2}{s+1}$, and $\lambda_m = 1-\frac{2}{m+1}$. Signals are determined as follows,

$$\begin{aligned} z_t &= +1 && \text{if} \quad MACD_t > Signal\,Line_t \\ z_t &= -1 && \text{if} \quad MACD_t < Signal\,Line_t \\ z_t &= z_{t-1} && \text{otherwise.} \end{aligned}$$

And an investor

$$\begin{aligned} &\text{Buys} && \text{if} \quad z_t - z_{t-1} = +2 \\ &\text{Sells} && \text{if} \quad z_t - z_{t-1} = -2. \end{aligned}$$

Moving average oscillator (MAOS)

The moving average oscillator is an alternative of the MACD while they are both computed similarly. The MAOS is calculated as the difference between a short MA and a long MA,

$$MA_t = m^{-1} \sum_{i=t-m}^{t-1} P_{i+1} - n^{-1} \sum_{i=t-n}^{t-1} P_{i+1}. \tag{3.4}$$

Signals are determined as follows

$$\begin{aligned} z_t &= +1 && \text{if} \quad MA_t > 0 \\ z_t &= -1 && \text{if} \quad MA_t < 0 \\ z_t &= z_{t-1} && \text{otherwise,} \end{aligned}$$

and an investor

$$\begin{aligned} &\text{Buys} && \text{if} \quad z_t - z_{t-1} = +2 \\ &\text{Sells} && \text{if} \quad z_t - z_{t-1} = -2. \end{aligned}$$

Stochastic oscillator (STOS)

The stochastic oscillator is a momentum indicator which shows the location of the price relative to the high-low range over a specific period. The stochastic oscillator is calculated as

$$\%K_t = 100 * \frac{P_t - min(P_{it-m}^{t-1})}{max(P_{it=t-m}^{t-1}) - min(P_{ii=t-m}^{t-1})} \tag{3.5}$$

$$\%D = \frac{1}{n} \sum_{i=t-m}^{n} \%K_{t-i+1}. \tag{3.6}$$

Buy and sell signals are determined as

$$
\begin{array}{llll}
\text{Buy} & \text{if} & \%K_{t-1} < \%D_{t-1} & \text{and} \quad \%K_t > \%D_t \\
\text{Sell} & \text{if} & \%K_{t-1} > \%D_{t-1} & \text{and} \quad \%K_t < \%D_t
\end{array}
$$

Channel breakout (CHB)

A channel occurs when the high price over a specific period is within $x\%$ of the low price over the same period. Therefore, an investor buys (sells) when the price goes above (below) the channel by $b\%$.

Relative strength indicator (RSI)

The RSI compares the magnitude of recent gains to recent losses to determine overbought and oversold conditions of a currency. The RSI is calculated as

$$RSI = 100 - \frac{100}{1 + RS^*}, \tag{3.7}$$

where

$$RS = \frac{average\,of\,x\,days'\,up\,prices}{average\,of\,x\,days'\,down\,prices}. \tag{3.8}$$

The RSI ranges from 0 to 100%. If it is above 70%, the currency is considered as overbought and the RSI generates a sell signal. On the other hand, if the RSI is below 30%, the currency is considered as oversold and buy signal is determined.

Bollinger band (BOLL)

Bollinger band is a band which is two standard deviations away from a simple moving average. It is developed by John Bollinger in the 1980's. Since standard deviation is a measure of volatility, the more volatile the market is, the wider the band gets. An investor buys (sells) the FX when the price is above (below) the upper (lower) band.

< Table 3.1 around here >

3.3 Methodology

3.3.1 Data snooping methods

Neely and Weller (2011) point out that data snooping, data mining and publication bias are three related but distinct problems that could lead to false conclusions about the profitability of TTRs. Data snooping is defined as a situation in which researchers, either consciously or unconsciously, study rules that are already proven to be profitable. Data mining is when researchers consider many rules but only interpret the most successful ones. In other words, negative results are ignored, while positive results are reported. Publication bias is the tendency of journals to accept papers with positive results rather than negative results. Technicians normally back-test rules which implies that they tend to study rules which are already profitable on past data. The basis of Neely and Weller (2011) shows that it is almost impossible to avoid some data snooping as data sets and rules are limited. They highlight that data snooping is broadly known but not always addressed as a significant issue in the finance literature.

There are a large number of methods suggested by the literature to control for data snooping bias. A very simple method is the Bonferroni correction which ensures that the overall Type I error rate of α is maintained when performing m independent hypothesis tests simultaneously. It rejects any hypothesis with a p-value smaller than α/m which is a excessively conservative criterion. As an example, if one sets Type I error to 0.05 and performs 1000 hypothesis tests, a p-value smaller than $5 * 10^{-5}$ is required to reject a hypothesis. Perneger (1998) suggests that Bonferroni adjustments are at best, unnecessary and at worst, deleterious to sound statistical inference.

3.3.2 White's (2000) reality check and extensions

White's (2000) RC is the first method which exploits the dependence structure of the individual test statistics. The RC approach tests the null hypothesis that the best rule does not outperform the benchmark. One disadvantage of White's method is that once it identifies an outperforming rule, the procedure stops. In other words, when the null hypothesis is rejected, it concludes that there is at least one model which beats the benchmark, but it is not able to identify the complete subset of outperforming rules. This drawback is addressed by Romano and Wolf (2005) who suggest a stepwise multiple testing procedure to asymptotically control the family-wise error rate (henceforth, FWER) at a chosen level. Using a stepwise-RC method, further outperforming rules can be detected in subsequent steps which makes this method more powerful than the RC.[4]

Another drawback of the RC from which the stepwise-RC also suffers, is that when the hypotheses involve inequality constraints, both methods are conservative. This is because they are both based on the least favorable configuration (henceforth, LFC).[5] Hansen (2005) shows that a test based on the LFC is likely to lose power dramatically when many poor and irrelevant rules are included in the test. He modifies White's RC by employing a studentized test statistic to reduce the impact of poor and irrelative rules, and to consider a sample dependent null distribution. He proposes the SPA test which is no longer based on the LFC but it addresses only the question of whether the best rule beats the benchmark. In other words, once it finds an outperforming rule, the procedure stops.

Similar to the extension from the RC test to the stepwise-RC test, it is logical to extend the SPA to the stepwise-SPA test. Hsu et al. (2010) prove analytically and by simulations that the stepwise-SPA is more powerful than the stepwise-RC. The main drawback of their proposed method is that the stepwise-SPA test controls for the FWER which guards against any false positives. In practice, investors tend to identify as many outperforming rules as possible without including too many false positives to diversify. In other words, investors do not consider the signal of one TTR at a time but a combination of multiple strategies which implies that the FWER is

[4]When testing multiple hypotheses, each test has type I and type II errors. A measure to determine the overall error rate is FWER which is the probability of making one or more type I errors. Therefore, instead of controlling the probability of a type I error at a chosen level for each test, the FWER is controlled at this level.

[5]That is least favourable to the alternative hypothesis.

too conservative given it leads to many missed findings.

3.3.3 False discovery rate

Benjamini and Hochberg (1995) address the issue of FWER by proposing the false discovery rate which is a less conservative method. The idea of the FDR is that by allowing a specific proportion of false discoveries, it significantly improves the power of detecting the outperforming rules. Barras et al. (2010) extend the FDR by introducing the $FDR^{+/-}$, which estimates separately the proportion of false discoveries among trading rules that perform better or worse than the benchmark. For instance, an FDR^{+} of 20% implies that among the rules selected as outperforming, on average 20% do not deliver positive returns while an FDR^{+} of 100% indicates that none of the selected models generates positive performance. Table 3.2 shows possible outcomes of hypothesis testing. While the FWER is calculated as $FWER = P(V >= 1)$, the FDR can be interpreted as the expected proportion of Type I error among the rejected hypothesis, $FDR = E(\frac{V}{R}|R > 0)P(R > 0)$. For multiple hypothesis tests, this implies that the FDR is much less conservative compared to the FWER and results in a significant improvement in the power of test.

< Table 3.2 around here >

The FDR identifies the outperforming rules, even if the performance of the best rule is due to luck, contrary to the stepwise-RC and the stepwise-SPA. Therefore, the FDR detects almost all outperforming rules, while keeping the amount of false discovery at a chosen level which provides investors with a diversified pool of signals.[6] Benjamini and Hochberg (1995) assume that the multiple hypotheses are independent when they propose the FDR approach. Storey (2003), Storey and Tibshirani (2003), and Storey et al. (2004) show that when the number of tests is large, the FDR still holds under weak dependence of the p-values. Bajgrowicz and Scaillet (2012) define week dependence as any form of dependence whose effect becomes negligible as the

[6] The advantages of the FDR approach over the stepwise-RC are demonstrated by the simulations of Bajgrowicz and Scaillet's (2012) paper.

number of tests increases to infinity. In this paper, although TTRs are dependent in small groups, each family of rules acts independently of the others. For example, a five-day MA with a 0.001 band is highly correlated with a five-day MA with a 0.005 band, however, it is less likely to be correlated with a two hundred-day MA. Such type of dependence is called block dependence and meets the weak dependence conditions.

The FDR^+ is estimated as follows

$$FDR^+_{(\gamma)} = \frac{\hat{F}^+_{(\gamma)}}{\hat{R}^+_{(\gamma)}} \tag{3.9}$$

where $\hat{R}^+$ is an estimator of R^+ and denotes the number of trading rules selected as significantly outperforming rules. $\hat{F}^+$ is an estimator of F^+ and denotes the number of trading rules that do not generate genuine performance but have been selected erroneously. The FDR method allows one to estimate π_A^+ and π_A^- which are the proportions of positive and negative TTRs in the population, respectively. π_0 is the proportion of rules without abnormal performance and as explained below, is required for estimating $\hat{F}^+$

$$\hat{F}^+_{(\gamma)} = \frac{1}{2}\hat{\pi}_0 l\gamma \tag{3.10}$$

$$\hat{\pi}_0(\lambda) = \frac{\#(p_k > \lambda; k = 1, \ldots, l)}{l(1-\lambda)}, \tag{3.11}$$

where l is the number of TTRs in our sample, γ is a threshold which is applied to determine the null and alternative p-values, and λ is a tuning parameter.

The estimation procedure for the values of λ^* and γ^* is proposed by Storey (2003) and Storey et al. (2004). To determine the value of λ^*, one considers a range of values ($\lambda = 0.05, 0.1, \ldots, 0.95$) and calculates the $\hat{\pi}_0$ for each value. Among them, choose the minimum $\hat{\pi}_0$ and for each possible value of λ and implement 1,000 bootstrap replications of $\hat{\pi}_0(\lambda)$ by drawing with replacement from TTR p-values. We denote them $\hat{\pi}_0^b(\lambda)$ for $b = (1, \ldots, 000)$. The estimated mean squared error of $\hat{\pi}_0^b(\lambda)$

for each value of λ is computed as follows

$$M\hat{S}E_{\lambda} = \frac{1}{1000}\sum_{b=1}^{1000}[\hat{\pi}_0^b(\lambda) - \min\lambda\hat{\pi}_0(\lambda)]^2. \quad (3.12)$$

The λ is determined such that $\lambda^* = \arg\min_{\lambda} M\hat{S}E_{(\lambda)}$. It is , however, worth noting that Barras et al. (2010) find that λ^* is not overly sensitive to the choice of λ^*.

The value of γ^* is determined in the same way. Consider a range of γ^* values ($\gamma = 0.30, 0.35, \ldots, 0.50$) and compute the relative $\hat{\pi}_A^-(\lambda)$ which is the proportion of rules with negative abnormal performance using the following formula:

$$\hat{\pi}_A^- = \frac{T_{(\gamma)}^- - A_{(\gamma)}^-}{l} \quad (3.13)$$

where $T_{(\gamma)}^-$ denotes the number of alternative rules with negative performance and a p-value smaller than γ, and $A_{(\gamma)}^-$ denotes the number of alternative models with negative performance and a p-value greater than γ. Then find the value of γ which maximises $\hat{\pi}_A^-$ and perform 1,000 bootstrap replications of $\hat{\pi}_A^-$ for each possible value of γ which are denoted by $\hat{\pi}_A^{b-}(\gamma)$ for $b = 1, \ldots, 1000$. Calculate the estimated mean squared error of $\hat{\pi}_A^-(\gamma)$ for each value of γ as follows,

$$M\hat{S}E_{(}^{-}\gamma) = \frac{1}{1000}\sum_{b=1}^{1000}[\hat{\pi}_A^{b-}(\gamma) - \max_{\gamma}\hat{\pi}_A^-(\gamma)]^2. \quad (3.14)$$

Therefore, γ^- is calculated such that $\gamma^- = \arg\min_{\gamma} M\hat{S}E_{(\gamma)}^-$. The same data driven procedure is used to determine γ^+. Note that if $\min_{\gamma} M\hat{S}E_{(\gamma)}^- < \min_{\gamma} M\hat{S}E_{(\gamma)}^+$, set $\hat{\pi}_A^-(\gamma^*) = \hat{\pi}_A^-(\gamma^-)$ and to preserve the equality, set $\hat{\pi}_A^+(\gamma^*) = 1 - \hat{\pi}_0 - \hat{\pi}_A^-(\gamma^*)$. Otherwise, set $\hat{\pi}_A^+(\gamma^*) = \hat{\pi}_A^+(\gamma^*)$ and $\hat{\pi}_A^-(\gamma^*) = 1 - \hat{\pi}_0 - \hat{\pi}_A^+(\gamma^*)$. Barras et al. (2010) explain that although this method is entirely data driven, there is some flexibility in the choice of γ^*, as long as it is sufficiently high.

3.3.4 FDR portfolio

Daily data are downloaded from Datastream for the following currency pairs relative to the U.S. dollar: the British pound (GBP), the Canadian dollar (CAD), the Japanese yen (JPY), the Norwegian krone (NOK), the Swedish Krona (SEK), and the

Swiss fran (CHF). For interest rates, we collect the overnight Euro-currency interest rates from Datastream. Table 3.3 reports descriptive statistics of the daily returns.[7] The mean return rates show that on average the U.S. dollar depreciates against all exchange rates. The maximum daily depreciation of the U.S. dollar relative to these six currencies lies between 4.47% (GBP) and 6.6% (JPY). The standard deviation of daily changes lies between 0.0053 and 0.0072 which implies substantial daily volatility. The JPY has the smallest Sharpe ratio, while the corresponding number is much larger for the CHF. With the exception of the GBP and CHF, all distributions appear to be right skewed, while all daily return distributions have excess kurtosis compared to the normal distribution.

$<$ Table 3.3 around here $>$

The FDR level determines the balance between wrongly selected underperforming trading rules and leaving out truly outperforming ones. Barras et al. (2010) explain that if a low $F\hat{D}R^{+}$ target is chosen, only a small proportion of lucky rules are allowed in the portfolio. On the other hand, a high $F\hat{D}R^{+}$ target decreases the expected performance of the portfolio, however, it improves the diversification of the portfolio. Bajgrowicz and Scaillet (2012) set $F\hat{D}R^{+} = 10\%$ while they find that results are qualitatively stable for values ranging from 5% to 20%. In this study, $F\hat{D}R^{+}$ is set equal to 20%. This implies that eighty percent of the rules included in the portfolio produce genuine performance which results in a pool of multiple outperforming TTRs. The algorithm described in section 3.3.3 is used to determine the parameters to get an $F\hat{D}R^{+}$ as close as possible to the $F\hat{D}R^{+}$ target level.[8] The relative π_A^{+} and π_A^{-}, the proportion of rules with positive and negative performance respectively, are then estimated.

The set of 7,650 trading rules is applied at each point in time and, after pooling signals of the selected rules based on the predetermined $F\hat{D}R^{+}$ level with equal weight, the investor decides whether to buy or sell the currency. Following Neely

[7]Note that returns are calculated as changes in the natural logarithm of the daily exchange rates

[8]Our findings remain consistent when $F\hat{D}R^{+} = 10\%$.

and Weller (2011), assume that the investor either buys or sells in the market and there are no neutral signals. It is worth noting that since neutral signals are not considered, if pooling the signals results in equal number of buy and sell signals, the signal obtained on time $t-1$ is used for time t. Compute excess returns for the trading rules and assume that if a trading rule generates a long (short) signal, the investor converts the borrowed dollar (foreign currency) to foreign currency (dollar) at closing rate and earns the foreign (U.S.) overnight rate. The excess return is computed as

$$r_{k,t+1} = [lnS_{t+1} - lnS_t + ln(1 + i_t^*) - ln(1 + i_i)] * signal_k, \tag{3.15}$$

where i_t and i_t^* denote domestic and foreign interest rates, respectively. The Sharpe ratio is a risk adjustment criterion to check whether technical trading returns compensate investors for bearing total risk. It is used to measure performance of the TTRs and is calculated as

$$SR_k = \frac{\bar{r}_k}{\sigma_k}, \tag{3.16}$$

where $\bar{r}_k$ denotes the excess return of the rule k,

$$\bar{r}_k = \frac{1}{N}\sum_{t=L}^{T} r_{k,t+1} \tag{3.17}$$

and σ_k is the standard deviation of the excess return generated by k-th trading rule,

$$\sigma_k = \sqrt{(\frac{1}{N-1})\sum_{t=L}^{T}(r_{k,t+1} - \bar{r}_k)^2}. \tag{3.18}$$

Figure 3.1 shows the proportion of outperforming (π_A^+), null (π_0) and underperforming (π_A^-) rules. These estimates are long-term in-sample results and obtained in the absence of transaction costs. With the exception of the CAD, the FDR finds that an important proportion of the rules exhibit a significant predictive power. The proportion of outperforming rules is above 70% for the SEK, CHF and JPY, while the corresponding number is much smaller for the GBP and NOK. To highlight the advantage of the FDR over other methodologies, the number of outperforming rules selected by the FDR are compared with those of the stepwise-SPA test. We find that the stepwise-SPA test does not detect any outperforming rule for any of the currencies. This could be interpreted as the FWER controlling for not making a type I error. This would imply that a TTR is selected by the stepwise-SPA test only if its p-value is smaller than $(1-\alpha)^{7,650}$.

< Figure 3.1 around here >

To check for the robustness of our results, transaction costs are taken into account. Neely et al. (1997) show that the excess returns earned by trading rules are very sensitive to the level of transaction costs and to the liquidity of the markets. They assume one basis point and two basis points as two fixed values of one-way transaction cost. Neely and Weller (2011), on the other hand, use Bloomberg data on one-month forward bid-ask spreads as the basis for estimating transaction costs. Comparing their data with those on actual trader's screen, they conclude that actual spreads are roughly one third of the quoted spreads. In this paper, we use fixed 2.5 basis points one-way transaction costs following Chang and Osler (1999), Qi and Wu (2006) and Hsu and Taylor (2013). Figure 3.2 shows the results with transaction cost. Interestingly, there is only a negligible decline in the proportion of outperforming rules which indicates that our previous findings are robust to transaction costs, with the exception of NOK where the outperformance disappears after taking into account transaction costs.

< Figure 3.2 around here >

3.4 Persistence analysis

Persistence analysis is used to evaluate the out-of-sample performance of the outperforming rules. The motivation is that not only a historical outperformance is no indication that an investor could have chosen the future outperforming rules ex-ante, but also in practice investors change their strategies in an attempt to adapt to the changes in economic environment. The question is whether an investor could have predicted that which trading rule would generate genuine returns. We test whether outperforming rules could have been selected by an investor who has just access to the information that would have been readily available to her. Therefore, the universe of 7,650 TTRs is used for the selected currencies on a monthly basis and the outperforming rules over the month are selected based on the FDR and in the presence of transaction costs. Each currency's outperforming rules is used over the following month and exploit returns from different currencies are pooled with equal weight to measure the total out-of-sample performance of the outperforming rules. To the best of our knowledge, this is the first time this type of persistence analysis is performed on TTRs in the FX.

< Figure 3.3 around here >

Similar to the in-sample estimation, the $F\hat{D}R^{+}$ target is set to 20%.[9] Setting the $F\hat{D}R^{+}$ target to a higher level has two opposing effects on a portfolio. While the advantage of a high $F\hat{D}R^{+}$ target is that it reduces the portfolio expected future performance, a high $F\hat{D}R^{+}$ target increases diversification of the portfolio. The results show that the target of 20% allows for a small proportion of false discoveries while resulting in a well-diversified portfolio of trading rules.[10] Figure 3.3 presents the selected FDR level $F\hat{D}R^{+}$ of each exchange rate at the end of each month. The average $F\hat{D}R^{+}$ is also reported which shows that the average does not always match its target. Barras et al. (2010) explain that it can be interpreted as, for instance, an average of 56% for the GBP instead of the targeted 20% indicates that the proportion of outperforming rules in the population is too low to achieve a 20% FDR target. Table 3.4 also shows that for the GBP, NOK and CAD more than 20% of the achieved FDR are higher than 70% which implies an increase in the proportion of TTRs included in the portfolio since our selection becomes less restrictive.

Figure 3.4 demonstrates the number of outperforming trading rules for the cur-

[9] The main findings remain qualitatively similar for $F\hat{D}R^{+} = 10\%$.

[10] It is important to note that at each step optimal values for λ and γ are estimated as explained in Eq. 3.12 and Eq. 3.14.

rencies at the end of each month. With the exception of GBP, we observe TTRs outperformance for the 1995-2000 period for all currencies where the number of outperforming rules reaches its maximum in the case JPY. The performance of TTRs is similar across all currencies for the 2000-2005 period where an outperformance in a month is often followed by an underperformance, making the corresponding number volatile. The performance of technical analysis over the period of the recent financial crisis is different among the currencies. While there seems to be underperformance in the case of SEK and GBP, our findings suggest that the initial outperformance of trading rules for the JPY and NOK breaks down by mid-2008. During this period of high uncertainty, TTRs perform well for the CHF and CAD currency pairs. The outperformance of technical analysis disappears for the post-crisis period across all currencies, with the exception of SEK.

< Figure 3.4 around here >

We seek to find the best performing technical trading family for each currency. The proportion of outperforming rules is computed for ten groups of TTRs. Figures 3.5-3.10 show that CHB, EWMA, LWMA MA and MAOS are often selected as the best strategies throughout the sample. In other words, our evidence suggests that the performance of other trading groups is not robust to the bootstrap and transaction costs. The findings show that while MAOS rules appear as the best rule most often for the GBP, SEK and CHF, the best rules are more likely to belong to MA family for the CAD and NOK. In the case of the JPY, EWMA rules are chosen more than other groups.

< Figures 3.5-3.10 around here >

Figure 3.11 shows results of a study of the outperforming TTR portfolio turnover. The blue bars represent the proportion of TTRs that continues to outperform for three consecutive months. The red line represents the number of outperforming trading rules at each point in time (see Figure 3.4). Our findings show that while there

are periods in which all the best models remain outperforming over three months, the average proportion of the rules that remain in the portfolio lies between 25% to 36%, with the exception of the GBP for which the corresponding number is 19%. This highlights the importance of rebalancing our portfolio on a monthly basis. Over the recent crisis the outperformance of the TTRs becomes more short-lived in the case of the GBP and SEK, while there is an improvement in the performance of trading strategies for CAD, JPY and NOK. This shows that over this period of high uncertainty, TTRs are reliable and persistent indicators for these currencies.

< Figure 3.11 around here >

The performance of TTRs in the persistence test is evaluated based on Sharpe ratio. Figure 3.12 shows the annualized Sharpe ratio at the end of each period for six currencies. Rebalancing the portfolio on a monthly basis based on FDR results in positive Sharpe ratio on average, with the exception of the GBP. The results are in line with previous findings that the performance of TTRs is volatile throughout the sample and periods with a positive Sharpe ratio are followed by those with a negative Sharpe ratio. This may explain why the average Sharpe ratio is small across all currencies. It is worth noting that during the recent crisis, TTRs generate a positive Sharpe ratio for the SEK, NOK and CAD, while during the post-crisis period, we only find support for profitability of TTRs for SEK and JPY.

Figure 3.13 plots the annualized Sharpe ratio of a portfolio consisting of six currencies. In other words, we find that when we combine generated returns from rebalancing the portfolio of TTRs for each currency, the profitability remains very volatile. The findings highlight the outperformance of technical analysis during the crisis where an investor could have generated a Sharpe ratio of 0.0375 over the 01/2007- 12/2009 period. Although on average this generates a Sharpe ratio of 0.027, there is no persistence in profitability. In other words, there are opportunities for an investor to make profit, but they are short-lived. Therefore, the results are in line with the adaptive market hypothesis by Lo (2004) which modifies the efficient market hypothesis. The adaptive market hypothesis proposes that the forces that drive prices to their efficient levels are weaker and operate over a longer time horizon. According to Neely and Weller (2011), the adaptive market hypothesis can be regarded as the most plausible explanation of TTR profitability.

< Figures 3.12 & 3.13 around here >

3.5 Conclusion

We investigate the profitability of TTRs in the FX market. A universe of 7,650 trading rules which consists of well-known rules and less-studied rules is applied to six currencies: SEK, CHF, GBP, NOK, JPY and CAD. To control for data snooping bias, the FDR methodology is employed. The motivation is that the previously used methodologies to control for data snooping either can only detect one outperforming rule (the RC and SPA tests) or they are too conservatives (the stepwise-RC and stepwise-SPA test). By contrast, while the FDR allows for a small (and specific) proportion of false findings, it detects almost all possible outperfoming rules. Our long-term in-sample results find predictive ability for an important proportion of the rules. Taking into account transaction cost, we find our results to be robust to the choice of one-way transaction cost of 2.5 basis points with the exception of NOK.

Persistence analysis is used to address the question whether investors could have predicted outperforming rules ex-ante in the presence of transaction costs. The $F\hat{D}R^{+}$ target is set to 20% and construct a portfolio of outperforming rules every month and use them for the following month. Our results show that at each step, FDR detects a large number of outperforming rules throughout the sample. This highlights the advantage of FDR since investors tend to combine the signals of multiple strategies rather than considering the signal of one TTR. Our findings suggest that among ten trading rule families, CHB, EWMA, LWAMA, MA and MAOS are the most selected strategies by FDR. However, a study of the portfolio turnover shows that, on average, the proportion of the rules that remain in the portfolio after three rebalancings lies between 0.19 and 0.36. Comparing these findings with the initial in-sample results shows that an investor should update the portfolio frequently to adopt changes in the economy rather than sticking to a specific set of TTRs.

The performance measure used in this paper is Sharpe ratio which measure the average excess return per unit total risk. The annualized Sharpe ratios at the end of each month show that the performance of the technical strategies are fairly volatile where a period with a positive Sharpe ratio is often followed by a period with a negative Sharpe ratio. If an investor constructed a portfolio of the outperforming rules and updated it on a monthly basis, she would obtain a positive Sharpe ratio for all currencies but for the GBP. Although our findings indicate that there are profitable opportunities for TTRs, their performance fluctuates throughout the sample which shows that the performance is not persistent at the monthly horizon. These findings support Lo's (2004) adaptive market hypothesis.

Table 3.1 Technical trading rule parameters

This table summerizes technical trading rule parameters used in this paper.

	Parameters	Description	Value
FR	x	Band for buy signal	0.0005, 0.001, 0.005, 0.01, 0.05, 0.1
	y	Band for sell signal	0.0005, 0.001, 0.005, 0.01, 0.05
	c	Number of days a position is held during which all other signals are ignored	5, 10, 25, 50
	d	Number of days for the time delay filter	2, 3, 4, 5
MA LWMA EWMA	m	Short run moving average	2, 5, 10, 15, 20, 25, 30, 40, 50, 75 100, 125, 150, 200
	n	Long run moving average	5, 10, 15, 20, 25, 30, 40, 50, 75, 100 125, 150, 200, 250
	b	Fixed band multiplication value	0.0005, 0.001, 0.005, 0.01, 0.05, 0.1
	c	As previous	5, 10, 25, 50
	d	As previous	2, 3, 4, 5
MACD	m	Short run moving average	12, 15, 20
	n	Long run moving average	26, 30, 35
	l	Length of histogram	7, 9, 12, 15
MAOS	m	Short run moving average	2, 5, 10, 15, 20, 25, 30, 40, 50, 75 100, 125, 150, 200
	n	Long run moving average	5, 10, 15, 20, 25, 30, 40, 50, 75, 100 125, 150, 200, 250
	b	Fixed band multiplication value	0.0005, 0.001, 0.005, 0.01, 0.05, 0.1
	c	As previous	5, 10, 25, 50
	d	As previous	2, 3, 4, 5
STOS	m	Number of days used for minimum calculation	3, 5, 10, 15
	n	Number of days used for maximum calculation	10, 15, 20, 25, 30
	c	As previous	5, 10, 25, 50
	d	As previous	2, 3, 4, 5
CHB	e	Evaluation period	5, 10, 15, 20, 25, 50, 100, 150, 200, 250
	$b1$	Band for buy signals	0.0005, 0.001, 0.005, 0.01, 0.05, 0.1
	$b2$	Band for buy signals	10% to 90% of $b1$
	c	As previous	5, 10, 25, 50
	d	As previous	2, 3, 4, 5
RSI	e	Evaluation period	14, 25, 30, 50
	$Lowerband$	Long run moving average	10, 20, 30, 50
	$Upperband$	Fixed band multiplicative value	60, 70, 80
	c	As previous	5, 10, 25, 50
	d	As previous	2, 3, 4, 5
BOLL	e	Evaluation period	2,5,10,15,20,25,30,40,50,75,100,125 150,200,250
	$nstd$	Number of standard deviations	2, 3, 4
	b	Fixed band multiplicative value	0.0005, 0.001, 0.005, 0.01, 0.05, 0.1
	c	As previous	5, 10, 25, 50
	d	As previous	2, 3, 4, 5

Table 3.2 Decision making in hypothesis testing

This tables shows possible decisions made in hypothesis testing regarding to true state of nature. V is type I error and T is type II error. Therefore, $FEWR = P(V \geq 1)$ is the probability of at least one type I error, while $FDR^{+} = E(V/R|R > 0)$ is the rate that discoveries are false.

	H_0 accepted	H_0 rejected	Total
H_0 true	U	V	n_0
H_1 false	T	S	$n - n_0$
	$n - R$	R	n

Table 3.3 Descriptive statistics

This table summerizes statistics for daily changes in the logarithm of exchange rates for the 01/03/1994-31/12/2013 period, with 5,217 observations for six currencies. Exchange rate is defined as the U.S. dollar price of one unit foreign currency. The Sharpe ratios are annualized.

	GBP	CAD	JPY	NOK	SEK	CHF
Mean*100	0.0022	0.0041	0.0013	0.0041	0.0050	0.0098
Min	-0.0392	-0.0434	-0.0371	-0.0502	-0.0354	-0.0847
Max	0.0447	0.0505	0.0658	0.0646	0.0555	0.0545
Standard deviation	0.0055	0.0053	0.0069	0.0072	0.0072	0.0068
Sharpe ratio	0.0039	0.0078	0.0019	0.0057	0.0070	0.0143
Skewness	-0.0415	0.0966	0.4695	0.0223	0.1685	-0.1371
Kurtosis	7.3438	10.1228	8.0239	8.3122	6.5587	10.2294

Table 3.4 Achieved false discovery rate

This table categorized the achieved FDR at the end of each month. Periods with high achieved FDR show that the proportion of outperforming rules in the population is too low to achieve a 20% FDR target.

Currency	20%	20-50%	50-70%	>70%
SEK	0.50	0.34	0.04	0.12
CHF	0.34	0.55	0.02	0.09
GBP	0.27	0.27	0.05	0.41
NOK	0.34	0.28	0.05	0.33
JPY	0.41	0.38	0.04	0.17
CAD	0.25	0.44	0.02	0.28

Fig. 3.1 In-sample performance without TC

Figure 3.1 displays the proportion of in-sample outperforming (π_A^+), neutral (π_0) and underperforming (π_A^-) rules for the 01/03/1994-31/12/2013 period. Theses results are obtained in the absence of transaction costs. The optimal λ and γ are computed as explained in Section 2.3.

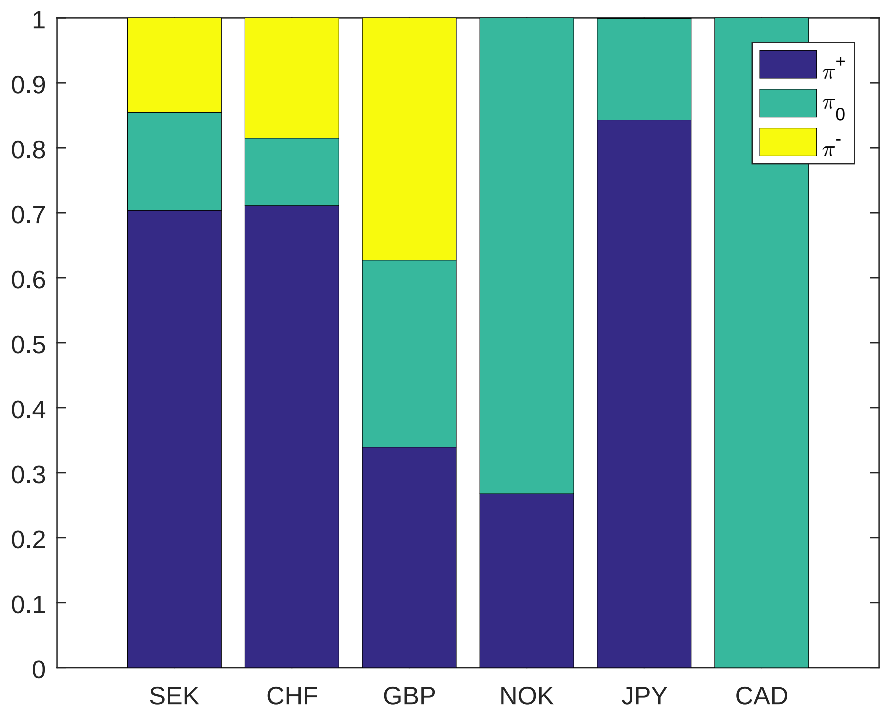


Fig. 3.2 Proportions of in-sample performance with TC

Figure 3.2 displays the proportion of in-sample outperforming (π_A^+), neutral (π_0) and underperforming (π_A^-) rules for the 01/03/1994-31/12/2013 period. Theses results are obtained in the presence of one-way 2.5 basis point transaction cost. The optimal λ and γ are computed as explained in Section 2.3.

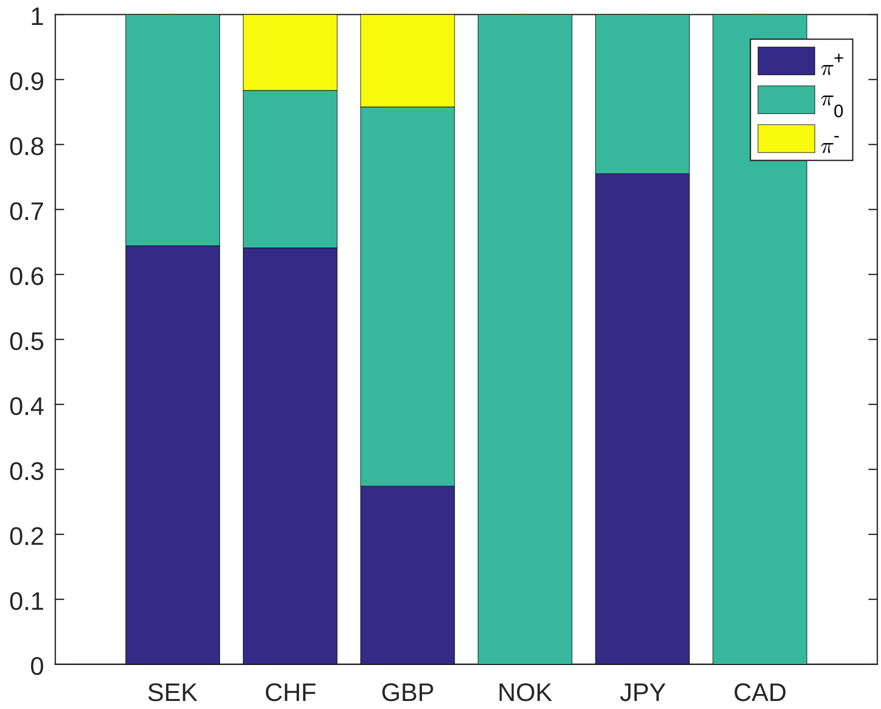


Fig. 3.3 In-sample performance with TC

Figure 3.3 displays the achieved FDR level at the end of each month throughout the sample. The average selected FDR is also reported for each currency. The target $F\hat{D}R^{+}$ is set to 20% which implies that 80% of the rules included in the portfolio generate genuine performance. If the selected FDR is higher than 20%, it shows that the proportion of outperforming rules in the population is too low to achieve the target.Theses results are obtained in the presence of one-way 2.5 basis point transaction cost.

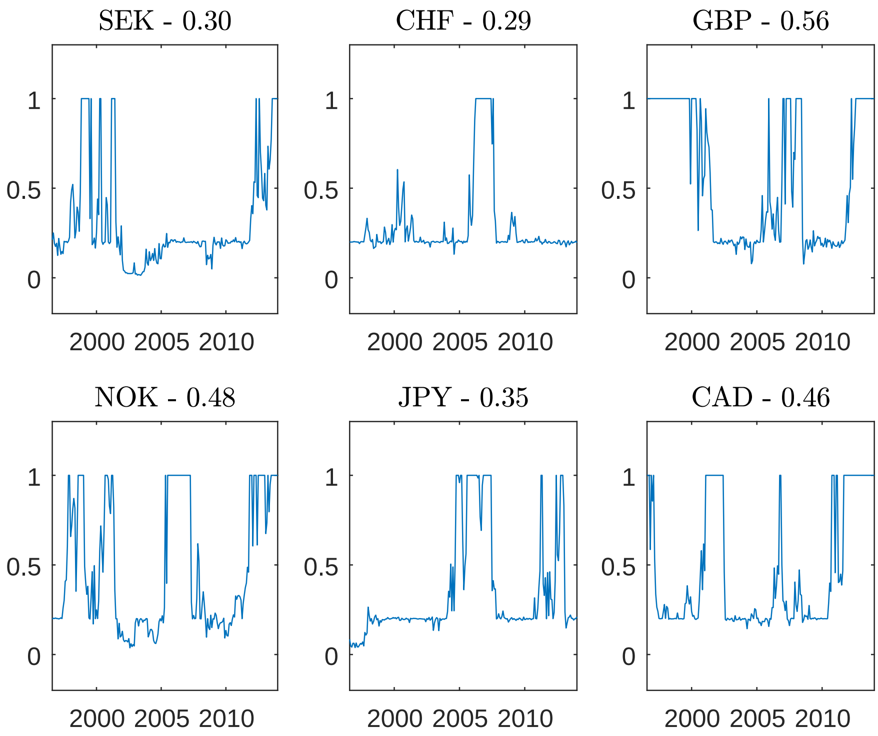


Fig. 3.4 Quantity of outperforming rules

Figure 3.4 displays the number of outperforming rules selected by FDR at the end of each month. The universe of 7,650 trading rules are evaluated on a monthly basis based on the Sharpe ratio criterion. It is important to note that these results are obtained while transaction cost is taken into account and data snooping bias is controlled for.

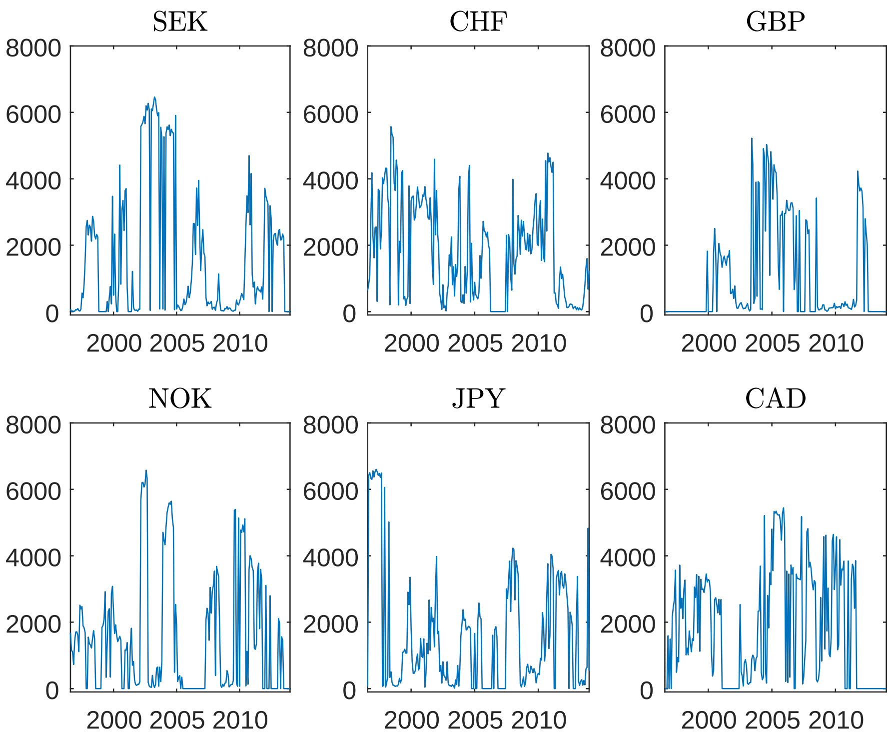


Fig. 3.5 Outperforming TTRs for SEK

Figure 3.5 displays the proportion of outperforming rules selected by FDR throughout the sample in the case of Swedish Krona (SEK) for ten trading strategy families: channel breakout (CHB), exponentially weighted moving average (EWMA), filter rule (FR), linearly weighted moving average (LWMA), moving average (MA), relative strength index (RSI), bollinger band (BOLL), stochastic oscillator (STOS), moving average oscillator (MAOS), moving average convergence-divergence (MACD).

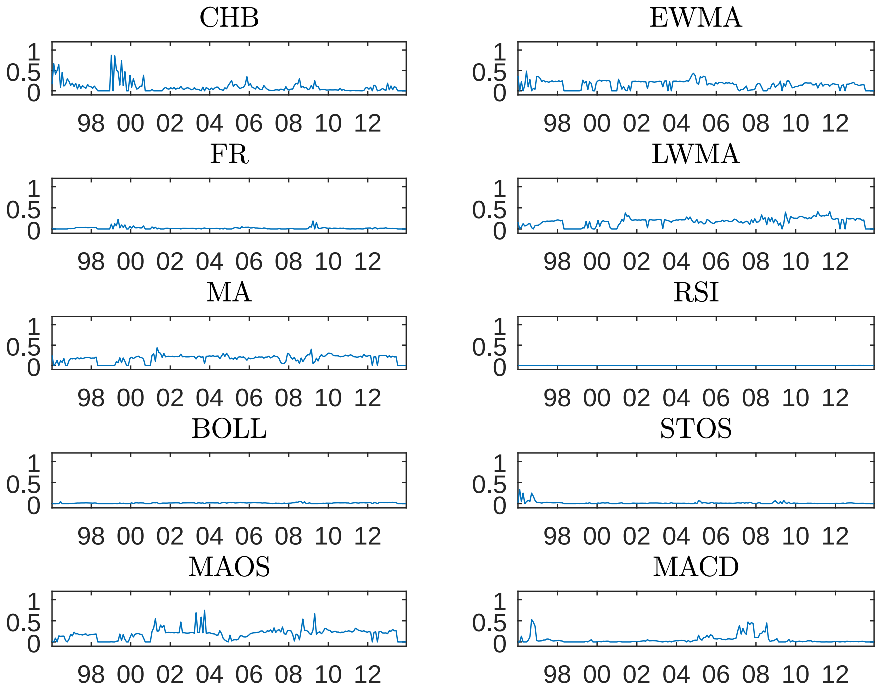


Fig. 3.6 Outperforming TTRs for CHF

Figure 3.6 displays the proportion of outperforming rules selected by FDR throughout the sample in the case of Swiss Franc (CHF) for ten trading strategy families: channel breakout (CHB), exponentially weighted moving average (EWMA), filter rule (FR), linearly weighted moving average (LWMA), moving average (MA), relative strength index (RSI), bollinger band (BOLL), stochastic oscillator (STOS), moving average oscillator (MAOS), moving average convergence-divergence (MACD).

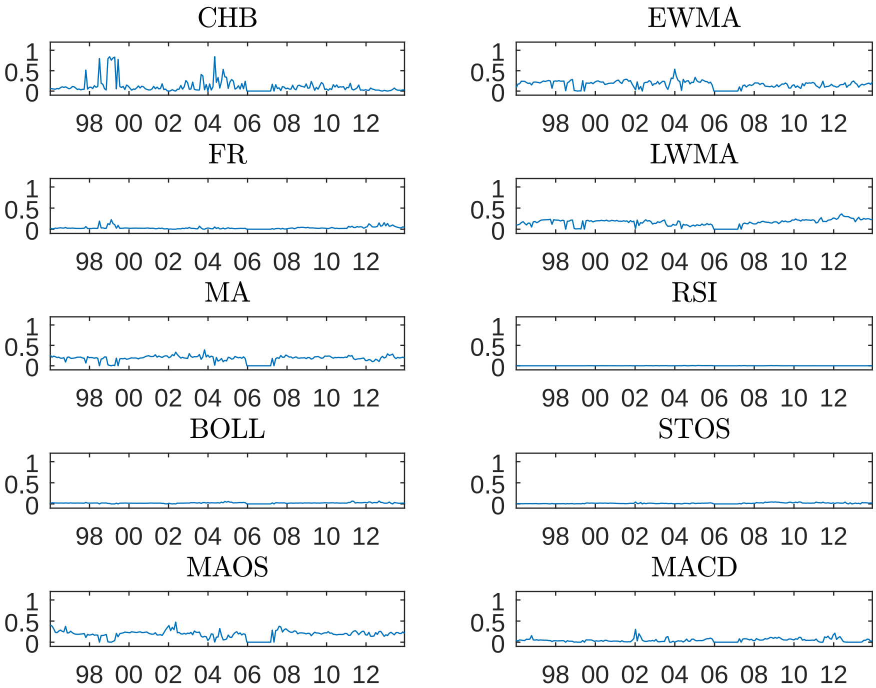


Fig. 3.7 Outperforming TTRs for GBP

Figure 3.7 displays the proportion of outperforming rules selected by FDR throughout the sample in the case of British Pound (GBP) for ten trading strategy families: channel breakout (CHB), exponentially weighted moving average (EWMA), filter rule (FR), linearly weighted moving average (LWMA), moving average (MA), relative strength index (RSI), bollinger band (BOLL), stochastic oscillator (STOS), moving average oscillator (MAOS), moving average convergence-divergence (MACD).

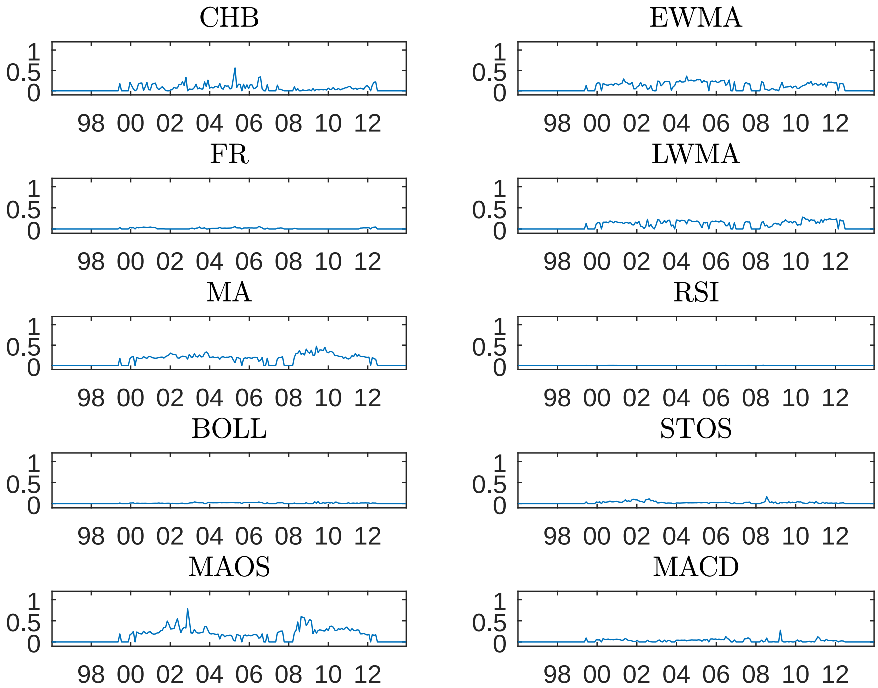


Fig. 3.8 Outperforming TTRs for NOK

Figure 3.8 displays the proportion of outperforming rules selected by FDR throughout the sample in the case of Norwegian Krone (NOK) for ten trading strategy families: channel breakout (CHB), exponentially weighted moving average (EWMA), filter rule (FR), linearly weighted moving average (LWMA), moving average (MA), relative strength index (RSI), bollinger band (BOLL), stochastic oscillator (STOS), moving average oscillator (MAOS), moving average convergence-divergence (MACD).

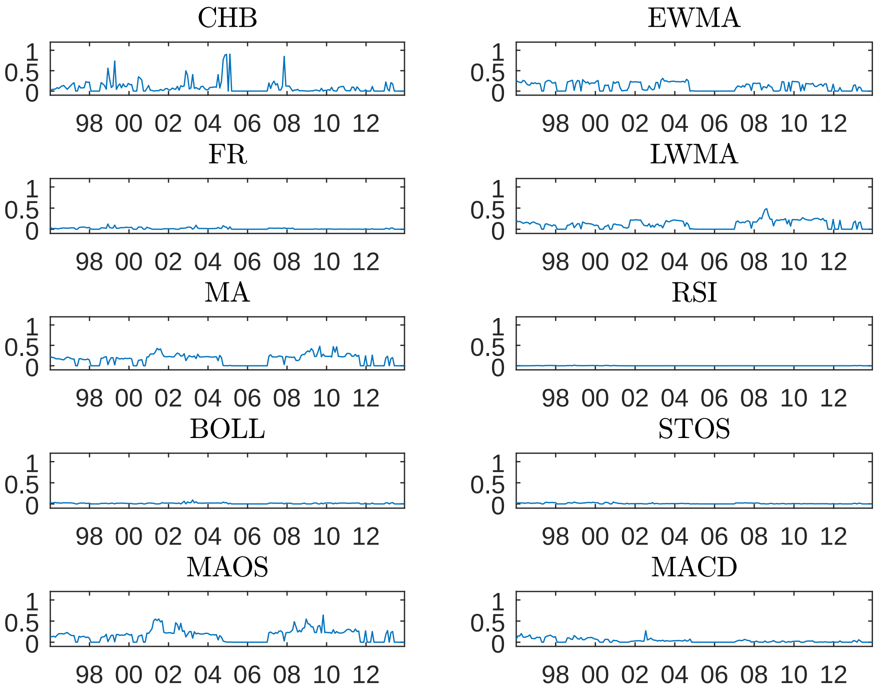


Fig. 3.9 Outperforming TTRs for JPY

Figure 3.9 displays the proportion of outperforming rules selected by FDR throughout the sample in the case of Japanese Yen (JPY) for ten trading strategy families: channel breakout (CHB), exponentially weighted moving average (EWMA), filter rule (FR), linearly weighted moving average (LWMA), moving average (MA), relative strength index (RSI), bollinger band (BOLL), stochastic oscillator (STOS), moving average oscillator (MAOS), moving average convergence-divergence (MACD).

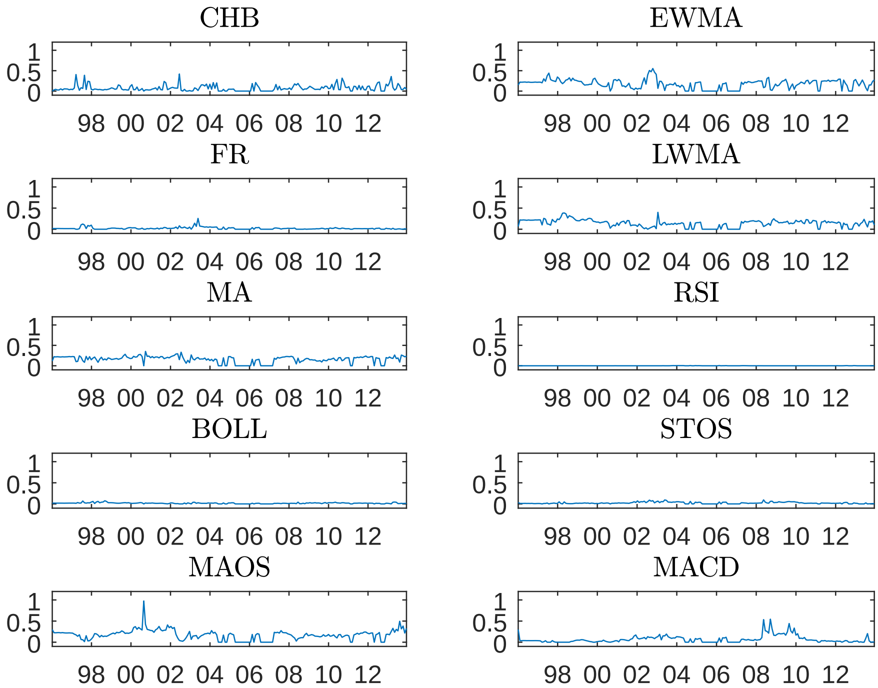


Fig. 3.10 Outperforming TTRs for CAD

Figure 3.10 displays the proportion of outperforming rules selected by FDR throughout the sample in the case of Canadian Dollar (CAD) for ten trading strategy families: channel breakout (CHB), exponentially weighted moving average (EWMA), filter rule (FR), linearly weighted moving average (LWMA), moving average (MA), relative strength index (RSI), bollinger band (BOLL), stochastic oscillator (STOS), moving average oscillator (MAOS), moving average convergence-divergence (MACD).

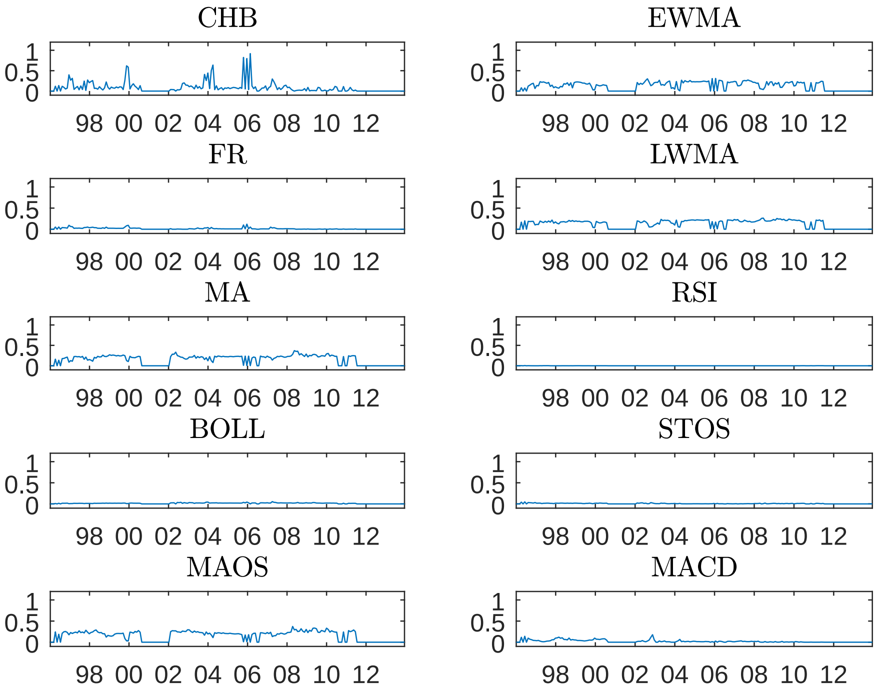


Fig. 3.11 Portfolio turnover

Figure 3.11 displays results of the portfolio turnover. The blue bars represent the proportion of rules remain in the portfolio after three rebalancings. The red line represents the number of outperforming rules selected by FDR. These results are obtained when one-way transaction costs of 2.5 basis points are taken into account.

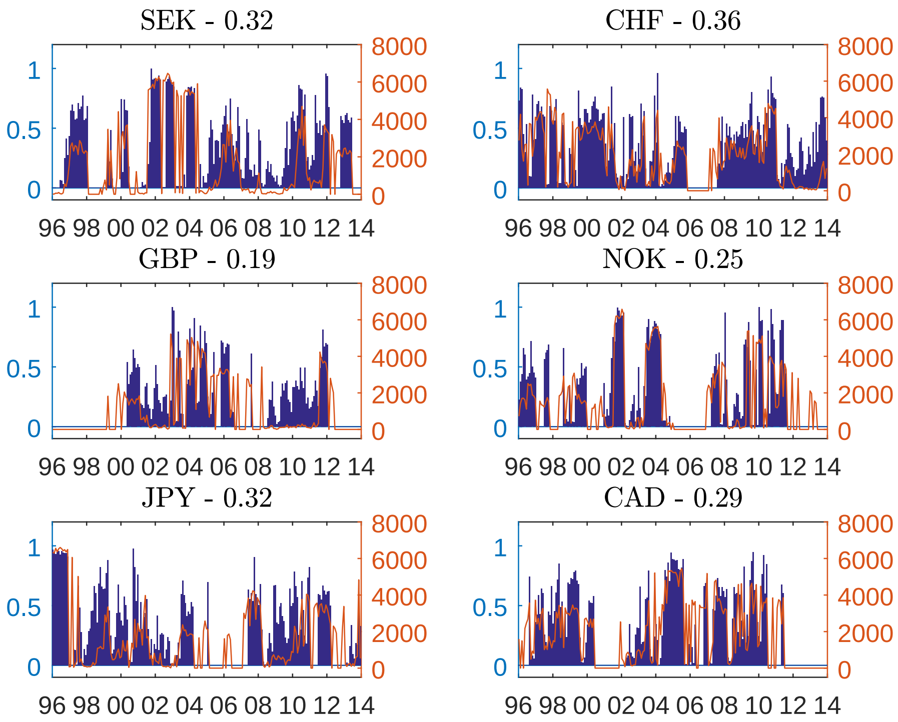


Fig. 3.12 Sharpe ratio of each currency

Figure 3.12 displays the Sharpe ratio criterion for each currency. The best performing models are selected at the end of each month and their performance is evaluated in the following month based on the Sharpe ratio criterion. The average Sharpe ratio throughout the sample is reported for each currency. The Sharpe ratios are annualized. It is important to note that one-way transaction costs of 2.5 basis point are taken into account.

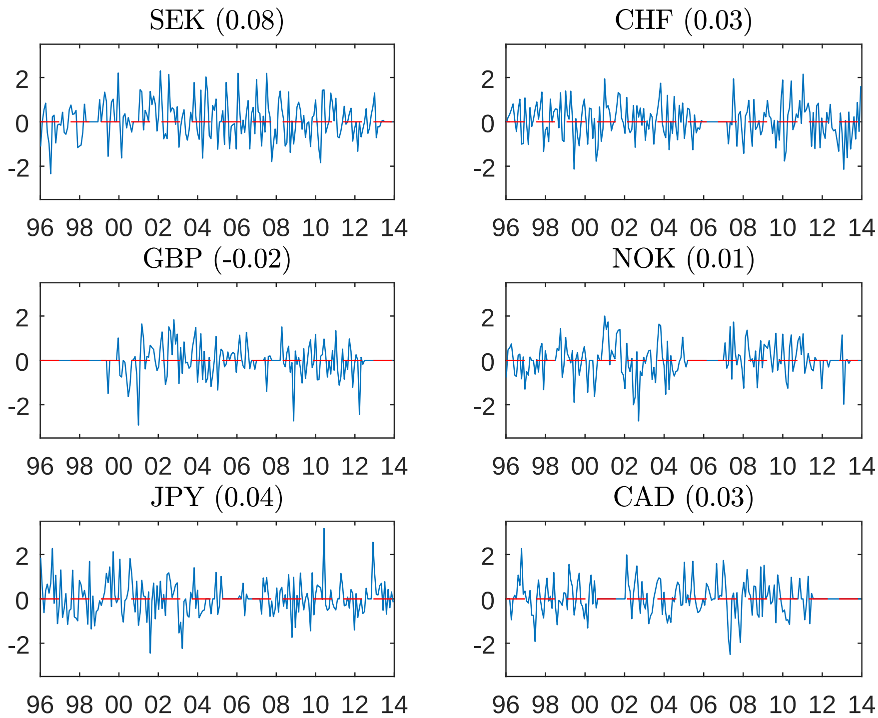
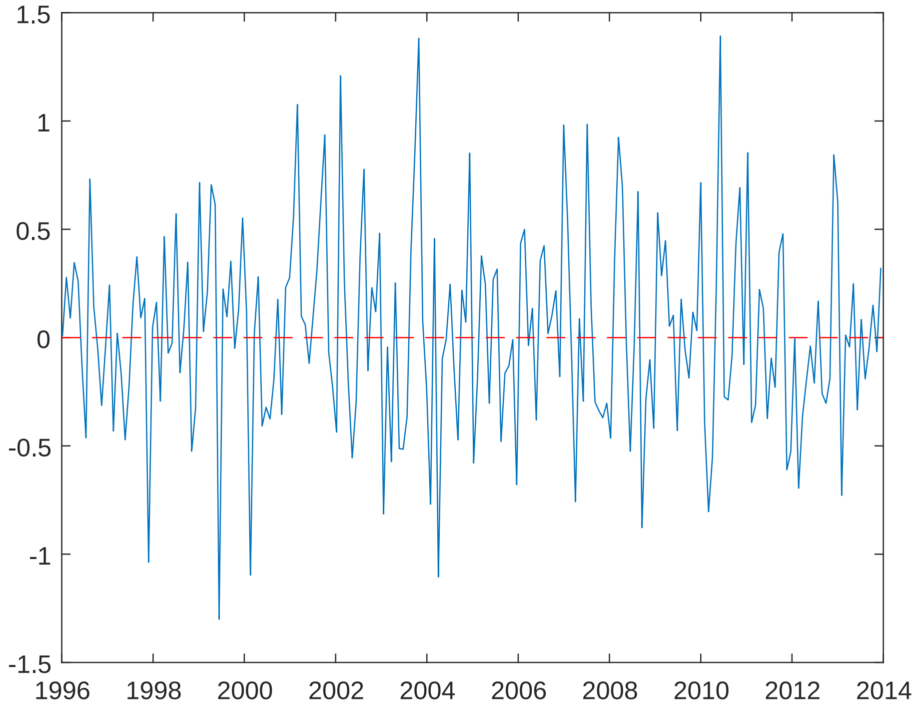


Fig. 3.13 Sharpe ratio of the portfolio

Figure 3.13 displays the Sharpe ratio criterion for a portfolio consisting of six currencies. The universe of 7,650 trading rule is used every month and the outperforming rules selected by FDR are used in the following month for the relevant currency. Their out-of-sample performance is evaluated by the Sharpe ratio. The Sharpe ratios are annualized. It is important to note that one-way transaction costs of 2.5 basis point are taken into account.



Chapter 4

The predictive ability of macroeconomic variables and technical indicators in the FX

The seminal paper by Meese and Rogoff (1983a) established that standard FX models could not outperform the random walk. This finding, known as Meese-Rogoff puzzle, is supported by many studies (see Berkowitz and Giorgianni, 2001; Chinn, 1991; Kilian, 1999), yet the recent literature has detected a series of economic fundamentals that seemingly have resolved the puzzle (see Clark and West, 2006; Molodtsova and Papell, 2012; Rime et al., 2010; Rossi, 2006). Rossi (2013) explains why the literature is controversial, suggesting that the answer to the question 'Are exchange rates predictable?' is, 'it depends'. She argues that the choice of predictor, forecast horizon, sample period, model, and forecast evaluation method are important factors that affect predictive ability.

A separate yet related literature discusses trading rules as alternative variables for forecasting exchange rates. Technical analysis has been extensively adopted by foreign exchange traders since the beginning of floating exchange rate in the early 1970s. Several surveys stress the broad use of technical analysis by practitioners (see Gehrig and Menkhoff, 2004; Lui and Mole, 1998; Menkhoff, 1997; Taylor and Allen, 1992). Menkhoff and Taylor (2007) conclude that amongst the different arguments to explain the performance of technical analysis in the foreign exchange market, the most plausible one is that it is able to inform traders about non-fundamental price determinants.

This chapter compares the predictive ability of a set of macroeconomic variables with that of a set of technical rules in forecasting exchange rates. It considers six macroeconomic variables and a constant to construct the full set of 126 economic models as well as 126 technical models each containing an individual technical indicator and a constant. To evaluate the out-of-sample performance of the models relative to the random walk benchmark, it employs the Theil (1971) mean square forecast error (henceforth, MSFE) decomposition, the out-of-sample R^2 suggested by Campbell and Thompson (2008), and the Clark and West (2007) test. The results show that while there are outperforming economic and technical models, there is more support in favor of the outperformance of economic models (89% of the economic models) in comparison to the technical models (30% of the technical models). Due to the drawbacks of the Clark and West (2007) test in case of nested models, the modified version of the White (2000) reality check by Clark and McCracken (2012) is also employed to check the robustness of our results.

The main contribution of this chapter is to find support for the USD/CAD exchange rate local predictability. We propose two approaches to show that an investor could choose models that generate more accurate forecasts than the random walk benchmark at the end of each month. Both methodologies are motivated by Timmermann (2008) and Rossi (2013) where the former reveals that predictive ability is often short-lived and a model is less likely to beat the benchmark for a long period, and the latter finds instabilities in models' forecasting performance over time. This sheds light on the importance of our choice to evaluate the performance of a pool of models over time, as there might be cases where outperformance of a model over a short period is followed by its underperformance over a long period, therefore looking at the performance of a single model by averaging across the sample period would ignore important local information about model performance.

The first methodology seeks to find the outperforming models at the end of each month using an information criteria approach. Three criteria are employed to assess the performance of the models at each point in time and compute their statistical significance. To control for the effects of data snooping, we run 2000 bootstraps at each point in time and therefore, select outperforming models with respect to the criteria and their relative bootstrapped p-values. Two approaches are considered to combining forecasts obtained from the outperforming models. One approach is to simply use equal weights while another approach is to use weights that are proportional to the inverse of the out-of-sample loss. We find that an investor could beat the random walk benchmark by choosing outperforming economic models according to their significant $\bar{R}^2$ and combining their forecasts with equal weight.

Following Rossi and Inoue (2012) critique regarding forecasting with a fixed window size we implement the dynamic model averaging by Raftery et al. (2010). This approach follows a recursive specification and allows the forecasting models to change over time and the coefficients of each model to evolve over time. To the best of our knowledge, this is the first study that employs this method for exchange rates. The results are in line with those of the former methodology, finding predictive ability for the economic models as well as the technical models. Our results for both economic and technical models show that the model with the highest posterior predictive probability is not constant throughout the sample which highlights the importance of considering a pool of models.

The remainder of the chapter is organised as follows. In Section 4.1 we briefly review the literature on predictive ability of economic and technical models in forecasting exchange rates. Section 4.2 describes the data set and discusses the framework used to construct pools of models. Section 4.3 presents the empirical results. Section 4.4 describes the dynamic model averaging and reports the relative results. Section 4.5 concludes.

4.1 Related literature

4.1.1 Meese-Rogoff puzzle

There has been an ongoing debate on the predictability of exchange rates following the seminal work of Meese and Rogoff (1983a,b) in which they claim that the random walk provides better forecasts of exchange rates than economic models. On one side of the debate, there are studies which find economic models with outperforming forecasts (see Clark and West, 2006; Engle, 1994; Molodtsova and Papell, 2012; Rime et al., 2010; Rossi, 2006; West et al., 1993). On the other, many studies support the Meese and Rogoff's (1983a,b) findings (see Bacchetta et al., 2010; Berkowitz and Giorgianni, 2001; Chinn, 1991; Kilian, 1999; Qi and Wu, 2003; Rossi and Sekhposyan, 2011).

Meese and Rogoff (1988) use uncovered interest rate parity (UIRP) to predict real exchange rates out-of-sample. The findings exploit the idea that real interest rate differentials do not generate better forecasts than the random walk. Similar studies are undertaken by Cheung et al. (2005) and Alquist and Chinn (2008) who find that UIRP forecasts for some countries at long horizons outperform those of the random walk, however, its performance is never significantly better. Clark and West (2006) and Molodtsova and Papell (2009) also study the predictive ability of UIRP. The latter find positive results for some countries and the former detect outperformance by UIRP at short horizons. Rossi (2013) states that the consensus on predictive ability of UIRP is a significantly negative slope estimate and a constant significantly different from zero. Several papers seek to find possible explanations for these findings; for instance estimation biases and imprecise standard errors are found by Bekaert and Hodrick (2001) and Baillie and Bollerslev (2000), respectively.

Cheung et al. (2005) study the forecasting ability of purchasing power parity (PPP) according to which the real price of comparable commodity baskets in two countries should be the same. They find that PPP outperforms the random walk at the longest horizons, however, its relative performance is significantly worse at shorter horizons. Rogoff (1996) demonstrates that deviations from PPP can be attributed to short-lived disturbances in the presence of nominal price stickiness, therefore, they should be transitory (i.e. 1-2 years). However, he finds that half-life deviations from PPP lie between three to five years which is known as Rogoff's PPP Puzzle. Rossi (2013) explains that there are variety of concerns in the case of PPP such as the underestimation of uncertainty around point estimates which is studied by Cheung and Lai (2000) and Lopez et al. (2003), and heterogeneity in disaggregating data which is considered by Imbs et al. (2005).

The empirical evidence on the monetary model of exchange rate determination by Frenkel (1976) and Mussa (1982) is mixed. MacDonald and Taylor (1993) and Mark and Sul (2001) find cointegration between exchange rates and monetary fundamentals. However, the out-of-sample performance of the monetary model is rejected by Meese and Rogoff (1983a,b) with this Chinn and Meese (1995), and Cheung et al. (2005) studies are in line. By contrast, Engel and West (2007) find forecastability for 5 out of 18 currencies one quarter ahead, outperforming the driftless random walk.

4.1.2 Parameter instability

Rossi (2006) studies parameter instability in term of the likelihood ratio test, Quandt (1960) likelihood ratio developed by Andrews (1993), the Andrews and Ploberger (1994) Exponential-Wald and Mean-Wald test, Diebold and Mariano (1995), and Clark and McCracken (2001) tests. She finds that for some countries the null hypothesis that the exchange rate is random walk can be rejected. She concludes that historical inability of economic models in outperforming the random walk might be due to the unstable relationship between fundamentals and exchange rates over time. In other words, she poses the question of whether the existence of parameter instability can explain the poor out-of-sample forecast of the economic models.

In a more recent paper, Bacchetta et al. (2010) evaluate whether parameter instability can explain the Meese-Rogoff puzzle. They explain that if parameters were constant and known the linear model would beat the random walk benchmark.

They conclude that in order to explain the Meese and Rogoff's (1983a,b) findings the assumptions about constant and known parameters should be relaxed. They explain that there are two offsetting effects for time-varying parameters when their reduced-form model is used. On one hand, it causes the weak out-of-sample performance of the model, and on the other hand, time variation in parameters increases the average explanatory power of fundamentals. The latter is due to the fact that sometimes parameters become high in absolute value, and therefore, fundamentals have more explanatory power. Nevertheless, they conclude that the Meese-Rogoff puzzle cannot be explained by time-varying parameters but rather by small-sample estimation bias.

Molodtsova and Papell (2009) study short-run out-of-sample predictability of Taylor Rules, interest rate, monetary and PPP fundamentals in forecasting exchange rate using rolling OLS regressions. They use the Clark and West (2006) statistic to evaluate the out-of-sample performance of exchange rate models and the superior predictive ability (SPA) test by Hansen (2005) to control for data snooping. They find significant forecastability for Taylor Rules fundamentals whose results improve when the model is symmetric with heterogeneous coefficients, smoothing and a constant. By contrast, models with interest rate, monetary and PPP fundamentals are less likely to outperform the random walk.

A growing number of papers in the 2000s reports short-term forecastability by using panel forecast methods, innovative estimation procedures, more powerful out of-sample test statistics and new structural models (see Engel et al., 2008; Gourinchas and Rey, 2007; Molodtsova and Papell, 2009). However, Rogoff and Stavrakeva (2008) question the robustness of the results of these influential papers. They state that one of the reasons of the optimistic results is the failure to check for robustness with respect to alternative out-of-sample test statistics. They apply the Clark and West (2006), and Clark and McCracken (2005) tests and find that in some cases these two test statistics are highly significant, while the Diebold and Mariano (1995) and West (1996) statistics are not. They interpret this as a sign of the presence of forecast bias.

They attribute the robustness of the results that support predictability of economic models to different forecast windows and find that even predictability of those models found to be robust to alternative out-of-sample test statistics disappears by changing the sample.

This has been addressed by Giacomini and Rossi (2010) where the entire time path of competing models' relative performance. They claim that previous approaches to forecast comparison are based on measure of global performance, whilst, they seek to test the stability of relative out-of-sample forecasting performance of monetary models over time by proposing the fluctuation test and the one-time reversal test. The former does not require specifying the nature of the instability under the alternative, however the alternative of the latter is of a single, permanent break in the relative performance of the two models. They find that the relative forecasting performance has changed over time and the British Pound and Deutsche Mark exchange rates were predictable in the late eighties which is in line with Molodtsova and Papell (2009) findings. However, this forecastability disappeared in more recent years.

Following Giacomini and Rossi (2010), Rossi and Inoue (2012) focus on the issue of robustness over different forecast windows. They demonstrate that the random window size used by researchers may result in insignificant predictability which is a concern as other window sizes would maybe find significant forecastability. Besides, considering many window sizes but only reporting the results for the successful ones raises another concern. Therefore, they consider the estimation window as a nuisance parameter in their paper and propose predictive ability tests which are robust to the choice of the estimation window size. They find that the predictive ability of economic models tends to appear at smaller window sizes. This implies that empirical evidence in favour of predictive ability may be due to the existence of instabilities in the predictive ability, for which rolling windows of small size are advantageous.

Sarno and Valente (2009) study short-term forecastability of quarterly economic fundamentals in predicting future exchange rate movements using real time data. They consider seven explanatory variables such as the once-lagged nominal exchange rate change, the deviation from a canonical monetary fundamentals model, the interest rate differential, net foreign assets, U.S. trade balance, the foreign country's trade balance and an intercept term. First, they apply the reality check by White (2000) to test whether any of the models constructed based on all possible combinations of the variables (overall 127 models), can generate better out-of-sample forecasts than those of the random walk. Finding significant evidence that the best model outperforms the random walk, they implement a set of different statistical and economic criteria to select ex ante the best model from a large pool of models available on a quarterly basis. They find that standard information criteria are unable to capture shifts in the parameters. They thus conclude that the exchange rate disconnect phenomenon is due to poor model selection criteria, and not because of lack of explanatory power in the fundamentals.

Due to inability of economic models to perform well consistently, Chinn and Moore (2011) introduce a model for exchange rates, which is a hybrid of the conventional specification with monetary fundamentals and the Evans-Lyons microstructure approach by Evans and Lyons (2002). They follow Evans and Lyons (2002) in which a combination of innovations in public and private information are determinants of changes in exchange rates. Evans and Lyons (2002) argue that private information is revealed through order flow. In other words, order flow is found to contemporaneously explain a significant proportion of the high-frequency variation in exchange rates. Chinn and Moore (2011) find that their hybrid model provides greater in-sample stability and out-of-sample forecasts compared to the basic macroeconomic and random walk models.

For an overview of predictors used in several papers in the literature, Rossi (2013) offers a critical survey of the literature on forecasting exchange rates in the last ten years. She establishes that both Taylor Rules fundamentals and net foreign asset positions have out-of-sample predictive ability. However the robustness of their predictors is questioned by some papers (see Alquist and Chinn, 2008; Rogoff and Stavrakeva, 2008). She also clarifies that the least successful models are non-linear specifications, while the most successful ones are linear specifications. Rossi (2013) explains that data transformations such as de-trending, filtering and seasonally adjusted data may explain differences in findings across studies. However, she claims that the frequency of the data and whether realized or forecasted fundamental is used, are less likely to affect predictability. She also points out the importance of the choice of the forecast evaluation method, where the benchmark should be the random walk without drift.

4.1.3 Technical analysis

Technical analysis generates forecasts on the basis of past prices, regardless of any underlying economic or fundamental analysis. Although technical analysis is not rooted in fundamentals, the broad use of technical analysis by practitioners in the foreign exchange market is well documented. An early survey conducted by Taylor and Allen (1992) reveals that at short horizons of one week or less, 90 percent of major leaders in the London foreign exchange market used some form of chartist input, with 60 percent confirming that they regarded such information as at least as important as economic fundamentals. These results are supported by findings of a survey by Menkhoff (1997). Later surveys include Lui and Mole (1998) and Cheung and Wong (2000) that both focus on Asian markets and a more recent one by Gehrig and Menkhoff (2004) considering foreign exchange dealers and fund managers in Germany. These surveys find that respondents considered technical analysis as an important tool which was more useful than fundamental analysis in forecasting trends, especially at horizons up to six months.

Menkhoff and Taylor (2007) offer a literature survey where they seek to explain the continued use of technical analysis. They report that the use of technical analysis cannot be seen as a not-fully-rational behaviour since all professionals in the market visually rely on this tool at least to a small degree. They also demonstrate that profitability of technical analysis can be related to foreign exchange interventions. The other explanation is that technical analysis is simply an instrument in the processing of market information, however, the main drawback here is that it does not explain the reason behind slow reaction to news. Nevertheless, they claim that the most satisfying explanation can be the ability of technical analysis in informing traders about non-fundamental price determinants. They conclude that academic researchers must understand technical analysis and integrate it into economic reasoning at both the macroeconomic and the microstructural levels.

The gap between macroeconomic variables mainly used in academic research and technical indicators extensively used by practitioners is filled by Neely et al. (2014) study. They compare the predictability of technical indicators in forecasting the equity risk premium to that of macroeconomic variables. Their results indicate that in-sample performance of technical indicators is as good as, or better than, macroeconomic variables. They implement the forecast encompassing test by Harvey et al. (1998), and find that considering information from both macroeconomic variables and technical indicators can result in more accurate forecasts. They employ different metrics to compare the relative forecasting performance of the models, for instance, CW test, out of sample R^2 (R^2_{OS}), and also Clark and McCracken (2012) reality check to control for data mining. Their findings point to statistically and economically significant out-of-sample forecasting ability of technical indicators, matching or exceeding that of macroeconomic variables.

To our best knowledge, such a study has not been conducted on the foreign exchange market. In other words, different studies in the literature of exchange rate forecasting consider a variety of macroeconomic variables or technical indicators separately, but none compares their performances, or utilizes information from both to improve forecasts. Therefore, our chapter fills this gap comparing the predictive ability of macroeconomic variables with that of technical indicators which is consistent with Menkhoff and Taylor (2007) results that technical analysis is able to capture information that is not reflected in macroeconomic variables.[1]

4.2 A pool of models

Chinn (2010) categorises the possible approaches to resolving the Meese-Rogoff puzzle into the following groups:

- Considering different functional form
- Considering regime switching
- Employing panel regressions
- Including additional variables

He explains that among these, the most popular approach is to include additional variables, which is the initial motivation for our construction of a pool of models.

4.2.1 Fundamentals

The main focus of this chapter is on the bilateral USD/CAD exchange rate since Canada has a sufficiently long history of a market-based floating exchange rates. We consider a U.S. investor who wishes to forecast the USD/CAD exchange rate by using a menu of economic variables and technical indicators from 1976:04 to 2014:12. We follow Sarno and Valente (2009) and assume that, at each point in time, the investor employs all possible combinations of all the fundamentals, in addition to a drift term, as possible OLS regressors. This generates a set of $2^n - 2$ different models, implying 126 economic models (ECON) by combining 7 different regressors.[2] Therefore, the simplest model is a univariate model containing a macro variable and a constant, and

[1] Section 3.1.2 reviews the use of technical analysis in the FX in more details.

[2] We follow Belsley et al. (1980) and use the variance inflation factor (VIF) to check for multi-collinearity between economic variables. We find weak multi-collinearity (<10%) which is consistent with the Gauss-Markov assumptions since there only requires that there is no severe multi-collinearity.

the richest model is a model with all n regressors, including the drift term.[3] Thus, our models can be expressed as follows:

$$\Delta s_{t+1} = \beta_i X_{i,t} + \epsilon_{t+1,i} \tag{4.1}$$

where $\Delta s_{t+1} = s_{t+1} - s_t$, β is a vector of parameters, $X_{t,i}$ is a $t \times n$ matrix of regressors available at time t, and $t = 1, \ldots, T$.

We obtain our data set from Rossi (2013) and update it which consists of monthly observations on overnight interest rates (i_t), producer price index (ppi_t), industrial production (y_t), consumer price index (cpi_t), the deviation from a canonical monetary model ($mf_t = (m_t - m_t^*) - (y_t - y_t^*)$) by Mark (1995), and spot exchange rates (s_t) for the U.S. dollar against Canadian dollar.[4] m_t, y_t, and cpi_t are seasonally adjusted. Seasonal adjustment allows us to avoid using information that was not available at the time the forecast was made.[5] We also take the log of all the variables, except overnight interest rates. We then construct the following regressors used in our models: interest rate differential ($i_{diff,t}$), producer price indices differential ($ppi_{diff,t}$), industrial production differential ($y_{diff,t}$), consumer price differential ($cpi_{diff,t}$), in addition to mf_t and s_t. For example, $i_{diff,t}$ is defined as the difference between the domestic (i_{US}) and foreign interest rate (i_{CAN}). To avoid the spurious regression problem in in-sample analysis, the predictors are differenced which results in a rejection of the Augmented Dickey Fuller (ADF) unit root test.

[3]This implies that our pool of models includes the UIRP, PPP, Taylor rule fundamentals, monetary model, and Balassa-Samuelson model.

[4]The average change in price of three indices such as finished goods, intermediate goods and crude commodities is measured by *ppi*. The money supply is denoted by m which includes physical money, demand deposits, checking accounts and negotiable order of withdrawal accounts. *ipi* tracks total output from the nation's factories, mines, and utilities. *cpi* estimates price changes in a basket of goods and services, reflecting consumption expenditure in an economy.

[5]Following Rossi (2013) we use a one-sided moving average seasonal adjustment filter instead of a two-sided one, as forecasts generated based on two-sided filtered data use information which would have not been available at the time of the forecast.

4.2.2 Technical trading rules

We consider a moving-average (MA) rule that is a trend-following technical indicator and generates buy or sell signals ($SIGNAL_{i,t} = 1$ or $SIGNAL_{i,t} = 0$, respectively) by comparing two moving averages at time t:

$$S_{i,t} = \left\{ \begin{array}{lll} 1 & if & MA_{S,t} \geq MA_{L,t} \\ 0 & if & MA_{S,t} < MA_{L,t} \end{array} \right\} \tag{4.2}$$

where

$$MA_{j,t} = (\frac{1}{j}) \sum_{i=0}^{j-1} P_{t-1} \quad for \quad j = S, L \tag{4.3}$$

P_t is the level of an exchange rate, and S (L) is the length of the short (long) MA ($S < L$). The MA rule seeks to find changes in exchange rate trends since the short MA is more sensitive to recent price changes than is the long MA. We consider monthly MA rules with $S = 3, 6, 9, 12, 18, 24, 36$ and $L = 6, 9, 12, 18, 24, 36, 48$.

The second technical strategy is based on momentum. A simple momentum rule generates the following signal:

$$S_{i,t} = \left\{ \begin{array}{lll} 1 & if & P_t \geq P_{t-M} \\ 0 & if & P_t < P_{t-M}. \end{array} \right\} \tag{4.4}$$

In other words, when a current price P_t is higher than its level M periods ago, it indicates positive momentum and relatively high expected excess returns which results in a buy signal. We compute monthly signals for $M = 6, 9, 12, 15, 18, 24, 36, 48$.

The third indicator is Moving Average Convergence/Divergence (MACD) which consists of the MACD line and signal line. The MACD line is the difference between the shorter Exponential Moving Average, EMA(S), and the longer EMA(L).[6] The signal line is the EMA(P) of the MACD line, where $L > S > P$. If the MACD line is above the signal line, it indicates a buy signal, and a sell signal if the MACD line is below the signal line. We consider monthly MACD with $L = 18, 24, 30$, $S = 9, 12, 15$ and $P = 3, 6, 9, 12$.

[6] An exponential moving average is similar to a simple moving average, except that more weight is allocated to the more recent data.

The forth indicator is the Moving Average of Oscillator (OsMA) which is defined as a modification of the MACD indicator. The OsMA reflects the difference between an oscillator (MACD) and its moving average (signal line). Therefore, when OsMA is positive, a buy signal is generated, and when it falls below zero, a sell signal is generated.

The fifth indicator is the Stochastic Oscillator (OsST) which is a momentum indicator whose dependency on market fluctuations can be decreased by using a moving average of the results as follows

$$K\% = \frac{C - MAX(N)}{MAX(N) - MIN(N)} * 100, \tag{4.5}$$

where C is the last close price and $MAX(n)$ ($MIN(N)$) is the maximum (minimum) price over the last N months. We define $D\%$ as the MA(M) of $K\%$. A buy signal is generated when the $K\%$ crosses over the $D\%$, and a sell signal is generated when the $K\%$ crosses under the $D\%$. We consider $N = 12, 15, 18, 24, 30, 48$ and $M = 3, 6, 9, 12, 15$.

The last technical indicator employed is Bollinger bands (BOL) which consist of a middle band, upper band and lower band where

$$Middle\ band = MA(N) \tag{4.6}$$

$$Upper\ band = Middle\ band + Standard\ Deviation \tag{4.7}$$

$$Lower\ band = Middle\ band - Standard\ Deviation. \tag{4.8}$$

When price is above the upper band, it implies a buy signal and when price is below the lower band, it generates a sell signal. We consider $N = 3, 6, 9, 12, 18, 24, 30, 36, 48$.

4.2.3 Forecasting models

Principal component analysis is employed to incorporate information from multiple macroeconomics variables. We follow convention in standardizing the individual predictors before computing the principal components. The predictive regression model is

$$\Delta s_{t+1} = \alpha + \sum_{k=1}^{K} \beta_k \hat{F}_{k,t}^{ECON} + \epsilon_{t+1} \tag{4.9}$$

This results in a more parsimonious model by filtering out much of the noise in the individual predictors. Eq.4.10 incorporates information from all of the technical indicators where $\hat{F}_{k,t}^{ECON}$ in (2) is replaced by $\hat{F}_{k,t}^{TECH}$

$$\Delta s_{t+1} = \alpha + \sum_{k=1}^{K} \beta_k \hat{F}_{k,t}^{TECH} + \epsilon_{t+1} \tag{4.10}$$

Similarly, we consider the entire set of macroeconomic variables and technical indicators by estimating the following model

$$\Delta s_{t+1} = \alpha + \sum_{k=1}^{K} \beta_k \hat{F}_{k,t}^{ALL} + \epsilon_{t+1} \tag{4.11}$$

Following Neely et al. (2014), we choose the number of principal components included in a regression model, based on adjusted R^2.[7]

4.3 Empirical results

4.3.1 In-sample estimation

In-sample evidence of predictive content is important and a necessary condition for out-of-sample forecastability (Inoue and Kilian (2006)). We estimate all the possible models constructed by combining different macroeconomic variables and different technical indicators using whole sample data. More specifically, we estimate heteroskedasticity-consistent t-statistics, $\bar{R}^2$ statistics and bootstrapped p-values of coefficients. The wild bootstrap procedure is implemented to calculate empirical p-values and control for heteroskedasticity and data snooping.

Following convention and in order to construct a pseudo-sample of observations for the exchange rate changes and the sets of macroeconomic variables and technical indicators, we assume that each macroeconomic variable follows an AR(1) process

[7]For instance, in the case of the PC for the macroeconomic variables, there are six principal components. We regress PC_1, $PC_{1,2}$, $PC_{1,2,3}$ and so on, on the exchange rate and choose the model with the minimum $\bar{R}^2$.

and each technical indicator follows a first-order, two-state Markov-switching process. Stambaugh (1999) points out that if the predictive variable of a model follows an AR(1) process and its residuals are correlated with the residuals of the predictive model, the OLS estimator of the predictive variable's coefficient will be biased in finite samples.[8] In order to address that, we calculate $\beta^* = \beta + (1+3\beta)/T + 3(1+3\beta)/T^2$, where β is estimated by OLS and T is the number of observations.

< Table 4.1 around here >

In order to save space, we do not report all the regression results, however, summarised results are reported in Table 4.1. We find that among economic variables, only Δs has significant in-sample predictive ability at the 5% significance level according to its bootstrapped p-value. However, four groups of technical indicators appear to have statistically significant coefficients in at least one technical model. The unreported $\bar{R}^2$ statistics of our models appear small at first glance. However, Campbell and Thompson (2008) show that the correct way to analyse the magnitude of $\bar{R}^2$ is to compare it with the squared Sharpe ratio, SR^2. If R^2 is large relative to SR^2, then the information in the predictive model can be used by an investor. As reported in Panel B of Table 4.1, we find that $SR^2 = 0.1\%$ which indicates that all the ECON and 92% of the technical models (TECH) have a larger R^2, implying that those models contain valuable information for an investor to use. All three PC models have a larger R^2 than SR^2.

Meese and Rogoff (1983a) show that a model's in-sample outperformance seldom translates to out-of-sample outperformance. Clark and McCracken (2005) also note

[8] More specifically, consider the following models

$$y_t = \alpha + \beta x_{t-1} + \nu_t$$
$$x_t = \gamma + \theta x_{t-1} + \varepsilon_t$$

where the errors (ν_t, ε_t) are serially independent and identically distributed with contemporaneous correlation which is $\Sigma = \left\{ \begin{array}{cc} \sigma_u^2 & \sigma_{\nu,\varepsilon} \\ \sigma_{\nu,\varepsilon} & \sigma_\varepsilon^2 \end{array} \right\}$. Stambaugh (1999) shows that if $\sigma_{\nu,\varepsilon} \neq 0$, the OLS estimator of β will be biased.

that in-sample predictability needs not necessarily indicate out-of-sample predictive ability. According to Rossi and Sekhposyan (2011), one of the possible reasons for this phenomenon is time-varying forecasting ability which might be due to changes in the model parameters. To test for parameter instability, we implement the Elliott and Muller (2006) test which has good size and power properties in the presence of heteroskedasticity. The null hypothesis is that all the coefficients (including the intercept) are constant. We compute $q\hat{L}L$ statistic, which is based on a quasi Local Level, and compare it to the critical values which are provided by Elliott and Muller (2006). There is strong evidence of structural instability for 57 ECON (45%) and 44 TECH (35%) at the 5% level. The results for ECON are tabulated in Table 4.2 with respect to number of predictors. Furthermore, none of the PC models is found to reject the null hypothesis of structural instability at the 5% level, implying more stability for the PCs.

< Table 4.2 around here >

4.3.2 Global out-of-sample forecasting

This section reports out-of-sample (henceforth, OOS) forecasting statistics for all the models defined in Section 2.[9] Following the literature (e.g. Clark and West, 2007; Molodtsova and Papell, 2009), 120 observations are used as the estimation period and the forecasts are calculated using a rolling specification. The forecast evaluation period runs from April 1986 to December 2014 and includes 344 forecasts. The out-of-sample forecast is given by

$$\Delta s_{t+1} = \hat{\beta}_{i,t} X_{i,t} \tag{4.12}$$

where $\hat{\beta}_{i,t}$ is the OLS estimate from regressing $\{\Delta s_k\}_{k=2}^{t}$ on $\{X_{i,k}\}_{k=1}^{t-1}$. Our benchmark model is the random walk without drift which is defined as the toughest benchmark by Rossi (2013).

[9] If a model exhibits no significant in-sample performance, its out-of-sample performance is more likely to be insignificant. However, as both global and local OOS performances are executed in this chapter, the entire set of models is considered.

To begin the forecast evaluation, we implement the Theil (1971) MSFE decomposition to check for bias-efficiency trade-offs in the forecasts. This decomposes the MSFE of a competing model against the random walk into the squared forecast bias and a remainder term which depends on the forecast volatility. Neely et al. (2014) state that limiting forecast volatility helps to reduce the remainder term.[10] Panel A of Table 4.3 shows that 84 of ECON and 24 of TECH have lower squared forecast bias than that of the benchmark, implying lower bias. On the other hand, the remainder term for 112 ECON and 44 TECH is less than that of the benchmark, indicating higher efficiency. Among the PC models, only the PC of economic models is less biased and more efficient than the benchmark, while the PC of technical models and the PC of all the variables are found to be as biased and efficient as the benchmark.

< Table 4.3 around here >

The Campbell and Thompson (2008) out-of-sample $R^2(R^2_{OS})$ measures the proportional decrease in mean squared forecast error (MSFE) for the predictive regression forecast compared to the random walk benchmark:

$$R^2_{OS} = 1 - \Big(\frac{\Sigma_{t=1}^{T}(\Delta s_t - \Delta\hat{s}_t)^2}{\Sigma_{t=1}^{T}\Delta s_t^2}\Big) \tag{4.13}$$

where $\Delta\hat{s}_t$ is the fitted value obtained from a predictive regression estimated through period $t-1$. We look for models with positive R^2_{OS}, implying that their out-of-sample performance is better than the benchmark. We find a positive R^2_{OS} for almost 90% of the ECON and 25% of the TECH (see Panel B of Table 4.3). In line with previous results, we find a positive R^2_{OS} for the PC of economic variables but not for the PC of technical indicators or the PC of all the variables.

The Clark and West (2007) test (henceforth, CW) is implemented to compare the predictive ability of our models to that of the benchmark. The CW test statistic is the MSFE-adjusted statistic which is the modified version of the Diebold and Mariano

[10]The Theil (1971) MSFE decomposition is given by $(\bar{\hat{r}} - \bar{r})^2 + (\sigma_{\hat{r}} - \rho\sigma_r)^2 + (1-\rho^2)\sigma_r^2$, where $\bar{r}$ $(\bar{\hat{r}})$ is the mean of the actual (forecasted) value, $\sigma_r(\sigma_{\hat{r}})$ is the standard deviation of the actual (forecasted) value, and ρ is the correlation coefficient between the actual and forecasted values. The remainder term is given by $(\sigma_{\hat{r}} - \rho\sigma_r)^2 + (1-\rho^2)\sigma_r^2$.

(1995) and West (1996) statistic. It is worth noting that the advantage of the CW over Diebold and Mariano (1995) and West (1996) is that it has an approximately standard normal asymptotic distribution when models are nested.[11]

Panel C in Table 4.3 shows that 112 of the ECON and 38 of TECH reject the nul of equal predictive ability at the 5%, implying that they produce more accurate forecasts than the benchmark. The CW results indicate that all three PCs outperform the random walk benchmark at the 5% level of significance. The difference between Panel B and C in terms of TECH is because the MSFE-adjusted statistic takes into account the negative expected difference between the benchmark MSFE and competing model MSFE under the null, so that the null can be rejected even if the R^2_{OS} is negative. Overall, all three panels of Table 4.3 find more support for outperformance of the economic models in comparison with the technical models. This could be interpreted as since technical indicators detect patterns in the data, they require a consistent and larger window of observations.

Rogoff and Stavrakeva (2008) demonstrate that one of the main drawbacks of using the CW test for nested models is that it cannot be always interpreted as the minimum MSFE test such as in Diebold and Mariano (1995) and West (1996).[12] They explain that it is more likely especially in the presence of scale bias where the CW tests whether the exchange rate is a random walk.[13] They suggest while using the asymptotic CW might seem appealing it is important that one checks the robustness of the results. Therefore, the Stepwise SPA test (henceforth, SSPA) by Hsu et al. (2010) is performed with 2000 bootstraps to test whether any of the models can produce more accurate forecasts than that of the random walk benchmark. The SSPA test is an extension of SPA test by Hansen (2005) which can identify models that violate the null of no abnormal performance, and the relative familywise error

[11]See Section 2.3.1 for more details on the Clark and West (2007) test.

[12]However, Clark and West (2007) suggest that the CW statistic should be interpreted as the minimum MSFE test statistic.

[13]Holden and Peel (1989) explain that scale bias occurs when a forecaster, on average, undershoots or overshoots the forecast by some percentage.

(FWE) rate can be asymptotically set at any pre-specified level.[14] The SSPA results contradict our previous findings, implying that none of the models could generate on average more accurate forecasts than those of the benchmark throughout the sample. In other words, we find no global predictability with the information embedded in a single economic model or trading rule unable to explain future movements in the exchange rate. However, this does not preclude these models performing well for part of the sample.

4.3.3 Local OOS performance

In order to test the difference between global and local predictability performance we implement two complementary tests which are proposed by Giacomini and White (2006) (henceforth, GW), and Giacomini and Rossi (2010) (henceforth, the fluctuation test). It is worth noting that these two tests directly account for the effect of parameter uncertainty. As the authors note the GW test evaluates the forecasting method rather than the forecasting model. The forecasting method contains the forecasting model as well as a number of parameters such as which estimation procedure to choose and what data to use for estimation. Not only can this framework be applied to both nested and non-nested models, but it also enables the researcher to take a conditional or unconditional perspective.[15] The GW test uses a rolling window method which not only affords significant generality but also is widely applied in the forecasting literature. The results indicate that the null hypothesis of equal average forecasting performance is rejected by 19 ECON, implying that there is strong evidence for global predictive ability of these models at the 5% significance level.

We apply the fluctuation test which measures the local relative forecasting performance of the models. This is useful for assessing whether macroeconomic variables

[14]FWE rate is defined as the probability of at least one Type I error. In other words, it is the probability of at least once erroneously rejecting the null.

[15]Giacomini and White (2006) demonstrate that on average the unconditional approach detects the more accurate forecast in the past. However, the conditional approach asks if one could use available information to predict which model will be more accurate for a specific future date and is thus the approach adopted here.

and technical indicators have predictive power at some specific points in time. This links to the Giacomini and Rossi (2010) view that, in the presence of structural instability, the relative performance of models may be time-varying, implying that averaging this evolution over time will result in a loss of information. For instance, a model may perform best on average over a specific period, however, the better local performance of the competing model may be ignored as for example, when considering only the recent past.[16] We find that 72 (57%) of ECON and 25 (20%) of TECH produce more accurate forecasts than the benchmark at least once over the selected period. The results show that 63% of the ECON contain Δppi_{diff} while OsST appears in 96% of the TECH. A graphical analysis of the local relative loss can be used to determine periods in which economic model is significantly more accurate (see Figure 4.1).

< Figure 4.1 around here >

The contrast between the results of the GW and the fluctuation tests sheds light on the importance of temporary predictability which may be ignored if one only looks at global forecasting performance. This is in line with Timmermann's (2008) findings which suggest that no single forecasting model consistently outperforms the benchmark, and thus, any predictability is, at best, short-lived and likely to disappear fairly quickly. He suggests that this is why forecasting approaches that stick to a single model tend to fail.

With this in mind, we employ a set of criteria to determine the best models at each point in time and use them to predict the value of the variable in the subsequent period. This approach is consistent with Giacomini and White (2006) where they explain that the performance of the models may be characterized by persistence, so that if a model performs well today, it may do so tomorrow.

We perform out-of-sample forecasting using our pool of models and apply three different criteria for model selection. The criteria employed in this study are the

[16] See Section 2.3.2 for more details on the fluctuation test.

adjusted coefficient of determination ($\bar{R}^2$), a sign criterion (SIGN), and the Mean Squared Error (MSE).[17] We follow Sarno and Valente (2009) and compute the set of criteria at the end of each month. Subsequently, models chosen in our model selection procedure are employed to forecast the variable of interest in the following month. This is similar to the procedure that an investor would be following in a real time. One of the contributions of our study is that we both compute these criteria and also determine their statistical significance. More specifically, we employ a wild fixed-regressor bootstrap with 2000 replications, similar to Clark and McCracken (2012), to estimate the relative p-value for each criterion.

Figure 4.2 displays the proportion of best economic models selected by each criterion.[18] $\bar{R}^2$ tends to pick more than 55% of the best models throughout our sample on average, however, between 2001 and 2003, there are four periods in which $\bar{R}^2$ is unable to pick any model. While the proportion of best models selected by MSFE is almost 36% on average, its performance throughout the sample is more volatile compared with the other two criteria. The SIGN criterion, on the other hand, fails to select almost any model for the periods 1997-2003 and 2005-2007, where the first period occurs just less two years after when the SIGN shows its best performance by selecting 34% of the best models in December 1995.

In the case of TECH, Figure 4.3 shows that, although the average proportion of best models selected by $\bar{R}^2$ is larger than other two criteria, its performance seems more volatile compared to Figure 4.2. As shown by Figure 4.3, there are periods in which $\bar{R}^2$ does not choose any model. The longest of such period is between 1997 and 2002 that is followed by a period in which 58% of the proportion of best models is on average selected by $\bar{R}^2$. The proportion of best models selected by the MSFE is highly fluctuating throughout our sample, specifically between 1995 and 2004 where usually fluctuates between 0 and 100%. The SIGN criterion seems to be able to select best models only for two periods of 1989-1997 and 2005-2014 and performs better in

[17] The SIGN criterion is based on directional accuracy of the in-sample prediction of the exchange rate change.

[18] Note that the best model is defined as the model with a significant bootstrapped p-value.

the former than the latter.

< Figures 4.2&4.3 around here >

Two approaches are implemented to combine the forecasts obtained from the outperforming models based on the criteria at each point in time. One approach is to take an equally weighted average, while the other approach uses weights which are proportional to the inverse of the out-of-sample loss. We implement the SSPA test by Hsu et al. (2010) where the results indicate that employing different criteria and different approaches to combine the forecasts does not result in beating the random walk benchmark in the case of TECH. When the equal-weighted approach is implemented in case of economic models, only the $\bar{R}^2$ results outperform the benchmark. However, when forecasts are combined with respect to their relative out-of-sample loss, the benchmark cannot be beaten. Our results are in line with the literature that equal-weighted forecasts perform better than more complicated approaches (see Stock and Watson (2001)).

< Figure 4.4 around here >

Panel A in Figure 4.4 displays the percentage of the models selected by both $\bar{R}^2$ and MSFE. Although MSFE is able to pick on average 42% of the models selected by $\bar{R}^2$, the relative performance is quite volatile throughout the sample. Comparing Panel B in Figure 4.4 with Figure 4.3, if the SIGN criterion detects the best models, its model selection profile is similar to that of the $\bar{R}^2$. The overall poor performance of this criterion is due to the fact that it is rarely able to detect the best models. If one applied our methodology and at each point in time and chose the best models according to the $\bar{R}^2$ and its relative bootstrapped p-value, those would have generated more accurate forecasts than those of the random walk benchmark.

< Figure 4.5 around here >

Figure 4.5 shows how often each of the competing models has been picked by $\bar{R}^2$. On average, each model appears in 191 periods among the best models. This supports the ability of our methodology to switch between different models and detect the best ones throughout the sample. The following models have been selected more than 80% of times by this criterion over our sample

$$\Delta s_t = c + \Delta i_{diff,t-1} + \Delta cpi_{diff,t-1} + \Delta mf_{t-1} + \Delta s_{t-1} \quad (4.14)$$

$$\Delta s_t = c + \Delta ppi_{diff,t-1} + \Delta i_{diff,t-1} + \Delta mf_{t-1} + \Delta s_{t-1} \quad (4.15)$$

$$\Delta s_t = c + \Delta ppi_{diff,t-1} + \Delta y_{diff,t-1} + \Delta cpi_{diff,t-1} + \Delta mf_{t-1} \quad (4.16)$$

$$\Delta s_t = c + \Delta y_{diff,t-1} + \Delta i_{diff,t-1} + \Delta cpi_{diff,t-1} + \Delta s_{t-1} \quad (4.17)$$

$$\Delta s_t = c + \Delta ppi_{diff,t-1} + \Delta i_{diff,t-1} + \Delta s_{t-1} \quad (4.18)$$

$$\Delta s_t = c + \Delta ppi_{diff,t-1} + \Delta y_{diff,t-1} + \Delta i_{diff,t-1} + \Delta cpi_{diff,t-1} + \Delta mf_{t-1} \quad (4.19)$$

$$\Delta s_t = c + \Delta ppi_{diff,t-1} + \Delta i_{diff,t-1} + \Delta cpi_{diff,t-1} + \Delta mf_{t-1} \quad (4.20)$$

$$\Delta s_t = c + \Delta ppi_{diff,t-1} + \Delta i_{diff,t-1} + \Delta cpi_{diff,t-1} + \Delta s_{t-1} \quad (4.21)$$

$$\Delta s_t = c + \Delta ppi_{diff,t-1} + \Delta cpi_{diff,t-1} + \Delta mf_{t-1} \quad (4.22)$$

$$\Delta s_t = c + \Delta cpi_{diff,t-1} + \Delta s_{t-1} \quad (4.23)$$

The above equations show significant predictive ability for all the macroeconomic variables among which Δcpi_{diff} appears in 80% of the above models, followed by Δi_{diff} and Δppi_{diff} that are in 70% of Models 4.14 - 4.23. We observe not only are there different variables in the above equations but also there are models with different number of predictors. It is interesting to note that there is no univariate nor bivariate models among the above models. This may indicate more support for predictive ability for larger models compared to parsimonious ones which could be interpreted as different variables are required to predict different dimensions of exchange rates changes. This finding demonstrates the importance of considering a pool of models which enables the criterion to explore and detect the best models among large number of possible models.

4.4 Dynamic model averaging

According to Rossi and Inoue (2012), our previous findings may be subject to chance as we have reported empirical results for one window size. They show that the results for a successful window size may have been achieved after many different window sizes, while the predictive ability of a model may have been ignored due to an ad-hoc window size. To check the robustness of our previous findings, we implement dynamic model averaging (DMA) method which does not require a specific window size. Bayesian model averaging (BMA) has received extensive attention in the econometrics literature. Wright (2008) demonstrates that the assumption of BMA is that, among many different models, one model is the true model. A researcher who does not know which model is the true model, starts from a prior about it and calculates the posterior probabilities that each model is the true model. The forecasts generated by the models are then weighted with respect to the posterior probabilities. Raftery et al. (2010) explains that in the presence of model uncertainty the BMA is a methodology for statistical inference of static datasets where there is uncertainty about variables to be included in the model. However, due to the fact that the BMA is restricted to static problems, DMA is proposed by Raftery et al. (2010). They show that the DMA methodology combines the BMA with hidden Markov models, and forgetting in state-space modeling.

In this chapter, we implement the DMA strategy since both the forecasting models and coefficients in each model can change over time. To explain the DMA, we consider our pool of models consisting of all macroeconomic and technical models, $M_t \in \{1, 2, \ldots, m\}$ where $m = 255$, and denote their predictors by z_{t-1} where the relative coefficients matrix is $\Psi = (\theta'^{(1)}_t, \ldots, \theta'^{(m)}_t)'$.[19] Therefore, we have

$$\begin{aligned} y_t &= z^{(m)}_{t-1}\theta^{(m)}_t + \varepsilon^{(m)}_t && (4.24) \\ \theta^{(m)}_{t+1} &= \theta^{(m)}_t + \nu^{(m)}_t && (4.25) \end{aligned}$$

[19]Our pool of models consists of 126 ECON, 126 TECH, and three principal components.

where

$$\varepsilon_t^{(m)} \sim N(0, H_t^{(m)}) \tag{4.26}$$

$$\nu_t^{(m)} \sim N(0, Q_t^{(m)}). \tag{4.27}$$

For given values of H_t and Q_t the variance of residuals in 4.24 and 4.25 repectively, standard filtering results are used to perform recursive forecasting. One of the elements of the DMA is Kalman filtering which is based on 4.28, 4.29 and 4.30

$$\Psi_{t-1}|M_{t-1} = m, y^{t-1} \sim N(\hat{\theta}_{t-1}^{(m)}, \Sigma_{t-1|t-1}^{(m)}) \tag{4.28}$$

$$\Psi_t|M_t = m, y^{t-1} \sim N(\hat{\theta}_{t-1}^{(m)}, \Sigma_{t|t-1}^{(m)}) \tag{4.29}$$

$$\Psi_t|M_t = m, y^t \sim N(\hat{\theta}_t^{(m)}, \Sigma_{t|t}^{(m)}). \tag{4.30}$$

To simplify notations we drop the m superscript. The Kalman filtering proceeds using

$$\Sigma_{t|t-1} = \frac{1}{\lambda}\Sigma_{t-1|t-1} \tag{4.31}$$

$$\hat{\theta}_{t|t} = \hat{\theta}_{t|t-1} + \Sigma_{t|t-1}x'_{t-1}(H_t + x_{t-1}\Sigma_{t|t-1}x'_{t-1})^{-1}(y_t - x_{t-1}\hat{\theta}_{t-1}) \tag{4.32}$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1}x'_{t-1}(H_t + x_{t-1}\Sigma_{t|t-1}x'_{t-1})^{-1}x_{t-1}\Sigma_{t|t-1}, \tag{4.33}$$

λ is defined as a forgetting factor. Koop and Korobilis (2012) state that the weight of observations j periods in the past is given by λ^j where $0 < \lambda^j \leqslant 1$. We set $\lambda = 0.99$, following Raftery et al. (2010), which implies that, for monthly data observations two years ago are allocated approximately 80% as much weight as last period's observation. Following Koop and Korobilis (2012), an exponentially weighted moving average estimate of $H_t^{(m)}$ is applied which is an appropriate estimator in case of time-varying volatilities. This approach allows for a recursive estimation which is used to forecast volatility. More specifically, the forecast of time $t+1$ using data up to time t is

$$\hat{H}_{t+1|t} = \kappa\hat{H}_{t|t-1} + (1-\kappa)(y_t - x_{t-1}\hat{\theta}_t)^2 \tag{4.34}$$

where κ is a decay factor and set to be $\kappa = 0.97$ for monthly data as proposed by

Riskmetrics. We set $H_{t-1|t-1} = var(y)/4$, however, using different values would not change our main results as the value of κ shrinks as more data are observed over time (see 4.32).

Subsequently, we perform recursive forecasting using the following predictive distribution

$$y_t|y^{t-1} \sim N(x_{t-1}\hat{\theta}_{t-1}, H_t + x_{t-1}\Sigma_{t|t-1}x'_{t-1}). \tag{4.35}$$

However, we do not wish our results to be conditional on a specific model. Koop and Korobilis (2012) demonstrate that defining a transition matrix using a Markov chain Monte Carlo algorithm is most likely to be computationally infeasible. We, thus, employ α which denotes a forgetting factor for the state equation for the models. In other words, α can be interpreted as the weight used in DMA which is allocated to model m. Therefore, DMA is implemented using

$$p(\Psi_{t-1}|y^{t-1}) = \sum_{k=1}^{K} p(\theta_{t-1}^{(m)}|M_{t-1} = m, y^{t-1})Pr(M_{t-1} = m|y^{t-1}), \tag{4.36}$$

where $p(\theta_{t-1}^{(m)}|M_{t-1} = m, y^{t-1})$ is given by (20), and then the model prediction equation is

$$Pr(M_{t|t-1,m} = \Pr(M_{t-1} = m|y^{t-1})^{\alpha} / \sum_{m=1}^{M} Pr(M_{t-1} = m|y^{t-1})^{\alpha}, \tag{4.37}$$

while we set $\alpha = 0.99$. It is worth noting that BMA is a special case of this methodology, and can be implemented by setting $\alpha = \lambda = 1$.

The models are initially equally weighted where the relative weight is equal to $\varphi_{0,m} = 1/M$, but weights are updated each month with respect to the relative posterior predictive model probabilities, using

$$\varphi_{t|t,k} = w_{t,m} / \sum_{m=i}^{M} w_{t,m}, \tag{4.38}$$

where

$$w_{t,m} = \varphi_{t|t-1} lf_m(y_t|y_{t-1}) \tag{4.39}$$

$$\varphi_{t|t-1,m} = \frac{\varphi_{t-1|t-1,m}^{\alpha}}{\sum_{m=1}^{K} \varphi_{t-1|t-1,m}^{\alpha}}. \tag{4.40}$$

Note that in 4.39, $lf_m(y_t|y_{t-1})$ is the log-likelihood density of 4.35.

We set $\hat{\theta}_0^{(m)} = 0$ and $\Sigma_0^{(m)} = 100$ to have an uninformative prior and obtain two series of forecasts where one is generated by the DMA, and the other by dynamic model selection (DMS) which selects the model with the highest posterior predictive model probability at each month. Figure 4.6 displays the DMS results for economic models. Panel A shows that model with Δcpi_{diff} and Δmf_{diff} as macroeconomic predictors, is found to be the best model most often. However, there are ten other models that are also found to have the highest posterior predictive probability at some points in time.

The frequency of changes in the best model is given in Panel B. More specifically, we define an index which varies between zero, if the best model at time t is the same as the one at time $t-1$, and one otherwise. A few changes are observed in the periods 1980-1983 and 1993-1998. The figure also shows that the best model changed post crisis and over the last few years. It is worth noting that DMS results for ECON provide support for the performance of parsimonious models as we find that 67% of the best ECON throughout the sample are univariate models. DMS results for TECH are shown in Panel A of Figure 4.7. We observe that here the DMS varies among 23 different TECH among which MOM(9), OsMA(6,7) and OsST(3,30) seem to achieve the highest probability most often. Panel B shows that the best model changes over the 1988-2001 period, after being consistent for almost ten years.

$<$ Figures 4.6&4.7 around here $>$

To evaluate the performance of this method against the random walk, we compute the R^2_{OS} by Campbell and Thompson (2008) (see Equation 4.13). We find that R^2_{OS} of DMA and DMS are positive for both sets of models, varying between 4.4% and 4.6% for ECON and 1.5% and 2% for TECH, implying support for predictability. In order to find the periods in which the models perform best, we compute the difference in the cumulative sum of squared out-of-sample forecast errors (SSE) between the competing models and the random walk benchmark, following Welch and Goyal

(2008). If the $\Delta CumSSE_t$ value, computed as follows, are positive and increasing, it implies that the model generates more accurate forecasts at that specific point in time.

$$\Delta CumSSE_{i,t} = \sum_{l=1}^{t}(e_{RW,l})^2 - \sum_{l=1}^{t}(e_{i,l})^2 \tag{4.41}$$

where $i = [DMA, DMS]$.

Figure 4.8 plots the SSE for DMA and DMS for economic and TECH in Panels A and B, respectively. Our results show that DMA and DMS of ECON perform similarly throughout our sample. In Panel A, we observe poor performance for both DMA and DMS until the end of 2003 where their performance starts fluctuating between positive and negative areas for almost three years. Outperformance, however, begins in 2007 and lasts until the end of our sample. In the case of TECH, the performance of DMA and DMS seem similar for the first few years, DMA outperforms random walk for the 1987-20014 period. DMA is also found to consistently beat the benchmark since 2009, after a long period of poor performance. Our results in Panel B reveal that not only can this method detect the predictive ability of TECH, but also it performs well during and post crisis.

$<$ Figure 4.8 around here $>$

Gargano and Pettenuzzo (2014) point out the drawback of R^2_{OS} and SSE is that both approaches ignore information on the full probability distribution. They suggest using predictive likelihood to evaluate the accuracy of the density forecasts. By taking log likelihood density of (25) at each point in time, we obtain log predictive scores for all the competing models and also the random walk benchmark, denoted by $LS_{m,t}$ and $LS_{RW,t}$, respectively. Our results show that the averages for both DMA and DMS methods and for both pools of models are positive which indicates that our approach outperforms the random walk. However, to investigate whether the differences in forecast accuracy are significant, we follow Clark and Ravazzolo (2014) where Diebold and Mariano (1995) is applied to test the equality of the average log scores. As stressed by Clark and McCracken (2011), the Diebold-Mariano test is a conservative test in the case of nested models, and since all of the competing models considered here nest the random walk, we report one-sided test p-values, suggested by Gargano and Pettenuzzo (2014). The p-values provide statistical support for our previous findings, indicating that our results are highly significant at the 1% level. The cumulative log score (LS) differentials between the benchmark and the ith competing model where $i = [DMA, DMS]$, is displayed in Figure 4.9. it is computed as follows

$$\Delta CumLS_{i,t} = \sum_{l=1}^{t} [LS_{i,l} - LS_{RW,l}]. \tag{4.42}$$

Panel A and B show that the values of the cumulative log score are positive for both approaches and both sets of models after 1985, which implies that forecasts obtained from the competing models are more accurate than those of the benchmark. It is worth noting that for both pools of models, outperformance of DMS starts before DMA.[20] The performance of DMS is slightly better than DMA throughout the sample. This can be interpreted as DMS allocates zero weight on all models other than the one with the highest posterior predictive model probability, while shrinking the contribution of all models except a single one towards zero. Koop and Korobilis (2012) explain that this additional shrinkage may provide some additional forecast benefits over DMA. Therefore, these results support our previous findings and indicate significant predictive ability for both economic and technical models.

< Figure 4.9 around here >

4.5 Conclusion

We analyze predictability of USD/CAD exchange rate. Two categories of predictors are employed: macroeconomic variables and technical indicators where the predictors are widely used in the literature (see Rossi (2013) and Neely et al. (2014)). We consider all possible combinations of seven macroeconomic variable and 126 individ-

[20]The outperformance of DMS for economic (technical) models starts 16 (52) months earlier than that of the DMA.

ual technical indicators, and compare their performance. Our pool of models also consists of three principal component models. Each model is employed to generate out-of-sample one-step-ahead forecasts for the exchange rate changes. To evaluate the global performance of the models, different metrics are applied, such as the Theil (1971) MSFE decomposition, the Campbell and Thompson (2008) R^2_{OS}, and the Clark and West (2007) test. The results show that significant number of ECON and TECH beat the benchmark where the number of outperforming economic models is considerably higher than TECH. To check the robustness of our results, we use the Clark and McCracken (2012) test which is a modified version of the White (2000) reality check and controls for data snooping. The reality check result confirms our previous findings, implying that at least one of the models generates more accurate forecasts than the random walk benchmark.

We use the Giacomini and White (2006) test and the Fluctuation test to highlight the difference between the global and local forecasting performance of our models. The idea is to show that when the performance of a model is evaluated globally, the local outperformance of the model may be neglected. This is in line with Timmermann's (2008) study where he shows that the outperformance of a model is not consistent. This, thus, motivates our first methodology that investigates whether an investor could use information embedded in the $\bar{R}^2$, MSFE or SIGN criteria at each point in time to choose the outperforming model in the following period. The main contribution of this chapter is that we show that one could beat the benchmark by using the $\bar{R}^2$ criterion in choosing the best economic models and putting an equal weight on their forecasts. It is worth nothing that while the results support the predictive ability of all the macroeconomic variables, the results are mostly in favor of economic models with more than three predictors where the best models tend to change frequently. This sheds light on the importance of considering a variety of models as well as variables.

We implement dynamic model averaging (DMA) which is a mixture of Bayesian model averaging, hidden Markov models and forgetting in state-space modeling. Our motivation is that this method generates forecasts recursively and therefore does not require a specific window size, and also allocates updating weights among all the forecasts with respect to the relative posterior predictive model probabilities. We also generate forecasts using the model with highest posterior predictive probability (DMS) at each month. The Campbell and Thompson (2008) R^2_{OS} results suggest that the forecasts generated by our method for both economic and technical models are more accurate than those of the random walk. The economic model with Δcpi_{diff} and Δmf as predictors, and MOM(9), OsMA(6,7) and OsST(3,30) are chosen as the DMS most often throughout the sample. To check the robustness of our results, we compute the cumulative sum of squared out-of-sample forecast error between the competing models and the benchmark. We find although there is a long period of underperformance for both DMA and DMS of economic models, they generate more accurate forecasts than the benchmark prior, during and after the crisis. The results for technical models imply that DMA outperforms the benchmark from 1987. The log score difference for DMA and DMS, compared to the benchmark, and also test statistics of the Diebold and Mariano (1995) support our previous findings in favor of the predictive ability of DMA and DMS. Our findings show that the cumulative log scores for both economic and technical models are positive after 1985, indicating more accurate forecasts.

Although predictive ability of the competing models tend to be short-lived, one could generate more accurate forecasts by considering a pool of models and evaluates their performance at each period, so that only the best models would be used for the following period. In this chapter, we show that this could be done by applying either R^2_{OS} criterion or DMA methodology. We demonstrate that while the former is only able to detect the predictive ability of economic models, the latter finds support for outperformance of both economic and technical models.

Table 4.1 Description of data

The table summarizes in-sample estimation results. Our pool of models consists of all possible combinations of six macroeconomic variables and a constant that generate 126 economic models, 126 individual technical indicators, and three principal component models (PCs). The second (fourth) column presents the number of models in which each of the macroeconomic variables (technical indicators) is found to be significant. The statistical significance of the models is computed by a wild bootstrap procedure using 2000 replications. Panel A shows the predictors with bootstrapped p-values smaller than 5%. Panel B reports the number of economic and technical models with significant $\bar{R}^2$. Following Campbell and Thompson (2008), we compare $\bar{R}^2$ to SR^2 and conclude that a model with a $\bar{R}^2$ larger than the relative SR^2 contains valuable information about the exchange rate changes.

Panel A

Macroeconomic Variables	# of ECON with significant coefficient for the Relevant Variable	Technical Indicators	# of TECH with significant coefficient for the Relevant Variable
Δppi_{diff}	0	OsMA	13
Δy_{diff}	0	MACD	0
Δi_{diff}	0	MA	12
Δcpi_{diff}	0	MOM	1
Δmf_{diff}	0	OsST	0
Δs	42	BOL	2

Panel B

# of ECON with Significant $\bar{R}^2$ (inc. PC_{ECON})	# of TECH with Significant $\bar{R}^2$ (inc. PC_{TECH})
126	117

Table 4.2 Parameter instability results

The table reports the result of the Elliott and Muller (2006) instability test for economic models. The values are in percentages which show the proportion of models with an identical number of explanatory variables (including a constant) that reject the null of parameter stability, for instance, 24% of the ECON with two explanatory variables reject the null hypothesis at the 1% level of significance, implying parameter instability. Note that for each row, the value in each column is independent from the ones in other columns.

# of Variables in the Model	1%	5%	10%
1	17	0	17
2	24	10	5
3	31	9	9
4	37	9	6
5	38	19	10
6	14	57	14
7	0	100	0

Table 4.3 Out-of-sample forecasting results

The tables summarizes the out-of-sample results for both sets of economic and technical models. Panel A reports the results for the Theil (1971) MSFE decomposition as explained in Section 4.3.2. Models with lower bias and also those with more efficiency than the random walk benchmark are reported. Panel B presents the number of models with positive out-of-sample R^2 (R^2_{OS}) suggested by Campbell and Thompson (2008). Panel C presents the results for the Clark and West (2007) test. More specifically, the number of models that reject the null hypothesis of equal forecast accuracy at three different significance levels is reported. Note that in Panel C and for each pool of models, the values in rows are independent from one another.

Panel A: Bias-Efficiency

# of ECON with Lower Bias	# of ECON with Higher Efficiency	# of TECH with Lower Bias	# of TECH with Higher Efficiency
84	112	24	44

Panel B: # of Models with Positive R^2_{OS}

ECON	TECH
112	31

Panel C: # of Models Outperforming the Benchmark According to the CW

ECON		TECH	
1%	112	1%	12
5%	0	5%	26
10%	4	10%	20

Fig. 4.1 Fluctuation test results

Figure 4.1 displays the Fluctuation test results for two arbitrary economic and technical models. The economic model includes Δcpi, Δms and Δs as predictors, and the technical model consists of OsST(6,30) signals and a constant. We choose 105 forecasts as the size of the rolling window to construct the test statistic. This is almost 30% of the number of observations in the out-of-sample period. The black dashed-dotted line represents the critical value and therefore when the blue line is above that, it indicates significant predictive ability. Note that these results contrast with those of the Giacomini and White (2006) test, since the latter evaluates the forecasts according to their global performance.

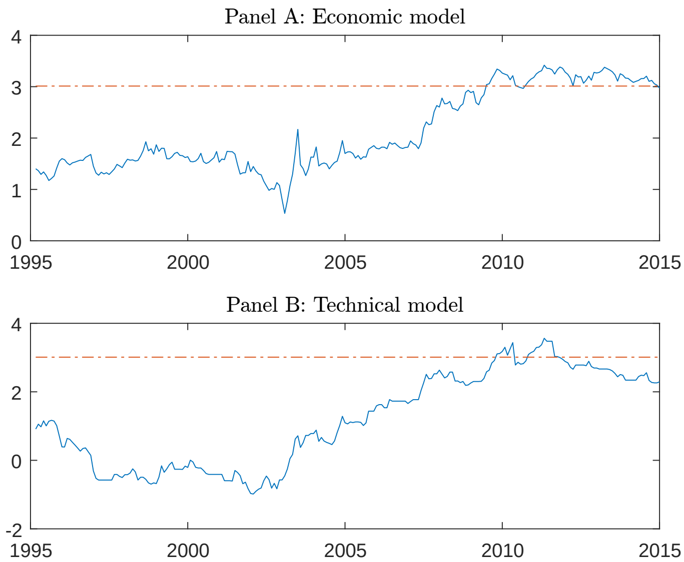


Fig. 4.2 Info criteria ECON model selection portfolio

Figure 4.2 displays the results of our first methodology for ECON where three criteria are employed to evaluate the predictive ability of the pool of ECON to detect the best models at each period. To do so, we take into account statistically significance of the model selection criteria by calculating their bootstrapped p-values. Subsequently, the best models are used to generate forecasts at the following period. Each panel, therefore, shows the proportion of the best models selected by each criterion, ranging from 0 to 100%. For instance, we observe that SIGN criterion barely contributes to the best models for the 1998-2008 period.

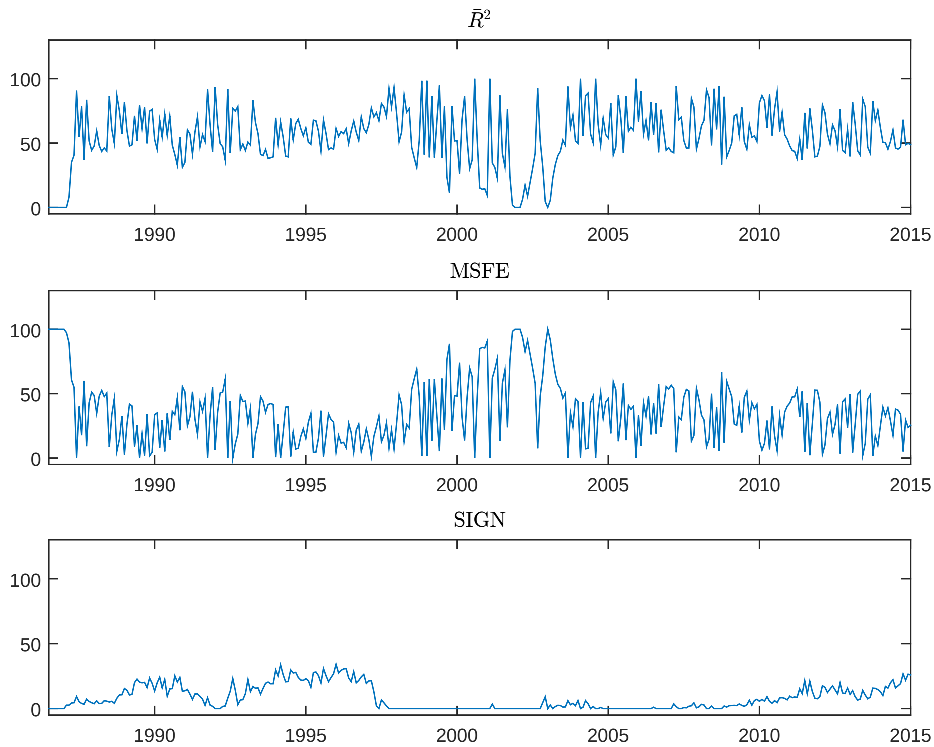


Fig. 4.3 Info criteria TECH model selection portfolio

Figure 4.3 displays the results of our first methodology for TECH where three criteria are employed to evaluate the predictive ability of the pool of TECH to detect the best models at each period. To do so, we take into account statistically significance of the model selection criteria by calculating their bootstrapped p-values. Subsequently, the best models are used to generate forecasts at the following period. Each panel, therefore, shows the proportion of the best models selected by each criterion, ranging from 0 to 100%. For instance, we observe that SIGN criterion barely contributes to the best models for the 1997-2005 period.

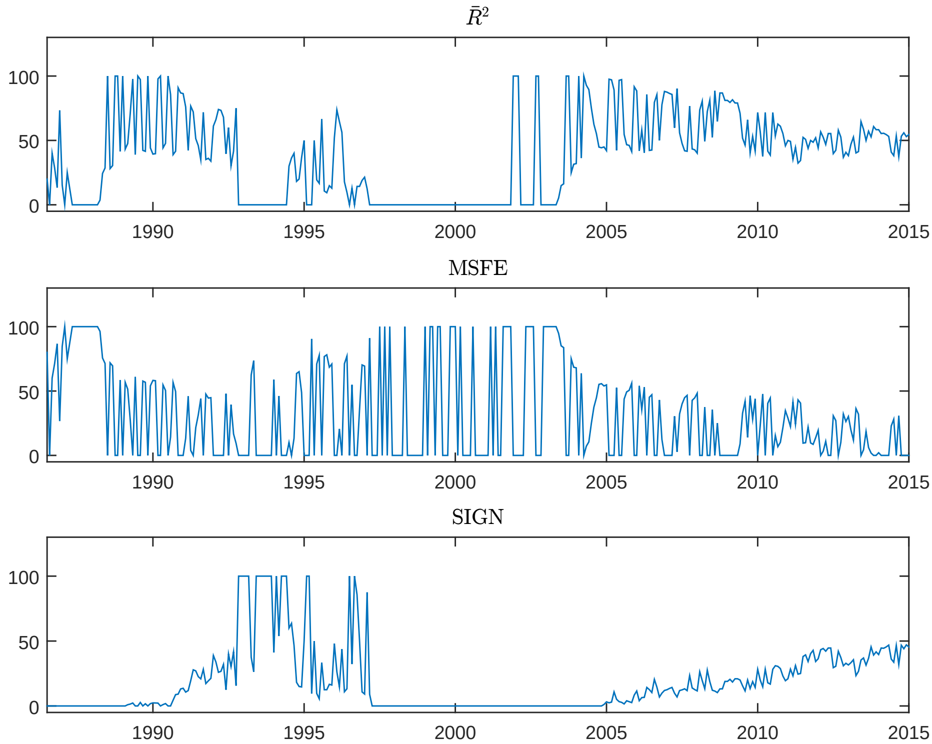


Fig. 4.4 Info criteria relative performance

Figure 4.4 displays proportion of the ECON selected by both MSFE and $\bar{R}^2$ in Panel A and by SIGN and $\bar{R}^2$ in Panel B. While fluctuation of this ratio in Panel A is due to volatile number of models selected by both $\bar{R}^2$ and MSFE, Panel B shows a ratio of zero for the 1998-2003 period which is solely due to inability of SIGN criterion in detecting any best model.

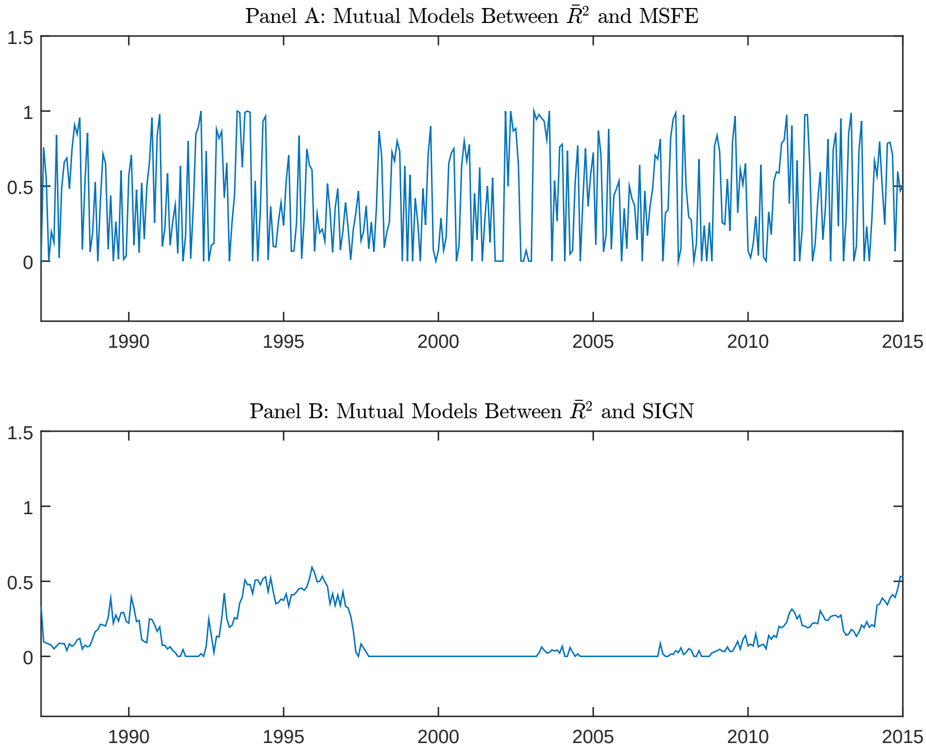


Fig. 4.5 $\bar{R}^2$ model selection portfolio

Figure 4.5 displays how many times each of the 126 ECON has been selected by $\bar{R}^2$. For instance, model 4.14 has been selected as one of the best models for more than 300 times. Note that there is no model which has not ever been chosen as the best model throughout our sample which implies that the best models tend to change frequently throughout the sample. It highlights the importance of considering a pool of models.

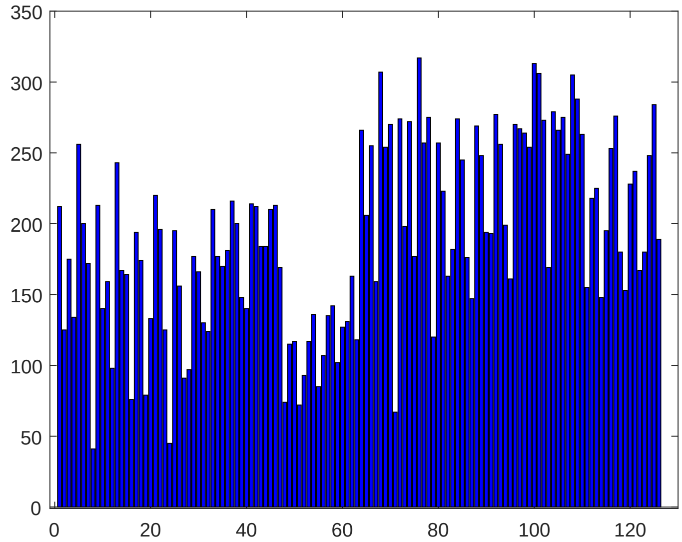
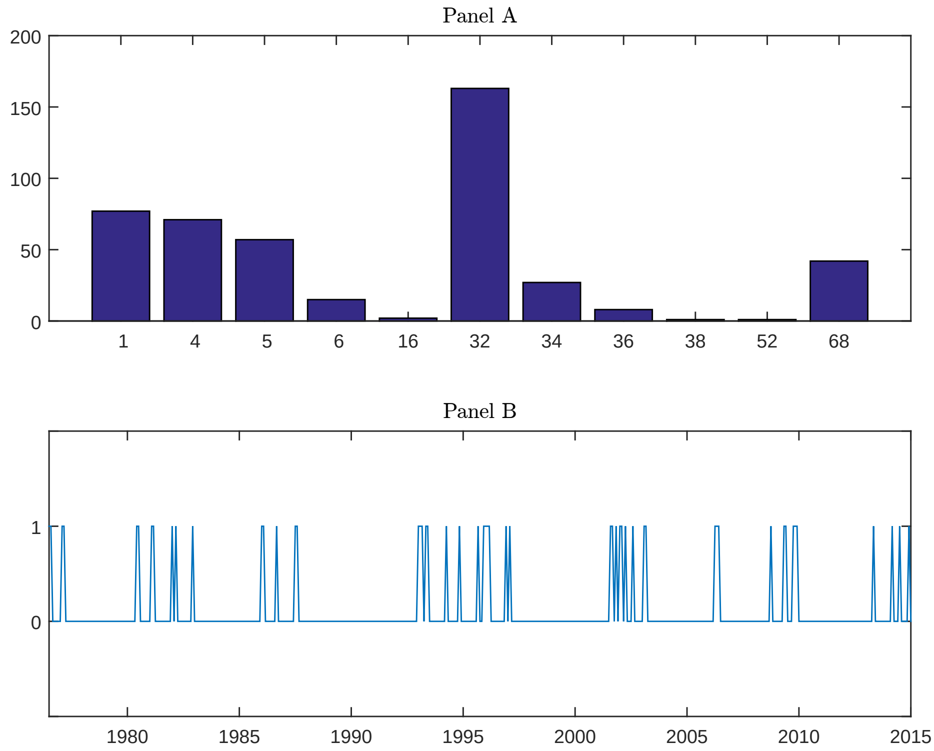


Fig. 4.6 DMS results for ECON models

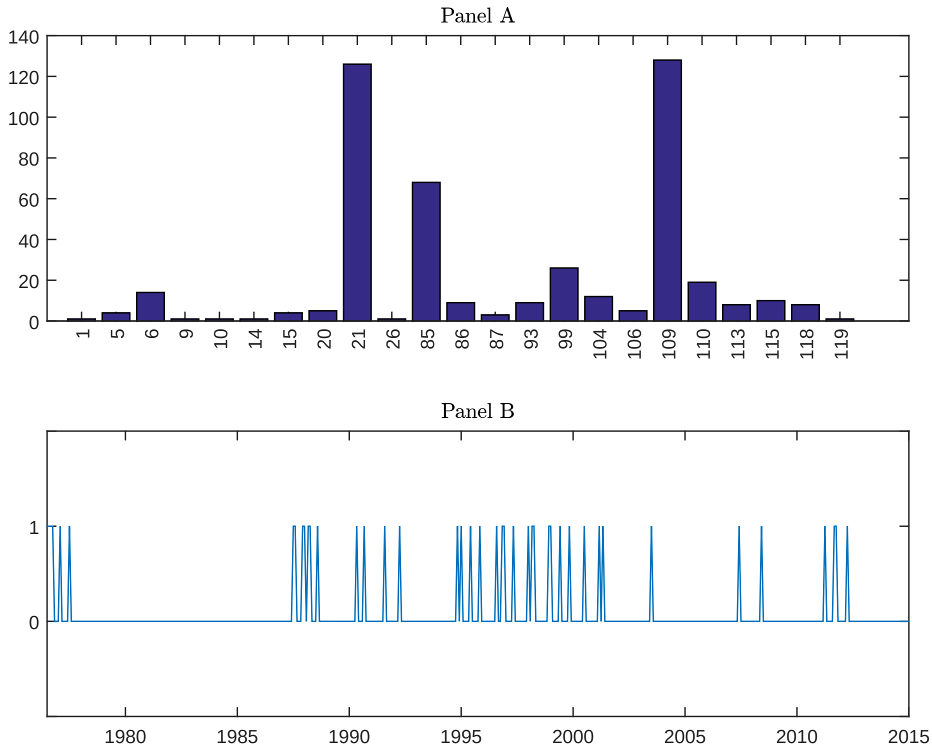
Figure 4.6 displays dynamic model selection (DMS) results for ECON where we assume an uninformative prior. We estimate the parameters of the competing models recursively and generate one-step-ahead forecasts at each month.Panel A represents models with highest posterior predictive probability at each month on X-axis, while Y-axis shows how many times a model has been selected as the best model throughout our sample. Panel B shows how often the best model changes. In other words, we define a vector of zero and one, where one represents the case when the best model at time t is different than the one at time $t-1$, and zero expresses otherwise. The table reports the predictors of the selected models.



Model	Predictors
1	Δs
4	Δcpi_{diff}
5	Δcpi_{diff}, Δs
6	Δcpi_{diff}, Δmf_{diff}
16	Δy_{diff}
32	Δppi_{diff}
34	Δppi_{diff}, Δmf_{diff}
36	Δppi_{diff}, Δcpi_{diff}
38	Δppi_{diff}, Δcpi_{diff}, Δmf_{diff}
52	Δppi_{diff}, Δcpi_{diff}, Δy_{diff}
68	c, Δcpi_{diff}, Δs

Fig. 4.7 DMS results for TECH models

Figure 4.7 displays dynamic model selection (DMS) results for TECH where we assume an uninformative prior. We estimate the parameters of the competing models recursively and generate one-step-ahead forecasts at each month.Panel A represents models with highest posterior predictive probability at each month on X-axis, while Y-axis shows how many times a model has been selected as the best model throughout our sample. Panel B shows how often the best model changes. In other words, we define a vector of zero and one, where one represents the case when the best model at time t is different than the one at time $t-1$, and zero expresses otherwise. The table reports the predictors of the selected models.



Models	Predictors
1, 5, 6, 9, 10, 14, 15, 20, 21, 26	OsMA
85, 86, 87	MOM
93, 99, 104, 106, 109, 110, 113, 115, 118	OsST
119	BOL

Fig. 4.8 Cumulative SSE results

Figure 4.8 displays the cumulative sum of squared forecasts errors for the random walk benchmark minus those of dynamic model average (DMA) and dynamic model selection (DMS) for economic and TECH in Panel A and B, respectively. We estimate the parameters of the competing models recursively and generate one-step-ahead forecasts at each month. Positive values indicate more accurate forecastability for the competing models, while negative values suggest the opposite. In both panels, the blue solid line represents DMA model and the blue dashed line tracks the DMS model results.

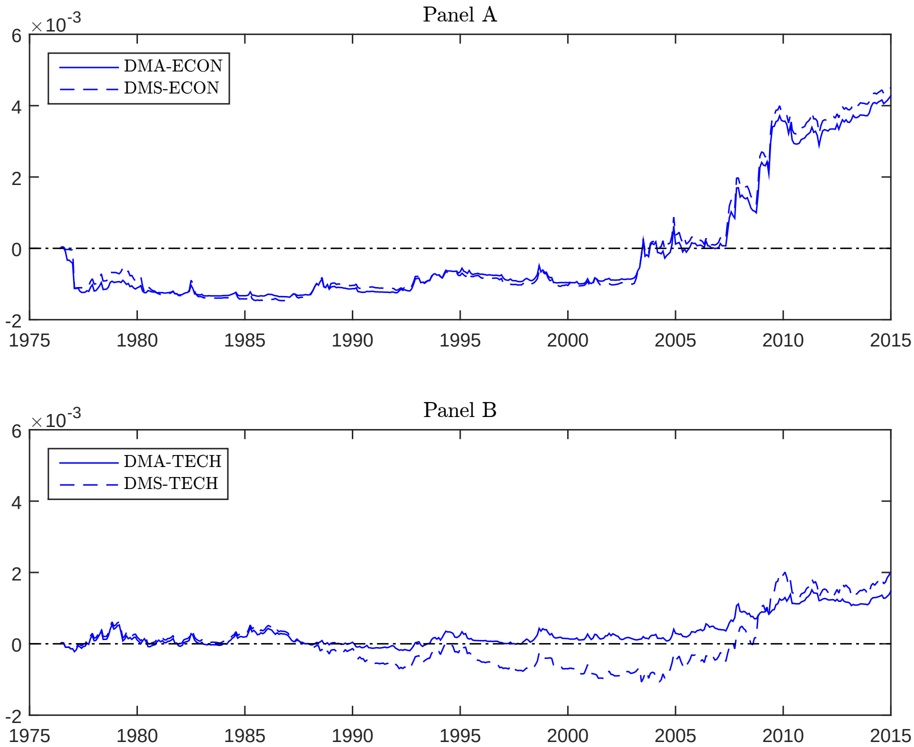
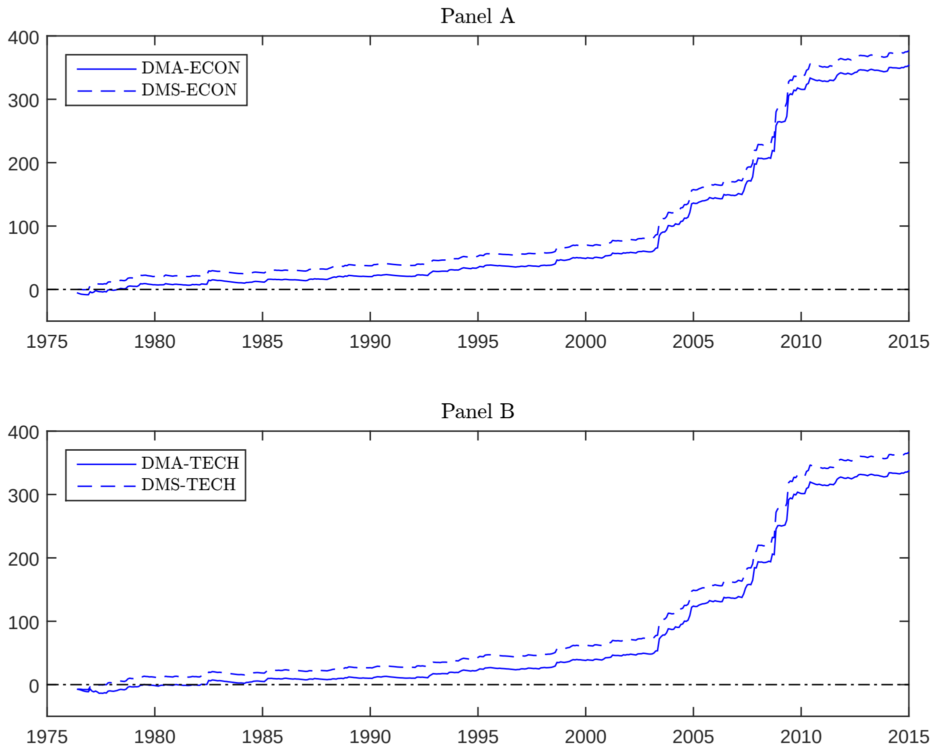


Fig. 4.9 Cumulative LS results

Figure 4.9 displays the cumulative sum of log score differences between the economic and technical models and the benchmark, plotted in Panel A and B, respectively. We estimate the parameters of the competing models recursively and generate one-step-ahead density forecasts. Subsequently, we use them to compute log predictive scores. Positive values indicate more accurate forecastability for the competing models, while negative values suggest the opposite. In both panels, the blue solid line represents dynamic model average (DMA) model and the blue dashed line tracks dynamic model selection (DMS) model.



Chapter 5

Concluding remarks

The empirical literature on nominal exchange rates shows that the current exchange rate is the best predictor for future exchange rates. This is known as Meese-Rogoff puzzle and has surprisingly remained robust after many further studies and development of more sophisticated econometric techniques (see Berkowitz and Giorgianni, 2001; Frankel and Rose, 1995). Recently, a literature has developed claiming that exchange rates are predictable (see Engel and West, 2007; Molodtsova and Papell, 2012; Rime et al., 2010).

This thesis makes several contributions. First, it contributes to the literature that commodities can forecast exchange rates. It uses eight commodities that are the dominant exports of three countries and also have significant shares in the world commodity market. The Clark and West (2007) test results show that realized commodity prices have strong predictive ability for both daily and monthly data, but the monthly outperformance is subject to data snooping. It investigates the forecasting performance of lagged prices by using the fluctuation test and finds that a few commodities show evidence of predictive ability at some points in time. This is the motivation to combine the whole information set to test whether considering all available information beats the random walk. It applies information combination and forecast combination. The former constructs a large model containing all the commodities, while the latter combines the forecasts generated by each commodity with equal weights. Principal component analysis is also applied which identifies patterns in the data and highlights their similarities and differences. It finds that both information combination and forecast combination generate more accurate forecasts than the random walk at the daily frequency.

Second, it contributes to the debate on the profitability of technical analysis in foreign exchange markets by employing a new bootstrap method. The false discovery rate by Barras et al. (2010) can detect almost all possible outperforming rules. The initial results support long-term in-sample profitability of trading rules, after taking into account transaction costs. To address the question whether investors could have predicted outperforming rules ex-ante, it performs a persistence analysis. It uses Sharpe ratio as the performance measure and finds that if a speculator constructed a portfolio of the outperforming rules and updated it on a monthly basis, she would obtain a positive Sharpe ratio on average. It then shows that despite these profit opportunities, the performance of trading rules is fluctuating over time which implies that the speculator could not persistently profit from technical analysis. Instead the results support the adaptive market hypothesis by Lo (2004).

Finally, it makes two major contributions to the forecasting exchange rate literature. It employs a set of economic models and a sets of technical models and compare their performance in forecasting exchange rates. The first contribution is that after detecting instability in the forecasting performance of the models, it constructs a portfolio of outperforming models based on three different criteria and update it every month. The empirical results also show that there is evidence for predictive ability of economic models for only one criterion. However, when forecasts obtained from outperforming technical and economic models are combined with equal weights, the random walk benchmark is beaten by all three criteria. This highlights the importance of considering both technical analysis and fundamental analysis in forecasting exchange rates since they can explain different aspects of exchange rate changes. The second contribution is that, to check that the results are not subject to the choice of rolling specification or in-sample window, it applies dynamic model averaging by Raftery et al. (2010). This improves the previous results and provides more support for the predictive ability of technical analysis.

5.1 Future research

An immediate avenue for future research emerges from the results in Chapter 2. Inoue and Kilian (2006) state that in-sample predictive ability is a necessary condition for out-of-sample forecastability. In other words, it would be interesting to see how improving the competing models internally (omitted variables, functional forms, endogeneity) could affect their forecasting performance at different frequencies.

An immediate avenue for future research emerges from the results in Chapter 3. One could investigate the performance of of trading rules in different markets. In other words, instead of considering only exchange rates, a portfolio of variety of assets such as commodities, stocks, bonds and exchange rates could be considered. This situation is more similar to what fund managers face in practice.

It would also be interesting to gain some further insights into the predictive ability results in Chapter 4. Given the unstable relationship between fundamentals and exchange rates, it would be interesting to investigate the optimal window size for estimating the economic models. Inoue et al. (2015) propose a novel method for selecting the estimation window size for forecasting. Since the forecasting performance is sensitive to the choice of window size, future work along this line of research could apply the optimal window size to our models and examine if this leads to more accurate forecasts. On the other hand, there is a growing literature that claims that statistical significance does not mechanically imply economic significance for risk averse investors (see Della Corte et al., 2008, 2009; Thornton and Valente, 2012). Therefore, a future work can investigate the economic value of investing in exchange rates by exploiting the predictive ability that we document (Abhyankar et al. (2005)). Moreover, since exchange rates are subject to significantly high transaction costs, it will further affect the profitability of real time trading strategies that exploit such predictive ability.

Appendix A

The model confidence set

This appendix provides details of the model confidence set in Section 2.3.1.

Hansen et al. (2011) introduces the model confidence set (henceforth MCS) which is a set of models containing the best model with a given level of confidence. This approach identifies the subset of models that includes the best model with a given level of significance, without having to specify a particular model as a benchmark. An advantage of the MCS is that it allows for the possibility that more than one model can be the best.

Let M_0 contain a finite number of models, denoted by $i = 1, \ldots, m_0$ and $L_{i,t}$ be the loss associated with object i at time t. The set of superior objects is calculated as follows:

$$M^* \equiv [i \in M_0 : E(d_{ij,t}) \leq 0 \quad \text{for all} \quad j \in M_0]$$

where $d_{ij,t} \equiv L_{i,t} - L_{j,t}$ for all $i, j \in M_0$. The MCS approach is based on an equivalent test, δ_M, and an eliminative rule, e_M. The equivalent test is implemented to test the following null hypothesis.

$$H_{0,M} : E(d_{ij,t}) = 0 \quad \text{for some} \quad i, j \in M$$

The eliminative rule determines which object is to be removed from M when the null is rejected. Hansen et al. (2011) define the MCS algorithm as follows

- Step 0: initially set $M = M_0$.
- Step 1: test the null by using the equivalent test at level α.

- Step 2: if the null is not rejected, it is concluded that $\hat{M}^*_{a-\alpha}$, otherwise e_M is used to eliminate an object from M and repeat the procedure from Step 1.

$\hat{M}^*_{(1-\alpha)}$, which is the set of objects which are not eliminated, is known as the model confidence set. The MCS gives p-values for each object. In other words, for a given object, $i \in M^0$, the MCS p-value, $\hat{p}_i$, is the threshold at which $i \in \hat{M}^*_{1-\alpha}$, if and only if $\hat{p}_i \geq \alpha$. Therefore, an object with a large MCS p-value is more likely to be one of the best alternative in M^0. Note that we employ the MCS to determine the set of models which are not eliminated at the 5% significance level.

The MCS results are reported in Tables A.1-A.2 for daily and monthly data, respectively. Following Hansen et al. (2003), the confidence level for the MCS is set to $\alpha = 0.25$. The block length and the number of bootstrap samples are set to 2 and 1,000, respectively. For the contemporaneous models at the daily frequency, the MCS includes oil for all three currencies. Moreover, it contains gold for the AUD and BRL, and copper for the AUD and CAD. Our results shows that wheat is also included in the MCS of BRL. The findings are similar across the currencies for monthly data where the MCS includes only copper. This can be interpreted as copper being the only commodity that performs consistently best (in comparison with other commodities) in forecasting changes in the currencies.

The MSC results for lagged models are quite large for all currencies at both frequencies. The implication is that none of the competing models is superior to the others. This is consistent with the SSPA results where no model outperforms the random walk benchmark. In particular, the MCS contains six models (oil, gas, copper, aluminum, coffee and sugar) for the AUD and all the models for the BRL and CAD at the daily frequency. The MCS results for the BRL at the monthly frequency are similar to those at the daily frequency. In the case of AUD and CAD, the MCS contains five and six models, respectively.

Table A.1 MCS for daily data

This table reports the results for the MCS for three currencies: Australian dollar (AUD), Brazilian real (BRL) and Canadian dollar (CAD). The set of selected commodities contains the best model. We set the confidence level for MCS to $\alpha = 0.25$, the block length to 2, and the number of bootstrap samples to 1,000, following Hansen (2003). It is important to note that there is no need to specify a benchmark.

	AUD	BRL	CAD
	Panel A. Contemporaneous		
Oil	✓	✓	✓
Gas			
Gold	✓	✓	
Sugar			
Coffee			
Aluminum			
Wheat		✓	
Copper	✓		✓
	Panel B. Lagged		
Oil	✓	✓	✓
Gas	✓	✓	✓
Gold		✓	✓
Sugar	✓	✓	✓
Coffee	✓	✓	✓
Aluminum	✓	✓	✓
Wheat		✓	✓
Copper	✓	✓	✓

Table A.2 MCS for monthly data

This table reports the results for the MCS for three currencies: Australian dollar (AUD), Brazilian real (BRL) and Canadian dollar (CAD). The set of selected commodities contains the best model. We set the confidence level for MCS to $\alpha = 0.25$, the block length to 2, and the number of bootstrap samples to 1,000, following Hansen (2003). It is important to note that there is no need to specify a benchmark.

	AUD	BRL	CAD
	Panel A. Contemporaneous		
Oil			
Gas			
Gold			
Sugar			
Coffee			
Aluminum			
Wheat			
Copper	✓	✓	✓
	Panel B. Lagged		
Oil	✓	✓	
Gas	✓	✓	✓
Gold	✓	✓	✓
Sugar		✓	
Coffee		✓	
Aluminum	✓	✓	✓
Wheat		✓	✓
Copper	✓	✓	✓

References

Abhyankar, A., Sarno, L., Valente, G., 2005. Exchange rates and fundamentals: evidence on the economic value of predictability. Journal of International Economics 66 (2).

Akram, Q. F., 2004. Oil prices and exchange rates: Norwegian evidence. The Econometrics Journal 7.

Akram, Q. F., 2009. Commodity prices, interest rates and the dollar. Energy Economics 31.

Alexander, S. S., 1961. Price movements in speculative markets: Trends or random walks. Industrial Management Review (pre-1986) 2.

Allen, H., Taylor, M. P., 1990. Chartist, noise and fundamentals in the london foreign exchange market. Economic Journal 100.

Aloui, R., Aïssa, M. S. B., Nguyen, D. K., 2013. Conditional dependence structure between oil prices and exchange rates: a copula-garch approach. Journal of International Money and Finance 32.

Alquist, R., Chinn, M. D., 2008. Conventional and unconventional approaches to exchange rate modelling and assessment. International Journal of Finance & Economics 13 (1).

Andrews, D. W. K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61.

Andrews, D. W. K., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62.

Bacchetta, P., Wincoop, E. v., Beutler, T., 2010. Can parameter instability explain the meese-rogoff puzzle?, Working Paper, NBER International Seminar on Macroeconomics 2009.

Baillie, R. T., Bollerslev, T., 2000. The forward premium anomaly is not as bad as you think. Journal of International Money and Finance 19.

Bajgrowicz, P., Scaillet, O., 2012. Technical trading revisited: False discoveries, persistence tests, and transaction costs. Journal of Financial Economics 106.

Bank for International Settlements, 2016. Triennial central bank survey of foreign exchange and otc derivatives markets in 2016. Monetary and Economic Department.

Barras, L., Scaillet, O., Wermers, R., 2010. False discoveries in mutual fund performance: Measuring luck in estimated alphas. The Journal of Finance 65.

Bekaert, G., Hodrick, R. J., 2001. Expectations hypotheses tests. Journal of Finance 56.

Belsley, D. A., Kuh, E., Welsch, R. E., 1980. Regression Diagnostics: Identifying Influential Data and Source of Collinearity. John Wiley.

Bénassy-Quéré, A., Mignon, V., Penot, A., 2007. China and the relationship between the oil price and the dollar. Energy Policy 35.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) 57.

Berkowitz, J., Giorgianni, L., 2001. Long-horizon exchange rate predictability? Review of Economics and Statistics 83.

Bessembinder, H., Chan, K., 1995. The profitability of technical trading rules in the asian stock markets. Pacific-Basin Finance Journal 3.

Bodart, V., Candelon, B., Carpantier, J.-F., 2012. Real exchanges rates in commodity producing countries: A reappraisal. Journal of International Money and Finance 31.

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

Brush, J. S., Boles, K. E., 1983. The predictive power in relative strength & capm. The Journal of Portfolio Management 9.

Campbell, J. Y., Thompson, S. B., 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? Review of Financial Studies 21.

Cashin, P., Céspedes, L. F., Sahay, R., 2004. Commodity currencies and the real exchange rate. Journal of Development Economics 75.

Cayen, J. P., Coletti, D., Lalonde, R., Maier, P., 2010. What drives exchange rates? new evidence from a panel of us dollar bilateral exchange rates, Working Paper, Bank of Canada.

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

Chen, Y. C., Rogoff, K., 2003. Commodity currencies. Journal of International Economics 60.

Chen, Y. C., Rogoff, K., Rossi, B., 2010. Can exchange rates forecast commodity prices? Quarterly Journal of Economics 125.

Cheung, Y. W., Chinn, M. D., 2001. Currency traders and exchange rate dynamics: A survey of the us market. Journal of International Money and Finance 20.

Cheung, Y. W., Chinn, M. D., Pascual, A. G., 2005. Empirical exchange rate models of the nineties: Are any fit to survive? Journal of International Money and Finance 24.

Cheung, Y. W., Lai, K. S., 2000. On the purchasing power parity puzzle. Journal of International Economics 52.

Cheung, Y. W., Wong, C. Y. P., 2000. A survey of market practitioners' views on exchange rate dynamics. Journal of International Economics 51.

Chinn, M. D., 1991. Some linear and nonlinear thoughts on exchange rates. Journal of International Money and Finance 10.

Chinn, M. D., 2010. Comment on "can parameter instability explain the meese-rogoff puzzle?". In: Reichlin, L., West, K. D. (Eds.), NBER Book Series NBER International Seminar on Macroeconomics. University of Chicago Press, Chicago.

Chinn, M. D., Meese, R. A., 1995. Banking on currency forecasts: How predictable is change in money? Journal of International Economics 38.

Chinn, M. D., Moore, M. J., 2011. Order flow and the monetary model of exchange rates: Evidence from a novel data set. Journal of Money, Credit and Banking 43.

Clark, T. E., McCracken, M. W., 2001. Tests of equal forecast accuracy and encompassing for nested models. Journal of Econometrics 105.

Clark, T. E., McCracken, M. W., 2005. The power of tests of predictive ability in the presence of structural breaks. Journal of Econometrics 124.

Clark, T. E., McCracken, M. W., 2011. Testing for unconditional predictive ability. In: Clements, M., Hendry, D. (Eds.), Oxford Handbook of Economic Forecasting. Oxford University Press, Oxford.

Clark, T. E., McCracken, M. W., 2012. Reality checks and comparisons of nested predictive models. Journal of Business & Economic Statistics 30.

Clark, T. E., Ravazzolo, F., 2014. Macroeconomic forecasting performance under alternative specifications of time-varying volatility. Journal of Applied Econometrics 30.

Clark, T. E., West, K. D., 2006. Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis. Journal of Econometrics 135.

Clark, T. E., West, K. D., 2007. Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics 138.

Cornell, W. B., Dietrich, J. K., 1978. The efficiency of the market for foreign exchange under floating exchange rates. The Review of Economics and Statistics 60.

Coudert, V., Mignon, V., Penot, A., 2008. Oil price and the dollar. Energy Studies Review 15.

Dale, C., Workman, R., 1981. Measuring patterns of price movements in the treasury bill futures market. Journal of Economics and Business 33.

Della Corte, P., Sarno, L., Thornton, D. L., 2008. The expectation hypothesis of the term structure of very short-term rates: Statistical tests and economic value. Journal of Financial Economics 89 (1).

Della Corte, P., Sarno, L., Tsiakas, I., 2009. An economic evaluation of empirical exchange rate models. Review of Financial Studies 22 (9).

Diebold, F. X., 1989. Forecast combination and encompassing: Reconciling two divergent literatures. International Journal of Forecasting 5 (4).

Diebold, F. X., Mariano, R. S., 1995. Comparing predictive accuracy. Journal of Business & Economic Statistics 13.

Diebold, F. X., Pauly, P., 1990. The use of prior information in forecast combination. International Journal of Forecasting 6 (4).

Dooley, M. P., Shafer, J., 1984. Analysis of short-run exchange rate behavior: March 1973 to november 1981. In: Bigman, D., Taya, T. (Eds.), Floating Exchange Rates and the State of World Trade and Payments. BeardBooks, Washington.

Dornbusch, R., 1976. Expectations and exchange rate dynamics. The journal of political economy.

Drożdż, S., Kwapień, J., Oświęcimka, P., Rak, R., 2010. The foreign exchange market: return distributions, multifractality, anomalous multifractality and the epps effect. New Journal of Physics 12 (10).

Efron, B., 1992. Bootstrap methods: another look at the jackknife. Springer.

Efron, B., Tibshirani, R., 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical Science.

Elliott, G., Muller, U. K., 2006. Efficient tests for general persistent time variation in regression coefficients. Journal of Econometric Studies 73.

Engel, C., Mark, N. C., West, K. D., 2008. Exchange rate models are not as bad as you think. NBER Macroeconomics Annual 2007 22.

Engel, C., M. N. C., West, K. D., 2007. Exchange rate models are not as bad as you think, Working Paper, NBER Macroeconomic Annual.

Engle, C., 1994. Can the markov switching model forecast exchange rates? Journal of International Economics 36.

Engle, R. F., Granger, C. W., Kraft, D., 1984. Combining competing forecasts of inflation using a bivariate arch model. Journal of Economic Dynamics and Control 8 (2).

Evans, M. D. D., Lyons, R. K., 2002. Order flow and exchange rate dynamics. Journal of Political Economy 110.

Fama, E. F., Blume, M. E., 1966. Filter rules and stock-market trading. The Journal of Business 39.

Fang, Y., Xu, D., 2003. The predictability of asset returns: an approach combining technical analysis and time series forecasts. International Journal of Forecasting 19.

Ferraro, D., Rogoff, K., Rossi, B., 2015. Can oil prices forecast exchange rates? an empirical analysis of the relationship between commodity prices and exchange rates. Journal of International Money and Finance 54.

Frankel, J. A., 1979. On the mark: A theory of floating exchange rates based on real interest differentials. The American Economic Review 69 (4).

Frankel, J. A., 2010. The natural resource curse: a survey, Working Paper, National Bureau of Economic Research.

Frankel, J. A., Froot, K. A., 1990. Chartists, fundamentalists, and trading in the foreign exchange market. The American Economic Review 80.

Frankel, J. A., Rose, A. K., 1995. Empirical research on nominal exchange rates. Handbook of International Economics 3.

Frankel, J. A., Rose, A. K., 1996. Currency crashes in emerging markets: An empirical treatment. Journal of International Economics 41.

Freedman, D. A., Peters, S. C., 1984. Bootstrapping a regression equation: Some empirical results. Journal of the American Statistical Association 79.

Frenkel, J. A., 1976. A monetary approach to the exchange rate: Doctrinal aspects and empirical evidence. The Scandinavian Journal of Economics 78.

Gargano, A., Pettenuzzo, Davide, a. T. A., 2014. Bond return predictability: economic value and links to the macroeconomy, Working Paper, SSRN.

Gehrig, T., Menkhoff, L., 2004. The use of flow analysis in foreign exchange: Exploratory evidence. Journal of International Money and Finance 23.

Giacomini, R., Komunjer, I., 2005. Evaluation and combination of conditional quantile forecasts. Journal of Business & Economic Statistics 23 (4).

Giacomini, R., Rossi, B., 2010. Model comparisons in unstable environments. Journal of Applied Econometrics 25.

Giacomini, R., White, H., 2006. Tests of conditional predictive ability. Econometrica 74.

Goodhart, C., 1988. The foreign exchange market: a random walk with a dragging anchor. Economica 55.

Goodman, S. H., 1979. Foreign exchange rate forecasting techniques: implications for business and policy. The Journal of Finance 34.

Gourinchas, P.-O., Rey, H., 2007. International financial adjustment. Journal of Political Economy 115.

Granger, C. W., 1989. Invited review combining forecasts—twenty years later. Journal of Forecasting 8 (3).

Habib, M. M., Kalamova, M. M., 2007. Are there oil currencies? the real exchange rate of oil exporting countries, Working Paper, ECB.

Hansen, P. R., 2003. Asymptotic tests of composite hypotheses, Working Paper, Brown University Economics.

Hansen, P. R., 2005. A test for superior predictive ability. Journal of Business and Economic Statistics 23.

Hansen, P. R., Lunde, A., Nason, J. M., 2003. Choosing the best volatility models: the model confidence set approach. Oxford Bulletin of Economics and Statistics 65 (s1).

Hansen, P. R., Lunde, A., Nason, J. M., 2011. The model confidence set. Econometrica 79.

Harvey, D. S., Leybourne, S. J., Newbold, P., 1998. Tests for forecast encompassing. Journal of Business and Economic Statistics 16.

Hendry, D. F., Clements, M. P., 2004. Pooling of forecasts. The Econometrics Journal 7 (1).

Holden, K., Peel, D. A., 1989. Unbiasedness, efficiency and the combination of economic forecasts. Journal of Forecasting 8.

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

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

Hsu, P. H., Taylor, M. P., 2013. Forty years, thirty currencies and 21,000 trading rules: a large-scale, data-snooping robust analysis of technical trading in the foreign exchange market, Working Paper, Thirty Currencies and 21.

Imbs, J., Mumtaz, H., Ravn, M. O., Rey, H., 2005. Aggregation bias does explain the PPP puzzle, Working Paper, National Bureau of Economic Research.

Inoue, A., Jin, L., Rossi, B., 2015. Rolling window selection for out-of-sample forecasting with time-varying parameters, Working Paper, Center for Economic Policy and Research.

Inoue, A., Kilian, L., 2006. On the selection of forecasting models. Journal of Econometrics 130.

Issa, R., Lafrance, R., Murray, J., 2008. The turning black tide: energy prices and the canadian dollar. Canadian Journal of Economics/Revue canadienne d'économique 41.

Ito, A., 1999. Profits on technical trading rules and time-varying expected returns: evidence from pacific-basin equity markets. Pacific-Basin Finance Journal 7.

Jacobs, B. I., Levy, K. N., 1988. Disentangling equity return regularities: New insights and investment opportunities. Financial Analysts Journal 44.

Jensen, M. C., 1967. Random walks: reality or myth-comment. Financial Analysts Journal 23.

Jensen, M. C., Benington, G. A., 1970. Random walks and technical theories: Some additional evidence. The Journal of Finance 25.

Kilian, L., 1999. Exchange rates and monetary fundamentals: What do we learn from long-horizon regressions? Journal of Applied Econometrics 14.

Koop, G., Korobilis, D., 2012. Forecasting inflation using dynamic model averaging. International Economic Review 53.

Koranchelian, T., 2005. The equilibrium real exchange rate in a commodity exporting country: Algeria's experience, Working Paper, IMF.

Lizardo, R. A., Mollick, A. V., 2010. Oil price fluctuations and us dollar exchange rates. Energy Economics 32.

Lo, A. W., 2004. The adaptive markets hypothesis. The Journal of Portfolio Management 30.

Lopez, C., Murray, C. J., Papell, D. H., 2003. State of the art unit root tests and the PPP puzzle, Working Paper, Department of Economics, University of Houston.

Lui, Y.-H., Mole, D., 1998. The use of fundamental and technical analyses by foreign exchange dealers: Hong kong evidence. Journal of International Money and Finance 17.

Lyons, R. K., 2001. New perspective on fx markets: Order-flow analysis. International Finance 4.

MacDonald, R., Taylor, M. P., 1993. The monetary approach to the exchange rate: rational expectations, long-run equilibrium, and forecasting, Working Paper, IMF Staff Papers.

Maier, P., DePratto, B., 2008. The Canadian dollar and commodity prices: Has the relationship changed over time? Bank of Canada.

Mark, N. C., 1995. Exchange rates and fundamentals: Evidence on long-horizon predictability. American Economic Review 85.

Mark, N. C., Sul, D., 2001. Nominal exchange rates and monetary fundamentals: Evidence from a small post-bretton woods panel. Journal of International Economics 53.

Meese, R. A., Rogoff, K., 1983. Empirical exchange rate models of the seventies: Do they fit out of sample? Journal of International Economics 14.

Meese, R. A., Rogoff, K., 1983a. Empirical exchange rate models of the Seventies: Do they fit out of sample? Journal of International Economics 14.

Meese, R. A., Rogoff, K., 1983b. The out-of-sample failure of empirical exchange rate models: Sampling error or misspecification. In: Frenkel, J. A. (Ed.), Exchange Rates and International Macroeconomics. University of Chicago, Chicago.

Meese, R. A., Rogoff, K., 1988. Was it real? the exchange rate-interest differential relation over the modern floating-rate period. Journal of Finance 43.

Menkhoff, L., 1997. Examining the use of technical currency analysis. International Journal of Finance and Economics 2.

Menkhoff, L., Taylor, M. P., 2007. The obstinate passion of foreign exchange professionals: Technical analysis. Journal of Economic Literature 45.

Molodtsova, T., Papell, D. H., 2009. Out-of-sample exchange rate predictability with Taylor rule fundamentals. Journal of International Economics 77.

Molodtsova, T., Papell, D. H., 2012. Taylor rule exchange rate forecasting during the financial crisis, Working Paper, National Bureau of Economic Research.

Murphy, J. A., 1986. Futures fund performance: A test of the effectiveness of technical analysis. Journal of Futures Markets 6.

Mussa, M., 1982. A model of exchange rate dynamics. Journal of Political Economy 90.

Narayan, P. K., Narayan, S., Prasad, A., 2008. Understanding the oil price-exchange rate nexus for the fiji islands. Energy Economics 30.

Neely, C., Weller, P., Dittmar, R., 1997. Is technical analysis in the foreign exchange market profitable? a genetic programming approach. Journal of Financial and Quantitative Analysis 32.

Neely, C. J., Rapach, D. E., Tun, J., Zhou, G., 2014. Forecasting the equity risk premium: The role of technical indicators. Management Science 60.

Neely, C. J., Weller, P. A., 2011. Technical analysis in the foreign exchange market, Working Paper, Federal Reserve Bank of St. Louis.

Oberlechner, T., Osler, C. L., 2008. Overconfidence in currency markets, Working Paper, SSRN.

Perneger, T. V., 1998. What's wrong with bonferroni adjustments. British Medical Journal 316.

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

Poole, W., 1967. Speculative prices as random walks: an analysis of ten time series of flexible exchange rates. Southern Economic Journal.

Pring, M. J., 1991. Technical analysis explained. New-York: McGraw-Hill.

Qi, M., Wu, Y., 2003. Nonlinear prediction of exchange rates with monetary fundamentals. Journal of Empirical Finance 10.

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 38.

Quandt, R. E., 1960. Tests of the hypothesis that a linear regression system obeys two separate regimes. Journal of American Statistical Association 55.

Raftery, A. E., Karny, M., Ettler, P., 2010. Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. Technometrics 52.

Rime, D., Sarno, L., Sojli, E., 2010. Exchange rate forecasting, order flow and macroeconomic information. Journal of International Economics 80.

Rogoff, K., 1996. The purchasing power parity puzzle. Journal of Economic Literature 34.

Rogoff, K., Stavrakeva, V., 2008. The continuing puzzle of short horizon exchange rate forecasting, Working Paper, National Bureau of Economic Research.

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

Rossi, B., 2006. Are exchange rates really random walk? some evidence robust to parameter instability. Macroeconomic Dynamics 10.

Rossi, B., 2013. Exchange rate predictability. Journal of Economic Literature 51.

Rossi, B., Inoue, A., 2012. Out-of-sample forecast test robust to the choice of window size. Journal of Business and Economic Statistics 30.

Rossi, B., Sekhposyan, T., 2011. Understanding models' forecasting performance. Journal of Econometrics 164.

Sager, M. J., Taylor, M. P., 2006. Under the microscope: the structure of the foreign exchange market. International Journal of Finance & Economics 11.

Sarno, L., Valente, G., 2009. Exchange rate and fundamentals: Footloose or evolving relationship? Journal of European Economic Association 7.

Skouras, S., 2001. Financial returns and efficiency as seen by an artificial technical analyst. Journal of Economic Dynamics and Control 25.

Stambaugh, R. F., 1999. Predictive regressions. Journal of Financial Economics 54.

Stock, J. H., Watson, M. W., 2001. A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series, Working Paper, SSRN.

Stock, J. H., Watson, M. W., 2004. Combination forecasts of output growth in a seven-country data set. Journal of Forecasting 23 (6).

Storey, J. D., 2003. The positive false discovery rate: a bayesian interpretation and the q-value. Annals of Statistics 31.

Storey, J. D., Taylor, J. E., Siegmund, D., 2004. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66.

Storey, J. D., Tibshirani, R., 2003. Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences 100.

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

Sweeney, R. J., 1986. Beating the foreign exchange market. The Journal of Finance 41.

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

Taylor, S. J., 1994. Trading futures using a channel rule: A study of the predictive power of technical analysis with currency examples. Journal of Futures Markets 14.

Theil, H., 1971. Applied Economic Forecasting. North Holland.

Thornton, D. L., Valente, G., 2012. Out-of-sample predictions of bond excess returns and forward rates: An asset allocation perspective. Review of Financial Studies 25 (10).

Timmermann, A., 2008. Elusive return predictability. International Journal of Forecasting 24.

Van Horne, J. C., Parker, G. G. C., 1967. The random-walk theory: an empirical test. Financial Analysts Journal 23.

Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21.

West, K. D., 1996. Asymptotic inference about predictive ability. Econometrica 64.

West, K. D., Edison, H. J., Cho, D., 1993. A utility-based comparison of some models of exchange rate volatility. Journal of International Economics 35.

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

Wright, J. H., 2008. Bayesian model averaging and exchange rate forecasts. Journal of Econometrics 146.