Journal of International Money and Finance 95 (2019) 332–362

Contents lists available at ScienceDirect

Journal of International Money and Finance

journal homepage: www.elsevier.com/locate/jimf

Exchange rate prediction redux: New models, new data, new currencies

Yin-Wong Cheung [a,1,*], Menzie D. Chinn [b,2], Antonio Garcia Pascual [c,3], Yi Zhang [d,4]

[a] City University of Hong Kong, Hong Kong
[b] University of Wisconsin, Madison and NBER, United States
[c] Barclays, United Kingdom
[d] Center for Healthcare Policy and Research, University of California, Davis, One Shields Avenue, Davis, CA 95616, United States

ARTICLE INFO

Article history:
Available online 21 March 2018

JEL classification:
F31
F47

Keywords:
Exchange rates
Monetary model
Interest rate parity
Behavioral equilibrium exchange rate model
Forecasting performance

ABSTRACT

Previous assessments of nominal exchange rate determination, following Meese and Rogoff (1983) have focused upon a narrow set of models. Cheung et al. (2005) augmented the usual suspects with productivity based models, and "behavioral equilibrium exchange rate" models, and assessed performance at horizons of up to 5 years. In this paper, we further expand the set of models to include Taylor rule fundamentals, yield curve factors, and incorporate shadow rates and risk and liquidity factors. The performance of these models is compared against the random walk benchmark. The models are estimated in error correction and first-difference specifications. We examine model performance at various forecast horizons (1 quarter, 4 quarters, 20 quarters) using differing metrics (mean squared error, direction of change), as well as the "consistency" test of Cheung and Chinn (1998). The purchasing power parity estimated in the error correction form delivers the best performance at long horizons by a mean squared error measure. Moreover, along a direction-of-change dimension, certain structural models do outperform a random walk with statistical significance. While one finds that these forecasts are cointegrated with the actual values of exchange rates, in most cases, the elasticity of the forecasts with respect to the actual values is different from unity. Overall, model/specification/currency combinations that work well in one period and one performance metric will not necessarily work well in another period and alternative performance metric.

1. Introduction

Nearly fifteen years ago, three of the authors embarked upon an assessment of the then dominant empirical exchange rate models of the time.[5] Over the past decade, the consensus – such as it was – regarding the determinants of exchange rate

* Corresponding author.
E-mail addresses: yicheung@cityu.edu.hk (Y.-W. Cheung), mchinn@lafollette.wisc.edu (M.D. Chinn), Antonio.GarciaPascual@barclays.com (A.G. Pascual), ychzhang@ucdavis.edu (Y. Zhang).
[1] Department of Economics and Finance.
[2] Robert M La Follette School of Public Affairs, and Department of Economics.
[3] Macro Research, London.
[4] Center for Healthcare Policy and Research.
[5] Published as Cheung et al. (2005). The title of that paper was appropriated from the original 1981 Meese and Rogoff *International Finance Discussion Paper* No. 184, subsequently published (1983a, 1983b).

https://doi.org/10.1016/j.jimonfin.2018.03.010
0261-5606/

movements has further disintegrated. The sources of this phenomenon can in part be traced to the realities of the new world economy, and in part to the development of new theories of exchange rate determination. Now seems a good time to re-visit in a comprehensive fashion the question posed in our title.

To motivate this exercise, first consider how different the world was then. The "New Economy" was an established phenomenon, with accelerated productivity growth in the US. Inflation and output growth, across the advanced economies, appeared to have entered a prolonged and durable period of relative stability, a development dubbed "The Great Moderation". If one were to ask a typical international finance authority what the most robust determinant of the dollar-based exchange rate (shown in Figs. 1–3) was, the likely answer would be "real interest differentials". Compare to the present situation of short term policy rates bound at zero (Fig. 4) and possibly unrepresentative of the actual stance of monetary policy (shadow rates in Fig. 5), slowing productivity growth, and repeated bouts of financial risk intolerance and illiquidity (VIX and TED spreads in Fig. 6). Observed real interest differentials at the short horizon are likely to be close to zero, given the zero lower bound, and low inflation worldwide.

It is against this backdrop that several new models have been forwarded in the past decade. Some explanations are motivated by new findings in the empirical literature, such the correlation between net foreign asset positions and real exchange rates. Others, such as those based on central bank reaction functions have now become well established in the literature. Or models that relate the exchange rate to interest rate differentials at several horizons simultaneously. But several of these models have not been subjected to comprehensive examination of the sort that Meese and Rogoff conducted in their original 1983 work. While older models have been ably reviewed (Engel, 2014; Rossi, 2013), we believe that a systematic examination of these newer empirical models is due, for a number of reasons.

First, while some of these models have become prominent in policy and financial circles, they have not been subjected to the sort of rigorous out-of-sample testing conducted in academic studies.

Second, the same criteria are often used, neglecting many alternative dimensions of model forecast performance. That is, the first and second moment metrics such as mean error and mean squared error are considered, while other aspects that might be of greater importance are often neglected. We have in mind the direction of change – perhaps more important from a market timing perspective – and other indicators of forecast attributes.

In this study, we extend the forecast comparison of exchange rate models in several dimensions.

- Eight models are compared against the random walk. Of these, four were examined in our previous study (Cheung et al., 2005). The new models include a real interest differential model incorporating shadow interest rates, Taylor rule fundamentals, a sticky price monetary model augmented with risk proxies, and an interest rate model incorporating yield curve factors. In addition, we implement a different specification for purchasing power parity.
- The behavior of US dollar-based exchange rates of the Canadian dollar, British pound, Japanese yen, Swiss franc, and the euro are examined. The German mark has dropped out, while the last two exchange rates are added.
- The models are estimated in two ways: in first-difference and error correction specifications.
- Forecasting performance is evaluated at several horizons (1-, 4- and 20-quarter horizons) and three sample periods: post-1982, post-dot.com boom and post-Crisis onset. We have thus evaluated out of sample periods, spanning the times that have witnessed notable changes in the global environment.
- We augment the conventional metrics with a direction of change statistic and the "consistency" criterion of Cheung and Chinn (1998).

It is worthwhile to stress that our study is *not* aimed at determining which model best *forecasts*, but rather aimed at determining which model appears to have the greatest empirical content, by which we mean the ability to reliably predict exchange rate movements. Were our objective the former, we would not conduct *ex post* historical simulations where we assume knowledge of the realized values of the right hand side variables.

Consistent with previous studies, we find that no model consistently outperforms a random walk according to the mean squared error criterion at short horizons. Somewhat at variance with some previous findings, we find that the proportion of times the structural models (as a group) outperform a random walk at long horizons is slightly greater than would be expected if the outcomes were merely random, 16%, using a 10% significance level. The aggregate average outperformance rate, however, does not convey the superior performance of individual specifications. The error correction specification of purchasing power parity, in particular, performs especially well at longer horizons of one year and 5 years; it beats the random walk benchmark 80% (21 out of 26 cases) of the time.

The direction-of-change statistics indicate more forcefully that the structural models do outperform a random walk characterization by a statistically significant amount. The structural models (as a group) outperform a random walk 29% of the time. Under this performance criterion, the first-difference specification of BEER model outperforms the random walk benchmark 70% (23 out of 33 cases) of the time, while the error correction specification of purchasing power parity outperforms 54% (21 out of 39 cases) of the time.

In terms of the "consistency" test of Cheung and Chinn (1998), some positive results are obtained. The actual and forecasted rates are cointegrated 60% of the time, and is much more often than would occur by chance for all the models. However, in most of these cases of cointegration, the condition of unitary elasticity of expectations is rejected, so very few instances of consistency are found.

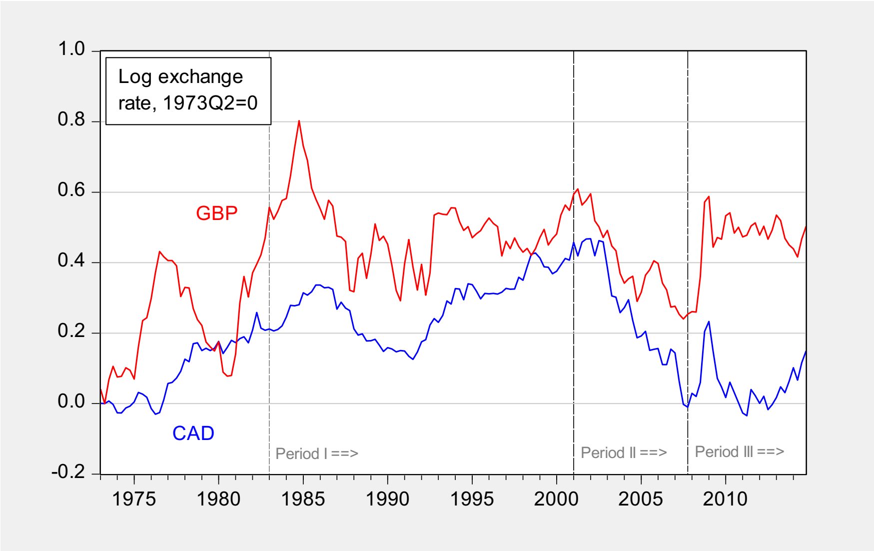

Fig. 1. Exchange rates for Canadian dollar and British pound, end of month.

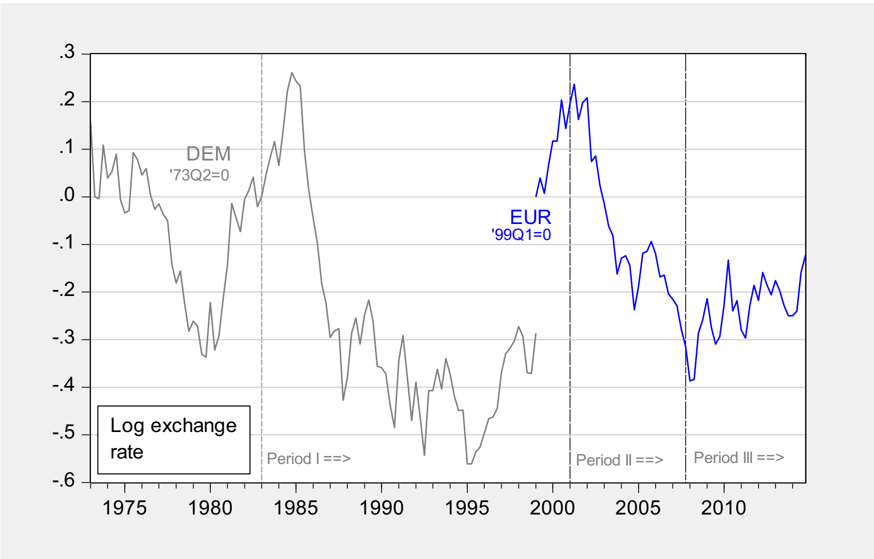

Fig. 2. Exchange rates for Deutsche mark and euro, end of month.

We conclude that the question of exchange rate predictability (still) remains unresolved. In particular, while the oft-used mean squared error and the direction of change criteria provide an encouraging perspective, more so than in our previous study, and the direction of change results are, relatively speaking, even more positive. As in our previous study, the best model and specification tend to be specific to the currency and out-of-sample forecasting period.

2. Theoretical models

The universe of empirical models that have been examined over the floating rate period is enormous, and evidenced in the introduction, ever expanding. Consequently, any evaluation of these models must necessarily be selective. Our criteria require that the models are (1) prominent in the economic and policy literature, (2) readily implementable and replicable, and (3) rarely evaluated in a comparative and systematic fashion. We use the random walk model as our benchmark naive model, in line with previous work. Two "models" are merely parity conditions.

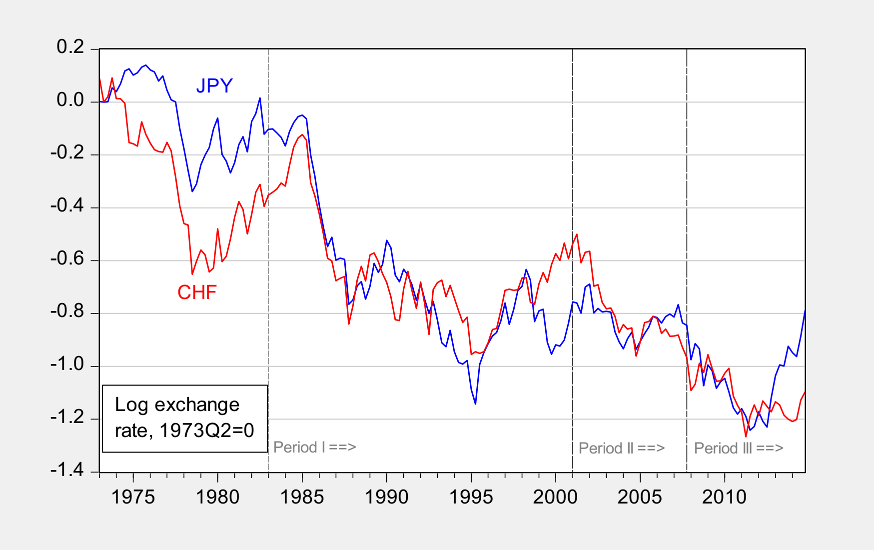

Fig. 3. Exchange rates for Japanese yen and Swiss franc, end of month.

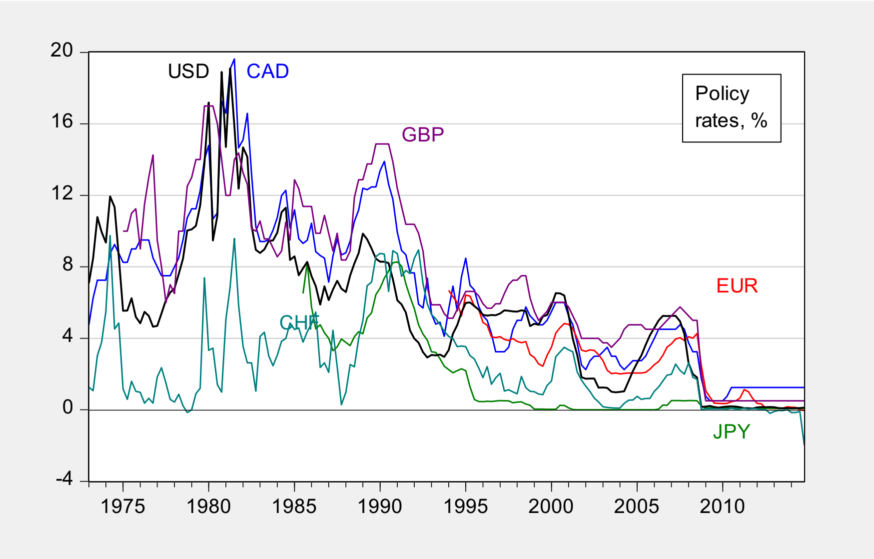

Fig. 4. Overnight interest rates.

2.1. Uncovered interest rate parity

$$s_{t+k} = s_t + \hat{i}_{t,k}, \quad (1)$$

where s is the (log) exchange rate, $i_{t,k}$ is the interest rate of maturity k, ^ denotes the intercountry difference. Unlike the other specifications, this relation involves no estimation in order to generate predictions.[6]

Interest rate parity might seem to be an unlikely candidate for predicting exchange rates, given the extensive literature documenting the failure of interest differentials to predict the right direction of exchange rate changes, let alone the levels. However, Chinn and Meredith (2004) found that long maturity interest rates do tend to correctly predict subsequent long horizon exchange rate changes. This result was verified, although in an attenuated form, in Chinn and Zhang (2015).

[6] Note that we use the exact formulation, rather than the log approximation, to calculate the predictions.

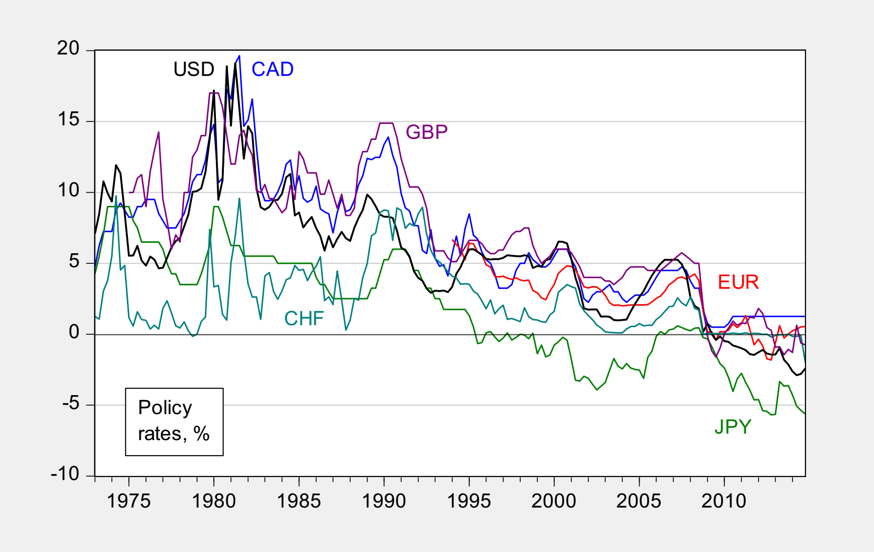

Fig. 5. Overnight interest rates and shadow rates.

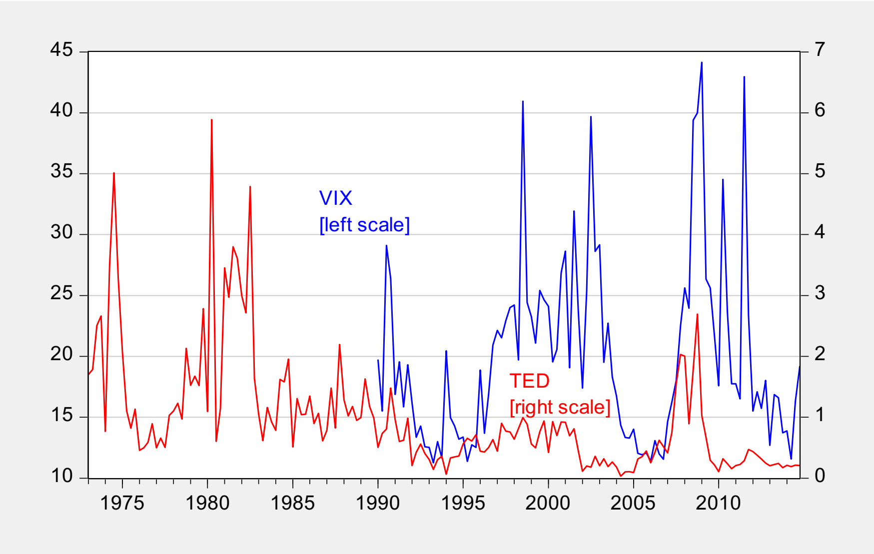

Fig. 6. VIX (left scale) and TED spread (right scale).

2.2. Relative purchasing power parity

$$s_t = \beta_0 + \hat{p}_t, \tag{2}$$

where p is log price level, and ^ denotes the intercountry difference. While the relationship between the exchange rate and the price level is not estimated, the adjustment process in the error correction specification over time is.[7] Recent work (Jordá and Taylor, 2012, Ca' Zorzi et al., 2016, among others) has documented the usefulness of PPP deviations for predicting exchange rate changes.[8]

[7] This contrasts with the procedure in Cheung et al. (2005). In that case the constant of the real exchange rate was iteratively estimated to generate a forecast for k steps ahead. In this paper, we *estimate* the adjustment pace in an error correction specification, or relationship between changes in exchange rate and changes in price differentials.

[8] Although Jordá and Taylor show the reversion is nonlinear in nature.

2.3. Sticky price monetary model

Our first "model" is included as a standard comparator – the workhorse model of Dornbusch (1976) and Frankel (1979). This approach still provides the fundamental intuition for how flexible exchange rates behave. The sticky price monetary model can be expressed as follows:

$$s_t = \beta_0 + \beta_1 \hat{m}_t + \beta_2 \hat{y}_t + \beta_3 \hat{i}_t + \beta_4 \hat{\pi}_t + u_t, \quad (3)$$

where m is log money, y is log real GDP, i and π are the interest and inflation rate, respectively, and u_t is an error term. The characteristics of this model are well known, so we do not devote time to discussing the theory behind the equation.

2.4. Behavioral equilibrium exchange rate (BEER) model

We examine a diverse set of models that incorporate a number of familiar variants. A typical specification is:

$$s_t = \beta_0 + \hat{p}_t + \beta_5 \hat{\omega}_t + \beta_6 \hat{r}_t + \beta_7 \hat{g}debt_t + \beta_8 tot_t + \beta_9 nfa_t + u_t, \quad (4)$$

where p is the log price level (CPI), ω is the relative price of nontradables, r is the real interest rate, *gdebt* the government debt to GDP ratio, *tot* the log terms of trade, and *nfa* is the net foreign asset. This specification can be thought of as incorporating the Balassa-Samuelson effect (by way of the relative price of nontradables), real interest differential model, an exchange risk premium associated with government debt stocks, and additional portfolio balance effects arising from the net foreign asset position of the economy. Clark and MacDonald (1999) is one exposition of this approach.

Models based upon this approach have been commonly employed to determining the rate at which currencies will gravitate to over some intermediate horizon, especially in the context of policy issues. This approach has been often used by market practitioners to assess the extent of currencies deviation from fair value.[9]

Next are four specifications not examined in our previous study.

2.5. Taylor rule fundamentals

One major empirical innovation of the 2000's involved taking endogeneity seriously, in particular the presence of central bank reaction functions. Given the use of Taylor rules by central banks, it is natural to substitute out policy rates with the objects in the Taylor rule – namely output and inflation gaps. This procedure is first implemented by (Molodtsova and Papell, 2009). The resulting specification is:

$$s_{t+k} - s_t = \beta_0 + \beta_1 \hat{\tilde{y}}_t + \beta_2 \hat{\pi}_t + u_t, \quad (5)$$

where $\tilde{y}_t$ is the output gap.[10]

2.6. Real interest differential

The real interest differential was one of the most widely used models of the real exchange rate, prior to the encounter with the zero lower bound in the US, Japan, the euro area and the UK. The innovation here is to use shadow rates for periods in which policy rates are effectively bound at zero.[11] These nominal rates are adjusted by inflation; we use lagged one year inflation as a proxy for expected inflation. Hence:

$$s_t = \beta_0 + \beta_1 (\hat{i}_t^{shadow} - \hat{\pi}_t) + u_t. \quad (6)$$

2.7. Sticky price monetary model augmented by risk and liquidity factors

One of the characteristics of the post-2007 period is the importance of the safe-haven character of the US dollar and liquidity concerns, the latter particularly during the period surrounding the Lehman bankruptcy. In order to account for these factors, we augment the standard monetary model with proxy measures – namely the VIX and the three-month Treasury-Libor (TED) spread.

$$s_t = \beta_0 + \beta_1 \hat{m}_t + \beta_2 \hat{y}_t + \beta_3 \hat{i}_t + \beta_4 \hat{\pi}_t + \beta_5 VIX_t + \beta_6 TED_t + u_t. \quad (7)$$

[9] We do not examine a closely related approach: macroeconomic balances approach of the IMF (see Faruqee, Isard and Masson, 1999). This approach, and the succeeding methodology incorporated into the External Balances Approach (EBA), requires extensive judgements regarding the trend level of output, and the impact of demographic variables upon various macroeconomic aggregates. We did not believe it would be possible to subject this methodology to the same out of sample forecasting exercise applied to the others. The NATREX approach is conceptually different from the BEER methodology. However, it shares a sufficiently large number of attributes with the latter that we decided not to separately examine it.

[10] We estimate the output gap using a Hodrick-Prescott filter applied to the full sample, extended by 6 quarters using an ARIMA model.

[11] The shadow rate is used only for those periods when it is calculated; otherwise the overnight money market or policy rate is used.

2.8. Yield curve slope

Recent work by Chen and Tsang (2013) emphasizes the information content in the slope and curvature of the yield curve. We implement a simpler version of their specification, incorporating the intercountry-difference in the level of the three month interest rate, and difference in the slope (10 year minus three month yields).[12]

$$s_{t+k} - s_t = \beta_0 + \beta_1(\hat{i}_t) + \beta_2(slope_t) + u_t. \tag{8}$$

3. Data, estimation and forecasting comparison

3.1. Data

The analysis uses quarterly data for the United States, Canada, UK, Japan, Germany, and Switzerland over the 1973q2 to 2014q4 period. The exchange rate, money, price and income variables are drawn primarily from the IMF's *International Financial Statistics*. The interest rates used to conduct the interest rate parity forecasts are essentially the same as those used in Chinn and Meredith (2004), Chinn and Quayyum (2012)). See the Appendix A for a more detailed description.

Three out-of-sample periods are used to assess model performance: 1983Q1-2014Q4, 2001Q1-2014Q4, and 2007Q4-2014Q4. The first period encompasses the period after the end of monetary targeting in the U.S., the second conforms to the post-dot.com period, while the third spans the period of financial turmoil associated with the end of the US housing boom. We term these Periods I, II, III, respectively.

Figs. 1–3 depict, respectively, the dollar-based exchange rates - an increase implies dollar appreciation - examined in this study. We include the Deutschemark in Fig. 2 to provide context for the evolution of the euro over the 1999–2014 period. The different dashed lines denote the beginnings of Period I, II, and III. In one sense, the longest out-of-sample period (Period I) subjects the models to a more rigorous test, in that the prediction takes place over several large dollar appreciations and subsequent depreciations. In other words, this longer span encompasses more than one "dollar cycle". The use of this long out-of-sample forecasting period has the added advantage that it ensures that there are many forecast observations to conduct inference upon.

In another sense, the shortest sample (Period III) confronts the models with a more challenging test – particularly the older models, as this period is dominated by the global financial crisis, which a priori conventional fundamentals such as money stocks, output and the like are unlikely to fully capture developments, which may be more related to market conditions such as volatility, risk premia and illiquidity.

3.2. Estimation and forecasting

We adopt the convention in the empirical exchange rate modeling literature of implementing "rolling regressions." That is, estimates are applied over a given data sample, out-of-sample forecasts produced, then the sample is moved up, or "rolled" forward one observation before the procedure is repeated. This process continues until all the out-of-sample observations are exhausted.[13]

Two specifications of these theoretical models were estimated: (1) an error correction specification, and (2) a first differences specification. Since implementation of the error correction specification is relatively involved, we will address the first-difference specification to begin with. Consider the general expression for the relationship between the exchange rate and fundamentals:

$$s_t = X_t\Gamma + u_t, \tag{9}$$

where X_t is a vector of fundamental variables under consideration. The first-difference specification involves the following regression:

$$\Delta s_t = \Delta X_t\Gamma + u_t. \tag{10}$$

These estimates are then used to generate one- and multi-quarter ahead forecasts. Since these exchange rate models imply joint determination of all variables in the equations, it makes sense to apply instrumental variables. However, previous experience indicates that the gains in consistency are far outweighed by the loss in efficiency, in terms of prediction (Chinn and Meese, 1995). Hence, we rely solely on OLS.[14]

[12] Eq. (8) can be taken as nesting Eq. (1) for the one quarter horizon. However, this is not true for the other horizons.

[13] The use of rolling estimates makes sense also in order to hold the sample size use for estimation constant, so that, among other benefits, the power of the tests is held constant in the forecast comparison exercise.

[14] Clearly, we have restricted ourselves to linear estimation methodologies, eschewing functional nonlinearities (Meese and Rose, 1991) and regime switching (Engel and Hamilton, 1990). We have also omitted panel regression techniques in conjunction with long run relationships, despite evidence suggests the potential usefulness of such approaches (Mark and Sul, 2001). Finally, we did not undertake systems-based estimation that has been found in certain circumstances to yield superior forecast performance, even at short horizons (e.g., MacDonald and Marsh, 1997).

The error correction estimation involves a two step procedure. In the first step, the long-run cointegrating relation implied by (5) is identified using the Johansen procedure. The estimated cointegrating vector ($\tilde{\Gamma}$) is incorporated into the error correction term, and the resulting equation

$$s_t - s_{t-k} = \delta_0 + \delta_1(s_{t-k} - X_{t-k}\tilde{\Gamma}) + u_t \tag{11}$$

is estimated via OLS. Eq. (7) can be thought of as an error correction model stripped of short run dynamics. A similar approach was used in Mark (1995) and Chinn and Meese (1995), except for the fact that in those two cases, the cointegrating vector was imposed a priori.[15]

One key difference between our implementation of the error correction specification and that undertaken in some other studies involves the treatment of the cointegrating vector. In some other prominent studies, the cointegrating relationship is estimated over the entire sample, and then out of sample forecasting undertaken, where the short run dynamics are treated as time varying but the long-run relationship is not. While there are good reasons for adopting this approach,[16] we allow our estimates of the *long-run* cointegrating relationship vary as the data window moves.

It is also useful to stress the difference between the error correction specification forecasts and the first-difference specification forecasts. In the latter, *ex post* values of the right hand side variables are used to generate the predicted exchange rate change. In the former, contemporaneous values of the right hand side variables are not necessary, and the error correction predictions are true *ex ante* forecasts. Hence, we are affording the first-difference specifications a tremendous informational advantage in forecasting.[17]

3.3. Forecast comparison

To evaluate the forecasting accuracy of the different structural models, the ratio between the mean squared error (MSE) of the structural models and a driftless random walk is used. A value smaller (larger) than one indicates a better performance of the structural model (random walk). We also explicitly test the null hypothesis of no difference in the accuracy of the two competing forecasts (i.e. structural model vs driftless random walk). In particular, we use the Diebold-Mariano-West statistic (Diebold and Mariano, 1995; West, 1996) which is defined as the ratio between the sample mean loss differential and an estimate of its standard error; this ratio is asymptotically distributed as a standard normal. The loss differential is defined as the difference between the squared forecast error of the structural models and that of the random walk. A consistent estimate of the standard deviation can be constructed from a weighted sum of the available sample autocovariances of the loss differential vector.[18] Following Andrews (1991), a quadratic spectral kernel is employed, together with a data-dependent bandwidth selection procedure.[19] See Diebold and Mariano (1995) and Andrews (1991) for a more detailed discussion on the test and quadratic spectral kernel.

We also examine the predictive power of the various models along different dimensions. One might be tempted to conclude that we are merely changing the well-established "rules of the game" by doing so. However, there are very good reasons to use other evaluation criteria. First, there is the intuitively appealing rationale that minimizing the mean squared error (or relatedly mean absolute error) may not be important from an economic standpoint.[20]A less pedestrian motivation is that the typical mean squared error criterion may miss out on important aspects of predictions, especially at long horizons. For instance, Christoffersen and Diebold (1998) point out that accounting for cointegration does not typically lead to forecast improvement when assessed using the standard mean squared error criterion vis a vis univariate predictions.[21]

Hence, our first alternative evaluation metric for the relative forecast performance of the structural models is the direction of change statistic, which it is computed as the number of correct predictions of the direction of change over the total number of predictions. A value above (below) 50 per cent indicates a better (worse) forecasting performance than a naive model that predicts the exchange rate has an equal chance to go up or down. Again, Diebold and Mariano (1995) and

[15] We could have included another specification including short run dynamics, hence encompassing both error correction and first difference specifications. We opted to exclude short-run dynamics in Eq. (11), first for the sake of brevity, and second because the inclusion of short-run dynamics creates additional issues on the generation of the right-hand-side variables and the stability of the short-run dynamics that complicate the forecast comparison exercise beyond a manageable level. Including short run dynamics would also mean that long horizon error correction results would not be distinguishable from integrating forecasts from a standard error correction model (Kilian and Taylor, 2003).

[16] In particular, one might wish to use as much information as possible to obtain estimates of the cointegrating relationships – the asymmetry in estimation approach is troublesome, and makes it difficult to distinguish quasi-ex ante forecasts from true ex ante forecasts.

[17] Note that excluding short run dynamics in the error correction model means that the use of Eq. (11) yields true *ex ante* forecasts and makes our exercise directly comparable with, for example, Mark (1995), Chinn and Meese (1995) and Groen (2000).

[18] Since we adopt the rolling regression setting, the Diebold-Mariano-West test is asymptotically valid for our forecast comparison exercise, under some regularity assumptions (Giacomini and White, 2006). Also, the use of bootstrapped or finite-sample critical values tends to reduce the rejection rate and, hence, reinforces the low rejection results reported in the following sections. The case of using the adjusted MSPE statistic (Clark and West, 2006) is discussed in Section 4.4.

[19] We also experienced with the Bartlett kernel and the deterministic bandwidth selection method. The results from these methods are qualitatively very similar.

[20] For example, Leitch and Tanner (1991) argue that a direction of change criterion may be more relevant for profitability and economic concerns, and hence a more appropriate metric than others based on purely statistical motivations.

[21] See Duy and Thoma (1998) for a contrasting assessment regarding the use of cointegrating relationships.

West (1996) provide a test statistic for the null of no forecasting performance of the structural model. The statistic follows a binomial distribution, and its studentized version is asymptotically distributed as a standard normal.

Finally, we believe that any reasonable criteria would put some weight the tendency for predictions from cointegrated systems to "hang together". The third metric we use to evaluate forecast performance is the consistency criterion proposed in Cheung and Chinn (1998). This metric focuses on the time-series properties of the forecast. The forecast of a given spot exchange rate is labeled as consistent if (1) the two series have the same order of integration, (2) they are cointegrated, and (3) the cointegration vector satisfies the unitary elasticity of expectations condition. Loosely speaking, a forecast is consistent if it moves in tandem with the spot exchange rate in the long run. Cheung and Chinn (1998) provide a more detailed discussion on the consistency criterion and its implementation.

4. Comparing the forecast performance

4.1. The MSE criterion

The comparison of forecasting performance based on MSE ratios is summarized in Table 1. The Table contains MSE ratios and the p-values from five dollar-based currency pairs, eight structural models, the error correction and first-difference specifications, three forecasting horizons, and three forecasting samples. The results for the three forecasting periods are presented under Tables 1a, 1b, and 1c, respectively. Each cell in the Table has two entries. The first one is the MSE ratio (the MSEs of a structural model to the random walk specification). The entry underneath the MSE ratio is the p-value of the hypothesis that the MSEs of the structural and random walk models are the same. Obviously, because the euro only comes into existence in 1999, there are no entries for the two earlier out-of-sample prediction periods. Moreover, because of the lack of data, the behavioral equilibrium exchange rate model is not estimated for the dollar-Swiss franc and dollar-yen exchange rates. Finally, the lack of earlier data for the risk and liquidity proxies means that the augmented sticky-price monetary model predictions are only available for the most recent sample. Altogether, there are 462 MSE ratios, with 195 (about 42%) pertaining to the latest sample. Of these 462 ratios, 285 are computed from the error correction specification and 177 from the first-difference one.

Note that in the tables, only "error correction specification" entries are reported for the interest rate parity model. In fact, this model is not "estimated"; rather the predicted spot rate is calculated using the uncovered interest parity condition. To the extent that long term interest rates can be considered the error correction term, we believe this categorization is most appropriate.

Overall, the MSE results are not particularly favorable to the structural models. Of the 462 MSE ratios, 266 are not significant (at the 10% significance level) and 196 are significant, about 42.4%. That is, for the majority cases one cannot differentiate the forecasting performance between a structural model and a random walk model. There is a higher rate of rejection than would be expected from random results. For the 196 significant cases, however, there are 124 cases in which the random walk model *is significantly* better than the competing structural models and only 72 cases in which the opposite is true. Still, the latter represents a 16.2% rate of statistical outperformance, using the 10% level of marginal significance. This means that we are rejecting the null at a rate higher than what one would expect from random chance. This outcome is much more positive than obtained in Cheung et al. (2005), in which case there were essentially no instances in which the random walk was significantly outperformed (specifically, 2 out of 216 ratios, or less than 1%).

Inspection of the MSE ratios reveals a few obviously consistent patterns, in terms of outperformance. The significant cases are not proportionally distributed across the three forecasting periods. The forecasting sample that starts in 1983 has the smallest number of total cases; of the 111 reported MSE ratios, 65 (approximately 58.6%) are significantly worse or better than random walk. This period has the highest proportion of successes: 25.2%, and the 28 success cases account for 38.9% of the total number of successes. In line with the results in Cheung et al. (2005), we also find some clustering of outperformance at the long horizon. 26 entries, or 36.1% of the total number of success cases, are at the 5 year horizon.[22]

In terms of the economic models, one finding is that relative purchasing power parity, estimated using an error correction specification, noticeably outperforms the random walk benchmark with a success rate of 72% (28 out of 39 cases). Recall, this is the case where the change in the exchange rate is related to the lagged real exchange rate; no contemporaneous information about price levels is included. The outperformance relative to the random walk is typically greater the longer the horizon, so that at the year horizon, the outperformance is statistically significant for all currencies for all periods (except for CAD and JPY in Period III; even then the MSE ratio is quite low). In contrast, this pattern does *not* extend to the first differences specification of relative PPP, wherein the exchange rate is allowed to move with the inflation differential plus a drift term that is updated by rolling. Hence, the inclusion of contemporaneous information (time t inflation differentials) does not offset the misspecification implicit in PPP in growth rates.[23]

Another result is that interest rate parity seldom works well, but if it does, it does so at a longer horizon, such as one year or 5 years. The statistically significant outperformances occur during Period II and Period III (3 successful cases out of 12 in

[22] However, at the 5 year horizon, the random walk model *significantly* outperforms the competing structural models in 70 cases (60.5% of the total outperformance cases).

[23] These findings are consistent with the results reported by Eichenbaum et al. (2017), which indicate that exchange rates respond to the real exchange rate when countries' central banks follow Taylor rules, and are consistent with those of Cheung et al. (2004).

Table 1a
The MSE ratios from the dollar-based exchange rates. *Period I: 1983q1-2014q4.*

Specification	Horizon	PPP	SPMM	BEER	IRP	RID	TRF	SPMA	YCS
Panel A: CAN$/$									
ECM	1 quarter	0.927	0.969	1.020	1.022	0.944	0.935		
		0.010	0.584	0.670	0.057	0.444	0.240		
	4 quarter	0.852	0.995	1.065	1.027	1.005	0.767		
		0.003	0.960	0.447	0.086	0.958	0.029		
	20 quarter	0.600	1.023	1.155	0.907	1.040	1.104		
		0.009	0.947	0.174	0.537	0.982	0.260		
FD	1 quarter	1.028	1.077	0.611		1.136			
		0.557	0.476	0.014		0.170			
	4 quarter	1.195	1.144	0.469		1.172			
		0.004	0.251	0.000		0.012			
	20 quarter	2.226	2.100	0.747		1.863			
		0.000	0.000	0.118		0.000			
Panel B: Yen/$									
ECM	1 quarter	0.917	0.932		1.016	0.957	0.899		0.878
		0.044	0.116		0.719	0.292	0.069		0.036
	4 quarter	0.825	0.895		1.110	0.875	0.883		0.754
		0.010	0.014		0.125	0.009	0.017		0.000
	20 quarter	0.524	0.950		1.050	0.935	0.946		0.729
		0.000	0.006		0.006	0.004	0.007		0.001
FD	1 quarter	1.062	1.158			1.050			
		0.224	0.073			0.337			
	4 quarter	1.084	1.195			1.074			
		0.507	0.170			0.608			
	20 quarter	1.857	2.176			1.884			
		0.058	0.003			0.038			
Panel C: SF/$									
ECM	1 quarter	0.941	0.991			0.967	0.926		
		0.088	0.755			0.300	0.059		
	4 quarter	0.813	0.969			0.986	0.940		
		0.004	0.278			0.362	0.068		
	20 quarter	0.398	1.080			1.024	1.171		
		0.000	0.002			0.000	0.048		
FD	1 quarter	1.062	1.306			1.074			
		0.088	0.050			0.046			
	4 quarter	1.156	1.656			1.160			
		0.032	0.000			0.041			
	20 quarter	2.309	1.615			1.975			
		0.000	0.185			0.022			
Panel D: BP/$									
ECM	1 quarter	0.900	0.943	0.997	1.015	0.952	0.916		0.938
		0.031	0.221	0.941	0.467	0.380	0.048		0.418
	4 quarter	0.733	0.952	1.106	1.022	1.070	0.923		1.062
		0.000	0.432	0.030	0.239	0.244	0.347		0.517
	20 quarter	0.339	1.192	1.496	0.958	1.650	1.273		1.306
		0.000	0.067	0.000	0.182	0.000	0.007		0.042
FD	1 quarter	1.023	1.177	1.006		1.112			
		0.628	0.002	0.971		0.068			
	4 quarter	1.095	1.145	1.056		1.188			
		0.040	0.093	0.758		0.001			
	20 quarter	1.616	1.512	1.389		2.335			
		0.000	0.000	0.008		0.000			

Note: Each cell in the Table has two entries. The first one is the MSE ratio (the MSEs of a structural model to the random walk specification). The entry underneath the MSE ratio is the p-value of the hypothesis that the MSEs of the structural and random walk models are the same (Diebold and Mariano, 1995). The notation used in the table is ECM: error correction specification; FD: first-difference specification; PPP: purchasing power parity; SPMM: sticky-price monetary model; BEER: behavioral equilibrium exchange rate model; IRP: interest rate parity model; RID: real interest differential model; TRF: Taylor rule fundamentals; SPMA: sticky-price monetary augmented model; YCS: yield curve slope model. The forecasting horizons (in quarters) are listed under the heading "Horizon."

Period II and 4 out of 15 in Period III, and none in Period I). It's important to recall that this is the only specification which involves absolutely no estimation. Thus, it's hard to discern whether the result is driven by model validity, or the absence of estimation uncertainty.

With respect to the new specifications, one can make the following observations. The real interest differential model, incorporating shadow policy rates does not do particularly well. The greatest success is in the longest prediction period

Table 1b
The MSE ratios from the dollar-based exchange rates. *Period II: 2001q1-2014q4.*

Specification	Horizon	PPP	SPMM	BEER	IRP	RID	TRF	SPMA	YCS
Panel A: CAN$/$									
ECM	1 quarter	0.958	0.983	0.938	0.996	0.970	0.955	0.911	0.987
		0.032	0.537	0.188	0.807	0.427	0.142	0.170	0.422
	4 quarter	0.849	1.039	1.057	0.981	1.055	0.828	1.061	1.070
		0.018	0.233	0.847	0.646	0.024	0.024	0.917	0.497
	20 quarter	0.234	1.353	1.744	1.227	1.435	1.464	1.835	1.712
		0.010	0.029	0.956	0.009	0.006	0.008	0.162	0.004
FD	1 quarter	0.959	1.026	0.435		1.039		1.005	
		0.415	0.453	0.004		0.032		0.990	
	4 quarter	1.009	1.122	0.361		1.104		1.094	
		0.997	0.006	0.000		0.000		0.920	
	20 quarter	1.345	1.574	0.475		1.590		3.699	
		0.005	0.000	0.000		0.000		0.000	
Panel B: Yen/$									
ECM	1 quarter	0.990	0.903	0.939	1.003	1.002	1.011	0.916	0.918
		0.616	0.114	0.380	0.691	0.610	0.642	0.120	0.744
	4 quarter	0.925	0.970	0.861	0.998	1.057	1.119	1.062	0.886
		0.999	0.857	0.043	0.937	0.407	0.097	0.493	0.159
	20 quarter	0.570	1.014	1.097	0.840	1.057	0.962	1.008	1.025
		0.016	0.059	0.007	0.005	0.087	0.061	0.071	0.014
FD	1 quarter	1.067	1.144	0.965		1.028		1.086	
		0.226	0.063	0.836		0.533		0.552	
	4 quarter	1.000	1.117	0.948		1.005		1.236	
		0.645	0.131	0.758		0.624		0.059	
	20 quarter	1.297	1.306	1.255		1.163		2.666	
		0.199	0.525	0.077		0.089		0.000	
Panel C: SF/$									
ECM	1 quarter	0.919	0.988	0.972	1.005	1.069	0.998	0.995	1.079
		0.242	0.628	0.572	0.536	0.504	0.702	0.628	0.576
	4 quarter	0.703	1.022	1.143	0.983	1.209	1.008	1.157	1.428
		0.042	0.022	0.417	0.031	0.979	0.014	0.783	0.220
	20 quarter	1.214	1.772	1.513	0.797	1.269	1.781	2.767	2.648
		0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000
FD	1 quarter	1.100	1.032	1.764		0.989		1.408	
		0.259	0.949	0.025		0.346		0.079	
	4 quarter	1.101	1.034	3.031		1.036		1.793	
		0.232	0.051	0.000		0.027		0.006	
	20 quarter	1.598	1.328	16.795		1.038		5.045	
		0.000	0.000	0.000		0.000		0.000	
Panel D: BP/$									
ECM	1 quarter	0.912	1.016	0.970	0.984	0.942	0.994	1.020	1.025
		0.099	0.514	0.362	0.734	0.490	0.900	0.663	0.166
	4 quarter	0.704	1.015	1.007	0.932	1.045	0.919	0.977	1.053
		0.026	0.900	0.731	0.897	0.124	0.490	0.585	0.053
	20 quarter	0.266	0.898	1.107	1.019	1.111	0.909	0.938	1.022
		0.000	0.097	0.003	0.804	0.009	0.090	0.843	0.254
FD	1 quarter	0.931	1.026	0.641		1.080		1.166	
		0.518	0.684	0.304		0.360		0.149	
	4 quarter	0.864	1.077	0.658		1.047		1.292	
		0.046	0.349	0.126		0.001		0.008	
	20 quarter	0.798	0.900	0.440		1.167		1.405	
		0.018	0.223	0.000		0.025		0.001	

Note: Each cell in the Table has two entries. The first one is the MSE ratio (the MSEs of a structural model to the random walk specification). The entry underneath the MSE ratio is the p-value of the hypothesis that the MSEs of the structural and random walk models are the same (Diebold and Mariano, 1995). The notation used in the table is ECM: error correction specification; FD: first-difference specification; PPP: purchasing power parity; SPMM: sticky-price monetary model; BEER: behavioral equilibrium exchange rate model; IRP: interest rate parity model; RID: real interest differential model; TRF: Taylor rule fundamentals; SPMA: sticky-price monetary augmented model; YCS: yield curve slope model. The forecasting horizons (in quarters) are listed under the heading "Horizon."

(Period I), with 2 significant cases (many entries – 11 – significantly above unity, though). Surprisingly, use of the shadow rates does not resurrect the real interest model for the latest prediction period.

What about augmenting the models with risk and liquidity factors? First, note that the workhorse model, the sticky price monetary model has an unremarkable performance in all three periods. Adding the VIX and TED spread to this model results in some improved performance for the JPY (error correction model), the Swiss franc at five years in Sample III. However, clearly adding these variables in is not a panacea for the poor prediction of the model.

Table 1c
The MSE ratios from the dollar-based exchange rates. *Period III: 2007q4-2014q4.*

Specification	Horizon	PPP	SPMM	BEER	IRP	RID	TRF	SPMA	YCS
Panel A: CAN$/$									
ECM	1 quarter	0.968	0.973	0.916	0.990	0.951	0.931	0.969	0.978
		0.251	0.300	0.698	0.145	0.338	0.131	0.683	0.662
	4 quarter	0.882	0.976	0.925	0.968	0.993	0.627	0.894	0.975
		0.066	0.430	0.798	0.037	0.683	0.002	0.313	0.575
	20 quarter	0.519	0.952	0.997	0.872	1.040	0.964	0.937	1.146
		0.144	0.866	0.679	0.604	0.770	0.698	0.405	0.115
FD	1 quarter	0.897	1.011	0.447		1.037		0.776	
		0.257	0.721	0.085		0.155		0.230	
	4 quarter	0.947	1.041	0.385		1.038		0.633	
		0.184	0.091	0.019		0.007		0.157	
	20 quarter	1.159	1.207	1.211		1.338		0.919	
		0.086	0.167	0.497		0.083		0.932	
Panel B: Yen/$									
ECM	1 quarter	0.989	0.946	0.912	0.979	0.971	1.058	0.832	0.943
		0.917	0.421	0.398	0.181	0.864	0.297	0.004	0.518
	4 quarter	0.923	0.966	0.854	0.940	0.933	1.118	0.762	0.885
		0.662	0.737	0.267	0.044	0.876	0.076	0.000	0.102
	20 quarter	0.509	1.019	0.852	0.810	0.869	0.884	0.924	0.918
		0.973	0.381	0.553	0.907	0.603	0.558	0.769	0.379
FD	1 quarter	1.042	1.122	0.650		1.002		1.018	
		0.578	0.137	0.039		0.825		0.872	
	4 quarter	0.889	1.041	0.634		0.954		1.130	
		0.218	0.412	0.000		0.914		0.177	
	20 quarter	0.948	1.338	0.697		1.165		1.643	
		0.480	0.144	0.842		0.253		0.019	
Panel C: SF/$									
ECM	1 quarter	0.919	1.042	0.971	0.989	1.025	0.973	0.999	1.025
		0.480	0.707	0.554	0.148	0.892	0.652	0.870	0.858
	4 quarter	0.771	0.950	1.019	0.971	1.229	0.943	1.024	1.135
		0.985	0.373	0.909	0.840	0.478	0.308	0.707	0.754
	20 quarter	0.246	2.047	1.489	0.981	1.021	1.491	0.940	1.061
		0.012	0.015	0.145	0.002	0.002	0.007	0.000	0.000
FD	1 quarter	1.048	0.981	1.065		0.989		1.730	
		0.729	0.668	0.725		0.648		0.119	
	4 quarter	1.004	1.016	1.047		0.990		1.621	
		0.630	0.549	0.927		0.601		0.077	
	20 quarter	1.050	1.130	2.327		1.208		3.545	
		0.002	0.002	0.238		0.014		0.002	
Panel D: BP/$									
ECM	1 quarter	0.896	0.968	0.935	0.960	0.943	0.962	1.015	1.023
		0.042	0.275	0.254	0.182	0.332	0.194	0.997	0.874
	4 quarter	0.568	0.938	1.011	0.885	1.045	0.961	0.978	1.028
		0.012	0.090	0.539	0.095	0.788	0.512	0.717	0.573
	20 quarter	0.040	0.753	1.019	0.870	1.025	0.665	0.874	1.105
		0.003	0.213	0.310	0.125	0.305	0.027	0.234	0.777
FD	1 quarter	0.848	1.102	0.515		1.114		1.159	
		0.201	0.363	0.329		0.289		0.401	
	4 quarter	0.911	1.392	0.554		1.041		1.078	
		0.071	0.226	0.250		0.697		0.999	
	20 quarter	1.085	1.579	0.141		1.035		0.760	
		0.440	0.306	0.284		0.568		0.891	
Panel E: EU/$									
ECM	1 quarter	0.876	0.979	0.905	0.986	0.996	0.960	0.985	0.949
		0.188	0.722	0.743	0.401	0.480	0.660	0.680	0.903
	4 quarter	0.558	1.043	1.072	0.946	1.142	1.153	0.849	1.276
		0.008	0.023	0.042	0.157	0.030	0.031	0.288	0.052
	20 quarter	0.310	1.253	1.205	0.945	1.169	1.897	1.308	1.431
		0.008	0.000	0.000	0.050	0.000	0.000	0.000	0.000

(*continued on next page*)

Table 1c (*continued*)

Specification	Horizon	PPP	SPMM	BEER	IRP	RID	TRF	SPMA	YCS
FD	1 quarter	0.994	1.197	1.341		1.079		1.589	
		0.669	0.096	0.211		0.283		0.053	
	4 quarter	1.101	1.269	6.023		1.201		1.202	
		0.009	0.017	0.027		0.005		0.200	
	20 quarter	1.476	2.020	11.974		1.582		3.391	
		0.000	0.000	0.001		0.000		0.053	

Note: Each cell in the Table has two entries. The first one is the MSE ratio (the MSEs of a structural model to the random walk specification). The entry underneath the MSE ratio is the p-value of the hypothesis that the MSEs of the structural and random walk models are the same (Diebold and Mariano, 1995). The notation used in the table is ECM: error correction specification; FD: first-difference specification; PPP: purchasing power parity; SPMM: sticky-price monetary model; BEER: behavioral equilibrium exchange rate model; IRP: interest rate parity model; RID: real interest differential model; TRF: Taylor rule fundamentals; SPMA: sticky-price monetary augmented model; YCS: yield curve slope model. The forecasting horizons (in quarters) are listed under the heading "Horizon."

The Taylor rule fundamentals model typically delivers outperformance relative to a random walk; however, unlike previous studies, we do not find statistically significant outperformance, except perhaps in Period I, which, out of a total of 12, registers 7 cases of outperformance (and 2 cases of underperformance).[24] We attribute this differing result to the fact that we impose the same model to all cases (in particular, we impose homogeneity of coefficients across countries, and omit an interest rate smoothing parameter).[25]

The yield curve model provides only a few statistically significant cases of outperformance. Out of 33 yield curve cases, there are three statistically significant outperformances (all involving the JPY in Period I).

The group performance of these four new specifications appears to be worse than the four specifications examined in Cheung et al. (2005). For instance, considering the reported MSE ratios, the four new specifications in total garner 20 significant success cases against 65 significant underperformance cases. The four "old" models, on the other hand, significantly outperformed the random walk 50 times, while being outperformed 59 times.

Notice that some of our models can only be compared during the most recent prediction period, starting in 2007 (Period III). Here, one noticeable result is that no structural model of the euro does particularly well in out of sample forecasting. In this period, out of the 37 statistically significant under-performances, 22 are associated with the euro. Rather than interpreting this as necessarily a failure of the models per se, we suspect this is largely due to the brevity of the sample period. Given the euro's inception in 1999, we only have 8 years of data in which to estimate the various models.

Consistent with the existing literature, our results are supportive of the assertion that it is difficult to consistently beat the random walk model using the MSE criterion. The current exercise further strengthens the assertion as it covers three different forecasting periods, and some structural models that have not been extensively studied before.

4.2. The direction of change criterion

Table 2 reports the proportion of forecasts that correctly predict the direction of the dollar-based exchange rate movement and, underneath these sample proportions, the p-values for the hypothesis that the reported proportion is significantly different from ½. When the proportion statistic is significantly larger than ½, the forecast is said to have the ability to predict the direct of change. On the other hand, if the statistic is significantly less than ½, the forecast tends to give the wrong direction of change. For trading purposes, information regarding the significance of incorrect prediction can be used to derive a potentially profitable trading rule by going again the prediction generated by the model. Following this argument, one might consider the cases in which the proportion of "correct" forecasts is larger than or less than ½ contain the same information. However, in evaluating the ability of the model to describe exchange rate behavior, we separate the two cases.

There is mixed evidence on the ability of the structural models to correctly predict the direction of change. Among the 462 direction of change statistics, 134 (27) are significantly larger (less) than ½ at the 10% level. The occurrence of the significant outperformance cases is higher (29%) than the one implied by the 10% level of the test.

Let us take a closer look at the incidences in which the forecasts are in the right direction. The 134 cases are unevenly split between the error correction and first-difference specifications – 89 from the former specification and 45 from the latter; that is, error correction specifications account for 66.4% of the success entries. It appears that the error correction specification – which incorporates the empirical long-run relationship – is a better specification for the models under consideration, according to the direction of change criterion.

[24] There are 3 cases of outperformance (4 cases of underperformance) and 2 cases of outperformance (4 cases of underperformance) in Period II and Period III, respectively.

[25] Obviously, to the extent that some central banks adhere to Taylor rules and others do not, we should expect cross country variation in the results. Also, the Taylor rule based exchange rate equation varies with the choice of the optimal interest rate rule that may not be the same across countries (Binici and Cheung, 2012).

Table 2a
Direction of change statistics from the dollar-based exchange rates. *Period I: 1983q1-2014q4.*

Specification	Horizon	PPP	SPMM	BEER	IRP	RID	TRF	SPMA	YCS
Panel A: CAN$/$									
ECM	1 quarter	0.656	0.531	0.523	0.461	0.539	0.594		
		0.000	0.480	0.596	0.377	0.377	0.034		
	4 quarter	0.664	0.680	0.592	0.480	0.568	0.632		
		0.000	0.000	0.040	0.655	0.128	0.003		
	20 quarter	0.624	0.661	0.569	0.578	0.624	0.578		
		0.010	0.001	0.151	0.103	0.010	0.103		
FD	1 quarter	0.547	0.531	0.664		0.453			
		0.289	0.480	0.000		0.289			
	4 quarter	0.504	0.584	0.784		0.496			
		0.929	0.060	0.000		0.929			
	20 quarter	0.294	0.514	0.560		0.431			
		0.000	0.774	0.213		0.151			
Panel B: Yen/$									
ECM	1 quarter	0.539	0.586		0.516	0.555	0.555		0.656
		0.377	0.052		0.724	0.216	0.216		0.000
	4 quarter	0.576	0.600		0.504	0.672	0.600		0.752
		0.089	0.025		0.929	0.000	0.025		0.000
	20 quarter	0.716	0.670		0.688	0.661	0.550		0.688
		0.000	0.000		0.000	0.001	0.292		0.000
FD	1 quarter	0.492	0.516			0.539			
		0.860	0.724			0.377			
	4 quarter	0.600	0.552			0.608			
		0.025	0.245			0.016			
	20 quarter	0.550	0.596			0.495			
		0.292	0.044			0.924			
Panel C: SF/$									
ECM	1 quarter	0.570	0.578			0.539	0.617		
		0.112	0.077			0.377	0.008		
	4 quarter	0.608	0.600			0.592	0.560		
		0.016	0.025			0.040	0.180		
	20 quarter	0.817	0.807			0.752	0.661		
		0.000	0.000			0.000	0.001		
FD	1 quarter	0.414	0.492			0.453			
		0.052	0.860			0.289			
	4 quarter	0.480	0.320			0.480			
		0.655	0.000			0.655			
	20 quarter	0.541	0.578			0.587			
		0.389	0.103			0.069			
Panel D: BP/$									
ECM	1 quarter	0.539	0.531	0.492	0.484	0.578	0.555		0.547
		0.377	0.480	0.860	0.724	0.077	0.216		0.289
	4 quarter	0.680	0.600	0.488	0.480	0.536	0.608		0.576
		0.000	0.025	0.788	0.655	0.421	0.016		0.089
	20 quarter	0.844	0.532	0.339	0.615	0.349	0.486		0.578
		0.000	0.503	0.001	0.017	0.002	0.774		0.103
FD	1 quarter	0.445	0.477	0.664		0.492			
		0.216	0.596	0.000		0.860			
	4 quarter	0.480	0.488	0.584		0.472			
		0.655	0.788	0.060		0.531			
	20 quarter	0.422	0.495	0.606		0.303			
		0.103	0.924	0.028		0.000			

Note: Each cell in the Table has two entries. The first one reports the proportion of forecasts that correctly predict the direction of the dollar-based exchange rate movement. Underneath each direction of change statistic are the p-values for the hypothesis that the reported proportion is significantly different from ½ is listed. When the statistic is significantly larger than ½, the forecast is said to have the ability to predict the direct of change. If the statistic is significantly less than ½, the forecast tends to give the wrong direction of change. The notation used in the table is ECM: error correction specification; FD: first-difference specification; PPP: purchasing power parity; SPMM: sticky-price monetary model; BEER: behavioral equilibrium exchange rate model; IRP: interest rate parity model; RID: real interest differential model; TRF: Taylor rule fundamentals; SPMA: sticky-price monetary augmented model; YCS: yield curve slope model. The forecasting horizons (in quarters) are listed under the heading "Horizon."

The forecasting period have an impact on prediction performance, as the success rate declines as the horizon gets shorter and shorter: 46.2% (62 cases) occurred in the 20-quarter horizon, 34.3.% (46 cases) in four-quarter horizon, and 19.4% (26 cases) in the one-quarter horizon. In addition, the significant underperformances are highest in the latest period (III), although most of these cases are all associated with the euro rate. Hence, this result might arise from the small sample we use to estimate the euro models.

Table 2b
Direction of change statistics from the dollar-based exchange rates. *Period II: 2001q1-2014q4.*

Specification	Horizon	PPP	SPMM	BEER	IRP	RID	TRF	SPMA	YCS
Panel A: CAN$/$									
ECM	1 quarter	0.625	0.464	0.536	0.464	0.536	0.536	0.625	0.643
		0.061	0.593	0.593	0.593	0.593	0.593	0.061	0.033
	4 quarter	0.642	0.434	0.509	0.585	0.377	0.528	0.566	0.566
		0.039	0.336	0.891	0.216	0.074	0.680	0.336	0.336
	20 quarter	0.622	0.541	0.568	0.405	0.541	0.568	0.730	0.838
		0.139	0.622	0.411	0.250	0.622	0.411	0.005	0.000
FD	1 quarter	0.571	0.464	0.750		0.482		0.571	
		0.285	0.593	0.000		0.789		0.285	
	4 quarter	0.528	0.396	0.811		0.340		0.585	
		0.680	0.131	0.000		0.020		0.216	
	20 quarter	0.568	0.216	0.892		0.243		0.622	
		0.411	0.001	0.000		0.002		0.139	
Panel B: Yen/$									
ECM	1 quarter	0.518	0.589	0.571	0.536	0.554	0.589	0.625	0.554
		0.789	0.181	0.285	0.593	0.423	0.181	0.061	0.423
	4 quarter	0.547	0.585	0.679	0.547	0.566	0.566	0.528	0.566
		0.492	0.216	0.009	0.492	0.336	0.336	0.680	0.336
	20 quarter	0.811	0.811	0.784	0.811	0.811	0.838	0.568	0.811
		0.000	0.000	0.001	0.000	0.000	0.000	0.411	0.000
FD	1 quarter	0.554	0.464	0.643		0.482		0.500	
		0.423	0.593	0.033		0.789		1.000	
	4 quarter	0.585	0.566	0.736		0.528		0.509	
		0.216	0.336	0.001		0.680		0.891	
	20 quarter	0.811	0.730	0.730		0.811		0.486	
		0.000	0.005	0.005		0.000		0.869	
Panel C: SF/$									
ECM	1 quarter	0.536	0.607	0.589	0.518	0.518	0.554	0.589	0.536
		0.593	0.109	0.181	0.789	0.789	0.423	0.181	0.593
	4 quarter	0.566	0.698	0.623	0.642	0.585	0.679	0.566	0.453
		0.336	0.004	0.074	0.039	0.216	0.009	0.336	0.492
	20 quarter	0.730	0.865	1.000	0.946	1.000	0.892	0.838	0.838
		0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000
FD	1 quarter	0.482	0.536	0.554		0.536		0.607	
		0.789	0.593	0.423		0.593		0.109	
	4 quarter	0.717	0.755	0.509		0.736		0.509	
		0.002	0.000	0.891		0.001		0.891	
	20 quarter	1.000	1.000	0.297		1.000		0.405	
		0.000	0.000	0.014		0.000		0.250	
Panel D: BP/$									
ECM	1 quarter	0.500	0.500	0.518	0.536	0.500	0.536	0.536	0.518
		1.000	1.000	0.789	0.593	1.000	0.593	0.593	0.789
	4 quarter	0.547	0.491	0.396	0.472	0.491	0.472	0.604	0.415
		0.492	0.891	0.131	0.680	0.891	0.680	0.131	0.216
	20 quarter	0.919	0.703	0.405	0.486	0.324	0.703	0.459	0.514
		0.000	0.014	0.250	0.869	0.033	0.014	0.622	0.869
FD	1 quarter	0.571	0.589	0.732		0.393		0.500	
		0.285	0.181	0.001		0.109		1.000	
	4 quarter	0.717	0.528	0.642		0.264		0.434	
		0.002	0.680	0.039		0.001		0.336	
	20 quarter	0.703	0.541	0.811		0.405		0.378	
		0.014	0.622	0.000		0.250		0.139	

Note: Each cell in the Table has two entries. The first one reports the proportion of forecasts that correctly predict the direction of the dollar-based exchange rate movement. Underneath each direction of change statistic are the p-values for the hypothesis that the reported proportion is significantly different from ½ is listed. When the statistic is significantly larger than ½, the forecast is said to have the ability to predict the direct of change. If the statistic is significantly less than ½, the forecast tends to give the wrong direction of change. The notation used in the table is ECM: error correction specification; FD: first-difference specification; PPP: purchasing power parity; SPMM: sticky-price monetary model; BEER: behavioral equilibrium exchange rate model; IRP: interest rate parity model; RID: real interest differential model; TRF: Taylor rule fundamentals; SPMA: sticky-price monetary augmented model; YCS: yield curve slope model. The forecasting horizons (in quarters) are listed under the heading "Horizon."

It is hard to make generalizations about which model performs the best. For instance, the BEER model accounts for 30 significant outperformances. Perhaps PPP does best among all the models, with 30 cases, with 21 cases pertaining to error correction models. Recall that this means time t to t + k information regarding differential inflation is less useful than reversion to the real rate in predicting the direction of change.

Table 2c
Direction of change statistics from the dollar-based exchange rates. *Period III: 2007q4-2014q4.*

	Horizon	PPP	SPMM	BEER	IRP	RID	TRF	SPMA	YCS
Panel A: CAN$/$									
ECM	1 quarter	0.517	0.448	0.517	0.483	0.414	0.586	0.483	0.586
		0.853	0.577	0.853	0.853	0.353	0.353	0.853	0.353
	4 quarter	0.731	0.423	0.385	0.692	0.308	0.808	0.423	0.423
		0.019	0.433	0.239	0.050	0.050	0.002	0.433	0.433
	20 quarter	0.600	0.500	0.600	0.500	0.500	0.600	0.600	0.600
		0.527	1.000	0.527	1.000	1.000	0.527	0.527	0.527
FD	1 quarter	0.586	0.448	0.759		0.448		0.655	
		0.353	0.577	0.005		0.577		0.095	
	4 quarter	0.462	0.462	0.769		0.231		0.538	
		0.695	0.695	0.006		0.006		0.695	
	20 quarter	0.400	0.400	0.700		0.400		0.600	
		0.527	0.527	0.206		0.527		0.527	
Panel B: Yen/$									
ECM	1 quarter	0.655	0.621	0.552	0.586	0.517	0.483	0.655	0.552
		0.095	0.194	0.577	0.353	0.853	0.853	0.095	0.577
	4 quarter	0.731	0.577	0.692	0.615	0.654	0.462	0.923	0.577
		0.019	0.433	0.050	0.239	0.117	0.695	0.000	0.433
	20 quarter	0.500	0.500	0.500	0.500	0.500	0.600	0.600	0.500
		1.000	1.000	1.000	1.000	1.000	0.527	0.527	1.000
FD	1 quarter	0.586	0.379	0.793		0.483		0.448	
		0.353	0.194	0.002		0.853		0.577	
	4 quarter	0.769	0.500	0.846		0.577		0.615	
		0.006	1.000	0.000		0.433		0.239	
	20 quarter	0.500	0.500	0.500		0.500		0.200	
		1.000	1.000	1.000		1.000		0.058	
Panel C: SF/$									
ECM	1 quarter	0.483	0.379	0.621	0.586	0.483	0.586	0.586	0.552
		0.853	0.194	0.194	0.353	0.853	0.353	0.353	0.577
	4 quarter	0.423	0.654	0.615	0.462	0.538	0.654	0.577	0.615
		0.433	0.117	0.239	0.695	0.695	0.117	0.433	0.239
	20 quarter	0.200	1.000	1.000	0.800	1.000	1.000	1.000	1.000
		0.058	0.002	0.002	0.058	0.002	0.002	0.002	0.002
FD	1 quarter	0.448	0.517	0.517		0.552		0.448	
		0.577	0.853	0.853		0.577		0.577	
	4 quarter	0.654	0.654	0.615		0.654		0.615	
		0.117	0.117	0.239		0.117		0.239	
	20 quarter	1.000	1.000	1.000		1.000		0.300	
		0.002	0.002	0.002		0.002		0.206	
Panel D: BP/$									
ECM	1 quarter	0.483	0.621	0.655	0.586	0.586	0.517	0.586	0.586
		0.853	0.194	0.095	0.353	0.353	0.853	0.353	0.353
	4 quarter	0.577	0.692	0.538	0.615	0.538	0.500	0.577	0.615
		0.433	0.050	0.695	0.239	0.695	1.000	0.433	0.239
	20 quarter	0.900	0.400	0.500	0.600	0.500	0.600	0.600	0.600
		0.011	0.527	1.000	0.527	1.000	0.527	0.527	0.527
FD	1 quarter	0.655	0.655	0.793		0.379		0.414	
		0.095	0.095	0.002		0.194		0.353	
	4 quarter	0.654	0.500	0.692		0.500		0.538	
		0.117	1.000	0.050		1.000		0.695	
	20 quarter	0.600	0.600	0.800		0.500		0.700	
		0.527	0.527	0.058		1.000		0.206	
Panel E: EU/$									
ECM	1 quarter	0.655	0.448	0.517	0.552	0.414	0.483	0.483	0.517
		0.095	0.577	0.853	0.577	0.353	0.853	0.853	0.853
	4 quarter	0.769	0.462	0.577	0.538	0.423	0.538	0.577	0.500
		0.006	0.695	0.433	0.695	0.433	0.695	0.433	1.000
	20 quarter	0.900	0.100	0.100	0.800	0.100	0.100	0.100	0.100
		0.011	0.011	0.011	0.058	0.011	0.011	0.011	0.011

(*continued on next page*)

Table 2c (*continued*)

	Horizon	PPP	SPMM	BEER	IRP	RID	TRF	SPMA	YCS
FD	1 quarter	0.414	0.379	0.414		0.379		0.483	
		0.353	0.194	0.353		0.194		0.853	
	4 quarter	0.423	0.462	0.500		0.423		0.500	
		0.433	0.695	1.000		0.433		1.000	
	20 quarter	0.100	0.100	0.100		0.100		0.400	
		0.011	0.011	0.011		0.011		0.527	

Note: Each cell in the Table has two entries. The first one reports the proportion of forecasts that correctly predict the direction of the dollar-based exchange rate movement. Underneath each direction of change statistic are the p-values for the hypothesis that the reported proportion is significantly different from ½ is listed. When the statistic is significantly larger than ½, the forecast is said to have the ability to predict the direct of change. If the statistic is significantly less than ½, the forecast tends to give the wrong direction of change. The notation used in the table is ECM: error correction specification; FD: first-difference specification; PPP: purchasing power parity; SPMM: sticky-price monetary model; BEER: behavioral equilibrium exchange rate model; IRP: interest rate parity model; RID: real interest differential model; TRF: Taylor rule fundamentals; SPMA: sticky-price monetary augmented model; YCS: yield curve slope model. The forecasting horizons (in quarters) are listed under the heading "Horizon."

In terms of innovations, the yield curve models work quite well in the first two periods (I, II): 8 out of 18 cases yield outperformance – but the outperformance is currency specific – there is only one successful case for the British pound, one for Swiss franc. In addition, the performance breaks down in the latest period (III).

The sticky price monetary model does particularly poorly in Period III (4 significant outperformances). When augmented with the VIX and the TED (the augmented sticky price monetary model), the model fails to improve noticeably in this dimension.

In terms of the economics, it is not clear that the newer exchange rate models decisively edge out the "old fashioned" sticky-price model.

The cases of correct direction prediction appear to cluster at the long forecast horizon. The 20-quarter horizon accounts for 62 of the 134 cases. This is about the same proportion than in Cheung et al. (where the long horizon accounted for about 36.5% of the successes) Mirroring the MSE results, it is interesting to note that the direction of change statistic tends to work for the interest rate parity model only at the 20-quarter horizon. This pattern is entirely consistent with the finding that uncovered interest parity holds better at long horizons.

4.3. The consistency criterion

The consistency criterion only requires the forecast and actual realization comove one-to-one in the long run. One may argue that the criterion is less demanding than the MSE and direct of change metrics. Indeed, a forecast satisfies the consistency criterion can (1) have a MSE larger than that of the random walk model, (2) have a direction of change statistic less than ½, or (3) generate forecast errors that are serially correlated. However, given the problems related to modeling, estimation, and data quality, the consistency criterion can be a more flexible way to evaluate a forecast. In assessing the consistency, we first test if the forecast and the realization are cointegrated.[26] If they are cointegrated, then we test if the cointegrating vector satisfies the (1, −1) requirement. The cointegration results are reported in Table 3. The test results for the (1, −1) restriction are reported in Table 4.

275 of 462 cases reject the null hypothesis of no cointegration at the 10% significance level. Thus, 275 forecast series (59.5% of the total number) are cointegrated with the corresponding spot exchange rates. The error correction specification and the first-difference specification account for, respectively, 158 (57.5%) and 117 (42.5%) of the cointegrated cases. On this count, the error correction specification imposing a cointegrating relationship performs better than the first-difference specification.

There is no real pattern in terms of findings of cointegration, across currencies and models, at least in Periods I and II. The largest difference is the decrease in number of cointegrated cases in Period III; the proportion drops from 71% to 53% moving from I and II, to period III. This is to be expected given the decrease in number of observations as one goes to the latest period.[27]

The results of testing for the long-run unitary elasticity of expectations at the 10% significance level are reported in Table 4. In total there are only 9 cases in which the (1, −1) restriction is not rejected. The condition of long-run unitary elasticity of expectations; that is the (1, −1) restriction on the cointegrating vector, is rejected by the data in almost all cases, for the longest period (I). Only when examining the shortest out-of-sample periods (III) is it the case that there are countable failures to reject (5 out of 104 cases). This indicates that the "consistency" criterion is a very difficult one to meet using the models and empirical methods we have adopted.

[26] The Johansen method is used to test the null hypothesis of no cointegration. The maximum eigenvalue statistics are reported in the manuscript. Results based on the trace statistics are essentially the same. Before implementing the cointegration test, both the forecast and exchange rate series were checked for the I(1) property. For brevity, the I(1) test results and the trace statistics are not reported.

[27] There will only be 8 observations in the five year ahead forecasts for Sample III, for instance.

Table 3a
Cointegration between dollar-based exchange rates and their forecasts. *Period I: 1983q1-2014q4.*

Specification	Horizon	PPP	SPMM	BEER	IRP	RID	TRF	SPMA	YCS
Panel A: CAN$/$									
ECM	1 quarter	7.564	45.286	55.842	4.890	45.442	32.772		
		0.629	0.001	0.001	0.898	0.001	0.001		
	4 quarter	48.271	49.208	46.940	43.647	46.900	26.173		
		0.001	0.001	0.001	0.001	0.001	0.001		
	20 quarter	5.670	7.628	10.794	5.976	10.134	12.603		
		0.824	0.623	0.299	0.791	0.367	0.154		
FD	1 quarter	62.396	92.624	72.232		109.964			
		0.001	0.001	0.001		0.001			
	4 quarter	44.767	43.573	47.154		44.987			
		0.001	0.001	0.001		0.001			
	20 quarter	6.651	9.037	11.493		6.370			
		0.722	0.479	0.228		0.751			
Panel B: Yen/$									
ECM	1 quarter	15.055	72.139		14.434	40.931	43.834		14.243
		0.067	0.001		0.084	0.001	0.001		0.089
	4 quarter	42.509	41.745		39.352	39.529	39.281		31.511
		0.001	0.001		0.001	0.001	0.001		0.001
	20 quarter	18.254	8.122		8.255	8.433	7.837		12.163
		0.021	0.572		0.559	0.540	0.601		0.177
FD	1 quarter	46.820	74.253			76.519			
		0.001	0.001			0.001			
	4 quarter	38.005	33.042			37.943			
		0.001	0.001			0.001			
	20 quarter	6.388	7.594			7.190			
		0.749	0.626			0.667			
Panel C: SF/$									
ECM	1 quarter	7.915	83.230			61.838	103.310		
		0.593	0.001			0.001	0.001		
	4 quarter	44.791	39.555			53.772	41.980		
		0.001	0.001			0.001	0.001		
	20 quarter	9.682	11.555			13.427	8.766		
		0.413	0.222			0.118	0.506		
FD	1 quarter	87.446	99.953			108.487			
		0.001	0.001			0.001			
	4 quarter	43.481	36.278			40.980			
		0.001	0.001			0.001			
	20 quarter	8.402	10.862			8.583			
		0.544	0.292			0.525			
Panel D: BP/$									
ECM	1 quarter	8.710	42.517	62.709	8.384	117.159	68.513		14.307
		0.512	0.001	0.001	0.545	0.001	0.001		0.087
	4 quarter	34.164	38.779	39.665	38.147	41.653	34.689		45.750
		0.001	0.001	0.001	0.001	0.001	0.001		0.001
	20 quarter	13.792	14.063	14.508	14.626	15.233	14.851		15.379
		0.104	0.095	0.081	0.078	0.063	0.072		0.060
FD	1 quarter	93.173	82.837	72.536		138.032			
		0.001	0.001	0.001		0.001			
	4 quarter	40.322	37.764	24.937		38.056			
		0.001	0.001	0.002		0.001			
	20 quarter	16.402	17.582	21.326		13.373			
		0.042	0.027	0.007		0.120			

Note: Each cell in the Table has two entries. The first one reports the Johansen maximum eigenvalue statistic for the null hypothesis that an exchange rate and its forecast are not cointegrated. The entry underneath reports the p-value for the null hypothesis. The notation used in the table is ECM: error correction specification; FD: first-difference specification; PPP: purchasing power parity; SPMM: sticky-price monetary model; BEER: behavioral equilibrium exchange rate model; IRP: interest rate parity model; RID: real interest differential model; TRF: Taylor rule fundamentals; SPMA: sticky-price monetary augmented model; YCS: yield curve slope model. The forecasting horizons (in quarters) are listed under the heading "Horizon."

4.4. Discussion

Several aspects of the foregoing analysis merit discussion. To begin with, even at long horizons, the performance of the structural models is less than impressive along the MSE dimension. This result is consistent with those Cheung et al. (2005), although the results are more promising, with higher proportions of outperformance.

Table 3b
Cointegration between dollar-based exchange rates and their forecasts. *Period II: 2001q1-2014q4.*

Specification	Horizon	PPP	SPMM	BEER	IRP	RID	TRF	SPMA	YCS
Panel A: CAN$/$									
ECM	1 quarter	6.582	39.598	12.932	5.446	30.306	10.464	28.521	4.620
		0.729	0.001	0.139	0.847	0.001	0.333	0.001	0.919
	4 quarter	21.253	19.694	22.786	20.919	20.641	16.612	24.633	21.517
		0.007	0.012	0.004	0.008	0.009	0.039	0.002	0.006
	20 quarter	9.840	9.648	10.261	10.275	9.966	9.876	12.341	10.314
		0.397	0.416	0.354	0.352	0.384	0.393	0.167	0.348
FD	1 quarter	46.488	37.776	26.238		49.788		38.463	
		0.001	0.001	0.001		0.001		0.001	
	4 quarter	20.063	20.601	23.016		20.516		18.377	
		0.010	0.009	0.004		0.009		0.020	
	20 quarter	9.711	10.068	10.458		9.875		10.987	
		0.410	0.373	0.334		0.393		0.280	
Panel B: Yen/$									
ECM	1 quarter	18.439	44.046	41.331	5.209	42.452	22.192	43.350	8.795
		0.020	0.001	0.001	0.869	0.001	0.005	0.001	0.503
	4 quarter	17.580	18.473	17.083	15.584	18.020	16.338	17.122	16.418
		0.027	0.019	0.033	0.056	0.023	0.043	0.032	0.041
	20 quarter	16.479	5.066	5.354	3.614	5.631	7.406	7.858	6.199
		0.040	0.882	0.856	0.974	0.828	0.645	0.599	0.769
FD	1 quarter	40.862	56.223	25.468		65.609		35.570	
		0.001	0.001	0.001		0.001		0.001	
	4 quarter	16.679	16.184	13.565		15.594		16.181	
		0.038	0.045	0.113		0.056		0.045	
	20 quarter	3.898	3.624	6.840		3.897		11.642	
		0.962	0.974	0.703		0.962		0.213	
Panel C: SF/$									
ECM	1 quarter	10.594	80.981	34.108	11.086	60.800	96.831	77.672	12.126
		0.320	0.001	0.001	0.269	0.001	0.001	0.001	0.179
	4 quarter	18.105	20.158	21.813	17.737	19.159	19.784	22.725	17.001
		0.022	0.010	0.005	0.025	0.015	0.012	0.004	0.033
	20 quarter	11.416	9.549	9.582	10.089	8.761	7.088	12.029	10.150
		0.236	0.426	0.423	0.371	0.507	0.678	0.185	0.365
FD	1 quarter	73.580	94.727	38.828		94.972		47.881	
		0.001	0.001	0.001		0.001		0.001	
	4 quarter	18.643	19.171	14.516		17.212		22.414	
		0.018	0.015	0.081		0.031		0.004	
	20 quarter	8.133	9.041	10.363		7.380		11.963	
		0.571	0.478	0.343		0.648		0.188	
Panel D: BP/$									
ECM	1 quarter	10.910	35.819	43.127	5.294	42.720	25.942	40.478	13.283
		0.287	0.001	0.001	0.861	0.001	0.001	0.001	0.123
	4 quarter	18.959	15.693	17.607	20.190	19.665	14.230	19.850	19.304
		0.016	0.054	0.027	0.010	0.012	0.090	0.012	0.014
	20 quarter	6.739	10.630	10.890	8.523	10.225	13.997	8.040	9.493
		0.713	0.316	0.289	0.531	0.357	0.097	0.581	0.432
FD	1 quarter	59.496	30.726	31.812		91.589		41.131	
		0.001	0.001	0.001		0.001		0.001	
	4 quarter	21.582	27.196	12.047		18.835		20.866	
		0.006	0.001	0.184		0.017		0.008	
	20 quarter	7.198	14.150	10.579		10.763		12.372	
		0.667	0.092	0.321		0.302		0.166	

Note: Each cell in the Table has two entries. The first one reports the Johansen maximum eigenvalue statistic for the null hypothesis that an exchange rate and its forecast are not cointegrated. The entry underneath reports the p-value for the null hypothesis. The notation used in the table is ECM: error correction specification; FD: first-difference specification; PPP: purchasing power parity; SPMM: sticky-price monetary model; BEER: behavioral equilibrium exchange rate model; IRP: interest rate parity model; RID: real interest differential model; TRF: Taylor rule fundamentals; SPMA: sticky-price monetary augmented model; YCS: yield curve slope model. The forecasting horizons (in quarters) are listed under the heading "Horizon."

Setting aside issues of statistical significance, it is interesting that the interest rate parity model at the 4- and 20-quarter horizons does particularly well in period III. This is true, despite the fact that interest rate parity does not appear to hold as well for interest rates bound at zero, of which there are several during the 2007–14 period.[28]

[28] Chinn and Quayyum (2012) document the fact that long horizon uncovered interest parity doesn't hold as well for Japan and Switzerland over the last decade.

Table 3c
Cointegration between dollar-based exchange rates and their forecasts. *Period III: 2007q4-2014q4.*

Specification	Horizon	PPP	SPMM	BEER	IRP	RID	TRF	SPMA	YCS
Panel A: CAN$/$									
ECM	1 quarter	9.631	18.087	10.139	10.425	12.024	16.608	8.453	13.376
		0.418	0.022	0.366	0.337	0.185	0.039	0.538	0.120
	4 quarter	11.975	11.403	12.915	12.340	11.890	10.599	10.592	12.574
		0.188	0.237	0.139	0.167	0.193	0.319	0.320	0.155
	20 quarter	11.822	7.100	9.421	18.486	17.090	17.265	18.585	17.538
		0.197	0.677	0.440	0.019	0.032	0.030	0.019	0.027
FD	1 quarter	29.545	21.602	17.886		29.725		26.240	
		0.001	0.006	0.024		0.001		0.001	
	4 quarter	11.582	12.495	16.740		11.889		11.681	
		0.219	0.159	0.037		0.193		0.209	
	20 quarter	14.935	16.248	14.591		17.213		16.778	
		0.070	0.044	0.079		0.031		0.036	
Panel B: Yen/$									
ECM	1 quarter	24.918	17.789	11.867	31.949	23.081	8.788	26.279	29.464
		0.002	0.025	0.194	0.001	0.004	0.504	0.001	0.001
	4 quarter	9.426	7.847	8.608	8.299	8.217	7.047	8.681	10.188
		0.439	0.600	0.523	0.554	0.562	0.682	0.515	0.361
	20 quarter	10.913	14.441	12.338	12.909	10.287	8.214	8.523	12.425
		0.287	0.083	0.168	0.140	0.351	0.563	0.531	0.163
FD	1 quarter	28.720	32.919	20.547		32.001		23.136	
		0.001	0.001	0.009		0.001		0.004	
	4 quarter	8.931	8.305	7.977		8.199		7.829	
		0.490	0.553	0.587		0.564		0.602	
	20 quarter	10.087	9.199	8.535		11.695		12.481	
		0.371	0.462	0.530		0.207		0.160	
Panel C: SF/$									
ECM	1 quarter	10.250	36.460	16.953	25.641	35.972	38.802	21.648	4.443
		0.355	0.001	0.034	0.001	0.001	0.001	0.006	0.931
	4 quarter	7.694	8.864	8.915	8.232	10.423	7.984	9.353	7.880
		0.616	0.496	0.491	0.561	0.337	0.586	0.446	0.597
	20 quarter	6.747	16.057	12.356	10.846	14.462	21.015	17.605	13.272
		0.713	0.047	0.167	0.294	0.083	0.008	0.027	0.124
FD	1 quarter	36.062	41.527	26.778		48.343		41.069	
		0.001	0.001	0.001		0.001		0.001	
	4 quarter	9.054	8.268	12.527		8.307		11.510	
		0.477	0.557	0.158		0.553		0.226	
	20 quarter	20.187	12.747	29.077		15.080		16.108	
		0.010	0.147	0.001		0.067		0.046	
Panel D: BP/$									
ECM	1 quarter	14.234	13.530	27.554	28.427	26.970	13.658	26.140	16.500
		0.089	0.114	0.001	0.001	0.001	0.109	0.001	0.040
	4 quarter	17.780	17.377	17.357	17.216	17.180	18.608	20.105	17.109
		0.025	0.029	0.029	0.031	0.031	0.018	0.010	0.032
	20 quarter	40.217	32.386	30.639	30.963	31.291	32.030	26.682	34.418
		0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
FD	1 quarter	35.498	20.861	17.396		43.574		42.512	
		0.001	0.008	0.029		0.001		0.001	
	4 quarter	16.863	23.915	26.191		17.659		23.508	
		0.035	0.003	0.001		0.026		0.003	
	20 quarter	33.814	26.296	14.840		33.347		25.529	
		0.001	0.001	0.073		0.001		0.001	
Panel E: EU/$									
ECM	1 quarter	10.477	14.699	15.838	22.366	13.999	22.202	9.550	10.540
		0.332	0.076	0.051	0.005	0.097	0.005	0.426	0.325
	4 quarter	13.438	12.428	13.449	13.776	14.174	11.296	16.253	15.708
		0.117	0.163	0.117	0.105	0.091	0.248	0.044	0.053
	20 quarter	4.641	13.729	10.087	12.623	20.576	32.626	14.331	4.881
		0.917	0.107	0.371	0.153	0.009	0.001	0.087	0.898

(*continued on next page*)

Table 3c (*continued*)

Specification	Horizon	PPP	SPMM	BEER	IRP	RID	TRF	SPMA	YCS
FD	1 quarter	24.265	28.753	28.399		33.951		24.612	
		0.003	0.00b1	0.001		0.001		0.002	
	4 quarter	12.605	13.094	13.049		13.635		11.459	
		0.154	0.131	0.133		0.110		0.231	
	20 quarter	12.451	11.739	19.181		24.208		13.483	
		0.162	0.203	0.015		0.003		0.116	

Note: Each cell in the Table has two entries. The first one reports the Johansen maximum eigenvalue statistic for the null hypothesis that an exchange rate and its forecast are not cointegrated. The entry underneath reports the p-value for the null hypothesis. The notation used in the table is ECM: error correction specification; FD: first-difference specification; PPP: purchasing power parity; SPMM: sticky-price monetary model; BEER: behavioral equilibrium exchange rate model; IRP: interest rate parity model; RID: real interest differential model; TRF: Taylor rule fundamentals; SPMA: sticky-price monetary augmented model; YCS: yield curve slope model. The forecasting horizons (in quarters) are listed under the heading "Horizon."

Table 4a
Results of the (1, −1) restriction test: dollar-based exchange rates. *Period I: 1983q1-2014q4*.

Specification	Horizon	PPP	SPMM	BEER	IRP	RID	TRF	SPMA	YCS
Panel A: CAN$/$									
ECM	1 quarter		24.849	25.471		16.648	23.444		
			0.000	0.000		0.000	0.000		
	4 quarter	47.280	45.221	38.112	42.459	41.450	21.411		
		0.000	0.000	0.000	0.000	0.000	0.000		
	20 quarter								
FD	1 quarter	28.609	62.635	70.450		58.588			
		0.000	0.000	0.000		0.000			
	4 quarter	38.600	33.900	45.362		39.986			
		0.000	0.000	0.000		0.000			
	20 quarter								
Panel B: Yen/$									
ECM	1 quarter	7.444	70.187		1.938	38.347	41.802		11.466
		0.006	0.000		0.164	0.000	0.000		0.001
	4 quarter	39.305	40.019		37.206	37.851	37.666		29.576
		0.000	0.000		0.000	0.000	0.000		0.000
	20 quarter	14.702							
		0.000							
FD	1 quarter	43.798	72.575			74.340			
		0.000	0.000			0.000			
	4 quarter	36.556	31.104			36.271			
		0.000	0.000			0.000			
	20 quarter								
Panel C: SF/$									
ECM	1 quarter		56.509			46.934	74.098		
			0.000			0.000	0.000		
	4 quarter	40.240	32.076			45.131	35.007		
		0.000	0.000			0.000	0.000		
	20 quarter								
FD	1 quarter	49.139	94.942			76.324			
		0.000	0.000			0.000			
	4 quarter	35.653	30.811			34.346			
		0.000	0.000			0.000			
	20 quarter								
Panel D: BP/$									
ECM	1 quarter		41.935	54.374		103.092	60.086		13.083
			0.000	0.000		0.000	0.000		0.000
	4 quarter	33.858	38.399	37.648	38.002	39.044	33.429		40.703
		0.000	0.000	0.000	0.000	0.000	0.000		0.000
	20 quarter		7.405	6.903	10.319	7.581	6.483		12.625
			0.007	0.009	0.001	0.006	0.011		0.000

Table 4a (*continued*)

Specification	Horizon	PPP	SPMM	BEER	IRP	RID	TRF	SPMA	YCS
FD	1 quarter	86.641	77.771	66.489		111.515			
		0.000	0.000	0.000		0.000			
	4 quarter	39.833	37.331	22.880		37.035			
		0.000	0.000	0.000		0.000			
	20 quarter	11.093	16.747	19.502					
		0.001	0.000	0.000					

Note: Each cell in the Table has two entries. The first entry is the likelihood ratio test statistic for the restriction of (1, −1) on the cointegrating vector. The entry underneath is its p-value. The test is only applied to the cointegration cases present in Table 3. The notation used in the table is ECM: error correction specification; FD: first-difference specification; PPP: purchasing power parity; SPMM: sticky-price monetary model; BEER: behavioral equilibrium exchange rate model; IRP: interest rate parity model; RID: real interest differential model; TRF: Taylor rule fundamentals; SPMA: sticky-price monetary augmented model; YCS: yield curve slope model. The forecasting horizons (in quarters) are listed under the heading "Horizon."

Table 4b
Results of the (1, −1) restriction test: dollar-based exchange rates. *Period II: 2001q1-2014q4.*

Specification	Horizon	PPP	SPMM	BEER	IRP	RID	TRF	SPMA	YCS
Panel A: CAN$/$									
ECM	1 quarter		23.355			18.437		9.501	
			0.000			0.000		0.002	
	4 quarter	13.204	11.951	13.386	13.630	12.866	8.518	14.256	12.810
		0.000	0.001	0.000	0.000	0.000	0.004	0.000	0.000
	20 quarter								
FD	1 quarter	31.512	31.503	22.126		11.973		24.435	
		0.000	0.000	0.000		0.001		0.000	
	4 quarter	12.487	12.735	15.697		12.733		5.565	
		0.000	0.000	0.000		0.000		0.018	
	20 quarter								
Panel B: Yen/$									
ECM	1 quarter	16.159	39.557	40.788		38.379	21.898	36.281	
		0.000	0.000	0.000		0.000	0.000	0.000	
	4 quarter	17.083	18.100	16.941	15.266	17.832	15.931	16.489	15.421
		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	20 quarter	3.513							
		0.061							
FD	1 quarter	39.875	45.822	25.435		64.889		35.433	
		0.000	0.000	0.000		0.000		0.000	
	4 quarter	16.442	15.492			15.414		14.837	
		0.000	0.000			0.000		0.000	
	20 quarter								
Panel C: SF/$									
ECM	1 quarter		49.744	15.483		33.032	55.732	32.060	
			0.000	0.000		0.000	0.000	0.000	
	4 quarter	4.088	5.534	3.346	4.274	6.125	4.072	5.947	2.483
		0.043	0.019	0.067	0.039	0.013	0.044	0.015	0.115
	20 quarter								
FD	1 quarter	31.427	64.961	19.402		28.778		30.701	
		0.000	0.000	0.000		0.000		0.000	
	4 quarter	4.106	5.328	0.671		3.929		4.347	
		0.043	0.021	0.413		0.047		0.037	
	20 quarter								
Panel D: BP/$									
ECM	1 quarter		35.037	43.091		41.114	22.122	39.803	
			0.000	0.000		0.000	0.000	0.000	
	4 quarter	17.851	15.564	17.521	19.948	19.534	14.087	19.437	19.088
		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	20 quarter						5.390		
							0.020		
FD	1 quarter	54.917	25.682	31.619		90.767		36.244	
		0.000	0.000	0.000		0.000		0.000	
	4 quarter	21.269	26.581			18.748		19.644	
		0.000	0.000			0.000		0.000	
	20 quarter		13.325						
			0.000						

Note: Each cell in the Table has two entries. The first entry is the likelihood ratio test statistic for the restriction of (1, −1) on the cointegrating vector. The entry underneath is its p-value. The test is only applied to the cointegration cases present in Table 3. The notation used in the table is ECM: error correction specification; FD: first-difference specification; PPP: purchasing power parity; SPMM: sticky-price monetary model; BEER: behavioral equilibrium exchange rate model; IRP: interest rate parity model; RID: real interest differential model; TRF: Taylor rule fundamentals; SPMA: sticky-price monetary augmented model; YCS: yield curve slope model. The forecasting horizons (in quarters) are listed under the heading "Horizon."

Table 4c
Results of the (1, −1) restriction test: dollar-based exchange rates. *Period III: 2007q4-2014q4.*

Specification	Horizon	PPP	SPMM	BEER	IRP	RID	TRF	SPMA	YCS
Panel A: CAN$/$									
ECM	1 quarter		13.318				14.330		
			0.000				0.000		
	4 quarter								
	20 quarter				3.621	4.723	1.454	13.163	9.883
					0.057	0.030	0.228	0.000	0.002
FD	1 quarter	26.424	19.233	14.603		28.359		24.203	
		0.000	0.000	0.000		0.000		0.000	
	4 quarter			12.821					
				0.000					
	20 quarter	6.509	4.227	4.139		5.549		12.730	
		0.011	0.040	0.042		0.018		0.000	
Panel B: Yen/$									
ECM	1 quarter	0.061	9.839		29.138	15.213		25.960	23.109
		0.805	0.002		0.000	0.000		0.000	0.000
	4 quarter								
	20 quarter		9.690						
			0.002						
FD	1 quarter	21.843	32.387	16.226		30.679		22.319	
		0.000	0.000	0.000		0.000		0.000	
	4 quarter								
	20 quarter								
Panel C: SF/$									
ECM	1 quarter		32.357	6.864	16.771	14.446	26.508	12.192	
			0.000	0.009	0.000	0.000	0.000	0.000	
	4 quarter								
	20 quarter		13.078			8.453	19.904	11.925	
			0.000			0.004	0.000	0.001	
FD	1 quarter	25.108	27.850	20.559		9.336		38.181	
		0.000	0.000	0.000		0.002		0.000	
	4 quarter								
	20 quarter	12.353		19.139		8.141		0.399	
		0.000		0.000		0.004		0.528	
Panel D: BP/$									
ECM	1 quarter	8.470		23.523	10.964	21.031		22.203	11.507
		0.004		0.000	0.001	0.000		0.000	0.001
	4 quarter	12.953	12.987	12.101	14.455	12.285	11.428	12.896	12.877
		0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000
	20 quarter	23.073	5.274	7.892	9.479	7.597	13.101	9.782	8.159
		0.000	0.022	0.005	0.002	0.006	0.000	0.002	0.004
FD	1 quarter	26.945	15.746	11.226		35.028		38.586	
		0.000	0.000	0.001		0.000		0.000	
	4 quarter	13.049	22.891	4.110		12.422		22.786	
		0.000	0.000	0.043		0.000		0.000	
	20 quarter	8.142	14.002	8.312		9.773		22.137	
		0.004	0.000	0.004		0.002		0.000	
Panel E: EU/$									
ECM	1 quarter		6.076	8.654	1.205	8.971	10.031		
			0.014	0.003	0.272	0.003	0.002		
	4 quarter					9.241		11.945	9.982
						0.002		0.001	0.002
	20 quarter					13.164	10.728	8.591	
						0.000	0.001	0.003	
FD	1 quarter	19.218	18.193	11.159		28.555		20.088	
		0.000	0.000	0.001		0.000		0.000	
	4 quarter								
	20 quarter			0.258		21.621			
				0.611		0.000			

Note: Each cell in the Table has two entries. The first entry is the likelihood ratio test statistic for the restriction of (1, −1) on the cointegrating vector. The entry underneath is its p-value. The test is only applied to the cointegration cases present in Table 3. The notation used in the table is ECM: error correction specification; FD: first-difference specification; PPP: purchasing power parity; SPMM: sticky-price monetary model; BEER: behavioral equilibrium exchange rate model; IRP: interest rate parity model; RID: real interest differential model; TRF: Taylor rule fundamentals; SPMA: sticky-price monetary augmented model; YCS: yield curve slope model. The forecasting horizons (in quarters) are listed under the heading "Horizon."

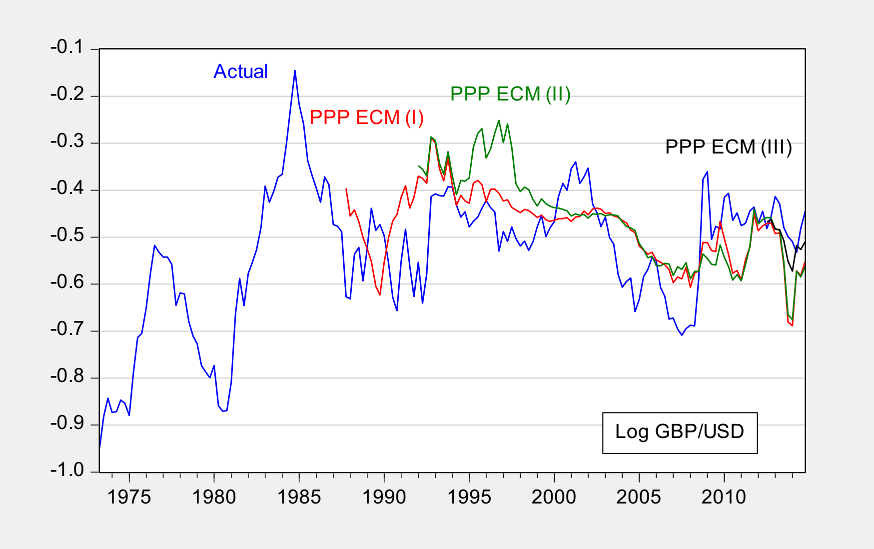

Fig. 7. GBP/USD exchange rate and 20 quarter ahead PPP ECM forecasts.

Expanding the set of criteria does yield some interesting surprises. In particular, the direction of change statistics indicate more evidence that structural models can outperform a random walk. However, the basic conclusion that no economic model is consistently more successful than the others remain intact, with the possible exception of relative purchasing power parity, couched in an error correction framework.

Even if we cannot glean from this analysis a consistent "winner", it may still be of interest to note the best and worst performing combinations of model/specification/currency. The best performance on the MSE criterion is turned in by the purchasing power parity model at the 20-quarter horizon for the British pound exchange rate (post-2007), with a MSE ratio of 0.04 (p-value of 0.003); other PPP forecasts for the other periods follow close behind. Fig. 7 plots the actual British pound exchange rate, and the 20 quarter ahead forecasts for the three periods. The graph shows that forecast performance of the parity model varies across time, but the forecasts track the actual exchange rate movements pretty well during 1985–1990 and 1993–1997.

The worst performances are associated with first-difference specifications; in this case the highest MSE ratio is for the first differences specification of the behavioral equilibrium exchange rate model at the 20-quarter horizon for the Swiss franc exchange rate for Period II. This outcome is partly due to the short sample of data used to estimate the model, so it's probably not the most relevant case to examine.

Perhaps more relevant is the sticky price monetary model augmented with the VIX and TED spread, in first differences, with a MSE ratio of 3.5. To graphically illustrate the failure, we graphed forecasts together with the actual exchange rate, in Fig. 8. Interestingly, the 20 step ahead forecast from the error correction model version of this economic model significantly outperforms a random walk. One might think it's a matter of the levels, but the ECM version of the unaugmented sticky price model does poorly as well, with ratio of 2.0.

Whether this divergence in results arising from inclusion of the VIX and Ted spread would obtain in a sample extending forward in time is an interesting question; the most recent ten years has been remarkable for its unique events involving risk, volatility and liquidity conditions, and it is exactly during this period one expects the variables to be helpful. In fact, neither this model nor the real interest rate differential (intended to capture some features of the data in post-crisis period) perform particularly well over period III.

This pattern of results is not atypical. The superior performance of a particular model/specification/currency combination does not typically carry over from one out-of-sample period to the other, nor from one specification to the other.[29]

It is known that the use of the Diebold-Mariano statistic may yield a conservation test against the random walk specification. To enhance the test power, Clark and West (2006) proposes an adjusted mean squared prediction error (MSPE) statistic. For brevity, we present in Appendix C a description of the Clark-West statistic, and the forecast comparison results based on this statistic in Appendix D in a layout similar to that of Table 1. As expected, the use of the Clark-West statistic increases the instances in which an exchange rate model outperforms the random walk specification. Nevertheless, these results do not change the basic observation that the performance of an exchange rate model varies across specifications, currencies, periods, forecasting horizon, and evaluation criteria.

[29] This pattern of results is consistent with De Grauwe and Grimaldi (2006), Spronk, et al. (2013), Bacchetta and van Wincoop (2013), as well as survey based results in Cheung and Chinn (2001).

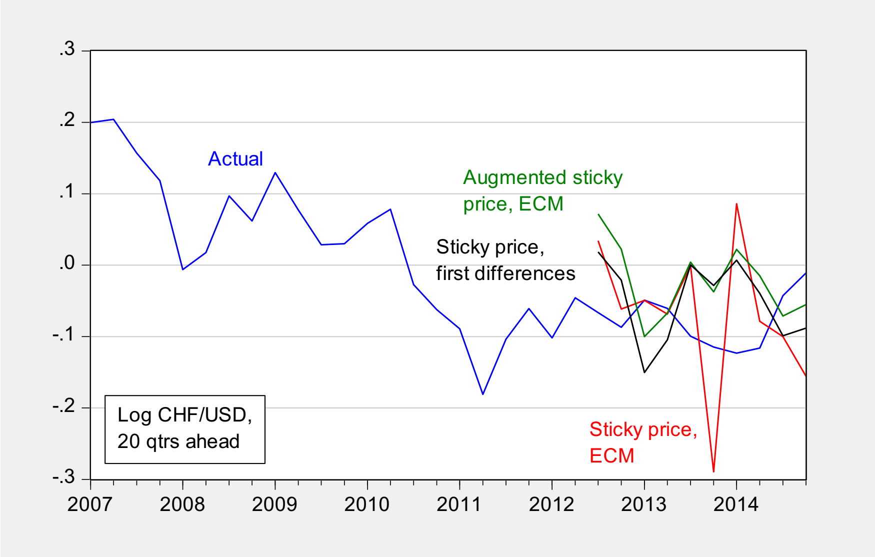

Fig. 8. CHF/USD exchange rate and 20 quarter ahead forecasts for Period III.

Finally, some additional insights can be gleaned by looking across all sample periods, and disregarding statistical significance. At the longest horizon of five years, where difficult to model transitory shocks are least likely to be important, interest rate parity does best in terms of the MSE criterion. Moreover, if one were to ask what model performs the best in terms of the most outperformances relative to a random walk, or exhibits the lowest MSE, it would be purchasing power parity.[30]

5. Concluding remarks

This paper has systematically assessed the predictive capabilities of models, including several developed over the last decade. These models have been compared along a number of dimensions, including econometric specification, currencies, out-of-sample prediction periods, and differing metrics.

In summarizing the evidence from this exhaustive analysis, we conclude that the models that have become popular in last fifteen years or so might not be much better than the older ones. Overall, the average results from all the models are not very successful, on either the MSE or consistency criteria. On the other hand, many models seem to do well, particularly using the direction of change criterion. Some model/specification combinations, however, can deliver superior forecast results at specific forecast horizon and under certain performance metrics. Of the economic models, for example, the error correction specification of purchasing power parity and first difference specification of BEER model outperform the random walk benchmark by a wide margin, using the MSE and direction-of-change criteria, respectively.

In the most recent period, accounting for risk and liquidity tends to improve the fit of the workhorse sticky price monetary model, even if the predictive power is still unimpressive. But in general the more recent models do not consistently outperform older ones, even when assessed on the recent, post-crisis period. Overarching these results, specifications incorporating long run (cointegrating) relationships tend to outperform first differences specifications, particularly along the MSE dimension.

Overall, a specific model/specification/currency combination may performance well in some periods under a performance metric, it will not necessarily work well in another period with an alternative performance metric.

Acknowledgements

We thank an anonymous referee, Luca Dedola, Michael Ehrmann, Philipp Hartmann, Nelson Mark, Barbara Rossi, Michele Ca' Zorzi, Kenneth West, James Yetman, and seminar participants at the ECB, the 2017 IEFS/AEA International Macro and Finance session, and the JIMF-City University HK-BIS Conference for very helpful comments. Cheung gratefully thanks The Hung Hing Ying and Leung Hau Ling Charitable Foundation for its support. Chinn and Zhang acknowledge the financial support of research funds of the University of Wisconsin. Part of this paper was written while Chinn was Wim Duisenberg Fellow at the ECB.

[30] The characterization of outperformance excludes the EUR exchange rates.

Appendix A. Data

Unless otherwise stated, we use seasonally-adjusted quarterly data from the *IMF International Financial Statistics* ranging from the second quarter of 1973 to the last quarter of 2014.

- The exchange rate data are end of period exchange rates.
- The output data are industrial production.
- Money is M2.
- Consumer price indices are used to calculate annual inflation, and along with the producer price index used to calculate the relative price of nontradables.
- Interest rates used in the monetary models are three month Treasury rates. Interest rates used in real interest differential model are overnight rates; shadow rates for US, UK, Euro area are from Wu and Xia, for Japan from IMF *Global Financial Stability Report* (2015), and Ichiue and Ueno (2006, 2007).
- The three-month, annual and five-year interest rates are end-of-period constant maturity interest rates, and are obtained from the IMF country desks, updated from Bloomberg. See Chinn and Meredith (2004), Chinn and Quayyum (2012) for details. Five year interest rate data were unavailable for Japan and Switzerland; hence data from Global Financial Data http://www.globalfindata.com/ were used, specifically, 5-year government note yields for Switzerland and 5-year discounted bonds for Japan.
- The net foreign asset (NFA) series is computed as follows. Using stock data for year 1995 on NFA (Lane and Milesi-Ferretti, 2001), and flow quarterly data from the IFS statistics on the current account, we generated quarterly stocks for the NFA series.
- To generate quarterly government debt data we follow a similar strategy. We use annual debt data from the *IFS* statistics, combined with quarterly government deficit (surplus) data. The data source for Canadian government debt is the Bank of Canada. For the UK, the IFS data are updated with government debt data from the public sector accounts of the UK Statistical Office. Data for Switzerland and Japan are from the BIS.

Appendix B. Diebold-Mariano-West statistics

The Diebold-Mariano-West statistics (Diebold and Mariano, 1995; West, 1996) are used to evaluate the forecast performance of the different model specifications relative to that of the *naive* random walk.

Given the exchange rate series x_t and the forecast series y_t, the loss function L for the mean square error is defined as:

$$L(y_t) = (y_t - x_t)^2. \tag{A1}$$

Testing whether the performance of the forecast series is different from that of the naive random walk forecast z_t, it is equivalent to testing whether the population mean of the loss differential series d_t is zero. The loss differential is defined as

$$d_t = L(y_t) - L(z_t). \tag{A2}$$

Under the assumptions of covariance stationarity and short-memory for d_t, the large-sample statistic for the null of equal forecast performance is distributed as a standard normal, and can be expressed as

$$\frac{\bar{d}}{\sqrt{\frac{1}{T^2}\sum_{\tau=-(T-1)}^{(T-1)} l(\tau/S(T)) \sum_{t=|\tau|+1}^{T} (d_t - \bar{d})(d_{t-|\tau|} - \bar{d})}}, \tag{A3}$$

where $l(\tau/S(T))$ is the lag window, $S(T)$ is the truncation lag, and T is the number of observations. Different lag-window specifications can be applied, such as the Barlett or the quadratic spectral kernels, in combination with a data-dependent lag-selection procedure (Andrews, 1991).

For the direction of change statistic, the loss differential series is defined as follows: d_t takes a value of one if the forecast series correctly predicts the direction of change, otherwise it will take a value of zero. Hence, a value of $\bar{d}$ significantly larger than 0.5 indicates that the forecast has the ability to predict the direction of change; on the other hand, if the statistic is significantly less than 0.5, the forecast tends to give the wrong direction of change. In large samples, the studentized version of the test statistic,

$$\frac{\bar{d} - 0.5}{\sqrt{0.25/T}}, \tag{A4}$$

is distributed as a standard Normal.

Appendix C. Clark-West statistic

To evaluate the forecasting accuracy of the different structural models, we use the adjusted mean squared prediction error (MSPE) statistic proposed by Clark and West (2006). Under the null hypothesis, the MSPE of a zero mean process is the same as the MSPE of the linear alternative. Despite the equality, one expects the alternative model's sample to be larger than the null's. To adjust for the downward bias, Clark and West propose a procedure that performs well in simulations.

We calculate the test statistic as the difference between the MSPE of the random walk model and the MSPE from the linear alternative, which is then adjusted downward to account for the spurious in-sample fit. Under the first model, the process is a zero mean martingale difference process; under the second model, the process is linear,

Model 1: $y_{t+1} = e_{t+1}$
Model 2: $y_{t+1} = X_{t+1}B + e_{t+1}$, $E(e_{t+1}|I_t) = 0$

Our inferences are based on a formal test for the null hypothesis of no difference in the accuracy, as measured by the MSPE, of the two competing forecasts, the linear (structural) model and the driftless random walk. Thus, the hypothesis test is

$H_0 : \sigma_1^2 - \sigma_2^2 = 0$
$H_A : \sigma_1^2 - \sigma_2^2 > 0$

A value larger (smaller) than zero indicates that the linear model (random walk) outperforms the random walk (linear model). The difference between the two MSPEs is asymptotically normally distributed. For forecast horizons beyond one period, one needs to account for the autocorrelation induced by the rolling regression. We use Clark and West's proposed estimator for the asymptotic variance of the adjusted mean between the two MSPEs, which is robust to the serial correlation.

Appendix D. Forecast comparison based on Clark-West statistics

Table A4a
Period I: 1983q1-2014q4.

Specification	Horizon	PPP	SPMM	BEER	IRP	RID	TRF	SPMA	YCS
Panel A: CAN$/$									
ECM	1 quarter	3.231	1.470	1.517		1.848	2.178		
		0.001	0.071	0.065		0.032	0.015		
	4 quarter	3.837	2.545	1.702		1.922	3.516		
		0.000	0.005	0.044		0.027	0.000		
	20 quarter	4.519	3.440	2.394		3.847	2.802		
		0.000	0.000	0.008		0.000	0.003		
FD	1 quarter	0.721	1.350	4.497		−0.790			
		0.236	0.089	0.000		0.785			
	4 quarter	−1.031	0.940	2.045		0.073			
		0.849	0.174	0.020		0.471			
	20 quarter	−2.019	−0.829	0.850		−1.371			
		0.978	0.796	0.198		0.915			
Panel B: Yen/$									
ECM	1 quarter	2.947	3.074			2.231	3.174		4.315
		0.002	0.001			0.013	0.001		0.000
	4 quarter	3.828	4.148			5.185	4.451		7.943
		0.000	0.000			0.000	0.000		0.000
	20 quarter	6.355	6.337			5.915	5.741		5.952
		0.000	0.000			0.000	0.000		0.000
FD	1 quarter	−0.209	0.818			0.185			
		0.583	0.207			0.426			
	4 quarter	0.236	1.061			0.565			
		0.407	0.144			0.286			
	20 quarter	0.408	1.204			0.474			
		0.342	0.114			0.318			

Table A4a (*continued*)

Specification	Horizon	PPP	SPMM	BEER	IRP	RID	TRF	SPMA	YCS
Panel C: SF/$									
ECM	1 quarter	3.211	1.752			2.058	3.084		
		0.001	0.040			0.020	0.001		
	4 quarter	4.356	3.144			3.248	3.397		
		0.000	0.001			0.001	0.000		
	20 quarter	7.033	8.830			7.277	4.864		
		0.000	0.000			0.000	0.000		
FD	1 quarter	−0.525	−0.599			−0.601			
		0.700	0.725			0.726			
	4 quarter	−0.629	−0.518			−0.324			
		0.735	0.698			0.627			
	20 quarter	−1.304	1.262			−0.104			
		0.904	0.103			0.541			
Panel D: BP/$									
ECM	1 quarter	2.990	2.270	1.221		2.158	2.761		2.193
		0.001	0.012	0.111		0.015	0.003		0.014
	4 quarter	5.607	2.346	−0.744		0.413	2.923		1.603
		0.000	0.009	0.771		0.340	0.002		0.054
	20 quarter	7.719	−0.025	−2.775		−4.028	−1.171		0.242
		0.000	0.510	0.997		1.000	0.879		0.404
FD	1 quarter	0.313	−1.491	3.890		−1.053			
		0.377	0.932	0.000		0.854			
	4 quarter	−1.137	−0.277	1.624		−1.118			
		0.872	0.609	0.052		0.868			
	20 quarter	−0.984	−1.279	−0.840		−0.923			
		0.837	0.900	0.799		0.822			

Note: Each cell in the Table has two entries. The first one is the Clark-West adjusted MSPE statistic, and the second is the one-sided p-value; rejection implies the forecast performance of the random walk model is "worse" than the competing specification. The notation used in the table is ECM: error correction specification; FD: first-difference specification; PPP: purchasing power parity; SPMM: sticky-price monetary model; BEER: behavioral equilibrium exchange rate model; IRP: interest rate parity model; RID: real interest differential model; TRF: Taylor rule fundamentals; SPMA: sticky-price monetary augmented model; YCS: yield curve slope model. The forecasting horizons (in quarters) are listed under the heading "Horizon."

Table A4b
Period II: 2001q1-2014q4.

Specification	Horizon	PPP	SPMM	BEER	IRP	RID	TRF	SPMA	YCS
Panel A: CAN$/$									
ECM	1 quarter	2.264	0.911	2.280		0.921	1.676	2.130	1.279
		0.012	0.181	0.011		0.179	0.047	0.017	0.100
	4 quarter	2.671	−0.771	1.904		−1.856	2.666	1.631	0.231
		0.004	0.780	0.028		0.968	0.004	0.051	0.409
	20 quarter	3.397	−1.766	1.619		−2.363	−1.894	4.054	6.050
		0.000	0.961	0.053		0.991	0.971	0.000	0.000
FD	1 quarter	1.068	−0.211	3.969		−1.971		1.711	
		0.143	0.584	0.000		0.976		0.044	
	4 quarter	−1.396	−0.665	1.609		−1.632		1.487	
		0.919	0.747	0.054		0.949		0.069	
	20 quarter	−2.432	−3.799	1.399		−3.096		−0.496	
		0.993	1.000	0.081		0.999		0.690	
Panel B: Yen/$									
ECM	1 quarter	0.270	2.151	1.246		0.119	0.386	2.072	1.147
		0.394	0.016	0.106		0.452	0.350	0.019	0.126
	4 quarter	1.457	1.227	3.292		0.912	0.505	0.084	2.119
		0.073	0.110	0.000		0.181	0.307	0.467	0.017
	20 quarter	4.567	4.386	4.410		4.370	4.469	2.795	6.002
		0.000	0.000	0.000		0.000	0.000	0.003	0.000
FD	1 quarter	−0.535	−1.492	3.000		−0.038		0.954	
		0.704	0.932	0.001		0.515		0.170	
	4 quarter	−0.190	−1.030	1.434		0.572		1.007	
		0.575	0.849	0.076		0.284		0.157	
	20 quarter	0.395	0.700	0.387		0.395		0.284	
		0.346	0.242	0.349		0.346		0.388	

(*continued on next page*)

Table A4b (*continued*)

Specification	Horizon	PPP	SPMM	BEER	IRP	RID	TRF	SPMA	YCS
Panel C: SF/$									
ECM	1 quarter	1.660	1.477	1.865		0.290	1.455	1.675	0.587
		0.048	0.070	0.031		0.386	0.073	0.047	0.279
	4 quarter	2.814	3.324	3.000		1.816	3.576	1.901	0.637
		0.002	0.000	0.001		0.035	0.000	0.029	0.262
	20 quarter	3.918	6.900	11.825		11.902	6.518	5.250	4.967
		0.000	0.000	0.000		0.000	0.000	0.000	0.000
FD	1 quarter	−0.306	1.353	−0.391		1.667		0.455	
		0.620	0.088	0.652		0.048		0.325	
	4 quarter	0.786	1.182	−0.823		1.758		−0.151	
		0.216	0.119	0.795		0.039		0.560	
	20 quarter	0.898	0.726	−0.854		1.289		−0.958	
		0.185	0.234	0.803		0.099		0.831	
Panel D: BP/$									
ECM	1 quarter	2.082	−0.108	1.082		0.929	0.575	0.821	−0.854
		0.019	0.543	0.140		0.176	0.283	0.206	0.803
	4 quarter	2.925	0.624	−0.148		−1.377	1.284	1.281	−1.406
		0.002	0.266	0.559		0.916	0.099	0.100	0.920
	20 quarter	7.684	2.137	−2.762		−2.101	2.345	0.861	−0.666
		0.000	0.016	0.997		0.982	0.010	0.194	0.747
FD	1 quarter	1.012	0.421	2.651		−0.645		−0.180	
		0.156	0.337	0.004		0.740		0.571	
	4 quarter	0.536	0.054	1.091		−1.098		−2.164	
		0.296	0.478	0.138		0.864		0.985	
	20 quarter	−0.396	−2.588	−0.920		−0.592		−0.414	
		0.654	0.995	0.821		0.723		0.660	

Note: Each cell in the Table has two entries. The first one is the Clark-West adjusted MSPE statistic, and the second is the one-sided p-value; rejection implies the forecast performance of the random walk model is "worse" than the competing specification. The notation used in the table is ECM: error correction specification; FD: first-difference specification; PPP: purchasing power parity; SPMM: sticky-price monetary model; BEER: behavioral equilibrium exchange rate model; IRP: interest rate parity model; RID: real interest differential model; TRF: Taylor rule fundamentals; SPMA: sticky-price monetary augmented model; YCS: yield curve slope model. The forecasting horizons (in quarters) are listed under the heading "Horizon."

Table A4c
Period III: 2007q4-2014q4.

Specification	Horizon	PPP	SPMM	BEER	IRP	RID	TRF	SPMA	YCS
Panel A: CAN$/$									
ECM	1 quarter	1.409	1.179	1.099		1.017	1.673	0.626	0.679
		0.079	0.119	0.136		0.154	0.047	0.266	0.248
	4 quarter	2.515	0.857	0.488		0.457	3.293	1.290	−0.078
		0.006	0.196	0.313		0.324	0.000	0.099	0.531
	20 quarter	1.707	0.522	1.084		0.022	0.750	1.272	0.994
		0.044	0.301	0.139		0.491	0.227	0.102	0.160
FD	1 quarter	1.221	−0.120	2.555		−1.431		1.994	
		0.111	0.548	0.005		0.924		0.023	
	4 quarter	−0.715	−0.245	−0.128		−1.362		2.716	
		0.763	0.597	0.551		0.913		0.003	
	20 quarter								
Panel B: Yen/$									
ECM	1 quarter	0.208	1.204	1.049		0.543	−0.643	3.184	0.441
		0.418	0.114	0.147		0.294	0.740	0.001	0.330
	4 quarter	1.059	0.410	1.830		0.976	−0.517	4.494	0.772
		0.145	0.341	0.034		0.164	0.697	0.000	0.220
	20 quarter	1.184	0.269	0.579		0.736	0.981	1.165	0.427
		0.118	0.394	0.281		0.231	0.163	0.122	0.335
FD	1 quarter	−0.185	−1.211	3.053		0.133		1.104	
		0.573	0.887	0.001		0.447		0.135	
	4 quarter	0.173	−0.185	3.427		0.770		0.580	
		0.432	0.573	0.000		0.221		0.281	
	20 quarter								

Table A4c (*continued*)

Specification	Horizon	PPP	SPMM	BEER	IRP	RID	TRF	SPMA	YCS
Panel C: SF/$									
ECM	1 quarter	1.091	0.212	1.076		0.319	1.017	0.618	0.646
		0.138	0.416	0.141		0.375	0.155	0.268	0.259
	4 quarter	0.820	1.729	1.300		0.466	1.785	1.020	0.780
		0.206	0.042	0.097		0.321	0.037	0.154	0.218
	20 quarter	−2.064	3.843	5.455		5.989	4.587	5.803	5.743
		0.981	0.000	0.000		0.000	0.000	0.000	0.000
FD	1 quarter	0.127	1.062	0.877		1.034		−0.442	
		0.449	0.144	0.190		0.151		0.671	
	4 quarter	−0.079	0.534	0.423		0.759		0.232	
		0.531	0.297	0.336		0.224		0.408	
	20 quarter								
Panel D: BP/$									
ECM	1 quarter	2.255	1.240	1.268		1.221	1.628	0.509	0.152
		0.012	0.107	0.102		0.111	0.052	0.306	0.439
	4 quarter	2.823	2.050	0.808		0.162	1.091	0.666	0.922
		0.002	0.020	0.209		0.436	0.138	0.253	0.178
	20 quarter	3.716	1.648	1.367		1.419	2.380	1.416	0.598
		0.000	0.050	0.086		0.078	0.009	0.078	0.275
FD	1 quarter	1.446	0.435	2.284		−1.022		−0.057	
		0.074	0.332	0.011		0.847		0.523	
	4 quarter	0.216	0.866	0.948		0.873		0.407	
		0.414	0.193	0.171		0.191		0.342	
	20 quarter								
Panel E: EU/$									
ECM	1 quarter	1.877	0.090	1.357		−0.246	0.225	0.283	1.099
		0.030	0.464	0.087		0.597	0.411	0.388	0.136
	4 quarter	3.161	−1.170	−0.412		−1.336	−1.037	0.192	−1.289
		0.001	0.879	0.660		0.909	0.850	0.424	0.901
	20 quarter	2.746	−3.873	−3.908		−4.063	−3.816	−3.917	−3.549
		0.003	1.000	1.000		1.000	1.000	1.000	1.000
FD	1 quarter	0.263	−0.859	1.032		−0.500		0.262	
		0.396	0.805	0.151		0.691		0.397	
	4 quarter	−0.954	−0.004	−0.261		−0.158		1.364	
		0.830	0.502	0.603		0.563		0.086	
	20 quarter								

Note: Each cell in the Table has two entries. The first one is the Clark-West adjusted MSPE statistic, and the second is the one-sided p-value; rejection implies the forecast performance of the random walk model is "worse" than the competing specification. The notation used in the table is ECM: error correction specification; FD: first-difference specification; PPP: purchasing power parity; SPMM: sticky-price monetary model; BEER: behavioral equilibrium exchange rate model; IRP: interest rate parity model; RID: real interest differential model; TRF: Taylor rule fundamentals; SPMA: sticky-price monetary augmented model; YCS: yield curve slope model. The forecasting horizons (in quarters) are listed under the heading "Horizon."

References

Andrews, Donald, 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.

Bacchetta, P., Van Wincoop, E., 2013. On the unstable relationship between exchange rates and macroeconomic fundamentals. J. Int. Econ. 91 (1), 18–26.

Binici, Mahir, Cheung, Yin-Wong, 2012. Exchange rate dynamics under alternative optimal interest rate rules. Pacific Basin Finance J. 20, 122–150.

Ca'Zorzi, Michele, Muck, Jakub, Rubaszek, Michal, 2016. Real exchange rate forecasting and PPP: this time the random walk loses. Open Econ. Rev. 27 (3), 1–25.

Chen, Yu-Chin, Tsang, Kwok Ping, 2013. What does the yield curve tell us about exchange rate predictability. Rev. Econ. Stat. 95 (1), 185–205.

Cheung, Yin-Wong, Chinn, Menzie David, 2001. Currency traders and exchange rate dynamics: a survey of the US market. J. Int. Money Finance 20 (4), 439–471.

Cheung, Yin-Wong, Chinn, Menzie, 1998. Integration, cointegration, and the forecast consistency of structural exchange rate models. J. Int. Money Finance 17 (5), 813–830.

Cheung, Yin-Wong, Chinn, Menzie, Pascual, Antonio Garcia, 2005. Empirical exchange rate models of the nineties: are any fit to survive? J. Int. Money Finance 24 (November): 1150–1175. Also NBER Working Paper no. 9393 (December 2002).

Cheung, Yin-Wong, Lai, Kon S., Bergman, Michael, 2004. Dissecting the PPP puzzle: theunconventional roles of nominal exchange rate and price adjustments. J. Int. Econ. 64(1), 135–150.

Chinn, Menzie, Meese, Richard, 1995. Banking on currency forecasts: how predictable is change in money? J. Int. Econ. 38 (1–2), 161–178.

Chinn, Menzie, Meredith, Guy, 2004. Monetary policy and long horizon uncovered interest parity, IMF Staff Papers 51(3) (November), 409–430.

Chinn, Menzie, Quayyum, Saad, 2012. Long Horizon Uncovered Interest Parity, Reassessed, NBER Working Paper No. 18482.

Chinn, Menzie, Zhang, Yi, 2015. Uncovered interest parity and monetary policy near and far from the zero lower bound, NBER Working Paper No. 21159 (May).

Christoffersen, Peter F., Diebold, Francis X., 1998. Cointegration and long-horizon forecasting. J. Business Econ. Statis. 16, 450–458.

Clark, Peter, MacDonald, Ronald, 1999. Exchange rates and economic fundametals: a methodological comparison of beers and feers. In: Stein, J., MacDonald, R. (Eds.), Equilibrium Exchange Rates. Kluwer, Boston, pp. 285–322.

Clark, Todd E., West, Kenneth D., 2006. Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis. J. Economet. 135 (1–2), 155–186.
De Grauwe, P., Grimaldi, M., 2006. Exchange rate puzzles: a tale of switching attractors. Eur. Econ. Rev. 50 (1), 1–33.
Dornbusch, Rudiger, 1976. Expectations and exchange rate dynamics. J. Polit. Econ. 84, 1161–1176.
Diebold, Francis, Mariano, Roberto, 1995. Comparing predictive accuracy. J. Business Econ. Stat. 13, 253–265.
Duy, Timothy, Thoma, Mark, 1998. Modeling and forecasting cointegrated variables: some practical experience. J. Econ. Busin. 50 (3), 291–307.
Eichenbaum, Martin, Johannsen, Benjamin K., Rebelo, Sergio, 2017. Monetary policy and the predictability of nominal exchange rates, mimeo (November).
Engel, Charles, 2014. Exchange Rates and Interest Parity. Handbook of International Economics, volume 4 (Elsevier), 453–522.
Engel, Charles, Hamilton, James, 1990. Long swings in the exchange rate: are THEY in the data and do markets know it? Am. Econ. Rev. 80 (4), 689–713.
Faruqee, Hamid, Isard, Peter, Masson, Paul R., 1999. A Macroeconomic Balance Framework for Estimating Equilibrium Exchange Rates. In: Stein, J., MacDonald, R. (Eds.), Equilibrium Exchange Rates. Kluwer, Boston, pp. 103–134.
Frankel, Jeffrey A., 1979. On the mark: a theory of floating exchange rates based on real interest differentials. Am. Econ. Rev. 69, 610–622.
Giacomini, Raffaella, White, Halbert, 2006. Tests of conditional predictive ability. Econometrica 74 (6), 1545–1578.
Groen, Jan J.J., 2000. The monetary exchange rate model as a long-run phenomenon. J. Int. Econ. 52 (2), 299–320.
Ichiue, Hibiki, Ueno, Yoichi, 2006. Monetary policy and the yield curve at zero interest: the macro-finance model of interest rates as options. Bank of Japan Discussion Paper No. 06-E-16.
Ichiue, Hibiki, Ueno, Yoichi, 2007. Equilibrium interest rate and the yield curve in a low interest rate environment, Bank of Japan Discussion Paper No. 06-E-18.
International Monetary Fund, 2015, Global Financial Stability Report: (IMF, October). http://www.imf.org/External/Pubs/FT/GFSR/2015/02/index.htm.
Kilian, Lutz, Taylor, Mark P., 2003. Why is it so difficult to beat the random walk forecast of exchange rates? J. Int. Econ. 60 (1), 85–107.
Lane, P., Milesi-Ferretti, G.M., 2001. The external wealth of nations: measures of foreign assets and liabilities for industrial and developing. J. Int. Econ. 55, 263–294 http://econserv2.bess.tcd.ie/plane/data.html.
Leitch, Gordon, Ernest Tanner, J., 1991. Economic forecast evaluation: profits versus the conventional error measures. Am. Econ. Rev. 81 (3), 580–590.
MacDonald, Ronald, Marsh, Ian, 1997. On fundamentals and exchange rates: a casselian perspective. Rev. Econ. Statis. 79(4), 655–664.
Mark, Nelson, 1995. Exchange rates and fundamentals: evidence on long horizon predictability. Am. Econ. Rev. 85, 201–218.
Mark, Nelson, Sul, Donggyu, 2001. Nominal exchange rates and monetary fundamentals: evdience from a small post-bretton woods panel. J. Int. Econ. 53 (1), 29–52. February.
Meese, Richard, Rogoff, Kenneth, 1983a. Empirical exchange rate models of the seventies: do they fit out of sample? J. Int. Econ. 14, 3–24.
Meese, Richard, Rogoff, Kenneth, 1983b. "The Out-of-Sample Failure of Exchange Rate Models: Sampling Error or Misspecification? In: Frankel, J. (Ed.), Exchange Rates and International Macroeconomics. Chicago University Press, Chicago, pp. 67–105.
Meese, Richard, Rose, Andrew K., 1991. An empirical assessment of non-linearities in models of exchange rate determination. Rev. Econ 58 (3), 603–619.
Molodtsova, T., Papell, D.H., 2009. Out-of-sample exchange rate predictability with Taylor rule fundamentals. J. Int. Econ. 77, 167–180.
Rossi, Barbara, 2013. Exchange rate predictability. J. Econ. Literat. 51 (4), 1063–1119.
Spronk, R., Verschoor, W.F., Zwinkels, R.C., 2013. Carry trade and foreign exchange rate puzzles. Eur. Econ. Rev. 60, 17–31.
West, Kenneth D., 1996. Asymptotic inference about predictive ability. Econometrica 64 (5), 1067–1084.