JIAN WANG
JASON J. WU

The Taylor Rule and Forecast Intervals for Exchange Rates

In this paper, we examine the Meese–Rogoff puzzle from a different perspective: out-of-sample interval forecasting. While most studies in the literature focus on point forecasts, we apply semiparametric interval forecasting to a group of exchange rate models. Forecast intervals for 10 OECD exchange rates are generated and the performance of the empirical exchange rate models are compared with the random walk. Our contribution is twofold. First, we find that in general, exchange rate models generate tighter forecast intervals than the random walk, given that their intervals cover out-of-sample exchange rate realizations equally well. Our results suggest a connection between exchange rates and economic fundamentals: economic variables contain information useful in forecasting distributions of exchange rates. We also find that the benchmark Taylor rule model performs better than the monetary, PPP and forward premium models, and its advantages are more pronounced at longer horizons. Second, the bootstrap inference framework proposed in this paper for forecast interval evaluation can be applied in a broader context, such as inflation forecasting.

JEL codes: C14, C53, F31
Keywords: Meese–Rogoff puzzle, exchange rate forecast, interval forecasting, Taylor rule model.

In a seminal paper, Meese and Rogoff (1983) find that economic fundamentals—such as the money supply, trade balance and national income—are of little use in forecasting out-of-sample exchange rates. This finding has been termed the Meese–Rogoff puzzle. In defense of fundamental-based exchange rate models, various combinations of economic variables and econometric methods have

We are grateful to Editor Ken West and two anonymous referees for valuable comments and suggestions that greatly improved the paper. We also thank Menzie Chinn, Charles Engel, Bruce Hansen, Jesper Linde, Enrique Martinez-Garcia, Tanya Moldtsova, David Papell, John Rogers, Mark Wynne, and seminar participants at the Dallas Fed and the University of Houston for helpful comments. All views are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Dallas, the Board of Governors, or the Federal Reserve System. All Matlab programs are available upon request.

Jian Wang *is at the Research Department, Federal Reserve Bank of Dallas* (*E-mail:* jian.wang@dal.frb.org). Jason J. Wu *is at the Board of Governors of the Federal Reserve System* (*E-mail:* jason.j.wu@frb.gov).

Received October 14, 2009; and accepted in revised form July 27, 2011.

Journal of Money, Credit and Banking, Vol. 44, No. 1 (February 2012)

been used in attempts to overturn Meese and Rogoff's finding. Mark (1995) finds greater exchange rate predictability at longer horizons.[1] Clarida and Taylor (1997) and Clarida et al. (2003) suggest that interest rate differentials contain information useful in predicting exchange rates out of sample. Groen (2000, 2005) and Mark and Sul (2001) detect exchange rate predictability by using panel data. Kilian and Taylor (2003) find that exchange rates can be predicted from economic models at horizons of 2 to 3 years, after taking into account the possibility of nonlinear exchange rate dynamics. Faust, Rogers, and Wright (2003) find that the economic models consistently perform better using real-time data than revised data, although they do not perform better than the random walk. Other recent studies on exchange rate predictability include Alquist and Chinn (2008) and Adrian, Etula, and Shin (2009).

This paper addresses the Meese–Rogoff puzzle from a different perspective: interval forecasting. A forecast interval captures a range in which the exchange rate may lie with a certain probability, given predictors available at the time of forecast. Our contribution is twofold. First, we find that for 10 OECD exchange rates, exchange rate models in general generate tighter forecast intervals than the random walk, given that the intervals cover the realized exchange rates (statistically) equally well. This finding suggests an intuitive connection between exchange rates and economic fundamentals beyond point forecasting: the use of economic variables as predictors helps narrow down the range in which future exchange rates may lie relative to random walk forecast intervals. Second, we propose a bootstrap inference framework for cross-model comparison of out-of-sample forecast intervals. The proposed framework can be used for forecast interval evaluation in a broader context. For instance, the framework can also be used to evaluate out-of-sample inflation forecasting.

We apply the semiparametric forecast intervals of Wu (Forthcoming) to a group of exchange rate models that includes Monetary, Purchasing power parity (PPP), and forward premium models.[2] We also consider a set of exchange rate models based on the Taylor rule for monetary policy. The benchmark Taylor rule model is from Engel and West (2005) and Engel, Wang, and Wu (2009). Several alternative Taylor rule setups presented in Molodtsova and Papell (2009) and Engel, Mark, and West (2007) are also considered. Recent studies find empirical support for the linkage between exchange rates and Taylor rule fundamentals. Engel and West (2005) derive the exchange rate as a present-value asset price from a Taylor rule model. Engel and West (2006) find positive correlations between the model-based and actual real exchange rates between the U.S. dollar and Deutschmark. Mark (2009) examines the role of Taylor rule fundamentals for exchange rate determination in a model with learning. He finds that the model is able to capture six major swings of the real Deutschmark–dollar exchange rate from 1973 to 2005. Kim and Ogaki (2009) find

1. Chinn and Meese (1995) and MacDonald and Taylor (1994) find similar results. Long-horizon exchange rate predictability in Mark (1995) has been challenged by subsequent studies such as Kilian (1999) and Berkowitz and Giorgianni (2001).

2. For instance, see Frenkel (1976), Mussa (1976), Bilson (1978), and more recently Mark (1995) and Engel and West (2005) for reference to the monetary model. See Froot and Rogoff (1995) for a review of studies on PPP.

that combining the Taylor rule and a standard exchange rate model in a system approach helps to produce more reasonable estimates of the real exchange rate's half-life. Chinn (2008) also finds that Taylor rule fundamentals do better than other models in in-sample prediction at the 1-year horizon. Using the inference procedure proposed by Clark and West (2006, 2007), Molodtsova and Papell (2009) find significant short-horizon out-of-sample predictability of exchange rates with Taylor rule fundamentals for 11 out of 12 currencies vis-á-vis the U.S. dollar over the post–Bretton Woods period. Molodtsova, Nikolsko-Rzhevskyy, and Papell (2008a, 2008b) find evidence of out-of-sample exchange rate predictability with forecasts based on Taylor rule fundamentals using real-time data. With a present-value asset pricing model as discussed in Engel and West (2005), Chen and Tsang (2009) find that information contained in the cross-country yield curves is useful in predicting exchange rates.[3]

We include 10 OECD exchange rates (relative to the U.S. dollar) over the post–Bretton Woods period in our data set. For these 10 exchange rates, the out-of-sample forecast intervals at different forecast horizons are generated from the exchange rate models and compared to the random walk results. Coverage accuracy and the length of forecast intervals are used as criteria to compare models.

Coverage accuracy refers to the difference between the probability that out-of-sample realizations fall into the forecast intervals and the nominal coverage. The length of the intervals is a measure of their tightness: the distance between their upper and lower bounds. Most evaluation methods of forecast intervals in the literature focus on comparing coverage accuracies across models, pioneered by the work of Christoffersen (1998). Given the importance of accurate coverage, we first test whether forecast intervals from exchange rate models and the random walk have equal coverage accuracies. The model with more accurate coverage is considered the better model. In cases where equal coverage accuracies cannot be rejected, we go on to test whether the lengths of forecast intervals are the same. All else equal, a wider forecast interval is more likely to cover out-of-sample realizations of the exchange rate. Testing the lengths of forecast intervals conditional on equal coverage accuracy helps us to compare interval lengths more meaningfully. Given equal coverage accuracy, the model with tighter forecast intervals is considered to be a better model.[4] In general, we find that most exchange rate models perform better than the random walk model, especially at long horizons: models either have more accurate coverages than the random walk, or in cases of equal coverage accuracies, the models have tighter forecast intervals. The improvement in lengths of forecast intervals is large especially at long horizons: economic models can reduce the lengths of intervals by 10% or

3. An earlier work of using term structure to forecast exchange rates is Clarida et al. (2003). More recently, Diez de los Rios (2009) also finds some evidence that the term structure of forward premia contains information for forecasting future spot exchange rates.

4. There are situations in which researchers/practitioners may be willing to exchange some coverage accuracy for a more informative interval, or vice versa. In this case, a joint test for coverage accuracy and interval length is more appropriate.

more compared to the random walk at the 12-month horizon. Among the exchange rate models, the benchmark Taylor rule model performs the best.

Our finding that economic models generate tighter forecast intervals can be economically useful. Consider, for instance, value-at-risk (or VaR).[5] VaR is a prevalent risk management tool used by investors. It is essentially a one-sided forecast interval measuring downside risks. The notion of coverage accuracy for VaR is the same as that for two-sided forecast intervals. A VaR estimate with nominal coverage α should cover realized losses 100% of the time over a sufficiently long period. Tighter VaR estimates may help investors avoid overhedging, or holding too much costly equity capital that is intended to buffer losses, as long as it covers losses accurately over time. Using one of our empirical results as an illustration, we demonstrate in Section 3.1 that if an investor has a short position on the Australian dollar, a VaR based on the benchmark Taylor rule model is on average (across 202 monthly, 1-year-ahead forecasts) 13% lower than a random-walk-based VaR, even though the random walk VaR suffers from more VaR breaches, which occur when realized losses exceed VaR.

Perhaps a more important contribution of our empirical findings is their implications for the theoretical molding of exchange rates. Random walk intervals are essentially intervals from the *unconditional* distribution of exchange rate changes, whereas intervals based on economic models are intervals from *conditional* distributions. This means that if the true model is a random walk model, then, asymptotically, the length of both the random walk and the economic model have to be exactly the same. So under what circumstances will conditional intervals be tighter than unconditional ones? Here is one example. Suppose economic fundamental X takes of values of 1 and 4 with equal probabilities. Suppose that conditional on X, exchange rate change Y follows a normal distribution: $Y = \sqrt{X}\varepsilon$, where $\varepsilon \sim N(0, 1)$ independent of X. Given realizations of X, symmetric 95% conditional forecast intervals are roughly $(-1.96\sqrt{X}, 1.96\sqrt{X})$.[6] The unconditional forecast interval is roughly $(-3.3, 3.3)$. The 95% conditional intervals have an average length of around 5.88,[7] which is approximately 11% shorter than the unconditional interval length. A tighter conditional interval implies a linkage between economic fundamentals and exchange rate changes, because no difference between the conditional and unconditional distributions will exist if exchange rates are unrelated to fundamentals. In contrast, a point forecast evaluation under the popular linear model $Y = \beta X + u$ will not be able to detect this linkage because any consistent estimate of β will be zero asymptotically,[8] rendering equal model and random walk mean square prediction errors (or any criteria based on forecast errors) in the limit. In this regard, comparisons of interval lengths may reveal connections between exchange rates and fundamentals that point forecast evaluation may overlook.

5. We thank an anonymous referee for suggesting VaR as an example.
6. The choice of 95% here is not critical. What follows holds for other probabilities as well.
7. Since $3.92E(\sqrt{X}) = 5.88$.
8. For instance, because $E(\varepsilon|X) = 0$, least square estimates of beta is zero in the limit.

Several papers have studied out-of-sample exchange rate density (distribution) forecasts. Diebold, Hahn, and Tay (1999) use the RiskMetrics model of J.P. Morgan (1996) to compute half-hour-ahead density forecasts for Deutschmark–dollar and yen–dollar returns. Christoffersen and Mazzotta (2005) provide option-implied density and interval forecasts for four major exchange rates. Boero and Marrocu (2004) obtain 1-day-ahead density forecasts for euro nominal effective exchange rate using self-exciting threshold autoregressive (SETAR) models. Sarno and Valente (2005) evaluate the exchange rate density forecasting performance of the Markov-switching vector equilibrium correction model that is developed by Clarida et al. (2003). They find that information from the term structure of forward premia helps the model to outperform the random walk in forecasting out-of-sample densities of the spot exchange rate. More recently, Hong, Li, and Zhao (2007) construct half-hour-ahead density forecasts for euro–dollar and yen–dollar exchange rates using a comprehensive set of univariate time series models that capture fat tails, time-varying volatility and regime switches.

There are several common features across the studies listed above that make them different from this paper. First, the focus of the above studies is not to make connections between the exchange rate and economic fundamentals. These studies use high frequency data, which are not available for most conventional economic fundamentals.[9] Also, with the exception of Sarno and Valente (2005), all studies focus on univariate time series models. Second, these studies do not consider multi-horizon-ahead forecasts.[10] Lastly, the above studies assume that the densities are analytically defined for a given model. The semiparametric method used in this paper does not impose such restrictions. Our choice of this method is motivated by the fact that macroeconomic models typically do not describe the future *distributions* of exchange rates.[11] By using semiparametric forecast intervals,[12] we do not require the linear models we use to be correctly specified, or contain parametric distributional assumptions.

It is also important to establish what this paper is *not* attempting. First, the inference procedure does not carry the purpose of finding the *correct* model specification. Rather, inference is on how useful models are in generating forecast intervals measured in terms of coverage accuracies and lengths. Second, this paper does not consider alternatives to using semiparametric forecast intervals. Some models might perform better if parametric distributional assumptions (e.g., the forecast errors follow a $t-$GARCH) or other assumptions (e.g., the forecast errors are independent of the predictors) are added. One could presumably estimate the forecast intervals

9. For instance, Diebold, Hahn, and Tay (1999) and Hong, Li, and Zhao (2007) use intraday data.

10. These studies do not consider such forecasts perhaps due to the fact that the models used are often highly nonlinear. Iterating nonlinear density models multiple horizons ahead is analytically difficult, if not infeasible.

11. For instance, the Taylor rule models considered in this paper do not describe any features of the data beyond the conditional means of future exchange rates.

12. For brevity, we omit semiparametric and simply use the term forecast intervals when we believe it causes no confusion.

differently based on the same models, but this exercise lies outside the scope of this paper.

The remainder of this paper is organized as follows. Section 1 describes the models we use, as well as the data. Section 2 illustrates how the semiparametric forecast intervals are constructed from a given model and proposes loss criteria to evaluate the quality of the forecast intervals, the corresponding test statistics, and method to bootstrap critical values. Section 3 presents results of out-of-sample forecast evaluation. Finally, Section 4 contains conclusions and suggestions for future research.

1. MODELS AND DATA

Eight models are used in this paper. Let $m = 1, 2, \ldots, 8$ be the index of these models and the first model be the benchmark Taylor rule model. A general setup of the models takes the form of:

$$s_{t+h} - s_t = \alpha_{m,h} + \beta'_{m,h}\mathbf{X}_{m,t} + \varepsilon_{m,t+h}, \tag{1}$$

where $s_{t+h} - s_t$ is h-period changes of the (log) exchange rate, and $\mathbf{X}_{m,t}$ contains economic variables that are used in model m. Following the literature of long-horizon regressions, both short- and long-horizon forecasts are considered. Models differ in economic variables that are included in matrix $\mathbf{X}_{m,t}$. In the benchmark Taylor rule model,

$$\text{Model 1: } \mathbf{X}_{1,t} \equiv \left[\, \pi_t - \pi_t^* \quad y_t^{gap} - y_t^{gap*} \quad q_t \,\right],$$

where $\pi_t(\pi_t^*)$ is the inflation rate, and $y_t^{gap}(y_t^{gap*})$ is the output gap in the home (foreign) country. The real exchange rate q_t is defined as $q_t \equiv s_t + p_t^* - p_t$, where $p_t(p_t^*)$ is the (log) consumer price index in the home (foreign) country. The next subsection describes the benchmark Taylor rule model in detail.

We also consider the following models that have been studied in the literature:

Model 2: $\mathbf{X}_{2,t} \equiv \left[\pi_t - \pi_t^* \quad y_t^{gap} - y_t^{gap*}\right]$

Model 3: $\mathbf{X}_{3,t} \equiv \left[\pi_t - \pi_t^* \quad y_t^{gap} - y_t^{gap*} \quad i_{t-1} - i_{t-1}^*\right]$, where i_t (i_t^*) is the short-term interest rate in the home (foreign) country.

Model 4: $\mathbf{X}_{4,t} \equiv [\pi_t - \pi_t^* \quad y_t^{gap} - y_t^{gap*} \quad q_t \quad i_{t-1} - i_{t-1}^*]$

Model 5: $\mathbf{X}_{5,t} \equiv q_t$

Model 6: $\mathbf{X}_{6,t} \equiv s_t - [(m_t - m_t^*) - (y_t - y_t^*)]$, where m_t (m_t^*) is the money supply and $y_t(y_t^*)$ is total output in the home (foreign) country.

Model 7: $\mathbf{X}_{7,t} \equiv i_t - i_t^*$

Model 8: $\mathbf{X}_{8,t} \equiv 0$

Models 2–4 are the Taylor rule models studied in Molodtsova and Papell (2009). Model 2 is the constrained benchmark model when PPP always holds. Molodtsova and Papell (2009) include an interest rate lag in models 3 and 4 to take into account potential interest rate smoothing rules of the central bank.[13] Model 5 is the PPP model and model 6 is the monetary model. Model 7 is the forward premium model and model 8 is the random walk model with a drift.[14] Given a date t and horizon h, the objective is to estimate forecast intervals for $s_{t+h} - s_t$ conditional on $\mathbf{X}_{m,t}$. Below we describe the benchmark Taylor rule model and the monetary model that motivate models 1 and 6.

1.1 Benchmark Taylor Rule Model

The benchmark model is the Taylor rule model derived in Engel and West (2005) and Engel, Wang, and Wu (2009). Following Molodtsova and Papell (2009), we focus on models that depend only on *current* levels of inflation and the output gap.[15] The Taylor rule in the home country takes the form of:

$$\bar{i}_t = \bar{i} + \delta_\pi(\pi_t - \bar{\pi}) + \delta_y y_t^{gap}, \tag{2}$$

where $\bar{i}_t$ is the central bank's target for the short-term interest rate at time t, $\bar{i}$ is the long-term equilibrium rate, π_t is the inflation rate, $\bar{\pi}$ is the target inflation rate, and y_t^{gap} is the output gap. We assume that the foreign country follows a similar Taylor rule. In addition, we follow Engel and West (2005) to assume that the foreign country targets the exchange rate in its Taylor rule:

$$\bar{i}_t^* = \bar{i} + \delta_\pi(\pi_t^* - \bar{\pi}) + \delta_y y_t^{gap*} + \delta_s(s_t - \bar{s}_t), \tag{3}$$

where $\bar{s}_t$ is the targeted exchange rate. Assume that the foreign country targets the PPP level of the exchange rate: $\bar{s}_t = p_t - p_t^*$, where p_t and p_t^* are logarithms of home and foreign aggregate prices. In equation (3), we assume that the policy parameters take the same values in the home and foreign countries. Molodtsova and Papell (2009) refer to this case as "homogeneous Taylor rules." Our results hold qualitatively in the case of heterogenous Taylor rules. To simplify the presentation, we assume that the home and foreign countries have the same long-term inflation and interest rates. Such restrictions have been relaxed in our econometric model after including a constant term.

13. The coefficients on lagged interest rates in the home and foreign countries can take different values in Molodtsova and Papell (2009).

14. We also tried the random walk without a drift. It does not change our results.

15. Clarida, Gali, and Gertler (1998) find empirical support for forward-looking Taylor rules. Forward-looking Taylor rules are ruled out because they require forecasts of predictors, which creates additional complications in out-of-sample forecasting.

We do not consider interest rate smoothing in our benchmark model. That is, the actual interest rate (i_t) is identical to the target rate in the benchmark model: $i_t = \bar{i}_t$.[16] Substituting the difference of equations (2) and (3) into uncovered interest-rate parity (UIP), we have:

$$s_t = E_t \left\{ (1-b) \sum_{j=0}^{\infty} b^j (p_{t+j} - p^*_{t+j}) \right. \\ \left. -b \sum_{j=0}^{\infty} b^j \left[\delta_y \left(y^{gap}_{t+j} - y^{gap*}_{t+j} \right) + \delta_\pi (\pi_{t+j} - \pi^*_{t+j}) \right] \right\}, \tag{5}$$

where the discount factor $b = 1/1 + \delta_s$. Under some conditions, Engel, Wang, and Wu (2009) show that the present value asset pricing format in equation (5) can be written into an error-correction form:[17]

$$s_{t+h} - s_t = \alpha_h + \beta_h z_t + \varepsilon_{t+h}, \tag{6}$$

where the deviation of the exchange rate from its equilibrium level is defined as:

$$z_t = s_t - p_t + p^*_t + \frac{b}{1-b} \left[\delta_y \left(y^{gap}_t - y^{gap*}_t \right) + \delta_\pi (\pi_t - \pi^*_t) \right]. \tag{7}$$

We use equation (6) as our benchmark setup in calculating h-horizon-ahead out-of-sample forecasting intervals. According to equation (7), the matrix $\mathbf{X}_{1,t}$ in equation (1) includes economic variables $q_t \equiv s_t + p^*_t - p_t$, $y^{gap}_t - y^{gap*}_t$, and $\pi_t - \pi^*_t$.

1.2 Monetary Model

Assume the money market clearing condition in the home country is:

$$m_t = p_t + \gamma y_t - \alpha i_t,$$

where m_t is the log of money supply, p_t is the log of aggregate price, i_t is the nominal interest rate, and y_t is the log of output. A symmetric condition holds in the foreign country and we use an asterisk in superscript to denote variables in the foreign country. Subtracting the foreign money market clearing condition from the

16. Molodtsova and Papell (2009) consider the following interest rate smoothing rule:

$$i_t = (1-\rho)\bar{i}_t + \rho i_{t-1}, \tag{4}$$

where ρ is the interest rate smoothing parameter. We include these setups in models 3 and 4.

17. While the long-horizon regression format of the benchmark Taylor model is derived directly from the underlying Taylor rule model, this is not the case for the models with interest rate smoothing (models 3 and 4). Molodtsova and Papell (2009) only consider the short-horizon regression for the Taylor rule models. We include long-horizon regressions of these models only for the purpose of comparison.

home, we have:

$$i_t - i_t^* = \frac{1}{\alpha}[-(m_t - m_t^*) + (p_t - p_t^*) + \gamma(y_t - y_t^*)]. \tag{8}$$

The nominal exchange rate is equal to its purchasing power value plus the real exchange rate:

$$s_t = p_t - p_t^* + q_t. \tag{9}$$

The UIP condition in financial market takes the form:

$$E_t s_{t+1} - s_t = i_t - i_t^*. \tag{10}$$

Substituting equations (8) and (9) into (10), we have

$$s_t = (1-b)[m_t - m_t^* - \gamma(y_t - y_t^*) + q_t] + bE_t s_{t+1}, \tag{11}$$

where $b = \alpha/(1+\alpha)$. In the standard monetary model, such as Mark (1995), purchasing power parity holds ($q_t = 0$). Furthermore, it is assumed that $\gamma = 1$. Under these assumptions, solving s_t recursively and applying the "no-bubbles" condition, we have:

$$s_t = E_t \left\{ (1-b) \sum_{j=0}^{\infty} b^j (m_{t+j} - m_{t+j}^* - (y_{t+j} - y_{t+j}^*)) \right\}.$$

1.3 Data

The models and corresponding forecast intervals are estimated using monthly data for 10 OECD countries. The United States is treated as the foreign country in all cases. For each country we synchronize beginning and end dates of the data across all models estimated. The 11 countries and periods considered are: Australia (73:m1-07:m6), Canada (75:m2-07:m6), Denmark (74:m1-07:m6), Germany (73:m1-07:m6), Japan (73:m1-07:m6), New Zealand (77:m10-07:m6), Norway (73:m1-07:m6), Sweden (73:m1-07:m6), Switzerland (75:m9-07:m6), and the United Kingdom (73:m1-06:m3). Data availability dictates the beginning of our sample in each country. We end our sample in June 2007 to avoid the effect of the recent global financial crisis on exchange rates.[18] The exchange rates during this period are affected by factors that are not captured by the exchange rate models in our paper. For instance, flight to quality may have played an important role in driving exchange rate movements during financial crises. Engel and West (2010) find that most of the strength of the

18. The UK data end in March 2006 because its money supply is not available after that date. The exchange rate of the Deutschmark is from FRB H.10 legacy currency. The exchange rate after 1998 is replaced with the euro at a constant exchange rate between the euro and Deutschmark.

dollar in 2008 and 2009 was mainly driven by changes in risk premium. In addition, the nominal interest rate of the United States hit the zero bound in 2009. Thus, the Taylor rule may not be a good measure of monetary policy during the crisis period. As a result of these factors, we do not extend our sample beyond June 2007.

Most of the data are taken from Molodtsova and Papell (2009).[19] The rest of the data are from the International Financial Statistics and the G10 data set of Haver Analytics. With the exception of interest rates, the data are transformed by taking natural logs and then multiplying by 100. The nominal exchange rates are end-of-month rates taken from the Federal Reserve Bank of St. Louis database. Output data (y_t) are proxied by Industrial Production (IP) from the International Financial Statistics (IFS) database. IP data for Australia and Switzerland are only available at a quarterly frequency, and hence are transformed from quarterly to monthly observations using the quadratic-match average option in Eviews. To avoid the look-ahead bias, the quarterly data are transformed to monthly observations in a recursive fashion. Following Engel and West (2006), the output gap (y_t^{gap}) is calculated by quadratically de-trending the IP for each country. The de-trending of output is also done in a recursive fashion.

Prices data (p_t) are proxied by Consumer Price Index (CPI) from the IFS database. Again, CPI for Australia is only available at a quarterly frequency and the quadratic-match average calculated recursively is used to impute monthly observations. Inflation rates are calculated by taking the first differences of the logs of CPIs. The money market rate from IFS (or call money rate) is used as a measure of the short-term interest rate set by the central bank. Finally, M1 is used to measure the money supply for most countries. M0 for the UK is used due to the unavailability of M1 data.

2. ECONOMETRIC METHOD

2.1 Estimating Forecast Intervals

For a given model m, the objective is to estimate the distribution of $s_{t+h} - s_t$ conditional on $\mathbf{X}_{m,t}$ utilizing equation (1). This is the h-horizon-ahead *forecast distribution* of the exchange rate, from which the corresponding *forecast intervals* can be derived. For a given α, the forecast interval of coverage $\alpha \in (0, 1)$ is an interval in which $s_{t+h} - s_t$ should lie with a probability of α, conditional on $\mathbf{X}_{m,t}$.

Models $m = 1, \ldots, 8$ in Section 1 are essentially point forecast models. To construct forecast intervals for a given model, we apply the methodology of Wu (Forthcoming) to all models. An α-coverage forecast interval of $s_{t+h} - s_t$ conditional on $\mathbf{X}_{m,t}$ can be obtained by the following three-step procedure:

Step 1. Estimate model m by OLS and obtain residuals $\widehat{\varepsilon}_{m,j+h} \equiv s_{j+h} - s_j - \widehat{\alpha}_{m,h} + \widehat{\beta}'_{m,h}\mathbf{X}_{m,j}$ for $j = 1, \ldots, t - h$.[20]

19. We thank the authors for the data, which we downloaded from David Papell's website. For the exact line numbers and sources of the data, see the data appendix of Molodtsova and Papell (2009).

20. In a rolling forecast (described in the next section), the sample does not necessarily have to start at period 1. The forecast interval at time $t + 1$ will be constructed using data from $j = 2, ..., t + 1$, for instance.

Step 2. For a range of values of ε,[21] estimate the conditional distribution of $\varepsilon_{m,t+h} \mid \mathbf{X}_{m,t}$ by:

$$\widehat{P}(\varepsilon_{m,t+h} \leq \varepsilon \mid \mathbf{X}_{m,t}) \equiv \frac{\sum_{j=1}^{t-h} 1(\widehat{\varepsilon}_{m,j+h} \leq \varepsilon)\mathbf{K}_b(\mathbf{X}_{m,j} - \mathbf{X}_{m,t})}{\sum_{j=1}^{t-h} \mathbf{K}_b(\mathbf{X}_{m,j} - \mathbf{X}_{m,t})}, \tag{12}$$

where $\mathbf{K}_b(\mathbf{X}_{m,j} - \mathbf{X}_{m,t}) \equiv b^{-d}\mathbf{K}((\mathbf{X}_{m,j} - \mathbf{X}_{m,t})/b)$ and $\mathbf{K}(\cdot)$ is a multivariate Gaussian kernel with a dimension the same as that of $\mathbf{X}_{m,t}$, and b is the smoothing parameter or bandwidth.[22]

Step 3. Find the $(1-\alpha)/2$ and $(1+\alpha)/2$ quantiles of the estimated distribution, $\widehat{\varepsilon}_{m,h}^{(1-\alpha)/2}(\mathbf{X}_{m,t})$ and $\widehat{\varepsilon}_{m,h}^{(1+\alpha)/2}(\mathbf{X}_{m,t})$, respectively. The estimate of the α-coverage forecast interval for $s_{t+h} - s_t$ conditional on $\mathbf{X}_{m,t}$ is:

$$\widehat{I}_{m,h}^{\alpha}(\mathbf{X}_{m,t}) \equiv \left(\widehat{\beta}'_{m,h}\mathbf{X}_{m,t} + \widehat{\varepsilon}_{m,h}^{(1-\alpha)/2}(\mathbf{X}_{m,t}), \quad \widehat{\beta}'_{m,h}\mathbf{X}_{m,t} + \widehat{\varepsilon}_{m,h}^{(1+\alpha)/2}(\mathbf{X}_{m,t})\right). \tag{13}$$

The above procedure uses the forecast models in equation (1) to estimate the location of the forecast distribution, while the nonparametric kernel estimator estimates the shape. Wu (Forthcoming) shows that under suitable regularity conditions, this method consistently estimates the forecast distribution conditional on the predictors, whether or not the model is functionally misspecified.[23] Stationarity of economic variables is one of the regularity conditions. In our models, exchange rate differences and inflation rates are widely recognized as stationary, while empirical tests for real exchange rates, interest rates and output gaps generate mixed results. We take the stationarity of these variables as given.

Model 8 is the random walk model. The estimator in equation (12) becomes the empirical distribution function (EDF) of the exchange rate changes (i.e., $\mathbf{X}_{8,t} = 0$). The forecast intervals of economic models are compared with those of the random walk. The goal is to test whether forecast intervals based on economic models are more accurate than those based on the random walk model. The criteria used are coverage accuracies and lengths of forecast intervals.

Following Christoffersen (1998) and related work, we first test the coverage accuracy of the intervals. Forecast intervals with nominal coverage of α should in population cover out-of-sample realization 100% of the time. If 95% forecast intervals contain out-of-sample observations only 50% of the time, say, these forecasts cannot be trusted and are referred to as *undercoverage*. In contrast, *overcoverage* implies that intervals can be improved in terms of tightness. Thus, we say that a

21. Usually sorted residuals $\{\widehat{\varepsilon}_{m,j+h}\}_{j=1}^{t-h}$.

22. b is selected using a modified version of Silverman's rule of thumb. We repeated the estimation and inference for the Australian dollar by selecting b using a more rigorous bootstrap procedure based on the suggestion of Hall, Wolff, and Yao (1999). We did not find any qualitative difference between this method and Silverman's rule of thumb for the Australian dollar. However, this method imposes considerable computational burden and therefore was not used for the remaining countries.

23. It is consistent in the sense of convergence in probability as t goes to infinity.

model outperforms the random walk if its forecast intervals cover more accurately than those of the random walk.[24]

If intervals from the model and random walk have equal coverage accuracies, the next logical step is to investigate whether model-based intervals are tighter than random walk intervals. As previously discussed, tighter forecast intervals suggest that the model is more informative, given that competing intervals have equally accurate coverages. The coverage accuracy and length tests are conducted at various horizons for the seven economic models relative to the random walk for each of the 10 OECD exchange rates.

2.2 *Coverage and Length Null Hypotheses*

In this section, the null hypotheses for the coverage accuracy and length tests are introduced. Let the sample size be T and all data are collected in a vector $\mathbf{W}_t$. The forecast intervals are based on a rolling estimation scheme. Let $T = R + N$ and R be the size of the rolling window. For each horizon h and model m, a sequence of $N(h) \equiv N + 1 - h$ forecast intervals are generated using rolling data: $\{\mathbf{W}_t\}_{t=1}^{R}$ for forecast for date $R + h$, $\{\mathbf{W}_t\}_{t=2}^{R+1}$ for forecast for date $R + h + 1$, and so on, until forecast for date T is generated using $\{\mathbf{W}_t\}_{t=N(h)}^{R+N(h)-1}$.

For each h we have $N(h)$ observations to compare the coverages and lengths across models out-of-sample. For each null, we define a forecast loss. As defined in equation (13), let $\widehat{I}_{m,h}^{\alpha}(\mathbf{X}_{m,t})$ be the h-horizon-ahead forecast interval of model m with a nominal coverage α, conditional on $\mathbf{X}_{m,t}$, using a rolling window of data that ends at t, and let $I_{m,h}^{\alpha}(\mathbf{X}_{m,t})$ be its population counterpart. The *coverage accuracy loss* is defined as:

$$CL_{m,h}^{\alpha} = \left[E1\left(Y_{t+h} \in I_{m,h}^{\alpha}(\mathbf{X}_{m,t})\right) - \alpha\right]^2, \tag{14}$$

where $1(\cdot)$ is the indicator function that takes on the value of one if the expression in the bracket is true and zero otherwise. For economic models ($m = 1, \dots, 7$), the goal is to compare the coverage accuracy loss of forecast intervals of model m with that of the random walk ($m = 8$). For $m = 1, \dots, 7$, the null and alternative hypotheses are:

$$H_0 : \Delta CL_{m,h}^{\alpha} \equiv CL_{8,h}^{\alpha} - CL_{m,h}^{\alpha} = 0 \tag{15}$$

$$H_A : \Delta CL_{m,h}^{\alpha} \neq 0. \tag{16}$$

This null is similar to the unconditional coverage test of Christoffersen (1998) but differs in that it does not require that the coverage accuracy loss from the model be

24. Under suitable regularity conditions, semiparametric forecast intervals of a model (as well as the random walk) cover 100% of out-of-sample observations as the estimation sample tends to infinity. Thus, the model and random walk will have equal coverage accuracies in the limit. In the context of rolling estimation used in this paper, we view the coverage accuracy test as providing inference on how well the intervals cover when estimated with a rolling window of finite size.

zero: it only tests that the coverage loss is the same for the model and random walk. Notice also that the loss function is symmetric for under- and overcoverage. In some cases, an asymmetric loss function may be more suitable and we leave the issue of asymmetric loss functions to future work.

Given the importance of coverage accuracy, the coverage accuracy test is used as a first pass for comparing models with the random walk. If the null hypothesis is rejected in favor of the model (random walk), we conclude that the model (random walk) generates better forecast intervals. If the the equal coverage accuracies null hypothesis cannot be rejected, we then test whether the model generates tighter intervals than the random walk. The comparison is valid because forecast intervals are equally accurate in covering realized exchange rates. To that end, we define the *length loss* as:

$$LL^{\alpha}_{m,h} \equiv E\left[leb\left(I^{\alpha}_{m,t}(\mathbf{X}_{m,t})\right)\right] \tag{17}$$

where $leb(\cdot)$ is the Lesbesgue measure. To compare the length loss of forecast intervals of economic models $m = 1, 2, \ldots, 7$ with that of the random walk ($m = 8$), the null and alternative hypotheses are:

$$H_0 : \Delta LL^{\alpha}_{m,h} \equiv LL^{\alpha}_{8,h} - LL^{\alpha}_{m,h} = 0 \tag{18}$$

$$H_A : \Delta LL^{\alpha}_{m,h} \neq 0. \tag{19}$$

For a given model, if we fail to reject (16), but go on to reject (18) in favor of the model (random walk), we conclude that the model (random walk) generates better forecast intervals.[25]

We use a bootstrap inference procedure to test coverage accuracy and length hypotheses.[26] Bootstrapping allows us to confront two important evaluation issues. First, it allows us to evaluate forecast intervals that are not parametrically derived. Evaluation methods developed in well-known studies such as Christoffersen (1998), Diebold, Gunther, and Tay (1998), Corradi and Swanson (2006a), and references within Corradi and Swanson (2006b) require the forecast density or intervals to be parametrically specified. The bootstrap is not limited to such restrictions and allows comparisons among parametric and semiparametric forecasts with different rates of convergence. Second, bootstrapping allows us to easily test joint hypotheses, such as whether the random walk has the tightest interval out of all models.

25. At the suggestion of an anonymous referee, we also conducted inference for interval lengths based on one-sided tests: $H_0 : \Delta LL^{\alpha}_{m,h} \leq 0$ and $H_A : \Delta LL^{\alpha}_{m,h} > 0$. The only difference between one-sided and two-sided tests is the definition of the rejection region. Results are available upon request.

26. In a previous version of the paper, we conducted asymptotic inference based on application of the results in Giacomini and White (2006). We thank the editor for suggesting the bootstrap alternative.

2.3 Bootstrap Inference

Define the sample analog of the coverage accuracy loss in equation (14):

$$\widehat{CL}^{\alpha}_{m,h} = \left(N(h)^{-1} \sum_{t=R}^{T-h} 1\big(Y_{t+h} \in \widehat{I}^{\alpha}_{m,h}(\mathbf{X}_{m,t})\big) - \alpha \right)^2 .$$

A studentized test statistic is used in our study.[27] We apply the Delta method to equation (14) and set the denominator of the test statistics as $\sqrt{\widehat{\Gamma}'_{m,h}\widehat{\Omega}_{m,h}\widehat{\Gamma}_{m,h}}$, where $\widehat{\Omega}_{m,h}$ is the Newey and West (1987) estimator[28] applied to the vectors $\{1(Y_{t+h} \in \widehat{I}^{\alpha}_{m,h}(\mathbf{X}_{m,t}))\}_{t=R}^{T-h}$ and $\{1(Y_{t+h} \in \widehat{I}^{\alpha}_{8,h}(\mathbf{X}_{8,t}))\}_{t=R}^{T-h}$, and

$$\widehat{\Gamma}_{m,h} \equiv \begin{bmatrix} 2\left(N(h)^{-1} \sum_{t=R}^{T-h} 1\big(Y_{t+h} \in \widehat{I}^{\alpha}_{m,h}(\mathbf{X}_{m,t})\big) - \alpha \right) \\ 2\left(N(h)^{-1} \sum_{t=R}^{T-h} 1\big(Y_{t+h} \in \widehat{I}^{\alpha}_{8,h}(\mathbf{X}_{8,t})\big) - \alpha \right) \end{bmatrix}.$$

The test statistic for the equal coverage accuracies null hypothesis in (16) is defined as:

$$Ct^{\alpha}_{m,h} \equiv \frac{\sqrt{N(h)}\Delta\widehat{CL}^{\alpha}_{m,h}}{\sqrt{\widehat{\Gamma}'_{m,h}\widehat{\Omega}_{m,h}\widehat{\Gamma}_{m,h}}}. \tag{20}$$

Similarly, the sample analog of the length loss in equation (17) is defined as:

$$\widehat{LL}^{\alpha}_{m,h} = N(h)^{-1} \sum_{t=R}^{T-h} leb\big(\widehat{I}^{\alpha}_{m,h}(\mathbf{X}_{m,t})\big).$$

The test statistic is studentized with the denominator $\widehat{\Sigma}_{m,h}$, which is the Newey and West (1987) estimator applied to the vector $\{leb(\widehat{I}^{\alpha}_{8,h}(\mathbf{X}_{m,t})) - leb(\widehat{I}^{\alpha}_{m,h}(\mathbf{X}_{m,t}))\}_{t=R}^{T-h}$. The test statistic for the equal length null hypothesis in (18) is defined as:

$$Lt^{\alpha}_{m,h} \equiv \frac{\sqrt{N(h)}\Delta\widehat{LL}^{\alpha}_{m,h}}{\sqrt{\widehat{\Sigma}_{m,h}}}. \tag{21}$$

27. In the context of bootstrapping predictive ability tests, Hansen (2005) and Romano and Wolf (2005) suggest that studentized test statistics perform well.

28. We use a window width of 12 since the longest horizon is 12 months.

Bootstrap critical values are obtained by calculating bootstrapped versions of equations (20) and (21) over 2000 bootstrapped samples of $\{\mathbf{W}_t\}_{t=1}^T$, using the steps below:[29]

Step 1. Resample blocks of $\{\mathbf{W}_t\}_{t=1}^T$ with replacement, using the stationary bootstrap of Politis and Romano (1994). Specifically, let $z_1, z_2, \ldots$ be random integers from $\{1, ..., T\}$. Pick the first bootstrap observation, $\mathbf{W}_1^* = \mathbf{W}_{z_1}$. Let $\mathbf{W}_2^* = \mathbf{W}_{z_2}$ with probability p and $\mathbf{W}_2^* = \mathbf{W}_{z_1+1}$ with probability $1 - p$.[30] The bootstrap is circular in that if the $T + 1$ observation from the original sample needs to be used, one takes the first observation instead. Continue in this manner until $\mathbf{W}_T^*$ has been chosen.

Step 2. For each pseudo-sample $\{\mathbf{W}_t^*\}_{t=1}^T$, re-estimate the forecast intervals, and calculate the bootstrap version of (20) and (21) as:

$$Ct_{m,h}^{\alpha*} = \frac{\sqrt{N(h)}\left(\Delta CL_{m,h}^{\alpha*} - \Delta\widehat{CL}_{m,h}^{\alpha}\right)}{\sqrt{\Gamma_{m,h}^{*'}\Omega_{m,h}^{*}\Gamma_{m,h}^{*}}}$$

$$Lt_{m,h}^{\alpha*} = \frac{\sqrt{N(h)}\left(\Delta LL_{m,h}^{\alpha*} - \Delta\widehat{LL}_{m,h}^{\alpha}\right)}{\sqrt{\Sigma_{m,h}^{*}}},$$

where $\Delta CL_{m,h}^{\alpha*}$, $\Delta LL_{m,h}^{\alpha*}$, $\Gamma_{m,h}^{*}$, $\Omega_{m,h}^{*}$ and $\Sigma_{m,h}^{*}$ are estimated in the same way as $\Delta\widehat{CL}_{m,h}^{\alpha}$, $\Delta\widehat{LL}_{m,h}^{\alpha}$, $\widehat{\Gamma}_{m,h}$, $\widehat{\Omega}_{m,h}$ and $\widehat{\Sigma}_{m,h}$, respectively, but using $\{\mathbf{W}_t^*\}_{t=1}^T$ instead of $\{\mathbf{W}_t\}_{t=1}^T$.

Step 3. Repeat steps 1 and 2 above 2,000 times, forming bootstrap null distributions for $Ct_{m,h}^{\alpha*}$ and $Lt_{m,h}^{\alpha*}$. Compare the test statistics $Ct_{m,h}^{\alpha}$ and $Lt_{m,h}^{\alpha}$ with the bootstrap null distributions. Reject the null hypotheses if $Ct_{m,h}^{\alpha}$ and $Lt_{m,h}^{\alpha}$ lie in the appropriate rejection regions of the distributions.

While showing the theoretical validity of the bootstrap tests is out of the scope of this paper, we note that White (2000), Romano and Wolf (2005), and Corradi and Swanson (2007) analyze theoretical aspects of block bootstrap inference in the context of tests on predictive ability. In particular, the mechanics of the procedure above is a combination of the procedure in White (2000) and Corradi and Swanson (2007), as it resamples the entire time series in order to repeat both forecasting and evaluation exercises in each bootstrap iteration.[31]

29. We also conducted inference based on the more restrictive Gaussian bootstrap, in which the exchange rate is assumed to follow a Gaussian random walk, and economic variables followed a vector autoregression. Results based on the Gaussian bootstrap do not differ qualitatively from those based on block bootstrap. Results are available upon request.

30. The data used exhibit high degrees of serial correlation. Therefore, we select $p = 0.1$.

31. We note that our framework is different from that of Corradi and Swanson's (2007) in terms of how we center the bootstrap test statistics, and that we use rolling instead of recursive estimation.

3. RESULTS

We estimate forecast intervals with a nominal coverage of $\alpha = 0.95$ from each model. Auxiliary results show that our qualitative findings do not depend on the choice of α, as long as its sufficiently large. The size of the rolling window (R) is set to 200 for all countries. From our experience, tampering with R does not change the qualitative results, unless R is chosen to be unusually big or small.

For time horizons $h = 1, 3, 6, 12$ and models $m = 1, \ldots, 8$, we construct a sequence of $N(h)$ 95% forecast intervals $\{\widehat{I}^{0.95}_{m,t+h}\}_{t=R}^{T-h}$ for the h-horizon change of the exchange rate $s_{t+h} - s_t$. Then we compare economic models and the random walk by computing empirical coverages, lengths and test statistics $Ct^{0.95}_{m,h}$ and $Lt^{0.95}_{m,h}$ described in Section 1. For a test at 10% confidence level, a p-value[32] above 95% indicates rejection of the null hypothesis in favor of the economic model and a p-value below 5% indicate rejection in favor of the random walk. Table 1 summarizes our main results. The table shows the comparison between the seven economic models and the random walk model for the horizons of one and twelve months. For each horizon, the "Coverage Test" column reports the results of coverage accuracy tests. The "Length Test" column reports results of length tests for currencies that fail to reject the null of equal coverage accuracy. The "Overall" column is the sum of the "Coverage Test" and "Length Test" columns. The "Model Better" column show the number of exchange rates (out of 10) for which the economic models perform better than the random walk under corresponding criterion. For instance, the first entry (5) of Table 1 means that in 5 out of 10 exchange rates, the benchmark Taylor rule model (Taylor 1) has more accurate coverages than the random walk in 1-month-ahead forecasts.

The benchmark Taylor rule model (Taylor 1) shows strong exchange rate predictability at both 1-month and 12-month horizons. In 8 out of 10 exchange rates, the benchmark Taylor rule model has either more accurate coverages or tighter forecast intervals given the same coverage accuracies compared with the random walk at 1 month. At a 12-month horizon, the benchmark Taylor rule model outperforms the random walk model in every case. Other Taylor rule models perform similarly to the benchmark model, though not quite as well, indicating that our findings are reasonably robust to different setups of the Taylor rule model. Other exchange rate models do not perform as well as the benchmark Taylor rule model. The PPP model shows exchange rate predictability only at the long horizon (12 months) and the performance of the Monetary model becomes worse at 12 months. The forward premium model is the worst among all economic models, which is not surprising given the empirical failures of UIP.

Notably, the evidence of exchange rate predictability is stronger in the length tests than in the coverage accuracy tests. There are 70 coverage accuracy tests and 70 length tests at each horizon.[33] At 1 month, the evidence of exchange rate predictability is

32. That is, the percentage of bootstrapped test statistics $Ct^{0.95*}_{m,h}$ and $Lt^{0.95*}_{m,h}$ that are *below* $Ct^{0.95}_{m,h}$ and $Lt^{0.95}_{m,h}$, respectively.

33. Due to 10 exchange rates and seven models.

TABLE 1

SUMMARY OF RESULTS

	Horizon = 1 month						Horizon = 12 months					
	Coverage test		Length test		Overall		Coverage test		Length test		Overall	
	Model better	RW better	Model better	RW better	Model better	RW better	Model better	RW better	Model better	RW better	Model better	RW better
Taylor 1	5	0	3	0	8	0	4	0	6	0	10	0
Taylor 2	2	3	5	0	7	3	3	2	3	0	6	2
Taylor 3	4	1	5	0	9	1	2	2	3	1	5	3
Taylor 4	7	2	1	0	8	2	2	0	6	0	8	0
PPP	2	2	3	1	5	3	2	1	5	0	7	1
Monetary	3	2	4	0	7	2	0	8	1	0	1	8
FP	0	7	3	0	3	7	1	5	4	0	5	5
Total	23	17	24	1	47	18	14	18	28	1	42	19

NOTE:
– This table summarizes the results of the coverage accuracy tests and length tests that are reported in Tables 2–8. These tests are for 95% forecast intervals and additional details can be found in Tables 2–8.
– Each test compares the random walk model with one of the following seven exchange rate models for 10 U.S. dollar-foreign currency exchange rates listed in Tables 2–8. Models "Taylor 1" to "Taylor 4" are the Taylor rule models. "PPP" is the purchasing power parity model. "Monetary" is the monetary model and "FP" is the forward premium model. See Section 1 for details of these models. RW is the abbreviation of Random Walk.
– The coverage test evaluates the coverage accuracy of forecast intervals. The length test evaluates the length of forecast intervals and is conducted only for the currencies that fail to reject the null of equal coverage accuracy. The bootstrap procedure used to conduct the tests is described in Section 1.3.
– For each horizon and model, the "Coverage Test" column reports the results of two-sided tests of equal empirical coverage accuracy for ten currencies at the 10% level. A better model is the one with more accurate empirical coverages. The "Length Test" column reports the results of two-sided length tests at the 10% level for currencies that fail to reject the null hypothesis in the coverage test. A better model is the one with tighter forecast intervals.
– For each horizon and test, entries in the columns labeled "Model Better" are the number of exchange rates for which the economic model performs better than the random walk under the corresponding test. For instance, the first entry (5) of the table means that in 5 out of 10 exchange rates, the null of equal coverage accuracy is rejected at the 10% level in favor of model Taylor 1 (p-value greater than 95%) in 1-month-ahead forecasts. Entries in the columns labeled "RW Better" are the number of exchange rates for which the null hypothesis is rejected in favor of the random walk (p-value less than 5%) at the 10% level.
– For each horizon and model, the "Overall" column is the sum of the "Coverage Test" and "Length Test" columns. Ten is the maximum value of the sum of the two figures in the "Overall" column because we have 10 currencies in total and the length test is conducted only for currencies that fail to reject the null hypothesis of equal coverage accuracy.
– The data are monthly data with most currencies starting in 1973m1 and ending in 2007m6. Some currencies have different sampling periods due to data unavailability. See Section 1.3 for exact dates. The out-of-sample forecasting starts after the 200th observation of each currency and about 210 (200) forecasts are generated at the 1-month (12-month) horizon.

not very strong based on coverage accuracy tests: economic models have more accurate coverages than the random walk in 23 cases, compared to 17 cases where the random walk has more accurate coverage. In contrast, economic models have tighter forecast intervals than the random walk, given equal coverage accuracies in 24 out of 30 cases, while the random walk has tighter forecast intervals in only 1 case. Similar patterns prevail at 12 months. Out of 38 cases where economic and random walk models have statistically equal coverage accuracies, economic models have tighter forecast intervals in 28 cases. In contrast, the random walk model has tighter intervals than economic models in only 1 cases. Overall, the length tests made important contributions in finding exchange rate predictability, particularly at long horizons. In 28 out of 42 cases where economic models outperform the random walk model at 12 months, the evidence of exchange rate predictability comes from length tests rather than coverage accuracy tests.

In the next subsection, we provide more details on our benchmark Taylor rule model results. After that, we report and discuss the results of alternative models.

3.1 Results of Benchmark Taylor Rule Model

Table 2 shows the results of the benchmark Taylor rule model. For each time horizon h and exchange rate, the first column (RW Cov.) reports coverages of random walk forecast intervals with a nominal coverage of 95%. The second column (Cov.) reports coverages of the benchmark Taylor rule model. The third column (RW Leng.) reports lengths of random walk forecast intervals. The last column (Leng./RWLeng.) reports lengths of the benchmark Taylor rule model intervals relative to those of the random walk intervals.[34] Superscripts * in the columns of "RW Cov." (Cov,) denote rejections of equal coverage accuracies in favor of the random walk model (the economic model) at a significance level of 10%. Superscripts * in the columns of "RW Leng." (Leng./RWLeng.) denote rejections of equal lengths in favor of the random walk model (the economic model) at a significance level of 10%. Note that these length tests are conducted only for currencies that fail to reject the null of equal coverage accuracy.

The middle panel of Table 2 summarizes the results of these exchange rates. In "Coverage" row, the "Model Better" columns report the number of exchange rates for which the Taylor rule model has more accurate coverages than the random walk. The "RW Better" columns report the number of exchange rates for which the random walk has more accurate coverages. In the "Length" row, a better model is the one with tighter forecast intervals given equal coverage accuracies. The "Overall" row is the sum of the "Coverage" and "Length" rows. We find strong evidence of predictability for the benchmark Taylor rule model: the model intervals either have more accurate coverages or are tighter given equal coverage accuracies for horizons $h = 1, 6, 12$.[35] In 30 coverage tests for 10 exchange rates at the three horizons, there are 11 rejections of equal coverage accuracies in favor of the Taylor rule model and 1 rejection in favor of the random walk. For the remaining 18 cases where the Taylor rule model and the random walk have statistically equal coverage accuracies, 12 length tests are rejected in favor of the Taylor rule model. In general, length tests contributed significantly in finding exchange rate predictability. For instance, at the 6-month horizon, in 3 out of the 5 cases where the Taylor rule model outperforms the random walk, evidence comes from length tests rather than coverage accuracy tests.

Note that the benchmark Taylor rule model improves forecast interval lengths without sacrificing coverage accuracies. For most exchange rates and time horizons, the Taylor rule model has tighter intervals, while its coverages are equally or even more accurate than those of the random walk. The only exception is for the Swedish Krona at 6 months, where the Taylor rule model has wider forecast intervals than the random walk. The improvement in the length of forecast intervals is also quantitatively substantial, especially at long horizons. At 12 months, the intervals of the Taylor rule model are usually more than 10% tighter than those of the random walk.

34. Empirical coverages and lengths reported in this and the following tables are averages across $N(h)$ out-of-sample trials.

35. Similar patterns exist for horizon $h = 3$. We do not report its results for brevity.

TABLE 2

RESULTS OF BENCHMARK TAYLOR RULE MODEL

	Horizon = 1 month				Horizon = 6 months				Horizon = 12 months			
	RW Cov.	Cov.	RW Leng.	$\frac{\text{Leng.}}{\text{RW Leng.}}$	RW Cov.	Cov.	RW Leng.	$\frac{\text{Leng.}}{\text{RW Leng.}}$	RW Cov.	Cov.	RW Leng.	$\frac{\text{Leng.}}{\text{RW Leng.}}$
Australia	96.24%	94.84%*	9.147	98.36%	97.12%	96.63%	24.939	96.37%	94.06%	94.06%	39.372	88.46%*
Canada	88.83%	88.30%	4.178	97.75%	90.16%*	85.79%	12.309	90.52%	88.14%	87.01%	18.732	80.65%*
Denmark	97.51%	97.51%	10.521	96.79%*	98.47%	97.45%*	29.784	95.14%	97.89%	96.84%*	46.495	88.56%
Germany	97.18%	96.71%*	10.993	96.81%	97.60%	98.08%	31.610	90.85%*	98.51%	97.52%*	48.683	85.41%
Japan	96.24%	95.31%*	11.680	97.77%	96.63%	96.15%	33.431	94.92%*	96.04%	96.04%	53.479	82.77%*
New Zealand	96.15%	94.87%*	10.899	98.79%	93.38%	92.05%	30.962	91.15%	91.03%	93.10%	49.853	86.81%*
Norway	95.31%	94.84%	9.436	93.89%*	93.75%	93.27%	28.843	92.90%	94.55%	92.08%	38.283	90.46%*
Sweden	94.37%	93.90%	9.289	97.91%*	94.23%	95.19%	29.447*	101.52%	91.09%	93.56%*	47.137	88.33%
Switzerland	98.90%	97.24%*	11.688	99.38%	98.86%	98.30%*	34.111	88.85%	99.41%	98.24%*	50.371	81.48%
UK	96.46%	95.96%	10.432	98.83%	94.82%	93.26%	32.689	87.40%*	98.40%	88.24%	43.290	85.86%*

	Model better	RW better	Model better	RW better	Model better	RW better
Coverage†	5	0	2	1	4	0
Length‡	3	0	3	1	6	0
Overall§	8	0	5	2	10	0

(*Continued*)

TABLE 2

CONTINUED

	Probability (out of the $N(h)$ out-of-sample forecasts) that the model forecast interval length < RW forecast interval length			
	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Australia	67.61%	74.41%	58.65%	100.00%
Canada	94.68%	100.00%	99.45%	100.00%
Denmark	86.57%	61.31%	76.02%	97.89%
Germany	91.08%	95.73%	100.00%	100.00%
Japan	86.38%	72.99%	97.12%	100.00%
New Zealand	52.56%	75.32%	96.03%	100.00%
Norway	98.12%	84.36%	96.63%	85.64%
Sweden	84.51%	81.52%	24.52%	100.00%
Switzerland	56.91%	74.30%	100.00%	100.00%
UK	71.21%	100.00%	96.89%	100.00%

NOTE:

– For each horizon, the first column (RW Cov.) reports empirical coverages of the random walk intervals with a nominal coverage of 95%. The second column (Cov.) reports empirical coverages of the economic model. The third column (RW Leng.) reports the lengths of random walk intervals. The fourth column (Leng./RW Leng.) reports the lengths of the economic model divided by the length of the random walk. Empirical coverages and lengths are averages across $N(h)$ out-of-sample trials.

– Superscripts $*$ in the "RW Cov." (Cov,) columns denote rejections of equal coverage accuracies in favor of the random walk model (the economic model) at a significance level of 10%.

– Superscripts $*$ in the "RW Leng." (Leng./RW Leng.) columns denote rejections of equal lengths in favor of the random walk model (the economic model) at a significance level of 10%. Note that this test is conducted only for currencies that fail to reject the null of equal coverage accuracy.

†– A better model is the one with more accurate coverages.

‡– A better model is the one with tighter forecast intervals for currencies that fail to reject the null of equal coverage accuracy.

– Sum of the "Coverage" and "Length" rows.

– h denotes forecast horizons for monthly data.

Given the trade-off between out-of-sample coverages and lengths, equal length tests may favor models that systematically under-cover. However, the length improvement of the benchmark Taylor rule model is not driven by systematic undercoverage. At 12 months, the benchmark Taylor rule model and random walk have exactly the same coverages for the Australian dollar and the yen, but the Taylor rule model has substantially tighter intervals. In addition, Taylor rule intervals for the New Zealand dollar have larger empirical coverages than random walk intervals, but still Taylor rule intervals are tighter.

As for each individual exchange rate, the benchmark Taylor rule model works best for the Danish krone, the Deutschmark, the Japanese yen, and the Swiss franc: for all horizons, the model has more accurate coverages or tighter forecast intervals given equal coverage accuracies compared to the random walk. The Taylor rule model performs better than the random walk in most horizons for the remaining exchange rates except the Canadian dollar, for which the Taylor rule model outperforms the random walk only at long horizons. To illustrate the economic usefulness of the benchmark Taylor rule model, we use VaR as an example. Consider an investor with a short position in the Australian dollar. The actual gain/loss of the short position depends on changes in the exchange rate and the interest rate differential between the Australian dollar and the U.S. dollar. For simplicity, we abstract from the interest rate differential in our discussion below. In this case, a depreciation in the Australian dollar (or $s_{t+h} - s_t > 0$) is desirable. Because the VaR is a one-sided measure, the "Model 2.5%" and "RW 2.5%" lines in Figure 2(c) later in this article display the 97.5% (instead of 95% confidence band in our two-sided tests), 1-year-ahead VaRs based on the benchmark Taylor rule and random walk models, respectively. Note that the random walk VaR is on average 13% larger than Taylor rule VaR over 202 1-year-ahead forecasts. Even so, the random walk VaR is not necessarily more conservative: it was breached by realized losses 12 out of 202 times ($\approx$6%), whereas the ideal number is 5 ($\approx$2.5%). This compares unfavorably to the Taylor rule VaR, which was breached only 8 times. Furthermore, the average size of the breach is on average 34% of the VaR for the random walk, compared to just 10% for the Taylor rule. Overall, the random walk VaR is too conservative in some periods (e.g., after 2005), and not conservative enough in others (e.g., from June 2003 to June 2004). An overly conservative VaR, over time, may induce the investor to overhedge his position or capitalize more than what is necessary to buffer potential losses,[36] consequently reducing potential gains from taking other investment opportunities.[37]

36. We use these concepts only heuristically and intuitively in this example. Formal assessments of whether a given position is overhedged or overcapitalized require more rigorous frameworks, such as the one in Hung, Chiu, and Lee (2006).

37. Consider also the long position in the Australian dollar. The corresponding Taylor rule and random walk VaRs are the lines "Model 97.5%" and "RW 97.5%" in Figure 2(c), respectively. In this case, the random walk VaR suffers no breaches, while Taylor rule VaR has four breaches. However, the fact that the random walk VaR has no breaches (the ideal number should be 5) and that it is on average more than 10% larger than Taylor rule VaR indicate that it is overly conservative. Furthermore, random walk VaR inadequately reflects the dynamics in the Australian dollar changes, as evident in the figure.

3.2 Results of Other Models

Tables 3–8 reportresults of the remaining six models (three alternative Taylor rule models, PPP, monetary, and forward premium models). In general, the results for the coverage accuracy tests do not show strong evidence that model intervals cover more accurately than random walk intervals, at both short and long horizons. However, after taking length tests into consideration, we find that economic models perform better than the random walk, especially at long horizons. Taylor rule model 4 (the benchmark model with interest rate smoothing in Table 5) and the PPP model (Table 6) perform the best among the six models. Results of these two models are very similar to that of the benchmark Taylor rule model. At 12 months, both models outperform the random walk for most exchange rates. The performances of Taylor rule models 2 (Table 3) and 3 (Table 4) are less impressive than other models at long horizons. Still, for more than half of the exchange rates, the models outperform the random walk at various horizons.

Comparing the benchmark Taylor rule model, the PPP model and the monetary model, the performance of the PPP model (Table 6) is worse than the other two models at short horizons. Compared to the Taylor rule and PPP models, the monetary model outperforms the random walk only for a smaller number of exchange rates at 6 and 12 months. Overall, the benchmark Taylor rule model seems to perform better than the monetary and PPP models, a finding that echoes the results of Molodtsova and Papell (2009).

Our results are also robust under heterogeneous Taylor rules. In this model, we relaxed the assumption that the Taylor rule coefficients are the same in the home and foreign countries, and replace $\pi_t - \pi_t^*$ and $y_t^{gap} - y_t^{gap*}$ in $\mathbf{X}_{1,t}$ with $\pi_t, \pi_t^*, y_t^{gap}$ and y_t^{gap*}. We find qualitatively similar results as in Taylor rule models 1–4.[38]

3.3 Discussion and Additional Results

After Mark (1995) finds exchange rate predictability at long horizons, long-horizon exchange rate predictability has become an active research area. With panel data, Engel, Mark, and West (2007) document that long-horizon predictability is relatively robust in the exchange rate forecasting literature. In this paper, the evidence of long-horizon predictability is robust across different models and currencies. At 12 months, all Taylor rule models outperform the random walk for the following exchange rates: the Danish krone, Deutschmark, and Swiss franc.

The lower panel of Tables 2–8 reports the probability (out of $N(h)$ out-of-sample forecasts) that the economic models produce tighter forecast intervals than the random walk.[39] For most exchange rates and model comparisons, economic models are more likely to have tighter forecast intervals than the random walk at longer horizons. In addition, the percentage reduction in the length of forecast intervals is also quantitatively larger at longer horizons. For instance, the forecast intervals of

38. Results are available upon request.

39. We thank an anonymous referee for suggesting that we document this.

TABLE 3

Results of Taylor Rule Model 2

	Horizon = 1 month				Horizon = 6 months				Horizon = 12 months			
	RW Cov.	Cov.	RW Leng.	$\frac{\text{Leng.}}{\text{RW Leng.}}$	RW Cov.	Cov.	RW Leng.	$\frac{\text{Leng.}}{\text{RW Leng.}}$	RW Cov.	Cov.	RW Leng.	$\frac{\text{Leng.}}{\text{RW Leng.}}$
Australia	96.24%	90.61%	9.147	90.99%*	97.12%	90.38%	24.939	92.91%	94.06%	87.13%	39.372	86.41%*
Canada	88.83%*	81.38%	4.178	87.88%	90.16%*	81.42%	12.309	82.69%	88.14%*	79.66%	18.732	70.33%
Denmark	97.51%	92.04%	10.521	90.69%*	98.47%	94.90%*	29.784	96.74%	97.89%	93.68%*	46.495	95.17%
Germany	97.18%	95.77%*	10.993	97.82%	97.60%	96.15%*	31.610	97.59%	98.51%	97.03%*	48.683	96.86%
Japan	96.24%	92.02%	11.680	93.40%*	96.63%	89.90%	33.431	95.17%	96.04%*	88.12%	53.479	89.68%
New Zealand	96.15%*	89.74%	10.899	88.32%	93.38%*	81.46%	30.962	90.08%	91.03%	91.03%	49.853	85.68%*
Norway	95.31%	92.96%	9.436	93.74%*	93.75%	93.75%	28.843*	100.51%	94.55%	92.57%	38.283	99.18%
Sweden	94.37%*	87.32%	9.289	93.54%	94.23%	90.87%	29.447*	105.05%	91.09%	88.12%	47.137	95.56%
Switzerland	98.90%	90.06%	11.688	89.67%*	98.86%	90.91%	34.111	85.53%*	99.41%	88.82%	50.371	91.17%*
UK	96.46%	94.44%*	10.432	97.35%	94.82%	93.78%	32.689	94.19%	98.40%	96.26%*	43.290	95.52%

	Model better	RW better	Model better	RW better	Model better	RW better
Coverage†	2	3	2	2	3	2
Length‡	5	0	1	2	3	0
Overall§	7	3	3	4	6	2

(*Continued*)

TABLE 3

CONTINUED

	Probability (out of the $N(h)$ out-of-sample forecasts) that the model forecast interval length < RW forecast interval length			
	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Australia	79.34%	70.62%	62.98%	81.68%
Canada	95.74%	98.92%	97.81%	100.00%
Denmark	75.62%	47.74%	36.22%	51.58%
Germany	82.63%	81.52%	68.75%	71.29%
Japan	71.36%	34.60%	36.54%	61.39%
New Zealand	80.77%	87.01%	71.52%	81.38%
Norway	91.08%	76.78%	11.54%	50.99%
Sweden	69.95%	57.82%	12.98%	50.50%
Switzerland	69.61%	45.81%	87.50%	63.53%
UK	61.62%	63.78%	79.79%	62.57%

NOTE:
– For each horizon, the first column (RW Cov.) reports empirical coverages of the random walk intervals with a nominal coverage of 95%. The second column (Cov.) reports empirical coverages of the economic model. The third column (RW Leng.) reports the lengths of random walk intervals. The fourth column (Leng./RW Leng.) reports the lengths of the economic model divided by the length of the random walk. Empirical coverages and lengths are averages across $N(h)$ out-of-sample trials.
– Superscripts * in the "RW Cov." (Cov,) columns denote rejections of equal coverage accuracies in favor of the random walk model (the economic model) at a significance level of 10%.
– Superscripts * in the "RW Leng." (Leng./RW Leng.) columns denote rejections of equal lengths in favor of the random walk model (the economic model) at a significance level of 10%. Note that this test is conducted only for currencies that fail to reject the null of equal coverage accuracy.
†– A better model is the one with more accurate coverages.
‡– A better model is the one with tighter forecast intervals for currencies that fail to reject the null of equal coverage accuracy.
§– Sum of the "Coverage" and "Length" rows.
– h denotes forecast horizons for monthly data.

TABLE 4

Results of Taylor Rule Model 3

	Horizon = 1 month				Horizon = 6 months				Horizon = 12 months			
	RW Cov.	Cov.	RW Leng.	$\frac{\text{Leng.}}{\text{RW Leng.}}$	RW Cov.	Cov.	RW Leng.	$\frac{\text{Leng.}}{\text{RW Leng.}}$	RW Cov.	Cov.	RW Leng.	$\frac{\text{Leng.}}{\text{RW Leng.}}$
Australia	96.24%	94.37%*	9.147	96.85%	97.12%	93.27%*	24.939	100.40%	94.06%	90.59%	39.372	93.46%
Canada	88.83%*	84.04%	4.178	89.97%	90.16%*	82.51%	12.309	80.91%	88.14%*	80.23%	18.732	71.18%
Denmark	97.51%	94.53%*	10.521	94.10%	98.47%	94.90%*	29.784	98.02%	97.89%	95.79%*	46.495	94.04%
Germany	97.18%	96.71%	10.993	98.41%*	97.60%	96.63%*	31.610*	99.74%	98.51%	95.05%*	48.683	97.95%
Japan	96.24%	93.43%	11.680	95.37%*	96.63%	93.75%	33.431	90.39%*	96.04%	96.04%	53.479	86.20%*
New Zealand	96.15%	92.31%	10.899	83.45%*	93.38%*	80.79%	30.962	86.24%	91.03%*	82.07%	49.853	78.01%
Norway	95.31%	93.43%	9.436	94.01%*	93.75%	93.27%	28.843*	100.51%	94.55%	91.58%	38.283*	100.60%
Sweden	94.37%	93.43%	9.289	94.90%*	94.23%	93.27%	29.447	96.67%	91.09%	86.63%	47.137	88.85%*
Switzerland	98.90%	93.37%*	11.688	92.50%	98.86%	90.91%	34.111	84.64%*	99.41%	86.47%	50.371	88.96%*
UK	96.46%	94.44%*	10.432	95.83%	94.82%	92.75%	32.689	93.53%*	98.40%	98.93%	43.290	100.05%

	Model better	RW better	Model better	RW better	Model better	RW better
Coverage†	4	1	3	2	2	2
Length‡	5	0	3	2	3	1
Overall§	9	1	6	4	5	3

(Continued)

TABLE 4

CONTINUED

	Probability (out of the $N(h)$ out-of-sample forecasts) that the model forecast interval length < RW forecast interval length			
	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Australia	56.81%	57.82%	48.56%	60.40%
Canada	84.57%	100.00%	98.91%	100.00%
Denmark	82.59%	49.25%	31.12%	69.47%
Germany	76.06%	70.14%	55.77%	62.38%
Japan	83.57%	80.09%	97.60%	97.52%
New Zealand	94.23%	90.91%	69.54%	100.00%
Norway	93.90%	80.09%	32.69%	46.53%
Sweden	78.40%	65.88%	47.60%	80.20%
Switzerland	57.46%	59.78%	78.98%	74.12%
UK	74.75%	60.71%	81.35%	45.45%

NOTE:
– For each horizon, the first column (RW Cov.) reports empirical coverages of the random walk intervals with a nominal coverage of 95%. The second column (Cov.) reports empirical coverages of the economic model. The third column (RW Leng.) reports the lengths of random walk intervals. The fourth column (Leng./RW Leng.) reports the lengths of the economic model divided by the length of the random walk. Empirical coverages and lengths are averages across $N(h)$ out-of-sample trials.
– Superscripts * in the "RW Cov." (Cov,) columns denote rejections of equal coverage accuracies in favor of the random walk model (the economic model) at a significance level of 10%.
– Superscripts * in the RW Leng. ("Leng./RW Leng.") columns denote rejections of equal lengths in favor of the random walk model (the economic model) at a significance level of 10%. Note that this test is conducted only for currencies that fail to reject the null of equal coverage accuracy.
† – A better model is the one with more accurate coverages.
‡ – A better model is the one with tighter forecast intervals for currencies that fail to reject the null of equal coverage accuracy.
§ – Sum of the "Coverage" and "Length" rows.
– h denotes forecast horizons for monthly data.

TABLE 5

Results of Taylor Rule Model 4

	Horizon = 1 month				Horizon = 6 months				Horizon = 12 months			
	RW Cov.	Cov.	RW Leng.	$\frac{\text{Leng.}}{\text{RW Leng.}}$	RW Cov.	Cov.	RW Leng.	$\frac{\text{Leng.}}{\text{RW Leng.}}$	RW Cov.	Cov.	RW Leng.	$\frac{\text{Leng.}}{\text{RW Leng.}}$
Australia	96.24%	95.31%*	9.147	99.00%	97.12%	94.71%*	24.939	95.74%	94.06%	92.08%	39.372	88.12%
Canada	88.83%	89.36%*	4.178	97.63%	90.16%	89.07%	12.309	88.77%	88.14%	85.31%	18.732	79.10%*
Denmark	97.51%	97.01%*	10.521	96.84%	98.47%	98.47%	29.784	94.22%*	97.89%	96.84%*	46.495	87.97%
Germany	97.18%	96.71%*	10.993	96.48%	97.60%	96.15%*	31.610	89.35%	98.51%	94.55%*	48.683	84.59%
Japan	96.24%	94.37%*	11.680	97.51%	96.63%	91.35%	33.431	83.78%*	96.04%	96.04%	53.479	69.82%*
New Zealand	96.15%*	92.31%	10.899	90.42%	93.38%*	84.77%	30.962	80.05%	91.03%	86.90%	49.853	69.91%*
Norway	95.31%*	93.43%	9.436	93.56%	93.75%	91.83%	28.843	92.45%	94.55%	90.59%	38.283	90.63%
Sweden	94.37%	93.90%	9.289	97.36%*	94.23%	95.19%	29.447	98.87%	91.09%	89.60%	47.137	86.07%*
Switzerland	98.90%	97.79%*	11.688	99.39%	98.86%	98.30%	34.111	89.27%	99.41%	100.00%	50.371	78.40%*
UK	96.46%	94.44%*	10.432	97.20%	94.82%*	88.08%	32.689	83.40%	98.40%	87.17%	43.290	80.27%*

	Model better	RW better	Model better	RW better	Model better	RW better
Coverage†	7	2	3	2	2	0
Length‡	1	0	2	0	6	0
Overall§	8	2	5	2	8	0

(Continued)

TABLE 5

CONTINUED

	Probability (out of the $N(h)$ out-of-sample forecasts) that the model forecast interval length < RW forecast interval length			
	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Australia	56.34%	77.73%	63.46%	98.51%
Canada	71.81%	100.00%	100.00%	100.00%
Denmark	66.67%	53.27%	94.90%	98.42%
Germany	94.84%	92.42%	100.00%	100.00%
Japan	59.15%	99.05%	100.00%	100.00%
New Zealand	90.38%	85.71%	100.00%	100.00%
Norway	100.00%	83.41%	100.00%	84.65%
Sweden	91.55%	71.09%	59.13%	100.00%
Switzerland	69.06%	70.39%	100.00%	100.00%
UK	85.35%	99.49%	100.00%	100.00%

NOTE:
– For each horizon, the first column (RW Cov.) reports empirical coverages of the random walk intervals with a nominal coverage of 95%. The second column (Cov.) reports empirical coverages of the economic model. The third column (RW Leng.) reports the lengths of random walk intervals. The fourth column (Leng./RW Leng.) reports the lengths of the economic model divided by the length of the random walk. Empirical coverages and lengths are averages across $N(h)$ out-of-sample trials.
– Superscripts * in the "RW Cov." (Cov,) columns denote rejections of equal coverage accuracies in favor of the random walk model (the economic model) at a significance level of 10%.
– Superscripts * in the "RW Leng." (Leng./RW Leng.) columns denote rejections of equal lengths in favor of the random walk model (the economic model) at a significance level of 10%. Note that this test is conducted only for currencies that fail to reject the null of equal coverage accuracy.
† – A better model is the one with more accurate coverages.
‡ – A better model is the one with tighter forecast intervals for currencies that fail to reject the null of equal coverage accuracy.
§ – Sum of the "Coverage" and "Length" rows.
– h denotes forecast horizons for monthly data.

TABLE 6

Results of Purchasing Power Parity Model

	Horizon = 1 month				Horizon = 6 months				Horizon = 12 months			
	RW Cov.	Cov.	RW Leng.	Leng./RW Leng.	RW Cov.	Cov.	RW Leng.	Leng./RW Leng.	RW Cov.	Cov.	RW Leng.	Leng./RW Leng.
Australia	96.24%	96.24%	9.147*	100.99%	97.12%	96.63%	24.939	97.62%	94.06%	91.09%	39.372	89.84%*
Canada	88.83%	88.83%	4.178	99.11%*	90.16%	90.16%	12.309	95.02%*	88.14%	87.57%	18.732	90.88%*
Denmark	97.51%*	98.51%	10.521	99.79%	98.47%*	99.49%	29.784	97.11%	97.89%	98.95%	46.495	92.34%*
Germany	97.18%	97.18%	10.993	97.57%*	97.60%*	99.04%	31.610	89.56%	98.51%	99.50%	48.683	85.48%*
Japan	96.24%	95.31%*	11.680	99.46%	96.63%	94.23%*	33.431	93.77%	96.04%	96.53%	53.479	83.72%*
New Zealand	96.15%	96.15%	10.899	99.56%	93.38%	94.04%	30.962	96.13%	91.03%	93.10%	49.853	96.18%
Norway	95.31%	95.31%	9.436	97.51%*	93.75%	93.75%	28.843	94.34%*	94.55%	93.56%	38.283	96.01%
Sweden	94.37%*	92.96%	9.289	97.87%	94.23%	95.67%	29.447	99.81%	91.09%	95.54%*	47.137	88.34%
Switzerland	98.90%	97.24%*	11.688	97.69%	98.86%	97.73%*	34.111	86.82%	99.41%	97.06%*	50.371	82.03%
UK	96.46%	96.97%	10.432	99.66%	94.82%	93.78%	32.689	89.36%*	98.40%*	87.17%	43.290	85.11%

	Model better	RW better	Model better	RW better	Model better	RW better
Coverage†	2	2	2	2	2	1
Length‡	3	1	3	0	5	0
Overall§	5	3	5	2	7	1

(Continued)

TABLE 6

CONTINUED

	Probability (out of the $N(h)$ out-of-sample forecasts) that the model forecast interval length < RW forecast interval length			
	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Australia	31.92%	61.14%	75.48%	99.50%
Canada	75.53%	91.40%	92.90%	100.00%
Denmark	55.72%	51.76%	66.84%	98.42%
Germany	91.08%	81.04%	100.00%	100.00%
Japan	38.50%	97.63%	100.00%	100.00%
New Zealand	52.56%	76.62%	100.00%	48.28%
Norway	95.31%	87.68%	96.63%	84.65%
Sweden	90.14%	91.94%	51.44%	100.00%
Switzerland	65.19%	59.22%	97.73%	100.00%
UK	51.01%	60.20%	95.34%	100.00%

NOTE:

– For each horizon, the first column (RW Cov.) reports empirical coverages of the random walk intervals with a nominal coverage of 95%. The second column (Cov.) reports empirical coverages of the economic model. The third column (RW Leng.) reports the lengths of random walk intervals. The fourth column (Leng./RW Leng.) reports the lengths of the economic model divided by the length of the random walk. Empirical coverages and lengths are averages across $N(h)$ out-of-sample trials.

– Superscripts * in the "RW Cov." (Cov,) columns denote rejections of equal coverage accuracies in favor of the random walk model (the economic model) at a significance level of 10%.

– Superscripts * in the "RW Leng." (Leng./RW Leng.) columns denote rejections of equal lengths in favor of the random walk model (the economic model) at a significance level of 10%. Note that this test is conducted only for currencies that fail to reject the null of equal coverage accuracy.

† – A better model is the one with more accurate coverages.

‡ – A better model is the one with tighter forecast intervals for currencies that fail to reject the null of equal coverage accuracy.

§ – Sum of the "Coverage" and "Length" rows.

– h denotes forecast horizons for monthly data.

TABLE 7

RESULTS OF MONETARY MODEL

	Horizon = 1 month				Horizon = 6 months				Horizon = 12 months			
	RW Cov.	Cov.	RW Leng.	$\frac{\text{Leng.}}{\text{RW Leng.}}$	RW Cov.	Cov.	RW Leng.	$\frac{\text{Leng.}}{\text{RW Leng.}}$	RW Cov.	Cov.	RW Leng.	$\frac{\text{Leng.}}{\text{RW Leng.}}$
Australia	96.24%	94.84%*	9.147	100.39%	97.12%*	88.94%	24.939	98.00%	94.06%*	84.65%	39.372	92.54%
Canada	88.83%*	88.30%	4.178	100.13%	90.16%*	84.70%	12.309	98.06%	88.14%	87.57%	18.732	96.13%
Denmark	97.51%	97.01%*	10.521	100.43%	98.47%	93.37%*	29.784	98.85%	97.89%*	87.37%	46.495	93.88%
Germany	97.18%	97.18%	10.993	97.80%*	97.60%*	90.38%	31.610	91.00%	98.51%*	84.16%	48.683	85.39%
Japan	96.24%	96.24%	11.680	101.11%	96.63%	92.79%	33.431	96.28%*	96.04%*	89.11%	53.479	83.11%
New Zealand	96.15%	95.51%	10.899	97.65%*	93.38%*	87.42%	30.962	95.06%	91.03%*	82.07%	49.853	92.10%
Norway	95.31%	94.84%	9.436	95.48%*	93.75%*	90.87%	28.843	95.06%	94.55%*	86.63%	38.283	94.73%
Sweden	94.37%*	91.55%	9.289	96.41%	94.23%*	88.94%	29.447	92.12%	91.09%*	84.65%	47.137	86.79%
Switzerland	98.90%	98.34%*	11.688	98.45%	98.86%*	84.09%	34.111	90.22%	99.41%*	77.65%	50.371	79.19%
UK	96.46%	96.46%	10.432	98.14%*	94.82%	94.30%	32.689	94.24%*	98.40%	90.91%	43.290	89.92%*

	Model better	RW better	Model better	RW better	Model better	RW better
Coverage†	3	2	1	7	0	8
Length‡	4	0	2	0	1	0
Overall§	7	2	3	7	1	8

(*Continued*)

TABLE 7

CONTINUED

	Probability (out of the $N(h)$ out-of-sample forecasts) that the model forecast interval length < RW forecast interval length			
	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Australia	37.56%	63.03%	74.52%	98.02%
Canada	42.02%	76.34%	78.69%	84.75%
Denmark	38.31%	66.83%	71.43%	91.05%
Germany	84.04%	68.25%	91.35%	98.02%
Japan	38.03%	69.19%	94.23%	99.50%
New Zealand	79.49%	73.38%	80.79%	78.62%
Norway	91.08%	91.47%	85.58%	65.35%
Sweden	89.67%	84.83%	77.40%	100.00%
Switzerland	64.09%	64.80%	92.61%	99.41%
UK	79.29%	90.31%	97.41%	94.12%

NOTE:
– For each horizon, the first column (RW Cov.) reports empirical coverages of the random walk intervals with a nominal coverage of 95%. The second column (Cov.) reports empirical coverages of the economic model. The third column (RW Leng.) reports the lengths of random walk intervals. The fourth column (Leng./RW Leng.) reports the lengths of the economic model divided by the length of the random walk. Empirical coverages and lengths are averages across $N(h)$ out-of-sample trials.
– Superscripts * in the "RW Cov." (Cov,) columns denote rejections of equal coverage accuracies in favor of the random walk model (the economic model) at a significance level of 10% or below.
– Superscripts * in the "RW Leng." (Leng./RW Leng.) columns denote rejections of equal lengths in favor of the random walk model (the economic model) at a significance level of 10% or below. Note that this test is conducted only for currencies that fail to reject the null of equal coverage accuracy.
†– A better model is the one with more accurate coverages.
‡ – A better model is the one with tighter forecast intervals for currencies that fail to reject the null of equal coverage accuracy.
§ – Sum of the "Coverage" and "Length" rows.
– h denotes forecast horizons for monthly data.

TABLE 8

RESULTS OF FORWARD PREMIUM MODEL

	Horizon = 1 month				Horizon = 6 months				Horizon = 12 months			
	RW Cov.	Cov.	RW Leng.	$\frac{\text{Leng.}}{\text{RW Leng.}}$	RW Cov.	Cov.	RW Leng.	$\frac{\text{Leng.}}{\text{RW Leng.}}$	RW Cov.	Cov.	RW Leng.	$\frac{\text{Leng.}}{\text{RW Leng.}}$
Australia	96.24%*	84.04%	9.147	88.14%	97.12%*	82.21%	24.939	91.93%	94.06%*	76.24%	39.372	86.60%
Canada	88.83%*	74.47%	4.178	88.30%	90.16%*	74.32%	12.309	84.18%	88.14%*	70.06%	18.732	72.29%
Denmark	97.51%	90.55%	10.521	90.84%*	98.47%	91.84%*	29.784	85.87%	97.89%	90.00%	46.495	85.57%*
Germany	97.18%*	85.92%	10.993	85.73%	97.60%*	82.69%	31.610	90.98%	98.51%	86.63%	48.683	91.59%*
Japan	96.24%*	90.14%	11.680	93.65%	96.63%*	87.98%	33.431	83.92%	96.04%	89.11%	53.479	76.44%*
New Zealand	96.15%*	85.26%	10.899	76.33%	93.38%*	74.17%	30.962	85.11%	91.03%*	71.03%	49.853	80.64%
Norway	95.31%*	83.57%	9.436	82.60%	93.75%*	75.00%	28.843	76.62%	94.55%*	73.76%	38.283	80.00%
Sweden	94.37%*	87.32%	9.289	86.56%	94.23%*	85.10%	29.447	79.39%	91.09%*	70.30%	47.137	68.35%
Switzerland	98.90%	90.61%	11.688	87.16%*	98.86%	93.75%*	34.111	88.55%	99.41%	91.76%*	50.371	89.42%
UK	96.46%	87.88%	10.432	82.73%*	94.82%	88.08%	32.689	74.72%*	98.40%	90.91%	43.290	78.59%*

	Horizon = 1 month: Model better	Horizon = 1 month: RW better	Horizon = 6 months: Model better	Horizon = 6 months: RW better	Horizon = 12 months: Model better	Horizon = 12 months: RW better
Coverage†	0	7	2	7	1	5
Length‡	3	0	1	0	4	0
Overall§	3	7	3	7	5	5

(*Continued*)

TABLE 8

CONTINUED

	Probability (out of the $N(h)$ out-of-sample forecasts) that the model forecast interval length < RW forecast interval length			
	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Australia	69.01%	59.72%	65.87%	61.39%
Canada	63.83%	64.52%	69.40%	70.62%
Denmark	79.60%	86.93%	71.94%	92.11%
Germany	84.98%	75.83%	71.15%	42.08%
Japan	66.20%	67.30%	96.63%	99.01%
New Zealand	94.87%	98.05%	66.89%	100.00%
Norway	84.51%	84.83%	64.42%	61.88%
Sweden	84.98%	90.05%	98.08%	99.50%
Switzerland	74.03%	93.85%	76.14%	64.71%
UK	95.45%	94.39%	90.67%	89.84%

NOTE:
– For each horizon, the first column (RW Cov.) reports empirical coverages of the random walk intervals with a nominal coverage of 95%. The second column (Cov.) reports empirical coverages of the economic model. The third column (RW Leng.) reports the lengths of random walk intervals. The fourth column (Leng./RW Leng.) reports the lengths of the economic model divided by the length of the random walk. Empirical coverages and lengths are averages across $N(h)$ out-of-sample trials.
– Superscripts * in the "RW Cov." (Cov,) columns denote rejections of equal coverage accuracies in favor of the random walk model (the economic model) at a significance level of 10%.
– Superscripts * in the "RW Leng." (Leng./RW Leng.) columns denote rejections of equal lengths in favor of the random walk model (the economic model) at a significance level of 10%. Note that this test is conducted only for currencies that fail to reject the null of equal coverage accuracy.
† – A better model is the one with more accurate coverages.
‡ – A better model is the one with tighter forecast intervals for currencies that fail to reject the null of equal coverage accuracy.
§ – Sum of the "Coverage" and "Length" rows.
– h denotes forecast horizons for monthly data.

economic models are usually more than 10% tighter than those of the random walk at 12 months, while the difference is typically 5% or less at 1 month. Figures 1 and 2 show the lengths of intervals generated by the benchmark Taylor rule model and the random walk for the Australian dollar and the Canadian dollar at 1 and 12 months, respectively.[40] It is clear that the improvement in the lengths is more pronounced at 12 months than at 1 month.

In addition to assessing the performance of models individually, it is also interesting to compare the performance of the random walk with several exchange rate models together. To that end, Corte, Sarno, and Tsiakas (2009) use Bayesian methods to estimate and rank a set of empirical exchange rate models. Based on the rankings, they construct combined forecasts of exchange rates and find that strategies based on combined forecasts yield larger economic gains over the random walk model. In a similar spirit, we also test jointly whether random walk forecast intervals have the shortest lengths relative to those of all other models.[41] Our method follows the idea of Reality Check in White (2000). By comparing all competing models simultaneously, we take into account correlations across models.

The joint null and alternative hypotheses are:

$$H_0 : max_{m=1,\ldots,7} \Delta LL^{\alpha}_{m,h} \leq 0 \tag{22}$$

$$H_A : max_{m=1,\ldots,7} \Delta LL^{\alpha}_{m,h} > 0. \tag{23}$$

If we reject (22), it is concluded that there is at least one model that generates better forecast intervals than the random walk. Unlike our two-sided coverage accuracy and length tests, the joint test we consider here is a one-sided test. Note that there is a fundamental difference between our joint test and the Bayesian model averaging method used in Corte, Sarno, and Tsiakas (2009). Their method is to find whether a combination of models can outperform the random walk, while the idea of the "reality check" in this paper is to conduct a multiple hypothesis test. The test statistic for the joint null in equation (22) is defined as:

$$Jt^{\alpha}_{h} \equiv max_{m=1,\ldots,7} \sqrt{N(h)} \Delta \widehat{LL}^{\alpha}_{m,h}. \tag{24}$$

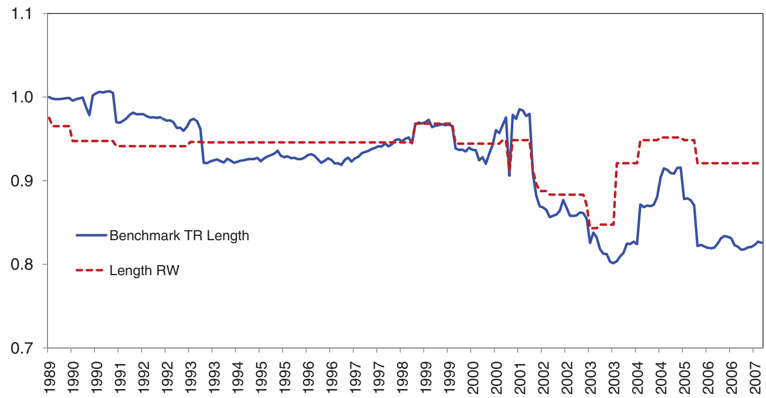
The critical values are obtained by calculating bootstrapped versions of equation (24) over 2,000 bootstrapped samples of $\{\mathbf{W}_t\}_{t=1}^{T}$ using the procedure described in Section 2.

Table 9 reports the results for the joint length test. In 38 of the 40 tests,[42] the null hypothesis that the random walk has the tightest forecast intervals among all models is rejected at a significance level of 10%. The only exceptions are the Australian

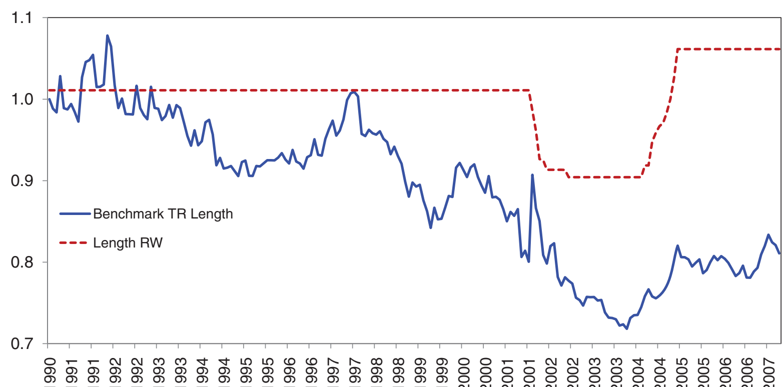
40. Figures in other countries show similar patterns. Results are available upon request.

41. We thank the editor and an anonymous referee for inspiring this test.

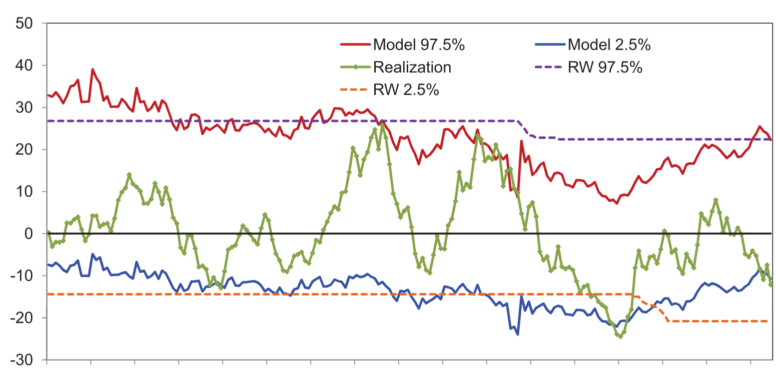
42. Permutations of 10 countries and four horizons.



(a) Lengths of 1-month-ahead forecast intervals

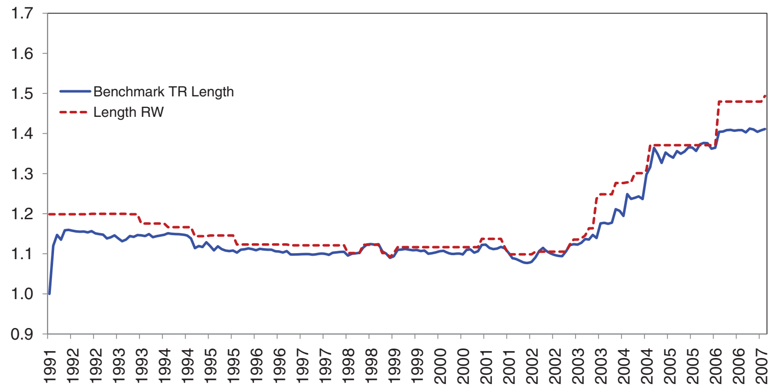


(b) Lengths of 12-month-ahead forecast intervals

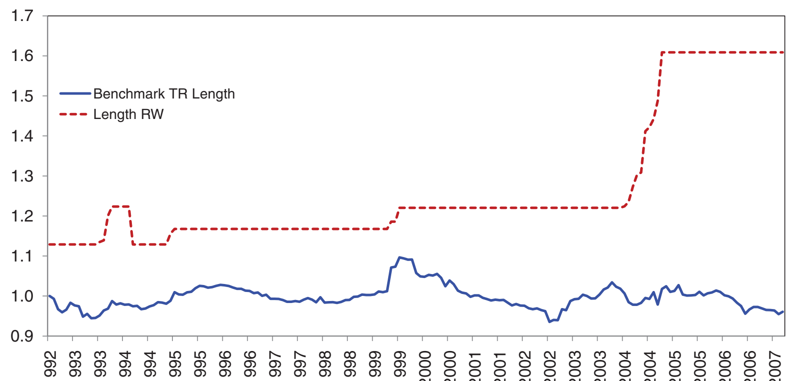


(c) Forecast interval sandrealizations of exchange rate changes (12-month-ahead)

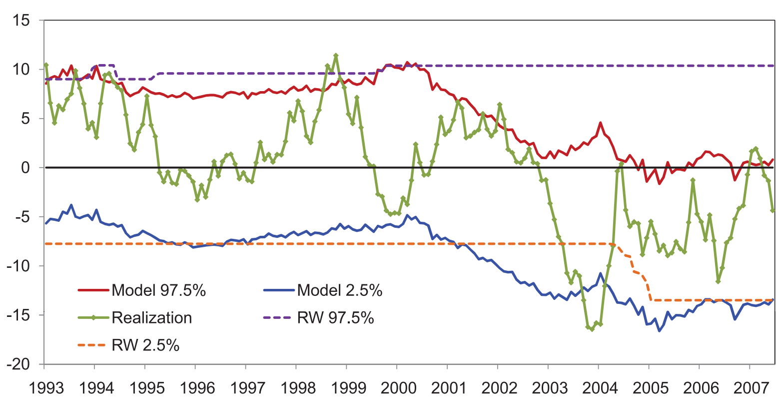
Fig. 1. Lengths of Forecast Intervals for Benchmark Taylor-rule and Random Walk Models: Australia. Notes: In charts (a) and (b), the lengths of forecast intervals are normalized by the first observation of the benchmark Taylor rule model.



(a) Lengths of 1-month-ahead forecast intervals



(b) Lengths of 6-month-ahead forecast intervals



(c) Forecast intervals and realizations of exchange rate changes (12-month-ahead)

Fig. 2. Lengths of Forecast Intervals for Benchmark Taylor Rule and Random Walk Models: Canada. Notes: In charts (a) and (b), the lengths of forecast intervals are normalized by the first observation of the benchmark Taylor rule model.

TABLE 9

RESULTS OF MIN-LENGTH TESTS

	Horizon = 1	Horizon = 3	Horizon = 6	Horizon = 12
Australia	3.85%	1.10%	42.35%	7.50%
Canada	23.90%	4.30%	9.25%	0.25%
Denmark	0.20%	0.05%	0.15%	0.90%
Germany	0.05%	0.05%	1.60%	0.90%
Japan	3.45%	0.40%	0.05%	0.05%
New Zealand	0.05%	0.15%	0.20%	0.05%
Norway	0.05%	0.05%	0.05%	0.05%
Sweden	2.25%	0.30%	0.50%	0.05%
Switzerland	0.15%	0.05%	0.10%	0.05%
UK	0.05%	0.05%	0.05%	0.15%

NOTE: Entries are p-values of one-tailed test for H_0: $\max_{m=1,\ldots,7} \Delta LL^{\alpha}_{m,h} \leq 0$ for $m = 1, 2, \ldots, 7$. Small p-values indicate the rejection of the null hypothesis.

dollar at 6 months and the Canadian dollar at 1 month. For all other currencies and horizons, at least one exchange rate model generates tighter forecast intervals than the random walk.

4. CONCLUSION

This paper contributes to the literature on the Meese–Rogoff puzzle by showing that economic fundamentals are useful in interval forecasting of exchange rates. We apply semiparametric forecast intervals of Wu (Forthcoming) to a group of models for 10 OECD exchange rates. In general, out-of-sample forecast intervals generated by these models are tighter than those generated by the random walk, given that the intervals cover realized exchange rates equally well. The evidence of exchange rate predictability is more pronounced at longer horizons, a result that echoes previous long-horizon studies such as Mark (1995). A benchmark Taylor rule model is found to perform better than PPP, monetary, and forward premium models. The reductions in the lengths of forecast intervals relative to random walk intervals can be substantial (10% or more) at long horizons. Tighter forecast intervals can be economically useful, as we have illustrated using a VaR example. In addition, the findings of this paper can hopefully shed light on the theoretical modeling of exchange rates in future research.

Several recent studies have approached the Meese–Rogoff puzzle from different angles. West, Edison, and Cho (1993), Abhyankar, Sarno, and Valente (2005), and Della Corte, Sarno, and Tsiakas (2009) study exchange rate predictability using utility-based criteria. Abhyankar, Sarno, and Valente (2005) investigate the utility-based value of exchange rate forecasts to an investor who allocates her wealth to assets denominated in different currencies. They find that optimal portfolio weights generated from the monetary model are substantially different (both qualitatively and quantitatively) from optimal weights generated under a random walk model. Della Corte, Sarno, and Tsiakas (2009) assess the economic value of the in-sample and out-of-sample forecasting power of the exchange rate models in the context of dynamic asset allocation strategies. They find that the predictive ability of forward exchange rate premiums has substantial economic value in a dynamic allocation strategy. Both Abhyankar, Sarno, and Valente (2005) and Della Corte, Sarno, and Tsiakas (2009) use a Bayesian framework which takes into account parameter uncertainties in forecasting models. Our paper differs from these two studies in several ways. For instance, we estimate the conditional distributions of exchange rate changes semi-parametrically, while Abhyankar, Sarno, and Valente (2005) and Della Corte, Sarno, and Tsiakas (2009) focus on Bayesian posterior distributions. Also, neither study considers Taylor rule fundamentals in its forecast, as we did in this paper. Utility-based evaluation of exchange rate forecasts is an important issue. It would be interesting to apply our forecasting methods in this direction.

Some recent studies investigate the possibilities that economic fundamentals cannot forecast the exchange rate, even if the exchange rate is determined by these fundamentals. Engel and West (2005) argue that as the discount factor gets close to one, the exchange rate approximately follows a random walk even though it is determined by economic fundamentals as in a present-value asset pricing model. Nason and Rogers (2008) generalize the Engel-West theorem to a class of open-economy dynamic stochastic general equilibrium (DSGE) models. Some reasonable modifications to the Engel–West assumptions may lead to reconciliations between the Engel–West explanation and empirical findings of exchange rate predictability. For instance, Engel, Wang, and Wu (2009) show that when there exist stationary, but persistent, unobservable fundamentals (e.g., a risk premium), long-horizon predictability prevails in *point forecasting*. In a similar vein, it would be of interest to study conditions under which our findings on *interval forecasting* can be reconciled with the Engel–West theorem.

Issues such as parameter instability (Rossi 2005), nonlinearity (Kilian and Taylor 2003), real-time data (Faust, Rogers, and Wright 2003, Molodtsova, Nikolsko-Rzhevskyy, and Papell 2008a, 2008b), and model selection (Sarno and Valente 2009) can all contribute to the Meese–Rogoff puzzle. In addition, panel data might be helpful in detecting exchange rate predictability, especially at long horizons (Mark and Sul 2001, Groen 2000, 2005, Engel, Mark, and West 2007, Rogoff and Stavrakeva 2008). It may be fruitful to incorporate these issues in interval forecasting. We leave these extensions for future research.

LITERATURE CITED

Abhyankar, Abhay, Lucio Sarno, and Giorgio Valente. (2005) "Exchange Rates and Fundamentals: Evidence on the Economic Value of Predictability." *Journal of International Economics*, 66, 325–48.

Adrian, Tobias, Erkko Etula, and Hyun Song Shin. (2009) "Risk Appetite and Exchange Rates." Federal Reserve Bank of New York Staff Report No. 361.

Alquist, Ron, and Menzie Chinn. (2008) "Conventional and Unconventional Models Approaches to Exchange Rate Modeling and Assessment." *International Journal of Finance and Economy*, 13, 2–13.

Berkowitz, Jeremy, and Lorenzo Giorgianni. (2001) "Long-Horizon Exchange Rate Predictability?" *Review of Economics and Statistics*, 83, 81–91.

Bilson, John. (1978) "Rational Expectations and the Exchange Rate." In *The Economics of Exchange Rates: Selected Studies*, edited by Jacob Frenkel and Harry Johnson, pp. 75–96. Reading, MA: Addison-Wesley.

Boero, Gianna, and Emanuela Marrocu. (2004) "The Performance of SETAR Models: A Regime Conditional Evaluation of Point, Interval and Density Forecasts." *International Journal of Forecasting*, 20, 305–20.

Chen, Yu-chin, and Kwok Ping Tsang. (2009) "What Does the Yield Curve Tell Us about Exchange Rate Predictability?" Manuscript, University of Washington, and Virginia Tech.

Chinn, Menzie. (2008) "Nonlinearities, Business Cycles and Exchange Rates." *Economics Notes*, 37, 219–39.

Chinn, Menzie, and Richard A. Meese. (1995) "Banking on Currency Forecasts: How Predictable Is Change in Money?" *Journal of International Economics*, 38, 161–78.

Christoffersen, Peter F. (1998) "Evaluating Interval Forecasts." *International Economic Review*, 39, 841–62.

Christoffersen, Peter F., and Stefano Mazzotta. (2005) "The Accuracy of Density Forecasts from Foreign Exchange Options." *Journal of Financial Econometrics*, 3, 578–605.

Clarida, Richard, Jordi Gali, and Mark Gertler. (1998) "Monetary Rules in Practice: Some International Evidence." *European Economic Review*, 42, 1033–67.

Clarida, Richard, and Mark Taylor. (1997) "The Term Structure of Forward Exchange Premiums and the Forecastability of Spot Exchange Rates: Correcting the Errors." *Review of Economics and Statistics*, 89, 353–61.

Clarida, Richard, Lucio Sarno, Mark Taylor, and Giorgio Valente. (2003) "The Out-of-Sample Success of Term Structure Models as Exchange Rate Predictors: A Step Beyond." *Journal of International Economics*, 60, 61–83.

Clark, Todd, and Kenneth D. West. (2006) "Using Out-of-Sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis." *Journal of Econometrics*, 135, 155–86.

Clark, Todd, and Kenneth D. West. (2007) "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models." *Journal of Econometrics*, 138, 291–311.

Corradi, Valentina, and Norman Swanson. (2006a) "Predictive Density and Conditional Confidence Interval Accuracy Tests." *Journal of Econometrics*, 135, 187–228.

Corradi, Valentina, and Norman Swanson. (2006b) "Predictive Density Evaluation." In *Handbook of Economic Forecasting*, edited by Elliott, Graham, Clive Granger, and Allan Timmerman. Amsterdam: Elsevier.

Corradi, Valentina, and Norman Swanson. (2007) "Nonparametric Bootstrap Procedures for Predictive Inference Based on Recursive Estimation Schemes." *International Economic Review*, 48, 67–109.

Della Corte, Pasquale, Lucio Sarno, and Ilias Tsiakas. (2009) "An Economic Evaluation of Empirical Exchange Rate Models." *Review of Financial Studies*, 22, 3491–530.

Diebold, Francis X., Todd A. Gunther, and Anthony S. Tay. (1998) "Evaluating Density Forecasts with Applications to Financial Risk Management." *International Economic Review*, 39, 863–83.

Diebold, Francis X., Jinyong Hahn, and Anthony S. Tay. (1999) "Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns of Foreign Exchange." *Review of Economics and Statistics*, 81, 661–73.

Diez de los Rios, Antonio. (2009) “Can Affine Term Structure Models Help Us Predict Exchange Rates?” *Journal of Money, Credit and Banking*, 41, 755–66.

Engel, Charles, Nelson Mark, and Kenneth D. West. (2007) “Exchange Rate Models Are Not as Bad as You Think.” *NBER Macroeconomics Annual*.

Engel, Charles, Jian Wang, and Jason J. Wu. (2009) “Long-Horizon Forecasts of Asset Prices when the Discount Factor is close to Unity.” Globalization and Monetary Policy Institute Working Paper No. 36.

Engel, Charles, and Kenneth D. West. (2005) “Exchange Rates and Fundamentals.” *Journal of Political Economy*, 113, 485–517.

Engel, Charles, and Kenneth D. West. (2006) “Taylor Rules and the Deutschmark-Dollar Real Exchange Rate.” *Journal of Money, Credit and Banking*, 38, 1175–94.

Engel, Charles, and Kenneth D. West. (2010) “Global Interest Rates, Currency Returns, and the Real Value of the Dollar.” *American Economic Review, Papers and Proceedings*, May.

Faust, Jon, John Rogers, and Jonathan Wright. (2003) “Exchange Rate Forecasting: The Errors We’ve Really Made.” *Journal of International Economics*, 60, 35–59.

Frenkel, Jacob. (1976) “A Monetary Approach to the Exchange Rate: Doctrinal Aspects and Empirical Evidence.” *Scandinavian Journal of Economics*, 78, 200–24.

Froot, Kenneth, and Kenneth Rogoff. (1995) “Perspectives on PPP and Long-Run Real Exchange Rates.” In *Handbook of International Economics*, Vol. 3, edited by G. Grossman and K. Rogoff. Amsterdam: Elsevier.

Giacomini, Raffaella, and Halbert White. (2006) “Tests of Conditional Predictive Ability.” *Econometrica*, 74, 1545–78.

Groen, Jan J. (2000) “The Monetary Exchange Rate Model as a Long-Run Phenomenon.” *Journal of International Economics*, 52, 299–319.

Groen, Jan J. (2005) “Exchange Rate Predictability and Monetary Fundamentals in a Small Multi-country Panel.” *Journal of Money, Credit, and Banking*, 37, 495–516.

Hall, Peter, Rodney C. L. Wolff, and Qiwei Yao. (1999) “Methods for Estimating a Conditional Distribution Function.” *Journal of American Statistical Association*, 94, 154–63.

Hansen, Peter R. (2005) “A Test for Superior Predictive Ability.” *Journal of Business and Economic Statistics*, 23, 365–80.

Hong, Yongmiao, Haitao Li, and Feng Zhao. (2007) “Can the Random Walk Be Beaten in Out-of-Sample Density Forecasts: Evidence from Intraday Foreign Exchange Rates.” *Journal of Econometrics*, 141, 736–76.

Hung, Jui-Cheng, Chiu, Chien-Liang, and Ming-Chih Lee. (2006) “Hedging with Zero-Value at Risk Hedge Ratio.” *Applied Financial Economics*, 16, 259–69.

J.P. Morgan. (1996) *RiskMetrics Technical Document*, retrieved from http://www.riskmetrics.com/system/files/private/td4e.pdf.

Kilian, Lutz. (1999) “Exchange Rates and Monetary Fundamentals: What do We Learn from Long-Horizon Regressions?” *Journal of Applied Econometrics*, 14, 491–510.

Kilian, Lutz, and Mark Taylor. (2003) “Why Is It So Difficult to Beat the Random Walk Forecast of Exchange Rate?” *Journal of International Economics*, 60, 85–107.

Kim, Hyeongwoo, and Masao Ogaki. (2009) “Purchasing Power Parity and the Taylor Rule.” Ohio State University Department of Economics Working Paper No. 09-03.

MacDonald, Ronald, and Mark P. Taylor. (1994) "The Monetary Model of the Exchange Rate: Long-Run Relationships, Short-Run Dynamics, and How to Beat a Random Walk." *Journal of International Money and Finance*, 13, 276–90.

Mark, Nelson C. (1995) "Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability." *American Economic Review*, 85, 201–18.

Mark, Nelson C. (2009) "Changing Monetary Policy Rules, Learning, and Real Exchange Rate Dynamics." *Journal of Money, Credit, and Banking*, 41, 1047–70.

Mark, Nelson, and Donggyu Sul. (2001) "Nominal Exchange Rates and Monetary Fundamentals Evidence from a Small Post-Bretton Woods Panel." *Journal of International Economics*, 53, 29–52.

Meese, Richard A., and Kenneth Rogoff. (1983) "Empirical Exchange Rate Models of the Seventies: Do They Fit Out-of-Sample?" *Journal of International Economics*, 14, 3–24.

Molodtsova, Tanya, Alex Nikolsko-Rzhevskyy, and David H. Papell. (2008a) "Taylor Rules with Real-Time Data: A Tale of Two Countries and One Exchange Rate." *Journal of Monetary Economics*, 55, S63–S79.

Molodtsova, Tanya, Alex Nikolsko-Rzhevskyy, and David H. Papell. (2008b) "Taylor Rules and the Euro." *Journal of Money, Credit and Banking*, 43, 535–52.

Molodtsova, Tanya, and David H. Papell. (2009) "Out-of-Sample Exchange Rate Predictability with Taylor Rule Fundamentals." *Journal of International Economics*, 77, 167–80.

Mussa, Michael. (1976) "The Exchange Rate, the Banlance of Payment, and Monetary and Fiscal Policy under a Regime of Controlled Floating." *Scandinavian Journal of Economics*, 78, 229–48.

Nason, James, and John Rogers. (2008) "Exchange Rates and Fundamentals: A Generalization." International Finance Discussion Papers, Number 948, Board of Governors of the Federal Reserve System.

Newey, Whitney K., and Kenneth D. West. (1987) "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, 55, 703–08.

Politis, Dimitris N., and Joseph P. Romano. (1994) "The Stationary Bootstrap." *Journal of the American Statistical Association*, 89, 1303–13.

Rogoff, Kenneth S., and Vania Stavrakeva. (2008) "The Continuing Puzzle of Short Horizon Exchange Rate Forecasting." NBER Working Paper No. 14071.

Romano, Joseph P., and Michael Wolf. (2005) "Stepwise Multiple Testing as Formalized Data Snooping." *Econometrica*, 74, 1237–82.

Rossi, Barbara. (2005) "Testing Long-horizon Predictive Ability with High Persistence, and the Meese–Rogoff Puzzle." *International Economic Review*, 46, 61–92.

Sarno, Lucio, and Giorgio Valente. (2005) "Empirical Exchange Rate Models and Currency Risk: Some Evidence from Density Forecasts." *Journal of International Money and Finance*, 24, 363–85.

Sarno, Lucio, and Giorgio Valente. (2009) "Exchange Rates and Fundamentals: Footloose or Evolving Relationship?" *Journal of the European Economic Association*, 7, 786–830.

West, Kenneth D., Hali J. Edison, and Dongchul Cho. (1993) "A Utility-based Comparison of Some Models of Exchange Rate Volatility." *Journal of International Economics*, 35, 23–45.

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

Wu, Jason J. (Forthcoming) "Semiparametric Forecast Intervals." *Journal of Forecasting*.