Exchange Rate Predictability

Barbara Rossi

February 2013

Barcelona GSE Working Paper Series

Working Paper nº 690

Exchange Rate Predictability

Barbara Rossi*

February 14, 2013

Abstract

The main goal of this article is to provide an answer to the question: "Does anything forecast exchange rates, and if so, which variables?". It is well known that exchange rate fluctuations are very difficult to predict using economic models, and that a random walk forecasts exchange rates better than any economic model (the Meese and Rogoff puzzle). However, the recent literature has identified a series of fundamentals/methodologies that claim to have resolved the puzzle. This article provides a critical review of the recent literature on exchange rate forecasting and illustrates the new methodologies and fundamentals that have been recently proposed in an up-to-date, thorough empirical analysis. Overall, our analysis of the literature and the data suggests that the answer to the question: "Are exchange rates predictable?" is, "It depends" – on the choice of predictor, forecast horizon, sample period, model, and forecast evaluation method. Predictability is most apparent when one or more of the following hold: the predictors are Taylor rule or net foreign assets, the model is linear, and a small number of parameters are estimated. The toughest benchmark is the random walk without drift.

Keywords: Exchange Rates, Forecasting, Instability, Forecast Evaluation

J.E.L. Codes: F3, C5

*Barbara Rossi (ICREA-UPF, Barcelona GSE, CREI). Mailing address: Universitat Pompeu Fabra-CREI, Carrer Ramon Trias Fargas, 25-27, Mercè Rodoreda bldg., 08005 Barcelona SPAIN. Tel.: +34-93-542-1655; e-mail: barbara.rossi@upf.edu

1 Introduction

The objective of this article is to offer a critical survey of the literature on predicting exchange rates in the last ten years. Since Meese and Rogoff (1983a,b, 1988), it has been well known that exchange rates are very difficult to predict using economic models; in particular, a simple, a-theoretical model such as the random walk is frequently found to generate better exchange rate forecasts than economic models. The latter finding is known as "the Meese and Rogoff puzzle". It is important to note that Meese and Rogoff's (1983a,b) finding that the random walk provides the best prediction of exchange rates should not be interpreted as a validation of the efficient market hypothesis. The efficient market hypothesis states that, in the absence of risk premia or when time variation in risk premia tends to be small relative to variation in fundamental pricing factors, bilateral exchange rates are the market's best guess of the relative, fundamental value of two currencies based on all available information at that time. The efficient market hypothesis does not mean that exchange rates are unrelated to economic fundamentals, nor that exchange rates should fluctuate randomly around their past values. Hence the puzzle. However, the recent literature has identified new macroeconomic and financial predictors that claim to forecast exchange rates. The goal of this article is to review both traditional as well as newly proposed exchange rate predictors and evaluate their ability to forecast exchange rates. The main goal is to provide an answer to the questions: "Are exchange rates predictable? And, if so, which predictors are the most useful to forecast exchange rates?".

When trying to answer these questions, a series of complications arise. First, a wide variety of predictors, models, estimation methods, measures of predictive content as well as evaluation tests have been used in the literature. Thus, researchers attempting to forecast exchange rates need to make several choices, such as: Which predictors to use? Which forecast horizon to predict? Which model to estimate? Which data frequency? Which sample? One of the goals of this paper is to provide guidance to researchers on navigating the existing literature as well as to provide a reliable overview of established findings that can be helpful in making these choices. Second, existing papers rely on different predictors, tests, samples or databases; it is possible that such predictors might have lost their forecasting ability, or may not be robust to other databases or samples. In addition, while a predictor might be successful according to a metric/test, it may not be so according to a different one. We therefore perform a thorough empirical evaluation of the success of the predictors identified in the literature using the most recent techniques and databases. Thus, our article

starts with a critical overview of existing predictors and the empirical stylized facts identified in the literature, with particular emphasis on the last ten years. Then, we illustrate the existing empirical evidence using the most up-to-date data and evaluation techniques in order to answer the question: "Does anything forecast exchange rates?".

This article should be of interest for several audiences. Economists and researchers in academia will find that the literature review and the empirical investigation provide guidance to navigate the literature, and will be useful for their work. Practitioners and forecasters at Central Banks and private businesses will also be interested in knowing which predictors, models and methodologies successfully predict exchange rates. Policymakers, for whom successful policy decisions crucially depend on successful forecasts, should also be interested in our assessment on where the literature stands. Finally, newspapers' frequent discussions of exchange rate forecasting suggest this literature review would be useful beyond academia and policy circles.

More in detail, why are exchange rate forecasts useful for Central Banks and policymakers? Wieland and Wolters (2011) provide a detailed review on how forecasts are used in policymaking. Typically, forecasts are used to project the consequences of particular policy measures for policymakers' targets. According to Greenspan (1994, p. 241), "implicit in any monetary policy action or inaction is an expectation of how the future will unfold, that is, a forecast". Wieland and Wolters (2011) provide empirical evidence that Central Bank policies in the US and Europe are described by interest rate rules, where interest rates respond to forecasts of inflation and economic activity, rather than outcomes. Not only economic policy relies on macroeconomic forecasts: the path of the policy may directly affect the forecasts ("projections") of macroeconomic aggregates. In the US, for example, prior to each Federal Open Market Committee meeting,[1] the Federal Reserve staff produces forecasts of several macroeconomic aggregates at horizons up to two years as a basis for their discussions. The variables forecasted by the staff include exchange rates, which influence current account projections as well as US real GDP growth, eventually. As pointed out in Edge et al. (2010), among others, the Federal Reserve staff forecast is derived from data, a variety of models and forecasting techniques, as well as expert judgment. Typically, these forecasts are conditioned on a specific future time path for the federal funds rate, the main instrument of monetary policy. The policy scenarios considered by the Federal Reserve staff may also include dollar depreciation/appreciation scenarios (i.e. scenarios in which the dollar appreciates or depreciates more than in the baseline forecasts, where typically it is assumed

---

[1]The Federal Open Market Committee is the meeting where US monetary policy is decided.

constant, according to the random walk model). Policy decisions are then taken on the basis of what the policymaker deems the most likely scenario. Exchange rate projections are also especially important for Central Banks of countries that are heavy importers/exporters of commodities. For example, one of the models used at the Bank of Canada is Amano and van Norden's (1995), where real exchange rates depend on terms of trade (see Coletti and Murchison, 2002).

At the same time, it should be noted that this review has an empirical, "reduced-form" focus. There are several reasons behind this choice. First, the majority of the empirical work in this area is done with a reduced-form approach; second, while there are theoretical structural models of exchange rate determination, typically they are too stylized to be literally taken to the data and successfully used for forecasting exchange rates. Moreover, fully developed structural models typically do not fit exchange rate data well, not to mention forecast them. Thus, while this article will sketch several "theoretical" models of exchange rate determination, this discussion is mainly provided to motivate the choice of economic predictors that have been considered in the literature. Furthermore, throughout the paper we focus on monthly and quarterly frequencies, as they are the ones of interest to economists; we will not consider very high frequency data analyses that are instead mostly of interest to risk management and finance. Finally, there is a large literature on in-sample estimation of exchange rate models. In-sample fit does not necessarily guarantee out-of-sample forecast success, as we will discuss. Thus, in this overview we will mainly focus on out-of-sample forecasts, although we will provide some discussion of in-sample fit. Note that, typically, real exchange rates are fitted in-sample, while nominal ones are forecasted out-of-sample; therefore, we will focus on the latter.[2]

Overall, our analysis of the literature and the data suggests that the answer to the question: "Are exchange rate predictable?" is, "it depends". In fact, it depends on the choice of predictor, forecast horizon, sample period, model, and forecast evaluation method. Predictability is most apparent when one or more of the following hold: the predictors are Taylor rule and net foreign assets fundamentals, the model is linear, and a small number of parameters are estimated. The toughest benchmark is the random walk without drift. There is some instability over samples for all models, and there is no systematic pattern across models in terms of which horizons or which sample periods the models predict best. Among the negative findings on which the literature has reached a consensus, typically, PPP and monetary models have no success at short (less than 2-3 years) horizons.

---

[2]See Rogoff (1996) for a review of in-sample fit of real exchange rate models.

More in detail, we draw five general conclusions.

First, the degree of success in forecasting exchange rates out-of-sample does depend on the choice of the predictor. Although there is disagreement in the literature, overall the empirical evidence is not favorable to traditional economic predictors (such as interest rates, prices, output and money).[3] Instead, Taylor-rule fundamentals and net foreign asset positions have promising out-of-sample forecasting ability for exchange rates. The consensus in the literature is that the latter fundamentals have more out-of-sample predictive content than traditional fundamentals; the disagreement in the literature is in the degree to which they can resolve the Meese and Rogoff puzzle.

Second, overall, among the model specifications considered in the literature, the most successful are linear ones.[4] Typically, in single-equation linear models, the predictor choice matters more than the model specification itself.[5]

Third, data transformations (such as de-trending, filtering and seasonal adjustment) may substantially affect predictive ability, and may explain differences in results across studies.[6] Another important factor that, for some fundamentals, may affect predictive ability is the use of real-time rather than revised data.[7] For a given model and predictor, predictive ability seems also to depend on the choice of the country. On the other hand, the frequency of the data does not seem to affect predictability.

Fourth, empirical results vary with the benchmark model, the sample period, forecast evaluation method and the forecast horizon. The random walk consistently provides the toughest benchmark. Different models vary in terms of which sample periods and forecast horizons work best, with no apparent overall pattern.

Finally, on the one hand, our empirical analysis confirms several findings in the literature: while several predictors display in-sample predictive ability for future exchange rates, only Taylor-rules display consistently significant out-of-sample forecasting ability at short horizons; and panel monetary models display some forecasting ability at long horizons. Furthermore, our analysis reveals instabilities in the models' forecasting performance: the predictability of fundamentals varies not only across countries, models and predictors, but

---

[3]Except possibly for monetary fundamentals at long horizons and interest rates at short horizons.

[4]Among them, error correction models (either single-equation or panel) are successful at long horizons, although there is disagreement among researchers regarding the degree of robustness of the result.

[5]For example, whether the researcher uses contemporaneous, realized or lagged fundamentals.

[6]For example, predictability of the monetary ECM model is much weaker or completely disappears after estimating the cointegrating parameters.

[7]This is a concern for monetary fundamentals but less of a concern for Taylor-rule fundamentals.

also over time. None of the predictors, models, or tests systematically find empirical support of superior exchange rate forecasting ability across all countries and time periods: when predictability appears, typically it occasionally does so for some countries and for short periods of time. These findings lead to new challenges: why does predictability change over time? Is it possible to design ways to exploit instabilities to improve exchange rates' forecasts? We discuss a few of these challenges in the last section.

There are several literature reviews on exchange rate predictability. How is this literature overview different from the existing ones? Frankel and Rose (1995) review the empirical literature on exchange rates up to 1995, whereas we focus on more recent contributions and include a thorough empirical analysis which includes recent data as well as predictors that have been identified in the last decade. Engel, Mark and West (2007) focus on explaining the fluctuations of exchange rates using selected models, countries and fundamentals; our analysis considers a broader set of fundamentals and more recent data. Melvin, Prins and Shand (2011) focus on forecasting exchange rates from an financial investor's point of view, e.g. carry trades; we focus instead on forecasting exchange rates using economic models and macroeconomic predictors. Lewis (1995) focuses on reviewing several puzzles in international financial markets, especially risk premia and home bias; we focus instead on assessing stylized facts on exchange rate predictability. Finally, Rogoff (1996) and Froot and Rogoff (1995) focus on purchasing power parity (PPP), according to which exchange rate fluctuations reflect fluctuations of countries' relative prices; while we do consider PPP as a predictor, our broader analysis includes several other predictors that have been considered in the literature to forecast exchange rates.[8]

The article is organized as follows. Section 2 provides a simple guiding example. Sections 3 to 6 review the literature: Section 3 discusses the predictors that have been used to forecast nominal exchange rates; Section 4 reviews a variety of models estimated in the literature; Section 5 examines the characteristics of the data used in the literature; and Section 6 overviews the forecast evaluation methods. At the end of each section, we provide a summary of the main findings in the literature. Finally, Section 7 revisits the empirical evidence in an up-to-date, thorough empirical exercise.

---

[8]For shorter reviews on exchange rate predictability that focus on specific topics, see: Neely (1997) on technical trading rules; Bailliu and King (2005) on exchange rate models and the Canadian experience; Neely and Dey (2010) on the effects of macroeconomic news announcements on exchange rates; Chinn (2011) for a survey on macroeconomic models of exchange rate determination; and Neely and Sarno (2002) on the empirical performance of the monetary model.

2 A Guiding Example

Before reviewing in detail the predictors, models, data and forecast evaluation methods proposed in the literature, let us consider a simple example to fix ideas about several basic concepts that we will discuss in this review. Readers that are familiar with basic in-sample fit and out-of-sample forecast methodologies can move directly to the next section.

Let the (log) of the exchange rate be denoted by s_t and let the (univariate) predictor (or fundamental) be denoted by f_t. Examples of predictors used in the literature are discussed in detail in the next section, and the choice of the data is discussed in Section 5.

The relationship between the exchange rate and its fundamental can be described by several models (see Section 4). For expositional purposes, let the model be linear and such that it does not include a constant term:

$$E_t\left(s_{t+h} - s_t\right) = \beta f_t, \quad t = 1, 2, ..., T,$$

where T is the total size of the available sample and h is the forecast horizon.

The model's performance is typically evaluated relative to that of a benchmark model. Let the benchmark model be the random walk without drift:

$$E_t\left(s_{t+h} - s_t\right) = 0.$$

In fact, we will argue in Section 6 that the random walk without drift is the appropriate benchmark for the analysis.

The predictive ability of the fundamental can be evaluated according to in-sample fit or out-of-sample forecast performance. In-sample fit is typically evaluated by estimating β over the full sample,

$$\widehat{\beta}_T = \left(\sum_{t=1}^{T} f_t^2\right)^{-1} \left(\sum_{t=1}^{T} f_t\left(s_{t+h} - s_t\right)\right),$$

and calculating a t-test on β: if the fundamental contains relevant information, then β should be different from zero. The latter is known as an in-sample (traditional) Granger-causality test. If the test rejects, it signals that the predictor contains useful information for explaining exchange rate fluctuations over the full sample. However, this does not necessarily mean that the predictor contains useful information to predict exchange rate fluctuations in real-time. To assess the latter, it is common to turn to forecasting.

To evaluate the models' out-of-sample forecasting ability, the sample is split into two parts: the in-sample portion, consisting of observations from 1 to R, and the out-of-sample

7

portion, of observations $R+h$ to $T+h$, of size $P \equiv T-R+1$. In the rolling window forecasting scheme, the parameter is re-estimated over time using the most recent R observations, where R is known as the estimation window size:

$$\widehat{\beta}_t = \left(\sum_{j=t-R+h+1}^{t} f_{j-h}^2 \right)^{-1} \left(\sum_{j=t-R+h+1}^{t} f_{j-h} \left(s_j - s_{j-h} \right) \right), \ t = R, R+1, ..., T,$$

to obtain a sequence of P h-step-ahead out-of-sample forecast errors, $\varepsilon_{t+h|t}^f \equiv s_{t+h} - s_t - \widehat{\beta}_t f_t$, $t = R, R+1, ..., T$. Note that the random walk forecast error is simply $\varepsilon_{t+h|t}^{rw} \equiv s_{t+h} - s_t$.

Under the rolling window forecast scheme, the model parameters are re-estimated progressively over time. An alternative forecast scheme is the recursive one, where the model parameters are always re-estimated using all the previous observations. That is, $\widehat{\beta}_t = \left(\sum_{j=h+1}^{t} f_{j-h}^2 \right)^{-1} \left(\sum_{j=h+1}^{t} f_{j-h} \left(s_j - s_{j-h} \right) \right), \ t = R, R+1, ..., T.$

The forecasting ability of the model is measured by a loss function; for example, a common choice is the Root Mean Squared Forecast Error (RMSFE), which will be the objective of our analysis unless otherwise noted:[9]

$$RMSFE_f \equiv \frac{1}{P} \sum_{t=R}^{T} \left(\varepsilon_{t+h|t}^f \right)^2.$$

The model forecasts better than the random walk if $RMSFE_f < RMSFE_{rw} \equiv \frac{1}{P} \sum_{t=R}^{T} \left(\varepsilon_{t+h|t}^{rw} \right)^2$. To judge whether the model forecasts significantly better, one typically tests whether $RMSFE_f - RMSFE_{rw}$ is equal to zero against the alternative that the difference is negative, i.e. using a t-test. Several methods to compute the standard errors and other available test statistics are discussed in Section 6.

3 Which Predictors To Use?

This section reviews several economic predictors of exchange rates that have been used in the literature. It explains why, according to economic theory, they should forecast exchange rates, reviews papers proposing/using them, and summarizes their empirical findings.[10] A convenient summary of the various predictors used in the literature is provided in Tables 1 and 2. Some empirical findings are contradictory: the next sections investigate in detail the reasons behind the differences among the papers.

---

[9]Other loss functions are sometimes used, such as the Mean Absolute Error, $MAE_f \equiv \frac{1}{P} \sum_{t=R}^{T} \left| \varepsilon_{t+h|t}^f \right|$.

[10]For a more detailed introduction to these models, see Obstfeld and Rogoff (1996) and Mark (2001).

3.1 Traditional Predictors

The traditional predictors used in the literature include interest rates, prices, money and output differentials.

(a) Interest Rate Differentials

Uncovered interest rate parity (UIRP) dates back to Fisher (1896) – see Dimand (1999). Fisher (1896) provided a general analysis of how interest rates can be related to expected changes in the relative value of units of account or commodities: one of the examples he considered concerned international currencies, and has become known as UIRP. UIRP states that, in a world of perfect foresight[11] with a nominal bilateral exchange rate S_t, investors can buy $1/S_t$ units of foreign bonds using one unit of the home currency, where S_t denotes the price of foreign currency in terms of home currency. Let the foreign bond pay one unit plus the foreign interest rate between time t and $t + h$, i^*_{t+h}. At the end of the period, the foreign return can be converted back in home currency and equals $S_{t+h}\left[\left(1 + i^*_{t+h}\right)/S_t\right]$ in expectation. By arbitrage and in the absence of transaction costs, this return must be in expectation equal to the return of the home bond, $1 + i_{t+h}$. That is, $\left(1 + i^*_{t+h}\right) E_t\left(S_{t+h}/S_t\right) = 1 + i_{t+h}$, where $E_t\left(.\right)$ denotes the expectation at time t. Finally, by taking logarithms and ignoring Jensen's inequality, the previous UIRP equation can be rewritten as:

$$E_t\left(s_{t+h} - s_t\right) = \alpha + \beta\left(i_{t+h} - i^*_{t+h}\right), \tag{1}$$

where $s_t \equiv \ln\left(S_t\right)$, $\alpha = 0$ and $\beta = 1$, and h is the horizon. Similarly, covered interest rate parity (CIRP) predicts exchange rates according to: $E_t\left(s_{t+h} - s_t\right) = \alpha + \beta\left(F_t - s_t\right)$, where F_t denotes the h-period ahead forward rate at time t.[12]

The empirical evidence is not favorable to UIRP. Meese and Rogoff (1988) use eq. (1) to forecast real exchange rates out-of-sample using real interest rate differentials, and compare its performance with the random walk, finding that the latter forecasts better. Similarly, Cheung, Chinn and Pascual (2005) and Alquist and Chinn (2008) find that, although for some countries UIRP forecasts better than the random walk at long horizons, its performance is never significantly better. Slightly more positive findings have been reported by Clark and West (2006) at short-horizons, and Molodtsova and Papell (2009) for some countries.[13] In-sample empirical evidence is not favorable to UIRP either. The consensus is that, typically,

---

[11] See Lewis (1995) for a discussion of UIRP without perfect foresight.

[12] CIRP states that the spread between forward and spot exchange rates equals the nominal interest differential between two countries, and was developed by Keynes (1923).

[13] Molodtsova and Papell (2009) estimate UIRP with unrestricted coefficients (both constant and slope) as well as without a constant and with an estimated slope. In the latter case, they only find marginal evidence

estimates of (1) display a negative and significant slope, and a constant significantly different from zero; see Froot and Thaler (1990) for a survey.[14]

(b) Price and Inflation Differentials

According to Purchasing Power Parity (PPP), the real price of comparable commodity baskets in two countries should be the same. That is, the price level in the home country, converted to the currency of the foreign country by the nominal exchange rate, should equal the price level of the foreign country. Thus, a unit of currency in the "home" country will have the same purchasing power in the "foreign" country. The theory, due to Cassel (1918), was developed during the debate on the appropriate value of nominal exchange rates among major industrialized countries after the hyper-inflations in World War I. Let the logarithm of the commodity price (CP) index in the home and foreign countries be denoted by p_t and p_t^*, respectively. Then, PPP implies that:

$$s_t = \alpha + \beta \left(p_t - p_t^* \right) + \varepsilon_t. \tag{2}$$

where $\alpha = 0$ and $\beta = 1$.

The out-of-sample empirical evidence is not favorable to PPP either: Cheung, Chinn and Pascual (2005) find that, although PPP forecasts better than the random walk at the longest horizons, its performance is never significantly better; at shorter horizons, it is significantly worse than the random walk. Whether PPP holds in-sample is also debated. In particular, two stylized facts emerge from Rogoff (1996). First, nominal exchange rates tend toward purchasing power parity in the long run, although the speed of convergence is remarkably slow. Second, short run deviations from PPP are substantially large. As Rogoff (1996) notes, deviations from PPP can be attributed to transitory disturbances in the presence of nominal price stickiness; thus, they should be short-lived (i.e., 1–2 years), while, in the data, half-life deviations from PPP range between three to five years.[15] Rogoff (1996, p. 647) called this empirical inconsistency the "PPP puzzle." Possible concerns and explanations include: underestimation of the uncertainty around point estimates (Cheung and Lai, 2000, Kilian

---

of predictive ability for Australia and Canada; in the former case, they find strong evidence in favor of Japan and Switzerland, as well as marginal evidence in favor of Australia and Canada.

[14] Possible explanations include: the presence of a time-varying risk premium (Fama, 1984; Backus, Foresi and Telmer, 2001); estimation biases (Bekaert and Hodrick, 2001); imprecise standard errors (Baillie and Bollerslev, 2000, Rossi, 2007a); and small samples (Chinn and Meredith, 2004, who find positive evidence in longer samples, and Chen and Tsang, 2011, who pool information from the whole term structure).

[15] The half-life measures how many time periods it takes for the effects of a shock to PPP to decrease by 50%.

and Zha, 2002, Murray and Papell, 2002, Gospodinov 2004, Lopez, Murray, and Papell, 2005, and Rossi, 2005c)[16] and heterogeneity in disaggregate data (Imbs, Mumtaz, Ravn and Rey, 2005).[17]

(c) Money and Output Differentials

According to the monetary model of exchange rate determination, bilateral nominal exchange rate fluctuations should reflect movements in countries' relative money, output, interest rates and prices. The monetary model was first introduced by Frenkel (1976) and Mussa (1976), and builds on a simple small open economy model where real output is exogenous. Real money demand is viewed as a function of income and the interest rate; by using UIRP and PPP to substitute relative interest rates and prices as function of exchange rates, one obtains a relationship between exchange rates, money and output differentials. More in detail, let m_t be the logarithm of nominal money, y_t be the logarithm of real output, and the horizon, h, equals 1. Then, real demand for money is modeled as:

$$m_t - p_t = -\eta i_{t+1} + \phi y_t.$$

Assuming that a similar equation holds for the foreign country ($m_t^* - p_t^* = -\eta i_{t+1}^* + \phi y_t^*$, where for simplicity of notation we assumed that the coefficients are symmetric and asterisks denote foreign country variables) and taking the difference between the two gives the relative money demand equation: $m_t - m_t^* - (p_t - p_t^*) = -\eta \left(i_{t+1} - i_{t+1}^* \right) + \phi \left(y_t - y_t^* \right)$. One approach (valid if prices and exchange rates are completely flexible) is to assume that PPP holds at every point in time, and substitute it in the relative money demand equation to get the *"flexible price version of the monetary model"*:[18]

$$s_t = \eta \left(i_{t+1} - i_{t+1}^* \right) - \phi \left(y_t - y_t^* \right) + \left(m_t - m_t^* \right). \tag{3}$$

In the presence of sticky price adjustment, either the relative price level or inflation differentials are included as regressors to obtain the *"sticky price version of the monetary model"*:

$$s_t = \eta \left(i_{t+1} - i_{t+1}^* \right) - \phi \left(y_t - y_t^* \right) + \left(m_t - m_t^* \right) + \zeta_2 \left(p_t - p_t^* \right); \tag{4}$$

---

[16]On the one hand, higher uncertainty implies that the PPP puzzle is even larger than previously thought; on the other hand, this partially reconciles PPP with models with sticky prices, since confidence intervals also include very short half-lives.

[17]Imbs, Mumtaz, Ravn and Rey (2005) find that, when heterogeneity is taken into account, the estimated persistence of real exchange rates and the half-life falls dramatically. Using panel techniques, Taylor and Sarno (1998) find evidence in favor of PPP in the long-run.

[18]This was also referred to as the Frenkel-Bilson model; the coefficient on money differential is unity, due to first degree homogeneity of relative money supply. See Meese and Rogoff (1983a).

in this case, PPP holds in the long run but does not hold in the short run.[19]

The empirical evidence on the monetary model is mixed. The in-sample evidence is somewhat positive,[20] while the out-of-sample evidence is less positive. On the one hand, Meese and Rogoff (1983a,b) demonstrate that the random walk forecasts exchange rates out-of-sample better than both the monetary models above, eqs. (3) and (4). Their finding has been confirmed by Chinn and Meese (1995) for short horizon (one-month to one-year-ahead) forecasts,[21] by Cheung, Chinn and Pascual (2005), who find that the monetary model does not predict well even at longer horizons (i.e. five years), and by Alquist and Chinn (2008).[22] Molodtsova and Papell (2009) also find very limited empirical evidence in favor of the model.[23] On the other hand, Mark (1995) finds strong and statistically significant evidence in favor of the monetary model at very long horizons (i.e. three to four years). The robustness of Mark's (1995) findings has, however, been questioned by Berkowitz and Giorgianni (2001), Kilian (1999), Groen (1999), Faust et al. (2003) and Rossi (2005a). Overall, the former four papers find less evidence in favor of predictive ability even at long horizons, whereas the latter finds more positive results. The next sections will shed light on the causes of the disagreement across the empirical findings.

(d) Productivity Differentials

More general monetary models that include additional predictors have been considered by several authors. For example, Cheung, Chinn and Pascual (2005) consider a model where PPP does not hold even in the long run; instead, relative prices $(p_t - p_t^*)$ are expressed as a function of productivity differentials z_t following Balassa (1964) and Samuelson (1964):

$$s_t = \eta \left(i_{t+1} - i_{t+1}^* \right) - \phi \left(y_t - y_t^* \right) + \zeta_1 \left(m_t - m_t^* \right) + \zeta_2 z_t + \varepsilon_t. \qquad (5)$$

Note that in the case of (5) the nominal exchange rate may depend on real variables. In some studies, the real price of non-tradeables is included instead of productivity differentials.

---

[19]Sometimes inflation differentials are used instead of prices (Cheung, Chinn and Pascual, 2005) and coefficients left unrestricted.

[20]MacDonald and Taylor (1993), Husted and MacDonald (1998), Groen (2000, 2002), Mark and Sul (2001) find cointegration between exchange rates and monetary fundamentals, while Sarantis (1994) does not. Rossi (2006) rejects that the coefficients of the monetary model are both constant and equal to zero, suggesting time-varying predictive ability.

[21]At longer horizons, Chinn and Meese's (1995) results are slightly more positive, although statistically significant only for the Yen/US dollar exchange rate among the five currencies they consider.

[22]Chinn and Meese (1995) consider eq. (3); Alquist and Chinn (2008) consider the sticky price monetary model (eq. 4) with differentials of money, real GDP, interest and inflation rates.

[23]They find evidence only for two countries among the twelve they consider.

Cheung, Chinn and Pascual (2005) measure productivity differentials by labor productivity indices (real GDP per employee). They find that the model with productivity differentials does not forecast better than the random walk.[24]

(e) Portfolio Balance

Traditional portfolio balance models (Frankel, 1982; Hooper and Morton, 1982) include a measure of stock balances:

$$s_t = \beta_0 + \beta_1 \left(i_t - i_t^* - E_t \left(s_{t+1} - s_t \right) \right) + b_t - b_t^*, \tag{6}$$

where b_t is the stock of home assets held by home and b_t^* is the stock of foreign assets held by home, and the unobservable term $E_t \left(s_{t+1} - s_t \right)$ is approximated by zero. Several measures of balances have been used in the literature as broad proxies: cumulated trade balance differentials, cumulated current account balance differentials, and government debt. Meese and Rogoff (1983a,b) find that, even after augmenting the monetary model by a measure of trade balance differentials, the model still does not forecast better than the random walk, a finding confirmed by Cheung, Chinn and Pascual (2005).[25]

Summary of Empirical Findings

Overall, the empirical evidence based on the traditional predictors is not favorable to the economic models. While the out-of-sample forecasting ability of the economic predictors occasionally outperforms that of a random walk in some studies for some countries/time periods, it definitely does not systematically do so. More importantly, with a few exceptions, their predictive ability is not significantly better than that of a random walk at short horizons. The main exception is the work by Clark and West (2006) regarding the out-of-sample predictive ability of UIRP; the next sections will investigate the reasons why their finding is different from the rest of the literature. At longer horizons, there is more evidence of predictive ability in favor of the monetary model, although the finding is contentious. At the same time, some predictors (i.e. interest rate differentials) show significant in-sample fit, although with coefficient signs that are inconsistent with economic theory.

---

[24] Wright (2008) also includes productivity differentials among his predictors. See Section 4 for a discussion.

[25] Cheung, Chinn and Pascual (2005) estimate a model where exchange rate fluctuations are functions of the government debt relative to GDP, the real interest rate, the net foreign asset position, the (log of the) terms of trade, the (log) price level differential and the relative price of non-tradeables.

3.2 Taylor Rule Fundamentals

Engel and West (2005, 2006) and Molodtsova and Papell (2009) propose fundamentals based on a Taylor rule for monetary policy (Taylor, 1993). Taylor (1993) formalizes the idea that the monetary authority sets the real interest rate as a function of how inflation differs from its target level (the higher the inflation, the more contractionary monetary policy will be) and also as a function of the output gap y_t^{gap} (if output is below its potential, monetary policy will be more expansionary). Taylor (1993) originally proposed the following specification: $r_{t+1} = \phi \left(\pi_t - \pi^\dagger \right) + \gamma y_t^{gap} + r^\dagger$, where π_t is the inflation rate, $\pi^\dagger$ is the target rate of inflation, y_t^{gap} is the output gap, $r_{t+1} \equiv i_{t+1} - \pi_{t+1}$ is the real interest rate (defined as the difference between the nominal interest rate, i_t, and the inflation rate, π_t), and $r^\dagger$ is the equilibrium real interest rate. Clearly, if one considers two economies, both of which set interest rates according to a Taylor rule, by UIRP their bilateral exchange rate will reflect their relative interest rates, and thus, as a consequence, their output gaps and their inflation levels. This basic idea is at the core of the Taylor-rule fundamental model of exchange rates. We will discuss this model in detail following the approach of Molodtsova and Papell (2009).

Molotdsova and Papell (2009) amend the Taylor rule to take into account two empirical facts. First, in an open economy setting, as the Central Bank attempts to maintain the nominal exchange rate at its purchasing power parity level (Svensson, 2000), monetary policy also depends on the real exchange rate, $q_t \equiv s_t - p_t + p_t^*$. Second, interest rate changes are sluggish since Central Banks prefer to avoid over-achieving their target (as in Clarida, Gali and Gertler, 1998). By adding these features to the original Taylor rule, they obtain:

$$i_{t+1} = (1 - \rho) \left(\mu + \lambda \pi_{t+1} + \gamma y_{t+1}^{gap} + \delta q_t \right) + \rho i_t + v_{t+1}$$

for all countries, whereas for the US $\delta = 0$, and v_{t+1} is the monetary policy shock. That is, using asterisks to denote foreign country variables:

$$i_{t+1}^* = (1 - \rho^*) \left(\mu^* + \lambda^* \pi_t^* + \gamma^* y_t^{gap*} + \delta^* q_t \right) + \rho^* i_t^* + v_{t+1}^*$$
$$i_{t+1} = (1 - \rho) \left(\mu + \lambda \pi_t + \gamma y_t^{gap} \right) + \rho i_t + v_{t+1}.$$

By taking the difference of the two equations, using UIRP and re-defining the coefficients, one obtains the specification in Molodtsova and Papell (2009):

$$E_t s_{t+1} - s_t = \widetilde{\mu} + \widetilde{\delta} q_t + \widetilde{\lambda}^* \pi_t^* + \widetilde{\gamma}^* y_t^{gap*} + \widetilde{\lambda} \pi_t + \widetilde{\gamma} y_t^{gap} + \rho i_t - \rho^* i_t^*, \tag{7}$$

which they refer to as the "asymmetric" Taylor rule. They also consider imposing the coefficient to be the same, a specification they refer to as the "symmetric" Taylor rule, and

typically the real exchange rate and the lagged interest rates are not included:

$$E_t s_{t+1} - s_t = \widetilde{\mu} + \widetilde{\lambda}\left(\pi_t - \pi_t^*\right) + \widetilde{\gamma}\left(y_t^{gap} - y_t^{gap*}\right). \tag{8}$$

Both the in-sample and the out-of-sample empirical evidence are mostly favorable to Taylor-rule fundamentals, although with exceptions. Regarding the in-sample evidence, Chinn (2008) estimates the Taylor model in-sample and finds that the coefficient signs are not consistent with theory, and that the choice of the gap measure is not innocuous.[26] Regarding the out-of-sample evidence, Molodtsova and Papell (2009) show that eq. (8) forecasts exchange rates out-of-sample significantly better than the random walk for several countries, although the performance depends on the exact specification.[27] Molodtsova et al. (2010), Giacomini and Rossi (2010) and Inoue and Rossi (2012) also find strong empirical evidence in favor of Taylor-rule fundamentals. On the other hand, Rogoff and Stavrakeva (2008) find that the empirical evidence in favor of Taylor-rule fundamentals is not robust.

Taylor rules are generally deemed to be a good description of monetary policy in the past three decades, but monetary policy may have changed during the recent 2007 financial crisis. Molodtsova and Papell (2012) study exchange rate forecasting during the financial crisis by including indicators of financial stress in the Taylor rule, such as the Libor-OIS/Euribor-OIS differential, the Bloomberg and OECD financial condition indices, and the TED spread differential. Adrian et al. (2011) use instead measures of liquidity such as funding liquidity aggregates of US financial intermediaries measured by stocks of US dollar financial commercial paper and overnight repos. Both the latter papers find positive evidence.[28]

3.3 External Imbalance Measures

Gourinchas and Rey (2007) argue that not only the current account, but the whole dynamic process of net exports, foreign asset holdings and return on the portfolio of net foreign assets are important predictors of exchange rates. When a country experiences a current account imbalance, the traditional intertemporal approach to the current account suggests that the

---

[26]Chinn (2008) argues that this model of exchange rate fluctuations builds on UIRP: if the latter does not hold in the data, it is surprising that the former holds. Engel and West (2006) calibrate the model in-sample for real exchange rates and find a correlation of 0.5 between fitted and realized real exchange rates.

[27]E.g. which output gap measure is used and whether a constant is estimated. Across the various specifications they consider, typically, exchange rates of four to seven countries out of twelve are significantly predictable.

[28]E.g. Adrian et al. (2011) find positive evidence for almost all the advanced countries as well as half of the emerging countries they consider.

country will need to run future trade surpluses to reduce this imbalance. Gourinchas and Rey (2007) argue instead that part of the adjustment can take place through a wealth transfer between that country and the rest of the world occurring via a depreciation of the value of its currency. Thus, they propose "net foreign assets" (NXA) as a potential predictor for future exchange rate fluctuations. NXA is the deviation from trend of a weighted combination of gross assets, gross liabilities, gross exports and gross imports, and measures the approximate percentage increase in exports necessary to restore external balance, that is, to restore the long run equilibrium of net exports and net foreign asset ratios.

The empirical evidence is overall favorable to external imbalance measures. Gourinchas and Rey (2007) and Della Corte, Sarno and Sestieri (2010) find that the net foreign asset model can predict (effective) exchange rates out-of-sample significantly better than the random walk at both long and short horizons.[29] Alquist and Chinn (2008) find that in some sub-sample the net foreign asset model forecasts (bilateral) exchange rates better than the random walk at short horizons for some countries; the results are however less favorable at longer horizons.[30]

3.4   Commodity Prices and Other Predictors

Chen and Rogoff (2003) focus attention on commodity prices as a potential new macroeconomic fundamental for exchange rates. They focus on "commodity currencies", that is exchange rates for countries where primary commodities constitute a significant share of exports (i.e. Australia, Canada and New Zealand). Their main idea is that, typically, exchange rates are endogenously determined in equilibrium together with other macroeconomic variables, so it is difficult to predict exchange rate changes based on reduced-form models. However, if it were possible to identify an exogenous shock to exchange rates, that would cleanly predict exchange rate fluctuations. Chen and Rogoff (2003) argue that commodity price changes act as "essentially exogenous" shocks for small open economies; the economies with a large share of exports in primary commodities will typically experience exchange rate

---

[29]Della Corte, Sarno and Sestieri (2010) use bilateral external imbalance measures, as opposed to the global measure used in Gourinchas and Rey (2007), which summarizes the net foreign assets position of a country vis-à-vis the rest of the world. They construct the measure of bilateral NXA based on Lane and Milesi Ferretti's (2007) database.

[30]Alquist and Chinn (2008) use a proxy for NFA position based on the end-of-year US foreign asset and liability data from the Bureau of Economic Analysis (BEA) and interpolated using US quarterly financial account data from IFS.

appreciations when the price of their commodity exports increases. Measures of commodity prices used in the literature to forecast exchange rate include commodity price indices (as in Chen and Rogoff, 2003, Chen, Rogoff and Rossi, 2010, and Ferraro, Rogoff and Rossi, 2011) and oil prices (as in Bacchetta, van Wincoop and Beutler, 2010, and Rossi and Sekhposyan, 2011). Chen and Rogoff (2003) find in-sample empirical evidence in favor of commodity prices as predictors of exchange rates; Chen, Rogoff and Rossi (2010) find that commodity prices are not significant out-of-sample predictors of exchange rates in quarterly data, and Ferraro, Rogoff and Rossi (2011) find that they are in daily data.[31]

3.5   What Have We Learned?

The literature has considered a wide variety of predictors. Table 1 summarizes the main predictors that have been used for out-of-sample exchange rate forecasting, and Table 2, Panel A, provides an overview of which predictors have been used out-of-sample in several of the papers we discussed. Overall, the empirical evidence is not favorable to traditional economic predictors, except possibly for the monetary model at very long horizons and the UIRP at short horizons, although there is disagreement in the literature. Both Taylor-rule fundamentals and net foreign asset positions have promising out-of-sample forecasting ability for exchange rates, although some papers question the robustness of the results. The consensus in the literature is that these fundamentals have more out-of-sample predictive content than traditional fundamentals; the disagreement in the literature is in the degree to which these new fundamentals can explain the Meese and Rogoff puzzle.

4   Which Models To Choose?

Several models have been used in the literature in the attempt to forecast exchange rates. Models can be divided into three representative groups: single-equation, multiple equations and panel models. The models in each of these groups can be either linear or non-linear, and may or may not allow for cointegration (i.e. by adding an error correction term) or time variation in the parameters. This section provides a guide to the various model specifications

---

[31]Additional predictors that have been considered mainly for in-sample fit of exchange rate models include order flows (defined as the difference between the number of buyer initiated transactions and the number of seller initiated transactions). See Rime, Sarno and Sojli (2010) and Chinn and Moore (2012). As discussed in Andersen et al. (2003, p. 59), order flows have the potential drawback that we remain ignorant about the macroeconomic determinants of high-frequency order flows.

that have been used in the literature, and evaluates both how the choice of the model may affect predictive ability as well as explain some of the differences in findings across papers. Table 2, Panel A, provides a convenient overview.

4.1   Single-equation, Linear Models

Meese and Rogoff (1983a, 1988) focus on models where exchange rate fluctuations are explained by the simple single-equation model:

$$E_t \left(s_{t+h} - s_t \right) = \beta_0 + \beta_1' f_{t+h}, \tag{9}$$

where the future, realized values of the fundamental f_{t+h} are used. We refer to (9) as the *"single-equation, contemporaneous, realized fundamental model"*. The actual, rather than the forecasted, value of the fundamentals is used as a predictor by Meese and Rogoff (1983a,b, 1988) to make sure that the lack of predictability of exchange rates is not due to poor forecasts of the fundamentals. The parameters are estimated either by simple OLS or by GMM (to deal with the endogeneity of the predictors). Meese and Rogoff (1983b) calibrate the parameter in a grid to explore the robustness of their results to possible inconsistencies in the parameter estimates. Cheung, Chinn and Pascual (2005) forecast exchange rates using eq. (9) either by using calibrated parameter values (based on economic theory)[32] or by estimating eq. (9) via OLS (ignoring endogeneity issues). Similarly, Bacchetta, van Wincoop and Beutler (2010) and Ferraro et al. (2011), among others, estimate eq. (9).[33]

To evaluate actual, ex-ante predictability of the fundamental, one might consider the following models. The *"single-equation, contemporaneous, forecasted fundamental"* model is:

$$E_t \left(s_{t+h} - s_t \right) = \beta_0 + \beta_1' E_t f_{t+h} \tag{10}$$

where $E_t f_{t+h}$ is estimated based on information available up to time t, and the endogeneity of the fundamentals requires instrumental variable estimation. Eq. (10) has been considered by Meese and Rogoff (1983a) and Chinn and Meese (1995). In the *"single-equation, lagged fundamental model"*:

$$E_t \left(s_{t+h} - s_t \right) = \beta_0 + \beta_1' f_t, \tag{11}$$

---

[32] For example, for PPP, they forecast future $s_{t+h} - s_t$ simply by using $p_{t+h} - p_{t+h}^*$.

[33] Ferraro et al. (2011) argue that the fundamental that they use (the rate of growth of commodity prices) can be considered essentially exogenous; thus, they can consistently estimate the parameters by OLS.

contemporaneous fundamentals are directly used for forecasting the h-step-ahead rate of growth of exchange rates.[34] Note that OLS can be used to estimate the parameters, since the regressors are lagged. Eq. (11) has been used by Wright (2008), Molodtsova and Papell (2009, 2012), Molodtsova et al. (2010) and Rossi (2005a), among others. A slight modification of eq. (11) is $E_t\left(s_{t+h} - s_t\right)/h = \beta_0 + \beta_1' f_t$, used by Gourinchas and Rey (2007) and Della Corte et al. (2010).

The empirical evidence based on the single-equation linear model is mixed. Typically, contemporaneous, realized fundamentals are not successful.[35] Authors disagree on whether using forecasted fundamentals improves the empirical evidence in favor of the monetary model (Meese and Rogoff, 1983a, find it doesn't and Faust, Rogers and Wright (2003) find that it does). The out-of-sample performance of lagged fundamentals depends on the predictor: it is poor for the monetary model with or without unemployment (see Rossi, 2006, and Rossi and Sekhposyan, 2011), but successful for the Taylor-rule model (Molodtsova and Papell, 2009, 2012, Molodtsova et al., 2010, and Giacomini and Rossi, 2010) and the net foreign asset model (Gourinchas and Rey, 2007, and Della Corte et al., 2010). Ferraro et al. (2011) find predictive ability using realized, contemporaneous commodity prices, though the predictive ability weakens considerably using lagged commodity prices. Overall, we conclude that, with the latter exception, it is the predictor that matters the most in determining the strength of the predictability rather than the exact specification of the single-equation linear model.

4.2   Single-equation, ECM models

Since the work by Mark (1995), a model that has been widely used includes an error correction term. The Error Correction Model (ECM) assumes that there is a long run relationship between the level of the exchange rate, s_t, and the level of the fundamentals, f_t; thus, in forecasting the rate of growth of the exchange rate, it includes a correction term that captures

---

[34]Note that the forecast is based on a parameter directly estimated from a regression of the current rate of growth of the exchange rate on the lagged value of the fundamental. See Marcellino, Stock and Watson (2006) for a review of direct and iterated forecast methods.

[35]See Meese and Rogoff (1983a,b) for the monetary model; Cheung, Chinn and Pascual (2005) for the monetary, productivity and net foreign assets measures; and Bacchetta, van Wincoop and Beutler (2010) for the monetary model augmented by unemployment and oil prices.

the long-run disequilibrium between the exchange rate and the fundamental levels:

$$E_t\left(s_{t+h} - s_t\right) \equiv \sum_{j=1}^{h} \Delta s_{t+j} = \beta_0 + \beta_1 \left(s_t - \gamma' f_t\right) \tag{12}$$

To help intuition, consider the monetary model, eq. (3). Note that by substituting the UIRP and PPP conditions in the relative money demand equation, we have: $m_t - m_t^* - s_t = -\eta\left(E_t s_{t+h} - s_t\right) + \phi\left(y_t - y_t^*\right)$, which leads to the popular approach of Mark (1995), where $E_t s_{t+h} - s_t = \alpha + \beta\left[\left(m_t - m_t^*\right) - \phi\left(y_t - y_t^*\right) - s_t\right]$. As long as $\beta < 1$, exchange rates revert back to their fundamental value $f_t \equiv \left(m_t - m_t^*\right) - \phi\left(y_t - y_t^*\right)$ over time. Thus, predictive ability should be stronger at longer horizons. The cointegration vector parameter (γ) can be calibrated – as in Mark (1995),[36] Chinn and Meese (1995), Abhyankar et al. (2005), Berkowitz and Giorgianni (2001) and Kilian (1999) – or estimated (typically by Stock and Watson's (1993) DOLS), either over the full-sample or recursively – as in Alquist and Chinn (2008), Chinn and Moore (2012) and Cheung, Chinn and Pascual (2005).[37]

Positive evidence in favor of the ECM model at long horizons has been found by Mark (1995), whereas Cheung, Chinn and Pascual (2005) and Alquist and Chinn (2008) find no predictive ability. Note that the former calibrates the cointegration parameters whereas the latter estimate them. On the other hand, using exactly the same ECM specification, Kilian (1999) and Groen (1999) find no predictive ability for the monetary model at long horizons, whereas Rossi (2005a) does. We will investigate in detail the reasons behind the disagreement over the predictive ability of the monetary model at long horizons in Sections 5.1, 6.4 and 6.5.[38]

4.3   Non-linear Models

Most of the literature focuses on linear models. Only a few papers focus on non-linear models, including: non-parametric methods (locally-weighted regressions), as in Diebold and Nason (1990), Meese and Rose (1991), Chinn (1991), Mizrach (1992) and Chinn and Meese (1995); neural networks, as in by Qi and Wu (2003); or (exponential) transition autoregressive models (ESTAR), as in Rapach and Wohar (2006). Again, the literature differs regarding whether actual, realized fundamentals or forecasted fundamentals are used – see Table 2.

---

[36] Mark (1995) calibrates $\gamma \equiv [1, -\phi] = [1, -1]$.

[37] The parameters β_0, β_1 are estimated by OLS.

[38] Amano and van Norden (1993) also use ECMs to predict the Canadian-US dollar real exchange rate using world commodity price indices and interest rates; and Amano and van Norden (1995, 1998a,b) study their cointegration properties.

The empirical evidence is not favorable to non-linear models. The vast majority of the literature finds that non-linear models forecast poorly,[39] with some exceptions.[40] This finding is consistent with the more general finding that non-linear models fit well in-sample,[41] but fail in out-of-sample forecasting exercises (Terasvirta, 2006).

4.4   Time-Varying Parameter (TVP) Models

A special form of non-linearity may be induced by time-variation in the parameters. There are several ways to deal with time variation in the parameters: either estimate a parametric model where the parameters change over time according to a rule (as in the Kalman filter approach by Wolff, 1987, and Schinasi and Swami, 1987; the Bayesian TVP model in Canova, 1993; Stock and Watson's (1998) random walk coefficient model; or the Markov Switching model of Engel and Hamilton (1990) and Engel (1994)) or take averages across models, either via a Bayesian Model Averaging (BMA) approach (Wright, 2008) or forecast combinations (Timmermann, 2006).[42]

The *"single-equation time-varying parameter"* model used in Schinasi and Swami (1987) and Wolff (1987) is:

$$s_t = \beta_t' f_t + u_t, \tag{13}$$
$$\beta_t = G\beta_{t-1} + K + Av_t,$$

where the parameter β_t changes over time according to an autoregressive process, K and A are constants, and u_t and v_t are unforecastable shocks.[43] The *"random walk coefficient time-varying parameter"* model (Stock and Watson, 1998) imposes $G = I$ and $K = 0$.

---

[39] Including Chinn (1991) and Chinn and Meese (1995) for the monetary model; Diebold and Nason (1990) for univariate models; Mizrach (1992) for locally weighted regression model across several currencies; Qi and Wu (2003) and Rapach and Wohar (2006).

[40] Meese and Rose (1991) find significant in-sample and out-of-sample predictability in the non-linear monetary model. Satchell and Timmermann (1995) show that, although the squared forecast errors of nonlinear models are higher than those of the random walk, non-linear models correctly predict a large proportion of the sign of exchange rate changes for several countries.

[41] Taylor and Peel (2000) and Kilian and Taylor (2003) are examples of in-sample estimation of non-linear models and out-of-sample forecasting, respectively, in the case of real exchange rate models.

[42] The difference between BMA and forecast combinations is that BMA estimates the weights in the forecast combination using Bayesian methods, whereas forecast combinations typically use equal weights or weights estimated with frequentist methods.

[43] Wolff (1987) estimated the TVP model using the Kalman filter.

In the "*Markov Switching*" model, the relationship between exchange rates and fundamentals depends on an unobservable variable. An example of the Markov Switching model is: $\left([s_t, f_t']' \,\middle|\, R_t\right) \sim N\left(\mu_{R_t}; \Omega_{R_t}\right)$, where $R_t = \{R_1, R_2, ..., R_k\}$ is an unobservable regime (or state). For example, for $k = 2$, R_1 could be a recession and R_2 could be an expansion; in this case, the relationship between exchange rates and fundamentals depends on the state of the business cycle. The regime evolves stochastically, according to a law of motion (see Engel and Hamilton, 1990, for details on the estimation and the model).

The empirical evidence is mixed. Some papers find out-of-sample forecast improvements over the random walk – see Schinasi and Swamy (1987) and Wolff (1987) for TVP models,[44] and Engel and Hamilton (1990) and Engel (1994) for a Markov Switching univariate model. In other cases, the forecast improvements are unclear: Rossi (2006) finds that a random walk coefficient TVP model outperforms the random walk for only one country; Bacchetta, van Wincoop and Beutler (2010) instead conclude that time variation cannot explain the Meese and Rogoff puzzle.[45]

4.5   Multivariate Models

Many different types of multivariate models have been used in the literature, typically generalizations of the single-equation models discussed above. Let $Y_t \equiv \left([s_t, f_t']'\right)$. Among the multivariate models, the following have been prominent.

(a) **VARs.** VARs have been considered in Meese and Rogoff (1983a,b):

$$A\left(L\right)Y_t = u_t, \tag{14}$$

where $A\left(L\right) = I - A_1 L - ... - A_p L^p$ and L is the lag operator. A special version of VARs are Bayesian VARs (BVARs), that is VARs estimated with a large number of variables imposing some Bayesian shrinkage for the parameters, which otherwise would be very imprecisely estimated in the small samples typically available to researchers. Meese and Rogoff (1983a,b) find that VARs with monetary fundamentals do not improve over the random walk. The BVARs with a large number of exchange rates in Carriero, Kapetanios and Marcellino (2009) outperform the random walk even at short horizons. Note that the BVAR considered by

---

[44]although the former did not test their significance and the latter finds significant evidence for one country out of three. Cheung and Erlandsson (2005) find in-sample empirical evidence in favor of a Markov Switching model for exchange rates.

[45]Although Chinn (2009) and Giannone (2009) debate their conclusion.

Carriero, Kapetanios and Marcellino (2009) does not include any economic fundamentals and therefore does not shed light on which fundamentals economists should focus on.

(b) Factor models. The "*factor ECM model*" proposed by Engel, Mark and West (2009) is:

$$E_t \left(s_{i,t+h} - s_{i,t} \right) = \beta_i + \beta_h \left(\delta_i' \varphi_t - s_{i,t} \right), \tag{15}$$

where φ_t is the factor extracted from the explanatory variables X_t, ($X_t = \lambda \varphi_t + u_{i,t}$, where the explanatory variables are a panel of exchange rates), i denotes the country, and the number of countries considered is large.[46] Engel, Mark and West (2009) find that, in some cases, the factor model improves forecasts at long horizons but does not improve short-horizon forecasts.[47]

(c) VECMs. The single-equation ECM model, eq. (12), is a simplification of the traditional "*multi-equation VECM model*":

$$\Delta s_{t+1} = \beta_0 + \beta_1 \left(s_t - \gamma' f_t \right) + \beta_2 \left(L \right) \Delta s_{t-1} + \beta_3 \left(L \right) \Delta f_{t-1} \tag{16}$$

where the short run dynamics is eliminated, and $\Delta \equiv 1 - L$. The empirical evidence on VECMs is mixed: some papers find positive evidence (i.e. MacDonald and Taylor, 1993, for the monetary model; Clarida and Taylor, 1997, for forward rates) while others are more negative (Rapach and Wohar, 2002, and Diebold, Gardeazabal and Yilmaz, 1994).[48] Overall, the literature suggests that single-equation ECM specifications are preferable to VECMs because the short-run dynamics of exchange rates is difficult to estimate (cfr. Cheung, Chinn and Pascual, 2005, p.1156)

(d) Multivariate time-varying parameter models. Multivariate models can also have time-varying parameters: a multivariate version of the time-varying parameter model (13) has been used in Canova (1993).[49] Due to the complexity, these models are estimated by Bayesian methods. The empirical evidence shows that multivariate TVP models may provide forecast improvements over the random walk (see Canova, 1993, for the multivariate

---

[46]Eq. (15) may include additional control variables, such the deviation of Taylor rules, monetary or PPP fundamentals from the current exchange rate of the country.

[47]They also consider the alternative specification: $E_t s_{i,t+h} = \beta_i' E_t \varphi_{t+h}$.

[48]Diebold, Gardeazabal and Yilmaz (1994) reconcile their negative finding with the fact that powerful cointegration tests reject cointegration. See also Baillie and Bollerslev (1989, 1994) for in-sample tests of cointegration among exchange rates.

[49]More in detail, Canova's (1993) model is: $B_t \left(L \right) Y_t = u_t$, where $u_t | \Im_t \sim N(0, V)$, $\Im_t$ is the information set, and $B_t = G B_{t-1} + K + A v_t$, where B_t is the vector containing all the parameters in $B_t \left(L \right)$ and $v_t \sim (0, \Sigma_t)$.

TVP model with interest rates, and Clarida, Sarno, Taylor and Valente, 2003, for the Markov Switching VECM model with forward rates).

(e) Bayesian Model Averaging (BMA). An alternative way to exploit predictability from many regressors is via Bayesian model averaging, which combines forecasts from competing models using weights that are estimated by posterior probabilities. Wright (2008) considers BMA with predictors that include both traditional predictors as well as stock prices, dividend yields and the current account. He finds that, for most currencies, BMA with sufficiently high shrinkage produces forecasts that are better than the random walk, although the magnitude of the improvements is marginal.

4.6   Panel Models

Several panel models have been estimated in the literature. The "*panel ECM*" considered by Mark and Sul (2001), Groen (2005), Engel, Mark and West (2007), Cerra and Saxena (2010) and Rapach and Wohar (2004) is:

$$E_t \left(s_{i,t+h} - s_{i,t} \right) = \beta_h \left(f_{i,t} - \gamma s_{i,t} \right),$$

where the error term contains both an individual, time-invariant component, an aggregate, time variant component, and an individual, time variant component.[50] Mark and Sul (2001), Groen (2005), Engel, Mark and West (2007) and Cerra and Saxena (2010) impose a known cointegrating parameter γ.[51] Rapach and Wohar (2004) estimate the cointegrating parameters recursively, but unfortunately do not compare the models' forecasts with the random walk. The "*panel, contemporaneous realized fundamental model*" used by Cerra and Saxena (2010) is: $E_t \Delta s_{i,t+1} = \beta f_{i,t+1}$, where $f_{i,t+1}$ is the actual, realized value of the fundamental. The "*panel, lagged fundamental model*" used by Adrian, Etula and Shin (2011) is: $E_t \Delta s_{i,t+1} = \beta f_{i,t}$, where $f_{i,t}$ is the lagged value of the fundamental.

The empirical evidence suggests that panel ECMs are quite successful for the monetary model (see Mark and Sul, 2001; Groen, 2005; Engel, Mark and West, 2007; and Cerra and Saxena, 2010) and for funding liquidity fundamentals (Adrian, Etula and Shin, 2011). They are less successful for PPP fundamentals (Mark and Sul, 2001).

---

[50]Its forecasts are: $\beta_h \left(f_{i,t} - \gamma s_{i,t} \right) + \zeta_i + \frac{1}{t} \sum_{j=1}^{t} \theta_j$, where ζ_i is the individual, time-invariant component and θ_t is the aggregate, time-variant component.

[51]In particular, Groen (2005) imposes knowledge on the number of cointegrating vectors as well as the parameters of the common cointegrating vector, and only re-estimates the intercept and the coefficients of short-run dynamics.

4.7 What Have We Learned?

Among the model specifications considered in the literature, the least successful are non-linear specifications, and the most successful are linear specifications. Among the linear models, the single-equation ECM and the panel ECM models are the most successful at long horizons, although there is disagreement among researchers about the degree of robustness of the results. Typically, but not always, for single-equation linear models the predictor choice matters more than whether the researcher uses contemporaneous, realized or lagged fundamentals; the linear monetary model does not perform well at any horizons, whereas Taylor-rule fundamentals and net foreign assets are successful predictors at short-horizons. Models with time-varying parameters show some degree of success, although most of the favorable empirical evidence is based on studies in the late 1980s-early 1990s, and the most recent analyses report less success (although with slightly different specifications). Among the multivariate models, the most successful specification is the panel ECM, although there is some evidence in favor of BVARs, BMA and factor ECM models.

The next section will investigate why, among researchers using the same model, there is substantial disagreement across empirical findings by examining in details other important dimensions in which the studies differ.

5 Which Data To Use?

Existing studies in the literature differ considerably depending on the characteristics of their data. In particular, they differ depending on the data transformation they perform, the countries they study, the sample they use, the frequency of the data, and whether the data are fully revised or real-time. This section reviews what choices have been made in the literature and what we know about how they potentially affect predictability; in particular, we include a discussion that revisits the empirical findings in the previous sections in this light. Panel B in Table 2 overviews data choices in selected papers in the literature.

5.1 End-of-sample versus Filtered Data and Calibrated Parameters

When conducting out-of-sample forecast evaluation, it is important to make sure that the information contained in future data is not trivially used to estimate the models' parameters.

To clarify, consider a forecaster estimating the model at time t and producing the h-step-ahead forecast for time $t + h$. It is important that when the forecaster produces the forecast at time t, the information contained between times t and $t + h$ (or later) is not used directly or indirectly in the estimation, otherwise the model has an unfair advantage relative to the random walk. In other words, the model would be given the advantage of "peeking" at data after time t, which would not have been available to the forecaster in real time. One example when this might happen is when the data are de-trended over the full sample or de-trended by any two-sided filtering procedure. For example, when using the output gap to forecast exchange rates, it is important that the output gap, usually defined as the difference between output and a linear or a quadratic trend, be de-trended using only information available at time t. Seasonal adjustment is another example where the treatment of the data is important. All the raw data used in Meese and Rogoff's (1983a) study were seasonally unadjusted: they performed the seasonal adjustment procedure themselves on the raw data. The reasons are twofold: using raw data allows researchers to seasonally adjust the data in a consistent way across different series; furthermore, it allows them to avoid using information that was not available at the time the forecast was made (for example, it is preferable to use a one-sided moving average seasonal adjustment filter, as opposed to a two-sided one, as forecasts based on two-sided filtered data implicitly use information which would have not been available at the time of the forecast). As a third example, one should focus on forecasting actual exchange rates, and not Hodrick-Prescott filtered exchange rates. As a fourth example, using calibrated fundamentals (e.g. linear combinations of macroeconomic predictors, where the coefficients are calibrated rather than estimated) is another example of "ad-hoc" filtering, which could potentially give the model an unfair advantage since it is unclear how the parameters are calibrated (in particular, what data inspired the calibrated parameters). As a fifth example, sometimes, auxiliary parameters are imposed in an "ad-hoc" manner, and it is unclear whether results are robust to such choices. Consider the case of a researcher using BMA techniques and shrinking the parameter towards the random walk (that is, imposing a rule to make the parameter smaller); if the rule is ad-hoc, it is unclear whether peeking at the whole dataset might have influenced the degree of shrinkage. Other examples (including the choice of the estimation sample or the use of real-time data) are discussed in detail below.

Let us critically overview how these "ad-hoc" procedures may affect the favorable findings of predictive ability reviewed in the previous sections. Typically, in the ECM estimation

literature, researchers impose a known cointegrating parameter.[52] The predictability found in VECMs with monetary fundamentals also typically relies on the cointegrating parameter estimated over the full sample:[53] Cheung, Chinn and Pascual (2005) and Alquist and Chinn (2008) show that the empirical evidence in favor of long horizon predictability of the monetary model is much weaker or completely disappears after estimating the cointegrating parameters. Thus, some of the disagreements over the empirical performance of the monetary model at long horizons can be explained by differences in models' specification: the model performs better by imposing an ad-hoc calibrated cointegration parameter than by estimating it.[54] In some panel analyses (i.e. Adrian, Etula and Shin, 2011) the fundamentals are de-trended over the full sample prior to forecasting as well. The favorable evidence for the BVAR relies on choosing the degree of Bayesian shrinkage, and it is unclear how different degrees of shrinkage may affect the results.

5.2 Forecasted versus Realized (or Ex-post) Fundamentals

The previous section highlights that the same model can be estimated either using forecasted or realized fundamentals. The latter also give an unfair advantage to the fundamentals-based economic model. However, it is important to note that the original Meese and Rogoff (1983a,b) analysis used future realized fundamentals, as opposed to fundamentals known at the time of the forecast, to predict exchange rates because they were proving a negative result. The result that Meese and Rogoff (1983a,b) wanted to prove is that no fundamental forecasts exchange rates better than a random walk; by showing that not even realized, h-period-ahead fundamentals were capable of predicting exchange rates, they made their findings really stark. In other words, if forecasted h-period-ahead fundamentals had no predictive content for h-period-ahead exchange rates, that might be due either to the lack of predictive content or to the fact that forecasts of fundamentals are poor. By demonstrating that even when using realized fundamentals it is not possible to improve exchange rate forecasts, Meese and Rogoff (1983a,b) were able to prove that the lack of predictive content in

---

[52]For example, Mark (1995) and Chinn and Meese (1995). Typically, papers in this literature also impose knowledge on the number of cointegrating vectors.

[53]For example, MacDonald and Taylor (1994) find that the sticky-price monetary model estimated via cointegration outperforms the random walk out-of-sample; however the cointegrating vector is estimated over the entire sample, generating forecasts that incorporate future realized values and hence they are not truly ex-ante.

[54]In fact, Groen (1999) argues that the cointegration relationship between exchange rates and monetary fundamentals is unstable and disappeared in the later sample.

the fundamentals is really the cause of the problem. Table 2 provides an overview on the use of forecasted versus realized fundamentals in the literature. As noted in the previous section, typically for single-equation linear models, the predictor choice matters more than whether the researcher uses contemporaneous, realized or lagged fundamentals. For panel models, Cerra and Saxena (2010) demonstrate that the use of realized versus lagged fundamentals does not matter either.

5.3   Countries and Samples

Existing studies in the literature differ with respect to the countries they consider as well as the sample. Most studies focus on bilateral exchange rates versus the US dollar, although there are exceptions.[55] Typically, the sample starts in 1973 to avoid the period of fixed exchange rates and typically the end of the sample is constrained by data availability; for example, individual European countries' exchange rates are available only prior to the Euro unification. More recent studies have replaced the exchange rate of the countries that are now part of the European Union with the euro-dollar exchange rate; its value can be back-casted to the period before the unification as well, although the effects of back-casting are unclear. A clear message in the literature is that models and predictors that forecast exchange rates well in one country do not necessarily provide competitive forecasts for other countries (see Cheung, Chinn and Pascual, 2005).

5.4   Frequency

Existing studies in the literature also differ with respect to the frequency of the data they consider, ranging from low frequency, yearly data, to quarterly, monthly, and even weekly, daily or very high-frequency data (such as those available at 30-minute intervals). There is clearly a trade-off between the frequency of the data and the span of the data, as lower frequency data (which are perhaps more informative regarding long-run trends in the data) are by construction available only for shorter samples.[56] Typically, studies interested in the long-term forecastability of exchange rates focus on quarterly data (i.e. Mark, 1995). There are no striking discrepancies between empirical results of studies that focus on monthly data

---

[55]For example, Meese and Rogoff (1983a,b) consider also the trade-weighted exchange rate.

[56]E.g. longer spans of data have been considered by Rapach and Wohar (2002) at the price of restricting attention to yearly frequencies and a smaller sample. See also Taylor (2002) on using longer span of data for the analyzing real exchange rates.

relative to those that focus on quarterly data; some of the studies that focus on daily data (for example, studies that use macroeconomic news announcements as predictors)[57] do generally find positive evidence for the model, but typically they are estimated in-sample, so the real difference seems to be whether the models are compared in-sample versus out-of-sample. We conclude that the frequency of the data does not typically affect findings of predictive ability.[58]

5.5   Data Revisions

Typically, researchers evaluate the out-of-sample forecasting ability of their models using data available at the time of their study. However, the data that a researcher has available in 2000Q1 for, say, GDP in 1998Q1 is not the same as that available to a researcher in 1998Q1. The reason is that data are continuously revised by statistical agencies. The revisions include those due to statistical agencies acquiring additional source information to update their initial estimates, changes in the aggregation methods (fixed to chain weighting), changes in base years used for calculating real variables, or changes in definitions of the concept being measured. Financial data, such as interest rates and exchange rates, are never revised, but GDP, money aggregates and inflation are. A real-time dataset collects vintages of data that were actually available to researchers at each point in time, before data revisions were applied to the data to obtain the finally revised data available at the end of the sample (or current vintage data). The conventional wisdom is that, for example, the use of revised data may lead researchers to include regressors that had little predictive content in real-time and may exaggerate the forecasting ability relative to predictors that were actually available at the time (see Koenig, Dolmas and Piger, 2003, and Croushore and Stark, 2001,2003).

The empirical evidence shows that the findings in favor of the long-horizon predictive ability of the monetary model are not robust to using real-time data. In particular, Faust, Rogers and Wright (2003) show that the favorable evidence of long-horizon predictive ability is present only over a two-year window of data vintages around those used in Mark (1995):

---

[57]Macroeconomic news announcements are defined as the difference between expected and realized ("announced") macroeconomic fundamental values. Expectations are typically measured by surveys. For example, Andersen, Bollerslev, Diebold and Vega (2003), Ehrmann and Fratzscher (2005), Faust, Rogers, Wang and Wright (2007) and Fratzscher (2009) study whether, in-sample, exchange rates react in a very short window around the announcement.

[58]One exception might seem Ferraro et al. (2011). However, even in their case, the main reason for the predictive ability is not just the use of daily data, but the use of both daily data and commodity prices as fundamentals.

most of the predictive ability eventually disappears after data revisions. Interestingly, they also find that the economic model performs better using real-time data than finally revised data. On the other hand, the empirical evidence shows that the predictive ability of Taylor-rule models is robust to using real-time data – see Molodtsova et al. (2010).

5.6   What Have We Learned?

Data transformations (such as de-trending, filtering and seasonally adjusted data) may crucially affect predictive ability, and may explain why, for the same models, some researchers do find predictive ability while others do not. Cheung, Chinn and Pascual (2005) and Alquist and Chinn (2008) find that the empirical evidence in favor of the single-equation ECM monetary model is much weaker or completely disappears after estimating the cointegrating parameters. The evidence in favor of the panel ECM model also relies on calibrated cointegration parameters, and it is unclear whether the findings would be robust to estimating the parameters; in addition, some panel ECM studies rely on data that have been previously de-meaned over the full sample, which may create an unfair advantage for the economic model over the random walk. Another important factor that, for some fundamentals, may weaken predictive ability is using real-time, rather than realized, data; this is the case for monetary fundamentals but less of a concern for Taylor-rule fundamentals. For a given model and predictor, predictive ability also depends on the choice of the country. On the other hand, the frequency of the data and whether the realized or the forecasted fundamental is used do not seem to affect predictability: the monetary model's forecasts do not beat a random walk either way. However, there are exceptions.[59]

6   Which Forecast Evaluation Methods To Choose?

This section provides an overview of the methods that have been used in the literature to assess the predictive content of macroeconomic fundamentals for exchange rate forecasts as well as important issues to keep in mind when evaluating the predictive content. We also critically revisit the empirical findings in the previous sections in the light of these issues. Table 2, Panel C, overviews the forecast evaluation methods used in selected papers.

---

[59]Sometimes forecasts perform better using real-time forecasts of future fundamentals instead of actual future fundamentals, as in Faust, Rogers and Wright (2003). Ferraro et al. (2011) find that predictability is stronger using realized (as opposed to lagged) fundamentals.

6.1 Choice of Benchmark Model

The majority of studies compare the out-of-sample forecasting performance of the predictors with those of a random walk without drift ("RW"), as the latter has been shown to be the best predictor of exchange rates since Meese and Rogoff (1983a,b) – see Table 2, Panel C. According to the random walk without drift, the best predictor of exchange rates tomorrow is the exchange rate today.[60] Thus, exchange rate changes are completely unpredictable:

$$E_t s_{t+h} - s_t = 0$$

An alternative benchmark is the random walk with drift ("RWWD"), according to which exchange rate changes are predictable but independent of other macroeconomic variables: $E_t s_{t+h} - s_t = \alpha \neq 0$.

It is important to note that the efficient market hypothesis does not imply that exchange rate changes should be unpredictable. That is, the Meese and Rogoff (1983a,b) finding that the random walk provides the best prediction of exchange rates should not be interpreted as a validation of the efficient market hypothesis. The efficient market hypothesis means that bilateral exchange rate is the market's best guess of the relative, fundamental value of two currencies based on all available information at that time. The efficient market hypothesis does not mean that exchange rates (like any asset prices) are unrelated to economic fundamentals, nor that exchange rates should fluctuate randomly around their past values.

Consistently, across papers, the random walk is used as the benchmark. Typically, when the random walk with drift forecasts are reported, they are worse than those of the random walk without drift. This may explain some puzzling results in the literature. For example, among papers that study non-linear models, Meese and Rose (1991) found more empirical evidence in favor of the nonlinear model, whereas Chinn (1991) and Chinn and Meese (1995) found the least. However, the benchmark in the former paper is the random walk with drift, whereas the benchmark in the latter two papers is the random walk without drift.

6.2 Choice of Forecast Horizon

Forecast evaluation also requires the choice of a forecast horizon, h. Forecast horizons typically considered in the literature range from short horizons (typically 1 month or 1 quarter ahead, depending on whether data are monthly or quarterly) to long horizons (horizons up

---

[60]When not specified otherwise, random walk means the random walk without drift.

to 4-5 years: for example, Mark (1995) considers forecast horizons up to 16 quarters, and Cheung, Chinn and Pascual (2005) up to 20 quarters).

The choice of the forecast horizon is important since it is possible that the models' predictive ability may depend on it. Typically, the empirical evidence in favor of fundamentals' predictability appears at horizons that differ depending on the predictor. For example, for monetary fundamentals, most studies agree on lack of short-horizon predictability (i.e. Cheung, Chinn and Pascual, 2005); however, there is evidence of long horizon predictability at the 3-4 year horizon (Mark, 1995). The predictability of Taylor-rule fundamentals is generally found at short horizons (Molodtsova and Papell, 2009), whereas the predictability of net foreign assets models is found at both short and long horizons by some studies (Gourinchas and Rey, 2007), and only at short horizons in some other studies (Alquist and Chinn, 2008).

6.3 Choice of Forecast Methodology

Assessing the forecasting ability of a model requires several choices. First, which forecasting method to use. The literature has been focusing mainly on rolling or recursive window forecasting schemes (see West, 1996), where parameters are re-estimated over time using a window of recent data – see Section 2. Shorter estimation window sizes allow the parameter to adapt more quickly to structural changes; on the other hand, the parameter is less efficiently estimated in a smaller sample.[61] The choice of how to split the sample between in-sample and out-of-sample periods as well as the choice of the rolling window size varies across papers. For example, among papers using a rolling window scheme, Meese and Rogoff (1983a) use a window of 93 observations, Chinn (1991) uses 48, Qi and Wu (2003) use 216, Cheung et al. (2005) consider 42 and 59, van Dijk and Franses (2003) use 128, Clark and West (2006, 2007) and Molodtsova and Papell (2009) use 120 observations. See Table 2, Panel C, for additional details.

Rogoff and Stavrakeva (2008) and Inoue and Rossi (2012) examine the robustness of Taylor-rule and panel ECM monetary models and find that their performance changes substantially depending on the estimation window size. Thus, it is not the case that the predictive ability is present no matter which window size is used, as pointed out in Rogoff and Stavrakeva (2008); on the other hand, there are window sizes for which the predictive ability is significant, as discussed in Inoue and Rossi (2012). Regarding the net foreign asset model, Rogoff and Stavrakeva (2008) find that its performance is less sensitive, although still not

---

[61]See e.g. Rossi (2011).

entirely robust, to the choice of the window size.

6.4   Choice of Evaluation Methods

The forecast evaluation process requires two main choices: which loss function to use to evaluate the forecast, and which test statistic to use to assess significance. Regarding the choice of loss function, researchers typically evaluate models' out-of-sample forecasting performance according to their mean squared forecast error ("MSFE") or root mean square forecast error ("RMSFE"), as in Meese and Rogoff (1983a,b, 1988). The former is simply the average, over the out-of-sample period, of the squared forecast errors of a model; the latter is its square root – see Section 2. Typically, researchers focus on either the RMSFE difference of competing forecasts or Theil's (1996) U statistics, which is the ratio of model RMSFEs to the RMSFEs of the benchmark model.[62] Note that this is only one of the available metrics/loss functions that can be used. Alternatively, researchers have used mean absolute errors ("MAE", as in Meese and Rogoff, 1983a) and asymmetric loss functions (Ito, 1990, West, Edison and Cho, 1993). Different loss functions give different weight to the deviations of the forecast from the target; for example, the MSFE gives equal weight to over- and under-predictions of the same magnitude.[63] Alternatively, the forecast evaluation exercise could target: (i) the direction of prediction, as in Engel (1994), Cheung et al. (2005) and Cerra and Saxena (2010); (ii) a utility-based measure, as in West, Edison and Cho (1993), Rime, Sarno and Sojli (2010), Abhyankar et al. (2005) and Della Corte et al. (2010); or (iii) the whole predictive density or interval forecasts, as in Diebold, Hahn and Tay (1999), Inoue and Rossi (2008) and Wang and Wu (2009). The direction of change statistic calculates the proportion of forecasts that correctly predict the direction of change of the exchange rate.[64] As opposed to the RMSFE criterion, it focuses on whether the sign of the forecast change is correct, and, unlike the RMSFE, is not influenced by the distance between the forecast and the actual realization. In fact, it is theoretically possible that a model could forecast perfectly the direction of change in all periods and yet forecast worse than the random walk according to the RMSFE criterion; this may happen, for example, if the model consistently over-predicts the magnitude of the change. For example, Satchell and Timmermann (1995) theoretically prove that, whereas in linear models the probability of predicting the sign of a

---

[62]While there are other forms of Theil U statistic, this is the one commonly used in the exchange rate literature.

[63]See Elliott and Timmermann (2008) for a review of asymmetric loss functions.

[64]See Pesaran and Timmermann (1992) for a test for the direction of prediction.

stochastic variable decreases when the MSFE increases, this monotonic relationship between the MSFE and the probability of correctly predicting the sign of the target variable does not need to hold in general non-linear models. Utility-based measures are typically interpreted as the transaction fee that a professional money manager can charge for providing estimates of the economic fundamental model to an investor that uses the benchmark model. The evaluation of the correct specification of density forecasts goes beyond point forecasts by evaluating the whole predictive density.[65]

The significance of superior forecast performance is typically assessed via out-of-sample predictive ability tests or in-sample Granger-causality tests. Both provide important insights, and are used for different goals. The former are used for assessing whether predictors would have improved exchange rate predictions in forecasting environments that mimic as closely as possible the one faced by forecasters in practice, and have been used in Meese and Rogoff (1983a,b). The latter are in-sample tests on whether the lagged predictor has significant explanatory power for exchange rates over the full sample, and have been used for example in Andersen et al. (2003).[66] It is important to keep in mind that the former is a much more challenging exercise than the latter: predictors which pass the latter may still not have predictive ability in a truly out-of-sample forecasting exercise. Indeed, the Meese and Rogoff puzzle is the finding that, although fundamentals are significant predictors of exchange rates in-sample, their out-of-sample predictive ability is not superior to that of the random walk. A version of Granger-causality test robust to instability is Rossi (2005b); unlike the traditional Granger-causality tests, it detects predictive ability even if it appears only in a sub-sample, or if the predictive relationship changes over time. The latter test has been used, for example, in Chen, Rogoff and Rossi (2010).

Traditional tests of out-of-sample predictive ability can be distinguished between absolute

---

[65] It typically relies on the Probability Integral Transform (see Diebold, Hahn and Tay, 1999, and Corradi and Swanson, 2006). See also Rossi and Sekhposyan (2012) for tests robust to instabilities.

[66] A different evaluation method is used in Engel and West (2005). They focus on evaluating whether exchange rates Granger-cause fundamentals in-sample (rather than vice-versa). The reason is because they show that, if fundamentals have a unit root and the discount factor is near one, the net present value condition implies that exchange rates are near–random walks. Since the present value condition implies that exchange rates should predict fundamentals, they tested whether exchange rates Granger-cause fundamentals in-sample. While this is an implication of their framework, and therefore a clever way to provide indirect empirical evidence to substantiate their theory, it is nevertheless not a necessary condition as exchange rates might Granger-cause future interest rates either because of the net present value condition or because interest rates are set by Central Banks to mitigate exchange rate fluctuations. Engel and West (2004) study the implications of a discount factor near unity for exchange rate volatility.

tests and relative tests of forecast evaluation. The former evaluate properties that optimal forecasts of a model should have, such as unbiasedness and uncorrelatedness.[67] The latter evaluate which, among competing models, forecasts the best, and include the tests proposed by Diebold and Mariano (1995), West (1996), Clark and McCracken's (2001) ENCNEW, Giacomini and White (2006), Clark and West (2006, 2007), among others. The Meese and Rogoff puzzle concerns tests of relative predictive ability, since Meese and Rogoff (1983a,b) compare the forecasting ability of fundamental-based exchange rate models with the random walk. While the tests for relative forecast performance developed in the literature are typically applied to MSFE differences between models, there is an important difference among them: on the one hand, West (1996), Clark and McCracken (2001) and Clark and West (2006,2007) test out-of-sample whether the benchmark model is equivalent to the competing model *in population*; instead, Diebold and Mariano (1995) and Giacomini and White (2006) test whether two models' forecasting ability is the same. I.e. the former test, using out-of-sample forecasts, whether, say, $\beta = 0$ in $E_t \left(s_{t+h} - s_t \right) = \beta f_t$; the latter test whether the forecasts of the fundamental model and that of the random walk are equivalent. Thus, the latter approach might be useful when the researcher is interested in evaluating forecasts, and the former when (s)he is interested in evaluating models in population. The difference between the two approaches is that, in nested models, the sample MSFE from the larger model is expected to be greater than that of the small model even when, in population, the two models have the same predictive ability, since the larger model introduces noise into its forecasts by estimating parameters that are useless in forecasting – see Clark and West (2006). Therefore, a finding that the small model has a smaller MSFE does not necessarily mean that the additional predictors present under the larger models are not useful for forecasting. The former tests take this into account in evaluating the relative predictive ability of competing models by re-centering the MSFE differential in the test statistic by a term that accounts for parameter estimation error and hence test whether, after adjusting for parameter estimation error, the larger model outperforms the benchmark – but parameter estimation error *is exactly one of the reasons* why there is a difference between in-sample fit and out-of-sample forecasts. Since the tests rely on different assumptions and null hypotheses, they have different power properties. See West (2006) and Diebold (2012) for insightful discussions. The majority of the papers in the exchange rate literature focus on the Diebold and Mariano (1995) and West (1996) tests (hereafter "DMW") or the Clark and West (2006,

---

[67]See Mincer and Zarnowitz (1969), West and McCracken (1998), and Patton and Timmermann (2006, 2007).

2007) tests (hereafter "CW"). Table 2, Panel C, provides an overview of which tests have been used in the literature.

The main conclusion that emerges from the literature is that the choice of the evaluation method matters, and a lot! For example, the use of different test statistics may explain the contradicting evidence on the empirical validity of UIRP. Typically, the majority of the studies that find lack of predictability for interest rate differentials either focus on RMSFEs or on the Diebold and Mariano (1995) test. Clark and West (2006) find that, based on the Diebold and Mariano (1995) and West (1996) tests, there is little evidence that UIRP beats the random walk; however, UIRP produces better forecasts than the random walk according to the Clark and West (2006) test. Similarly, using the Clark and West (2006) test, Alquist and Chinn (2008) conclude that UIRP can significantly outperform the random walk at long horizons.

The choice of the forecast evaluation method may also shed some light on the controversy over the empirical evidence on long horizon predictability of monetary fundamentals. On the one hand, Kilian (1999) and Berkowitz and Giorgianni (2001) dispute the finding of predictive ability of the monetary model at long horizons due to issues with the bootstrap implementation of the test and the fact that high persistence would distort inference. In particular, Berkowitz and Giorgianni (2001) and Berben and van Dijk (1998) show, via Monte Carlo simulations, that predictive ability measures of two completely unrelated time series could display large R-square and DMW statistics at long horizons if they are both highly persistent, a finding reminiscent of spurious regression. Kilian (1999) instead argues that, when bootstrapping the DMW statistic, it is important that the critical values of the statistic be calculated under the null hypothesis of the absence of a drift in the random walk, otherwise the bootstrap model is not consistent with the model used as the benchmark. On the other hand, Rossi (2005a) shows that the high persistence in exchange rates and fundamentals biases the estimation of the economic models, but not the random walk: when this bias outweighs the benefits from exploiting economic predictors, the random walk forecasts better.[68] Her findings show that the poor forecasting ability of economic models does not imply that the models are not a good description of the data.[69]

The choice of the evaluation metric may also be important in explaining differences in empirical results across studies. While Meese and Rogoff (1983a,b) show that their negative

---

[68] This happens even if the economic model is the true data generating process.

[69] A further limitation of traditional forecast evaluation tests is that they typically rely on stationarity assumptions, and are not robust to the presence of instabilities – see Section 6.5.

finding is robust to using either mean squared forecast errors or mean absolute errors, several papers note that other metrics are more successful in finding out-of-sample evidence in favor of macroeconomic predictors: Engel (1994) and Cheung, Chinn and Pascual (2005) find that the direction of change statistics provides more empirical evidence that models can outperform the random walk than MSFE comparisons; Abhyankar et al. (2005) finds similar results for utility-based measures.

6.5   Choice of the Forecast Sample

The sample utilized for out-of-sample forecast evaluation differs substantially across studies, not only because the studies have been performed at several points in time and differ in the overall samples they consider, but also because they differ in the choice of the window size used for estimation. Clearly, in the presence of instabilities, the models' performance could potentially be sensitive to the sample used for forecast evaluation.

Again, the literature is very clear on the following: the choice of the forecast sample matters enormously! In fact, one of the most interesting results in the literature concerning the robustness of results to the forecast period is the lack of robustness. For example, long horizon predictability of the monetary model is very sensitive to the forecast period: Kilian (1999) and Groen (1999) find that the long horizon predictability of the monetary model disappears once one extends Mark's (1995) sample. Cheung, Chinn and Pascual (2005) consider two out-of-sample periods and note that the superior performance of a particular combination of monetary fundamental, model specification and country combination does not necessarily carry over from one out-of-sample period to the other. Rogoff and Stavrakeva (2008) note that neither the out-of-sample forecasting ability of panel VECM models nor that of Taylor rules are robust to the choice of the out-of-sample forecast period. Giacomini and Rossi (2010) note substantial changes over time in the predictive ability of Taylor-rule and UIRP fundamentals relative to the random walk.

The lack of robustness of results across sample periods is consistent with the possibility that the parameters be unstable: in fact, Rossi (2006) shows that parameter instabilities are empirically important in exchange rate prediction. This suggests that it is crucial to consider evaluation methodologies that are robust to instabilities, such as Giacomini and Rossi's (2010) test, rather than traditional methods.[70] For example, using their proposed test, Giacomini and Rossi (2010) uncover a sharp worsening in the forecasting ability of UIRP

---

[70]See Rossi (2011) for an overview of forecast tests in the presence of instabilities.

in the late 1980s, to the point that traditional measures of predictive ability (that evaluate the performance of the model on average over the out-of-sample period) would overstate the recent predictive ability of the UIRP model. Rossi and Sekhposyan (2011) apply new methodologies to better understand why the economic models' performance is poor. They find that lack of predictive content is the major explanation for the lack of short-term forecasting ability of the economic models, whereas instabilities play a role especially for medium term (one-year ahead) forecasts.

6.6   What Have We Learned?

What we learn from this section is that the choice of the forecast evaluation method matters a lot. In particular, we note the following facts. The choice of the benchmark model is very important: choosing an inappropriate benchmark model overstates the empirical evidence in favor of the economic model's predictive ability; the random walk (without drift) is the toughest benchmark to beat and should be the one used in practice. The forecast horizon also matters a lot: typically, monetary fundamentals do not have short-horizon predictive ability (1 month or 1 quarter), although the monetary fundamentals in an ECM model may forecast better than a random walk at long horizons (3-4 years). In general, both the evaluation method as well as the forecast sample may potentially and substantially matter. For example, UIRP may outperform the random walk if certain tests are used (CW rather than DMW); in addition, its performance is unstable over time and worsened in the last decades. Rogoff and Stavrakeva (2008) find that the out-of-sample forecasting ability of the panel ECM model is not robust either to the choice of the forecast window nor to the choice of the out-of-sample forecast period. To conclude, Panel D in Table 2 summarizes whether the predictors in the selected papers we consider are successful in forecasting exchange rates.

7   Empirical Analysis

To summarize, there is substantial disagreement in the literature regarding whether exchange rates are forecastable. The literature clearly reached a consensus on several "negative" stylized facts, namely the fact that the monetary and PPP fundamentals have no predictive ability at short horizons. There is also agreement on the fact that non-linear models are the least successful models. Finally, there is also agreement on the fact that, should monetary models have any predictive ability at long horizons, it only appears in single-equation ECM

and panel ECM model specifications. On the other hand, there is substantial disagreement on whether monetary fundamentals do have predictive ability at long horizons and whether UIRP has predictive ability at short horizons. There is some evidence in favor of BMA models, although not strong. The only "positive" finding the literature reached a consensus on seems the fact that Taylor-rules and net foreign assets models have significant forecasting ability at short horizons.

In this last section we undertake an empirical analysis to illustrate several of the different approaches, models and methods used in the literature, and revisit the empirical findings using the most up-to-date sample, methods and fundamentals we have available. We focus on univariate linear models with either traditional or newly proposed fundamentals, the single-equation ECM and panel ECM monetary models as well as the BMA. Since non-linear models fit well in-sample, but produce poor out-of-sample forecasts, we do not consider them in our empirical exercise; similarly, the most recent studies report lack of empirical support for time-varying parameter models, so we will not consider them either.

Given that the literature differs on the data transformations, we will focus on a unifying framework where the only data transformation is seasonal adjustment, which we perform consistently for all series. We use several of the methods reviewed in Section 6 in the empirical analysis.[71]

7.1   Data Description

We collect data on exchange rates (relative to the US) and several economic fundamentals for a variety of countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, New Zealand, Portugal, Spain, Sweden, Switzerland, UK and the US.[72] For all countries, we collect monthly and quarterly data on overnight interest rates, 3-month Treasury Bills, 5-year Treasury Bonds, GDP/ industrial production, CPI and the money stock. We also collect data on the current account, the trade balance, public debt and government deficit/surplus – the latter data are available only at the annual or quarterly frequency. Data are from the IMF database via Datastream as well as

---

[71]Data limitations prevent us from studying the importance of real-time versus revised data for the countries that we consider.

[72]While we collected data also for Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Hong Kong, Mexico, Peru, Portugal, Russia, Singapore, Taiwan, Thailand and Turkey, the sample sizes of either the exchange rate or the fundamentals were severely limited, or potentially severely affected by measurement error. These countries were discarded from the analysis.

Philip Lane's website (http://www.philiplane.org/EWN.html). Since countries' geographical definitions have changed over time (for example, after the start of the euro), the sample size differs across countries. Also, slightly different fundamentals are available depending on the country (for example, the exact definition of money supply available might differ across countries, and the sample sizes for our measure of output for the same country differ depending on the frequency since GDP is only available at the quarterly frequency, and industrial production is available at the monthly frequency). We selected the most homogeneous definitions we could find, while striving to collect data that have the longest available sample. Table 3 provides information on the available samples. A not-for-publication Appendix provides additional details on the data, as well as their mnemonics and sources. The series are originally not seasonally adjusted, and seasonal adjustment was performed using one-sided moving averages with backward, equal weights.[73]

7.2   Methodology

We present empirical results using a few of the most used predictors and successful methodologies. Regarding the predictors, we consider interest rate differentials, price differentials, money and output differentials, Taylor rule fundamentals and measures of external imbalance such as the current account and the trade balance.[74]

Regarding the models, first we consider the performance of selected traditional single-equation linear models (UIRP, eq. 1; PPP, eq. 2, and monetary) as well as Taylor-rule fundamentals (eq. 8) and portfolio balance models (eq. 6).[75] The monetary model is considered either as in eq. (3) with variables in first differences, or in the ECM form, eq. (12).[76] We will then turn to two representative multivariate models (the BMA and the panel VECM models). We will compare the models' forecasting performance at short (one month or one quarter) and long (four years) horizons with that of a random walk without drift using RMSFE comparisons. The models' parameters are recursively re-estimated with a rolling window whose size is equal to half of the total sample size in the baseline analysis,

---

[73]For quarterly data, the filter is $1/3+1/3L+1/3L^2$; for monthly data it is $(1/12)+(1/12)L+...+(1/12)L^{11}$. Empirical results based on seasonally unadjusted data are qualitatively similar.

[74]The analysis will focus on monthly data when possible, given that both monthly and quarterly data have the same span but the former provides larger sample sizes.

[75]Unfortunately, net foreign assets data are available only at the annual frequency, so an analysis with the techniques currently used in this paper is not possible due to the small sample sizes.

[76]We report results for calibrated cointegration parameters: $\phi = 1$, as in Mark (1995) – see the discussion in Section 4.

although we investigate the consequences of varying both the size of the window and the forecast sample. Since the literature review highlights the importance of the forecast sample, we formally investigate the robustness (or lack thereof) of the empirical findings to both the choice of the out-of-sample forecast period and the choice of the estimation window size by using the tests proposed by Giacomini and Rossi (2010) and Inoue and Rossi (2012), respectively.[77]

We will compare models' performance using both in-sample and out-of-sample tests. As discussed in the previous section, out-of-sample forecast tests may have different power properties, so in the empirical application, we shed some light on this issue by considering both the traditional DMW and the CW tests.

7.3   Empirical Evidence

Results are reported in Tables 4 and 5. More in detail, for each model we consider, the first column reports the country whose nominal exchange rate growth we are forecasting (relative to the US dollar). The second column (labeled "GC") reports p-values of Rossi's (2005b) Granger-causality test robust to instabilities.[78] The third column reports the ratio of the root mean squared forecast error of the model relative to that of the random walk without drift (labeled "RMSFER"): values smaller than unity denote that the model forecasts better than the random walk benchmark. Finally, the last two columns assess the significance of the test for out-of-sample forecasting ability: the column labeled "DMW" reports p-values of the Diebold and Mariano (1995) and West (1996) test and the column labeled "CW" reports p-values of the Clark and West (2006) test.[79] The benchmark model in the latter two tests is the random walk without drift. "h" denotes the forecast horizon. All tests are implemented with Newey and West's (1987) heteroskedasticity and serial correlation robust covariance matrix.[80] Note that p-values lower than, say, 0.05 denote significance at the 5% level, which is the definition of significance we adopt in what follows when discussing the results. In all the tables, "n.a." denotes cases in which the sample size was too small and

---

[77]Note that since the total sample size varies by country, the size of the rolling window as well as the out-of-sample forecast sample vary by country, too.

[78]Detailed empirical results based on traditional Granger-causality tests are reported in the not-for-publication appendix and their findings summarized in this section.

[79]While the DMW test is typically used for non-nested models' forecast comparisons, in this article we interpret its critical values following Giacomini and White (2006), which show that the DMW test can be used to compare forecasts of nested models provided a rolling window is used for estimation.

[80]The truncation parameter is $T^{1/4}$, where T is the available sample size.

the test statistic is not reported.

Traditional Predictors' Performance in Monthly Data. Table 4, Panels A-D report the performance of the traditional predictors. The panels show some in-sample predictive ability at short (one-month-ahead) horizons for UIRP and the monetary ECM models,[81] but very limited at long (four-year-ahead) horizons for any of the models.[82] The out-of-sample predictive ability is almost inexistent: typically, models' MSFEs are larger than the random walk's,[83] and neither DMW nor CW find significant predictive ability. Why is it that our results for the monetary ECM model are so different from those in the literature? We investigate this issue at the end of this section.

Alternative Predictors' Performance. Panel E in Table 4 shows empirical evidence in favor of in-sample predictability of Taylor rule fundamentals for several countries. The CW test finds empirical evidence in favor of the Taylor-rule model only at short horizons: it is strongly significant for four countries and marginally for five others at the one-month-ahead horizon, and for none of the countries at the four-year horizon. Among the fundamentals that we consider, then, they are among the most successful out-of-sample at short horizons. Importantly, note however that the DMW test does not find predictive ability. The discrepancy between the CW and DMW tests emphasizes that the way we treat parameter estimation error matters: if we compare models and correct inference for the fact that the Taylor model estimates more parameters than the random walk, we conclude that the Taylor model, when evaluated "in population", has better predictive ability; however, if we evaluate its forecasts "at the actual estimated parameter values", their forecasting ability is not superior to that of the random walk.

Panel F in Table 4 reports results for the model that includes real interest rates, the trade balance and the current account as predictors (at the quarterly frequency). Interestingly, among the models we consider, this is one with the strongest in-sample predictive ability; however, its out-of-sample performance is extremely poor (possibly due to the large number of parameters to estimate).

Multivariate Models. Results for multivariate models are reported in Table 5. Panel A

---

[81] In fact, seven countries are significant for UIRP, three for PPP, one for the monetary and nine for the monetary ECM models. These results show that, for example, interest rates have been capable, at some point in time, to predict one-month-ahead exchange rates. According to the traditional Granger-causality test, reported in the not-for-publication appendix, traditional fundamentals have more limited predictive ability, significant only for a couple of countries for each model/predictor.

[82] PPP is significant for four countries, the other models/predictors perform much worse.

[83] MSFER is bigger than or equal to one.

in Table 5 shows that the monetary panel model displays some in-sample predictive ability (at some point in time) and almost inexistent out-of-sample forecasting ability at short horizons; interestingly, however, there is some evidence of out-of-sample forecasting ability at long horizons for four countries. Again, note the very different results for CW and DMW: once more, the latter never finds predictive ability. Panel B in Table 5 shows instead that the BMA is not significantly better than the random walk benchmark in forecasting out-of-sample at any horizon and for any test statistics.

Instabilities in Predictive Performance. Our results are broadly consistent with those in the literature, although in some cases they differ. For example, we find less out-of-sample forecasting ability at long horizons than other papers in the literature for the monetary model in error-correction form (ECM). Two of the possible reasons why our results may differ from those reported by other papers are that either the out-of-sample period is different, or the window size is different, or both. Let us consider each of these two effects separately. We report results based on the Giacomini and Rossi (2010) and the Inoue and Rossi (2012) tests in Figures 1 to 6.

Figures 1, 2, 5, and 6 plot Giacomini and Rossi's (2010) Fluctuation test statistic (solid line) along with its critical value (dotted line) for selected countries, models and forecast horizons.[84] The x-axis denotes time. The Fluctuation test examines how the forecast performance changed over time: whether the out-of-sample predictive ability of the traditional predictors ever showed up in the data, and, if so, when. For example, in Figure 1 the Fluctuation test (solid line) for Canada is never above its critical value (dotted line), thus the predictor never displays significant forecasting ability; on the other hand, the Fluctuation test does detect predictive ability for France in the late 2000s, Japan in mid-2000 and the UK in 2009. Overall, Figure 1 shows that monetary fundamentals have occasional and very short-lived predictive ability at the one-month horizon for some countries at some point in time. Figure 2 shows similar, occasional predictive ability at long horizons.[85]

A second reason why the results may differ is because of the use of a different estimation window size.[86] We study the behavior of the predictive ability as a function of the window size in Figures 3 and 4. The figures plot Inoue and Rossi's (2012) test statistic (solid line)

---

[84]The Fluctuation test is implemented with $m = 1/3$ in Giacomini and Rossi's (2010) notation.

[85]See Giacomini and Rossi (2010) for time variation in Taylor-rule fundamentals.

[86]In our exercise, we choose a rolling window with a size equal to half of the total sample size: while the fact that we select a window size equal to half of the total sample size is similar to what other papers have done, still our total sample is longer, and therefore our results may differ from those in the literature; furthermore, the sample is different too.

along with its critical value (horizontal line) for selected countries, models and forecast horizons. The x-axis denotes the window size, R.[87] For example, Figure 3 reports the DMW test calculated in rolling windows over the out-of-sample period for various window sizes R, reported on the x-axis. When the DMW test (solid line) is above Inoue and Rossi's (2012) critical values (dotted line), this signals predictive ability for that window size. Overall, Figures 3 and 4 show that indeed the window size strongly affects predictability for some countries.[88]

However, note that the out-of-sample period changes as we change the estimation window size: a larger estimation window size implies that a larger proportion of the sample is used for estimation and a smaller proportion is used for out-of-sample forecast evaluation. To analyze how the choice of the estimation window size and the out-of-sample period interact with each other, Figures 5 and 6 report the Fluctuation test for various window sizes for Germany and Japan, respectively. Very interesting findings emerge from the figures. Figure 5 shows that only small estimation window sizes can detect forecasting ability of monetary fundamentals in Germany, and their predictive power is concentrated in the late 1980s; Figure 6 shows instead that only large estimation window sizes can detect predictability of monetary fundamentals in Japan, and that the predictability emerges in the late 2000s.

Clearly, for some window sizes, it is possible to find some evidence of predictive ability for monetary fundamentals, thus confirming previous results that have been reported in the literature for selected window sizes and/or out-of-sample evaluation periods. At the same time, our results highlight the lack of robustness of these analyses to small changes in the procedures for forecast estimation and evaluation.

Summary

To summarize our results, most traditional predictors show in-sample forecasting ability, but none of them shows strong out-of-sample forecasting ability across all countries and tests. Turning to the most recent Taylor-rule predictors, they are among the most successful fundamentals out-of-sample at short horizons. Exploiting information on fundamentals using panels does not seem to help at short horizons, although it provides forecast improvements over the random walk at long horizons for a few countries. Figures 7 and 8 summarize the empirical findings. Figure 7 is a scatterplot of the p-values of Clark and West's (2006) test

---

[87]We implement the sup-type version of Inoue and Rossi's (2012) test.

[88]Figure 3 shows predictability at the one-month-ahead horizon for Switzerland, Japan and the UK: in the former, predictability is enhanced by a large window size, in the latter two it is the opposite. Figure 4 shows similar results for 48–month-ahead forecasts for Switzerland and Japan.

at short horizons (reported on the x-axis) and at long horizons (reported on the y-axis) for several predictors, denoted by different markers. Each point in the figure corresponds to a country/ predictor/ model combination. P-values smaller than 0.05 denote statistical significance at the 5% level. The figure shows that, for several countries, Taylor-rule fundamentals are concentrated in the left, mid-to-upper corner of the figure; therefore they are significant predictors at short horizons but not at long horizons. Conversely, the monetary panel model shows up on the right side of the figure; in some cases in the lower bottom and in several other cases on the top, right corner. Therefore, the monetary panel model is significant at long horizons (but not at short horizons) for some countries, and not significant at any horizon for several other countries. Figure 8 does the same for Rossi's (2005b) robust Granger-causality test. It shows that Taylor-rule and monetary panel models perform the best in-sample (the former at both horizons, the latter at short horizons only); however, comparing Figures 7 and 8 clearly shows that only the in-sample predictive ability of Taylor-rules survives out-of-sample.

Only when we allow the predictability to be varying over time using the Fluctuation test, we see a few episodes when the economic predictors' forecasting ability was stronger than that of the random walk benchmark even in traditional predictors, although they remain sporadic and short-lived in most countries. So, forecasting ability, when present, is an occasional and short-lived phenomenon.

8   Conclusions

To conclude, our analysis of the literature review has uncovered several stylized facts, which lead to five main conclusions.

First, a consensus in the literature is that the Taylor-rule and net foreign assets fundamentals have more out-of-sample predictive content than traditional fundamentals (such as interest rate, inflation, output and money differentials); in fact, monetary fundamentals at very long horizons and interest rate differentials at short horizons display forecasting ability according to some papers but not others. However, the disagreement in the literature is in the extent to which the former can explain the Meese and Rogoff puzzle.

Second, overall, among the model specifications considered in the literature, the most successful are linear specifications. Among them, the single-equation ECM and the panel ECM models are the most successful at long horizons, although there is disagreement among researchers about the degree of robustness of the results. Typically, but not always, for

single-equation linear models the predictor choice matters more than whether the researcher uses contemporaneous, realized or lagged fundamentals. Among the multivariate models, the most successful specification is the panel ECM.

Third, data transformations (such as de-trending, filtering and seasonal adjustment) may substantially affect predictive ability, and may explain some differences in results across studies. For example, the forecasting ability of the monetary model at long horizons is much weaker or completely disappears after estimating the cointegration parameters. Another important factor that, for some fundamentals, may affect predictive ability is using realized or real-time data.[89] For a given model and predictor, predictive ability seems also to depend on the choice of the country; on the other hand, with few exceptions, the frequency of the data and whether the realized or the forecasted fundamental is used do not seem to affect predictive ability.

Fourth, the choice of the benchmark, horizon, sample period and forecast evaluation method matters a lot. In particular, choosing an inappropriate benchmark model may over-state the empirical evidence in favor of the economic model's predictive ability: the random walk without drift is the toughest benchmark to beat. Empirical results depend on the forecast horizon,[90] the choice of the evaluation method,[91] and the forecast sample, as the performance of predictors is typically unstable over time.[92]

Fifth, on the one hand, the empirical analysis confirms several of the findings in the literature: several predictors display in-sample predictive ability for future exchange rates; however, only few predictors display out-of-sample forecasting ability: Taylor-rules at short horizons for several countries, and monetary panel models at long horizons for some countries. *On the other hand, our literature review and empirical evidence suggest slightly less out-of-sample predictive ability* in favor of some of the models and predictors used in the literature.[93] The reason is that *there are substantial instabilities in the models' forecasting performance:* the predictive power of fundamentals varies not only across countries, models and predictors, but also over time periods; it may appear in some periods and disappear in others.

---

[89]This is a concern for monetary fundamentals and less of a concern for Taylor-rule fundamentals.

[90]For example, the literature agrees that monetary fundamentals do not have predictive ability over short horizons, but disagrees on whether they might have over long horizons.

[91]For example, whether interest rate differentials outperform the random walk may depend on the test that is used in practice.

[92]For example, the out-of-sample forecasting ability of the panel ECM model is not robust to either the choice of the forecast window or the forecast period.

[93]For example, monetary fundamentals at long horizons.

Overall, although some predictors (Taylor-rule fundamentals and net foreign assets) do exhibit some predictive ability at short horizons, and others (monetary fundamentals, especially in panel models) reveal some predictive ability at long horizons, none of the predictors, models, or tests systematically find empirical support for superior exchange rate forecasting ability of a predictor for all models, countries and time periods: typically, when predictability appears, it does so occasionally for some countries and for short periods of time. Thus, Meese and Rogoff's (1983a,b) finding does not seem to be entirely and convincingly overturned. We also find another interesting puzzle: the predictive ability of the fundamentals is time-varying and occasional, yet existing time-varying parameter models are not successful in capturing it. Our findings lead to new challenges: why does the predictability of exchange rate models change over time? Is it possible to design ways to exploit instabilities and improve exchange rates' forecasts? An answer to these questions would require an answer to another important question: why are exchange rates poorly forecasted by economic models? In their papers, Meese and Rogoff (1983a,b) conjectured that sampling error, model misspecification and instabilities could potentially explain the poor forecasting performance of the economic models. Insights on the latter may be particularly relevant to understand and resolve the Meese and Rogoff puzzle.

9 References

Abhyankar, A., L. Sarno and G. Valente (2005), "Exchange Rates and Fundamentals: Evidence on the Economic Value of Predictability," *Journal of International Economics* 66(2) 325-348.

Adrian, T., E. Etula, and H.S. Shin (2011), "Risk Appetite and Exchange Rates," *mimeo*.

Alquist, R. and M.D. Chinn (2008), "Conventional and Unconventional Approaches to Exchange Rate Modelling and Assessment," *International Journal of Finance and Economics* 13, 2-13.

Amano, R.A. and S. van Norden (1993), "A Forecasting Equation for the Canada-US Dollar Exchange Rate," in: *The Exchange Rate and the Economy*, Proceedings of a conference held at the Bank of Canada.

Amano, R.A. and S. van Norden (1995), "Terms of Trade and Real Exchange Rates: the Canadian Evidence," *Journal of International Money and Finance* 14(1), 83-104.

Amano, R.A. and S. van Norden (1998a), "Oil prices and the Rise and Fall of the US Real Exchange Rate," *Journal of International Money and Finance* 17(2), 299-316.

Amano, R.A. and S. van Norden (1998b), "Exchange Rates and Oil Prices," *Review of International Economics* 6(4), 683-94.

Andersen, T., T. Bollerslev, F.X. Diebold and C. Vega (2003), "Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange," *American Economic Review* 93(1), 38-62.

Bacchetta, P., E. van Wincoop and T. Beutler (2010), "Can Parameter Instability Explain the Meese-Rogoff Puzzle?," in: L. Reichlin and K. West (eds.), *NBER International Seminar on Macroeconomics*, University of Chicago Press.

Backus, D., S. Foresi and C. Telmer (2001), "Affine Term Structure Models and the Forward Premium Anomaly," *Journal of Finance* 56, 279–304.

Baillie, R.T. and T. Bollerslev (1989), "Common Stochastic Trends in a System of Exchange Rates," *Journal of Finance* 44, 167–181.

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

Baillie, R.T. and T. Bollerslev (2000), "The Forward Premium Anomaly is Not as Bad as You Think," *Journal of International Money and Finance* 19(4), 471-488.

Bailliu, J. and M.R. King (2005), "What Drives Movements in Exchange Rates?", *Bank of Canada Review*, Autumn.

Balassa, B. (1964), "The Purchasing Power Parity Doctrine: A Reappraisal," *Journal of Political Economy* 72, 584-596.

Bekaert, B. and R. Hodrick (2001), "Expectations Hypotheses Tests," *Journal of Finance* 56(4), 1357-1394.

Berben, R.B. and D.J. van Dijk (1998), "Does the Absence of Cointegration Explain the Typical Findings in Long Horizon Regressions?," *Erasmus University Econometrics Institute, Report 9814.*

Berkowitz, J. and L. Giorgianni (2001), "Long-Horizon Exchange Rate Predictability?," *Review of Economics and Statistics* 83(1), 81-91.

Canova, F. (1993), "Modelling and Forecasting Exchange Rates with a Bayesian Time-Varying Coefficient Model," *Journal of Economic Dynamics and Control* 17(1/2), 233-61.

Carriero, A., G. Kapetanios and M. Marcellino (2009), "Forecasting Exchange Rates with a Large Bayesian VAR," *International Journal of Forecasting* 25, 400-417.

Cassel, G. (1918), "Abnormal Deviations in International Exchanges," *Economic Journal* 28,413–15.

Cerra, V. and S.C. Saxena (2010), "The Monetary Model Strikes Back: Evidence from the World," *Journal of International Economics* 81, 184-196.

Chen Y.C. and K.S. Rogoff (2003), "Commodity Currencies," *Journal of International Economics* 60, 133-169.

Chen, Y.C., K.S. Rogoff and B. Rossi (2010), "Can Exchange Rates Forecast Commodity Prices?," *Quarterly Journal of Economics* 125(3), 1145-1194.

Chen, Y.C. and B.K.P. Tsang (2011), "What Does the Yield Curve Tell Us About Exchange Rate Predictability?," *Review of Economics and Statistics,* forthcoming.

Cheung, Y.W., M.D. Chinn and A.G. Pascual (2005), "Empirical Exchange Rate Models of the Nineties: Are Any Fit to Survive?," *Journal of International Money and Finance* 24, 1150-1175.

Cheung, Y.W. and U.G. Erlandsson (2005), "Exchange Rates and Markov Switching Dynamics," *Journal of Business and Economic Statistics* 23, 314-320.

Cheung, Y. and K. Lai (2000), "On the Purchasing Power Parity Puzzle," *Journal of International Economics* 52, 321–330.

Chinn, M. (1991), "Some Linear and Nonlinear Thoughts on Exchange Rates," *Journal of International Money and Finance* 10, 214-230.

Chinn, M. (2008), "Nonlinearities, Business Cycles and Exchange Rates," *Economic Notes* 37(3), 219-239.

Chinn, M. (2009), "Comment on: Can Parameter Instability Explain the Meese-Rogoff Puzzle?," in: L. Reichlin and K. West (eds.), *NBER International Seminar on Macroeconomics*, University of Chicago Press.

Chinn, M. (2011), "Macro Approaches to Foreign Exchange Determination," *La Follette School Working Paper* 2011-013.

Chinn, M.D. and R. Meese (1995), "Banking on Currency Forecasts: Is Change in Money Predictable?," *Journal of International Economics* 38 (1-2), 161-78.

Chinn, M.D. and G. Meredith (2004), "Monetary Policy and Long-Horizon Uncovered Interest Parity," *IMF Staff Papers* 51(3).

Chinn, M.D. and M.J. Moore (2012), "Order Flow and the Monetary Model of Exchange Rates: Evidence from a Novel Data Set," *Journal of Money, Credit and Banking*, forthcoming.

Clarida, R., J. Gali and M. Gertler (1998), "Monetary Rules in Practice: Some International Evidence," *European Economic Review* 42, 1033-1067.

Clarida, R.H., L. Sarno, M.P. Taylor and G. Valente (2003), "The Out-of-sample Success of Term Structure Models as Exchange Rate Predictors: A Step Beyond," *Journal of International Economics* 60(1), 61-83.

Clarida, R.H. and M.P. Taylor (1997), "The Term Structure Of Forward Exchange Premiums And The Forecastability Of Spot Exchange Rates: Correcting The Errors," *Review of Economics and Statistics* 79(3), 353-361.

Clark, T.E. and M.W. McCracken (2001), "Tests of Equal Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics* 105, 85-110.

Clark, T.E. and K.D. West (2006), "Using Out-of-Sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis," *Journal of Econometrics* 135, 155-186.

Clark, T.E. and K.D. West (2007), "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models," *Journal of Econometrics* 138, 291-311.

Coletti, D. and S. Murchison (2002), "Models in Policy-Making," *Bank of Canada Review*.

Corradi, V. and N. Swanson (2006), "Predictive Density Evaluation," in: Elliott, G., C. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting* Vol. 1, Elsevier-North Holland, 100-134.

Croushore, D. and T. Stark (2001), "A Real-time Data Set for Macroeconomists," *Journal of Econometrics* 105(1), 111-130.

Croushore, D. and T. Stark (2003), "A Real-Time Data Set for Macroeconomists: Does

the Data Vintage Matter?," *Review of Economics and Statistics* 85(3), 605-617.

Della Corte, P., L. Sarno and G. Sestieri (2010), "The Predictive Information Content of External Imbalances for Exchange Rate Returns: How Much Is It Worth?," *Review of Economics and Statistics*, forthcoming.

Diebold, F.X., J. Hahn and A. Tay (1999), "Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns on Foreign Exchange," *Review of Economics and Statistics* 81, 661-673.

Diebold, F., J. Gardeazabal and K. Yilmaz (1994), "On Cointegration and Exchange Rate Dynamics," *Journal of Finance* 49 (2), 727– 735.

Diebold, F.X. and R.S. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13, 253-263.

Diebold, F.X. and J.A. Nason (1990), "Nonparametric Exchange Rate Prediction?," *Journal of International Economics* 28(3-4), 315-332.

Diebold, F.X. (2012), "Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests," *mimeo*, U. of Pennsylvania.

Dimand, R.W. (1999), "Irving Fisher and the Fisher Relation: Setting the Record Straight," *The Canadian Journal of Economics* 32 (3), 744-750.

Edge, R.M., M.T. Kiley and J.P. Laforte (2010), "A Comparison of Forecast Performance between Federal Reserve Staff Forecasts, Simple Reduced Form Models and a DSGE Model," *Journal of Applied Econometrics* 25(4), 720–754.

Ehrmann, M. and M. Fratzscher (2005), "Exchange Rates and Fundamentals: New Evidence from Real-time Data," *Journal of International Money and Finance* 24, 317-341.

Elliott, G. and A. Timmermann (2008), "Economic Forecasting," *Journal of Economic Literature* 46(1): 3–56.

Engel, C.M. (1994), "Can the Markov Switching Model Forecast Exchange Rates?," *Journal of International Economics* 36, 151– 165.

Engel, C.M. and J.D. Hamilton (1990), "Long Swings in the Exchange Rate: Are They in the Data and Do Markets Know It?," *American Economic Review* 80, 689-713.

Engel, C., N.C. Mark and K.D. West (2007), "Exchange Rate Models Are Not as Bad as You Think," in: *NBER Macroeconomics Annual 2007*, Acemoglu, D., Rogoff, K., Woodford, M., eds., University of Chicago Press, 381-441.

Engel, C., N.C. Mark and K.D. West (2009), "Factor Model Forecasts of Exchange Rates," *mimeo*.

Engel, C. and K.D. West (2004), "Accounting for Exchange-Rate Variability in Present-

Value Models When the Discount Factor is Near 1," *American Economic Review* 94(2), 119-125.

Engel, C. and K.D. West (2005), "Exchange Rate and Fundamentals," *Journal of Political Economy* 113, 485-517.

Engel, C. and K.D. West (2006), "Taylor Rules and the Deutschemark-Dollar Real Exchange Rate," *Journal of Money, Credit and Banking* 38, 1175-1194.

Fama, E. (1984), "Forward and Spot Exchange Rates," *Journal of Monetary Economics* 14, 319–338.

Faust, J., J.H. Rogers, S.Y.B. Wang and J.H. Wright (2007), "The High-Frequency Response of Exchange Rates and Interest Rates to Macroeconomic Announcements," *Journal of Monetary Economics* 54, 1051-1068.

Faust, J., J.H. Rogers and J.H. Wright (2003), "Exchange Rate Forecasting: the Errors We've Really Made," *Journal of International Economics* 60(1), 35-59.

Ferraro, D., K.S. Rogoff and B. Rossi (2011), "Can Oil Prices Forecast Exchange Rates?," *ERID Working Paper No. 95*, Duke University.

Fisher, I. (1896), *Appreciation and Interest*, New York: Macmillan for the American Economic Association; reprinted in: The Works of Irving Fisher, Vol. 1, W.J. Barber (ed.), London: Pickering & Chatto.

Frankel, J.A. (1982), "The Mystery of the Multiplying Marks: A Modification of the Monetary Model," *Review of Economics and Statistics* 64, 515-519.

Fratzscher, M. (2009), "What Explains Global Exchange Rate Movements During the Financial Crisis," *Journal of International Money and Finance* 28, 1390-1407.

Frankel, J.A. and A.K. Rose (1995), "Empirical Research on Nominal Exchange Rates," in: Grossman, G. and K.S. Rogoff (eds.), *Handbook of International Economics* Vol. 3, Elsevier-North Holland.

Frenkel, J.A. (1976), "A Monetary Approach to the Exchange Rate: Doctrinal Aspects and Empirical Evidence," *Scandinavian Journal of Economics* 78, 200-224.

Froot, K. and K. Rogoff (1995), "Perspectives on PPP and Long-run Real Exchange Rates," in: Grossman, G. and K.S. Rogoff (eds.), *Handbook of International Economics* Vol. 3, Elsevier-North Holland.

Froot, K. A. and R. H. Thaler (1990), "Anomalies: Foreign Exchange," *Journal of Economic Perspectives* 4, 179–192.

Giacomini, R. and B. Rossi (2010), "Forecast Comparisons in Unstable Environments," *Journal of Applied Econometrics* 25(4), 595-620.

Giacomini, R. and H. White (2006), "Tests of Conditional Predictive Ability," *Econometrica* 74(6), 1545-1578,

Giannone, D. (2009), "Comment on: Can Parameter Instability Explain the Meese-Rogoff Puzzle?," in: L. Reichlin and K. West (eds.), *NBER International Seminar on Macroeconomics*, University of Chicago Press.

Gospodinov, N. (2004), "Asymptotic Confidence Intervals for Impulse Responses for Near-Integrated Processes: An Application to Purchasing Power Parity," *Econometrics Journal* 7(2), 505-527.

Gourinchas, P.O. and H. Rey (2007), "International Financial Adjustment," *Journal of Political Economy* 115(4), 665-703.

Greenspan, A. (1994), "A Discussion of the Development of Central Banking," in: Capie, F., C. Goodhart, S. Fischer and N. Schnadt (eds.), *The Future of Central Banking*, Cambridge: Cambridge University Press.

Groen, J.J. (1999), "Long Horizon Predictability of Exchange Rates: is it for Real?," *Empirical Economics* 24, 451–469.

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

Groen, J.J. (2002), "Cointegration and the Monetary Exchange Rate Model Revisited," *Oxford Bulletin of Economics and Statistics* 64(4), 361-380.

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

Hooper, P. and J.E. Morton (1982), "Fluctuations in the Dollar: A Model of Nominal and Real Exchange Rate Determination," *Journal of International Money and Finance* 1, 39-56.

Husted, S. and R. MacDonald (1998), "Monetary-based Models of the Exchange Rate: a Panel Perspective," *Journal of International Financial Markets, Institutions and Money* 8(1), 1-19.

Imbs, J., H. Mumtaz, M. Ravn and H. Rey (2005), "PPP Strikes Back: Aggregation and the Real Exchange Rate," *Quarterly Journal of Economics* 120(1), 1-43.

Inoue, A. and B. Rossi (2008), "Monitoring and Forecasting Currency Crises," *Journal of Money, Credit and Banking* 40(2-3), 523-534.

Inoue, A. and B. Rossi (2012), "Out-of-Sample Forecast Tests Robust to the Choice of Window Size," *Journal of Business and Economic Statistics* 30(3), 432-453.

Ito, T. (1990), "Foreign Exchange Rate Expectations: Micro Survey Data," *American*

Economic Review, 80, 434-449.

Keynes, J.M. (1923), *A Tract on Monetary Reform*, London: Macmillan. Reprinted in: D.E. Moggridge and E.A.G. Robinson (eds.), *Collected Writings of John Maynard Keynes* Vol. 4, Cambridge University Press for the Royal Economic Society, 1971.

Kilian, L. (1999), "Exchange Rates and Monetary Fundamentals: What Do We Learn From Long-Horizon Regressions?," *Journal of Applied Econometrics* 14(5).

Kilian, L. and M.P. Taylor (2003), "Why Is It so Difficult to Beat the Random Walk Forecast of Exchange Rates?," *Journal of International Economics* 60(1), 85-107.

Kilian, L. and T. Zha (2002), "Quantifying the Uncertainty About the Half-Life of Deviations From PPP," *Journal of Applied Econometrics* 17, 107–125.

Koenig, E., S. Dolmas and J. Piger (2003), "The Use and Abuse of "Real-time" Data in Economic Forecasting," *Review of Economics and Statistics* 85(3), 618-628.

Lane P. and G.M. Milesi-Ferretti (2007), "The External Wealth of Nations II; Revised and Extended Estimates of Foreign Assets and Liabilities 1970–2004," *Journal of International Economics* 73(2), 223-250.

Lewis, K. (1995), "Puzzles in International Finance," in: Grossman, G.M. and K.S. Rogoff (eds.), *Handbook of International Economics* Vol. 3, Elsevier-North Holland, 1913-1949.

Lopez, C. C. Murray, and D. Papell (2005), "State-of-the-Art Unit Root Tests and the PPP Puzzle," *Journal of Money, Credit and Banking* 37, 361-69.

MacDonald, R. and M.P. Taylor (1993), "The Monetary Approach to the Exchange Rate: Rational Expectations, Long-run Equilibrium, and Forecasting," *IMF Staff Papers* 40.

MacDonald, R. and M.P. Taylor (1994), "The Monetary Model of the Exchange Rate: Long-run Relationships, Short-run Dynamics, and How to Beat a Random Walk," *Journal of International Money and Finance* 13, 276–290.

Marcellino, M., J.H. Stock and M.W. Watson (2006), "A Comparison of Direct and Iterated AR Methods for Forecasting Macroeconomic Series h-steps Ahead," *Journal of Econometrics* 135, 499-526.

Mark, N.C. (1995), "Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability," *American Economic Review* 85(1), 201-218.

Mark, N.C. (2001), *International Macroeconomics and Finance: Theory and Econometric Methods*, Malden, MA: Blackwell Publishers.

Mark, N.C. and D. Sul (2001), "Nominal Exchange Rates and Monetary Fundamentals: Evidence from a Small Post-Bretton Woods Sample," *Journal of International Economics*

53, 29-52.

Meese, R.A. and K.S. Rogoff (1983a), "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?," *Journal of International Economics* 14(1-2), 3-24.

Meese, R.A. and K.S. Rogoff (1983b), "The Out-of-Sample Failure of Empirical Exchange Rate Models: Sampling Error or Mis-specification?," in: *Exchange Rates and International Macroeconomics*, Jacob Frenkel, eds., Chicago: NBER and University of Chicago Press.

Meese, R.A. and K.S. Rogoff (1988), "Was it Real? The Exchange Rate-Interest Differential Relation Over the Modern Floating Rate Period," *Journal of Finance* 43, 923-948.

Meese, R.A. and A. Rose (1991), "An Empirical Assessment of Non-linearities in Models of Exchange Rate Determination," *Review of Economic Studies* 58, 603-619.

Melvin, M., J. Prins and D. Shand (2011), "Forecasting Exchange Rates: An Investor Perspective," in: Elliott, G. and A. Timmermann, *Handbook of Economic Forecasting Vol. 2*, Elsevier North-Holland.

Mincer, J. and V. Zarnowitz (1969), "The Evaluation of Economic Forecasts," in J. Mincer (ed.), *Economic Forecasts and Expectations*, New York: National Bureau of Economic Research.

Mizrach, B.M. (1992), "Multivariate Nearest-Neighbor Forecasts of EMS Exchange Rates," *Journal of Applied Econometrics* 7, S151– S163.

Molodtsova, T. and D.H. Papell (2009), "Out-of-Sample Exchange Rate Predictability with Taylor Rule Fundamentals," *Journal of International Economics*, 77(2), 167-180.

Molodtsova, T. and D.H. Papell (2012), "Taylor Rule Exchange Rate Forecasting During the Financial Crisis," *NBER International Seminar on Macroeconomics*.

Molodtsova, T., A. Nikolsko-Rzhevskyy and D.H. Papell (2010), "Taylor Rules and the Euro," *Journal of Money, Credit and Banking*, 535-552.

Murray, C. and D.H. Papell (2002), "The Purchasing Power Parity Persistence Paradigm," *Journal of International Economics* 56, 1–19.

Mussa, M. (1976), "A Model of Exchange Rate Dynamics," *Scandinavian Journal of Economics* 78, 229-248.

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

Neely, C.J. and S.R. Dey (2010), "A Survey of Announcement Effects on Foreign Exchange Returns," *Federal Reserve Bank of St. Louis Review* 92(5), 417-463.

Neely, C.J. and L. Sarno (2002), "How Well Do Monetary Fundamentals Forecast Exchange Rates?", *Federal Reserve Bank of St. Louis Review*, September-October.

Newey, W. and K.D. West (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, 703–708.

Obstfeld, M. and K.S. Rogoff (1996), "Foundations of International Macroeconomics," *Cambridge, MA: MIT Press.*

Patton, A. and A. Timmermann (2006), "Testing Forecast Optimality Under Unknown Loss," *Journal of American Statistical Association* 102(480), 1172-1184.

Patton, A. and A. Timmermann (2007), "Properties of Optimal Forecasts under Asymmetric Loss and Nonlinearity," *Journal of Econometrics* 140, 884-918.

Pesaran, M.H. and A. Timmermann (1992), "A Simple Nonparametric Test of Predictive Performance," *Journal of Business and Economic Statistics* 10(4), 461-465.

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

Rapach, D.E. and M.E. Wohar (2002), "Testing the Monetary Model of Exchange Rate Determination: New Evidence From a Century of Data," *Journal of International Economics* 58, 359– 385.

Rapach, D.E. and M.E. Wohar (2004), "Testing the Monetary Model of Exchange Rate Determination: A Closer Look at Panels," *Journal of International Money and Finance* 23(6), 867-895

Rapach, D.E. and M.E. Wohar (2006), "The Out-of-Sample Forecasting Performance of Nonlinear Models of Real Exchange Rate Behavior," *International Journal of Forecasting* 22(2), 341-361.

Rime D., L. Sarno and E. Sojli (2010), "Exchange Rate Forecasting, Order Flow and Macroeconomic Information," *Journal of International Economics* 80(1), 72-88.

Rogoff, K. (1996), "The Purchasing Power Parity Puzzle," *Journal of Economic Literature* XXXIV, 647-668.

Rogoff, K. and V. Stavrakeva (2008), "The Continuing Puzzle of Short Horizon Exchange Rate Forecasting," *NBER Working Paper* 14071.

Rossi, B. (2005a), "Testing Long-Horizon Predictive Ability with High Persistence, and the Meese-Rogoff Puzzle," *International Economic Review* 46(1), 61-92.

Rossi, B. (2005b), "Optimal Tests for Nested Model Selection with Underlying Parameter Instability," *Econometric Theory* 21(5), 962-990.

Rossi, B. (2005c), "Confidence Intervals for Half-Life Deviations from Purchasing Power Parity," *Journal of Business and Economic Statistics* 23(4), 432-442.

Rossi, B. (2006), "Are Exchange Rates Really Random Walks? Some Evidence Robust

to Parameter Instability," *Macroeconomic Dynamics* 10(1), 20-38.

Rossi, B. (2007a), "Expectations Hypotheses Tests and Predictive Regressions at Long Horizons," *Econometrics Journal* 10(3), 1-26

Rossi, B. (2011), "Advances in Forecasting Under Instabilities," in: Elliott, G. and A. Timmermann, *Handbook of Economic Forecasting* Vol. 2, Elsevier-North Holland.

Rossi, B. and T. Sekhposyan (2011), "Understanding Models' Forecasting Performance," *Journal of Econometrics* 164(1), 158-172.

Rossi, B. and T. Sekhposyan (2012), "Conditional Predictive Density Evaluation in the Presence of Instabilities", *mimeo.*

Samuelson, P. (1964), "Theoretical Notes on Trade Problems," *Review of Economics and Statistics* 46, 145-154.

Sarantis, N. (1994), "The Monetary Exchange Rate Model in the Long Run: An Empirical Investigation," *Weltwirtschaftliches Archiv* 130, 698–711.

Satchell, S. and A. Timmermann (1995), "An Assessment of the Economic Value of Non-linear Foreign Exchange Rate Forecasts", *Journal of Forecasting* 14(6), 477-497.

Schinasi, G.J. and P.A.V.B. Swamy (1987), "The Out-of-Sample Forecasting Performance of Exchange Rate Models when Coefficients are Allowed to Change," *Journal of International Money and Finance* 8(3), 375-390.

Stock, J.H. and M.W. Watson (1993), "A Simple Estimator of Cointegrating Vectors in Higher-Order Integrated Systems," *Econometrica* 61(4), 783-820.

Stock, J.H. and M.W. Watson (1998), "Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model," *Journal of the American Statistical Association* 93(441), 349-358.

Svensson, L.E.O. (2000), "Open Economy Inflation Targeting," *Journal of International Economics* 50, 155-183.

Taylor, A.M. (2002), "A Century of Purchasing Power Parity," *Review of Economics and Statistics* 84(1), pages 139-150.

Taylor, J.B. (1993), "Discretion Versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy* 39, 195-214.

Taylor, M. P. and D.A. Peel (2000), "Nonlinear Adjustment, Long-Run Equilibrium and Exchange Rate Fundamentals," *Journal of International Money and Finance* 19(1), 33-53.

Taylor, M.P. and L. Sarno (1998), "The Behavior of Real Exchange Rates During the Post-Bretton Woods Period", *Journal of International Economics* 46(2), 281-312.

Terasvirta, T. (2006), "Forecasting Economic Variables with Nonlinear Models," in G.

Elliott, C. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting* Vol. 1, Elsevier-North Holland.

Theil, H. (1966), *Applied Economic Forecasting*, North Holland.

Timmermann, A. (2006), "Forecast Combinations," in: Elliott, G., C. Granger and A. Timmermann, *Handbook of Economic Forecasting Vol. 1*, Elsevier-North Holland North-Holland.

Van Dijk, D. and P.H. Franses (2003), "Selecting a Nonlinear Time Series Model Using Weighted Tests of Equal Forecast Accuracy," *Oxford Bulletin of Economics and Statistics* 65, 727-744.

Wang, J. and J.J. Wu (2009), "The Taylor Rule and Interval Forecast for Exchange Rates," *Federal Reserve Board International Finance Discussion Papers* 963.

West, K.D. (1996), "Asymptotic Inference about Predictive Ability," *Econometrica* 64, 1067-1084.

West, K.D. (2006), "Forecast Evaluation," in: Elliott, G., C. Granger and A. Timmerman (eds.), *Handbook of Economic Forecasting* Vol. 1, Elsevier-North Holland, 100-134.

West, K.D., H.J. Edison and D. Cho (1993), "A Utility-based Comparison of Some Models of Exchange Rate Volatility," *Journal of International Economics* 35, 23-46.

West, K.D., and M.W. McCracken (1998), "Regression-Based Tests of Predictive Ability," *International Economic Review* 39, 817-840.

Wieland, W. and M.H. Wolters (2011), "Forecasting and Policy Making," in: Elliott, G. and A. Timmermann (eds.), *Handbook of Economic Forecasting*, Volume 2, Elsevier North Holland, forthcoming.

Wolff, C.P (1987), "Time-Varying Parameters and the Out-of-Sample Forecasting Performance of Structural Exchange Rate Models," *Journal of Business and Economic Statistics* 5(1), 87-97.

Wright, J. (2008), "Bayesian Model Averaging and Exchange Rate Forecasting," *Journal of Econometrics* 146, 329-341.

Tables and Figures

Table 1. Literature Review: Predictors and Economic Models

Predictors (f_t)	Economic Fundamentals	Mnemonics
$i_t - i_t^*$	Interest Rate Differentials	i
$F_t - s_t$	Forward Discount	F
$p_t - p_t^*$	(Log) Price Differentials	p
$\pi_t - \pi_t^*$	Inflation Differentials	π
$y_t - y_t^*$	(Log) Output Differentials	y
$m_t - m_t^*$	(Log) Money Differentials	m
z_t	Productivity Differentials	z
$b_t - b_t^*$	Asset Differentials	b
$y_t^{gap} - y_t^{gap*}$	Output Gap Differentials	y^{gap}
nxa_t	Net Foreign Assets	nxa
CP_t	Commodity Prices	CP

Model	f_t	Mnemonics
UIRP (CIRP)	$i_t - i_t^*;\ (F_t - s_t)$	i, F
PPP	$p_t - p_t^*$ or $\pi_t - \pi_t^*$	p, π
Monetary Model with Flexible Prices (I)	$\left[(i_t - i_t^*), (y_t - y_t^*), (m_t - m_t^*)\right]'$	i, y, m
Monetary Model with Flexible Prices (II) (or Frenkel-Bilson Model)	$\left[(y_t - y_t^*), (m_t - m_t^*)\right]'$	y, m
Monetary Model with Sticky Prices (I)	$\left[(i_t - i_t^*), (y_t - y_t^*), (m_t - m_t^*), (p_t - p_t^*)\right]'$	i, y, m, p
Monetary Model with Sticky Prices (II) (or Dornbush-Frankel Model)	$\left[(i_t - i_t^*), (y_t - y_t^*), (m_t - m_t^*), (\pi_t - \pi_t^*)\right]'$	i, y, m, π
Model with Productivity Differentials (or Balassa-Samuelson (1964) Model)	$\left[(i_t - i_t^*), (y_t - y_t^*), (m_t - m_t^*), z_t\right]'$	i, y, m, z
Portfolio Balance Model (or Hooper and Morton (1982) Model)	$\left[(i_t - i_t^*), (b_t - b_t^*)\right]'$	i, b
Taylor Rule Model	$\left[(\pi_t - \pi_t^*), (y_t^{gap} - y_t^{gap*})\right]'$	π, y^{gap}
Net Foreign Asset Model	nxa_t	nxa
Commodity Prices	CP_t	CP

Table 2. Literature Review: Models, Predictors and Data

| | A. Models and Predictors | | | | B. Data | | |
Reference	Linear	Eq.	Details	Predictors	Freq.	Sample	Countries
				Single Equation Models			
Abhyankar et al. (2005)	L	ECM	calibr. coeff.	m,y	M	1977:1-2000:12	CA,JP,UK
Alquist and Chinn (2008)	L	ECM	estim. coeff.	$i;(m,y,\pi,i)$	Q	1970:1-2005:Q4	CA,UK,EU
Bacchetta et al. (2010)	L	contemp.	real. fund.	m,y,u,i,CP	M	1975:9-2008:9	UK,SWI,JP,GER,AU,CA,NZ
Berkowitz and Giorgianni (2001)	L	ECM	calibr. coint.	m,y	Q	1973Q2-1991Q4	CA,GER,SWI,JP
Chen, Rogoff and Rossi (2010)	L	lag. fund.		ΔCP	Q	varies by country	AU,CAN,NZ,CHI,SA
Cheung et al. (2005)	L	diff.; ECM	estim. coint.	$i;p;(m,y,i,\pi)$;others(z,b)	Q	1973Q2-2000Q4	CAN,UK,GER,JP,SWI
Chinn (1991)	NL	levels,diff,ECM	non-parametric	$m,y,i;\pi$,wealth	Q	1974Q1-1988Q4	GER,JP
Chinn and Meese (1995)	L/NL	ECM/LWR	calibr. coeff.	m,y,i; others(π,TB,q)	M	1973:3-1990:11	GER,CA,UK,JP
Chinn and Moore (2012)	L	ECM	estim. coint.	order flow	M	1999:1-2001:1	EU,JP
Clark and West (2006)	L	lag fund.		i	M	varies by country	CAN,JP,SWI,UK
Della Corte et al. (2010)	L	lag fund.		nxa	Q (real t)	1973Q1-2007Q4	CAN,GER,UK,JP
Diebold and Nason (1990)	NL	univ.	LWR	Δs	W	1973-1987	10 countries
Engel and Hamilton (1990)	NL	MS		s,i	Q	1973Q3-1988Q1	GER,FR,UK
Engel (1994)	NL	MS		s	Q	1973-1991	18 countries
Engel, Mark and West (2009)	L	ECM factor		factor from s	Q	1973-2007	17 countries
Faust et al. (2003)	L	ECM	calibr. coint.	$y,m;(m,y,\pi,i,b)$	Q (real t)	1973Q2-1991Q4	GER,JP,CAN,SWISS
Ferraro et al. (2011)	L	lag/contemp.		ΔCP	D (real t)	1984:12-2010-11	CA,NOR,SA,AUS,CHI
Giacomini and Rossi (2010)	L	lag fund.		π, y^{gap}	M	1973:1-2008:1	12 countries
Gourinchas and Rey (2007)	L	lag fund.		nxa	Q	1973:Q1-2004Q1	US effective exch.
Groen (1999)	L	ECM		m,y	M	1973:4-1994:12	GER,FR,NET,CA
Inoue and Rossi (2012)	L	ECM		$\pi, y^{gap}, ; i$	M	1973:3-2008:1	10 countries
Kilian (1999)	L	ECM	calibr. coint.	m,y	Q	1973Q2-1991Q4	CAN,GER,SWI,JP
Mark (1995)	L	ECM	calibr. coint.	m,y	Q	1973Q2-1991Q4	CAN,GER,SWI,JP
Meese and Rogoff (1983a)	L	levels	OLS,GLS,IV	(m,y,i);others(π,TB,F)	M	1973:3-1981:6	UK,GER,JP,trade-w.
Meese and Rogoff (1983b)	L	contemp.	calib. coeff.	(m,y,i);others(π,TB)	M	1973:3-1981:6	UK,GER,JP,trade-w.
Meese and Rose (1991)	NL	lag/contemp.	LWR	(m,y,p);others(TB)	M	1974-1987	CA,GER,JP,UK
Molodtsova et al. (2010)	L	lag fund.	neural net.	π, y^{gap}	Q (real t)	1991Q1-2007Q3	EU
Molodtsova and Papell (2009)	L	lag fund.		$\pi, y^{gap}; m,y; i; p$	M	varies by country	12 countries
Molodtsova and Papell (2012)	L	lag fund.		π, y^{gap}, FCI	Q (real t)	1992Q4-2012Q1	EU
Qi and Wu (2003)	NL	ECM	neural net.	m,y,i	M	1973:3-1997:7	JP,UK,CA,GER
Rapach and Wohar (2006)	NL	univ.	ESTAR	s	M	1980:1-1994:12	UK,GER,FR,JP
Rime, Sarno and Sojli (2010)	L	ECM		order flows	D	2004:2-2005:2	UK,EU,JP
Rogoff and Stavrakeva (2008)	L	ECM		$(m,y);(\pi,y^{gap})$;nxa	Q	1973:1-2005:4	18 countries
Rossi (2005a)	L	ECM		m,y	M	1973Q3-1998Q2	GER,JP
Rossi (2006)	L/NL	lag fund.	TVP	$\Delta y, \Delta m, i$	M	1973:3-1998:12	CA,FR,GER,IT,JP
Rossi and Sekhposyan (2011)	L	lag fund.		$\Delta m, \Delta u, \Delta y, \Delta i, \Delta CP$	M	1975:9-2008:9	SWI,UK,CA,JP,GER
Schinasi and Swami (1987)	L	contemp.	TVP-AR	m,y,i,TB	M	1973:3-1980:3	UK,GER,JP
West, Edison and Cho (1993)	L/NL	univ.	GARCH/NP	s	W	1973:3-1989:9	CA,FR,GER,JP,UK
Wolff (1987)	L	contemp.	TVP-AR	$m,y,i;\pi$	M	1973:3-1984:4	GER,JP,UK
Wu and Wang (2009)	L	lag fund.	semi-param.	$(\pi,y^{gap},q);(m,y)$	M	varies by country	12 countries
				Multiple Eqs. Models			
Canova (1993)	L	TVP-VAR		i	W	1979 to 1987	FR,SWI,GER,UK,JP
Clarida et al. (2003)	NL	VECM-MS		F	W	1979:1-1998:52	FR,GER,JP,UK
Clarida and Taylor (1997)	L	VECM		F	W	1977:1-1993:52	UK,GER,JP
Diebold et al. (1994)	L	VECM		s	D	3/1/1980-1/28/1985	CAN,FR,GER,IT,JP,SWI,UK
Diebold, Hahn and Tay (1999)	L	density		s	30 min.	1996:1-1996:12	GER,JP
MacDonald and Taylor (1993)	L	VECM		m,y,i	M	1976:1-1990:12	GER
Mizrach (1992)	NL	LWR		Δs	D	1986:1-1988:12	FR,IT,GER
Rapach and Wohar (2002)	L	VECM		$(m,y); i$	Y	115 years	14 countries
Wright (2008)	L	contemp/lag	BMA	p,y,i,z,m,SP,D,CA	M,Q (real t)	1973-2005	CA,GER,JP,UK
				Panel Models			
Adrian, Etula and Shin (2011)	L	lag. fund.		repo;comm.paper	M	1993:1-2007:12	23 countries
Carriero et al. (2009)	NL	large BVAR		s	M	1995:1-2008:4	33 countries
Cerra and Saxena (2010)	L	level,demean		m,y	Y (real t)	1960-2004	98 countries
Engel, Mark and West (2007)	L	VECM	ECM; diff.	$(m,y,p); (\pi, y^{gap})$	Q	1973-2005	18 countries
Groen (2005)	L	VECM	calibr. coeff.	m,y	Q	1975Q1-2000Q4	CA,JP,GER
Mark and Sul (2001)	L	ECM	calibr. coint.	$m,y;p$	Q	1973Q1-1997Q1	19 countries
Rapach and Wohar (2004)	L	ECM	calibr. coint.	m,y	Q	1973Q1-1997Q1	18 countries

Table 2 (continued). Forecast Evaluation and Empirical Conclusions

Reference	Benchmark	Estim.	h	Forec.	Eval. Method	D. Successful Predictive Ability?
Single Equation Models						
Abhyankar et al. (2005)	VAR		3-48 mo.	1991:1	utility	Yes
Alquist and Chinn (2008)	RW	roll	1,4,20 qrs.	1987Q2	CW	No (except UIRP at long h and NFA for some countries at short h)
Bacchetta et al. (2010)	RW	roll	1 mo.	varies	MSFE	No
Berkowitz and Giorgianni (2001)	RW	roll	1-16 qrs.	1981Q4	R2,DMW	No
Chen, Rogoff and Rossi (2010)	RW,AR	roll	1 q.		ENCNEW	Yes
Cheung et al. (2005)	RW	roll(42,59)	1,4,20 qrs	1987Q2	MSE,sign	No
Chinn (1991)	RW	roll(48)	1, 4 qrs.	1986Q1		No
Chinn and Meese (1995)	RW		1,12,36 mo.	1983:01		No (except monetary at long h for some countries)
Chinn and Moore (2012)	RW		1-3-6 mo.	3 years	CW	Yes (but significant only for one country)
Clark and West (2006)	RW	roll(120)	1 mo.	varies	DMW,CW	Yes (for UIRP at short horizons)
Della Corte et al. (2010)	RW	roll(80)		1993Q1	utility-based	Yes (for NXA)
Diebold and Nason (1990)	RW		1-4 qrs. (iter.)	1986W1	MSE	No
Engel and Hamilton (1990)	RW		1-4 qrs. (iter.)	1984Q1		Yes
Engel (1994)	RW,F		1-12 qrs.	1986	DMW	Yes
Engel, Mark and West (2009)	RW,Taylor	roll(40)	1-16 qrs.	1999		Yes, somewhat (at long h)
Faust et al. (2003)	RW	roll	1 day			No
Ferraro et al. (2011)	RW,RWWD	roll	1 mo.	varies	DMW	Yes
Giacomini and Rossi (2010)	RW	roll(50)	1-24 qrs.	1983:2	GR,CW	Somewhat (it depends on time period)
Gourinchas and Rey (2007)	RW		1-48 mo.	1978Q2	CW	Yes
Groen (1999)	RW	rec	1 mo.	1981:10	DMW	No
Inoue and Rossi (2012)	RW	roll	1-16 qrs.	varies		Yes
Kilian (1999)	RW	rec	1-16 qrs.	1981Q4	R2,DMW	No
Mark (1995)	RW	rec	1,6,12 mo.	1981Q4	R2,DMW	Yes (only at long h)
Meese and Rogoff (1983a)	RW,VAR	roll(93)	1,6,12 mo.	1976:11	MSFE,MAE	No
Meese and Rogoff (1983b)	RW,VAR		48 mo.	1976:11	MSFE,MAE	No
Meese and Rose (1991)	RWWD	rec	1-4 qrs.	1984:1	MSFE	Yes
Molodtsova et al. (2010)	RW	roll(26)	1 mo.	1999Q4	CW	Yes for Taylor; no for UIRP and monetary
Molodtsova and Papell (2009)	RW	roll(120)	1 mo.	1982:3	CW	Yes
Molodtsova and Papell (2012)	RW	roll(,75T)	1 q.	2007Q1	DMW	No
Qi and Wu (2003)			1,6,12 mo. (iter)	1990:1	DMW	No
Rapach and Wohar (2006)			1 to 3 mo.	2004:6	DMW,PIT	Yes
Rime, Sarno and Sojli (2010)			1 day	varies	utility-based	Mostly no
Rogoff and Stavrakeva (2008)			varies	varies	various	Yes, based on GC-robust tests; somewhat; based on TVP models
Rossi (2005a)		roll,rec	1-12qrs.	1986	DMW	Yes (for some countries)
Rossi (2006)	RW,RWWD	roll	1 mo.	varies	DMW,ENCNEW	No
Rossi and Sekhposyan (2011)	RW	rec	1-12mo.	1980:4	DMW	Yes (but no statistical significance analysis)
Schinasi and Swami (1987)			15 mo.	1981:7	MSFE,MAE	Yes
West, Edison and Cho (1993)	RW	rec	1-24 mo. (iter)		utility-based	Somewhat (only for one country)
Wolff (1987)	RW	rec	1-12 mo. (iter)		MSFE,MAE	Yes
Wu and Wang (2009)	RW	roll(200)			interval,GW	Yes
Multiple Eqs. Models						
Canova (1993)	RW	roll,rec	1-52 w. (iter)	1986	Theil	Yes
Clarida et al. (2003)	RW	roll	4-52 w.	1996:1	DMW	Yes
Clarida and Taylor (1997)	RW	rec	1-12 mo.	1977:1	MSFE,MAE	Yes
Diebold et al. (1994)	absolute		1-26 days			No
Diebold, Hahn and Tay (1999)	absolute			1989-1990	PIT	n.a.
MacDonald and Taylor (1993)	RW	rec	1-12 mo. (iter)	1974:1	MSFE	Yes
Mizrach (1992)	RW		5 days		MSFE,MAE	No
Rapach and Wohar (2002)	RW	rec			ENCNEW	Somewhat (depends on country)
Wright (2008)	RW	rec	3-12 mo.	1990	MSFE	Yes (although magnitude of success is marginal)
Panel Models						
Adrian, Etula and Shin (2011)	RW,AR(1)	roll(48)	1-12 mo.	1997:1	CW	Yes for advanced countries, somewhat for emerging countries
Carriero et al. (2009)	RW	roll(84)	1-12 mo.	2001:1	GW	Yes
Cerra and Saxena (2010)	RW,RWWD	rec	1-5 yrs.	1984	various	Yes
Engel, Mark and West (2007)	RW	rec	1-16 qrs.	1983	CW	No
Groen (2005)	RW	rec	1-16 (iter)	1989Q1	bootstrap	Yes (but not always)
Mark and Sul (2001)	RW	rec	1, 16 qrs.	1983Q2	MSFE,Theil	Yes (at long horizons)
Rapach and Wohar (2004)	heter.	rec(50)	1-16 qrs.			n.a.

Table 3. Available Sample Sizes, By Country and Model

Models in Monthly Data

Country	Exchange Rate Differential			Interest Rate Differential			CPI Differential			Money Differential			Output Differential		
	Start	End	N.Obs.	Start	End	N.Obs.	Start	End	N.Obs.	Start	End	N.Obs.	Start	End	N.Obs.
Australia	1957:02	2011:06	653	1969:08	2011:06	503	n.a.	n.a.	n.a.	1975:03	2011:06	436	1957:02	2007:09	608
Austria	1957:02	1998:12	503	1967:02	1998:12	383	1995:02	2011:06	197	1997:10	2011:06	165	1957:02	2011:05	652
Belgium	1957:02	1998:12	503	1957:02	1999:01	504	1957:02	2011:06	653	2002:02	2011:06	113	1957:02	2011:03	650
Canada	1957:02	2011:06	653	1975:02	2011:06	437	1957:02	2011:06	653	1975:04	2011:06	435	1957:02	2011:05	652
Denmark	1957:02	2011:06	653	1972:02	2011:06	469	1980:02	2011:06	377	1991:02	2011:06	245	1957:02	2011:05	652
Finland	1957:02	1998:12	503	1978:01	2011:06	402	1960:02	2011:06	617	1960:02	2011:06	617	1957:02	2011:06	653
France	1957:02	1998:12	503	1964:02	1999:03	418	1990:02	2011:06	257	1978:01	2011:06	402	1957:02	2011:05	652
Germany	1957:02	1998:12	503	1960:02	2011:06	617	1957:02	2011:06	653	1973:02	2011:06	461	1958:02	2011:05	640
Greece	1957:02	2000:12	527	1998:02	1999:10	021	1959:02	2011:06	629	1980:02	2011:06	377	1957:02	2011:05	652
Ireland	1957:02	1998:12	503	1972:04	2011:06	450	1969:02	2011:06	509	1999:02	2011:06	149	1976:02	2011:05	424
Italy	1957:02	1998:12	503	1971:02	2011:06	485	1957:02	2011:06	653	1980:02	2011:06	377	1957:02	2011:05	652
Japan	1957:02	2011:06	653	1957:02	2011:06	653	2000:02	2011:06	137	1960:02	2011:06	617	1957:02	2011:06	653
New Zealand	1957:02	2011:06	653	1985:02	2011:06	317	n.a.	n.a.	n.a.	1977:04	2011:06	411	1987:02	2007:12	251
Portugal	1957:02	1998:12	503	1983:02	2000:03	206	1960:02	2011:06	617	1997:10	2011:06	165	1957:02	2006:04	591
Spain	1957:02	1998:12	503	1974:02	1998:12	449	1957:02	2011:06	653	1997:10	2011:06	165	1961:02	2011:05	604
Sweden	1957:02	2011:06	653	1966:01	2011:06	546	1960:02	2011:06	617	1998:02	2011:06	161	1957:02	2011:05	652
Switzerland	1957:02	2011:06	653	1975:10	2011:06	429	1957:02	2011:06	653	1985:01	2011:06	318	1995:02	2007:12	155
UK	1957:02	2011:06	653	1972:02	2011:06	473	1988:02	2011:06	281	1986:10	2011:06	297	1957:02	2011:05	652

Models in Quarterly Data

Country	Int. Rate Diff.			CPI Diff.			TB Diff.			Output Diff.			CA Diff.		
	Start	End	N.Obs.	Start	End	N.Obs.	Start	End	N.Obs.	Start	End	N.Obs.	Start	End	N.Obs.
Australia	1969:04	2011:02	167	1957:02	2011:02	217	1973:02	2010:04	151	1959:04	2011:02	207	1973:02	2010:04	151
Austria	1967:02	1998:04	127	1957:02	2011:02	217	1973:02	2011:01	152	1988:02	2011:01	092	1973:02	2011:01	152
Belgium	1957:02	1998:04	167	1957:02	2011:02	217	2002:02	2011:01	036	1980:02	2011:02	125	2002:02	2011:01	036
Canada	1975:02	2011:02	145	1957:02	2011:02	217	1973:02	2011:01	152	1961:02	2011:02	201	1973:02	2011:01	152
Denmark	1972:02	2011:02	155	1980:02	2011:02	125	1975:02	2011:01	144	1991:02	2011:02	081	1975:02	2011:01	144
Finland	1978:02	2011:02	133	1960:02	2011:02	205	1975:02	2011:01	144	1975:02	2011:02	145	1975:02	2011:01	144
France	1957:02	1999:01	168	1990:02	2011:02	085	1975:02	2010:04	143	1957:02	2011:02	217	1975:02	2010:04	143
Germany	1957:02	2011:02	217	1957:02	2011:02	217	1973:02	2011:01	152	1991:02	2011:02	081	1973:02	2011:01	152
Greece	1998:02	1999:03	006	1959:02	2011:02	209	1976:02	2011:01	135	2000:02	2011:01	044	1976:02	2011:01	135
Ireland	1969:02	2011:02	147	1969:02	2011:02	169	1981:02	2011:01	120	1997:02	2011:01	056	1981:02	2011:01	120
Italy	1971:02	2011:02	161	1957:02	2011:02	217	1973:02	2011:01	152	1981:02	2011:02	121	1973:02	2011:01	152
Japan	1957:02	2011:02	217	2000:02	2011:02	045	1977:02	2011:01	136	1980:02	2011:02	125	1977:02	2011:01	136
New Zealand	1985:02	2011:02	105	1957:02	2011:02	217	1980:02	2011:01	124	1987:03	2011:01	095	1980:02	2011:01	124
Portugal	1981:02	2000:01	076	1960:02	2011:02	205	1975:02	2011:01	144	1995:02	2011:02	065	1975:02	2011:01	144
Spain	1974:02	2011:02	149	1957:02	2011:02	217	1975:02	2011:01	144	1995:02	2011:02	065	1975:02	2011:01	144
Sweden	1966:02	2011:02	181	1957:02	2011:02	205	1975:02	2011:01	144	1993:02	2011:02	073	1975:02	2011:01	144
Switzerland	1976:01	2011:02	142	1957:02	2011:02	217	1999:02	2011:01	048	1980:02	2011:02	125	1999:02	2011:01	048
UK	1972:02	2011:02	157	1988:02	2011:02	093	1973:02	2011:01	152	1957:02	2011:02	217	1973:02	2011:01	152

Table 4. Relative Models' Forecasting Ability (Single Eq. Models)

h=1 Month (h=1 Quarter for Portfolio Balance)

Country	A. UIRP Model GC	RMSFE	DMW	CW	B. PPP Model RMSFE	DMW	CW	GC	C. Monetary Model RMSFE	DMW	CW	GC	D. Monetary ECM Model RMSFE	DMW	CW	GC	E. Taylor Model GC	RMSFE	DMW	CW	F. Portfolio Balance RMSFE	DMW	CW	GC
Australia	0.44	1.00	0.52	0.58	n.a.	n.a.	n.a.	n.a.	1.00	0.51	0.27	0.06	1.01	0.55	0.89	0.00	0.51	1.00	0.51	0.04	1.06	0.56	0.31	0.66
Austria	0.10	1.01	0.53	0.48	1.07	0.74	0.99	0.27	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0.01	1.00	0.47	0.06	1.06	0.61	0.77	1.00
Belgium	0.26	1.01	0.54	0.56	1.01	0.56	0.91	1.00	n.a.	n.a.	n.a.	n.a.	1.00	0.57	0.99	0.01	0.02	1.00	0.46	0.36	n.a.	n.a.	n.a.	n.a.
Canada	0.00	1.00	0.51	0.40	0.99	0.45	0.00	0.09	1.01	0.54	0.88	1.00	1.00	0.52	0.50	0.05	1.00	0.99	0.48	0.04	1.03	0.55	0.65	0.30
Denmark	1.00	1.01	0.56	0.95	1.00	0.52	0.55	0.11	1.01	0.56	0.58	0.34	1.00	0.51	0.35	0.00	0.08	1.00	0.50	0.38	1.04	0.59	0.72	0.00
Finland	0.63	1.02	0.57	0.86	1.00	0.53	0.76	0.50	1.01	0.55	0.72	0.38	1.01	0.54	0.50	0.39	0.00	0.99	0.46	0.02	1.08	0.64	0.84	0.00
France	0.78	1.02	0.54	0.75	1.02	0.59	0.88	0.42	1.01	0.55	0.73	0.72	1.00	0.54	0.76	0.02	0.21	1.00	0.43	0.04	1.27	0.70	0.77	0.00
Germany	0.00	1.01	0.53	0.47	1.00	0.52	0.57	0.82	1.01	0.53	0.34	0.37	1.00	0.52	0.53	0.00	0.28	1.00	0.45	0.08	1.09	0.62	0.85	0.01
Greece	0.00	1.00	0.51	n.a.	1.00	0.49	0.02	0.13	0.99	0.48	0.00	0.54	1.00	0.52	0.53	0.14	0.00	1.00	0.67	0.66	n.a.	n.a.	n.a.	n.a.
Ireland	0.00	1.03	0.54	0.76	1.01	0.54	0.58	0.00	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0.00	1.01	0.54	0.10	1.12	0.58	0.45	0.00
Italy	0.28	1.02	0.55	0.80	1.01	0.55	0.76	0.00	1.00	0.51	0.28	0.08	1.01	0.56	0.63	0.00	0.14	1.00	0.52	0.37	1.07	0.58	0.77	0.08
Japan	0.05	1.01	0.54	0.81	1.01	0.51	0.31	0.47	1.01	0.52	0.25	0.26	1.00	0.52	0.59	0.06	0.45	1.00	0.50	0.06	1.24	0.65	0.31	0.00
New Zealand	1.00	1.01	0.53	0.53	n.a.	n.a.	n.a.	n.a.	1.00	0.50	0.32	0.01	1.01	0.57	0.94	0.00	0.26	0.99	0.46	0.49	1.04	0.56	0.50	0.02
Portugal	0.00	1.02	0.58	0.91	1.01	0.56	0.77	0.00	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0.01	1.01	0.50	0.00	1.15	0.63	0.76	0.00
Spain	0.74	1.02	0.56	0.81	1.01	0.58	0.97	0.07	1.04	0.62	0.98	0.17	0.99	0.46	0.13	0.00	0.10	1.01	0.49	0.07	1.04	0.55	0.44	0.00
Sweden	0.00	1.04	0.53	0.88	1.00	0.50	0.30	0.19	1.01	0.58	0.91	0.78	1.02	0.63	0.99	0.02	0.00	1.01	0.52	0.71	1.03	0.55	0.38	0.00
Switzerland	0.45	1.01	0.55	0.84	1.00	0.50	0.38	0.36	1.01	0.58	0.38	0.02	1.01	0.57	0.26	0.09	0.21	1.00	0.51	0.26	1.12	0.57	0.09	0.00
UK	1.00	1.01	0.58	1.00	1.01	0.54	0.70	1.00	1.01	0.58	0.89	0.61	1.01	0.55	0.61	0.20	0.19	1.00	0.50	0.58	1.03	0.57	0.29	0.20

h=4 Years

Country	A. UIRP Model GC	RMSFE	DMW	CW	B. PPP Model RMSFE	DMW	CW	GC	C. Monetary Model RMSFE	DMW	CW	GC	D. Monetary ECM Model RMSFE	DMW	CW	GC	E. Taylor Model GC	RMSFE	DMW	CW	F. Portfolio Balance RMSFE	DMW	CW	GC
Australia	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	1.00	0.52	0.29	0.21	1.01	0.57	0.93	0.57	0.21	1.01	0.53	0.72	1.03	0.60	0.86	0.05
Austria	0.53	1.00	0.53	0.67	1.01	0.58	0.97	0.47	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0.18	1.01	0.50	0.60	1.05	0.59	0.11	0.00
Belgium	1.00	1.01	0.53	0.50	1.01	0.52	0.40	0.28	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0.04	1.01	0.54	0.54	n.a.	n.a.	n.a.	n.a.
Canada	0.43	1.01	0.53	0.36	1.00	0.51	0.44	0.72	1.00	0.53	0.73	0.50	1.00	0.51	0.38	0.50	0.86	1.00	0.46	0.18	1.03	0.57	0.65	0.26
Denmark	0.71	1.00	0.52	0.43	1.00	0.55	0.84	0.83	1.02	0.57	0.70	0.29	1.02	0.57	0.71	0.29	0.05	1.00	0.51	0.14	1.04	0.59	0.59	0.00
Finland	0.31	1.01	0.55	0.78	1.00	0.61	0.78	0.53	1.01	0.55	0.50	0.79	1.01	0.54	0.66	0.79	0.00	1.02	0.53	0.41	1.09	0.65	0.77	0.26
France	1.00	1.02	0.57	0.76	1.04	0.54	0.69	0.04	1.02	0.55	0.73	0.51	1.02	0.55	0.80	0.51	0.12	1.02	0.55	0.75	1.21	0.65	0.65	0.01
Germany	0.82	1.01	0.54	0.60	1.01	0.49	0.03	0.22	1.01	0.54	0.53	0.22	1.01	0.56	0.86	0.48	0.36	1.01	0.51	0.57	1.11	0.61	0.60	0.01
Greece	0.77	1.00	0.53	0.49	1.00	0.56	0.83	0.40	1.00	0.49	0.04	0.43	1.00	0.51	0.16	0.78	n.a.	1.00	0.49	0.15	n.a.	n.a.	n.a.	n.a.
Ireland	1.00	1.02	0.56	0.85	1.01	0.51	0.36	0.04	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0.00	1.01	0.53	0.15	1.11	0.71	0.52	0.00
Italy	1.00	1.02	0.56	0.70	1.00	0.57	0.64	0.22	1.04	0.62	0.97	0.35	1.03	0.59	0.80	0.31	0.17	1.00	0.53	0.31	1.06	0.56	0.42	0.00
Japan	0.64	1.01	0.55	0.85	1.02	0.53	0.47	0.00	1.01	0.52	0.36	0.13	1.01	0.53	0.80	0.39	0.45	1.00	0.50	0.40	1.09	0.61	0.22	0.00
New Zealand	1.00	1.01	0.55	0.71	n.a.	n.a.	n.a.	n.a.	1.01	0.54	0.72	0.22	1.00	0.52	0.28	0.89	0.53	1.03	0.58	0.76	1.06	0.63	0.90	0.02
Portugal	1.00	1.02	0.58	0.84	1.01	0.53	0.63	0.05	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0.00	1.12	0.53	0.78	1.06	0.59	0.28	0.00
Spain	0.70	1.02	0.55	0.61	1.00	0.53	0.46	0.59	1.02	0.59	0.62	0.80	1.02	0.59	0.81	1.00	0.00	1.08	0.49	0.16	1.06	0.58	0.41	0.00
Sweden	0.44	1.04	0.54	0.75	1.00	0.52	0.40	0.13	1.01	0.54	0.54	0.39	1.00	0.52	0.27	1.00	0.00	1.00	0.53	0.79	1.04	0.59	0.69	0.02
Switzerland	0.19	1.01	0.53	0.46	1.00	0.52	0.40	0.39	1.00	0.54	0.54	0.39	1.00	0.52	0.27	1.00	0.76	n.a.	0.56	0.74	1.04	0.60	0.60	0.00
UK	0.24	1.01	0.57	0.92	1.00	0.50	0.21	0.63	1.00	0.54	0.67	0.56	1.01	0.54	0.68	1.00	0.32	1.00	0.47	0.15	1.07	0.63	0.91	0.61

Table 5. Relative Models' Forecasting Ability (Multiple Eqs. Models)

Country	A. Monetary Panel Model			B. BMA Model		
	RMSE	DMW	CW	RMSE	DMW	CW
	h=1 Month			h=1 Month		
	h=1 Month					
Australia	1.01	0.53	0.34	1.00	0.53	0.72
Austria	1.30	0.90	n.a.	n.a.	n.a.	n.a.
Belgium	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Canada	1.01	0.53	0.63	1.00	0.53	0.83
Denmark	1.02	0.61	0.99	1.00	0.53	0.66
Finland	1.01	0.55	0.78	1.01	0.57	0.91
France	1.00	0.52	0.31	1.00	0.51	0.38
Germany	1.01	0.53	0.46	1.00	0.51	0.55
Greece	1.00	0.50	0.02	1.09	0.67	n.a.
Ireland	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Italy	1.00	0.51	0.29	1.00	0.53	0.59
Japan	1.01	0.53	0.62	1.00	0.49	0.21
New Zealand	1.01	0.54	0.44	1.00	0.54	0.78
Portugal	1.13	0.71	n.a.	n.a.	n.a.	n.a.
Spain	1.01	0.56	n.a.	n.a.	n.a.	n.a.
Sweden	1.00	0.54	0.69	1.01	0.57	0.89
Switzerland	1.03	0.61	0.97	1.00	0.52	0.56
U.K.	1.01	0.55	0.75	1.00	0.52	0.53
	h=4 Years					
Australia	1.35	0.70	0.31	1.00	0.48	0.18
Austria	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Belgium	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Canada	1.34	0.70	1.00	1.00	0.53	0.84
Denmark	1.49	0.76	1.00	1.01	0.55	0.83
Finland	1.84	0.80	0.00	1.01	0.55	0.87
France	1.72	0.78	1.00	1.00	0.54	0.79
Germany	1.75	0.76	1.00	1.00	0.52	0.50
Greece	1.14	0.60	0.00	n.a.	n.a.	n.a.
Ireland	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Italy	1.11	0.55	0.00	1.00	0.54	0.73
Japan	1.79	0.74	0.50	1.00	0.52	0.59
New Zealand	1.52	0.69	1.00	1.00	0.49	0.27
Portugal	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Spain	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Sweden	1.10	0.61	0.02	1.00	0.49	0.21
Switzerland	1.27	0.77	0.87	1.00	0.49	0.20
U.K.	1.60	0.82	1.00	1.00	0.48	0.16

Notes to the Tables.

Note to Table 1. The table reports the name of the model "Model", the fundamental predictors used in the model ("f_t") and the mnemonics used to refer to these fundamentals in Table 2.

Note to Table 2, Panel A. "L" denotes linear model and "NL" denotes non-linear model. Mnemonics for the predictors are as in Table 1. If a paper considers several models at the same time, their predictors are separated by a semicolon; for example, a paper that considers both UIRP and the monetary model with interest rates will be characterized by: $i; (m, \pi, y, i)$. In addition, "SP" denotes stock prices; "q" denotes the real exchange rate; "CP" denotes commodity prices (either oil price or the commodity price); "D" denotes the dividend yield; "CA" denotes the current account; "TB" denotes the trade balance; "s" is the lagged exchange rate; "FCI" denotes Financial Condition Indices; "r" denotes the real interest rate; "u" denotes unemployment; a "Δ" before a variable denotes the first difference of that variable. Note that Cheung et al. (2005) also consider the model with productivity (m, y, i, z) as well as a model with net foreign assets $(p, r, b$, terms of trade, net foreign assets and the relative price of non-tradeables). "Calibr. coeff." means calibrated coefficients; "estim. coeff." denotes estimated coefficients; "calibr. coint." means calibrated values for the cointegration vector; "estim. coint." denotes estimated cointegration vector; "diff." means that the model is estimated with variables in differences; "semi-param." means that the model is estimated with semi-parametric methods; "contemp." denotes contemporaneous predictors; "lag fund." denotes lagged fundamentals; "real. fund." means that the model uses ex-post, realized values of the contemporaneous fundamentals for prediction; "univ." denotes a univariate model (with possible lags of exchange rates and no fundamentals); "ECM" denotes Error Correction Model; "NP" denotes non-parametric; "LWR" denotes locally weighted regression; "MS" denotes Markov Switching model; "neural net." denotes neural network model. Canova's (1993) model includes stochastic volatility; in Cheung, Chinn and Pascual (2005), the contemporaneous fundamental model is estimated with OLS whereas the cointegrating vector in the ECM model is estimated via the Johansen procedure; Wright (2008) also includes the ratio of the current account to output among the regressors; in Adrian et al. (2011), the repo- and commercial paper variables (here labeled "repo;comm.paper") are detrended over the full sample.

Note to Table 2, Panel B. "Freq." denotes the frequency of the data: "Y" for yearly,

"Q" for quarterly, "M" for monthly, "D" for daily, "W" for weekly, and "30 min." for 30-minutes data. "real t" indicates that the study uses real-time vintages of data. "varies by country" means that typically the sample is 1973:3-2006:6, but for EU countries it may stop in 1998; "trade-w." denotes the trade-weighted exchange rate; "US effective exch." denotes the US effective exchange rate. Country mnemonics are as follows: "JP" is Japan, "GER" is Germany, "CA" is Canada, "UK" is the United Kingdom, "EU" is the Euro area, "SWI" is Switzerland, "AU" is Australia, "NZ" is New Zealand, "CHI" is Chile, "SA" is South Africa, "FR" is France, "NET" is the Netherlands, "IT" is Italy; "NOR" is Norway. Molodtsova et al. (2010) use quarterly data interpolated from annual data; Carriero et al. (2009) use average exchange rates rather than end-of-period data.

Note to Table 2, Panel C. The benchmark models include "Taylor" (the Taylor-rule model), the random walk ("RW"), the random walk with drift ("RWWD"), the forward discount ("F"), the autoregressive ("AR"), the vector autoregressive ("VAR") and the heterogeneous coefficient panel model ("heter."). "Absolute" denotes cases in which there is no benchmark model and the forecast evaluation is based on absolute (rather than relative) terms. "Estim" denotes the parameter estimation method, and includes rolling ("roll") and recursive ("rec") estimation; when available, the size of the rolling estimation window is included in parentheses; "T" denotes the total sample size. "h" is the forecast horizon; "1-16 qrs." typically denotes 1,4,8,12,16 quarters horizons; "(iter.)" denotes that forecasts at horizons bigger than one are made with an iterated method; otherwise it is assumed that the forecast method is direct. "Forec." denotes the starting date of the sample used for out-of-sample forecast evaluation (the end of the forecast sample is typically the end of the sample from Table 3); "varies" means that various window sizes have been used so several out-of-sample periods have been investigated; Alquist and Chinn (2008) also consider 1999Q1-20005:Q4; Chinn and Meese (1995) also consider 1987-1990:11; Groen (1999) also considers 1981:10-1991:12; Meese and Rogoff (1983a) also consider 1978:11-1981:6; Mizrach (1992) also considers 1974:1-1979:3; Cheung, Chinn and Pascual (2005) also consider post 1982; sample information for daily data is only approximate. "Eval. Method" denotes the method used for the forecast evaluation; "DMW" denotes the Diebold and Mariano (1995) and West (1996) test; "CW" denotes the Clark and West (2006) test; "ENCNEW" denotes the Clark and McCracken (2001) test; "GR" denotes the Giacomini and Rossi (2010) test; "Theil" denotes Theil's (1966) U statistic"; "GW" denotes the Giacomini and White (2006) test; "sign" denotes the direction of change statistics; "utility-based" denotes utility-based statistics; "MSFE" denotes the Mean Square Forecast Error; "MAE" denotes the Mean

66

Absolute Error; "R2" denotes the R-square statistic; "interval" denotes methods based on interval evaluation; "PIT" denotes the Probability Integral Transform; "various" indicates that several forecast evaluation methods have been used. Note that Rogoff and Stavrakeva (2008) consider all of the following statistics: CW, Theil, DMW, ENCNEW; Cerra and Saxena (2010) consider Theil, sign, DMW, CW.

Note to Table 2, Panel D. The information in the panel succintly summarizes whether the author(s)' preferred specification beats the benchmark and provides evidence in favor of successful predictive ability.

Note to Table 3. The tables report the sample sizes available for each predictor and each country. "Start" denotes the start of the sample; "End" denotes the end of the sample; "N. Obs." denotes the total number of observations.

Note to Tables 4 and 5. The tables report p-values of the following tests: Rossi's (2005) Granger-causality robust ("GC"), Diebold and Mariano (1995) and West (1996) ("DMW") and Clark and West (2006) ("CW"); the benchmark model in the latter two tests is the random walk without drift. "h" denotes the forecast horizon. "RMSFE$_{RW}$" denotes the root mean squared error of the random walk model; "RMSFER" denotes the ratio of the root mean squared forecast error of the model relative to that of the random walk without drift: values smaller than unity denote that the model forecasts better than the random walk benchmark.

Notes to the Figures.

Notes to Figures 1, 2, 5 and 6. The figures plot Giacomini and Rossi's (2010) Fluctuation test statistic (solid line) along with its critical value (dotted line) for selected countries, models and forecast horizons. The x-axis denotes time. When indicated, R denotes the window size used for parameter estimation; if not indicated, R equals half of the total sample size. The window size used to smooth out-of-sample predictive ability in the Fluctuation test is a third of the total number of forecasts.

Notes to Figure 3 and 4. The figures plot Inoue and Rossi's (2012) sup-test statistic (solid line) along with its critical value (dotted line) for selected countries, models and forecast horizons. The x-axis denotes the window size, R.

Notes to Figure 7. The figure is a scatterplot of the p-values of Clark and West's (2006) test at short-horizons (on the x-axis) and at long-horizons (on the y-axis) for several predictors, denoted by different markers, across multiple countries (each point denotes a country). P-values smaller than 0.05 denote statistical significance at the 5% level.

Figure 1. Fluctuation Test   (Monetary-ECM Model, Short-Horizon)

68
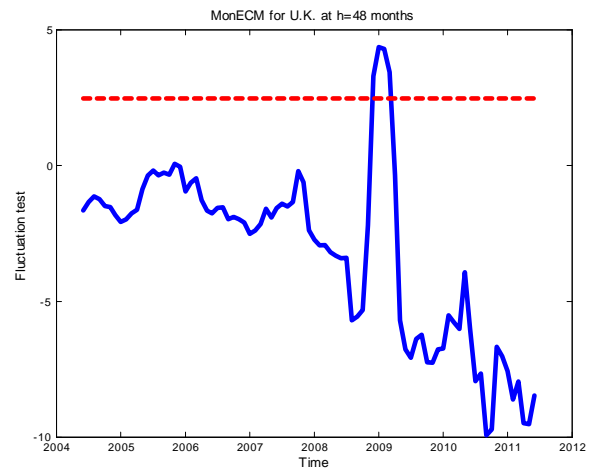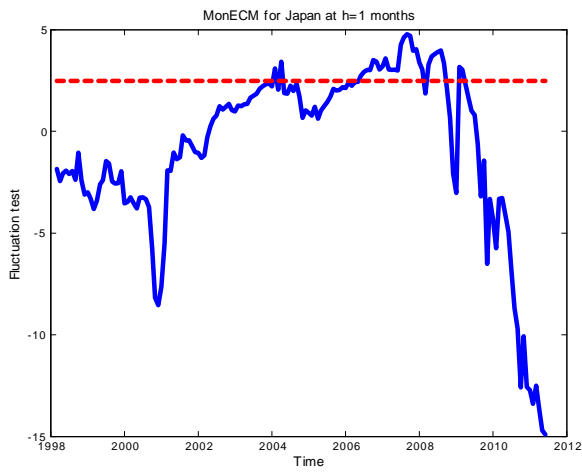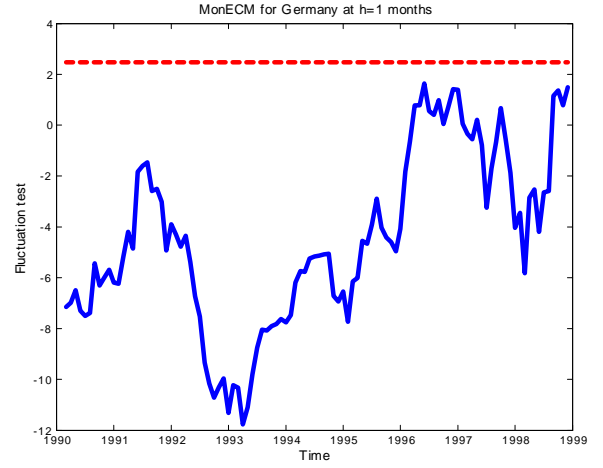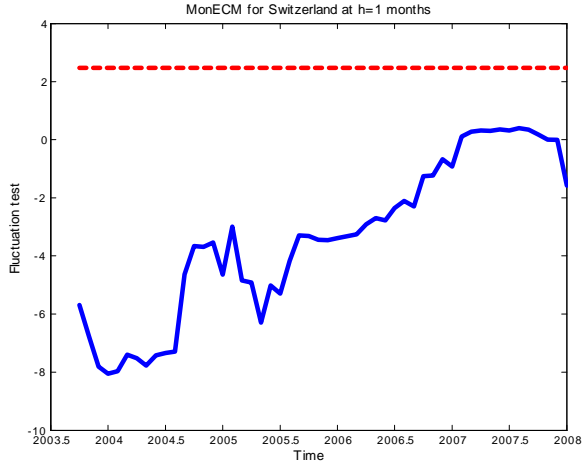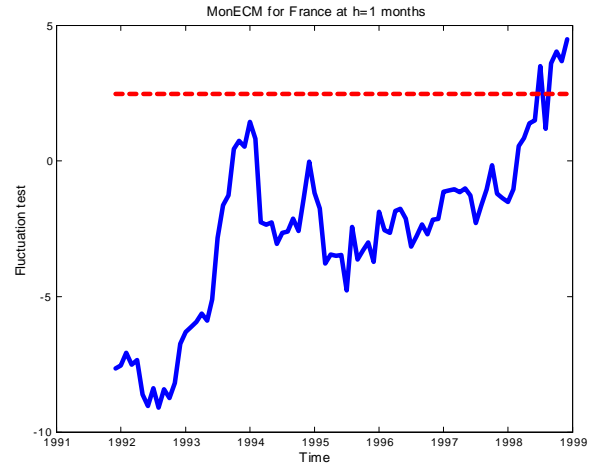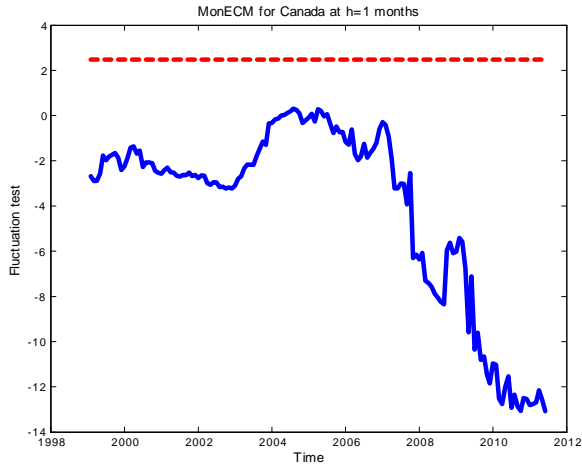
Figure 2. Fluctuation Test   (Monetary-ECM Model, Long-Horizon)
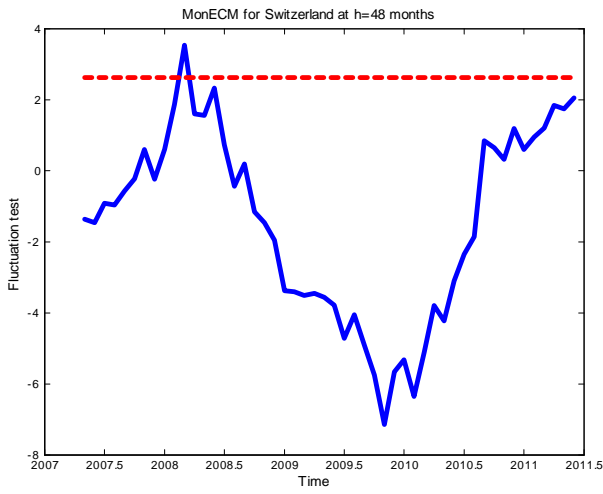

MonECM for Canada at h=48 months

MonECM for France at h=48 months

MonECM for Switzerland at h=48 months

MonECM for Germany at h=48 months

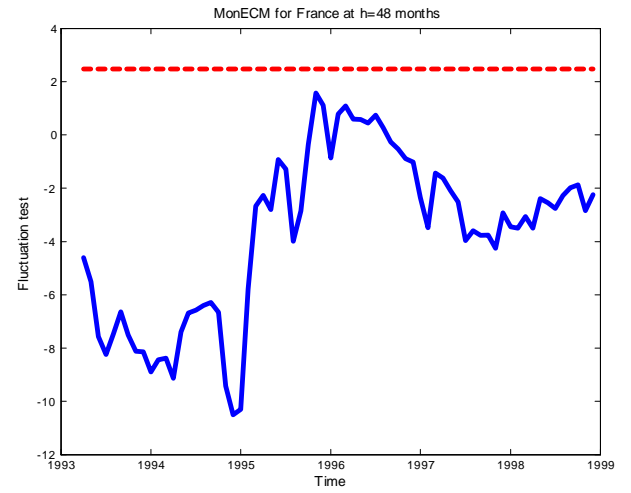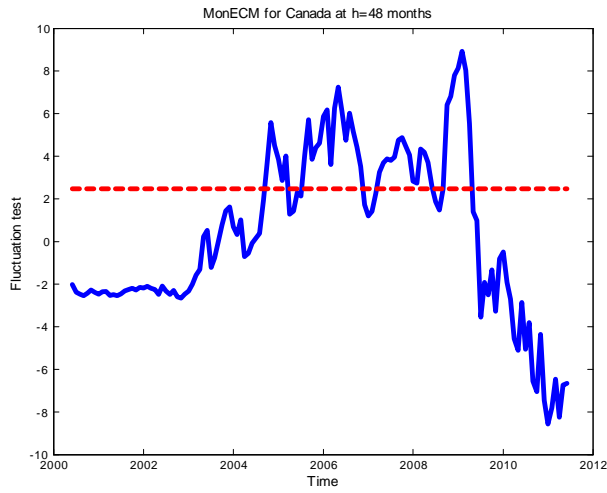MonECM for Japan at h=48 months

MonECM for U.K. at h=48 months
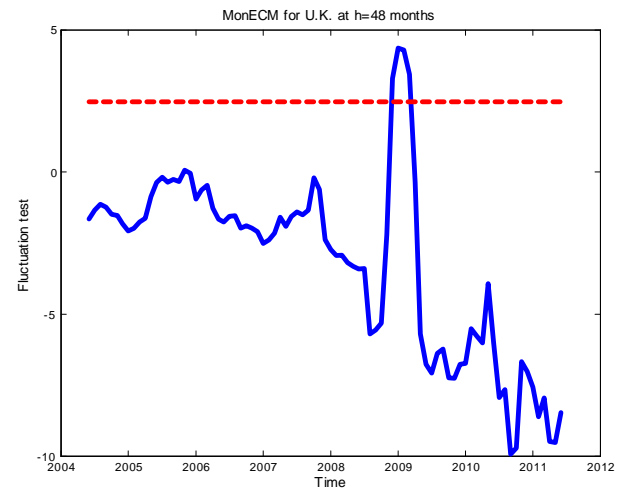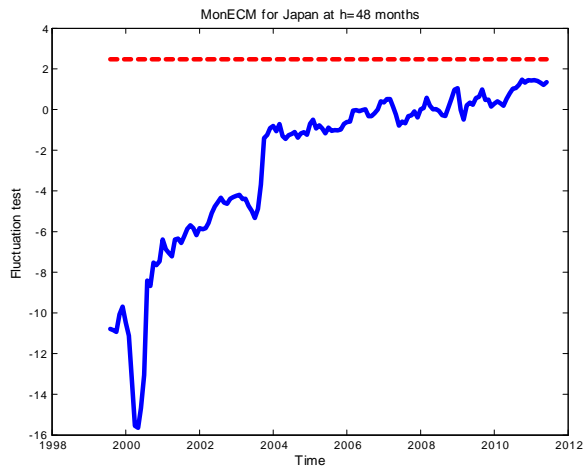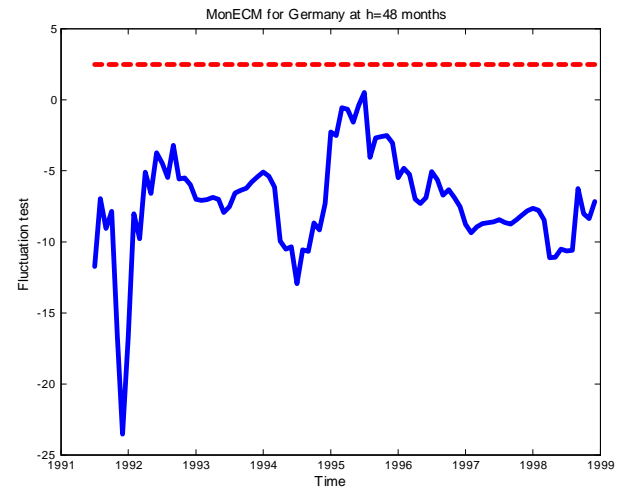
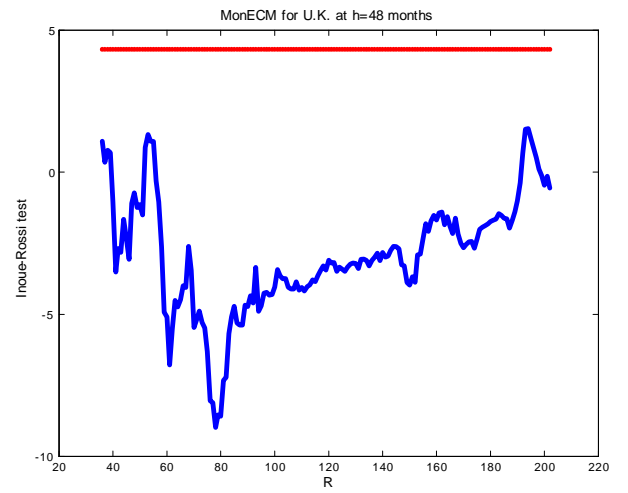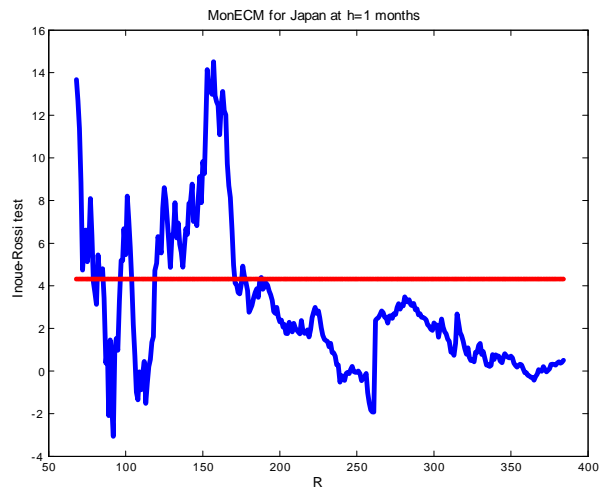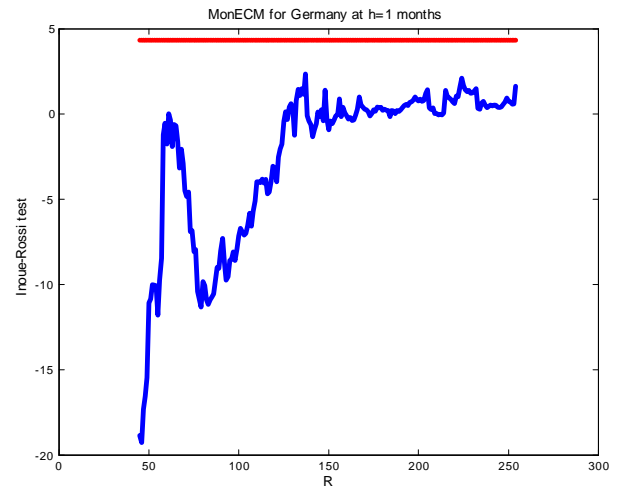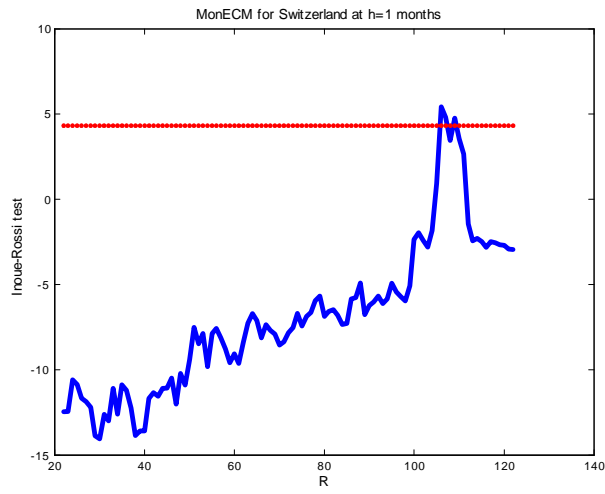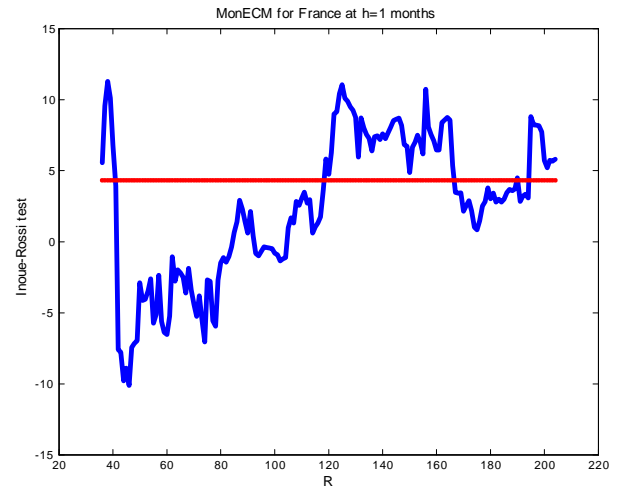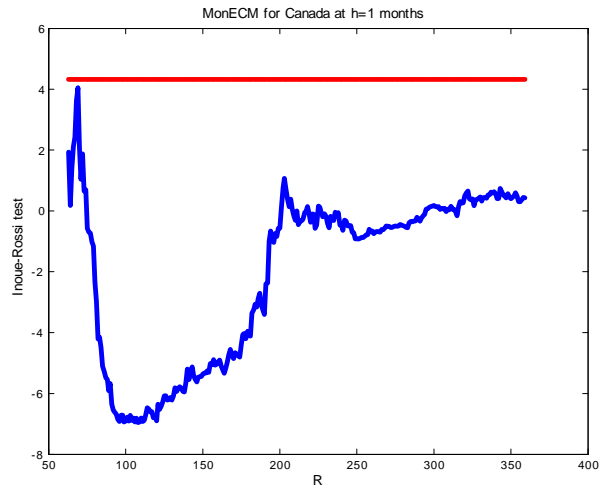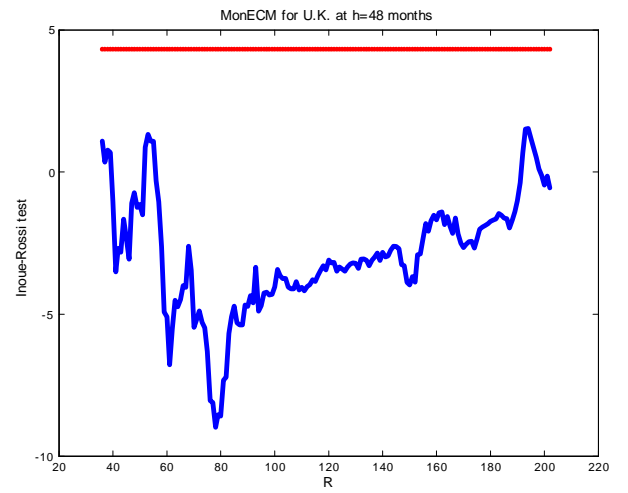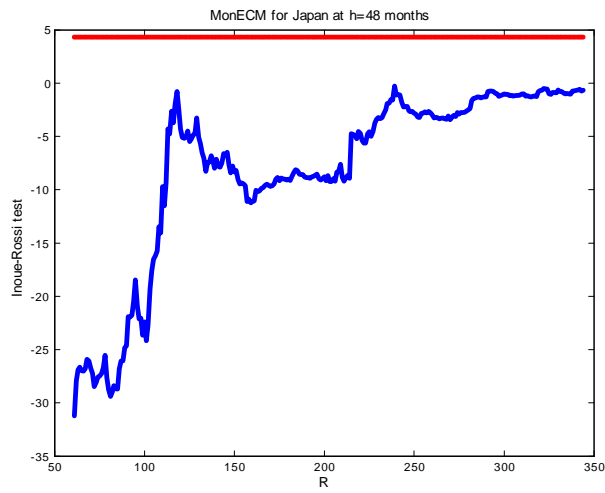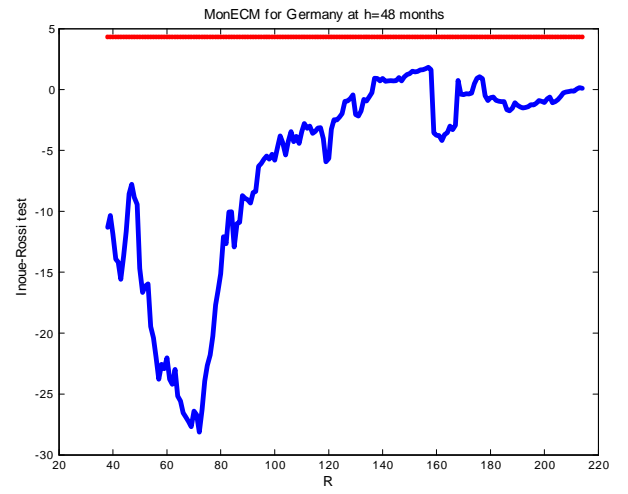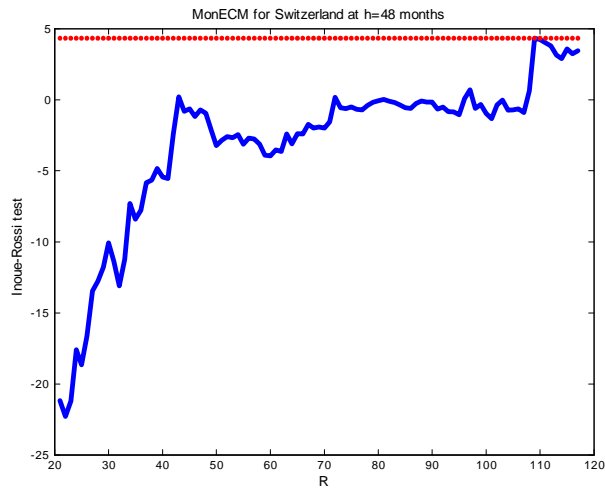# Figure 3. Robustness to Window Size (Monetary-ECM Model, Short-Horizon)

MonECM for Canada at h=1 months

MonECM for France at h=1 months

MonECM for Switzerland at h=1 months

MonECM for Germany at h=1 months

MonECM for Japan at h=1 months

MonECM for U.K. at h=48 months

Figure 4. Robustness to Window Size   (Monetary-ECM Model, Long-Horizon)
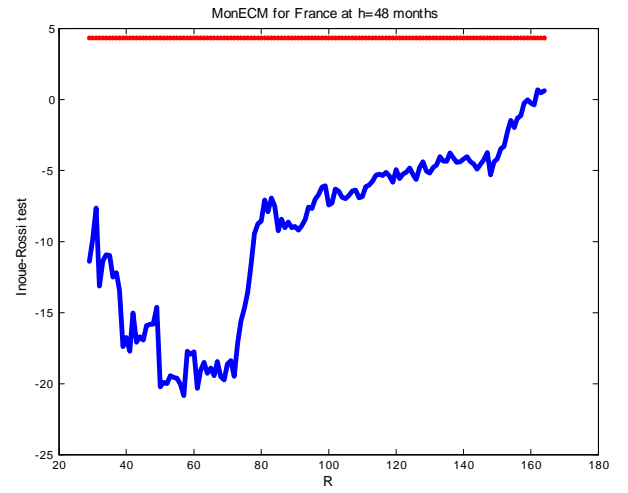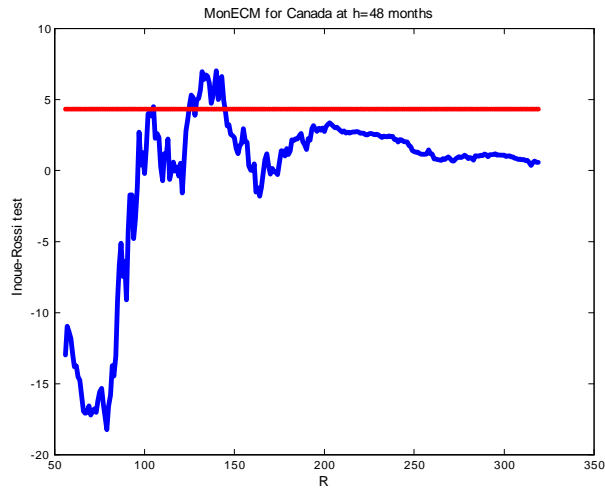
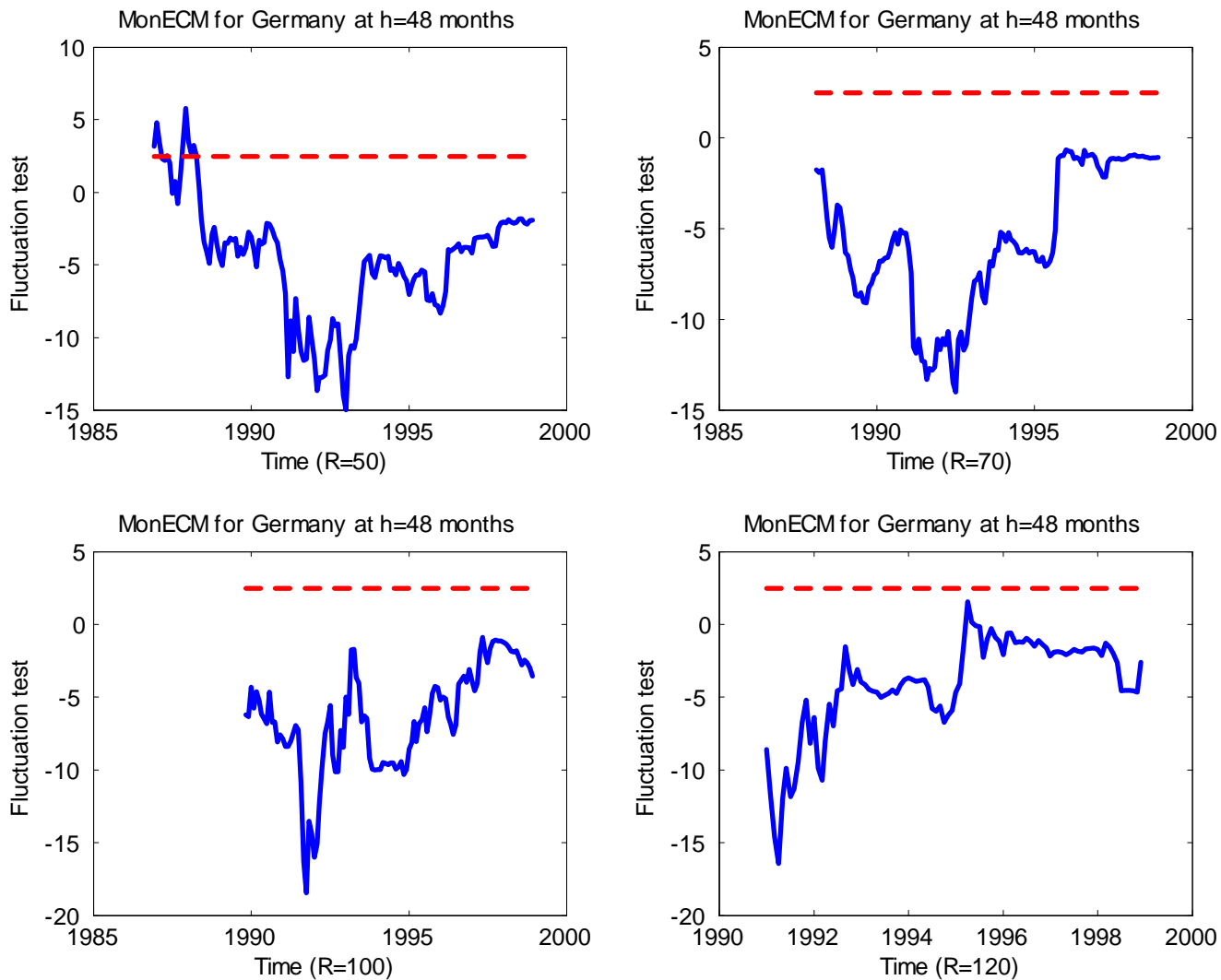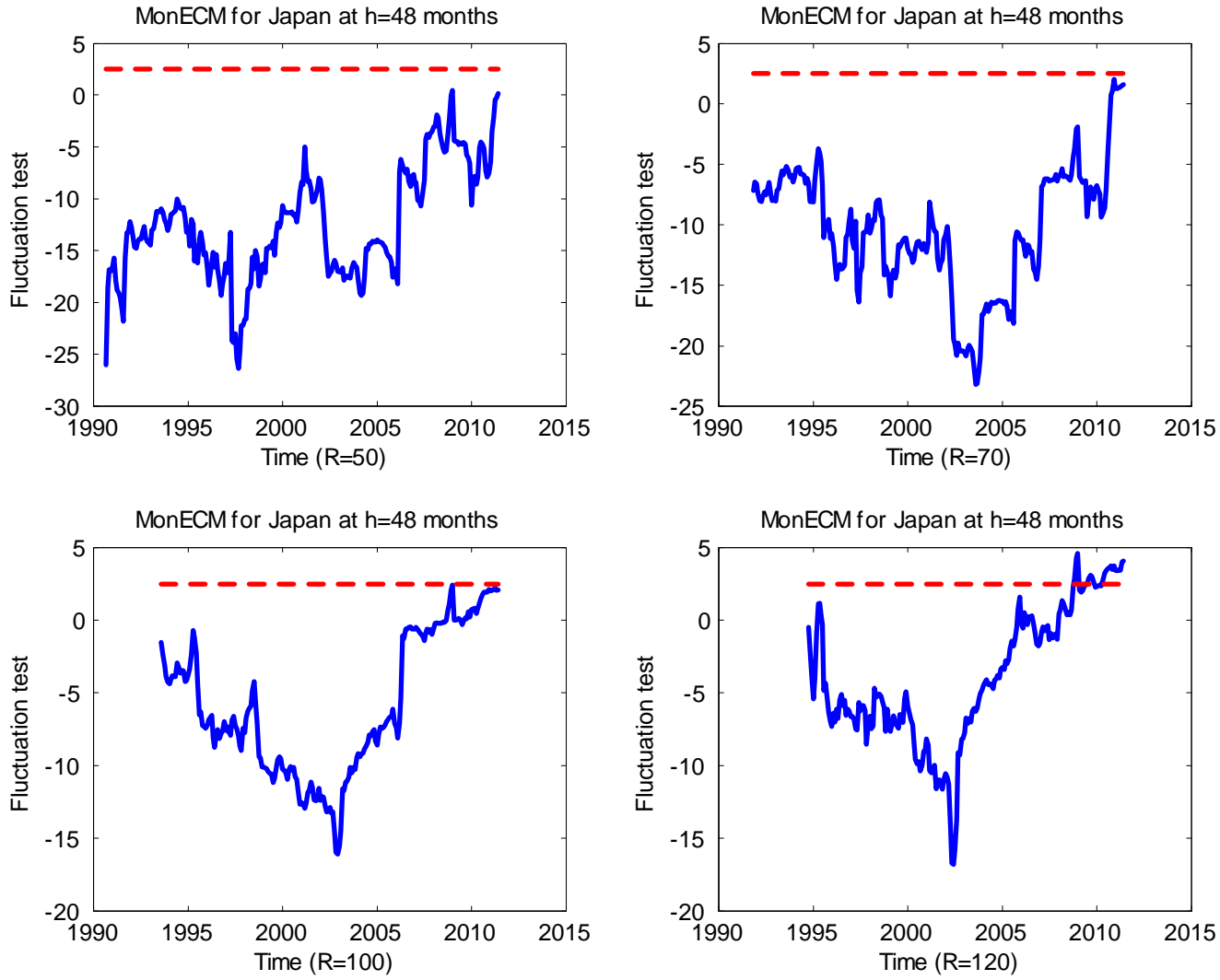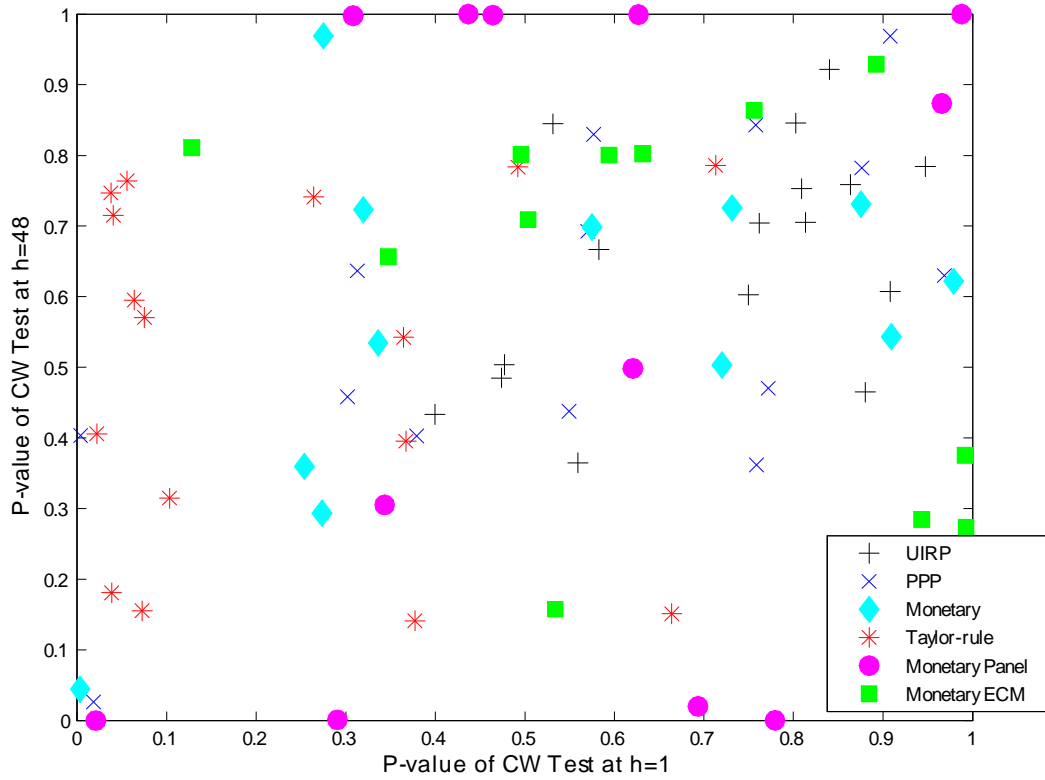**Figure 5. Interaction between Window Size and Out-of-Sample Period – Germany**

Figure 6. Interaction between Window Size and Out-of-Sample Period – Japan

Figure 7. Out-of-sample Predictive Ability of Economic Models

Figure 8. In-sample Predictive Ability of Economic Models

74