

# A Reconsideration of the Meese-Rogoff Puzzle: An Alternative Approach to Model Estimation and Forecast Evaluation

Kelly Burns\*

*Curtin University, Australia*

This study revisits the Meese-Rogoff puzzle by estimating the traditional monetary models of exchange rate determination in state-space form and comparing the accuracy of these forecasts against the naïve random walk model using a wide range of conventional and alternative measures of forecasting accuracy. The results demonstrate that incorporating stochastic movements in the parameters of exchange rate models does not enable the Meese-Rogoff puzzle to be overturned. However, estimating these models in state-space form substantially improves forecasting accuracy to the extent that the model and random walk produce an equivalent magnitude of error. Furthermore, the results prove that the Meese-Rogoff puzzle can be overturned if the forecasts are evaluated by alternative criteria. These criteria include direction accuracy, profitability, and measures that jointly take into account both magnitude and direction accuracy. (JEL: F31, F37, C53)

**Keywords:** forecasting; random walk; exchange rate models; time-varying parameters

**Article history:** Received: 13 June 2014, Received in final revised form: 3 November 2015, Accepted: 9 November 2015, Available online: 10 March 2016

## I. Introduction

In 1983, Meese and Rogoff published empirical results demonstrating

---

\* Corresponding author. Address: Curtin Graduate School of Business, Curtin University, 78 Murray Street, Perth, 6000, Australia. E-mail: [Kelly.Burns@curtin.edu.au](mailto:Kelly.Burns@curtin.edu.au). The author gratefully acknowledges the editor of this journal, two conference discussants, three anonymous referees and Professor Imad Moosa for useful comments on this paper.

(*Multinational Finance Journal*, 2016, vol. 20, no. 1, pp. 41–83)

© *Multinational Finance Society*, a nonprofit corporation. All rights reserved.

DOI: 10.17578/20-1-2

that the random walk cannot be outperformed by exchange rate models in terms of the magnitude of error in out-of-sample forecasting. One possible reason for these results is that static models of exchange rate determination cannot incorporate stochastic movements in the underlying parameters (Meese and Rogoff, 1983). The Meese and Rogoff (1983) results continue to stimulate significant research in the area and countless attempts have been made to overturn the findings using a variety of data, sample periods, methodologies and model specifications. Most of these attempts have, however, either been unsuccessful or fraudulent in their claims of success (in that appropriate testing was not conducted, or an augmented random walk component was introduced to the model). The conclusion now widely held by the profession is that the random walk model cannot be outperformed in forecasting exchange rates and that exchange rate models have little to no explanatory power. This is what is commonly known as the Meese-Rogoff puzzle.

This study employs a state-space model to assess the proposition that a possible reason for the failure of monetary models to outperform the random walk is the use of constant, rather than time-varying, parameters. A model estimated with constant coefficients is unable to incorporate stochastic movements in exchange rates and the explanatory macroeconomic variables. A time-varying parameter (TVP) model is a valid approach to compensate for structural changes, time-varying relationships and model instability in the context of exchange rate forecasting. With the advent of improvements in econometrics, TVP estimation has been employed by some economists to compensate for these possibilities (see, for example, Wolff, 1987; Schinasi and Swamy, 1989; Canova, 1993). There is some evidence that estimating monetary models in state-space form improves forecasting accuracy compared to static model forecasts (see, for example, Schinasi and Swamy, 1989; Rossi, 2006; Moosa and Burns, 2013a). However, these previous studies have several limitations that are addressed in this paper.

Unlike previous studies that use TVP estimation to generate exchange rate forecasts and revisit the Meese-Rogoff puzzle, this study is differentiated in three ways. First, conclusions reached about relative forecasting accuracy (in terms of the magnitude of error) are based on a statistical test of the numerical difference in the magnitude of error (measured using two different approaches: the AGS test and the Diebold and Mariano test). Although a small number of studies use the Diebold-Mariano test for this purpose, very few studies test whether the difference in the RMSEs is statistically significant (remembering that

Meese and Rogoff drew their conclusions based on a comparison of the numerical difference in the RMSEs, not the loss function defined by the Diebold-Mariano test). Any study that claims to outperform the random walk, without testing whether the difference in the RMSEs is statistically different, cannot legitimately claim to overturn the Meese-Rogoff puzzle. By formally testing whether the difference in magnitude only measures of forecasting accuracy is statistically significant, this study exposes the dangers and pitfalls of previous studies in the area. The results demonstrate that just because the random walk has a numerically smaller RMSEs, this does not mean that the random walk outperforms a competing model. What matters is whether the difference is statically significant. To confirm the robustness of hypothesis testing in relation to the difference in the RMSEs (using the AGS test), the Diebold-Mariano test, which tests for statistical difference in the loss function, is also used here.

The second way in which this study is differentiated to previous studies is that forecasting accuracy is assessed using several alternative measures that do not rely exclusively on the magnitude of the forecasting error.<sup>1</sup> While the author acknowledges that the use of alternative criteria to reconsider the Meese-Rogoff puzzle is not unique to this study, there are some important distinctions to the approach presented here. The small number of previous studies that have employed alternative measures of forecasting accuracy to reconsider the Meese-Rogoff puzzle employ one, or at most two, alternative measures. These are either profitability, direction accuracy, or both. A major problem with this approach is that direction accuracy and profitability are closely related, meaning that the findings drawn from either of these measures is likely to corroborate the other. Here, six such measures and tests (direction accuracy, the Pesaran and Timmerman (1992) test, mean returns, risk-adjusted returns, the adjusted root mean square error and proximity to perfect forecast) are used to demonstrate the proposition that the Meese-Rogoff puzzle can be overturned if forecasting accuracy is assessed by these alternative measures. The use of such a broad range of alternative measures of forecasting accuracy demonstrates the robustness of the proposition that the Meese-Rogoff puzzle can be outperformed if forecasting accuracy is assessed by measures that do not rely exclusively on the magnitude of error.

The third way in which this study makes a distinct contribution is by

---

1. The author is not aware of any other study that employs as many alternative measures of forecasting accuracy as presented here.

estimating the monetary models of exchange rate determination in state-space form using a more contemporary data sample. Most studies use a data sample from the 1970s and 1980s (see, for example, West, Edison and Cho, 1993; Wolff, 1987; Schinasi and Swamy, 1989) and a small number use data spanning up until the early 2000s (see, for example, Frommel, MacDonald and Menkhoff, 2005; Junttila and Korhonen, 2011). This study uses a more contemporary sample, spanning up to 2011 which therefore provides insight into the forecasting accuracy of TVP models well into the new millennium. Furthermore, while most studies utilise up to four bilateral exchange rates (and these are generally expressed against the USD), this study utilises six cross rates (only two of which are measured against the USD), thereby overcoming any dollar bias that may have existed in previous studies.

The results of this study show that stochastic movements in the underlying parameters of exchange rate models cannot explain the Meese-Rogoff puzzle. However, incorporating stochastic movements in the parameters brings about a substantial improvement in the magnitude of error of the model forecasts, to the extent that the model forecasts are as equally accurate as the random walk forecasts. More importantly, the study reveals that the random walk can be outperformed by any measure of forecasting accuracy that does not rely exclusively on the magnitude of error. It achieves this by being the first study to utilise such a large number of alternative measures of forecasting accuracy. Overall, the findings debunk the proposition that the inability of static models to incorporate time-varying relationships can explain the Meese-Rogoff results. It also contributes to the growing body of contemporary literature that employs alternative measures to assess the forecasting accuracy of exchange rate models.

In the next section, a discussion of the relevant literature is provided. A description of the methodology and data is presented in section III, and section IV describes the empirical findings. This is followed by a discussion of the results and conclusion.

## **II. Literature review**

The Meese-Rogoff puzzle refers to the widely accepted phenomena that the naïve random walk model cannot be outperformed by exchange rate models in out-of-sample forecasting. In their original study, Meese and

Rogoff (1983) analyse the relationship between exchange rates and contemporaneous macroeconomic fundamentals using a variety of models, including the Frenkel-Bilson model (flexible price monetary model), Dornbusch-Frankel model (sticky-price monetary model) and Hooper-Morton model (sticky-price monetary model incorporating current account effects).<sup>2</sup> Meese and Rogoff (1983) estimate the exchange rate models using various techniques, including using ordinary least squares (OLS), generalised least squares (GLS),<sup>3</sup> and Fair's (1970) instrumental variables technique (IV).<sup>4</sup> The forecasts generated by the models are compared against those of the naïve random walk, with and without drift, using the mean absolute error (MAE), mean square error (MSE) and root mean square error (RMSE). Based on a numerical comparison of these measures, Meese and Rogoff conclude that the random walk cannot be outperformed by monetary models in forecasting exchange rates. However, no formal testing of the statistical difference in these measures is performed.<sup>5</sup> Despite several attempts to overturn these results over the past 30 years, it is now widely accepted that the random walk model cannot be outperformed in forecasting exchange rates.

One of the reasons suggested in the literature for the failure of exchange rate models to outperform the random walk is stochastic movements in model parameters (see, for example, Meese and Rogoff, 1983; Schinasi and Swamy, 1989). Schinasi and Swamy (1989) present three compelling arguments to support this contention. First, model parameters can change over time because traders do not use information in the same way over all policy regimes and all time horizons. Second, market participants are heterogeneous, thus macroeconomic variables are not related to the exchange rate by a simple fixed-coefficient relationship. Third, the use of fixed coefficients implies the imposition of a restriction that may or may not be valid and empirical evidence suggests that a fixed coefficient restriction is invalid for macroeconomic

---

2. The sample comprised monthly data spanning the period 1973 to 1981 for three different bilateral exchange rates (USD/DEM, USD/JPY and USD/GBP). A rolling regression with a fixed sample size is used to generate out-of-sample forecasts from 1976 to 1981.

3. To correct for serial correlation in the error term.

4. To correct for simultaneous equation bias.

5. The AGS test suggested by Ashley, Granger and Schmalensee (1980) is available to test for the statistical difference in the RMSEs.

models (see, for example, Moosa and Kwiecien, 2002).<sup>6</sup>

Several studies support the proposition that the role played by macroeconomic fundamentals in driving exchange rate movements varies due to changing economic conditions. Junttila and Korhonen (2011), for example, use an error correction model with time-varying parameters to show empirically that the significance of the coefficients on the explanatory variables may differ according to the monetary regime. They find that the coefficient on relative interest rates has similar magnitudes, but different signs, depending on the current regime. Similarly, Wolff (1987) uses a Markov-switching model and suggests that parameter instability in exchange rate models may arise from instability in the money demand function, policy regime shifts and changes in the long-run real exchange rate (such as changes in oil prices or global trade patterns). Frommel, MacDonald and Menkhoff (2005) also use a Markov-switching framework and present evidence of an unstable relationship between macroeconomic fundamentals and exchange rates over time. They demonstrate that the impact of macroeconomic variables (such as short-term interest rates, inflation differentials and levels of economic growth) on exchange rates is closely related to the monetary policy regime and regime switches. Likewise, Rossi (2006) concludes that there is some sort of relationship between the exchange rate and its fundamentals, “but it is not stable over time.” Rossi (2006) suggests that if the nature of parameter instability is exploited, it may be possible to find economic models that forecast better than the random walk in terms of the magnitude of forecasting error. The proposition that the Meese-Rogoff puzzle can be overturned by incorporating stochastic movements in the parameters is explicitly considered in this study.

Time-varying parameters can be incorporated into exchange rate models through a variety of estimation procedures. Markov-switching models have been popular in this context because they capture the exchange rate adjustment process through a transition probability which is a function of the lagged deviation of the real exchange rate from its “equilibrium” level (see, for example, Frommel, MacDonald and Menkhoff, 2005; Engel, 1994). However, very few studies provide

---

6. Moosa and Kwiecien (2002) show that the nominal interest rate is more capable of predicting inflation if the assumption of fixed coefficients is relaxed. The implications of representing the Fisher equation by a fixed coefficient regression equation are that the real interest rate is fixed and that the response of the nominal interest rate to inflationary expectations does not change over time. These assumptions are unrealistic.

evidence that the improvement in forecasting accuracy relative to the random walk is statistically significant. For instance, Engel (1994) uses a Markov-switching model but is unable to outperform the random walk in terms of the magnitude of the error (although he does provide evidence indicating that the model is superior at predicting the direction of change in the exchange rate). Frommel, MacDonald and Menkhoff (2005) also use a Markov-switching model to forecast the DEM/USD, JPY/USD and GBP/USD, and report lower numerical RMSEs, relative to the random walk with drift. However, no formal testing of whether this difference is statistically significant is undertaken. Adopting a different approach, Canova (1993) uses a multivariate Bayesian time-varying coefficients autoregressive model because it has the benefit of utilising information from five spot rates (FFR, CHF, DEM, GBP and YEN vis-à-vis the USD) and the corresponding six interest rates. For all exchange rates considered, Canova claims to outperform the random walk based on a numerical comparison of conventional measures of magnitude of forecasting accuracy (Theil's U and MAD). Similar to the aforementioned studies, Canova (1993) fails to test whether the difference in the measures reported for the random walk and model is statistically different.

Adaptive expectation models have also been used to address time-variant issues in the context of exchange rate forecasting. These models largely became popular because they are relatively easy to employ. Molodstova and Papell (2009, 2013), for example, revisit the Meese-Rogoff puzzle using an adaptive expectation model. Using the test of superior predictive ability (SPA) to compare the out-of-sample predictability of the model and random walk forecasts, they conclude that the model outperforms the random walk. However, adaptive expectation models specify the current exchange rate to be a weighted average of the previous values of the exchange rate, plus an adjustment for any error in the previous periods forecast which, therefore, effectively nests the random walk model within the exchange rate determination model. Outperforming a random walk model with a model that nests the random walk model is tantamount to intellectual fraud and any study claiming to resolve the Meese-Rogoff puzzle using this approach should be entirely discounted.

Alternatively, some studies adopt the Kalman filter technique to incorporate time-varying parameters into exchange rate models. Yet again, almost all such studies neglect to test whether the difference in the RMSEs is statistically different. For instance, shortly after the

publication of the Meese and Rogoff results, Schinasi and Swamy (1989) re-worked the Meese-Rogoff results by re-estimating the same exchange rate models in a TVP framework using the Kalman filter technique. Based on a numerical comparison of conventional measures of forecasting accuracy, they report that the random walk is outperformed by exchange rate models estimated using TVP. They conclude that “while our results on fixed coefficient models support most of the Meese and Rogoff conclusions, we find that when coefficients are allowed to change, an important subset of conventional models.....can outperform forecasts of a random walk model”. However, Schinasi and Swamy (1989) do not test whether the numerical difference in the RMSEs of the random walk and exchange rate models is statistically different. Wolff (1987) similarly employs a recursive application of the Kalman filter to estimate the Frenkel-Bilson and Dornbusch-Frankel models in a TVP framework for USD/GBP, USD/JPY and USD/DEM using a sample from 1973 to the beginning of the floating exchange rate period in 1984. For the USD/DEM, Wolff (1987) presents evidence indicating that ex-post forecasts compare favourably well to that of the random walk, in that Theil’s U is less than one for both models at the 1, 3, 6, 12, 24 and 36 month forecasting horizon. However, Wolff (1987) also neglects to test whether the difference in the RMSEs of the model and random walk is statistically different.

Moosa and Burns (2013a) similarly replicate the Meese-Rogoff study using a TVP framework although, unlike the aforementioned studies, they formally test whether the difference in the RMSEs is statistically different.<sup>7</sup> Using the AGS test, their results show that the forecasts generated by exchange rate models cannot produce a statistically smaller RMSEs to the random walk.<sup>8</sup> Moosa and Burns (2013a) also assess forecasting accuracy using a small number of alternative criteria (including direction accuracy and profitability). They conclude that the random walk can be outperformed by exchange rate models if forecasting accuracy is measured in terms of the ability to

---

7. Moosa and Burns (2013a) use the same exchange rates and time period as used by Meese and Rogoff (1983), although they estimate the models by maximum likelihood using the Kalman filtering technique.

8. The author is not aware of any paper other than Moosa and Burns (2013a) where the AGS test is employed to determine if the improvement in the RMSE that arises from estimating the monetary model in a TVP framework is sufficient to outperform the random walk.

predict direction, measures that take into account magnitude and direction, and in terms of profitability. As sample selection is raised by Meese and Rogoff (1983) as one possible explanation for the puzzle, this study assesses the robustness of the Moosa and Burns (2013a) findings using a more contemporary sample period and alternative cross rates.

The use of alternative approaches to evaluate relative forecasting accuracy is not a novel or controversial approach. Cheung, Chinn and Pascual (2005) argue that using alternative criteria does not boil down to “changing the rules of the game.” In fact, there is a growing recognition that forecasting accuracy should be assessed by criterion that relates to the purpose of the forecast. Relying entirely on conventional measures of forecasting accuracy is inappropriate if it bears no relation to the purpose of the forecast and this approach may overlook important aspects of prediction, particularly at long horizons. Several authors advocate the use of other criteria because minimising the RMSE may not be important from an economic standpoint (see, for example, Christoffersen and Diebold, 1998; Cheung, Chinn and Pascual, 2005). Faust, Rogers and Wright (2003), for example, argue that “the absence of statistically significant predictive power need not indicate that an optimal decision-maker should ignore the model.” Moosa (2006) demonstrates that notions of forecasting accuracy are heterogeneous and should be defined and measured depending on the underlying decision rule, and that the most important consideration when assessing exchange rate forecasts is whether they lead to better decisions (Moosa, 2000).<sup>9</sup> Clearly, the methodology used to evaluate forecasting accuracy will impact conclusions reached about the relative accuracy of competing models in an out-of-sample forecasting exercise (Rossi, 2013). Corte, Lucio and Tsiakas (2009) actually attribute the Meese-Rogoff puzzle to the use of improper criteria that fail to take into account the real economic gains produced by the forecasts. Likewise, Moosa and Burns (2012a) argue that the random walk appears to be superior to exchange rate models solely because forecasting accuracy is measured by metrics that rely entirely on the magnitude of the forecasting error, while overlooking the ability of the models (and the random walk) to predict the direction of change.

Although economists have grappled with the inability of exchange

---

9. Moosa (2000) goes on to state that the forecasts should be accurate and timely, however a methodology for measuring or assessing these characteristics is not suggested.

rate models to outperform the random walk in terms of the magnitude of error for years, Moosa (2013) demonstrates that this finding is to be expected. He argues that as the period to period change in the spot rate is small – and this is, by definition, the magnitude of error of the random walk forecasts – the random walk will always have a very small magnitude of error, making it almost impossible for the model forecasts to generate a numerically smaller and statistically different RMSE. Similar reasoning put forward by Engel, Mark and West (2008) led them to conclude that beating the random walk in terms of the RMSE is “too strong a criterion for accepting a model”.

Based on these arguments, the question arises as to what other criteria are appropriate to assess the accuracy of exchange rate model forecasts. In some circumstances, the direction of change is the only criterion of importance when assessing forecasting accuracy. For instance, Engel (1994) suggests direction accuracy is the only relevant criterion in the case of central banks under a pegged exchange rate system, where intervention is required if the currency is expected to depreciate “regardless of the size of the expected depreciation.” Moosa (2006) suggests intraday trading (where the interest rate factor is negligible) as a further example where direction accuracy is the only important criterion. Alternatively, the magnitude of error can be the only important criterion in other circumstances (for example, betting on market volatility using straddles and strangles). There are also situations where both the magnitude of error, and direction accuracy, are important criteria (for example, carry trade).

Increasingly, direction accuracy is being used to assess the predictive power of exchange rate models. The reason for this is that a correct prediction of the direction of change is almost always more important than the magnitude of the error in the context of forecasting exchange rate movements (Engel, 1994; Engel and Hamilton, 1990). Several studies have unquestionably demonstrated that the random walk model can be outperformed by exchange rate models in terms of the direction of change (Boothe and Glassman, 1987; Engel, 1994; Abhyankar, Sarno and Valente, 2005; Cheung, Chinn and Pascual, 2005; Corte, Lucio and Tsiakas, 2009). As exchange rate forecasts are an input in financial decision making, another important test of forecasting accuracy is the ability to generate profit by basing decisions on these forecasts (such as currency trading). For instance, West, Edison and Cho (1993) propose a utility-based evaluation of exchange rate predictability (utility in this context is measured by profitability).

Overall, the literature lends support to the superiority of model forecasts when profitability is used to assess forecasting performance. Abhyankar, Sarno and Valente (2005), for example, find that the relative performance of exchange rate models improves when profit-based criterion are used. Li (2011) also finds that the profitability of carry trade and risk-return measures is enhanced by using model forecasts over and above the random walk forecasts.

Although economists have moved towards assessing exchange rate forecasts in terms of both the magnitude of error and direction accuracy, the literature almost exclusively examines these characteristics in isolation to each other. However, Moosa and Burns (2012a) propose a measure of forecasting accuracy that takes into account both magnitude and direction accuracy, without bias to either (the adjusted root mean square error or ARMSE). Using simulated data, they show that when magnitude and direction are jointly considered, the random walk can be outperformed by static exchange rate models. In a subsequent study, Moosa and Burns (2013a) use the same data sample as the one utilised in Meese and Rogoff (1983) to demonstrate that the random walk can be outperformed by exchange rate models estimated with both stochastic and time-varying coefficients if forecasting accuracy is assessed using the ARMSE.

Another method for jointly testing magnitude and direction accuracy is by measuring the proximity of the forecasts to a perfect forecast. Moosa and Burns (2013b) propose running a regression on the predicted change against the actual change and using a Wald test to determine the proximity of the forecasts to a perfect forecast (a set of perfect forecasts is represented by the general equation  $\hat{S}_t - S_{t-1} = \alpha + \beta(S_t - S_{t-1})$  with a general restriction that  $(\alpha, \beta) = (0,1)$ ). The numerical value of the Wald test statistic is indicative of the deviation of the line of best fit of the model forecasts from a perfect forecast, invariably involving magnitude and direction errors. As the test statistic is indicative of the deviation of the forecasts from a set of perfect forecasts, the numerical value of the Wald test statistic for this joint coefficient restriction can be compared (the smaller the numerical value of the Wald test statistic, the closer the forecasts are to a perfect fit) and a conclusion reached as to which model produces the better forecasts in terms of magnitude and direction accuracy. Using a sample of data from 1990 to 2010, Moosa and Burns (2013b) show that the random walk can be outperformed by the static version of the flexible price monetary model if forecasting accuracy is measured by the proximity of the model forecasts to a perfect forecast.

### III. Data and methodology

#### *A. Model estimation*

The approach adopted in this study is to estimate the three monetary models exchange rate determination used originally by Meese and Rogoff (1983) in state-space form. Following the methodology of Koopman, Shephard and Doornik (1999),<sup>10</sup> Koopman et al. (2006),<sup>11</sup> Moosa (2006) and Mendelssohn (2011), these models are estimated with time-varying parameters using the Kalman filter technique and incorporating unobserved components.<sup>12</sup> There are several motivations underpinning this approach. First, the incorporation of unobserved components provides a generalised model in which explanatory variables that do not appear explicitly on the right-hand side of the equation are accounted for by the trend and cycle components. It follows that the traditional specification of exchange rate determination models is adequate for the purposes of this study because unobserved components can account for any missing variables. Second, the results of the unit root tests show that, for the sample used in this study, the exchange rates are not co-integrated with the macroeconomic variables because the variables are not integrated of the same order (tables 1 and 2). Therefore, an error correction model (although legitimately used in other studies, such as Rapach and Wohar, 2002) is unsuitable for this study. Third, due to the lengthy sample used here, structural breaks are present in the data.<sup>13</sup> Evidence shows that allowing for time-varying parameters in the presence of multiple structural breaks produces statistically significant gains in forecasting accuracy (see, for example, Barnett, Mumtaz and Theodoridis, 2012; Lu and Ito, 2008). Fourth, in order to allow for any non-stationarity in the time series, the various

---

10. Koopman, Shephard and Doornik (1999) demonstrates how a model is represented in state-space form.

11. Koopman et al. (2006) describes the statistical treatment of the structural time series model (represented by equations (1) to (3)) based on the state-space form.

12. A full description of the specification of the measurement and transition equations for structural time series models, which form the state-space representation required for TVP estimation by the Kalman filter, can be found in Moosa (2006). TVP estimation is carried out by using the STAMP software (Structural Time Series Analyser, Modeller and Predictor).

13. The results of the Bai and Perron (1998) test for global multiple breaks show that the time series used in this study are characterised by multiple break points. These results are available from the author upon request.

components of the model are characterised as stochastic time-varying parameters (Young, Diego and Wlodek, 1999). Fifth, there are a range of studies that advocate using a time-varying parameter model as a valid approach to modelling and forecasting exchange rates (see, for example, Schinasi and Swamy, 1989; Wolff, 1987; Moosa and Burns, 2013a). Sixth, diagnostic testing of the residuals indicates that the model is a good fit.<sup>14</sup>

To estimate the Frenkel-Bilson, Dornbusch-Frankel and Hooper-Morton models in state-space form with unobserved components, the equations are represented as:

$$s_t = \mu_t + \varphi_t + \alpha_{1t}(m_t - m_t^*) + \alpha_{2t}(y_t - y_t^*) + \alpha_{3t}(i_t - i_t^*) + \varepsilon_t \quad (1)$$

$$s_t = \mu_t + \varphi_t + \alpha_{1t}(m_t - m_t^*) + \alpha_{2t}(y_t - y_t^*) + \alpha_{3t}(i_t - i_t^*) + \alpha_{4t}(q_t^e - q_t^{e*}) + \varepsilon_t \quad (2)$$

$$s_t = \mu_t + \varphi_t + \alpha_{1t}(m_t - m_t^*) + \alpha_{2t}(y_t - y_t^*) + \alpha_{3t}(i_t - i_t^*) + \alpha_{4t}(q_t^e - q_t^{e*}) + \alpha_{5t}B_t + \alpha_{6t}B_t^* + \varepsilon_t \quad (3)$$

where  $s_t$  is measured as the log of the domestic per foreign exchange rate (i.e. the price of the foreign currency in terms of the domestic currency),  $m$  is the log of the money supply,  $y$  is the log of real income,  $i$  is the short-term interest rate,  $q^e$  is the long-run expected inflation rate and  $B$  is the trade balance (an asterisk indicates the variable from a foreign perspective) and  $\varepsilon_t$  is the error term. The time series components of  $s_t$  are the trend ( $\mu_t$ ), the cyclical component ( $\varphi_t$ ) and the random component ( $\varepsilon_t$ ).

The trend variable ( $\mu_t$ ) represents the long-term component in the series and indicates the general direction of movement of the dependent variable. The trend consists of two parts, level and slope, both of which are specified to be stochastic. The level and slope are constructed by putting greater weight on more recent observations. The stochastic trend is represented by the general specification of the level and slope,

---

14. The results of the diagnostic testing are presented in tables 4 to 8 and discussed in section IV, part B.

TABLE 1. Results for unit root tests (data in log levels)

Variable		ADF	PP	KPSS	ERS Point Optimal	Ng-Perron Modified
CAD/AUD	Spot rate	-3.55***	-3.56***	0.35	8.70	-3.29
	Relative interest	-3.02**	-3.08**	0.35	3.91*	-7.22*
	Relative inflation	-2.35	-3.00**	0.39*	3.06**	-8.06*
	Relative money	-1.12	-1.20	1.46***	59.68	-0.25
	Relative income	-0.28	-0.50	1.43***	14.81	0.32
CAD/GBP	Spot rate	-2.00	-2.02	0.37*	6.50	-4.06
	Relative interest	-4.66***	-4.05***	0.07	1.80***	-12.09**
	Relative inflation	-1.65	-2.77*	0.21	5.34	-4.89
	Relative money	-0.63	-1.19	0.50**	20.96	-1.09
	Relative income	0.98	1.14	1.91***	184.15	2.20
CAD/USD	Spot rate	-0.58	-0.68	0.47**	9.26	-2.11
	Relative interest	-1.72	-2.14	1.02***	4.13*	-5.92*
	Relative inflation	-2.23	-4.05***	0.35	8.56	-2.20
	Relative money	-0.84	-0.47	2.01***	1.07	1.24
	Relative income	-0.62	-0.68	0.49**	11.57	-1.32
GBP/AUD	Spot rate	-2.04	-2.08	0.38*	17.46	-1.56
	Relative interest	-2.39	-2.57*	0.44*	2.42**	-11.72**
	Relative inflation	-2.23	-2.19	0.24	2.54**	-10.29**
	Relative money	-0.63	-0.79	1.49***	16.48	-0.56
	Relative income	-0.12	-1.18	1.96***	106.99	1.09

( Continued )

TABLE 1. (Continued)

Variable		ADF	PP	KPSS	ERS Point Optimal	Ng-Perron Modified
GBP/USD	Spot rate	-2.85**	-2.94**	0.41*	4.30*	-6.37*
	Relative interest	-2.55*	-2.26	0.76***	4.34*	-6.01*
	Relative inflation	-1.67	-2.19	0.35	8.91	-2.52
	Relative money	-1.76	-1.77	1.97***	114.77	0.34
	Relative income	0.37	-0.04	2.10***	193.55	1.50
	Trade balance					
	Canada	-1.94	-1.79	0.35	1.64***	-21.44***
	Australia	-1.85	-2.23	1.50***	6.17	-3.67
	United Kingdom	-2.35	-5.37***	1.26***	8.58	-1.27
	Japan	-2.57*	-4.26***	1.63***	9.20	-1.20
	U.S.A	-1.25	-1.15	1.67***	15.77	-1.24

**Note:** For the ADF test, the lag length was selected by using SIC (Modified SIC for ERS & Ng-Perron) values. For the PP and KPSS tests, the optimal bandwidth was selected by Newey-West method using Bartlett kernel. All unit root tests are performed with the assumption of constant term in the logarithm of the series, with the null hypothesis of unit root for all tests except for KPSS test where the null hypothesis is stationarity. The spectral estimation method for the ERS point optimal test and Ng-Perron modified unit root test is AR spectral OLS and AR GLS (detrended), respectively. An \*, \*\*, \*\*\* denote statistical significance at the 10, 5 and 1 per cent levels, respectively.

respectively:

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \quad (4)$$

$$\beta_t = \beta_{t-1} + \zeta_t \quad (5)$$

where  $\eta_t \sim NID(0, \sigma_\eta^2)$  and  $\zeta_t \sim NID(0, \sigma_\zeta^2)$ .

The stochastic cycle ( $\varphi_t$ ) is constructed from a sine-cosine wave which is impacted by shocks from disturbances and by a damping factor. The statistical specification of the cyclical component is as follows:

$$\varphi_t = \rho(\varphi_{t-1} \cos \theta + \varphi_{t-1}^* \sin \theta) + \omega_t \quad (6)$$

$$\varphi_t^* = \rho(-\varphi_{t-1} \sin \theta + \varphi_{t-1}^* \cos \theta) + \omega_t^* \quad (7)$$

TABLE 2. Results for unit root tests (first differenced variables)

Variable	ADF	PP	KPSS	ERS Point Optimal	Ng-Perron Modified	
CAD/AUD	Spot rate	-	-	-	0.33***	-34.27***
	Relative interest	-	-	-	0.17***	-12.40**
	Relative inflation	-14.84***	-	0.05	-	0.35
	Relative money	-10.78***	-19.30***	0.34	0.34***	-54.91***
	Relative income	-14.00***	-21.90***	0.35	0.30***	21.11***
CAD/GBP	Spot rate	-18.92***	-18.91***	-	0.89***	-4.02
	Relative interest	-	-	-	-	-
	Relative inflation	-8.88***	-18.01***	-	4.04*	-11.36**
	Relative money	-16.32***	-16.90***	0.38*	0.23***	-38.63***
	Relative income	-22.76***	-22.87***	0.43*	0.55***	0.91
CAD/USD	Spot rate	-18.54***	-18.54***	0.28	0.17***	-163.88***
	Relative interest	-11.24***	-20.75***	0.09	0.02***	-1,279.95***
	Relative inflation	-10.46***	-	-	13.46	-1.97
	Relative money	-6.28***	-18.00***	0.12	1.07***	-6.95*
	Relative income	-22.90***	-22.31***	0.27	1.81***	-0.79
GBP/AUD	Spot rate	-17.41	-17.40***	0.45*	0.33***	-4.14
	Relative interest	-4.46***	-17.09*	0.09	-	-
	Relative inflation	-8.45***	-15.41***	-	-	-
	Relative money	-17.51***	-17.51***	0.13	0.20***	106.71***
	Relative income	-12.93***	-20.80***	0.06	0.06***	-6,088.85***

(Continued)

TABLE 2. (Continued)

Variable		ADF	PP	KPSS	ERS Point Optimal	Ng-Perron Modified
GBP/JPY	Spot rate	-16.08***	-16.09***	0.10	0.64***	-1.82
	Relative interest	-12.92***	-12.76***	0.15	0.17***	-146.89***
	Relative inflation	-8.27***	-17.24***	-	-	-
	Relative money	-18.85***	-18.89***	-	0.16***	-72.54***
	Relative income	-21.16***	-	-	0.50***	-4.07
	Spot rate	-	-	0.05	0.87***	-0.62
GBP/USD	Relative interest	-12.92***	-12.77***	0.12	0.18***	-141.53***
	Relative inflation	-9.43***	-14.90***	-	6.89	-9.71**
	Relative money	-7.49***	-17.23***	0.30	0.38***	-65.88***
	Relative income	-11.20***	-24.10***	0.09	0.14***	-1.12
	Canada	-4.49***	-7.20***	-	-	-
	Australia	-6.73***	-7.89***	0.06	25.58	-11.14**
Trade balance	United Kingdom	-5.37***	-	0.09	13.19	-0.06
	Japan	-5.47***	-	0.05	14.34	-0.70
	U.S.A	-4.95***	-4.89***	0.10	0.64***	-10.90**

**Note:** For the ADF test, the lag length was selected by using SIC (Modified SIC for ERS & Ng-Perron) values. For the PP and KPSS tests, the optimal bandwidth was selected by Newey-West method using Bartlett kernel. All unit root tests were performed with the assumption of constant term in the logarithm of the series, with the null hypothesis of unit root for all tests, except for the KPSS test, where the null hypothesis is stationarity. The spectral estimation method for the ERS point optimal test and Ng-Perron modified unit root test is AR spectral OLS and AR GLS (detrended), respectively. An \*, \*\*, \*\*\* denote statistical significance at the 10, 5 and 1 per cent levels, respectively. Where the null hypothesis is rejected for the time series in levels at the 5 per cent level, no further unit root tests were performed on the data in first differences.

where  $\varphi_t^*$  appears by construction such that  $\omega_t$  and  $\omega_t^*$  are uncorrelated white noise disturbances with variances  $\sigma_\omega^2$  and  $\sigma_{\omega^*}^2$ , respectively. The frequency of the cycle and the damping factor on the amplitude are  $0 \leq \theta \leq \pi$  and  $0 \leq \rho \leq 1$ , respectively. The period of the cycle, which is the time taken by the cycle to go through its complete sequence of values, is  $2\pi/\theta$  (Koopman et al., 2006).

The Kalman filter (KF) is a set of vectors and matrix recursions which are used to estimate the unknown parameters via recursive estimation so as to minimise the MSE. In this sense, the KF plays the same role in state-space modelling as least squares estimation does for a conventional regression model. The KF can take a variety of forms and the form used in this study is represented as:

$$a_{t|t-1} = E(\alpha_t | Y_{t-1}, \delta = 0) \quad (8)$$

$$\sigma^2 P_{t|t-1} = E\{(\alpha_t - a_{t|t-1})(\alpha_t - a_{t|t-1})' | Y_{t-1}, \delta = 0\} \quad (9)$$

This is referred to as the mean and the MSE of the state space respectively, given past information and setting  $\delta$  to zero.<sup>15</sup> Filtering and smoothing algorithms are then applied on known state and error system matrices. The unknown values within the matrices are treated as parameters to be estimated using maximum likelihood estimation.

Once the model is estimated in state-space form using the aforementioned methodology, a recursive regression is used to generate one-step ahead forecasts at the one month horizon, meaning that the model uses all available information to generate the model forecasts (which is in line with what is recommended by Marcellino, 2002, and Marcellino, Stock and Watson, 2003). In out-of-sample forecasting, the model is estimated over part of the sample period,  $t=1,2,\dots,k$ , then a one-period-ahead forecast is generated for the point in time  $k+1$ . The process is then repeated by estimating the model over the period  $t=1,2,\dots,k+1$  to generate a forecast for the point in time  $k+2$ ,  $\hat{s}_{k+2}$ , and so on until  $\hat{s}_n$ , where  $n$  is the total sample size. The forecast log exchange rate generated using the Frenkel-Bilson model is:<sup>16</sup>

$$\hat{s}_t = \hat{\mu}_t + \hat{\varphi}_t + \hat{\alpha}_{1t}(m_t - m_t^*) + \hat{\alpha}_{2t}(y_t - y_t^*) + \hat{\alpha}_{3t}(i_t - i_t^*) \quad (10)$$

Hence the forecast level of the exchange rate is:

$$\hat{S}_t = \exp(\hat{s}_t) \quad (11)$$

---

15. These terms are corrected by the use of an augmented KF to allow for  $\delta$  not equal to zero.

16. For simplicity, only the Frenkel-Bilson model is described here.

**TABLE 3. Estimates drift factor and statistical significance**

CAD/AUD	GBP/AUD	CAD/GBP	CAD/USD	GBP/USD	GBP/JPY
0.28	0.59**	-0.25	-0.35	-0.06	0.34
(0.2817)	(0.0366)	(0.3616)	(0.1854)	(0.8519)	(0.3195)

**Note:** P-values are in parentheses. An \*, \*\*, \*\*\* denote statistical significance at the 10, 5 and 1 per cent level, respectively.

Turning now to the naïve random walk model, this model is represented as follows:

$$s_t = s_{t-1} + \varepsilon_t \quad (12)$$

or alternatively:

$$s_t - s_{t-1} = \Delta s_t = \varepsilon_t \quad (13)$$

where  $\varepsilon_t$  is completely random and displays no pattern over time. The time series follows a random walk with drift if the change between the current and previous period spot rate is equal to a drift factor ( $d$ ), plus the purely random component  $\varepsilon_t$ . In this case, the random walk with drift is represented as:

$$s_t = d + s_{t-1} + \varepsilon_t \quad (14)$$

$$s_t - s_{t-1} = \Delta s_t = d + \varepsilon_t \quad (15)$$

Following the approach of Engel and Hamilton (1990), Engel (1994), and Moosa and Burns (2013a), the choice between the random walk without and with drift depends on the statistical significance of the drift factor. If the drift factor is significant in-sample, then the appropriate out-of-sample standard is the random walk with drift. As per the Meese and Rogoff (1983) approach, the drift factor of the random walk is estimated as the average value of the percentage change in the exchange rate. As demonstrated in table 3, the drift factor is only significant in one instance.

### *B. Conventional measures of forecasting accuracy*

Once the time series for the forecast exchange rates ( $\hat{s}_t$ ) are obtained for the exchange rate models and naïve random walk model, a variety of forecasting accuracy measures are calculated. First, the root mean

square error (RMSE) is used to assess the magnitude of error of the exchange rate forecasts. The RMSE of the model forecasts is calculated as follows:

$$RMSE = \sqrt{\frac{1}{n-k} \sum_{t=k+1}^n (\hat{s}_t - s_t)^2} \quad (16)$$

where  $s_t$  is the realised exchange rate at time  $t$  and  $\hat{s}_t$  is the forecast of the exchange rate at time  $t$ . The root mean square error of the random walk is calculated as:

$$RMSE = \sqrt{\frac{1}{n-k} \sum_{t=k+1}^n (s_{t-1} - s_t)^2} \quad (17)$$

Unlike Meese and Rogoff (1983), a formal test for the statistical difference in the RMSE of the model and the random walk is undertaken using the AGS test, as suggested by Ashley, Granger and Schmalensee (1980). This test requires estimation of the linear regression:

$$D_t = \alpha_0 + \alpha_1 (M_t - \bar{M}) + u_t \quad (18)$$

where  $D_t = w_{1t} - w_{2t}$ ,  $M_t = w_{1t} + w_{2t}$ ,  $\bar{M}$  is the mean of  $M$ ,  $w_{1t}$  is the forecasting error at time  $t$  of the model with the numerically higher RMSE and  $w_{2t}$  is the forecasting error at time  $t$  of the model with the numerically lower RMSE. If the sample mean of the errors is negative, the observations of the series are multiplied by  $-1$  prior to running the regression. The estimates of the intercept term ( $\alpha_0$ ) and the slope ( $\alpha_1$ ) from equation (18) are used to test the statistical difference between the RMSEs of two different models. If  $\alpha_0$  and  $\alpha_1$  are both positive, then a Wald test of the joint null hypothesis  $H_0 : \alpha_0 = \alpha_1 = 0$  is appropriate.<sup>17</sup>

The Diebold and Mariano (1995) test is employed to evaluate whether the forecasts produced by the random walk and exchange rate model exhibit equality in their respective forecasting errors, as measured by a loss function. To calculate the Diebold and Mariano test

---

17. However, if one of the estimates is negative and statistically significant, then the test is inconclusive. But if one of the coefficients is negative and statistically insignificant the test remains valid (although significance is determined by the upper-tail of the t-test on the positive coefficient estimate).

statistic, the accuracy of each forecast is measured by the loss function:

$$L(s_{t+1}, s_{t+1|t}^i) = L(\varepsilon_{t+1|t}^i) \quad (19)$$

where  $s$  is the log of the exchange rate,  $\varepsilon$  is the error term and  $i$  takes a value of 1 or 2 and represents the model and random walk forecasts, respectively. The null hypothesis that the models have equal predictive accuracy is represented as:

$$H_0 : E[L(\varepsilon_{t+1|t}^1)] = E[L(\varepsilon_{t+1|t}^2)] \quad (20)$$

or alternatively:

$$H_0 : E[r_t] = 0 \quad (21)$$

where the loss differential is defined as:

$$r_t = L(\varepsilon_{t+1|t}^1) - L(\varepsilon_{t+1|t}^2) \quad (22)$$

The Diebold-Mariano test statistic is calculated as:

$$DM = \frac{\bar{r}}{\left(\widehat{LRV}_{\bar{r}} / (n-k)\right)^{1/2}} \quad (23)$$

where:

$$\bar{r} = \frac{1}{n-k} \sum_{t=k+1}^n r_t \quad (24)$$

and:

$$LRV_{\bar{r}} = \gamma_0 + 2 \sum_{j=k+1}^n \gamma_j \quad (25)$$

where  $\gamma_0$  is the variance of the loss differential ( $r_t$ ) and  $\gamma_j$  is the covariance between  $r_t$  and  $r_{t-j}$ . The null hypothesis of equal predictive accuracy is rejected at the 5 per cent level if  $|DM| > 1.96$ . A two-tailed test is used such that the alternative hypothesis is that the expected value of the loss differential is non-zero. If the null hypothesis is

rejected, the sign of the test statistic indicates whether the model forecasts produce a statistically larger or smaller loss differential.<sup>18</sup>

### C. Direction accuracy

Direction accuracy is defined as the proportion of times the forecast correctly predicts the direction of change of the exchange rate. The calculation of direction accuracy takes the form:

$$DA = \frac{1}{n-k} \sum_{t=k+1}^n \alpha_t \quad (26)$$

where:

$$\alpha = \begin{cases} 1 & \text{if } (\hat{s}_{t+1} - s_t)(s_{t+1} - s_t) > 0 \\ 0 & \text{if } (\hat{s}_{t+1} - s_t)(s_{t+1} - s_t) < 0 \end{cases} \quad (27)$$

Whether the direction accuracy of the model forecasts exceeds that of the random walk is tested under the null hypothesis  $H_0 : DA = \theta$  against the alternative  $H_1 : DA > \theta$ .<sup>19</sup> The test statistic is as follows:

$$z = \frac{DA - \theta}{\sqrt{DA(1-DA)/(n-k)}} \quad (28)$$

Initially, the parameter is set to  $\theta = 0$ , such that the test becomes that of the model against the random walk without drift (because this is a no

18. That is, a negative test statistic which is statistically significant indicates that the model outperforms the random walk (and vice versa). A two-tailed test is preferable over a one-tailed test because it enables a determination of whether the model or random walk have equal accuracy and, if not, which is superior in terms of forecasting accuracy as measured by the loss differential.

19. If forecasting accuracy is to be assessed according to direction accuracy, the question arises as to what is the appropriate benchmark value against which the direction accuracy of forecasts should be assessed. Most studies that include direction accuracy as a criterion to assess forecasting performance consider whether or not direction accuracy ( $DA$ ) exceeds zero, implying that the benchmark is the random walk without drift. In contrast, Cheung, Chinn and Pascual (2005) erroneously use the higher benchmark of  $DA = 0.5$  to assess the direction accuracy of the forecasts on the basis that the random walk “predicts the exchange rate has an equal chance to go up or down”. Evans and Lyons (2005) state explicitly that if the ex-ante forecast follows a random walk without drift, “there is no forecast change in the spot rate.” Hence, a random walk without drift has nil direction accuracy and the appropriate null hypothesis to be tested is  $H_0 : DA = 0$  rather than  $H_0 : DA = 0.5$ .

TABLE 4. ADF test for stationarity in the model residuals

	CAD/AUD	GBP/AUD	CAD/GBP	CAD/USD	GBP/USD	GBP/JPY
Frenkel-Bilson H(107)	-11.55***	-12.05***	-13.12***	-10.21***	-10.66***	-12.37***
Dornbusch-Frankel H(106)	-11.41***	-12.28***	-12.72***	-9.98***	-11.11***	-12.07***
Hooper-Morton H(105)	-12.47***	-10.88***	-13.19***	-11.99***	-10.04***	-12.50***

**Note:** The lag length was selected by using SIC (Modified SIC for ERS & Ng-Perron) values. The unit root tests are performed with the assumption of constant term in the logarithm of the series, with the null hypothesis of unit root. An \*, \*\*, \*\*\* denote statistical significance at the 10, 5 and 1 per cent levels, respectively.

change forecast and, therefore, has nil direction accuracy). As a supplementary test, the parameter is also set at  $\theta = 0.5$  to assess if the model correctly predicts the direction of change on at least 50 per cent of occasions.

The Pesaran and Timmermann (1992) test is used as a further test of direction accuracy of the model forecasts. The PT test is a non-parametric test of association between two variables that is used to test for independence between the predicted and actual change in the exchange rate. The null hypothesis is that there is no predictable relationship between the predicted and actual change (that is, the variables are statistically independently distributed). The PT test statistic is calculated as follows:

$$PT = \left\{ \frac{P_*(1-P_*)}{n-k} \right\}^{-1/2} (\hat{P} - P_*) \quad (29)$$

where  $\hat{P}$  is the proportion of times that the direction of change is predicted correctly,  $P_* = \Pr(b_i \cdot c_i > 0)$  and  $b_i$  and  $c_i$  represent the actual and predicted change, respectively.

#### D. Magnitude and direction accuracy

The adjusted root mean square error (ARMSE) is constructed by adjusting the conventional RMSE to take into account direction

**TABLE 5. Breusch-Pagan test for heteroscedasticity**

	CAD/AUD	GBP/AUD	CAD/GBP	CAD/USD	GBP/USD	GBP/JPY
Frenkel-Bilson	0.83	0.39	0.57	6.31*	0.53	0.92
Dornbusch-Frankel	0.81	0.42	0.53	6.23*	0.50	0.91
Hooper-Morton	0.18	0.07	0.58	0.50	0.49	0.88

**Note:** Test statistic is distributed approximately as  $F(h,h)$ . The value of  $h$  is 107, 106 and 105 for the Frenkel-Bilson, Dornbusch-Frankel and Hooper-Morton models, respectively. The critical value is 1.38 at the 5 per cent level. An asterisk indicates statistical significance at the 5 per cent level.

accuracy. If two models have equal RMSEs, the model with the higher confusion rate (CR) has a higher ARMSE. The formula for the ARMSE is as follows:

$$ARMSE = \sqrt{\frac{CR}{n-k} \sum_{t=k+1}^n (\hat{s}_t - s_t)^2} \quad (30)$$

where CR is the confusion rate, calculated as:

$$CR = 1 - DA \quad (31)$$

A useful property of the ARMSE as defined by equation (30) is that it is not biased towards measures of either magnitude (RMSE) or direction (CR).<sup>20</sup> Despite the unavailability of a formal test for whether the difference between two ARMSEs is statistically different, if either the corresponding differences in the RMSEs or  $DA$  are statistically different, it can be inferred that the ARMSEs are also statistically different.

Following Moosa and Burns (2013b), the proximity of the forecasts to a perfect forecast is also tested. This approach is used to jointly assess forecasting accuracy in terms of the magnitude of error and direction accuracy. Errors in magnitude and direction are jointly

---

20. By using simulated data, Moosa and Burns (2012a) show that the rank correlation between ARMSE and RMSE and between ARMSE and CR are close in value at 0.571 and 0.551, respectively.

TABLE 6. Durbin-Watson test statistics

	CAD/AUD	GBP/AUD	CAD/GBP	CAD/USD	GBP/USD	GBP/JPY
Frenkel-Bilson	1.99	2.13	2.18	1.98	2.05	2.02
Dornbusch-Frankel	1.99	2.17	2.14	1.95	2.09	2.05
Hooper-Morton	2.73	2.49	2.73	2.80	2.08	2.04

measured by the deviation of the line of best fit associated with a regression of the predicted change on the actual change, compared to a perfect forecast. The line of best fit has the general equation  $\hat{S}_t - S_{t-1} = \alpha + \beta(S_t - S_{t-1})$  and a set of perfect forecasts has the restriction that  $(\alpha, \beta) = (0, 1)$ . Any violation of the coefficient restrictions defining the line of perfect forecast implies less than perfect forecasts, invariably involving magnitude and direction errors. A Wald test of the joint coefficient restriction is conducted to determine if the violation is statistically significant as implied by the  $\chi^2$  statistic. If both the model and random walk violate this condition, relative forecasting superiority can be assessed by comparing the numerical value of the  $\chi^2$  statistic. That is, the bigger the value of the Wald test statistic, the greater the violation of the coefficient restriction and the worse the model is with respect to predictive power, as judged by magnitude and direction. For the random walk to outperform the model it must produce a smaller test static relative to the model.<sup>21</sup> Evans and Lyons (2005) propose a similar test which only assesses whether  $\beta = 1$ . The test proposed by Moosa and Burns (2013b) has a higher threshold and is therefore adopted in this study.

### E. Profitability

In this study, profitability is measured by mean returns and the Sharpe ratio. Profitability is the rate of return generated from period-by-period trading using two alternative strategies: pure carry trade and forecasting-based trading. In forecasting-based trading, the investment

21. Following the methodology of Moosa and Burns (2013b), the forecasts of the random walk model must be estimated using OLS so that a Wald test can be conducted.

TABLE 7. Box-Ljung Q-statistics (16,14)

	CAD/AUD	GBP/AUD	CAD/GBP	CAD/USD	GBP/USD	GBP/JPY
Frenkel-Bilson	13.66	17.79	45.07*	27.12*	13.35	14.59
Dornbusch-Frankel	14.80	13.93	45.96*	26.95*	11.09	13.06
Hooper-Morton	101.99*	43.83*	66.17*	72.77*	13.47	9.95

**Note:** The Box-Ljung Q-statistic is based on the first 16 residual autocorrelations and is distributed approximately as a chi squared distribution with 14 degrees of freedom. The critical value at the 5 per cent level is 23.68. An asterisk indicates statistical significance at the 5 per cent level.

decision depends on the forecast percentage change in the exchange rate and the expected return. By taking a short position on currency  $x$  and a long position on currency  $y$ , the expected return,  $f^e$ , is defined as:

$$f^e = (i_{y,t} - i_{x,t}) + \hat{S}_{t+1}^e \quad (32)$$

where  $\hat{S}^e$  is the expected percentage change in the exchange rate,  $i_y$  is the interest rate on currency  $y$  and  $i_x$  is the interest rate on currency  $x$ . The decision rule is to take a short position on  $x$  and a long position on  $y$  if the expected return is positive, and vice versa. The realised return from forecasting-based trading is given by:

$$f = \begin{cases} (i_{y,t} - i_{x,t}) + \hat{S}_{t+1}^e & \text{if } f^e > 0 \\ (i_{x,t} - i_{y,t}) - \hat{S}_{t+1}^e & \text{if } f^e < 0 \end{cases} \quad (33)$$

In a carry trade exercise where the drift factor is insignificant, the investment decision depends on the interest rate differential alone (since the random walk implies that  $\hat{S}_{t+1}^e = 0$ ). This involves taking a long position on the high interest currency and a short position on the low interest currency. The realised return on a carry trade operation is therefore given by:<sup>22</sup>

22. Where the drift factor is statistically significant, profitability is calculated on a modified carry trade operation where the assumption is that the exchange rate is expected to

**TABLE 8. Normality test statistics**

	CAD/AUD	GBP/AUD	CAD/GBP	CAD/USD	GBP/USD	GBP/JPY
Frenkel-Bilson	10.58*	26.75*	12.20*	164.11*	12.57*	19.09*
Dornbusch-Frankel	8.85*	27.14*	5.63	157.79*	10.69*	24.80*
Hooper-Morton	65.46*	37.81*	4.96	13.30*	1.01	23.76*

**Note:** The normality test statistics follow a chi squared distribution and have a critical value of 5.99 at the 5 per cent level, with two degrees of freedom. An asterisk indicates statistical significance at the 5 per cent level.

$$f = \begin{cases} (i_{y,t} - i_{x,t}) + \hat{S}_{t+1} & \text{if } i_{y,t} > i_{x,t} \\ (i_{x,t} - i_{y,t}) - \hat{S}_{t+1} & \text{if } i_{y,t} < i_{x,t} \end{cases} \quad (34)$$

Once profitability is measured, the mean value of the return is given by:

$$\bar{f} = \frac{1}{n-k} \sum_{t=k+1}^n f_t \quad (35)$$

A conventional test of the difference between two means (i.e. the mean returns) is also performed. The null hypothesis of equality in mean returns is rejected if:

$$\frac{f^1 - f^2}{s\sqrt{\frac{1}{n-k} + \frac{1}{n-k}}} > t(n-k+n-k-2) \quad (36)$$

where superscripts 1 and 2 represent the model and random walk, respectively, and:

---

change by the drift factor. In this case, the expected return is calculated by replacing the forecast percentage change in the exchange rate with the drift factor, which gives:  $f^e = (i_{y,t} - i_{x,t}) + d$ . The realised return from a carry trade strategy, where the drift factor is statistically significant, is given by:  $f = \begin{cases} (i_{y,t} - i_{x,t}) + d & \text{if } d > 0 \\ (i_{x,t} - i_{y,t}) - d & \text{if } d < 0 \end{cases}$ .

**TABLE 9. Proportion of variation in exchange rate accounted for by trend and cycle**

	CAD/AUD	GBP/AUD	CAD/GBP	CAD/USD	GBP/USD	GBP/JPY
Frenkel-Bilson	54.43	42.82	61.35	20.67	70.51	45.45
Dornbusch-Frankel	49.49	32.03	56.28	19.82	64.53	41.87
Hooper-Morton	42.66	25.49	46.05	14.07	44.94	37.56

**Note:** Calculated as the amount of variation in the exchange rate explained by the trend and cycle components, as a proportion of the amount of variation explained by the full model specification as described in equations (1) to (3) (as measured by the coefficient of determination).

$$s = \sqrt{\frac{(n-k-1)\sigma^2(f^1) + (n-k-1)\sigma^2(f^2)}{n-k+n-k-2}} \quad (37)$$

The standard deviation of the rate of return is calculated as:

$$SD = \sqrt{\frac{1}{n-k-1} \sum_{t=k+1}^n (f_t - \bar{f})^2} \quad (38)$$

The Sharpe ratio measures risk-adjusted returns. Following Burnside et al. (2010) and Gyntelberg and Remolona (2007), the Sharpe ratio is calculated as the ratio of the mean to the standard deviation of the rate of return:

$$SR = \frac{\bar{f}}{SD} \quad (39)$$

## IV. Data and empirical results

### A. Data

The empirical work of this study is performed on six exchange rates, two against the US dollar (CAD/USD and GBP/USD) and four cross

**TABLE 10. Root mean square error and hypothesis testing**

	CAD/AUD	GBP/AUD	CAD/GBP	CAD/USD	GBP/USD	GBP/JPY
Random walk	0.029	0.030	0.030	0.029	0.027	0.036
Frenkel-Bilson	0.029* (0.0993)	0.032* (0.0583)	0.036** (0.0004)	0.030** (0.0185)	0.029 (0.2982)	0.034 (0.6318)
Dornbusch-Frankel	0.029 (0.2670)	0.032** (0.0274)	0.034** (0.0148)	0.031*** (0.0116)	0.029 (0.3969)	0.035 (0.9372)
Hooper-Morton	0.053*** (0.0000)	0.054*** (0.0000)	0.046*** (0.0000)	0.048*** (0.0000)	0.037*** (0.0000)	0.036 (0.9807)

**Note:** The AGS test statistic follows a chi square distribution. P-values are in parentheses and \*, \*\*, \*\*\* denote statistical significance at the 10, 5 and 1 per cent level, respectively.

rates (CAD/AUD, CAD/GBP, GBP/AUD and GBP/JPY).<sup>23</sup> The sample spans the period from January 1984 to June 2011 and all data is sourced from International Financial Statistics (CD-ROM), published by the International Monetary Fund. Monthly data frequencies are used and, where a time series is only available on a quarterly frequency (for example, GDP for all countries and CPI for some countries), the data is converted into a monthly time series using a continuous time-dynamic interpolation method (Moosa and Burns, 2012b). The models (as specified in equations (1) to (3)) are estimated from January 1984 to June 2001, and one month ahead out-of-sample forecasts are generated from July 2001 to June 2011.

### B. Diagnostic testing

A range of diagnostic tests performed on the standardised residuals generated from the monetary model forecasts endorse the estimation of the models in state-space form conducted in this study. Unit root testing reports that the residuals are stationary in every instance (table 4).<sup>24</sup> The

23. The currencies selected account for 86 per cent of global currency turnover (Bank for International Settlements, 2013).

24. Only the ADF test results are presented here for the purposes of simplicity. Additional unit root test results (using the Phillips-Perron, Kwiatkowski-Phillips-Schmidt-Shin, Elliott-Rothenberg-Stock Point Optimal and modified Ng-Perron unit root tests) are available from the author upon request.

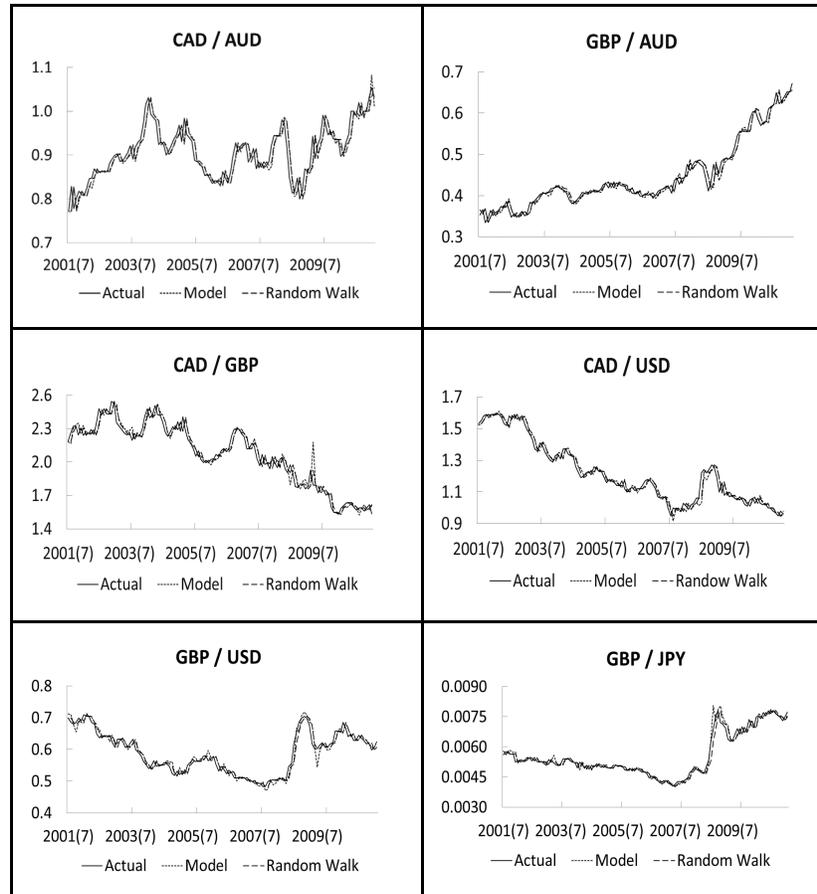


FIGURE 1.— Time plots

Breusch-Pagan test reveals that, in almost every case, the null hypothesis of homoscedasticity cannot be rejected (table 5). Although the assumption that the residuals are homoscedastic impacts the validity of statistical tests and confidence intervals associated with the regression model, there is no reason that this will detrimentally impact forecasting accuracy (Makridakis, Wheelwright and Hyndman, 1998). Therefore, even in the very small number of cases where the errors are heteroscedastic, this is not fatal to the analysis. The Durbin-Watson statistic reveals little evidence of positive or negative autocorrelation at lag one in the standardised residuals (table 6). For completeness, the

**TABLE 11. Diebold and Mariano test results**

	CAD/AUD	GBP/AUD	CAD/GBP	CAD/USD	GBP/USD	GBP/JPY
Frenkel-Bilson	1.97** (0.0512)	2.28** (0.0245)	0.94 (0.3507)	1.11 (0.2700)	1.17 (0.2425)	-0.51 (0.6110)
Dornbusch-Frankel	0.55 (0.5814)	2.20** (0.0299)	0.97 (0.3350)	1.18 (0.2391)	0.99 (0.3237)	-0.50 (0.6180)
Hooper-Morton	3.17*** 42.66	5.08*** 25.49	4.36*** 46.05	3.19*** 14.07	1.63 44.94	-0.01 37.56

**Note:** The Diebold-Mariano test statistic follows a t-distribution. P-values are in parenthesis and \*, \*\*, \*\*\* denote statistical significance at the 10, 5 and 1 per cent level, respectively.

Ljung-Box Q-statistics are also presented (table 7). Again, the results show that there is no autocorrelation in the residuals in the majority of cases. The Doornik and Hansen (2008) test is used to test for normality in the standardised residuals and the null hypothesis of normally distributed standardised errors cannot be rejected in three of the 18 cases (table 8). This finding is not a cause for concern because a regression model may still be robust, even if the errors are non-normal (Osborne and Waters, 2002).

As a further exercise, table 9 reports the amount of variation in the exchange rate that can be explained by the full model specification (as described in equations (1) to (3)) compared to the amount of variation in the exchange rate that can be explained by the trend and cycle components only.<sup>25</sup> On average, more than half of the variation in the exchange rate explained by equations (1) to (3) is attributable to the economic fundamentals. In other words, the trend and cycle components account for less than half of the total variation in the exchange rate which is explained by the exchange rate determination models.

### C. Out-of-sample forecasts

Figure 1 presents the time plots of the out-of-sample forecasts of the random walk and the Dornbusch-Frankel model, estimated with

25. Calculated as the amount of variation in the exchange rate explained by the trend and cycle components only, as a proportion of the amount of variation explained by the full model specification as described in equations (1) to (3) (as measured by the coefficient of determination).

**TABLE 12. Direction accuracy and hypothesis testing**

	CAD/AUD	GBP/AUD	CAD/GBP	CAD/USD	GBP/USD	GBP/JPY
Frenkel- Bilson	0.5210	0.5462	0.5966	0.5210	0.5042	0.5882
$H_0$ : DA = 0.00	(11.42)	(12.02)	(13.32)	(11.42)	(11.05)	(13.09)
$H_0$ : DA = 0.50	(0.46)	(1.02)	(2.16)	(0.46)	(0.09)	(1.96)
Dornbusch- Frankel	0.5210	0.5294	0.5966	0.5126	0.4958	0.6471
$H_0$ : DA = 0.00	(11.42)	(11.62)	(13.32)	(11.23)	(10.86)	(14.83)
$H_0$ : DA = 0.50	(0.46)	(0.65)	(2.16)	(0.28)	(-0.09)	(3.37)
Hooper- Morton	0.4202	0.5210	0.5126	0.4874	0.4202	0.6387
$H_0$ : DA = 0.00	(9.33)	(11.42)	(11.23)	(10.68)	(9.33)	(14.56)
$H_0$ : DA = 0.50	(-1.77)	(0.46)	(0.28)	(-0.28)	(-1.77)	(3.16)

**Note:** The critical value of the t-test statistic at the 5 per cent level of significance is 1.96. Test statistics are in parentheses.

time-varying parameters for each of the cross rates.<sup>26</sup> It is evident from these time plots that the actual exchange rate, the random walk forecasts and the exchange rate model forecasts are almost indistinguishable. Based on a visual inspection, both the model and random walk appear to provide a good approximation of the actual exchange rate. In the next section, the competing forecasts are assessed using conventional and alternative measures of forecasting accuracy, as presented in section III.

#### *D. Conventional measures of forecasting accuracy*

When forecasting accuracy is measured by conventional means (that is, by the magnitude of error only), the results reinforce the Meese-Rogoff

26. The time plots have been presented for all cross rates and only for the Dornbusch-Frankel model. There are two reasons for this. First, the time plots of the other model forecasts are extremely similar and presenting all 18 time plots would provide no real additional information to the reader. Second, in an exercise like this where a number of currencies and models are considered, the author must necessarily be selective about what evidence is presented.

**TABLE 13. PT Test for direction accuracy**

	CAD/AUD	GBP/AUD	CAD/GBP	CAD/USD	GBP/USD	GBP/JPY
Frenkel-Bilson	1.24 (0.2662)	-1.70 (0.1919)	1.61 (0.2045)	2.28 (0.1309)	1.12 (0.2906)	1.28 (0.2586)
Dornbusch-Frankel	0.82 (0.3646)	-1.04 (0.3078)	0.94 (0.3334)	2.10 (0.1476)	1.97 (0.1606)	2.64* (0.1041)
Hooper-Morton	-1.84 (0.1750)	-0.50 (0.4799)	-1.28 (0.2579)	1.10 (0.2938)	-0.98 (0.3231)	3.47* (0.0625)

**Note:** The PT test statistic follows a chi squared distribution and has a critical value of 3.84 at the 5 per cent level of significance, with one degree of freedom. P-values are also presented in parenthesis and \*, \*\*, \*\*\* denote statistical significance at the 10, 5 and 1 per cent level, respectively.

findings and show that no exchange rate model can produce a statistically smaller RMSE compared to the naïve model forecasts (table 10). In around half of the cases, the RMSE of the model and random walk are numerically the same (to three decimal places). Even where the model produces a slightly higher or smaller numerical RMSE, the AGS test proves the RMSEs are not statistically different in most cases.

These findings demonstrate the pitfalls of reaching conclusions about relative forecasting accuracy by comparing the numerical value of the RMSE. Just because the RMSEs are numerically different does not imply that they are statistically different. More importantly, the results show that exchange rate models estimated using time-varying parameters often produce a magnitude of forecasting error equivalent to the naïve random walk model. This overturns the suggestion by Meese and Rogoff (and others) that exchange rate models incorporating stochastic movements in parameters may be able to outperform the random walk in terms of the RMSE. The results also dismiss stochastic movements in parameters as a possible explanation for the Meese-Rogoff puzzle. Likewise, the Diebold and Mariano (1995) test proves that exchange rate models forecasts generated using a TVP framework produce an equivalent magnitude of error to the random walk (table 11). That is, there is no statistical difference in the loss function of the model and random walk forecasts. The exchange rate models perform best for the GBP/USD and GBP/JPY rates, where the null hypothesis of equality in the loss function cannot be rejected for all three models. Based on these results, the proposition that the random walk is superior to exchange rate models in terms of the magnitude of

**TABLE 14. Mean return and hypothesis testing**

	CAD/AUD	GBP/AUD	CAD/GBP	CAD/USD	GBP/USD	GBP/JPY
Random walk	6.14	7.97	-1.57	0.94	0.94	-0.77
Frenkel-Bilson	6.39 (0.9556)	9.73 (0.7101)	2.64 (0.3664)	1.50 (0.9017)	1.67 (0.8623)	5.78 (0.2519)
Dornbusch-Frankel	6.39 (0.9556)	9.25 (0.7873)	5.51 (0.1287)	1.37 (0.9237)	3.53 (0.5384)	8.43 (0.1067)
Hooper-Morton	1.98 (0.3520)	2.48 (0.2535)	1.79 (0.4714)	2.14 (0.7904)	1.26 (0.9405)	8.79* (0.0935)

**Note:** P-values for the null hypothesis  $H_0: \pi^M = \pi^{RW}$  are presented in parenthesis and \*, \*\*, \*\*\* denote statistical significance at the 10, 5 and 1 per cent level, respectively.

forecasting error is not supported. However, this does not infer that the model forecasts are able to outperform the random walk in terms of the RMSE.

Two conclusions can be reached on the basis of these results. First, parameter instability as a possible explanation for the Meese-Rogoff puzzle is entirely discounted because exchange rate models estimated with time-varying parameters cannot produce a statistically smaller RMSE compared to the random walk. Second, the random walk is unable to outperform exchange rate models if these models are estimated using time-varying parameters. This is demonstrated by showing that the small difference in the RMSEs of the model and random walk is not statistically significant.

#### *E. Alternative measures of forecasting accuracy*

The direction accuracy and hypothesis testing results verify the proposition that the random walk can be outperformed when exchange rate models are estimated using TVP, and forecasting accuracy is assessed by alternative measures (table 12). In all cases, the model outperforms the random walk because the null hypothesis of nil direction accuracy is rejected (and the random walk without drift is a no change forecast and, therefore, implies nil direction accuracy). This demonstrates that the Meese-Rogoff puzzle can be overturned if forecasting accuracy is assessed by direction accuracy and the exchange rate models are estimated with TVP. More importantly, the supplementary test results show that the model correctly predicts the

**TABLE 15. Sharpe ratio**

	CAD/AUD	GBP/AUD	CAD/GBP	CAD/USD	GBP/USD	GBP/JPY
Random walk	0.18	0.22	-0.04	0.03	0.03	-0.02
Frenkel-Bilson	0.19	0.27	0.07	0.04	0.05	0.13
Dornbusch-Frankel	0.19	0.25	0.15	0.04	0.11	0.19
Hooper-Morton	0.06	0.07	0.05	0.06	0.04	0.20

direction of change on at least 50 per cent of occasions in all cases. In some cases, direction accuracy of the model forecasts actually exceeds 50 per cent (four cases relating to CAD/GBP and GBP/JPY).

The conclusions drawn from the direction accuracy results corroborate the PT test results (table 13). The PT test rejects the null hypothesis of independence between the actual and predicted change in the exchange rate in two cases, both relating to the GBP/JPY. These are the same cases where direction accuracy exceeds 60 per cent. The results suggest that direction accuracy of at least 60 per cent is necessary for there to be a statistical relationship between the actual movement in the exchange rate and the change predicted by the model. In the majority of cases, however, the actual and predicted changes are statistically independent. The reason for this is that the PT test has a higher threshold than the t-test on direction accuracy and this highlights the importance of using a wide variety of measures to assess forecasting accuracy.

The mean returns and hypothesis testing results are presented in table 14. In almost every case, the model outperforms the random walk based on the numerical mean return (either a larger positive return or a smaller negative return). Nonetheless, hypothesis testing overwhelmingly shows that the mean returns of the random walk and model are not statistically different. This is largely due to the high variance in both the model and random walk forecasts and is most likely attributable to the lengthy sample used in this study. Notwithstanding this, the results show that the random walk is not superior to exchange rate models when forecasting accuracy is measured by profitability. The Sharpe ratio results, however, lend overwhelming support to the

**TABLE 16. Adjusted root mean square error**

	CAD/AUD	GBP/AUD	CAD/GBP	CAD/USD	GBP/USD	GBP/JPY
Random walk	0.029	0.030	0.030	0.029	0.027	0.036
Frenkel-Bilson	0.020	0.021	0.023	0.021	0.020	0.022
Dornbusch-Frankel	0.020	0.022	0.022	0.021	0.020	0.021
Hooper-Morton	0.040	0.038	0.032	0.035	0.028	0.022

superiority of the monetary models over the random walk (table 15). In almost every instance, the forecasts generated by the TVP models generate a larger risk-adjusted return compared to the random walk. Therefore, when forecasting accuracy is measured by risk-adjusted returns, the random walk model is easily outperformed by exchange rate determination models estimated in state-space form.

The ARMSE results verify the proposition that the random walk is outperformed by exchange rate models estimated with time-varying parameters when forecasting accuracy is assessed in terms of both errors in magnitude and direction (table 16). In almost every instance, the model performs as well as, or better than, the random walk in terms of the ARMSE. Most noteworthy are the results for the GBP/JPY where the error of the random walk forecast is 50 per cent larger than that of the model forecasts. The results prove that by taking errors in both magnitude and direction into account, the forecasts generated by monetary models estimated in a TVP framework are superior to the random walk, and the Meese-Rogoff puzzle can be overturned.

The results of the Wald test for proximity to a perfect forecast further support the proposition that the random walk model can be outperformed by exchange rate models estimated with time-varying parameters (table 17). Recall that the Wald test statistic measures the deviation of the forecasts from the perfect forecast. As the model forecasts have a much lower Wald test statistic compared to the random walk for every currency and model, this means that the model forecasts deviate from the perfect forecast by a lesser amount relative to the random walk forecasts. In fact, the random walk forecasts report a Wald test statistic that is between 34 and 3,416 times that of the exchange rate

**TABLE 17. Test for proximity to perfect forecast**

	CAD/AUD	GBP/AUD	CAD/GBP	CAD/USD	GBP/USD	GBP/JPY
Random walk	2857.40	2057.40	3213.10	201317.00	2525.90	22047.70
Frenkel-Bilson	1912.30	1902.50	302.42	571.86	261.51	272.94
Dornbusch-Frankel	1313.90	1275.60	352.58	533.58	234.23	66.36
Hooper-Morton	72.87	60.91	67.13	58.94	75.24	133.12

**Note:** The test statistic follows a chi squared distribution and has a critical value of 5.99 at the 5 per cent level of significance, with two degrees of freedom.

model forecasts. Although the null hypothesis that the forecasts are perfectly accurate is rejected in every case, the large numerical differences in the Wald test statistic demonstrates the superiority of exchange rate models estimated using TVP compared to the random walk.

## V. Discussion

The Meese-Rogoff puzzle posits that the random walk cannot be outperformed by exchange rate determination models in out-of-sample forecasting. Time-varying parameters have been suggested by Meese and Rogoff (1983) and others as a possible explanation for this result. Another reason put forward in more recent literature is that forecasting accuracy is traditionally assessed by conventional approaches that exclusively measure the magnitude of the forecasting error. The results of this study prove that estimating the monetary model of exchange rate determination using time-varying parameters does not, by itself, resolve the Meese-Rogoff puzzle. That is, the model forecasts are unable to produce a smaller and statistically different magnitude of forecasting error compared to the random walk forecasts. In the majority of cases, however, there is no statistical difference in the RMSE of the exchange rate model and random walk forecasts. Therefore, the models and random walk have equivalent forecasting accuracy as measured by the RMSE, if the models are estimated using TVP. On this basis, it can be concluded that parameter stability cannot explain the Meese-Rogoff

puzzle. However, incorporating stochastic movements in the parameters evidently brings about a vast improvement in the accuracy of the model forecasts to the extent that the random walk does not produce a superior forecast as measured by the RMSE. In fact, the gains arising from using a TVP framework are so substantial that the model and random walk forecasts generate statistically equivalent RMSEs.

The inability of monetary models of exchange rate determination to outperform the random walk in terms of the RMSE, even when estimated in state-space form, warrants further discussion. Some economists suggest that exchange rates are forward-looking and reflect market participants' expectations of both current and future fundamentals, often with a greater weight being placed on future economic fundamentals (see, for example, Bachetta and van Wincoop, 2011; Engel and West, 2005). The forward looking nature of exchange rate movements effectively disconnects the exchange rate from current economic fundamentals and, hence, monetary models of exchange rate determination perform poorly in terms of the magnitude of error. Alternatively, it has been suggested that unobserved shocks (such as structural instability as a result of policy changes) are the primary drivers of exchange rate movements and (as they are unobservable) are unable to be incorporated into exchange rate models (see, for instance, Balke, Ma and Wohar, 2013). However, both of these considerations are taken into account by adopting the unobserved components model used in this study (Koopman et al., 2006). Therefore, neither of these suggestions can adequately explain why exchange rate models cannot outperform the random walk in terms of the RMSE. The reason that exchange rate models estimated using time-varying parameters cannot outperform the random walk is that the period to period change in the exchange rate (which is by definition the error of the random walk) is so small that it is almost impossible for a model to produce forecasts with a statistically smaller magnitude of error unless the model produces a perfect forecast with nil error. It is for this reason that Engel, Mark and West (2008) suggest that beating the random walk in terms of the RMSE is "too strong" a criterion.

However, if forecasting accuracy is measured by any metric that does not exclusively rely on the magnitude of error, the random walk can be outperformed by exchange rate models estimated in state-space form in out-of-sample forecasting. The model forecasts are capable of outperforming the random walk forecasts in terms of direction accuracy, measures that take into account magnitude and direction (ARMSE),

risk-adjusted returns and proximity to a perfect forecast. When criteria other than the magnitude of error are considered, it is not difficult for the model forecasts to outperform the random walk. The reason is that the random walk by definition predicts no period-to-period change in the exchange rate, whereas the exchange rate model predicts a change in the exchange rate based on macroeconomic fundamentals. Although the random walk and model perform equivalently in terms of the RMSE, the model forecasts are so superior in terms of direction accuracy that, when both magnitude and direction accuracy are taken into account, the model overwhelmingly outperforms the naïve random walk model. Therefore, by reconsidering the Meese-Rogoff puzzle using alternative measures of forecasting accuracy (that do not rely exclusively on the magnitude of error), the results show that the random walk can be outperformed by exchange rate models estimated with time-varying parameters.

By using six alternative measures of forecasting accuracy, this study has unquestionably demonstrated that the Meese-Rogoff puzzle can be overturned if forecasting accuracy is measured by metrics that take into account other important characteristics of forecasting accuracy. No other study utilises such a wide variety of alternative measures of forecasting accuracy to demonstrate the robustness of the proposition that the Meese-Rogoff puzzle can be overturned by using measures of forecasting accuracy that do not rely exclusively on the magnitude of error. In addition, the results show that the naïve random walk model, and monetary models estimated in state-space form, produce equivalent errors in magnitude. The findings demonstrate the gains arising from estimating exchange rate models in a TVP framework over and above the random walk model. In addition, these results support three conclusions. First, parameter instability cannot explain the Meese-Rogoff puzzle because incorporating stochastic movements in the parameters does not enable the model to produce a statistically smaller RMSE compared to the naïve model. Second, if the Meese-Rogoff puzzle is reconsidered by assessing forecasting accuracy using alternative measures that do not rely exclusively on the magnitude of error, the random walk can be outperformed. Third, the forecasts produced by exchange rate determination models estimated in a TVP framework are more useful than those of the random walk and there are significant gains to using exchange rate models for out-of-sample forecasting.

## VI. Conclusion

This study comprehensively demonstrates that the Meese and Rogoff (1983) findings cannot be overturned by incorporating stochastic movements in the parameters of monetary models of exchange rate determination. That is, exchange rate models estimated in a TVP framework cannot produce a numerically smaller and statistically different magnitude of error compared to the random walk. This empirical result is to be expected because of the nature of exchange rate volatility. This does not mean that the random walk produces a superior forecast, or that macroeconomic fundamentals cannot predict movements in bilateral exchange rates. There are substantial benefits to be gained by estimating exchange rate models using a TVP framework, evinced by the superiority of the model forecasts over and above the random walk in terms of direction accuracy, profitability, proximity to a perfect forecast and the ARMSE. In addition, a model estimated using TVP can produce an equivalent RMSE to the random walk, meaning that the gains obtained in using a TVP model to forecast exchange rates comes at no cost in terms of the magnitude of forecasting error. The findings prove that the Meese-Rogoff puzzle can be overturned by using any alternative measure of forecasting accuracy that does not rely exclusively on the magnitude of error, and by incorporating stochastic movements in the parameters of the model.

*Accepted by: P.C. Andreou, PhD, Editor-in-Chief(Pro-Tem), November 2015*

## References

- Abhyankar, A.; Sarno, L.; and Valente, G. 2005. Exchange rates and fundamentals: Evidence on the economic value of predictability. *Journal of International Economics* 66(2): 325-348.
- Ashley, R.; Granger, C.; and Schmalensee, R. 1980. Advertising and aggregate consumption: An analysis of causality. *Econometrica* 48(5): 1149-67.
- Bai, J., and Perron, P. 1998. Estimating and testing linear models with multiple structural breaks. *Econometrica* 66(1): 47-78.
- Bacchetta, P., and van Wincoop, E. 2011. On the unstable relationship between exchange rates and macroeconomic fundamentals. *Journal of International Economics* 91(1): 18-26.
- Balke, N.; Ma, J.; and Wohar, M. 2013. The contribution of economic fundamentals to movements in exchange rates. *Journal of International Economics* 90(1): 1-16.

- Bank for International Settlements. 2013. Triennial central bank survey foreign exchange turnover in April 2013: Preliminary global results. Available at <http://www.bis.org/publ/rpfx13fx.pdf>.
- Barnett, A.; Mumtaz, H.; and Theodoridis, K. 2012. Forecasting UK GDP growth, inflation and interest rates under structural change: A comparison of models with time-varying parameters. Bank of England Working Paper No. 450, London, U.K., Bank of England.
- Boothe, P., and Glassman, D. 1987. Comparing exchange rate forecasting models: Accuracy versus profitability. *International Journal of Forecasting* 3(1): 65-79.
- Burnside, C.; Eichenbaum, M.; Kleshchelski, I.; and Rebelo, S. 2010. Do Peso Problems Explain the Returns to the Carry Trade? *Review of Financial Studies* 24(3): 853-891.
- Canova, F. 1993. Modelling and forecasting exchange rates with a Bayesian time-varying coefficient model. *Journal of Economic Dynamics and Control* 17(1-2): 233-261.
- Cheung, Y.-W.; Chinn, M.; and Pascual, A. 2005. Empirical exchange rate models of the nineties: Are they fit to survive? *Journal of International Money and Finance* 24(7): 1150-1175.
- Christoffersen, P., and Diebold, F. 1998. Cointegration and long-horizon forecasting. *Journal of Business & Economic Statistics* 16(4): 450-58.
- Corte, P.; Lucio, S.; and Tsiakas, I. 2009. An economic evaluation of empirical exchange rate models. *Review Financial Studies* 22(9): 3491-3530.
- Diebold, F., and Mariano, R. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13(3): 253-263.
- Doornik, J., and Hansen, H. 2008. An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics* 70(S1): 927-939.
- Engel, C. 1994. Can the Markov-switching model forecast exchange rates? *Journal of International Economics* 36(1-2): 151-165.
- Engel, C., and Hamilton, J. D. 1990. Long swings in the dollar: Are they in the data and do markets know it? *American Economic Review* 80(4): 689-713.
- Engel, C.; Mark, N.; and West, K. 2008. Exchange rate models are not as bad as you think. In: Acemoglu, D.; Rogoff, K.; and Woodford, M. (eds), NBER Macroeconomics Annual 2007 Volume 22. Illinois: University of Chicago Press.
- Engel, C., and West, K. 2005. Exchange Rates and Fundamentals. *Journal of Political Economy* 113(3): 485-517.
- Evans, M., and Lyons, R. 2005. Meese-Rogoff redux: Micro-based exchange rate forecasting. *American Economic Review* 95(2): 405-414.
- Faust, J.; Rogers, J.; and Wright, J. 2003. Exchange rate forecasting: the errors we've really made. *Journal of International Economics* 60(1): 35-59.
- Frommel, M.; MacDonald, R.; and Menkhoff, L. 2005. Do fundamentals matter for the D-Mark/Euro-Dollar? A regime switching approach. *Global Finance*

- Journal* 15(3): 321-335.
- Gyntelberg, J., and Remolona, E. 2007. Risk in Carry Trades: A Look at Target Currencies in Asia and the Pacific. *BIS Quarterly Review December*: 73-82.
- Junttila, J., and Korhonen, M. 2011. Nonlinearity and time-variation in the monetary model of exchange rates. *Journal of Macroeconomics* 33(2): 288-302.
- Koopman, S. J.; Shephard, N.; and Doornik, J. A. 1999. Statistical algorithms for models in state space form using SsfPack 2.2. *Econometrics Journal* 2(1): 113-166.
- Koopman, S.; Harvey, A.; Doornik, J.; and Shephard, N. 2006. Structural Time Series Analyser, Modeller and Predictor. London: Timberlake Consultants Ltd.
- Li, M. 2011. An evaluation of exchange rate models by carry trade. *Journal of Economics and International Finance* 3(2): 72-87.
- Lu, S., and Ito, T. 2008. Structural breaks and time-varying parameter: A survey with application. *Communications of the IBIMA* 2(18): 129-135.
- Marcellino, M. 2002. Instability and non-linearity in the EMU. C.E.P.R. Discussion Papers 3312.
- Marcellino, M.; Stock, J.; and Watson, M. 2003. Macroeconomic Forecasting in the Euro Area: Country Specific versus Area-Wide Information. *European Economic Review* 47(1): 1-18.
- Makridakis, S.; Wheelwright, S.; and Hyndman, R. 1998. Forecasting Methods and Applications (3rd ed). Chichester: John Wiley.
- Mark, N., and Sul, D. 2001. Nominal exchange rates and monetary fundamentals, Evidence from a small post-Bretton woods panel. *Journal of International Economics* 53(1): 29-52.
- Meese, R., and Rogoff, K. 1983. Empirical exchange rate models of the seventies: Do they fit out-of-sample? *Journal of International Economics* 14(1-2): 3-24.
- Mendelsohn, R. 2011. The STAMP software for state space models. *Journal of Statistical Software* 41(2): 1-18.
- Molodtsova, T., and Papell, D. 2009. Out-of-sample exchange rate predictability with Taylor rule fundamentals. *Journal of International Economics* 77(2): 167-180.
- Molodtsova, T., and Papell, D. 2013. Taylor rule exchange rate forecasting during the financial crisis. NBER International Seminar on Macroeconomics, University of Chicago Press, vol. 9(1): 55-97
- Moosa, I. 2000. Exchange Rate Forecasting: Techniques and Applications. London: Macmillan Business.
- Moosa, I. 2006. On the notion of forecasting accuracy as applied to financial decision rules involving exchange rates. *Forecasting Letters* 1(2): 5-9.
- Moosa, I. 2013. Why is it so difficult to outperform the random walk in exchange rate forecasting? *Applied Economics* 45(23): 3340-46.
- Moosa, I., and Burns, K. 2012a. Can exchange rate models outperform the

- random walk? Magnitude, direction and profitability as Criteria. *Economia Internazionale* 65(3): 473-490.
- Moosa, I., and Burns, K. 2012b. Interpolating flow and stock variables in a continuous-time dynamic framework. *Applied Economics Letters* 20(7): 621-625.
- Moosa, I., and Burns, K. 2013a. A reappraisal of the Meese-Rogoff puzzle. *Applied Economics* 46(1): 30-40.
- Moosa, I., and Burns, K. 2013b. The monetary model of exchange rates is better than the random walk in out-of-sample forecasting. *Applied Economics Letters* 45(23): 1293-1297.
- Moosa, I., and Kwiecien, J. 2002. Cross-country evidence on the ability of the nominal interest rate to predict inflation. *The Japanese Economic Review* 53(4): 478-495.
- Osborne, J., and Waters, E. 2002. Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research & Evaluation* 8(2): 1-5.
- Pesaran, H., and Timmermann, A. 1992. A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics* 10(4): 461-465.
- Rapach, D., and Wohar, M. 2002. Testing the monetary model of exchange rate determination: New evidence from a century of data. *Journal of International Economics* 58(2): 359-385.
- Rossi, B. 2006. Are exchange rates really random walks? Some evidence robust to parameter instability. *Macroeconomic Dynamics* 10(1): 20-38.
- Schinasi, G., and Swamy, P. 1989. The out-of-sample forecasting performance of exchange rate models when coefficients are allowed to change. *Journal of International Money and Finance* 8(3): 375-390.
- West, K.; Edison, H.; and Cho, D. 1993. A utility based comparison of some models of exchange rate volatility. *Journal of International Economics* 35(1): 23-45.
- Wolff, C. 1987. Time-varying parameters and the out-of-sample forecasting performance of structural exchange rate models. *Journal of Business and Economic Statistics* 5(1): 87-97.
- Young, P.; Diego, P.; and Wlodek, T. 1999. Dynamic harmonic regression. *Journal of Forecasting* 18(6): 369-394.