DISCUSSION PAPER SERIES

No. 6638

**EXCHANGE RATES AND
FUNDAMENTALS: FOOTLOOSE
OR EVOLVING RELATIONSHIP?**

Lucio Sarno and Giorgio Valente

INTERNATIONAL MACROECONOMICS

Centre for Economic Policy Research

www.cepr.org

EXCHANGE RATES AND FUNDAMENTALS: FOOTLOOSE OR EVOLVING RELATIONSHIP?

Lucio Sarno, University of Warwick, AXA Investment Managers and CEPR
Giorgio Valente, University of Leicester

This Discussion Paper is issued under the auspices of the Centre's research programme in **INTERNATIONAL MACROECONOMICS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as a private educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. Institutional (core) finance for the Centre has been provided through major grants from the Economic and Social Research Council, under which an ESRC Resource Centre operates within CEPR; the Esmée Fairbairn Charitable Trust; and the Bank of England. These organizations do not give prior review to the Centre's publications, nor do they necessarily endorse the views expressed therein.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

ABSTRACT

Exchange Rates and Fundamentals: Footloose or Evolving Relationship?*

Using novel real-time data on a broad set of economic fundamentals for five major US dollar exchange rates over the recent float, we employ a predictive procedure that allows the relationship between exchange rates and fundamentals to evolve over time in a very general fashion. Our key findings are that: (i) the well-documented weak out-of-sample predictive ability of exchange rate models may be caused by poor performance of model-selection criteria, rather than lack of information content in the fundamentals; (ii) the difficulty of selecting the best predictive model is largely due to frequent shifts in the set of fundamentals driving exchange rates, which can be interpreted as reflecting swings in market expectations over time. However, the strength of the link between exchange rates and fundamentals is different across currencies.

Lucio Sarno
Finance Group
Warwick Business School
University of Warwick
Coventry
CV4 7AL
Email: lucio.sarno@wbs.ac.uk

Giorgio Valente
Department of Economics
University of Leicester
University Road
Leicester
LE1 7RH
Email: giorgio.valente@le.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=138519

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=157421

1   Introduction

This paper employs an empirical framework for modelling and forecasting exchange rates that explicitly takes into account the survey and empirical evidence that there are changing parameters in economic fundamentals and that no fundamentals model appears to perform well for long periods of time. This framework allows for a menu of fundamentals that comprises not only the standard monetary fundamentals most commonly used in the literature, but also other variables suggested by exchange rate determination theory, including net foreign assets, interest rates, the trade balance, and lagged values of exchange rate changes. We focus on two key questions. First, as a preliminary to the forecasting exercise, we perform an 'empirical search' for the best performing exchange rate model given a broad set of fundamentals available in real time over the recent float for dollar exchange rates. In other words, we ask whether, allowing for shifts in the weights attached to the fundamentals over time, the best model that optimizes such information is indeed capable of predicting the exchange rate with a reasonable degree of accuracy. In this setting, the best model is chosen using *ex post* information on the realized exchange rate to select the optimal combination of fundamentals within a set of misspecified fundamentals models. Second, when the previous exercise suggests that there is a (time-variant) fundamentals model capable of explaining and predicting exchange rate movements, we ask whether it is possible to recover the best model obtained earlier when the out-of-sample forecasting exercise is conducted using conventional model selection criteria.

It is a stylized fact that out-of-sample forecasts of exchange rates produced by structural models based on economic fundamentals are no better than those obtained using a naive random walk or no-change model of the nominal exchange rate. These results, first noted in the seminal work of Meese and Rogoff (1983), have been proved to be very robust. Some recent research suggests that models based on canonical fundamentals can explain a small amount of the variation in exchange rates (e.g., Mark, 1995; Mark and Sul, 2001; Groen, 2005). Sweeney (2006) also provides evidence that challenges the conventional wisdom that industrial-country floating exchange rates contain unit roots and that, in out-of-sample forecasts, mean-reversion models beat random walks on average, in some forecast periods significantly. However, others remain skeptical (e.g., Kilian, 1999; Berkowitz and Giorgianni, 2001; Faust, Rogers and Wright, 2003; Engel and West, 2005), so that evidence that exchange rate forecasts obtained using fundamentals models are better than forecasts from a random walk remains elusive (e.g., Cheung, Chinn and Garcia Pascual, 2003; Sarno, 2005).

Prior research has also documented that economic fundamentals do not contain information useful for forecasting exchange rates partly because the behavior of the 'empirical fundamentals', that is the fundamentals that would be capable of explaining exchange rate movements, is radically different from the behavior of the fundamentals suggested by exchange rate determination theory. Notably, observing the marked increase in volatility of exchange rates which occurs when moving from a fixed or pegged exchange rate regime to a floating regime, Baxter and Stockman (1989) and Flood and Rose (1995) argue that any tentatively adequate exchange rate model should have fundamentals which are also much more volatile during floating rate regimes. This evidence may be seen as suggesting that there are speculative forces at work in the foreign exchange market which are not reflected in the usual menu of economic fundamentals. These constitute key elements of the 'exchange rate disconnect puzzle' (Obstfeld and Rogoff, 2000).[1]

---

[1] Quoting Obstfeld and Rogoff (2000, p. 373): '[...] the *exchange-rate disconnect puzzle* [...] alludes broadly to the exceedingly

Notwithstanding this evidence, it seems hard to believe that broad knowledge of the state of the economy at a point in time is useless information to forecast exchange rate movements. It may be that exchange rate models perform poorly not (only) because the information in the fundamentals is deficient, but because volatile expectations and departures from rationality are likely to account for the failure of exchange rate models. For example, Frankel (1996) argues that exchange rates are detached from fundamentals because of 'swings' in expectations about future values of the exchange rate, listing several pieces of evidence suggesting that expectations are to blame for such behavior. In this line of reasoning, Bacchetta and van Wincoop (2004) provide a theory of exchange rate determination which incorporates the fact that practitioners in the foreign exchange market regularly change the weight they attach to different economic variables—as evidenced in a variety of survey studies (e.g. Cheung and Chinn, 2001)—in the context of a stylized rational-expectations model of exchange rate determination. In this model, as the market rationally searches for an explanation of the observed exchange rate change, it may attribute it to some macroeconomic indicator, which in turn becomes the 'scapegoat' and influences trading strategies. This implies that agents attach an excessive weight to the 'scapegoat' variable to explain exchange rate movements during some periods, so that the exchange rate is unrelated to other observed economic fundamentals. The model is also capable of rationalizing parameter instability in empirical exchange rate models of the kind often documented in the relevant literature (e.g. Rossi, 2005, 2006). Over time different observed variables may be taken as the scapegoat, so that the weights attributed to economic variables change.

Surprisingly little attention has been directed towards assessing the potential of these considerations for establishing an economically meaningful relationship between exchange rates and fundamentals. The present paper fills this gap. Specifically, we employ a recursive procedure originally developed by Pesaran and Timmermann (1995) which allows us to select, quarter by quarter, the best model on the basis of a variety of statistical criteria within all possible combinations of fundamentals, allowing for the explanatory variables and the parameters to change over the sample examined. The out-of-sample forecasting exercise is carried out using a novel real time data set for the fundamentals for five major US dollar exchange rates during the recent float. The data incorporate historically available information for a US investor who wishes to select the best exchange rate model over time. The results of this out-of-sample exercise are compared with the results that would have been obtained if one had knowledge of the best-performing fundamentals model. This comparison sheds light on the usefulness of the information embedded in the fundamentals for forecasting the exchange rate as well as on the ability of the model selection procedure to use such information optimally.

The results, considering short-term forecasts of one-quarter ahead, are as follows. First, the information embedded in the economic fundamentals can explain future exchange rate movements with a remarkable degree of accuracy and allows us to outperform a random walk model for three out of five exchange rates. This finding is obtained using methods that are shown to account reliably for potential forecasting biases. However, this

---

weak relationship (except, perhaps, in the longer run) between the exchange rate and virtually *any* macroeconomic aggregates. It manifests itself in a variety of ways. For example, Meese and Rogoff (1983) showed that standard macroeconomic exchange-rate models, even with the aid of ex-post data on the fundamentals, forecast exchange rates at short to medium horizons no better than a naive random walk.' This is the aspect of the puzzle on which we focus in this paper. However, it is worth noting that the puzzle also refers to 'the remarkably weak short-term feedback links between the exchange rate and the rest of the economy' (Obstfeld and Rogoff, 2000, p. 380), that is the fact that fluctuations in the exchange rate seem to have little effect on aggregate activity. We do not focus on this aspect of the puzzle in this paper.

requires the ability to select the best model from the various set of models that can be used on the basis of information available, i.e. this requires that the investor must have available a reliable model selection criterion to discriminate among different (misspecified) models. Second, if conventional model selection criteria are used to choose *ex ante* the best model from a large pool of models available, the same set of economic fundamentals is not useful in forecasting exchange rates out of sample. This finding appears to be due to the inability of existing model selection criteria to identify the predictive variables to be used *ex ante* and to detect the frequent shifts in the best model capturing the evolving dynamic relationship between exchange rates and fundamentals.

We find that models which optimally use the information in the fundamentals change often and this implies frequent shifts in the parameters. Our interpretation of these results is that standard model selection criteria appear to be unable to capture such shifts, yielding empirical exchange rate models that cannot forecast the exchange rate better than a random walk model. Therefore, while the stylized fact that fundamentals models cannot beat a no-change model is–yet again–confirmed in this paper, the reason for this result is, on the basis of the evidence presented here, different from many other studies. Our evidence suggests that the exchange rate disconnect phenomenon is unlikely to be caused by lack of information in the fundamentals, and more likely to be the result of poor model selection criteria in this context. Consequently, we conclude that theories of exchange rate determination linking the exchange rate to fundamentals are not necessarily invalidated by the lack of strong forecasting performance in empirical exchange rate models that is routinely recorded in the literature, and that progress in this research agenda is likely to be made by investing research effort and resources on improving the current state of knowledge of model selection procedures.

The remainder of the paper is organized as follows. In Section 2 we briefly review the body of research to which this paper is related and further motivate our approach. Section 3 discusses the framework used to analyze exchange rate predictability allowing for shifts in fundamentals and parameter instability. Section 4 describes the data set and a preliminary empirical analysis of the time series of interest. Section 5 presents empirical results relating to the empirical search for the best exchange rate model. In Section 6 we report the results from the out-of-sample forecasting exercise based on the same set of fundamentals information and employing conventional model selection criteria to choose the best fundamentals model period by period. Section 7 provides a discussion of possible extensions of this research and reports some robustness checks. Section 8 concludes. Details of the construction of the real time data set and of the simulation procedure employed in the empirical analysis are provided in the Appendix.

2 Exchange Rates and Fundamentals: Overview and Motivation

2.1 The Exchange Rate Disconnect Puzzle

A large literature has investigated the relationship between exchange rates and fundamentals by focusing on the deviation, say u, of the nominal exchange rate from its fundamental value:

$$u_t = s_t - f_t, \tag{1}$$

where s denotes the log-level of the nominal bilateral exchange rate (the domestic price of the foreign currency); f is the long-run equilibrium of the nominal exchange rate determined by economic fundamentals; and t is a time subscript. The fundamentals term f is, most commonly, given by a parsimonious set of 'monetary fundamentals',

comprising the differential in money supply and the differential in output (e.g., Mark, 1995). However, it is clear that a broader set of fundamentals could be employed on the basis of international macroeconomics theory. Additional fundamentals include, for example, net foreign assets, as suggested by overlapping-generations general equilibrium models (e.g. Cavallo and Ghironi, 2002; Gourinchas and Rey, 2007), and the trade balance, as suggested by portfolio balance models (e.g. Branson, 1984; Kumhof and Van Nieuwerburgh, 2005) and elasticity models of the balance of trade (e.g. Krueger, 1983; Rose and Yellen, 1989; Obstfeld and Rogoff, 2004).[2]

While it has been difficult to establish the empirical significance of the link between monetary fundamentals and the exchange rate due to a number of cumbersome econometric problems (e.g. Kilian, 1999; Berkowitz and Giorgianni, 2001), some recent research suggests that macroeconomic fundamentals move together with the nominal exchange rate over long periods of time (Groen, 2000, 2005; Mark and Sul, 2001; Rapach and Wohar, 2002; Abhyankar, Sarno and Valente, 2005). The analysis of exchange rate predictability generally relies on long-horizon regressions of the following form:

$$\Delta_q s_{t+q} = c + \psi_q u_t + \varepsilon_{t+q},\tag{2}$$

where Δ_q denotes the q-difference operator. However, the literature often ignores the fact that data for economic fundamentals may not be available at the time forecasts are made or may suffer from measurement errors. Employing data in real time–that is data that would have been available to researchers at the time forecasts would have been produced–further complicates analyses of the predictive power of fundamentals on the exchange rate (Faust, Rogers and Wright, 2003).

Flood and Rose (1995) shed light on the relation (or lack of it) between exchange rates and fundamentals in what they term 'a virtual quest for fundamentals.' The idea is to compare the volatility of the traditional set of monetary fundamentals typically employed in the literature to the volatility of the fundamentals that would be capable of explaining the volatility of foreign exchange returns. Observing the marked increase in volatility of exchange rates which occurs when moving from a fixed exchange rate regime to a floating regime, Flood and Rose argue that any tentatively adequate exchange rate model should have fundamentals which are also much more volatile during floating rate regimes. In their empirical work, they find little change in the volatility of economic fundamentals suggested by monetary models across different nominal exchange rate regimes for a number of major exchange rates.[3]

2.2   Model Instability and Swings in Expectations

A number of authors have argued that the poor forecasting performance recorded in this literature may be due to the fact that the parameters in the estimated equations are unstable. This instability may be rationalized on a number of grounds, including policy regime changes, instabilities in the money demand or purchasing-power-

---

[2]See also the broad set of macroeconomic news suggested by Andersen, Bollerslev, Diebold and Vega (2003).

[3]Taking an alternative view, Engel and West (2005) demonstrate that in a rational-expectations present value model, under the assumptions that fundamentals are nonstationary and the discounting factor is near unity, the exchange rate will behave as a 'near' random walk process. This implies that the difficulty to predict exchange rates using fundamentals may well be consistent with conventional exchange rate determination models. Put differently, this theory offers reasons why forecasting with fundamentals can be very hard and why lack of forecasting power does not *per se* imply a rejection of conventional exchange rate determination theories. Yet, the arguments and evidence in this paper, while confirming the Engel-West view that the lack of forecasting power in fundamentals does not necessarily imply rejection of exchange rate theory, suggest that this result may be due to reasons different from and potentially complementary to the ones indicated by Engel and West.

parity equations, or also agents' heterogeneity leading to different responses to macroeconomic developments over time (e.g. see Schinasi and Swamy, 1989; Rossi, 2005, 2006).

Other researchers have claimed that volatile expectations or departures from rationality are likely to account for the failure of exchange rate models. For example, Frankel (1996) argues that exchange rates are detached from fundamentals by swings in expectations about future values of the exchange rate. Several pieces of evidence suggest that expectations are to blame for such behavior: 1) Survey measures of exchange rate expectations are very poor forecasters and the expectations, themselves, are frequently not internally consistent (Frankel and Froot, 1987; Sarno, 2005). 2) Failure of rational expectations is often blamed for the failure of uncovered interest parity (Engel, 1996). 3) Trend following trading rules appear to make positive risk-adjusted excess returns, in apparent violation of the efficient markets hypothesis (e.g. Levich and Thomas, 1993; Neely, Weller, and Dittmar, 1997). 4) Switching from a fixed exchange rate to a floating rate—which changes the way expectations are formed—changes the behavior of exchange rates and the ability of interest rate parity to explain exchange rate changes (Baxter and Stockman, 1989; Flood and Rose, 1995, 1999).

In this line of reasoning, Bacchetta and van Wincoop (2004, 2006) provide a theory of exchange rate determination which incorporates the fact that practitioners in the foreign exchange market regularly change the weight they attach to different economic variables—as evidenced in a variety of survey studies (e.g. Cheung and Chinn, 2001)—in the context of a stylized rational-expectations model. This model is capable of explaining parameter instability in empirical exchange rate models in terms of a 'scapegoat' story, where some variable is given excessive weight during some period, implying movements in the exchange rate that are unrelated with observed economic fundamentals, for example due to unobserved liquidity trades. As the market rationally searches for an explanation for the observed exchange rate change, it may attribute it to some macroeconomic indicator, which in turn becomes the scapegoat and influences trading strategies. Over time different observed variables may be taken as the scapegoat, so that the weights attributed to macroeconomic variables change.

2.3   Introducing Our Approach

In this paper, we build an empirical framework that explicitly takes into account the survey evidence that there are changing weights in fundamentals and that no model appears to perform well for long periods of time. Our framework allows for a menu of fundamentals that comprises not only the standard monetary fundamentals most commonly used in the literature since Mark (1995), but also other variables suggested by exchange rate determination theory, including net foreign assets, interest rates, the trade balance, and lagged values of exchange rate changes. We employ a recursive model selection procedure where we select, quarter by quarter, the best model on the basis of a variety of criteria across all possible combinations of fundamentals, allowing for the model specification and the parameters to change.

We perform an out-of-sample forecasting exercise using real time data for the fundamentals, for each criterion used in model selection and for five major dollar exchange rates, and then compare the results of this purely out-of-sample exercise with the results that would have been obtained if one knew the best performing fundamentals model within a set of misspecified fundamentals models. This comparison will shed light on the usefulness of the information embedded in the fundamentals for forecasting the exchange rate as well as on the ability of the model selection procedure to use such information optimally. We confine the analysis to one-step-ahead

updating and prediction since our procedure is designed to allow the update of available information over time and forecast period by period, conditional on such information, in a realistic fashion similar to the procedure that an investor would be following in real time. The next section describes in detail this empirical framework.

3   Methodology

3.1   The Pool of Models

Let us consider a US investor who wishes to forecast in real time a bilateral dollar exchange rate and believes that the exchange rate can be predicted by a menu of economic variables that are suggested by exchange rate determination theory. Since the investor does not know the true data generating process (DGP) linking such fundamentals to the future exchange rate, a reasonable way to produce the exchange rate forecasts involves searching for an adequate model specification among all possible models believed to be capable of predicting the exchange rate. The search ought to be general enough to encompass all possible models of exchange rate determination considered plausible by the investor as well as to allow also for the possibility that none of the theoretical specifications is indeed useful for forecasting, so that the framework ought to allow for the possibility that the exchange rate follows a random walk (with or without a drift). The selection of the best exchange rate model is based on one or more criteria. As the investor repeats this exercise over time and new information (data) becomes available, the investor may change the specification. These changes in the model may be due to the learning process of the investor with the arrival of new information and/or to the changing nature of the DGP of the exchange rate.

In order to characterize this behavior econometrically, we follow the stock price predictive approach of Pesaran and Timmermann (1995, 2000) and Bossaerts and Hillion (1999), which we adapt to the context of exchange rate forecasting. At each point in time t, the investor estimates all possible combinations of $k-1$ fundamentals which may predict future movements in exchange rates, in addition to a drift term in exchange rate changes as a possible regressor. Thus, the full set of possible models comprises 2^k different models. The simplest model is a random walk without drift (no regressors) and the richest model is a fully unrestricted model where all k regressors, inclusive of the drift term, are considered. Denote a model M_i for $i = 1, \ldots, 2^k$, and let the set of regressors be $X_j = \mathbf{1}, x_1, \ldots, x_{k-1}$. Consider a $k \times 1$ selection vector, say v_i, composed of ones and zeros where a one in its j-th element indicates that the regressor in question is included in the model. Then, Model M_i may be represented by the k-digit string of zeros and ones corresponding to the binary code of its number. If k_i is the number of regressors included in model M_i, $k_i = e'v_i$, where e' is a $k \times 1$ vector of ones. Thus, M_i may be written as follows:

$$\Delta s_{t+1} = \beta_i' X_{t,i} + \varepsilon_{t+1,i} \tag{3}$$

where $\Delta s_{t+1} \equiv s_{t+1} - s_t$; β is a vector of parameters; $X_{t,i}$ is a $k_i \times 1$ vector of regressors (available to the investor at time t) under model M_i; and $t = 1, \ldots, \tau, \ldots, T$. Estimation of all candidate models may be carried out by ordinary least squares (OLS)–see Pesaran and Timmermann (1995, 2000). Applying a model selection procedure, the investor will conduct the first estimation of all 2^k models on the basis of the first τ observations and selects one model. Then, as new data become available, the investor repeats the procedure from time $\tau + 1$ up to time $T - 1$, yielding time-varying estimates of β_i throughout the sample for each candidate model.

In essence, starting from time $t = \tau$, the procedure delivers at each point in time estimation of 2^k models with corresponding 2^k one-period-ahead forecasts of the change in the exchange rate. Out of the 2^k models estimated, at each point in time $t = \tau, \tau + 1, \ldots, T - 1$, only one is selected as a result of a selection procedure based upon a set of conventional model selection criteria. The models selected yield forecasts of the exchange rate at time $\tau + 1, \tau + 2, \ldots, T$ on the basis of information sets defined at time $\tau, \tau + 1, \ldots, T - 1$ respectively.

3.2 Model Selection

We employ a number of criteria in order to select the forecasting model, within the set of possible models that can be obtained from all combinations of the fundamentals considered, on the basis of available information: the Mean Absolute Error (MAE); Root Mean Square Error (RMSE); the Last Absolute Error (LAE)[4]; the adjusted coefficient of determination ($\overline{R}^2$); the Akaike Information Criterion (AIC); the Schwarz Information Criterion (a Bayesian information criterion, say BIC); a sign criterion (SIGN) based on directional accuracy of the in-sample predictions of the exchange rate change; the Posterior Information Criterion (PIC), due to Phillips and Ploberger (1996); and the Fisher Information Criterion (FIC), due to Wei (1992).

The primary aim is to select a forecasting equation that could be viewed as a reasonable approximation to the DGP at the time when this procedure would have been used, making the exercise as realistic as possible. Because our first estimation is done with quarterly data up to 1987, when we produce the first one-quarter-ahead forecast, a US investor would have been able to employ the procedure described in this section as OLS was obviously known and easy to employ at that point in time. Also all of the criteria employed were known at that time, with the exception of the PIC and the FIC. In addition, as discussed in detail below, we base our analysis on a suitably constructed set of real time data which correspond exactly to the data available at the time forecasts were made for each of the economic indicators employed.

3.3 Forecast Evaluation

In terms of evaluation criteria of out-of-sample forecast accuracy, we rely both on statistical and economic criteria. The statistical criteria include the MAE and the RMSE, which are the most common forecast accuracy measures used in the literature on exchange rate forecasting since Meese and Rogoff (1983). We also utilize two economic criteria of forecast evaluation. First, we use the recursive end-of-period wealth criterion (EOPW), which maximizes cumulated wealth using forecasts from model M_i in a switching portfolio:

$$W_{t,i} = W_0 \prod_{t=\tau}^{T-1} \left(1 + R_{p,t+1}^i - TC_t \right), \tag{4}$$

where W_0 is initial wealth; the portfolio return $R_{p,t+1}^i = \left[\left(1 - \omega_t^i \right) r_t + \omega_t^i \left(r_t^* + \Delta s_{t+1} \right) \right]$, where ω_t^i denotes the weight associated with the investment in the foreign asset within a portfolio comprising two bonds which are identical in all respects but the currency of denomination; r_t and r_t^* are the one-period returns on domestic and foreign riskless bonds respectively. The weight ω_t^i is computed at any point in time on the basis of

---

[4]The LAE is simply the criterion which selects the model that delivers the smallest in sample prediction error in the last observation available in real time. This criterion is often used in the 'Bayesian learning' literature (e.g. Evans and Honkapohja, 2001).

one-step ahead forecasts $\Delta s_{t+1,i}$ generated by Model M_i, and TC_t denotes transactions costs.[5]   Hence, $W_{t,i}$ denotes wealth obtained under Model M_i at the end of the forecasting period calculated using the realizations of exchange rate changes ($\Delta s_{t+1,i}$) and the corresponding one-period interest rates at home (r_t) and abroad (r_t^*); τ is as defined in Section 3.1.

The second economic criterion we employ is the recursive excess return criterion (RR), calculated as the ratio of the mean portfolio excess returns to their standard deviations realized over the forecasting period:

$$RR_{t,i} = \frac{\frac{1}{T-\tau-1} \sum_{t=\tau}^{T-1} \left(R_{p,t+1}^i - r_t - TC_t \right)}{\sigma_{R_p^i}}, \tag{5}$$

where $\sigma_{R_p^i}$ is the standard deviation of the portfolio return under Model M_i calculated over the forecasting period (from time τ to $T-1$).   This criterion is similar to the Sharpe ratio criterion used by Pesaran and Timmermann (1995, 2000), adapted to the context of the foreign exchange market.

3.4   Data Snooping

In evaluating the forecasting models in our setup, it is important to be aware of data-snooping (or data-mining) biases.   Data snooping involves searching through the database for correlations and patterns that differ from results that would be anticipated to occur by chance or in random conditions.   When models are discovered that outperform a benchmark, there are a number of potential problems in making the leap from a back-tested model to successfully using the model out-of-sample in real world conditions.   The main problem is determining the probability that the relationships identified in the forecasting exercise have occurred at random or whether the predictability may be unique to the specific sample that was tested.

We employ a Reality Check procedure of the kind proposed by White (2000) and used, among others, by Sullivan, Timmermann and White (1999) for testing the null hypothesis that the best model selected in a specification search has no predictive superiority over a given benchmark model–in our case the random walk. This procedure is specifically designed to allow aggressive model searching to be undertaken with confidence that one will not mistake results that could have been generated by chance for genuinely good results, effectively correcting for data snooping biases in forecast evaluation.   The procedure implemented for testing the null hypothesis of equal forecast accuracy between the selected best model and the random walk (no predictability) benchmark is described in Sections 5.1-5.2 and Appendix B.   We also provide evidence on the adequateness of the Reality Check's performance in terms of empirical size and power in the present context in Section 5, where the procedure is employed under simulated data drawn from a pure random walk and several alternative specifications for the process of the exchange rate.   The results given below confirm that the Reality Check satisfactorily addresses the data snooping biases that arise in this context, by considering the distribution of the

---

[5]In our calculations we allow for short selling and treat the weight ω_t^i as a binary variable.   This means that the values that ω_t^i can take are 2 (short sell the domestic asset and invest the proceedings in the foreign asset when the forecast excess returns from currency investment are positive) and -1 (i.e. short sell the foreign asset and invest in the domestic asset when the forecast excess returns from currency investment are negative) respectively. Further, as showed in equation (4), we have explicitly taken into account transaction costs by deducting from the gross return $R_{p,t+1}^i$ a fixed amount $TC = 0.0004$ only when there is a shift in the portfolio allocation, yielding a binary time series TC_t which is equal to 0.0004 if $\Delta \omega_t^i \neq 0$ and equal to 0 if $\Delta \omega_t^i = 0$. We also carried out the calculations with short selling constraints–in which case the binary variable TC_t takes the values 1 and 0 respectively–and obtained qualitatively identical results.

test statistic of the null hypothesis of no predictability that accounts for the broad search across models carried out in the empirical work–see Section 5 and Appendix B.

4   Data and Preliminary Analysis

Our data set comprises quarterly observations on money supply, income (gross domestic product), trade balance, 3-month eurodeposit rates for the US, Japan, UK, Canada, Switzerland and Germany; and spot exchange rates for the US dollar *vis-à-vis* the Japanese yen, UK sterling, Canadian dollar, Swiss franc and German mark. We also have data on net foreign assets between the US and each of Japan, UK, Canada, Switzerland and Germany. The sample period covers most of the recent floating exchange rate regime, from 1977Q1 to 2003Q3 (except for Germany, which ends in 1998Q4); the start date of the sample was dictated by data availability (for the net foreign assets). The choice of countries reflects our intention to examine exchange rate data for a set of major industrialized economies that have been governed by a pure float over the sample.[6]

For the purposes of our analysis we assembled and compiled real time data of quarterly frequency. In particular, we constructed four vintages for each year considered with the first real time observation starting at 1987Q3. We performed the first estimation of the set of models examined using data from 1977Q1 to 1987Q3 and hence the first exchange rate forecast produced refers to the change in the exchange rate from 1987Q3 to 1987Q4. A detailed description of the construction of the data set is given in Appendix A.[7]

Some of the data were transformed in natural logarithms prior to beginning the empirical analysis to yield time series for the log of the nominal exchange rate, s_t, US money, m_t, foreign money, m_t^*, US output, y_t, foreign output, y_t^*. We did not take logs for the domestic and foreign interest rates, r_t and r_t^*, and the domestic and foreign trade balance, tb_t and tb_t^*. Our measure of NFA_t is calculated as the difference between the log-detrended purchases and sales of foreign assets, consistent with the definition of Cavallo and Ghironi (2002, p. 1074).

The regressors we consider in employing the procedure described in Section 3 are all suggested by conventional theories of exchange rate determination and have been used in prior research in this context. They include: 1) The once-lagged nominal exchange rate change, in order to allow for any potential slight persistence in exchange rate movements (Backus, Gregory and Telmer, 1993). 2) The deviation from a canonical monetary fundamentals model. This deviation, termed mf_t, is constructed as $mf_t = (m_t - m_t^*) - (y_t - y_t^*)$ (Mark, 1995). 3) The interest rate differential, $(r_t - r_t^*)$, in the spirit of the uncovered interest rate parity condition. 4) The net foreign asset position of the US relative to the foreign country in question, NFA_t. 5) The US trade balance, tb_t. 6) The foreign country's trade balance, tb_t^*. 7) An intercept term. All variables (2)-(6) are

---

[6]Note that, while Canada, Japan and Switzerland have experienced a free float since the collapse of the Bretton Woods system in the early 1970s, the UK was in the Exchange Rate Mechanism (ERM) of the European Monetary System (EMS) for about two years in the early 1990s. However, given the short length of this period, we consider sterling as a freely floating exchange rate in this paper. The only mixed-regime country during this sample is Germany, which was in the ERM for most of the sample period under investigation and in fact joined the European Monetary Union on 1 January 1999, when the euro replaced the German mark as Germany's national currency.

[7]Faust, Rogers and Wright (2003) first employed a real-time data set for exchange rate forecasting, focusing primarily on long-horizon regressions based on monetary fundamentals alone–equation (2) above. Our real-time data set is broader both in terms of variables and in terms of exchange rates examined, and for some of the variables and vintages used here we had no prior study to follow. We are indebted to John Rogers for guidance and advice on the construction of some of the real time data.

lagged once in the empirical investigation discussed below.[8]

As a preliminary exercise, we test for unit root behavior of each of the time series to be used in the predictive framework described in Section 3. Specifically, we test for a unit root in Δs_t, mf_t, $(r_t - r_t^*)$, tb_t, tb_t^* and NFA_t over the full sample. We employ two efficient unit root tests (MZ_α and MZ_t) proposed by Ng and Perron (2001); these tests use generalized least squares-detrending to maximize test power and a modified information criterion to select the lag truncation in order to minimize size distortion. The results–not reported to conserve space but available upon request–suggest that, for each of these time series and for each country examined, the unit root hypothesis is rejected at conventional nominal significance levels. In turn, these results have a clear economic interpretation, confirming that the change in the exchange rate is stationary, the level of the exchange rate co-moves (or cointegrates) with standard monetary fundamentals over the long run, the interest rate differential across these major economies is stationary, and trade balances and net foreign asset positions also revert to some long-run equilibrium level.[9]

5 Empirical Results I: The Empirical Search for the Best Model

5.1 The setup of the empirical search

In our empirical work, we use the procedure described in Section 3 to shed light on several issues. The first question we address is whether fundamentals do contain information that is useful in predicting exchange rate movements. The literature and theoretical arguments presented in Section 2 suggests that the relationship between exchange rate movements and the fundamentals is likely to be characterized by some fundamentals having much more impact than others for certain periods of time and by severe parameter instability. Our predictive framework allows both for the possibility that the set of fundamentals changes over time and for the model parameters to change period by period. Hence, our procedure ought to be able to shed light on the role of these phenomena in understanding the link between exchange rates and fundamentals. The only caveat is that one needs reliable model selection criteria that are capable of selecting, within the broad set of fundamentals considered in the information set, the best empirical exchange rate model. Before carrying out a pure out-of-sample forecasting exercise in real time where models are selected with specific criteria, we carry out a simpler exercise. Specifically, we ask whether the best model (BM) constructed at time t, on the basis of information on the fundamentals available at time t, is a model consistent with the exchange rate disconnect results. In order to carry out this exercise, we implement the predictive procedure described in Section 3 and choose BM to be the model that minimizes the absolute error between the *ex post* realization of the exchange

---

[8]Although the menu of fundamentals used here is broader than the conventional set of fundamentals used in the literature, there are other variables one might have considered. For example, Gourinchas and Rey (2007) suggest that asset to liabilities ratios and export to import ratios are useful predictors of US dollar movements. Also, differentials in prices and in inflation may be used to capture absolute and relative purchasing power parity effects, respectively. While the price level could be thought of as a function of the monetary fundamentals used in this paper, inflation could move in a different way than the level and be linked to exchange rate fluctuations via central banks' reaction functions (e.g., Engel and West, 2006). We leave the exploration of these further fundamentals to future research.

[9]Note, however, that validity of the analysis in this paper *does not require* cointegration. This is because our focus is not on in-sample comparisons of predictive regressions, where spurious regressions invalidate conventional criteria of model comparisons based on statistical significance of parameters and goodness of fit measures. Our forecast evaluation is based purely on out-of-sample criteria based on forecast errors, which remain valid regardless of the existence of cointegration between exchange rates and fundamentals (see Berkowitz and Giorgianni, 2001).

rate change and the model forecast exchange rate change, i.e. we select BM as the model M_i such that

$$\min_i \left| \Delta s_{t+1} - \left(\widehat{\Delta s}_{t+1,i} \mid \Omega_t \right) \right| \qquad t = \tau, \tau + 1, \ldots, T - 1 \qquad (6)$$

where $\widehat{\Delta s}_{t+1,i}$ is the one-quarter-ahead out-of-sample forecast produced by model M_i conditional on the information set available at time t, Ω_t. BM, constructed according to this rule, is the model which performs better in forecasting the one-quarter-ahead exchange rate and is, therefore, the model that would have been selected if one had a criterion which could perfectly select, within a pool of (misspecified) models, the best combination of the seven regressors available.

We carry out the exercise described above for each dollar exchange rate considered, and compare the results for the BM obtained from our set of fundamentals to the scenario that one would obtain from using a simple random walk (RW) model. This scenario is of course an abstraction from reality. However, this analysis is useful in establishing how well (poorly) one can forecast the exchange rate using economic fundamentals within our framework relative to the case where the exchange rate is assumed to follow a RW. For each scenario, we calculate the MAE and the RMSE. If it is not possible to forecast exchange rates (better than a RW model) with such empirical search, then there would be no point implementing our procedure in real time since we would already know that the outcome would be a failure of our broad set of fundamentals to forecast exchange rates.

Note that this setup is different from assuming perfect foresight because the BM is a model made of fundamentals information and is potentially (indeed almost certainly) misspecified. In other words, the exchange rate forecasts $\widehat{\Delta s}_{t+1,i}$ are generated from each model M_i conditional only on information available at time t so that these models could have been estimated and the forecasts used in real time by an investor during the sample. If such fundamentals information and the exchange rate models are of no value in predicting the exchange rate, or if the true DGP of the exchange rate is a random walk, then *none* of the models examined will outperform a random walk benchmark in terms of forecast accuracy even using model selection criterion (6). In fact, in our empirical work we find cases where the BM significantly outperforms a RW model, as well as cases where this does not occur. The exercise is also different from an in-sample evaluation since the models are used to forecast the exchange rate conditional only on information available at the time of the forecasts. However, this setup is subject to forecasting biases, to which we now turn.

5.2 Forecasting Biases in the Empirical Search

The setup of the empirical search uses conventional forecasting methods for generating out-of-sample exchange rate forecasts from $2^7 = 128$ models.[10] However, what is distinct in this exercise is that the BM selection is exploiting *ex post* information on the basis of criterion (6), so that one can guarantee that the BM is selected, out of the 128 models considered, with no uncertainty. As discussed earlier, the problem of data snooping bias needs to be taken into account seriously in order to be sure that this setup does not bias the results into finding favorable forecasts. In our framework, where there are 128 models estimated and evaluated at each point in time, the effects of data snooping can be dramatic, especially in the case of the empirical search from the best fundamentals model.

---

[10] For the Swiss franc the number of models is $2^6 = 64$ as we only have six regressors due to the unavailability of Swiss trade balance (tb^*) data over the full sample.

In order to intuitively explain the danger of data snooping biases when employing criterion (6) in our context, let us use an example.[11] Suppose that we are trying to predict the outcome of a fair coin toss (i.e. we are trying to predict something unpredictable). Let the coin to be predicted be called coin A. The two competing predictors are two other fair coins, say coins B and C. All three coin flips are independent. Then using our methodology and analyzing forecasts *ex post*, we would find that we can predict the outcome of a flip in coin A 75 percent of the time. This is because the only time we won't have a correct prediction from at least one of coins B and C is when A=heads and B=C=tails, or A=tails and B=C=heads. There are 8 possible configurations of coin flip results for the three coins, so those 2 events occur with probability 2/8; the implication is that, *ex post*, the model search works 6/8=75 percent of the time. Naturally, however, if one performed real time, *ex ante* forecasting, it would not be possible to beat a 50 percent success rate. Therefore, evaluation of forecasting performance when forecasts are obtained from a large pool of models from which a single model is selected requires care in allowing for these biases.

Our framework is a special case of the class of data snooping problems studied by White (2000), whose Reality Check procedure we employ in this paper. The Reality Check implemented is described in detail in Appendix B. In essence, this procedure is designed to allow aggressive model searching to be undertaken with confidence that one will not mistake results that could have been generated by chance for genuinely good results. The Reality Check test evaluates the distribution of a performance measure accounting for the full set of models that lead to the best-performing model and for the specific model selection criterion employed in a forecasting setup, where one is interested in testing the null hypothesis that the BM selected in a specification search has no predictive superiority over a given benchmark model (in our case the RW model). This is the null hypothesis of equal forecast accuracy. White (2000) proposes and experiments with several simulation methods to implement the Reality Check. The procedure used in this paper is based on a Monte Carlo algorithm which allows to determine the p-value of the test statistic for the null hypothesis by incorporating the effects of data snooping from the search within a large pool of models.[12] One important feature of White's procedure is its generality, in the sense that when there is no data snooping bias (e.g. there is only one model under consideration) the procedure delivers an asymptotic p-value that is identical to the p-value one would obtain with procedures of the kind suggested by Diebold and Mariano (1995) and West (1996). However, the Reality Check corrects for the downwards bias in the p-value induced by the latter procedures in the presence of data snooping. Hence, the procedure is general enough to make researchers confident of the p-value calculated in forecasting exercises in a variety of settings.

In order to re-assure ourselves that there are no foresight or data snooping biases when we apply our empirical search and the selection criterion (6), before moving to the empirical work, we examine whether the Reality

---

[11]This example is due to Ken West, who used it in private correspondence and kindly allowed us to report it here.

[12]Other methods to carry out the Reality Check include various bootstrap methods for dependent data. Our choice to employ the Monte Carlo method is motivated by the fact that with quarterly data the dependence in exchange rates data disappears easily with one lag and that the distribution of exchange rate changes does not display statistically significant departures from normality. Under these conditions–lack of dependence and normality–Monte Carlo is expected to be both reliable and computationally efficient. This is consistent with the simulation results presented below, which suggest very satisfactory size properties of the test in our context. Also, although we employ White's procedure in this paper to protect ourselves from data snooping, in practice the danger of data snooping is unlikely to be particularly severe when the number of models is small. In essence, there are only a few theoretical models encompassed by the variables used in our set of fundamentals, whereas White (2000) used thousands of models and over a thousand observations. In the case of the present paper, it may well be that other methods for assessing forecasting accuracy which are less computationally demanding could be equally appropriate (e.g., see West, 2006).

Check procedure suggested by White (2000) works in our context. To this end, we employ the empirical search replacing the observed exchange rate data with artificial data simulated under a driftless random walk with the same number of observations as in the empirical analysis, with the variance calibrated on the Japanese yen.[13] In other words, we generate data from a DGP that is a driftless RW and apply the empirical search and the Reality Check to these data, following the steps described in Appendix B. Using these artificial data, if the Reality Check works well in accounting for forecasting biases, one should not find any evidence of superior predictability of the BM relative to the RW benchmark. This is because, if the true DGP is a RW, then the procedure should find that no model can outperform the RW model. Essentially, this exercise gives the empirical size of the test, i.e. the probability to reject the null hypothesis of equal forecast accuracy between the BM and the random walk model when the null hypothesis is true. These results are presented in Table 1 at three conventional significance levels. The empirical size is, in each case, close to the significance level examined, indicating at most a very tenuous size distortion. This is a crucial robustness exercise on the validity of the Reality Check procedure implemented in our specific context, which indicates that if exchange rates are indeed not predictable (random walks) then, even with our large pool of models, one would not be able to find statistical evidence against the RW benchmark when accounting for foresight and data snooping biases with the Reality Check.

We also document in Table 1 the empirical power of the test procedure for various potential departures from the null hypothesis–that is the percentage of rejections of the null hypothesis of equal forecast accuracy when the null is false. First, we examine the case where the random walk has a non-zero drift term–in one case (RWD1) a very small drift, calibrated on the Japanese yen, and in another case (RWD2) a drift term that is three time larger than the one in RWD1. The results suggest that for RWD1 the empirical power is extremely low for RWD1, presumably because the drift is very close to zero. This interpretation is consistent with the finding of a higher empirical power for RWD2, which has a larger drift term. Second, we examine the power when the DGP for the changes in the exchange rate is first-order autoregressive of order one, or AR(1). The power of test is now higher, and appears to increase with the size of the AR parameter (the size of the departure from the null hypothesis), as one would expect. Third, we consider three stationary AR(1) processes for the DGP of the level of the exchange rate. In these cases, the exchange rate ought to be predictable since it displays mean reversion to a constant equilibrium level over time, but the degree of predictability will depend on how close the AR(1) coefficient is to unity. We examine this possibility in light of the recent results of Sweeney (2006), who provides evidence against the notion that floating exchange rates contain unit roots and that, in some out-of-sample forecasts, mean-reversion models beat random walks significantly. Consistent with our priors, the empirical power of the testing procedure is relatively high when the autoregressive parameter is 0.3 (a large departure from the null hypothesis) and, to a lesser extent, 0.5, but it is much lower for a parameter of 0.8 (a relatively small departure from the RW null). Finally, we examine the case where the true DGP is a driftless random walk with a time-varying, ARCH error variance (RW-ARCH). In this scenario, the conditional variance of the exchange rate is predictable but the conditional mean is not. Therefore, although the RW-ARCH is different from the homoskedastic RW considered as the benchmark model under the null, the power

---

[13]Indeed, all of the simulation exercises in this section are calibrated on the dollar-yen data. However, since there is nothing particularly distinct about the properties of the dollar-yen relative to the other dollar exchange rates examined, we see no reason why these results should suffer from problems of specificity.

of the testing procedure should in this case be equal to the size of the test if the procedure works correctly. This is interesting to investigate since one may be concerned that the predictability of the conditional variance may confound tests of predictability on the conditional mean, generating spurious predictability. However, the results, reported for three sets of ARCH parameters, indicate that the rejection rate is very low and close to the test size.

Overall, these simulation results suggest that the empirical size of the Reality Check procedure proposed by White (2000) does not display size distortions, and that the empirical power is moderately high for large departures from the null hypothesis of equal forecast accuracy. This increases our confidence that the empirical search procedure is not prone to forecasting biases towards favorable results for the fundamentals models. We now turn to the results from implementing the empirical search on actual exchange rate data.

5.3 Results

The empirical results from applying the empirical search exercise to our dollar exchange rate data are presented in Table 2. The MAE and RMSE for the BM are lower than the corresponding MAE and RMSE from a RW model. However, the improvement in terms of MAE and RMSE is statistically significant only for the Japanese yen and the UK sterling at the 5 percent significance level, and at the 10 percent level for the Swiss franc, when p-values are calculated using the Reality Check. Therefore, on the one hand, for three major dollar exchange rates, our exercise suggests results in contrast with the view that the information embedded in economic fundamentals have no predictive power for the exchange rate. On the other hand, for two exchange rates, namely the Canadian dollar and the German mark, we find that the RW cannot be outperformed by our recursive procedure, even with the huge informational advantage used in our procedure. While for the German mark some results in favor of predictability exist in the literature (e.g. Mark, 1995), with respect to the Canadian dollar our results are not surprising. It is well documented that conventional fundamentals do not play an important role in explaining short-term fluctuations in the Canadian dollar and that a key short-run explanatory variable is commodity prices, reflecting the nature of this exchange rate as a classic commodity currency (e.g. see Amano and van Norden, 1995; and Chen and Rogoff, 2003). Given these results, it would be impossible to outperform the RW when doing the exercise *ex ante*, i.e. without the informational advantage provided by the use of model selection criterion (6). From this point onwards, therefore, we do not attempt to forecast the Canadian dollar and the German mark and accept the fact that our broad of set of fundamentals will not be capable of producing a model that is better than a RW. Instead, we shall focus entirely on the analysis of the remaining three exchange rates.

The accuracy of the forecasts achieved from the BM for the dollar against the Japanese yen, UK sterling and Swiss franc is made visually clear by the graphs in Figure 1, which presents the actual exchange rate change over the forecast horizon and the forecasts produced by the BM for each of these exchange rates. The graphs make apparent how the exchange rate forecasts from the BM track the exchange rate closely even though they display lower variability. This lower variability is of course expected as a conditional expectation has lower variability than the realization of the variable being forecast, and can be attributed to either noise that is unrelated to fundamentals or to a missing fundamental. However, while the figure confirms the stylized fact that the exchange rate is more volatile than the exchange rate predicted with fundamentals information, such

excess volatility is much lower than what is typically recorded in the relevant literature and does not prevent the BM fundamentals model from predicting fairly accurately exchange rate changes and more accurately than a RW benchmark.

5.4 How Frequently Does the Best Exchange Rate Model Change?

A logical question to ask relates to the implications of these results for the identification of the true DGP governing the dynamics of these three exchange rates. To address this issue we calculated the number of times the procedure outlined above selected the random walk (with or without drift) as the BM. We find that, over the forecasting period and for all exchange rates examined, the RW model was selected at most once; and in one case, the UK sterling, the RW model was never selected as the BM. This finding suggests that the true DGP of the exchange rate is not represented by a pure RW and there are gains from appropriately combining economic fundamentals over time in order to exploit their information content to predict the exchange rate out of sample.

Given the apparent ability to produce accurate exchange rate forecasts under the BM scenario, one then wonders which fundamentals are driving this result and how often they are switching. Figure 2 plots the 'inclusion frequency' (IF) for each of the regressors for the dollar-yen exchange rate, which we use as a representative exchange rate. The IF is a binary time series which indicates when a regressor is included in the BM, taking the value of unity, and when it is not included, taking the value of zero. The graphs of the IF for each of the seven regressors suggest that all of them have some importance in the BM, but they are hardly ever important for a very long time. In other words, the model using the information set available to the investor (forecaster) optimally changes quite often, as suggested by the high number of switches that characterize each of the explanatory variables in the regression. This is consistent both with the frequent swings in expectations that Frankel (1996) blames for the difficulty to explain exchange rate movements with standard fundamentals models, as well as with the story about higher-order beliefs that is behind the scapegoat model of Bacchetta and van Wincoop (2004) and their conjecture that agents in the foreign exchange market often change the weights attached to fundamentals in their models. Investigation of the persistence of the IFs shows that on average, across economic fundamentals employed and exchange rates investigated, the variables in the information set are selected for about two quarters. However, this persistence exhibits a very diverse behavior when looking at different variables one at a time. In the case of the Japanese yen, for example, the highest persistence is recorded by the IF of the deviation from monetary fundamentals (about 4 quarters), followed by the net foreign asset position and the Japanese trade balance (2.4 and 2.3 quarters respectively). The least persistent variables are the interest rate differential and the US trade balance, which exhibit an average IF of 1.9 and 1.3 quarters respectively.

Overall, across the three exchange rates for which the BM beats a RW, we can identify two groups of variables: one, which we may label 'short-term fundamentals', that has a frequent impact on the exchange rate but only for very short periods of time (i.e. one or two quarters). In this group we find the US trade balance and, in most cases, the interest rate differential. There is another group of variables (say, 'long-term fundamentals') whose switches are less frequent but more persistent (more than 2 quarters and up to 1 year). In this group we find the deviations from monetary fundamentals, net foreign asset positions and the foreign

15

trade balance. This simple descriptive evidence suggests that investors revise their expectations at discrete intervals, generating over time frequent switches in the predictive regressions.

Given that switches for some fundamental variables appear to occur over time at a relative high pace, it is interesting to understand if they exhibit persistence of some kind or are independently distributed. This is important because, if the shifts are independently distributed, it would be impossible to predict the BM. To address this issue we used a fairly general test for the null hypothesis that the switches are independent and identically distributed against an unspecified form of dependence, namely the Brock, Dechert, Scheinkman and LeBaron (1996) test (BDSL). The results, reported in Table 3, show that for all exchange rates examined, the null hypothesis of independence of all switches is strongly rejected with p-values virtually equal to zero. This evidence is also generally confirmed for individual variables, where the null hypothesis of independence is rejected in the majority of cases.[14]

5.5   Summing Up

To summarize, the results in this section suggest that the information in the fundamentals, when used in a framework that allows for the set of fundamentals included in the model to change over time, can forecast some exchange rates with a reasonable degree of accuracy. However, this requires the ability to select the best model from the various set of models that can be used on the basis of information available, i.e. this requires that the investor must have a reliable model selection criterion to discriminate among different specifications of the fundamentals model. The selection criterion would have to be able to generate a large number of switches in the model since the BM scenario clearly indicates that the weight attached to a fundamental can vary frequently over time.

6   Empirical Results II: Real-Time Forecasting

6.1   Can We Select the Best Exchange Rate Model in Real Time?

We now turn to the empirical implementation of the recursive predictive procedure employing the model selection criteria described in Section 3. Given the results from the exercise carried out in the previous section, if these model selection criteria perform satisfactorily, an investor forecasting three major dollar exchange rates should have been able to perform rather well over the sample 1987-2003. In essence, we generate conventional out-of-sample forecasts one quarter ahead according to the recursive procedure of Pesaran and Timmermann (1995) described in Section 3, namely conditional only upon information up to the data of the forecast and with successive re-estimation as the date on which forecasts are conditioned moves through the data set. In each period, the forecast model is selected using a model selection criterion.

The results from implementing the recursive predictive procedure provide very different evidence from the empirical search conducted in the previous section. In Tables 4 and 5 we report the out-of-sample results, for each model selection criterion considered. We report in Table 4 the MAE and the RMSE, while in Table

---

[14]The p-value for the null hypothesis of joint independence is calculated using Bonferroni bounds. In particular, for arbitrary correlations among BDSL statistics, the upper bound p-value for the joint test is given by $p \leq \min(k \cdot \min(p_1, \ldots, p_k), 1)$ providing the strongest evidence against the null hypothesis (see Kounias, 1968; Hunter, 1976; Galambos and Simonelli, 1996; Sullivan, Timmermann and White, 1999 and the references therein).

5 we report the forecasting results on the basis of the economic criteria of evaluation described in Section 3 (i.e. EOPW and RR). The results indicate that none of the model selection criteria (listed in the top row of Tables 4-5) can replicate, even remotely, the best model that uses optimally the information in the fundamentals available in the information set of the investor. The differences in MAEs and RMSEs relative to a RW are miniscule, and the two economic criteria (EOPW and RR) deliver similar unsatisfactory outcomes.[15]

While the comparison of the recursive predictive procedure with the best model may be seen as unfair given the huge informational advantage offered to the empirical search carried out in Section 5, the comparison with the RW model is a reasonable one, given that the RW model is the standard benchmark in the relevant literature. The latter comparison reveals that, in general, the performance of the RW model is very similar to the performance of the recursive predictive procedure (irrespective of the model selection criterion) for all exchange rates examined. We formally test the null hypothesis that the recursive procedure and the RW model have equal forecast accuracy and report the p-value for this test in parentheses. The p-values–calculated as described in Appendix B–are, for each exchange rate and selection criterion, comfortably larger than conventional nominal significance levels, suggesting that the recursive predictive procedure based on our set of fundamentals is, apart from few scattered exceptions, indistinguishable in terms of out-of-sample performance from a RW model that uses no information about the state of economy. In short, we confirm the stylized fact in the empirical literature that fundamentals models cannot beat a random walk model in exchange rate forecasting. In fact, we refine this stylized fact and prove its robustness even in the context of a recursive procedure that allows for a rich set of fundamentals and allows the model specification (fundamentals) to change period by period.

6.2 Why Can't We Use Effectively the Information in the Fundamentals?

These results beg the question: why is there such a striking contrast between the results in the empirical search (Table 2) and the results in the real-time procedure? The most logical answer ought to be that the model selection criteria employed fail to select the BM produced in the empirical search because the analysis in Section 5 differs from the analysis in Section 6 only due to the model selection criterion (i.e. the pool of models and forecasts are the same in these two exercises).

In order to shed light on the key differences between these two sets of results, we present in Figure 3 plots of the inclusion frequency (IF) for each of the regressors and for the case where the $\overline{R}^2$ is the model selection criterion, which may be seen as a representative criterion. The resulting IFs, calculated as in Section 5, suggest that all of the regressors except the lagged exchange rate are included in the selected model over the forecast sample. However, the comparison with the corresponding IFs from the BM given in Figure 2 reveals clearly that the model selection criterion tends to keep a fundamental in the model for many years, inducing very high persistence in the IFs. The number of switches using the $\overline{R}^2$ criterion is much smaller than the number of switches recorded under the BM. This is especially true for the interest rate differential and net foreign assets, which are included in the model throughout the forecast period, and also for the drift term, which is included ever since 1988. The persistence of the IFs suggests that using conventional model selection criteria it is not

---

[15]Some of the statistically significant MAE and RMSE may surprise, at first glance, in the sense that differences in MAE and RMSE which relate to the third decimal point may be found to be statistically significant. However, it is worth noting that the empirical distribution of the test of equal forecast accuracy is highly non-normal and that the literature often finds statistical significant differences when these relate to the fourth decimal point (e.g., see Mark, 1995; Kilian, 1999; McCracken and Sapp, 2005).

possible to reproduce the frequent swings that, according to our empirical search, characterize the sort of model that can explain future movements in the exchange rate. The results obtained using standard criteria would be in contrast with the existence of frequent shifts in expectations advocated by Frankel (1996), or the scapegoat story of exchange rate determination of Bacchetta and van Wincoop (2004), as well as with the survey evidence that agents in the foreign exchange market *often* change the weight they attach to a fundamental (Cheung and Chinn, 2001). The model selection criteria tend to select over-parameterized, larger exchange rate models than they should do.

6.3   Summing Up

Overall, taking together the results in Sections 5 and 6, it is clear that the information embedded in the conventional menu of fundamentals may be useful in forecasting exchange rates out of sample. In fact, the forecasts that can be produced by selecting the best set of fundamentals at each point in time can be very satisfactory on the basis of conventional criteria of forecast evaluation. However, the model that optimally uses the information in the fundamentals changes often and is consistent with frequent shifts in the parameters. Standard real-time model selection criteria appear to be unable to generate such shifts, yielding empirical exchange rate models that cannot forecast the exchange rate better than a random walk model. Therefore, while the stylized fact that fundamentals models cannot beat a no-change model is confirmed in this paper, the reason for this result is, on the basis of the evidence presented here, different from many other studies. Our evidence suggests that the exchange rate disconnect puzzle arising from our results is unlikely to be caused by lack of information in the fundamentals, and more likely to be the result of poor model selection criteria in this context.

7   Robustness

In this section we report some of the robustness checks carried out to assess the sensitivity of the results reported in Sections 5-6. First, we repeat the exercises carried out in Sections 5-6 using historical data for all fundamentals that are not in real time, i.e. using the revised data available at the present time but not at the time the investor was making the forecasts. Faust, Rogers and Wright (2003) first examined the importance of using data in real time in the context of long-horizon predictability regressions of the type used by Mark (1995). Using dozens of different data vintages for money and output, Faust, Rogers and Wright (2003) show that the predictability found in Mark's study tends to be present only during a two-year window of data around the time of that study. Predictability appears to be weaker during other time periods. However, Faust, Rogers and Wright also show that models using data in real time perform better in out-of-sample forecasting than models using revised data. In Table 6, we report the results for the dollar-yen exchange rate, as a representative case. Panel A gives the results for the empirical search and hence these results should be compared to the corresponding results in Table 2. Such comparison reveals that using revised data has no qualitative impact on the results for the best model. Also, the results when the forecasting model is selected using standard statistical criteria (Panel B of Table 5) reveals that using revised data generally tends to produce similar results to the ones in Table 4, so that it is difficult to decide whether the results using revised data are better or worse than the results using vintages. However, the differences are quantitatively small and qualitatively unimportant.

A second robustness check involves using a rolling, rather than recursive, procedure for model selection. One may argue that, if the recursive procedure employed in the core of the paper fails to forecast exchange rates because of its inability to generate enough switches in the forecasting model, then this problem may be mitigated by employing a rolling scheme for forecasting since this relies less on the distant past. On the other hand, the rolling scheme is subject to a constant estimation risk (parameter uncertainty), whereas in the recursive procedure the estimation risk reduces as the number of observations increases. We repeated the empirical analysis using a rolling forecasting scheme with a rolling window of 5 years (i.e. 20 quarterly observations). The results, given in Table 7 using again the dollar-yen as representative exchange rate, suggest that there are negligible differences in the case of the empirical search (Panel A), while for the real-time exercise (Panel B) the rolling forecast scheme tends, in general, to perform worse than the recursive scheme. These differences are, however, very marginal.

A third robustness exercise involves adding a second lag in each of the regressors used in the information set. In order to avoid a drastic increase in the number of models to be estimated, we assume that, although the parameters on the first and second lag for a fundamental are different, they switch together–i.e. switches remain associated with each fundamental, and each fundamental is now identified as two lags of the variable in question. The results from this exercise, reported again for the dollar-yen exchange rate (Table 8), suggest that the addition of a second lag induces at best marginal improvements in the forecasting performance.

Overall, the robustness exercises reported in this section indicate that the key results reached in Sections 5-6 are robust. These results are qualitatively unaffected by the use of revised data, by the use of a rolling–rather than recursive–forecasting scheme, and by the addition of an extra lag in the fundamentals.[16]

8   Discussion and Conclusions

This study focused on an empirical framework which explicitly takes into account the fact that there are changing weights in the economic fundamentals driving exchange rates and that no model of fundamentals appears to perform well for long periods of time. Our framework allows for a menu of fundamentals that goes beyond the standard monetary fundamentals most commonly used in the literature since Mark (1995), comprising other variables suggested by exchange rate determination theory, including net foreign assets, interest rates differential and the trade balance. Using a new set of real-time quarterly data spanning the period from 1977Q1 and 2003Q3, we employed a recursive procedure which selects, quarter by quarter, the best model on the basis of a variety of criteria within all possible combinations of fundamentals, allowing for the fundamentals model and the parameters to change over time.

Our main findings are as follows. First, the information embedded in the economic fundamentals, when used in a framework that allows for the set of fundamentals in the model to change over time and for parameter instability, can forecast the exchange rate with a remarkable degree of accuracy in some cases. However, this requires the ability to select the best model from the various set of models that can be used on the basis of information available, i.e. this requires that the investor must have a reliable model selection criterion to

---

[16] A further check we carried out involves estimating the exchange rate equations using the seemingly unrelated regressions (SUR) estimator in a panel for all of the exchange rates examined in order to obtain more precise estimates of the parameters by exploiting the cross-correlation in the covariance matrix (see Mark and Sul, 2001). However, the results–not reported to conserve space–were qualitatively identical to the core results reported earlier.

discriminate among different specifications of the fundamentals model. Second, if the conventional set of model selection criteria is used to pick up the best model from the large pool of models available, the same set of economic fundamentals is not useful in forecasting exchange rates out of sample.

We show that models that optimally use the information in the fundamentals change often and this implies frequent shifts in the parameters. Standard model selection criteria appear to be unable to generate such shifts, yielding empirical exchange rate models that cannot forecast the exchange rate better than a random walk model. Therefore, while the stylized fact that fundamentals models cannot beat a no-change model is–yet again–confirmed in this paper, the reason for this result is, on the basis of the evidence presented here, different from many other studies. Our evidence suggests that the exchange rate disconnect puzzle arising from our result is unlikely to be caused by lack of information in the fundamentals, and more likely be the result of poor model selection criteria in this context.

These findings can be connected to much existing literature. Engel and Hamilton (1990) recorded that long swings characterize the behavior of nominal exchange rates and, by taking them into account and estimating regime-switching models, it is possible to generate better forecasts than a random walk. This result has been confirmed in other contributions where different fundamentals have been employed (see, *inter alia*, Clarida, Sarno, Taylor and Valente, 2003, 2006, and the references therein). However, these forecasting results appear to be somewhat fragile. In fact, the literature on nonlinear modelling of exchange rates has produced models that, albeit fitting satisfactorily in sample, generally fail to beat simple random walk models or linear specifications in out-of-sample forecasting (e.g. see Diebold and Nason, 1990; Engel, 1994; Meese and Rose, 1990, 1991). This last evidence can be interpreted in light of our empirical results in that standard model selection criteria appear to be unable to detect and correctly identify shifts in the state variables (i.e. economic fundamentals). In fact, the existing empirical literature has focused primarily on the identification of long-swings in exchange rate data. Our results suggest that in order to enhance the forecasting performance of predictive regressions it is necessary to combine long-swings with short and recurrent swings in the predictive variables. Unfortunately, to date, econometric techniques able to thoroughly address this issue are not available and new techniques are awaited. Understanding the exact nature of these shifts remains an important challenge in this research agenda.

Exchange rate forecasting remains just as hard today as in the seminal work of Meese and Rogoff (1983), and the exchange rate is often well approximated by a driftless random walk for prediction purposes. Even trying our best in maximizing information available in a broad set of economic variables, we were unable to produce satisfactory out-of-sample results and use the available information optimally. While this indicates the inability of model selection criteria to discriminate among models, it also provides a potential explanation for the relative success of forecasting methods that combine the information in a large number of models, such as simple averaging of the forecasts of different models. Averaging across models is likely to produce better forecasts than using a model selected using standard selection criteria if combining models somehow 'averages out' the errors made by various misspecified models.[17] Finally, while there has been considerable progress in the design of sophisticated theoretical and empirical models of exchange rates, the evidence presented here suggests that overturning the Meese-Rogoff results requires that more research effort and resources be devoted to improving estimation and model selection procedures.

---

[17] This is consistent with recent evidence provided by Wright (2003) and Della Corte, Sarno and Tsiakas (2008), using Bayesian model averaging methods to out-of-sample exchange rate forecasting.

Table 1. Empirical size and power properties of the Reality Check procedure

		Size			Power									
		RW	RWD1	RWD2	DAR			AR				RW-ARCH		
					$\phi_1=-0.2$	$\phi_1=0.2$	$\phi_1=0.3$	$\rho_1=0.3$	$\rho_1=0.4$	$\rho_1=0.5$	$\rho_1=0.8$	$\varphi_1=0.3$	$\varphi_1=0.5$	$\varphi_1=0.8$
MAE	1%	0.012	0.014	0.031	0.078	0.083	0.249	0.846	0.643	0.429	0.056	0.016	0.013	0.014
	5%	0.058	0.058	0.111	0.187	0.226	0.435	0.947	0.852	0.680	0.162	0.051	0.045	0.043
	10%	0.102	0.103	0.176	0.271	0.317	0.547	0.974	0.906	0.779	0.251	0.093	0.089	0.091
RMSE	1%	0.010	0.011	0.014	0.057	0.063	0.209	0.861	0.684	0.493	0.047	0.010	0.010	0.010
	5%	0.055	0.057	0.085	0.164	0.191	0.419	0.943	0.876	0.721	0.183	0.020	0.051	0.054
	10%	0.125	0.131	0.155	0.259	0.287	0.562	0.976	0.933	0.827	0.291	0.080	0.111	0.114

Notes: The table reports the results of a Monte Carlo experiment where the null hypothesis that the log-level of the exchange rate follows a driftless random walk with constant variance is true (Empirical Size) or false (Empirical Power). Figures reported are rejection rates of the null hypothesis of no predictability under various possible DGPs across replications–i.e. they are equal to $\frac{\text{number of rejections of the null}}{\text{number of replications}}$. All the simulations are calibrated on data for the Japanese yen, and the number of replications is $1,000$. MAE and RMSE denote the mean absolute forecast error and root mean square forecast error respectively. RW, RWD1, RWD2 denote the DGPs of a driftless random walk (RW) and two random walks with drift; the drifts under RWD1 and RWD2 are set equal to the sample mean and three times the sample mean, respectively, of the first-difference of the dollar-yen exchange rate. The DAR DGP is a first-order autoregression for the first difference of the exchange rate, i.e. $\Delta s_{t+1} = \phi_0 + \phi_1 \Delta s_{t+1} + error$. AR is a first-order autoregression for the level of the exchange rate, i.e. $s_{t+1} = \rho_0 + \rho_1 \Delta s_{t+1} + error$. RW-ARCH is a driftless random walk (as in RW) with errors following an ARCH(1) process, i.e. $\Delta s_{t+1} = \eta_t; \eta_t = error \cdot \sqrt{1 + \varphi_1 \eta_{t-1}^2}$. The error terms under all DGPs have been simulated from a normal distribution with zero mean and variance calibrated on dollar-yen data. The p-values are computed by using the White's (2000) Reality Check procedure as indicated in the text and described in Appendix B.

Table 2. Results from the empirical search

		BM	RW
Japan	MAE	0.023	0.053
		(0.008)	(−)
	$RMSE$	0.039	0.065
		(0.019)	(−)
UK	MAE	0.017	0.039
		(0.037)	(−)
	$RMSE$	0.032	0.051
		(0.029)	(−)
Canada	MAE	0.012	0.019
		(0.450)	(−)
	$RMSE$	0.020	0.026
		(0.526)	(−)
Switzerland	MAE	0.030	0.054
		(0.075)	(−)
	$RMSE$	0.048	0.068
		(0.081)	(−)
Germany	MAE	0.026	0.052
		(0.281)	(−)
	$RMSE$	0.040	0.064
		(0.232)	(−)

Notes: MAE and RMSE denote mean absolute forecast errors and root mean square forecast errors respectively. BM is the best model selected by the empirical search; RW is the random walk model. Figures in parentheses are p-values of the null hypothesis of equal predictive accuracy between BM and RW. p-values are computed by using a Reality Check procedure as indicated in the text and described in Appendix B.

Table 3. Independence of the inclusion frequencies

	ALL	Δs_{t-1}	mf_{t-1}	$(i-i^*)_{t-1}$	NFA_{t-1}	tb_{t-1}	tb^*_{t-1}	c
				$m=3$				
Japan	0	0	0.034	0.551	0	0.959	0.204	0.033
UK	0	0.333	0.059	0	0	0.267	0.001	0.774
Switzerland	0	0	0	0	0	0.896	NA	0
				$m=4$				
Japan	0	0	0.030	0.946	0	0.660	0	0.003
UK	0	0.470	0	0.977	0	0	0	0.835
Switzerland	0	0.003	0	0	0.527	0.626	NA	0

Notes: The figures reported are p-values of the BDSL test for the null hypothesis that the switches relative to the variables in each column are iid (Brock, Dechert, Scheinkman, LeBaron, 1996). The critical values, from the normal distribution, are 1.960 and 2.576 at the five percent and one percent nominal levels of significance respectively. Given that the choice of the 'embedding dimension' m is crucial for the power of the test, we report the results for $m = 3, 4$, as suggested by Brock, Hsieh and Lebaron (1991) and Brock, Dechert, Scheinkman, LeBaron (1996). ALL denotes the p-value for the null hypothesis that the switches in all fundamentals are jointly independent; these p-values are calculated by using Bonferroni bounds as discussed in the text. 0 denotes a p-value lower than 10^{-8}. NA denotes cases where the variable was not included in the model due to data availability.

Table 4. Results from the real-time out-of-sample exercise: statistical criteria of evaluation

		MAE	RMSE	LAE	$\overline{R}^2$	AIC	BIC	SIGN	PIC	FIC	RW
Japan	MAE	0.052	0.050	0.057	0.052	0.051	0.056	0.056	0.050	0.050	0.053
		(0.015)	(0.003)	(0.534)	(0.007)	(0.003)	(0.829)	(0.298)	(0.003)	(0.003)	(—)
	RMSE	0.068	0.063	0.071	0.063	0.064	0.071	0.068	0.062	0.062	0.065
		(0.170)	(0.017)	(0.713)	(0.062)	(0.105)	(0.969)	(0.313)	(0.018)	(0.018)	(—)
UK	MAE	0.043	0.045	0.470	0.045	0.045	0.040	0.044	0.045	0.045	0.039
		(0.678)	(0.904)	(0.995)	(0.939)	(0.973)	(0.754)	(0.908)	(0.900)	(0.900)	(—)
	RMSE	0.056	0.059	0.061	0.059	0.057	0.056	0.057	0.059	0.059	0.051
		(0.682)	(0.923)	(0.996)	(0.945)	(0.917)	(0.949)	(0.903)	(0.920)	(0.920)	(—)
Switzerland	MAE	0.0546	0.053	0.056	0.054	0.058	0.054	0.054	0.053	0.053	0.054
		(0.223)	(0.043)	(0.372)	(0.075)	(0.769)	(0.404)	(0.110)	(0.046)	(0.046)	(—)
	RMSE	0.072	0.066	0.073	0.066	0.071	0.069	0.068	0.065	0.065	0.068
		(0.558)	(0.080)	(0.752)	(0.055)	(0.682)	(0.698)	(0.114)	(0.071)	(0.071)	(—)

Notes: The first row gives the model selection criteria while the second column indicates the forecast evaluation criteria. MAE, RMSE and LAE denote mean absolute error, root mean square error and the last absolute error respectively. SIGN is a sign criterion based on the directional accuracy of the in-sample predictions of the exchange rate change. AIC, BIC, PIC, FIC are the Akaike Information Criterion, the Schwarz Information Criterion, the Posterior Information Criterion and the Fisher Information Criterion respectively. RW denotes the benchmark random walk model. Figures in parentheses are p-values of the null hypothesis of equal forecast accuracy between the models selected according to the criterion indicated in each column and RW. p-values are computed by using the Reality Check procedure described in Appendix B.

Table 5. Results from the real-time out-of-sample exercise: economic criteria of evaluation

		MAE	RMSE	LAE	$\overline{R}^2$	AIC	BIC	SIGN	PIC	FIC	RW
Japan	EOPW	3.106	3.492	2.044	5.082	3.571	2.434	2.563	3.492	3.492	3.379
		(0.486)	(0.599)	(0.218)	(0.904)	(0.505)	(0.737)	(0.309)	(0.588)	(0.588)	(−)
	RR	0.307	0.337	0.202	0.434	0.341	0.248	0.258	0.337	0.337	0.328
		(0.512)	(0.592)	(0.267)	(0.814)	(0.596)	(0.241)	(0.369)	(0.581)	(0.581)	(−)
UK	EOPW	1.970	1.438	2.035	1.979	1.828	3.578	1.629	1.438	1.438	3.165
		(0.257)	(0.147)	(0.251)	(0.242)	(0.163)	(0.677)	(0.924)	(0.149)	(0.149)	(−)
	RR	0.230	0.136	0.236	0.231	0.207	0.414	0.171	0.136	0.136	0.374
		(0.255)	(0.119)	(0.260)	(0.263)	(0.150)	(0.649)	(0.123)	(0.125)	(0.125)	(−)
Switzerland	EOPW	2.075	3.694	1.361	3.800	1.780	2.038	4.038	3.694	3.694	3.964
		(0.138)	(0.489)	(0.081)	(0.503)	(0.096)	(0.123)	(0.586)	(0.497)	(0.497)	(−)
	RR	0.201	0.338	0.105	0.346	0.167	0.197	0.358	0.338	0.338	0.357
		(0.181)	(0.519)	(0.057)	(0.519)	(0.093)	(0.149)	(0.571)	(0.519)	(0.519)	(−)

Notes: The first row gives the model selection criteria while the second column indicates the forecast evaluation criteria. MAE, RMSE and LAE denote mean absolute error, root mean square error and the last absolute error change. AIC, BIC, PIC, FIC are the Akaike Information Criterion, the Schwarz Information Criterion, the Posterior Information Criterion and the Fisher Information Criterion respectively. EOPW and RR denote the recursive end-of-period wealth criterion and the recursive return criterion calculated as described in Section 3. RW denotes the benchmark random walk model. Figures in parentheses are p-values of the null hypothesis of equal forecast accuracy between the models selected according to the criterion indicated in each column and RW. p-values are computed by using the Reality Check procedure described in Appendix B.

Table 6. Results with revised data: dollar-yen

Panel A) Empirical search

	BM	RW
MAE	0.023	0.053
	(0.023)	(—)
RMSE	0.039	0.065
	(0.019)	(—)

Panel B) Real-time out-of-sample exercise

	MAE	RMSE	LAE	$\overline{R}^2$	AIC	BIC	SIGN	PIC	FIC	RW
MAE	0.053	0.051	0.056	0.052	0.052	0.056	0.054	0.052	0.052	0.053
	(0.013)	(0.007)	(0.591)	(0.009)	(0.004)	(0.732)	(0.304)	(0.004)	(0.002)	(—)
RMSE	0.069	0.063	0.071	0.062	0.067	0.070	0.065	0.062	0.062	0.065
	(0.152)	(0.021)	(0.672)	(0.064)	(0.112)	(0.884)	(0.356)	(0.015)	(0.016)	(—)

Notes: The results reported in this table replicate the ones in Tables 2 and 4 where historical (revised) data are used. See notes to Tables 2 and 4. *Panel A)* MAE and RMSE denote mean absolute forecast errors and root mean square forecast errors respectively. BM is the best model chosen for the dollar-yen; RW is the random walk model. Figures in parentheses are *p*-values of the null hypothesis of equal forecast accuracy between BM and RW. *Panel B)* The first row gives the model selection criteria while the second column indicates the forecast evaluation criteria. MAE, RMSE and LAE denote mean absolute error, root mean square error and the last absolute error respectively. AIC, BIC, PIC, FIC are the Akaike Information Criterion, the Schwarz Information Criterion, the Posterior Information Criterion and the Fisher Information Criterion respectively. RW denotes the benchmark random walk model. Figures in parentheses are *p*-values of the null hypothesis of equal forecast accuracy between the models selected according to the criterion indicated in each column and RW. For both Panels A-B, *p*-values are computed by using the Reality Check procedure described in Appendix B.

Table 7. Results using a rolling forecasting scheme: dollar-yen

Panel A) Empirical search

	BM	RW
MAE	0.017	0.053
	(0.025)	(−)
RMSE	0.035	0.066
	(0.020)	(−)

Panel B) Real-time out-of-sample exercise

	MAE	RMSE	LAE	$\overline{R}^2$	AIC	BIC	SIGN	PIC	FIC	RW
MAE	0.067	0.051	0.065	0.050	0.052	0.061	0.063	0.052	0.052	0.053
	(0.014)	(0.006)	(0.673)	(0.008)	(0.008)	(0.389)	(0.354)	(0.003)	(0.003)	(−)
RMSE	0.085	0.065	0.082	0.062	0.069	0.077	0.078	0.064	0.064	0.066
	(0.195)	(0.024)	(0.605)	(0.067)	(0.132)	(0.791)	(0.383)	(0.012)	(0.015)	(−)

Notes: The results reported in this table replicate the ones in Tables 2 and 4 where a rolling forecasting scheme is used. The rolling window is set up to 5 years (i.e. 20 quarters). *Panel A)* MAE and RMSE denote mean absolute forecast errors and root mean square forecast errors respectively. BM is the best model chosen for the dollar-yen; RW is the random walk model. Figures in parentheses are *p*-values of the null hypothesis of equal forecast accuracy between BM and RW. *Panel B)* The first row gives the model selection criteria while the second column indicates the forecast evaluation criteria. MAE, RMSE and LAE denote mean absolute error, root mean square error, and the last absolute error respectively. SIGN is a sign criterion based on the directional accuracy of the in-sample predictions of the exchange rate change. AIC, BIC, PIC, FIC are the Akaike Information Criterion, the Schwarz Information Criterion, the Posterior Information Criterion and the Fisher Information Criterion respectively. RW denotes the benchmark random walk model. Figures in parentheses are *p*-values of the null hypothesis of equal forecast accuracy between the models selected according to the criterion indicated in each column and RW. For both Panels A-B, *p*-values are computed by using the Reality Check procedure described in Appendix B.

Table 8. **Results using extra lags in the fundamentals: dollar-yen**

Panel A) Empirical search

	BM	RW
MAE	0.019	0.053
	(0.030)	(−)
RMSE	0.034	0.065
	(0.035)	(−)

Panel B) Real-time out-of-sample exercise

	MAE	RMSE	LAE	$\overline{R}^2$	AIC	BIC	SIGN	PIC	FIC	RW
MAE	0.050	0.048	0.054	0.049	0.051	0.053	0.054	0.048	0.048	0.053
	(0.015)	(0.020)	(0.496)	(0.005)	(0.007)	(0.404)	(0.630)	(0.004)	(0.008)	(−)
RMSE	0.066	0.064	0.069	0.064	0.064	0.070	0.068	0.064	0.064	0.065
	(0.205)	(0.027)	(0.550)	(0.078)	(0.164)	(0.702)	(0.421)	(0.011)	(0.013)	(−)

Notes: The results reported in this table replicate the ones in Tables 2 and 4 where estimations are carried out using an additional second lag in each of the regressors as discussed in Section 7. *Panel A)* MAE and RMSE denote mean absolute forecast errors and root mean square forecast errors respectively. BM is the best model chosen for the dollar-yen; RW is the random walk model. Figures in parentheses are *p*-values of the null hypothesis of equal forecast accuracy between BM and RW. *Panel B)* The first row gives the model selection criteria while the second column indicates the forecast evaluation criteria. MAE, RMSE and LAE denote mean absolute error, root mean square error and the last absolute error respectively. SIGN is a sign criterion based on the directional accuracy of the in-sample predictions of the exchange rate change. AIC, BIC, PIC, FIC are the Akaike Information Criterion, the Schwarz Information Criterion, the Posterior Information Criterion and the Fisher Information Criterion respectively. RW denotes the benchmark random walk model. Figures in parentheses are *p*-values of the null hypothesis of equal forecast accuracy between the models selected according to the criterion indicated in each column and RW. For both Panels A-B, *p*-values are computed by using the Reality Check procedure described in Appendix B.

A Appendix: Real Time Data Set

A.1 Macroeconomic Data

For the purposes of our analysis we assembled and compiled real time data of quarterly frequency. In particular, we constructed four vintages for each year considered with the first real time observation starting at 1987Q3. We performed the first estimation of the set of models examined using data from 1977Q1 to 1987Q3 and hence the first exchange rate forecast produced refers to the change in the exchange rate from 1987Q3 to 1987Q4. The real time data are taken from past issues of the OECD's *Main Economic Indicators* (MEI), which is published in January, April, July and October. Starting from observations for 1987Q3 (the first real time observation), we have a total of 65 data vintages until 2004Q1 for Canada, United States (US), Japan, Germany, Switzerland and the UK. Each series consists of quarterly historical data of the variables considered until 1987Q2, whereas the vintages data from 1987Q3 are taken from the corresponding MEI issue. The historical (revised) data cover the period 1973Q1-1987Q2 and were retrieved from the *Main Economic Indicators Historical Statistics 1969-1988* issue published by the OECD.

For Canada, US, Japan, Germany, Switzerland and UK we use the Gross Domestic Product (GDP), expressed in national currency, seasonally adjusted. The price level is the seasonally adjusted Consumer Price Index (CPI, all items) for each country.

Turning to the monetary base data, we retrieved seasonally adjusted data for M1, wherever possible. Only seasonally unadjusted data were available for Switzerland. For the UK we use M1 plus quasi money up to the 1989Q4 vintage and M2 thereafter because of lack of data for M1 for the full sample.

Seasonally adjusted data for the trade balance expressed in the national currency of the countries examined were also retrieved from the MEI. Data for the Japanese trade balance were expressed in US dollars up to the 1996Q3 vintage, and reported in Japanese yen at the subsequent vintages. Hence, we converted the data into Japanese yen using the dollar-yen spot exchange rate reported in the MEI. Similarly, data for Germany up to the 2002Q1 vintage were expressed in German marks (DM) and in euros for subsequent vintages. Therefore, we used the 1999 conversion rate for the mark-euro (1 euro = 1.95583 DM) in order to convert German data after the 2002Q1 vintage in marks.

A.2 Data on International Capital Movements

With respect to international capital movements, we constructed data vintages for the US transactions with foreigners in long-term foreign and domestic securities using past issues of the *US Treasury Bulletin*, which is published in March, June, September and December each calendar year by the US Department of the Treasury. The data cover transactions carried out in the US for the accounts of foreigners and transactions executed abroad for the accounts of institutions and individuals resident in the US. These transactions involve Long-Term Domestic and Foreign Securities and they are classified by type and country of the foreign buyers and sellers who deal with the US reporting institutions. Each issue reports the total purchases and sales by foreigners from and to US residents, respectively during the previous quarter.

Starting from the December 1987 issue to the December 2003 issue of the *Treasury Bulletin* we constructed four vintages per calendar year. This yields 65 vintages. Each vintage for each country considered consists of quarterly historical data on total purchases and sales. Monthly historical data on capital movements are retrieved from the US Department of the Treasury's website.[18] We converted these monthly historical data into quarterly by summing up the values in total purchases that correspond to the period between the issues of the *Treasury Bulletin*. The same task was performed for the total sales of foreign and domestic securities in the US in order to construct the necessary quarterly data for total sales.

B   Appendix: Reality Check to Evaluate the Best Model

We employ the Reality Check procedure proposed by White (2000), building on previous work by Diebold and Mariano (1995) and West (1996), for testing the null hypothesis that the best model (BM) selected in the specification search using the criterion indicated in equation (6) has no predictive superiority over a given benchmark model. This procedure is specifically designed to allow aggressive model searching to be undertaken with confidence that one will not mistake results that could have been generated by chance for genuinely good results. Put differently, the Reality Check explicitly takes into account any forecasting biases arising from the specific model selection criterion adopted and other possible biases related to data snooping. The test evaluates the distribution of a performance measure accounting for the full set of models that lead to the BM, and is based on the $L \times 1$ performance statistic:

$$\overline{f} = N^{-1} \sum_{t=R}^{T} \widehat{f}_{t+1},$$

where L is the number of "configurations" (or possible models, equal to $2^7 = 128$ each step ahead in this paper); N is the number of prediction periods indexed from R through T so that $T = R + N - 1$; $\widehat{f}_t$ is the observed performance measure for period t (such as the difference in RMSEs between the benchmark and a model), obtained on the basis of a model specification, set of predictor variables and parameter estimates. In our application, the L models generate L one-quarter-ahead forecasts of the exchange rate that are used to measure performance and, since the first set of models are estimated from 1977Q1 to 1987Q3, the first set of L forecasts is for the change in the exchange rate from 1987Q3 to 1987Q4.

We are interested in testing the null hypothesis that the BM selected in a specification search across the L possible models has no predictive superiority over a given benchmark model–in our case the random walk model. This null can be expressed as:

$$H_{0:} \max_{i=1,\dots,L} \left\{ E\left(f^i\right) \right\} \leq 0,$$

where f^i is the performance measure for model i. Rejection of this null hypothesis implies that the best model achieves performance that is superior to the random walk benchmark.[19] White (2000) proposes a 'Reality Check' algorithm which allows us to evaluate this null hypothesis.

---

[18] http://www.ustreas.gov/tic

[19] This null hypothesis is relevant, for example, for the case when the performance measure is the RMSE, where a lower number implies better performance. The inequality changes sign under the null hypothesis when the performance measure used improves for larger positive values, such as, for example, the RR (Sharpe ratio) criterion. In general, the procedure can be tailored to suit the criterion for comparing models adopted by the researcher as well as the criterion for model selection.

Following White (2000), the Reality Check algorithm used in this paper to determine the p-values of the test statistic of equal point forecast accuracy reported in Table 2 consists of the following steps:

1. Given the sequence of observations $\{Y_{t+1}\}$, where $Y_{t+1} = (\Delta s_{t+1}, X_t)'$ and X_t denotes the explanatory variables, estimate each of the possible configurations of the fundamentals model described in Section 3. The total number of configurations at each point in time is 128 $(= 2^7)$.

2. Select the BM according to the criterion indicated in Equation (6) in the text. Compute for the best model the statistics of interest $\overline{f}^{BM} = N^{-1} \sum_{t=R}^{T} \widehat{f}_{t+1}^{BM}$.

3. Postulate a data generating process (DGP), where the exchange rate is assumed to follow a driftless random walk under the null hypothesis H_0 and the innovations are assumed to be iid. The fundamentals variables (in each data vintage) are treated as exogenous in that we utilize the realized data in each iteration, without postulating a DGP for the fundamentals variables X_t for each vintage.

4. Based on the DGP specified in Step 3, generate a sequence of pseudo observations for the nominal exchange rate $\{Y_{t+1}^*\}$, where $Y_{t+1} = (\Delta s_{t+1}^*, X_t)'$, of the same length as the original data series $\{Y_{t+1}\}$. Repeat this step $1,000$ times.

5. For each of the $1,000$ Monte Carlo replications $\{Y_{t+1}^*\}$, estimate $L = 128$ configurations for each forecasting step and select the BM as in Step 2. Use N one-step ahead forecasts to construct the test statistic of interest, $\overline{f}^* = N^{-1} \sum_{t=R}^{T} \widehat{f}_{t+1}^*$ for each of the configurations (including the BM).

6. Construct the following statistics: $\overline{V}_j^* = \max_{i=1,\ldots,L} \left\{ \sqrt{N} \left(\overline{f}^{i,*} \right) \right\}_j$ where $j = 1, \ldots, 1,000$ is the number of replications.

7. Use the empirical distribution of the $1,000$ replications of the test statistic, $\overline{V}_j^*$ calculated under H_0 that the exchange rate follows a driftless random walk to determine the p-value of the test statistic $\overline{f}^{BM}$. In other words, this Reality Check by employing the maximum value over all models $\overline{V}_j^*$, allows the calculated p-value to incorporate the effects of data snooping from the search over the L models and forecasting biases due to the specific model selection criterion used in Step 2.

Table 1 in the text reports the size and power properties of this procedure replacing the observed data for the exchange rate with artificial data simulated under a driftless random walk with the same number of observations as in our empirical analysis, with the variance calibrated on the Japanese yen. Using these artificial data the empirical size is, in each case, close to the three significance levels examined, indicating at most a very tenuous size. This is, in our opinion, a crucial robustness exercise on the validity of the Reality Check procedure implemented in our specific context, which indicates that, if exchange rates are indeed not predictable (i.e. they follow a random walk), then even with our large pool of models and using the selection criterion in equation (6) one would not find statistical evidence against the random walk benchmark when accounting for data snooping and forecasting biases. At the empirical level, in fact, the Reality Check results for the empirical search reported in Table 2 indicate that for three dollar exchange rates the BM outperforms a random walk, whereas for two exchange rates this is not the case.

In the text, the analysis of the power properties is carried out using the same scheme where the DGP is not

given by a driftless random walk but postulated to be equal to different alternatives (i.e. random walk with drift, AR(1) for exchange rate returns, AR(1) for exchange rate levels etc.). The results reported in Table 1 document that the Reality Check is able to detect departures from the no-predictability benchmark when the null hypothesis that exchange rates follow a random walk is violated.

Finally, the p-values for testing the null hypothesis that the best model selected has no predictive superiority over the random walk model reported in Tables 4, 5, 6B and 7B are constructed using the same procedure described under points 1 to 7. However, the model selection criterion reported under Step 2 is replaced by a series of standard criteria (MAE, RMSE, LAE, $\overline{R}^2$, etc.)–see Section 3.2 in the main text for further details.

References

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

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

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

Bacchetta, P. and Van Wincoop, E. (2004), "A Scapegoat Model of Exchange-Rate Fluctuations," *American Economic Review Papers and Proceedings*, 94, 114-118.

Bacchetta, P. and Van Wincoop, E. (2006), "Can Information Heterogeneity Explain the Exchange Rate Determination Puzzle?", *American Economic Review*, 96, 552-576.

Backus, D., Gregory, A. and Telmer, C. (1993), "Accounting for Forward Rates in Markets for Foreign Currency," *Journal of Finance*, 48, 1887-1908.

Baxter, M. and Stockman, A.C. (1989), "Business Cycles and the Exchange Rate Regime: Some International Evidence," *Journal of Monetary Economics,* 23, 377-400.

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

Bossaerts, P. and Hillion, P. (1999), "Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?," *Review of Financial Studies*, 12, 405-428.

Branson, W.H., (1984), "A Model of Exchange Rate Determination With Policy Reaction: Evidence from Monthly Data," in P. Malgrange and P.A. Muet (eds.) *Contemporary Macroeconomic Modelling*, Oxford: Basil Blackwell.

Brock, W.A., Dechert, W.D., Scheinkman, J.A. and LeBaron, B. (1996), "A Test for Independence Based on the Correlation Dimension," *Econometric Reviews*, 15, 197-235.

Brock, W.A.,, Hsieh, D. and LeBaron, B. (1991), *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence.* MIT Press Cambridge, Massachusetts.

Cavallo, M. and Ghironi, F. (2002), "Net Foreign Assets and the Exchange Rate: Redux Revived," *Journal of Monetary Economics*, 49, 1057-1097.

Chen, Y.-C. and Rogoff, K. (2003), "Commodity Currencies," *Journal of International Economics*, 60, 133-160.

Cheung, Y., and Chinn, M. 2001, "Currency Traders and Exchange Rate Dynamics: A Survey of the U.S. Market," *Journal of International Money and Finance*, 20, 439-471.

Cheung, Y.-W., Chinn, M.D. and Garcia Pascual, A. (2003), "Recent Exchange Rate Models: In-Sample Fit and Out-of-Sample Performance," in: De Grauwe, P. (ed.), *Exchange Rate Modeling: Where Do Stand?*, Cambridge and New York: Cambridge University Press.

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

Clarida, R.H., Sarno, L., Taylor, M.P. and Valente, G. (2006), "The Role of Asymmetries and Regime Shifts in the Term Structure of Interest Rates," *Journal of Business*, 79, 1193-1225.

Della Corte, P., Sarno, L. and Tsiakas, I. (2008), "An Economic Evaluation of Empirical Exchange Rate Models," *Review of Financial Studies*, forthcoming.

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

Diebold, F.X. and Nason, J. (1990), "Nonparametric Exchange Rate Prediction?," *Journal of International Economics*, 28, 315-322.

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

Engel, C. (1996), "The Forward Discount Anomaly and the Risk Premium: A Survey of Recent Evidence," *Journal of Empirical Finance*, 3, 123-192.

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

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

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

Evans, G.W. and Honkapohja, S. (2001), *Learning and Expectations in Macroeconomics*, Princeton, NJ: Princeton University Press.

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

Flood, R.P. and Rose, A.K. (1995), "Fixing Exchange Rates A Virtual Quest for Fundamentals," *Journal of Monetary Economics*, 36, 3-37.

Flood, R.P. and Rose, A.K. (1999), "Understanding Exchange Rate Volatility Without the Contrivance of Macroeconomics," *Economic Journal*, 109, 660-672.

Frankel, J.A. (1996), "How Well Do Foreign Exchange Markets Function: Might a Tobin Tax Help?," in ul Haq, M., Kaul, I. and Grunberg, I. (eds.), *The Tobin Tax: Coping with Financial Volatility*, Oxford University Press, New York, 41-81.

Frankel, J. and Froot, K.A. (1987), "Using Survey Data to Test Standard Propositions Regarding Exchange Rate Expectations," *American Economic Review*, 77, 133-153.

Galambos, J. and Simonelli, I. (1996), *Bonferroni - type Inequalities with Applications,* Springer, New York.

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

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

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

Hunter, D. (1976), "An Upper Bound for the Probability of a Union," *Journal of Applied Probability*, 13, 597-603.

Kilian, L. (1999), "Exchange Rates and Monetary Fundamentals: What Do We Learn from Long- Horizon Regressions?," *Journal of Applied Econometrics*, 14, 491-510.

Kounias, S. (1968), "Bounds for the Probability of the Union With Applications," *Annals of Mathematical Statistics,* 39, 2154-2158.

Krueger, A.O. (1983), *Exchange Rate Determination*, Cambridge University Press.

Kumhof, M. and Van Nieuwerburgh, S. (2005), "Monetary Policy in an Equilibrium Portfolio Balance Model," International Monetary Fund and New York University, mimeo.

Levich, R.M. and Thomas, L. (1993), "The Significance of Technical Trading-Rule Profits in the Foreign Exchange Market: A Bootstrap Approach," *Journal of International Money and Finance*, 12, 451-474.

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

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

McCracken, M. and Sapp, S.G. (2005), "Evaluating the Predictability of Exchange Rates Using Long-Horizon Regressions: Mind Your p's and q's!" *Journal of Money, Credit and Banking*, 37, 473-494.

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

Meese, R.A. and Rose, A.K. (1990), "Nonlinear, Nonparametric, Nonessential Exchange Rate Estimation," *American Economic Review*, 80, 192-196.

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

Neely, C.J., Weller, P. and Dittmar, R. (1997), "Is Technical Analysis in the Foreign Exchange Market Profitable? A Genetic Programming Approach," *Journal of Financial and Quantitative Analysis*, 32, 405-426.

Ng, S. and Perron, P. (2001), "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power," *Econometrica*, 69, 1519-1554.

Obstfeld, M. and Rogoff, K. (2000), "The Six Major Puzzles in International Macroeconomics: Is there a Common Cause?," in: Bernanke, B. and Rogoff, K. (eds.), *NBER Macroeconomics Annual 2000*, Cambridge, Mass: MIT Press, 339-390.

Obstfeld, M. and Rogoff, K. (2004), "The Unsustainable US Current Account Position Revisited," *National Bureau of Economic Research Working Paper* No. 10869.

Pesaran, H.M. and Timmermann, A. (1995), "Predictability of Stock returns: Robustness and Economic Significance," *Journal of Finance*, 50, 1201-1228.

Pesaran, H.M. and Timmermann, A. (2000), "A Recursive Modelling Approach to Predicting UK Stock Returns," *Economic Journal*, 110, 159-191.

Phillips, P.C.B. and Ploberger, W. (1996), "Posterior Odds Testing for a Unit Root With Data Based Model Selection," *Econometrica*, 64, 381-412.

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

Rose, A.K. and Yellen, J.L. (1989), "Is there a J-curve?," *Journal of Monetary Economics*, 24, 53-68.

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

Rossi, B. (2006), "Are Exchange Rates Really Random Walks? Some Evidence Robust to Parameter Instability," *Macroeconomic Dynamics*, 10, 20-38.

Sarno, L. (2005), "Towards a Solution to the Puzzles in Exchange Rate Economics: Where Do We Stand?" *Canadian Journal of Economics*, 38, 673-708.

Schinasi, G.J. and Swamy, P.A.V.B. (1989), "The Out-of-Sample Forecasting Performance of Exchange Rate Models When Coefficients Are Allowed to Change," *Journal of International Money and Finance*, 8, 375-390.

Sullivan, R., Timmerman, A. and White, H. (1999), "Data-Snooping, Technical Trading Rule Performance, and the Bootstrap," *Journal of Finance,* 54, 1647-1692.

Sweeney, R.J. (2006), "Mean Reversion in G-10 Nominal Exchange Rates," *Journal of Financial and Quantitative Analysis*, 41, 685-708.

Wei, C. (1992), "On Predictive Least Squares Principles," *Annals of Statistics*, 20, 1-42.

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

West, K.D. (2006), "Forecast Evaluation," in Elliott, G., Granger, C.W.J. and Timmermann, A. (eds.), *Handbook of Economic Forecasting*, Vol. I, Amsterdam: Elsevier.

West, K.D. and McCracken, M.W. (1998), "Regression-Based Tests of Predictive Ability," *International Economic Review*, 39, 817-40.

White, H. (2000), "A Reality Check For Data Snooping," *Econometrica*, 68, 1097-1127.

Wright, J.H. (2003), "Bayesian Model Averaging and Exchange Rate Forecasts," Federal Reserve Board of Governors, Working Paper 2003-779.

Figure 1. Best Fundamentals Models

Japan

UK

Canada

Switzerland

Germany
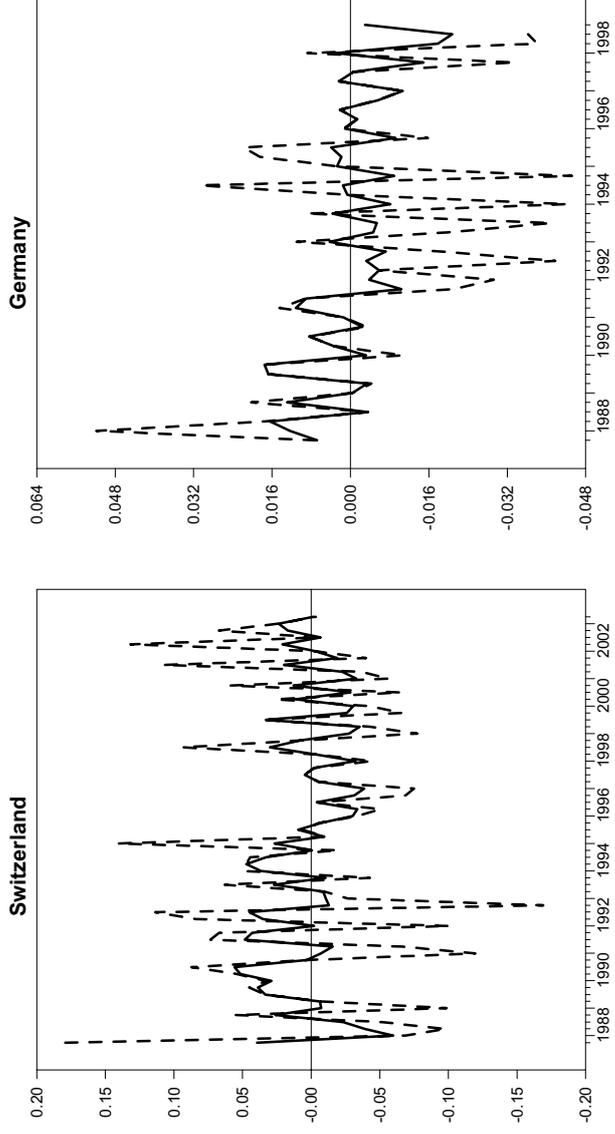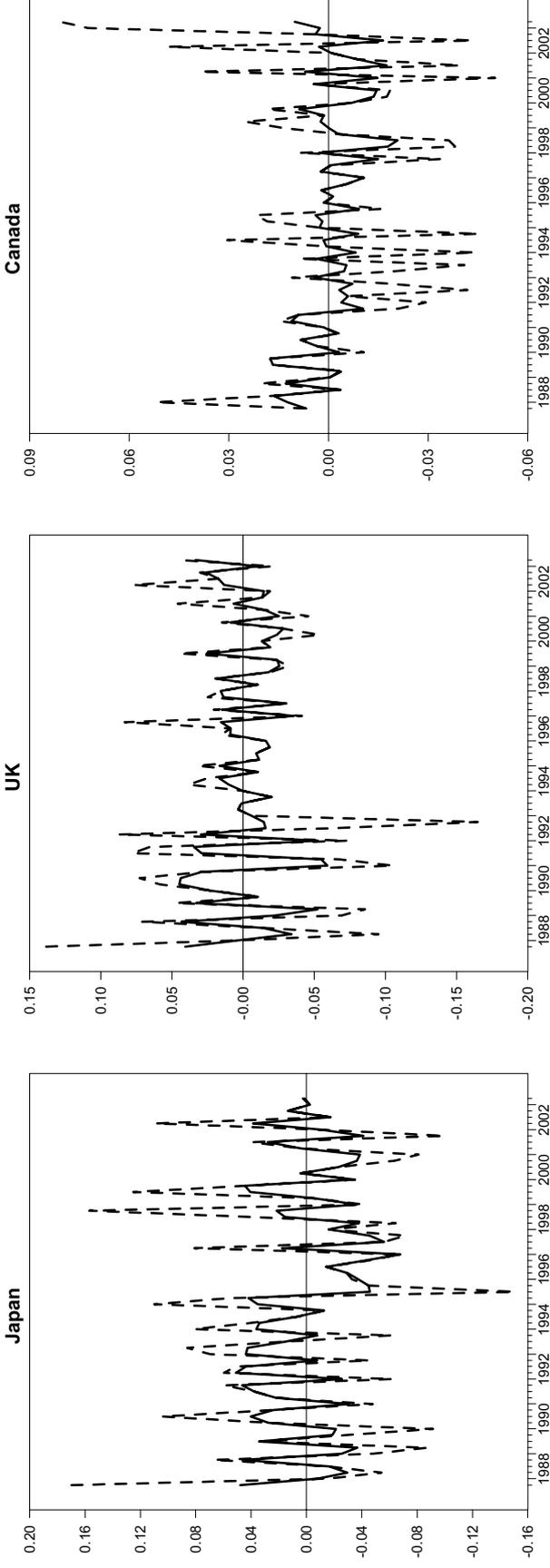
Figure 2. Empirical Search

Inclusion Frequencies for US dollar/Japanese yen
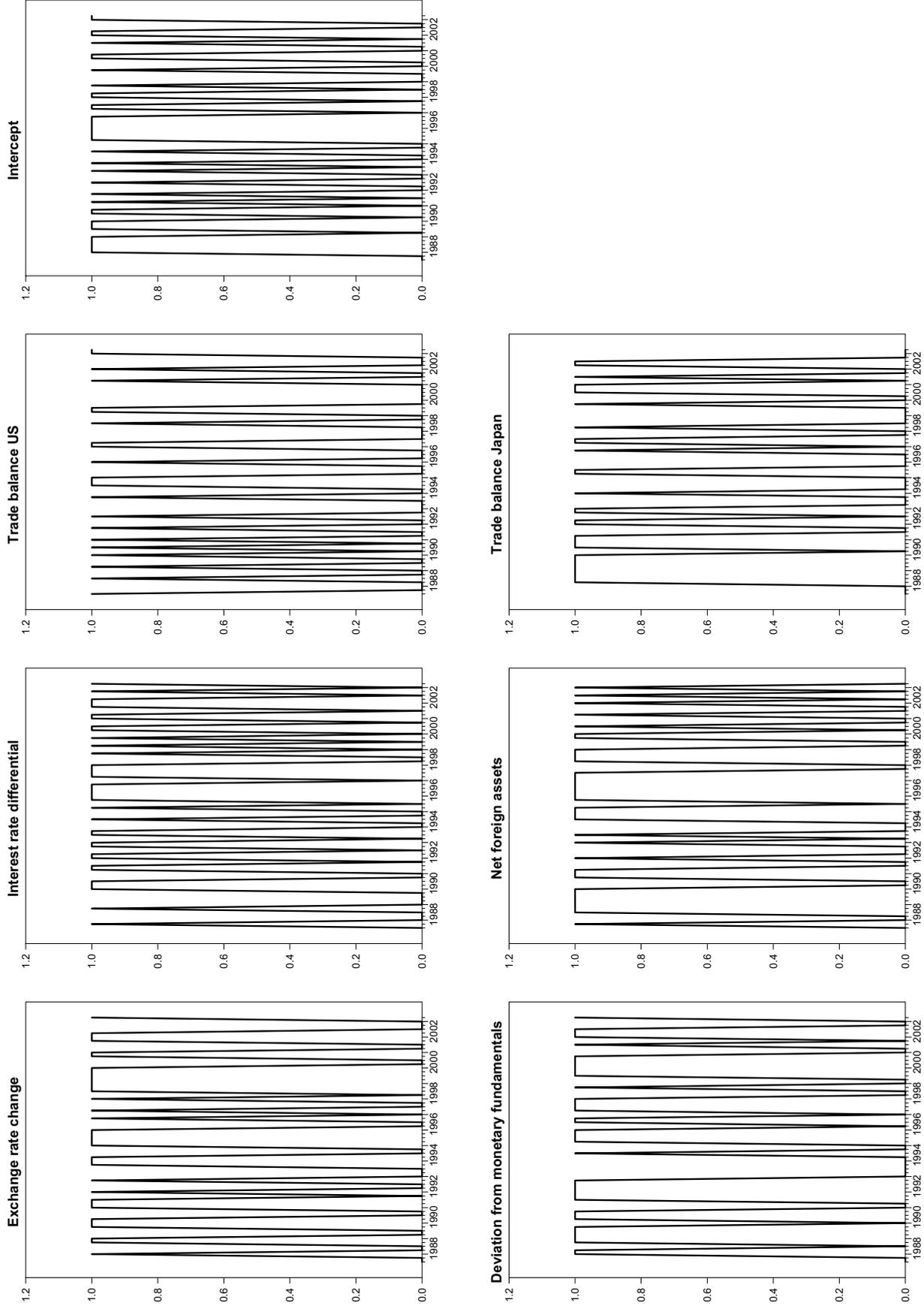
Figure 3. Real-time Search

Inclusion Frequencies for US dollar/Japanese yen

Exchange rate change

Interest rate differential

Trade balance US

Intercept

Deviation from monetary fundamentals

Net foreign assets

Trade balance Japan