Ribeiro, Pinho J. (2016) *Forecasting exchange rates in the presence of instabilities.* PhD thesis.

# Forecasting Exchange Rates in the Presence of Instabilities

Pinho J. Ribeiro

A Dissertation

Submitted in fulfilment of the
requirements for the degree of
Doctor of Philosophy

Adam Smith Business School
College of Social Sciences
University of Glasgow

August 4, 2016

## Abstract

Many exchange rate papers articulate the view that instabilities constitute a major impediment to exchange rate predictability. In this thesis we implement Bayesian and other techniques to account for such instabilities, and examine some of the main obstacles to exchange rate models' predictive ability. We first consider in Chapter 2 a time-varying parameter model in which fluctuations in exchange rates are related to short-term nominal interest rates ensuing from monetary policy rules, such as Taylor rules. Unlike the existing exchange rate studies, the parameters of our Taylor rules are allowed to change over time, in light of the widespread evidence of shifts in fundamentals - for example in the aftermath of the Global Financial Crisis. Focusing on quarterly data frequency from the crisis, we detect forecast improvements upon a random walk (RW) benchmark for at least half, and for as many as seven out of 10, of the currencies considered. Results are stronger when we allow the time-varying parameters of the Taylor rules to differ between countries.

In Chapter 3 we look closely at the role of time-variation in parameters and other sources of uncertainty in hindering exchange rate models' predictive power. We apply a Bayesian setup that incorporates the notion that the relevant set of exchange rate determinants and their corresponding coefficients, change over time. Using statistical and economic measures of performance, we first find that predictive models which allow for sudden, rather than smooth, changes in the coefficients yield significant forecast improvements and economic gains at horizons beyond 1-month. At shorter horizons, however, our methods fail to forecast better than the RW. And we identify uncertainty in coefficients' estimation and uncertainty about the precise degree of coefficients variability to incorporate in the models, as the main factors obstructing predictive ability.

Chapter 4 focus on the problem of the time-varying predictive ability of economic fundamentals for exchange rates. It uses bootstrap-based methods to uncover the time-specific conditioning information for predicting fluctuations in exchange rates. Employing several metrics for statistical and economic evaluation of forecasting performance, we find that our approach based on pre-selecting and validating fundamentals across bootstrap replications generates more accurate forecasts than the RW. The approach, known as bumping, robustly reveals parsimonious models with out-of-sample predictive power at 1-month horizon; and outperforms alternative methods, including Bayesian, bagging, and standard forecast combinations.

Chapter 5 exploits the predictive content of daily commodity prices for monthly commodity-currency exchange rates. It builds on the idea that the effect of daily commodity price fluctuations on commodity currencies is short-lived, and therefore

harder to pin down at low frequencies. Using MIxed DAta Sampling (MIDAS) models, and Bayesian estimation methods to account for time-variation in predictive ability, the chapter demonstrates the usefulness of suitably exploiting such short-lived effects in improving exchange rate forecasts. It further shows that the usual low-frequency predictors, such as money supplies and interest rates differentials, typically receive little support from the data at monthly frequency, whereas MIDAS models featuring daily commodity prices are highly likely. The chapter also introduces the random walk Metropolis-Hastings technique as a new tool to estimate MIDAS regressions.

# Contents

# List of Tables

# List of Figures

# Acknowledgements

Before seeing the evidence, I had no idea on how little I knew about Bayesian methods and issues concerning predicting exchange rates. I then updated my priors during my studies, and now, my posterior reassures me that I know nothing. I remain satisfied though, as I reached this conclusion after seeing the evidence, and more importantly, for being able to make just a tiny dent in this vast literature.

I wish to express my gratitude to my supervisors, Prof. Joseph P. Byrne and Prof. Dimitris Korobilis, for their guidance and support throughout these discoveries. Byrne and Korobilis stimulated my research ideas, reiterated on conciseness, emphasized accuracy, encouraged me to work and interact with other researchers, solidified independence in doing research, and have given me moral support. If I were to select supervisors in a Bayesian way, now I would benefit from this informative prior.

I had the advantage of working in one of the thesis's project (Chapter 5) at Norges Bank Research Department. I gratefully acknowledge the Bank's hospitality and thank the researchers at Norges Bank for such stimulating and excellent working environment. Dr. Claudia Foroni and Prof. Francesco Ravazzolo deserve a special reference, for their guidance and coauthorship. Of course, the usual disclaimer applies: the views expressed in this thesis are my own and should not be interpreted as reflecting the views of the Norges Bank. I am equally thankful to Prof. Raffaella Giacomini and Prof. James H. Stock for their captivating comments on the project's conceptual framework.

I owe gratitude to Prof. Craig Burnside for his advice, and to the examiners of my Ph.D thesis, Prof. Ronald MacDonald and Prof. Pasquale Della Corte, for their excellent suggestions. I would like to thank the professors and the foreign exchange dealers who gave me constructive suggestions on earlier versions of some of the thesis' chapters, including: Roberto Casarin, Herman K. van Dijk, Charles Engel, John Hydeskov, Gary Koop, Michael Melvin, Dagfinn Rime, Barbara Rossi, Christian Schumacher, Allan Timmermann, Shaun Vahey, and Kenneth D. West. I also benefited from valuable comments from participants at several conferences, including the International Association for Applied Econometrics Conference (IAAE, 2014), European Conference of the Econom[etr]ics Community (EC2, 2014), Workshop on Financial Determinants of Exchange Rates (2014), Workshop on Econometric Modeling and Forecasting in Central Banks (SIRE, 2014), Symposium of the Society for Nonlinear Dynamics and Econometrics (SNDE, 2015), Norges Bank Research Spring Meeting (2015), and Norges Bank Seminar (2015).

I will remain immensely grateful to the Banco de Moçambique for granting me the opportunity to broaden my knowledge. Without the Bank's support, my interest of pursing a Ph.D degree would have remained a mirage.

At the Adam Smith Building I enjoyed the company and friendship of Shuo Cao and Reinhold Kamati. I would also like to extend my sincere thanks to our postgraduate programme officers, in particular Christine Athorne, Hannah Engleman, and Jane Brittin.

Last but not least, I thank my parents, Jose Ribeiro and Julia Daniel, for their unconditional support to my education. I further thank my Nelinha and our children, Altenir and Eloia, for coping with my Ph.D-related mood swings, weird dreams about mathematical derivations, Ph.D bunker-mode, and so forth. The best part is that sometimes when research ideas are blinking, there is no better fix than listening to an audio-recorded bed-time story read by your kids to you.

*Glasgow, August 4, 2016*

To my parents, Jose Ribeiro and Julia Daniel

# Author's Declaration

I declare that, except where explicit reference is made to the contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Chapter 2 and 3 are drawn from collaborative work with my supervisors, Joseph P. Byrne and Dimitris Korobilis. Chapter 5 follows from a research project conducted as a PhD Intern at Norges Bank in collaboration with Claudia Foroni and Francesco Ravazzolo. In all cases, the core tasks were undertaken by myself. The joint works are listed below.

1. Byrne, J.P., Korobilis, D. and Ribeiro, P.J. (2016). "Exchange Rate Predictability in a Changing World". *Journal of International Money and Finance* 62, $1-24$. doi: 10.1016/j.jimonfin.2015.12.001

2. Byrne, J.P., Korobilis, D. and Ribeiro, P.J. (2014). "On the Sources of Uncertainty in Exchange Rate Predictability". Working Paper No. 2014-16, Adam Smith Business School.

3. Foroni, C., Ravazzolo, F. and Ribeiro, P.J. (2015). "Forecasting Commodity Currencies: The Role of Fundamentals with Short-Lived Predictive Content". Norges Bank Working Paper 2015/14.

Signature: ..................................................................

Printed name: Pinho Jose Ribeiro

# Chapter 1

# Introduction

## 1.1 Background and Motivation

A country's exchange rate is one of the most closely monitored indicators, as fluctuations in exchange rates can have far-reaching economic consequences. At a micro level, for instance, exchange rate movements constitute a major source of risk exposure for an investor's portfolio return or her assets and liabilities denominated in foreign currency. At a macro level, exchange rates affect a country's level of trade, competitiveness, and external debt burden. It is not surprising, therefore, that academics, policy makers, and market practitioners have both sought to predict exchange rate fluctuations. In this regard, an enduring view initiated by Meese and Rogoff (1983), proposed that forecasts based upon sound exchange rate models could not improve upon a simple no-change forecast.[1]

When rationalizing their findings, Meese and Rogoff (1983) conjecture that sampling error, model misspecification, and parameter instability could explain the poor forecasting performance. In samples of finite length, parameters may be estimated with error, resulting in inaccurate predictions even in the cases where predictors do exhibit predictive content. Similarly, model misspecification (e.g., unexplained nonlinearities), as well as possible biases arising from including (omitting) irrelevant (relevant) regressors may lead to poor forecasting power. The inclusion of irrelevant regressors may improve the in-sample fit of the model but penalizes the model when forecasting out-of-sample, a phenomenon also known as over-fitting (Rossi and Sekhposyan, 2011). Finally, parameter uncertainty, interpreted as shifts in model coefficients due to, for example, macroeconomic shocks can equally hinder forecasting performance.

The Meese-Rogoff's results, often referred to as the Meese-Rogoff puzzle, triggered a large literature examining the weak predictive power of the empirical exchange rates models. Thirty years on, and despite the availability of larger samples and sophisticated forecasting techniques, the issues they have put forward prevail and continue to catalyze research. For instance, in a recent survey of the literature, Rossi (2013) reiter-

[1] The no-change forecast is often called the random walk forecast, defined as one in which the level of the nominal exchange rate in the next period is predicted to stay at the current level.

ates that the predictive ability of macroeconomic models remains ephemeral; see also Cheung et al. (2005) and Rogoff and Stavrakeva (2008), for similar assertions. In other words, although some macro variables or exchange rate fundamentals exhibit predictive content in some periods, in others they fail. Ultimately, the predictive ability of exchange rate models remains plagued with instabilities (Bacchetta and van Wincoop, 2013 and Rossi, 2013).

## 1.2 Thesis Contribution and Outline

In this thesis we implement Bayesian and other techniques to account for several sources of instability that might affect the out-of-sample forecasting performance of exchange rate models. We also examine some of the major obstacles to exchange rate models' predictive ability. The contribution is laid out in four self-contained chapters.

Chapter 2 advances a framework where the parameters that drive fundamentals, and the interaction of these fundamentals with exchange rates, change over time. Precisely, we estimate time-varying parameters Taylor rules and examine their predictive content for exchange rates in a setting that further allows for the coefficients of the forecasting regression to change over time. In this sense, our approach accounts for instabilities by incorporating the notion of fast-changing economic conditions and policy actions. Our use of Bayesian methods in this context also sets us apart from the earlier maximum likelihood-based analysis. In a Bayesian framework, we can potentially estimate the parameters more efficiently, as we can treat them as jointly random variables. In our results, we find that allowing for such dynamics improves out-of-sample forecast relative to the driftless Random Walk (RW), principally in the forecast sample that begins just before the 2008 financial crisis and for horizons beyond four quarters.

In Chapter 3 we employ a systematic approach to examine the role of time-variation in the regressions coefficients, and other model-specific characteristics, in explaining the extant poor forecasting performance of exchange rate models. Our Bayesian approach allows for changing sets of explanatory variables at each point in time, but also for varying degrees of coefficients adaptivity over time. Within it, we investigate for example how relevant is the issue of allowing for time-variation in coefficients relative to keeping them constant. Using information in the likelihood, we also check whether the relevant set of macro variables for predicting exchange rates varies over time. Further, we scrutinize the major causes of impediment to forecasting ability, using a variance decomposition procedure that tracks the uncertainty associated with the model characteristics.

Using statistical and economic measures of forecasting performance we first find that, at longer horizons, our flexible approach provides substantial improvements in forecasting performance, leading to economic gains in a stylized dynamic asset allocation strategy. However, at 1-month forecasting horizon our flexible models fail to improve upon a typical no-change forecast – a frequent finding in the exchange rate literature. When examining the origins of the failure, we identify issues such as uncertainty in coefficients' estimation and uncertainty about the precise degree of coefficients variability to incorporate in the models as the main obstacles to forecasting ability. For the first time, we establish in a data-based manner that time-variation in the parameters of the forecasting regression constitute a major obstacle to exchange rate predictability.

Chapter 4 focus on the problem of the time-varying predictive ability of economic fundamentals for exchange rate (see, e.g., Cheung et al., 2005 and Rossi, 2013). It proposes bootstrap-based methods to reveal or identify fundamentals that carry predictive information at each point in time. One variant of the bootstrap method, known as bagging, generates forecasts from models based on the identified fundamentals and averages the forecasts across sample replications. An alternative method, called bumping, uses the best fundamental-based exchange rate model revealed and trained across sample replications to forecast. While bagging has been applied to forecasting other macro and financial variables, we are the first to introduce the two methods in exchange rate economics.

Using a variety of statistical and economic metrics of predictability, we find that bumping uncovers sets of fundamentals with strong and significant predictive content for 1-month ahead change in exchange rates throughout the forecast sample. None of the competing methods we consider outperforms bumping, including the random walk, bagging, standard Bayesian methods, classic forecast combinations, simple linear regressions based on the typical or more recent proposed fundamentals (e.g., relative yield curves), and the kitchen-sink regression. Inspection of why bumping produces better outcomes than bagging shows that bumping reveals parsimonious models with out-of-sample (OOS) forecasting ability, whereas bagging tends to over-fit, thereby uncovering exchange rate fundamentals with good in-sample fit but poor OOS forecasting power.

Chapter 5 considers a mixed-frequency approach to modeling dynamics in exchange rates while accounting for time-variation in predictive ability. It uses MIxed DAta Sampling (MIDAS) models in a Bayesian setting to exploit the predictive content of daily commodity price changes for monthly commodity-currency exchange rates. The chapter builds upon the recent evidence highlighting the short-lived, yet robust contemporaneous effect of daily commodity price fluctuations on commodity currencies; see Ferraro et al. (2015). The MIDAS approach allows each daily observation on price fluctuations to have a different weight or impact on the end-of-month observation on the exchange rate change, thereby taking advantage of the predictors' sampling properties. The chapter also introduces the random walk Metropolis-Hastings technique as a novel tool to estimate the class of MIDAS regressions we consider.

In our results, we first document the typical finding that commodity prices sampled at low (monthly)-frequency are devoid of forecasting power for exchange rates (see, e.g., Chen et al., 2010 and Ferraro et al., 2015). In contrast, exploiting the short-lived effect of commodity price changes on exchange rates leads to forecast improvements for most of the commodity currencies in our sample. Moreover, the low-frequency commodity prices and the commonly used macro variables, such as money supplies and interest rates differentials, typically receive little support from the data at 1-month forecasting horizon. On the contrary, MIDAS models with daily commodity prices are highly likely at this horizon.

All in all, we find evidence of the usefulness of accounting for instabilities in improving exchange rate predictability. Our data-based procedure to track the sources of uncertainty in exchange rate models offers novel insights regarding the causes of their weak predictive power.

# Chapter 2

# A Time-Varying Parameter Model for Exchange Rates and Fundamentals

## 2.1 Introduction

Predicting movements in exchange rates have long been a subject of interest to market practitioners, academics, and policy makers. A long-standing result, first documented by Meese and Rogoff (1983), is that predictions based upon fully-fledged macroeconomic models are no better than a no-change forecast. Rossi (2013) provides a survey of the subsequent literature that examined the predictive content of macroeconomic models, using theoretical and empirical innovations. Theoretical improvements have included studying the behavior of exchange rates in present-value models (Engel and West, 2005). Separately, empirical advances have included nonlinear methods, such as the exponential smooth transition auto-regressive model of Kilian and Taylor (2003) and time-varying parameter models (e.g., Rossi, 2006; Wolff, 1987).[1] This chapter seeks to combine these theoretical and empirical innovations in predicting exchange rates, in a dynamic world.

Engel and West (2005) and Engel et al. (2008) illustrate that models that can be cast in the standard present-value asset pricing framework imply that exchange rates are approximately random walks. This result holds under the assumptions of non-stationary fundamentals and a near unity discount factor. However, Engel and West (2004) present evidence that even when the discount factor is near one, a class of models based on observable fundamentals can still account for a fairly large fraction of the variance in exchange rates. An example in this class includes structural exchange rate models in which monetary policy follows the Taylor (1993) rule. Engel et al. (2008), Molodtsova and Papell (2009), and Rossi (2013) find that empirical exchange

[1] Other empirical approaches have included: long-horizon methods, see Mark (1995); panel models, see for example Papell (1997), Groen (2000), MacDonald and Nagayasu (2000), Mark and Sul (2001) and Engel et al. (2008); and factor exchange rate models, see Engel et al. (2015).

rate models conditioned on an information set from Taylor rules outperform the random walk benchmark in out-of-sample forecasting, particularly at short-horizons.

Despite the optimism instilled by this emerging research, one area remains unresolved. Exchange rate forecasting models are subject to parameter instability. Rossi and Sekhposyan (2011), for example, detect significant instabilities in models that employ classic and Taylor rule fundamentals. In their study, Meese and Rogoff (1983) had already conjectured that parameter instability may rationalize the poor forecasting of exchange rate models. To address the issue, several researchers have attempted to account for time-variation in parameters when forecasting exchange rates. Nonetheless, as Rossi (2013) and Rogoff and Stavrakeva (2008) point out, the problem has not yet been fully resolved. In fact, Rossi (2013) questions whether instabilities can be exploited to improve exchange rate forecasts.

In this chapter we revisit the issue of forecasting exchange rates with time-varying parameter models. In a major break with the earlier literature, our starting point is that macroeconomic conditions and policy actions evolve, often suddenly.[2] Thus, our modeling strategy allows for fast changing dynamics in the process that determine macroeconomic fundamentals, which in turn influence the path of the exchange rate. Only after these dynamics have been accounted for, we then proceed and allow for time-variation in parameters when predicting exchange rates. To help achieve efficiency in estimating the parameters, we use information in the likelihood based upon Bayesian methods. As Kim and Nelson (1999) refer, Bayesian methods treat all the unknown parameters in the system as jointly distributed random variables, such that each parameter estimate reflects uncertainty about the other parameters. In contrast, estimates based on classical maximum likelihood are prone to errors, since a large number of likelihood functions have to be evaluated. Therefore, unlike the previous literature, we do not rely on classical maximum likelihood methods (as in Rossi, 2006) or calibration (e.g. Wolff, 1987; Bacchetta et al., 2010), which can also be subjective and may give less accurate parameter estimates and inferior forecasting performance.[3]

It is straightforward to recognize the relevance of allowing for time-evolving macroeconomic fundamentals. If the process underlying macroeconomic fundamentals changes rapidly over time, their predictive content may depend upon statistically modeling it; and empirically, there is widespread evidence pointing out to time-evolving dynamics in fundamentals. In the context of fundamentals determined by Taylor rules, Boivin (2006), Kim and Nelson (2006), and Cogley et al. (2010) find that the U.S. Federal Reserve conduct of monetary policy is better characterized by a changing-coefficients Taylor rule. Trecroci and Vassalli (2010) present similar findings for the U.S., U.K., Germany, France, and Italy.

[2]See for example, Stock and Watson (1996) for evidence on structural instabilities in macroeconomic time series in general.

[3]Giannone (2010) provides a helpful critique of the results based on Bacchetta's et al.(2010) calibration, and shows how using the full maximum likelihood setup in a Bayesian framework is important in accounting for instabilities. Balke et al. (2013) also use Bayesian methods and focus upon modeling exchange rates in-sample with monetary fundamentals.

There is also a large literature documenting time-evolving relationships between fundamentals and exchange rates. Bacchetta and van Wincoop (2004), for example, explain this relationship on the basis of a scapegoat theory. Traders in foreign exchange markets seek explanations for fluctuations in the exchange rate, such that even when an unobservable variable is the cause of the fluctuation, they explain it on the basis of something they can observe, the macro variable. The macro variable is therefore a scapegoat, which in turn influences trading behavior and the exchange rate. Over time, fluctuations in exchange rates are then explained by time-varying weights attributed to scapegoat variables. In a recent application, Balke et al. (2013) and Park and Park (2013) show that allowing for such dynamics in the monetary model, improves in-sample fit and out-of sample predictive power for exchange rates.

Putting together these observations, we advance a framework where fundamentals themselves and their interaction with exchange rates change over time. In particular, we estimate time-varying parameter Taylor rules and examine their predictive content in a setting that allows for the parameters of the forecasting regression to change over time.[4] If we further consider the recent events in the world economy, our approach is also timely and topical. For example, Mumtaz and Sunder-Plassmann (2013) study the dynamics of exchange rates before and after the 2008 turmoil, and observe a markedly high volatility in recent years. Similarly, Taylor (2009) argues that prior to the Global Financial Crisis the U.S. Federal Reserve conduct of monetary policy was characterized by a non-linear Taylor rule. After the Crisis, central banks around the world have adopted unconventional monetary policy when confronted with the zero lower bound constraint on nominal interest rates. Furthermore, there has been considerable heterogeneity in country-specific fundamentals, which required bespoke policy measures (Draghi, 2014). All these developments suggest that the constant-parameter forecasting approach used in studies focusing in the samples before the recent turmoil may be ill-suited to capture the dynamics in the recent turbulent times. Our study, therefore, extends the results in the previous papers, including Engel et al. (2008), Engel and West (2005, 2006), Molodtsova and Papell (2009), and Rogoff and Stavrakeva (2008).[5]

In particular, this chapter's dataset consists of quarterly exchange rates from 1973Q1 to 2013Q1, on up to 17 OECD countries relative to the U.S. dollar. We calculate Theil's U-statistic from Root Mean Squared Forecast Error (RMSFE) recursively out-of-sample, while focusing in three forecast samples and four quarterly forecasting horizons ($h = 1, 4, 8, 12$). We assess the significance of the differences in the forecasts using the Diebold and Mariano (1995) and West (1996) tests, with bootstrapped critical

[4]Although in principle forecasting in a rolling regression scheme allows for parameters to change over time, a TVP model allows for instabilities to be updated systematically and more flexibly.

[5]Focusing in a sample before the 2008 financial crisis, Mark (2009) uses a non-linear modeling strategy. He employs a Vector Autoregressive model and least-squares learning techniques to update Taylor rules estimates, inflation and output gap which are then used to compute the exchange rate value. Using in-sample evidence, he finds that allowing for time-variation in parameters is relevant to account for the volatility of the Deutschemark and the Euro, relative to the U.S dollar. Our approach differs from Mark (2009) in that we focus upon out-of-sample predictability of non-linear Taylor rules.

values.

To preview our results, allowing for time-varying Taylor rules improves upon the driftless random walk at horizons beyond 1-quarter. In fact, for periods during and after the Global Financial Crisis, our approach yields a lower RMSFE than the benchmark for at least half of the currencies, and for as many as seven out of 10 currencies. When we examine the performance of a Fixed-Effect panel model, as a variant of models with constant-parameters, we find improvement over the RW mostly in the early, but marginally less in the late parts of our dataset.

The rest of the chapter proceeds as follows. The next section sets out the time-varying parameter regression we consider. Section 2.3 discusses the choice of fundamentals, and Section 2.4 covers data description and the mechanics of our forecasting exercise. The main empirical results are reported in Section 2.5, and Section 2.6 summarizes the findings and deals with robustness checks. We conclude in Section 2.7.

## 2.2 The Time-Varying Parameter Regression

A common practice in forecasting exchange rates is to model the change in the exchange rate as a function of its deviation from its fundamental implied value. As put forward by Mark (1995), this accords with the notion that exchange rates frequently deviate from the level implied by fundamentals, particularly in the short-run. More precisely, define $s_{t+h} - s_t \equiv \Delta s_{t+h}$ as the $h$-step-ahead change in the log of exchange rate, and $\Omega_t$ as a set of exchange rate fundamentals. Then,

$$\Delta s_{t+h} = \beta_{0t} + \beta_{1t} z_t + \varepsilon_{t+h}, \qquad \varepsilon_{t+h} \sim N(0, R); \tag{2.1}$$

$$z_t = \Omega_t - s_t. \tag{2.2}$$

As Eq. (2.2) suggests, $\Omega_t$ signals the exchange rate's fundamental value, hence $z_t$ is the deviation from the fundamental's implied level. When the spot exchange rate is lower than the level implied by the fundamentals, i.e., $s_t < \Omega_t$, then the spot rate is expected to increase.

In Eq. (2.1), the time-subscripts $t$ attached to the coefficients $\beta_t = [\beta_{0t}, \beta_{1t}]$, characterize them as changing over time. The exact coefficient's law of motion is inspired, among others, by Stock and Watson (1996), Rossi (2006), and Boivin (2006). We assume a Random Walk Time-Varying Parameter (RW-TVP) process:

$$\beta_t = \beta_{t-1} + v_t, \tag{2.3}$$

where the error term, $v_t$, is assumed homoskedastic, uncorrelated with $\varepsilon_{t+h}$ in Eq. (2.1) and with a diagonal covariance matrix $Q$. Putting together equations (2.1) and (2.3) results in a state-space model, where (2.1) is the measurement equation and (2.3) the transition equation.

We use Bayesian methods to estimate the parameters of the state-space model. Using the Kalman filter with maximum likelihood is another potential method, but the evaluation of a large number of likelihood functions may undermine the estimates (Kim and Nelson, 1999). With the method of maximum likelihood there is potential for accumulation of errors, as estimation of the state variables is conditional upon maximum likelihood estimates of the other parameters of the system. There is also the issue of identifying objective priors to initialize the Kalman filter. The solution to this latter issue involves setting diffuse priors or using a training sample, but solving the problem of obtaining efficient parameter estimates is more challenging. Bayesian methods, in contrast, treat all the unknown parameters in the system as jointly distributed random variables, such that the estimate of each of them reflects uncertainty about the others (Kim and Nelson, 1999).

In particular, we employ the Carter and Kohn (1994) algorithm and the Gibbs sampler to simulate draws from the parameters' posterior distribution. The Gibbs sampler, which falls in the class of Markov Chain Monte Carlo (MCMC) methods, is a numerical method that uses draws from conditional distributions to approximate joint and marginal distributions. To implement the method we need to (i) elicit priors for the unknown parameters, (ii) specify the form of their posterior conditional distributions, and finally (iii) draw samples from these posterior distributions. To parameterize the prior distributions we use pre-sample information. We do so largely because we are comparing the forecasting performance of several models, at a number of forecast samples and horizons. By setting priors based on a training sample we ensure that all the models are based on the same prior elicitation setting, and hence their performance is not influenced by the model's particular prior parameterization choice. This approach also provides natural shrinkage based on evidence in the likelihood, which in turn ensures that TVP estimates will be more accurate, with smaller variance, resulting in a sharper inference and potentially more precise forecasts. The remainder of the details about priors' elicitation and the steps of the algorithm are provided in Appendix C at the end of this thesis.

## 2.3 Taylor Rule Fundamentals

Having defined the form and the method to estimate the parameters of our main forecasting regression, an additional modeling issue relates to the exact specification of the fundamental information contained in $\Omega_t$. In this regard, our approach is consistent with models that relate the exchange rate to macroeconomic variables within the asset pricing framework (Engel and West, 2005). In this framework, the exchange rate is expressed as the present-value of a linear combination of economic fundamentals and unexpected shocks. Assuming rational expectations and a random walk process for the fundamentals, the framework implies that the spot exchange rate is determined by current observable fundamentals and unobservable noise - see Appendix A for derivations. We focus primarily on observable fundamentals derived from the Taylor (1993) rule.

The Taylor (1993) rule postulates that the monetary authority should set the policy interest rate considering inflation and its deviation from some target, output deviation from potential, and the equilibrium real interest rate. Then, it follows that the authority increases the policy rate when inflation is above the target and/or output is above its potential level.

An emerging research considers the implications of this policy setting for exchange rates, including Engel and West (2005), Engel et al. (2008), Mark (2009), and Molodtsova and Papell (2009, 2013). The premise is that the home and the foreign central banks conduct monetary policy following a Taylor rule. In line with this framework, the foreign monetary authority (the U.S. in our empirical section) is concerned with inflation and output deviations from their target values. In addition, Engel and West (2005) assume that the home country targets the real exchange rate. Following Clarida et al. (1998), it is also common to consider that central banks adjust the actual interest rate to eliminate a fraction of the gap between the current interest rate target and its recent past level, known as interest rate smoothing. By subtracting the foreign Taylor rule from the home, the following interest rate differential equation is obtained:

$$i_t - i_t^* = \phi_0 + \phi_1 \pi_t - \phi_1^* \pi_t^* + \phi_2 \overline{y}_t - \phi_2^* \overline{y}_t^* + \phi_3 q_t + \phi_4 i_{t-1} - \phi_4^* i_{t-1}^* + \mu_t, \tag{2.4}$$

where $i_t$ is the short-term nominal interest rate set by the central bank, asterisks indicate foreign (U.S.) variables; $\pi_t$, is inflation; $\overline{y}_t$, denotes the output gap; $q_t$ is the real exchange rate defined as $q_t = s_t + p_t^* - p_t$; $p_t$, is the log of the price level; $\phi_l$ for $l = 1, ..., 4$, are regression coefficients, and $\mu_t$ is a Gaussian error term. For a detailed derivation of Eq. (2.4) see Appendix B at the end of the thesis.

The link from monetary policy actions to exchange rates occurs through Uncovered Interest Rate Parity (UIRP), under distortions in beliefs about future interest rates as in Gourinchas and Tornell (2004). Molodtsova and Papell (2009) discuss at length such mechanisms. Under UIRP and rational expectations, any circumstance that causes the home (foreign) central bank to increase its policy rate relative to the foreign (home), will lead to an expected depreciation of its currency. However, the empirical evidence frequently rejects the UIRP condition and this is known as the forward premium puzzle (Engel, 1996). In Gourinchas and Tornell (2004) the puzzle arises due to a systematic distortion in investors' beliefs about the interest rate path. They show theoretically and empirically that under these distorted beliefs, an increase in the home country's interest rate can lead to its currency appreciation, instead of a depreciation as predicted by UIRP.[6] Assuming this evidence an increase in the home country's inflation above the

[6]Consider, for instance, that the central bank increases its policy rate, and allows it to revert back to equilibrium gradually. The adjustment in the country's currency exchange rate will depend on the extent to which investors perceive the path of the interest rate. If investors know the specific interest rate path, the currency will immediately appreciate until the interest rate differential equals the expected depreciation. If investors misjudge the shock as being transitory and believe that the

target, a rise in the output gap, or a deviation of the real exchange from the target will lead to an increase in its interest rate, cause appreciation, and a forecast of additional appreciation.

Using Eq. (2.4) to derive the forecasting regression or estimate Taylor rule fundamentals is valid when parameters are constant over time. In a dynamic world, Taylor rule parameters may be subject to structural instabilities. In this context, rather than estimating or assuming Taylor rules fundamentals from models with constant or calibrated parameters, we allow for the possibility of monetary policies that respond to macroeconomic conditions in a time-varying fashion. We estimate fundamentals from Taylor rules using a TVP regression of the following form:

$$i_t - i_t^* = \phi_{0t} + \phi_{1t}\pi_t - \phi_{1t}^*\pi_t^* + \phi_{2t}\overline{y}_t - \phi_{2t}^*\overline{y}_t^* + \phi_{3t}q_t + \phi_{4t}i_{t-1} - \phi_{4t}^*i_{t-1}^* + \mu_t, \quad (2.5)$$

from which we compute the fundamentals as:

$$\Omega_t = \widehat{\phi}_{0t} + \widehat{\phi}_{1t}\pi_t - \widehat{\phi}_{1t}^*\pi_t^* + \widehat{\phi}_{2t}\overline{y}_t - \widehat{\phi}_{2t}^*\overline{y}_t^* + \widehat{\phi}_{3t}q_t + \widehat{\phi}_{4t}i_{t-1} - \widehat{\phi}_{4t}^*i_{t-1}^* + s_t, \quad (2.6)$$

where $\widehat{\phi}_{lt}$, for $l = 1, ..., 4$, denotes the time $t$ coefficient's estimate. Note that the model in (2.5) is identical to (2.4), except for the time-varying coefficients. Thus, both, the information set from Taylor rules and the exchange rate forecasts are generated from TVP regressions.

The exact form of the Taylor rule and hence of Eq. (2.5) varies depending upon several assumptions. In all Taylor rules, the equilibrium real interest rate and the inflation target of the home and foreign country are assumed identical. This corresponds to setting $\phi_{0t} = 0$ in Eq. (2.5).[7] In addition, all specifications are asymmetric, implying that the home country also targets the real exchange rate.

Our models differ in some ways too, see Table 2.1. The first Taylor rule specification, which we denote $TR_{on}$, assumes homogeneous coefficients and no interest rate smoothing. This restricts the coefficients on inflation ($\phi_{1t} = \phi_{1t}^*$) and the output gap ($\phi_{2t} = \phi_{2t}^*$) of the home and foreign country Taylor rules. Engel and West (2006) find that it is reasonable to assume parameter homogeneity across countries. In addition, central banks do not smooth interest rates ($\phi_{4t} = \phi_{4t}^* = 0$). The assumption of no interest rate smoothing is in line with Engel and West's (2005) formulation. Molodtsova and Papell (2009) also use an identical Taylor rule.

reversion to equilibrium will be fast, the appreciation will be moderate. In the subsequent period, when they observe high than expected interest rates, they revise their perceptions about the duration of the interest rate shock up originating further appreciation. If this last effect surpasses the first, where they know the path of the shock, the currency will appreciate until the true magnitude of duration is perceived, at which point the currency will depreciate to its equilibrium value.

[7] This is a typical assumption in this literature including in Engel and West (2005), Engel et al. (2008), Rogoff and Stavrakeva (2008). As Molodtsova and Papell (2013) also note, whether to include a constant that captures differences in the equilibrium real interest rate and inflation target is irrelevant, because the forecasting regression includes a constant.

Table 2.1: Empirical Exchange Rate Models and Forecasting Approaches

| Fundamental's Exchange Rate Model | TVP Approach | | Constant Parameter Approach | | Forecast Sample and Number of Currencies (N) |
|---|---|---|---|---|---|
| | Information set (Fundamental) | Forecasting model | Information set (Fundamental) | Forecasting model | |
| | Estimated with a random walk Time-Varying Parameter model using Bayesian methods: | Random walk Time Varying Parameter (TVP) model, estimated using Bayesian methods: | Estimated with a single-equation fixed-parameter model, via Ordinary Least Square estimator: | Fixed-effect Panel model, estimated via Least Square Dummy Variable estimator: | A: 1992Q4+h - 1998Q4+h; N=17 (all currencies in the sample); |
| Taylor Rule (TR) | TR$_{on}$:<br>$z_t \equiv i_t - i_t^* = \alpha_{1t}(\pi_t - \pi_t^*) + \alpha_{2t}(\overline{y}_t - \overline{y}_t^*) + \phi_{3t}q_t + \mu_t$<br>TR$_{os}$:<br>$z_t \equiv i_t - i_t^* = \alpha_{1t}(\pi_t - \pi_t^*) + \alpha_{2t}(\overline{y}_t - \overline{y}_t^*) + \alpha_{3t}(i_{t-1} - i_{t-1}^*) + \phi_{3t}q_t + \mu_t$<br>TR$_{en}$:<br>$z_t \equiv i_t - i_t^* = \phi_{1t}\pi_t - \phi_{1t}^*\pi_t^* + \phi_{2t}\overline{y}_t - \phi_{2t}^*\overline{y}_t^* + \phi_{3t}q_t + \mu_t$ | $\Delta s_{t+h} = \beta_{0t} + \beta_{1t}z_t + \varepsilon_{t+h}$ | TR$_{on}$:<br>$z_t \equiv i_t - i_t^* = \alpha_1(\pi_t - \pi_t^*) + \alpha_2(\overline{y}_t - \overline{y}_t^*) + \phi_3 q_t + \mu_t$<br>TR$_{os}$:<br>$z_t \equiv i_t - i_t^* = \alpha_1(\pi_t - \pi_t^*) + \alpha_2(\overline{y}_t - \overline{y}_t^*) + \alpha_3(i_{t-1} - i_{t-1}^*) + \phi_3 q_t + \mu_t$<br>TR$_{en}$:<br>$z_t \equiv i_t - i_t^* = \phi_1\pi_t - \phi_1^*\pi_t^* + \phi_2\overline{y}_t - \phi_2^*\overline{y}_t^* + \phi_3 q_t + \mu_t$ | $\Delta s_{it+h} = \omega_i + \beta z_{it} + \varepsilon_{it+h}$ | B: 1999Q1+h - 2013Q1; N=10 (non-Euro area currencies and the Euro);<br>C: 2007Q1+h - 2013Q1; N=10 (non-Euro area currencies and the Euro). |

Notes: Summary of the models considered and forecasting approaches. The variables are defined as: $i$ = interest rate, $\pi_t$= inflation rate, $y_t$ = output, $\overline{y}_t$ = output gap, $q_t$ = real exchange rate, $m_t$= money, $p_t$= price level, and $s_t$ = nominal exchange rate. The subscripts $i$ and $t$ denote country and time, respectively. The asterisk defines the foreign country (U.S.). Three variants of Taylor rules (TR) are considered: (i) TRon: asymmetric rule with homogeneous coefficients and no interest rate smoothing, (ii) TRos: asymmetric rule with homogeneous coefficients and interest rate smoothing, and (iii) TRen: asymmetric rule with heterogeneous coefficients and no interest rate smoothing. The forecasts are computed for one-, four-, eight-, and 12-quarters-ahead horizons.

A second Taylor rule specification is similar to the above, except that it includes lagged interest rates. This is an asymmetric rule, with homogeneous coefficients and interest rate smoothing ($TR_{os}$). Since the assumption of coefficients' homogeneity between countries is maintained, then $\phi_{4t} = \phi^*_{4t}$ in Eq. (2.5). The inclusion of lagged interest rates implies that central banks limit interest rate variability in the spirit of Engel et al. (2008), Mark (2009), and Molodtsova and Papell (2009).

The third variant relaxes the assumption of homogeneous coefficients across countries, and central banks do not smooth interest rates. In terms of Eq. (2.5), $\phi_{4t} = \phi^*_{4t} = 0$; and is an asymmetric rule, with heterogeneous coefficients and no interest rate smoothing ($TR_{en}$). Molodtsova and Papell (2009) find that models of this type exhibit a strong forecasting performance.

To estimate each of these variants we set up a state-space representation as in Section 2.2, but here the measurement equation is defined by (2.5) and the transition process also follows a random walk. That is, as in Eq. (2.3) but with $\beta_t$ replaced by $\phi_t$. The estimation procedure is equally based on Bayesian methods and details about priors' elicitation, posterior distributions, and sampling algorithm are provided in Appendix C. Like in the forecasting regression, our results rely on data-based information to parameterize priors and the initial conditions.

Apart from our main forecasting regression which allows the coefficients to vary over time, we also forecast with a linear regression, i.e., $\beta_{it} = \beta_i$, for $i = \{0, 1\}$ in Eq. (2.1). To be precise, we use a Fixed-Effect (FE) panel regression, since Engel et al. (2008) argue that panel data methods forecast better than single-equation methods (see also Ince, 2014 and Engel et al., 2015). In any event, in a robustness analysis, we verify whether using a simple linear regression alters the results. In both cases, the information set from Taylor rules is obtained by estimating, via OLS, a single-equation fixed-parameter model similar to Eq. (2.4).

## 2.4 Data and Forecasting Mechanics

### 2.4.1 Data

We use quarterly data spanning 1973Q1 - 2013Q1. Exchange-rates are end-of-quarter values of the national currencies relative to the U.S dollar for the following OECD countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Japan, Italy, Korea, The Netherlands, Norway, Spain, Sweden, Switzerland, and the United Kingdom. The main source is the IMF's *International Financial Statistics* (IFS). Some of the countries in our sample period moved from their national currencies to the Euro. To generate the exchange rate series for these countries, the irrevocable conversion factors adopted by each country on the $1^{st}$ of January 1999 were employed, in the spirit of Engel et al. (2015).

To estimate Taylor rules we need the short-run nominal interest rates set by central banks, inflation rates and the output gap or the unemployment gap.[8] We use the central bank's policy rate when available for the entire sample period, or alternatively the discount rate or the money market rate. The proxy for quarterly output is industrial production (IP) in the last month of the quarter. The output gap is obtained by applying the Hodrick and Prescott (1997) filter recursively, to the IP series. However, to correct for the uncertainty about these estimates at the sample end-points, we follow Watson's (2007) methodology. We estimate bivariate VAR($\ell$) models that include the first difference of inflation and the change in the log IP, with $\ell$ determined by Akaike Information Criterion. These models are then used to forecast and backcast three years of quarterly data-points of IP, and the HP filter is applied to the resulting extended series.[9] The price level consists of the consumer price index (CPI) and the inflation rate is defined as the (log) CPI quarterly change. The data on money supply, IP, unemployment rate and CPI were seasonally adjusted by taking the mean over four quarters following Engel et al. (2015).

### 2.4.2 Forecast Implementation

Our forecasting exercise covers the short and the long horizons. Following Engel et al. (2008, 2015), we use a direct rather than an iterative method to forecast the $h$-quarter-ahead change in the exchange rates for $h = 1, 4, 8, 12$. The benchmark model is the driftless random walk. Since the seminal contribution by Meese and Rogoff (1983) it has been found that it is challenging to improve upon this benchmark (see Rossi, 2013, for a survey of the evidence to date).

The models' parameters are recursively re-estimated in an expanding window and using lagged fundamentals, as in Engel et al. (2015). For concreteness, let $T+h = R+P$ be the sample size comprising a proportion of $R$ observations for in-sample estimation, and $P$ for prediction at $h$-step-forecast horizon. Thus, $T + h$ constitutes the total number of observations after discarding data-points used to parameterize priors for the TVP models. We first use $R$ observations to compute the information set and to generate the parameters of the exchange rate forecasting regression. With these parameters we generate the first $h$-step-ahead forecast and compute the forecast error. We then add one observation at a time to the end of the in-sample period and repeat the same procedure until all $P$ observations are used. This suggests that allowing

[8] In estimating Taylor rules and due to possible endogeneity issues, several authors emphasize the timing of the data employed. The discussion centers on the idea that Taylor rules are forward-looking, and hence ex-post data might reflect policy actions taken in the past. Kim and Nelson (2006) note two approaches that can be employed to account for this. The first comprises using historical real-time forecasts that were available to policy-makers. The second consists in using ex-post data to directly model the policy-maker's expectations. Since historical real-time forecasts are unavailable for our sample of countries, we follow Molodtsova and Papell's (2009) approach, and use data that were observed (as opposed to the real-time forecasts) at time $t$, while forecasting $t + h$ period.

[9] We have also experimented with estimating an AR($\ell$) model for $\Delta \ln(IP_t)$ instead of a VAR($\ell$) model. The resulting output gap series were similar to the those based on the VAR forecasts, suggesting small differences in the forecast precision between the two models. Note that we use the standard smoothing parameter for quarterly data (i.e., 1600).

for time-variation in parameters in a recursive forecasting approach ultimately results in two potential sources of variation in parameters. The first is due to our recursive algorithm when computing the optimal parameter at each time of the in-sample period. The second source arises from extending the sample as observations are added to the end of the in-sample period (recursions). Our TVP forecasting approach is, therefore, highly flexible.[10]

We examine the forecasting performance of our models in three sub-samples. The first out-of-sample forecasts are for the period 1992Q4+$h$ - 1998Q4+$h$. In this sample, forecasts for all the 17 countries' currencies are generated. Since towards the end of the sample the realization occurs during the Euro area, we use the rescaled exchange rate to compare against the forecast. A second forecast sample covers the post-Euro period: 1999Q1+$h$ - 2013Q1. In this case we compute the forecast of the Euro currency as an average of the forecasts of the Euro-area countries in our sample. The forecast error is constructed as the difference between each of the country's realized value and the computed average. We therefore generate forecasts for the nine non-Euro area countries plus the Euro. These procedures draw from Engel et al. (2015). The last out-of-sample forecast period begins just before the recent financial turmoil and extends to the end of the sample, i.e., 2007Q1+$h$ - 2013Q1. Considering this window is particularly important, given the substantial instabilities that characterized the period, with consequences for the monetary policy reaction functions and the variance of the exchange rate. In this sample, we also compute forecasts for 10 currencies, following the procedure just described.

### 2.4.3 Forecast Evaluation

We employ the sample RMSFE to compare the out-of-sample forecasting performance of our models. We compute the ratio of the RMSFE of the fundamentals-based exchange rate model relative to the RMSFE of the driftless random walk, known as the Theil's U-statistic. Hence, models that perform better than the benchmark have a Theil's U less than one.[11]

To evaluate the significance of the differences in the forecasts of competing models, typically, the tests proposed by Diebold and Mariano (1995), West (1996) (hereafter DMW), and Clark and West (2006, 2007) (hereafter CW) are employed. However, Clark and West (2006) show that when comparing nested models, the DMW test is undersized, and hence the RMSFE differential should be adjusted by a term that

[10]Note that R corresponds to 1979Q1 - 1994Q4 for the first forecast sample, 1979Q1 - 1998Q4 for the second forecast sample and 1979Q1 - 2006Q4 for the last forecast sample. The forecast samples are summarised in the last column of Table 2.1. To illustrate the mechanics described in the paragraph, consider the case of h=4 and for the first forecast sample. We use $t$ = 1979Q1 to 1994Q4 to estimate regressions of the form: $s_{t+4} - s_t = \beta_{0t} + \beta_{1t} z_t + \varepsilon_{t+4}$. Using the parameters estimated from this regression, we use data from 1995Q4 to forecast the h=4 change in the exchange rate: $s_{1996Q4} - s_{1995Q4} = \widehat{\beta}_{0t} + \widehat{\beta}_{1t} z_{1995Q4}$. One observation is then added to the end of sample and the procedure is repeated.

[11]By fundamentals-based exchange rate model, we refer to any of the models given in Table 2.1.

accounts for the bias introduced by the larger model. On the other hand, Rogoff and Stavrakeva (2008) make the case for using the bootstrapped DMW test, rather than the CW test, arguing that the latter does not always test for minimum mean square forecast error. They also recall the applicability of the asymptotics of the CW test when forecasting in a rolling window, rather than recursive framework. Hence, we use a semi-parametric bootstrap to construct $p$-values of the DMW test-statistic in the spirit of Kilian (1999) and Rogoff and Stavrakeva (2008). We use this bootstrap to evaluate the forecasts from the FE panel regression and from an additional OLS regression that we consider in a robustness analysis.

To evaluate the forecasts from the TVP regressions we employ a procedure equivalent to the bootstrap above (see also Garratt et al., 2009 and Korobilis, 2013). Since for each draw in our MCMC algorithm we can compute the DMW-test, we can as well obtain the empirical distribution of the test from which we can calculate the critical values. We proceed in this fashion due to the high computational requirements to implement the bootstrap referred above with MCMC methods. See Appendix K for details on these bootstrap procedures.[12,13]

## 2.5 Empirical Results

Tables 2.2, 2.3, and 2.4 present results from the TVP forecasting regression and the Fixed-Effect (FE) panel regression. Each table corresponds to a different forecasting sample and the entries are the relative RMSFE, or simply the U-statistic. Statistically significant differences in the RMSFE, based on our bootstrapped critical values, are marked with asterisks.

Focusing first on the TVP regression, results indicate improvements upon the RW benchmark in the first and most notably, the last forecast sample. In the first forecast sample in Table 2.2, the TVP regression conditioned on fundamentals from Taylor rules with homogenous coefficients and no interest rate smoothing (TRon) outperforms the RW for 11 out of 17 currencies at $h$ =8, and nine out of 17 at $h$ =12. At shorter horizons, however, its performance deteriorates when conditioned on any of the Taylor rule specification. This is the case at $h$ =1 and $h$ =4, where the RW does better for over half of the currencies in either case and regardless of the Taylor rule variant. In contrast, in the last forecast sample in Table 2.4, the TVP regression with either Taylor rule specification beats the RW for at least half of the currencies at $h$ =4, 8, and 12 quarters. The variant with heterogeneous coefficients and no interest rate smoothing (TRen), exhibits the strongest performance; it outforecasts the RW for seven out of 10 currencies at $h$ =4, and six out of ten at $h$ =8 and $h$ =12. Nevertheless, in most cases, the differences in forecasting accuracy relative to the RW are statistically insignificant.

[12]In computing the sample long-run variance for the DMW test, we use the Newey and West's (1987) HAC standard errors, with a lag truncation parameter of $int\{Sample^{0.25}\}$, as in Rossi (2013).

[13]Note that the bootstrap procedure we implement can also be applied to construct bootstrapped critical values and $p$-values for the Clark and West (2006, 2007) test-statistic.

Table 2.2: Theil's U and DMW test, 1992Q4+h - 1998Q4+h

| | TVP Regression | | | | Fixed-Effect Panel Regression | | | |
|---|---|---|---|---|---|---|---|---|
| | U(1) | U(4) | U(8) | U(12) | U(1) | U(4) | U(8) | U(12) |
| | TRon: Homogenous rule, without interest rate smoothing | | | | | | | |
| Australia | **0.978** | **0.932** | **0.884** | **0.806** | **0.990** | 1.016 | **0.971** | **0.875** |
| Canada | **0.963** | **0.929** | **0.791** | **0.909** | **0.981**** | **0.964** | **0.928** | **0.824** |
| Denmark | **0.996** | 1.019 | **0.974** | 1.061 | 1.005 | 1.025 | **0.984** | 1.025 |
| UK | **0.986** | **0.900** | **0.921** | 2.234 | 1.048 | 1.289 | 1.489 | 1.420 |
| Japan | 1.023 | 1.083 | 1.194 | 1.564 | 1.021 | 1.088 | 1.197 | 1.555 |
| Korea | **0.998** | **0.984** | **0.951** | **0.899** | **0.998** | **0.982** | **0.946** | **0.836** |
| Norway | **1.000** | 1.024 | **0.987** | **0.926** | 1.001 | 1.031 | **0.953** | **0.835** |
| Sweden | 1.018 | 1.058 | **0.969** | **0.799** | 1.018 | 1.087 | **0.979** | **0.741** |
| Switzerland | 1.016 | 1.175 | 1.367 | 2.074 | 1.012 | 1.064 | 1.162 | 1.589 |
| Austria | 1.032 | 1.068 | 1.085 | 1.340 | 1.015 | 1.037 | 1.037 | 1.172 |
| Belgium | 1.013 | 1.036 | **0.965** | **0.964** | 1.003 | 1.019 | **0.981** | **0.990** |
| France | 1.042 | 1.076 | **0.993** | **0.912**** | 1.003 | 1.026 | **0.962** | **0.866** |
| Germany | 1.021 | 1.080 | 1.197 | 1.453 | 1.019 | 1.038 | 1.043 | 1.230 |
| Spain | **0.984** | 1.010 | **0.990** | **0.636** | **0.981** | 1.024 | **0.901** | **0.659** |
| Italy | 1.015 | 1.032 | 1.011 | 1.035 | 1.009 | **0.994** | **0.771** | **0.560** |
| Finland | 1.018 | 1.071 | **1.000** | **0.811** | 1.015 | 1.038 | **0.949** | **0.827**** |
| The Netherlands | 1.026 | 1.072 | 1.096 | 1.414 | 1.011 | 1.034 | 1.043 | 1.189 |
| *No. U's <1* | 7 | 4 | 11 | 9 | 4 | 3 | 11 | 10 |
| *No. DMW** | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 |
| | TRos: Homogenous rule, with interest rate smoothing | | | | | | | |
| Australia | 1.008 | 1.117 | 1.089 | **0.830** | **0.996** | 1.071 | 1.048 | **0.881** |
| Canada | **0.964** | **0.934** | **0.943** | **0.864** | **0.967*** | **0.922** | **0.867** | **0.709** |
| Denmark | 1.003 | **0.953** | **0.838** | **0.966** | 1.001 | **0.975** | **0.781**** | **0.966** |
| UK | 1.074 | 1.443 | 1.958 | 1.821 | 1.055 | 1.350 | 1.610 | 1.487 |
| Japan | 1.008 | 1.038 | 1.162 | 1.479 | 1.010 | 1.061 | 1.170 | 1.575 |
| Korea | **0.999** | **0.987** | **0.959** | **0.904** | 1.000 | **0.986** | **0.955** | **0.918** |
| Norway | **1.000** | **0.977** | **0.958** | **0.948** | 1.000 | **0.990** | **0.752**** | **0.697**** |
| Sweden | 1.046 | 1.058 | 1.041 | **0.715** | 1.009 | 1.059 | **0.832** | **0.489*** |
| Switzerland | 1.013 | 1.032 | 1.146 | 2.093 | **0.997** | **0.986** | **0.984** | 1.511 |
| Austria | 1.032 | 1.089 | 1.224 | 1.755 | 1.013 | 1.009 | **0.965** | 1.374 |
| Belgium | **1.000** | **0.907*** | **0.903** | 1.647 | **0.991** | **0.941*** | **0.775**** | **0.954** |
| France | 1.019 | **0.975** | **0.655*** | **0.444*** | **0.997** | **0.959** | **0.708**** | **0.613**** |
| Germany | 1.036 | 1.110 | 1.143 | 1.605 | 1.012 | 1.015 | **0.986** | 1.429 |
| Spain | **0.989** | 1.005 | **0.819** | **0.381*** | **0.994** | 1.021 | **0.752** | **0.381*** |
| Italy | 1.018 | 1.035 | 1.039 | 1.022 | 1.009 | 1.024 | **0.791** | **0.426** |
| Finland | 1.029 | 1.017 | **0.991** | **0.689*** | 1.000 | **0.983** | **0.787*** | **0.704**** |
| The Netherlands | 1.028 | 1.048 | 1.075 | 1.408 | 1.004 | **0.988** | **0.926** | 1.297 |
| *No. U's <1* | 5 | 6 | 8 | 9 | 6 | 9 | 14 | 11 |
| *No. DMW** | 0 | 1 | 1 | 3 | 1 | 1 | 5 | 5 |

Table 2.2: (continued)

| Country | TVP Regression | | | | Fixed-Effect Panel Regression | | | |
|---|---|---|---|---|---|---|---|---|
| | U(1) | U(4) | U(8) | U(12) | U(1) | U(4) | U(8) | U(12) |
| | TRen: Heterogeneous rule, without interest rate smoothing | | | | | | | |
| Australia | 0.984 | **0.955** | **0.889** | **0.818** | **0.980** | **0.983** | **0.958** | **0.933** |
| Canada | **0.957** | **0.929** | **0.783** | **0.879** | **0.971**** | **0.962** | **0.950** | **0.702** |
| Denmark | 1.006 | **0.961**** | **0.942**** | 1.074 | 1.009 | 1.002 | **0.924** | 1.179 |
| UK | 1.078 | 1.343 | 1.436 | 1.855 | 1.050 | 1.251 | 1.414 | 1.387 |
| Japan | 1.032 | 1.118 | 1.219 | 1.629 | 1.024 | 1.085 | 1.190 | 1.596 |
| Korea | **0.999** | **0.987** | **0.955** | **0.889** | **0.998** | **0.985** | **0.970** | **0.912** |
| Norway | 1.017 | 1.094 | 1.192 | 1.371 | **0.991** | **0.994** | 1.020 | 1.319 |
| Sweden | 1.010 | **0.972** | **0.984** | **0.829** | **0.986** | **0.963** | **0.743**** | **0.686**** |
| Switzerland | 1.019 | 1.130 | 1.350 | 2.273 | 1.018 | 1.070 | 1.191 | 1.802 |
| Austria | 1.070 | 1.314 | 1.476 | 1.776 | 1.030 | 1.093 | 1.095 | 1.234 |
| Belgium | 1.074 | 1.275 | 1.435 | 1.832 | **0.992** | **0.979** | **0.950** | 1.114 |
| France | 1.017 | 1.189 | 1.287 | 1.079 | **0.991** | **0.906**** | **0.719**** | **0.874** |
| Germany | 1.040 | 1.134 | 1.184 | 1.456 | 1.044 | 1.099 | 1.102 | 1.453 |
| Spain | **0.987** | **0.955** | **0.702*** | **0.481*** | 1.000 | **0.968** | **0.713*** | **0.663*** |
| Italy | 1.008 | 1.016 | 1.013 | **0.901** | 1.004 | **0.959** | **0.680** | **0.470*** |
| Finland | 1.021 | 1.018 | **0.974** | **0.661*** | **0.986*** | **0.936*** | **0.723*** | **0.846** |
| The Netherlands | 1.084 | 1.205 | 1.270 | 1.463 | 1.023 | 1.095 | 1.147 | 1.353 |
| *No. U's <1* | 4 | 6 | 7 | 7 | 8 | 10 | 10 | 8 |
| *No. DMW** | 0 | 1 | 2 | 2 | 2 | 2 | 4 | 3 |

**Notes**: Forecasting performance of the TVP forecasting regression and the Fixed-effect (FE) panel regression with Taylor rule fundamentals defined as TRon, TRos and TRen. The benchmark model for both forecasting regressions is the driftless Random Walk (RW). The U(h) is the U-statistic for quarterly forecast horizons, h. For example, U(1) is the U-statistic for one-quarter-ahead forecast. Values less than one (in **bold**), indicate that the fundamentals-based regression generates a lower RMSFE than the RW, and hence forecasts better than the RW. The table also reports the DMW test-statistic, with critical values based on a bootstrap procedure, see Section 2.4.3. Thus, asterisks (* 10% , ** 5%, *** 1%) denote the level of significance at which the null hypothesis of equal RMSFE is rejected. This is equivalent to a better average accuracy of the forecasts of the fundamentals-based regression relative to the benchmark. The last two rows at the bottom of the table summarise the results by counting the number of U's less than one - "No. of U's¡1", and the number of rejections of the null under the bootstrapped DMW-test, "No. of DMW*". The prediction sample is 1992Q4+h - 1998Q4+h.

Shifting the focus to the FE panel regression, results show that it forecasts well in the first and last forecast samples, especially at $h$ =4, 8, and 12 quarters. In the first forecast sample for example, 1992Q4+$h$ - 1998Q4+$h$, it outperforms the RW for at least half of the currencies at $h$ =8 and $h$ =12, when conditioned on TRon and TRos information sets. The strongest performance occurs at $h$ =8 and with TRos fundamentals, where it yields a lower RMSFE for as many as 14 out of the 17 currencies. However, as in the TVP regression, the differences in forecast errors are mostly statistically insignificant. For instance, of the 14 currencies for which the FE panel regression generates lower RMSFEs, for only five currencies the null of equal RMSFE is rejected. In the last forecast sample, 2007Q1+$h$ - 2013Q1, the regression produces a lower RMSFE than the RW for a minimum of five out of 10 currencies mainly at $h$ =4, 8, and 12 quarters. The best performance is achieved when the regression is conditioned on TRon fundamentals, where it outforecasts the RW for eight and seven currencies at $h$ =8 and $h$ =12, respectively. And it is again the case that the differences in forecast accuracy are statistically insignificant. We also note that both, the FE panel regression and the TVP regression performed unsatisfactorily in the forecast sample spanning 1999Q1+$h$ - 20013Q1, in Table 2.3. In this sample, forecasts based on the naive RW are more accurate for most currencies/horizons.

On balance, the FE panel regression had a better average performance in the first (early) sample, while the TVP regression outperformed the RW for a large number of currencies in the last sample. To illustrate what determines a U-statistic of certain magnitude for each regression, Figure 2.1 depicts the predicted path of the change in exchange rate based on fundamentals from a TVP Taylor rule *versus* those from a

Table 2.3: Theil's U and DMW test, 1999Q1+h - 2013Q1

| | TVP Regression | | | | Fixed-Effect Panel Regression | | | |
|---|---|---|---|---|---|---|---|---|
| | U(1) | U(4) | U(8) | U(12) | U(1) | U(4) | U(8) | U(12) |
| | TRon: Homogenous rule, without interest rate smoothing | | | | | | | |
| Australia | 1.024 | 1.104 | 1.222 | 1.358 | 1.014 | 1.047 | 1.115 | 1.253 |
| Canada | 1.014 | 1.052 | 1.116 | 1.457 | 1.006 | 1.023 | 1.064 | 1.110 |
| Denmark | 1.002 | 1.011 | 1.047 | 1.033 | 1.006 | 1.023 | 1.054 | 1.058 |
| UK | 1.026 | 1.068 | 1.049 | 1.278 | 1.003 | 1.025 | 1.075 | 1.156 |
| Japan | 1.003 | **0.977** | **0.929** | **0.861** | 1.002 | **0.993** | **0.950** | **0.857** |
| Korea | 1.073 | 1.131 | 1.085 | 1.165 | 1.029 | 1.085 | 1.128 | 1.243 |
| Norway | 1.014 | 1.044 | 1.030 | 1.130 | 1.010 | 1.045 | 1.131 | 1.238 |
| Sweden | 1.012 | 1.044 | 1.093 | 1.250 | 1.015 | 1.064 | 1.160 | 1.416 |
| Switzerland | **0.994** | **0.974** | **0.891** | **0.738** | **0.997** | **0.989** | **0.966** | **0.928** |
| Euro | 1.009 | 1.101 | 1.177 | 1.431 | 1.012 | 1.063 | 1.144 | 1.243 |
| *No. U's <1* | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 2 |
| *No. DMW** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | TRos: Homogenous rule, with interest rate smoothing | | | | | | | |
| Australia | 1.018 | 1.049 | 1.174 | 1.410 | 1.014 | 1.038 | 1.120 | 1.319 |
| Canada | 1.017 | 1.051 | 1.163 | 1.495 | 1.011 | 1.030 | 1.082 | 1.160 |
| Denmark | **1.000** | 1.018 | 1.044 | 1.056 | 1.004 | 1.026 | 1.071 | 1.112 |
| UK | 1.009 | 1.032 | 1.098 | 1.227 | 1.008 | 1.029 | 1.064 | 1.145 |
| Japan | 1.002 | **0.997** | **0.969** | **0.936** | **1.000** | **0.988** | **0.955** | **0.900** |
| Korea | 1.034 | 1.091 | 1.105 | 1.205 | 1.018 | 1.043 | 1.052 | 1.161 |
| Norway | 1.005 | 1.004 | **0.963** | **0.961** | 1.006 | 1.021 | 1.059 | 1.098 |
| Sweden | 1.014 | 1.051 | 1.092 | 1.298 | 1.020 | 1.082 | 1.208 | 1.519 |
| Switzerland | **0.994** | **0.937** | **0.789** | **0.598** | **0.990*** | **0.938*** | **0.826**** | **0.711**** |
| Euro | 1.008 | 1.064 | 1.174 | 1.339 | 1.006 | 1.039 | 1.113 | 1.203 |
| *No. U's <1* | 2 | 2 | 3 | 3 | 2 | 2 | 2 | 2 |
| *No. DMW** | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| | TRen: Heterogeneous rule, without interest rate smoothing | | | | | | | |
| Australia | 1.027 | 1.106 | 1.231 | 1.346 | 1.018 | 1.058 | 1.144 | 1.306 |
| Canada | 1.015 | 1.052 | 1.121 | 1.494 | 1.010 | 1.035 | 1.098 | 1.184 |
| Denmark | 1.006 | 1.018 | 1.040 | 1.036 | 1.000 | **0.996** | 1.000 | **0.984** |
| UK | 1.014 | 1.019 | 1.042 | 1.256 | 1.001 | 1.037 | 1.120 | 1.201 |
| Japan | 1.008 | **0.999** | **0.950** | **0.960** | **0.998** | **0.973** | **0.938** | **0.901** |
| Korea | 1.076 | 1.137 | 1.121 | 1.198 | 1.033 | 1.088 | 1.136 | 1.268 |
| Norway | **0.975** | **0.958** | **0.949** | **0.968** | **0.981***** | **0.985** | 1.037 | 1.093 |
| Sweden | 1.004 | 1.005 | 1.099 | 1.123 | 1.005 | 1.040 | 1.126 | 1.314 |
| Switzerland | **0.996** | **0.952** | **0.879** | **0.692** | **0.989*** | **0.950*** | **0.910*** | **0.870**** |
| Euro | 1.009 | 1.065 | 1.190 | 1.341 | 1.002 | 1.017 | 1.065 | 1.091 |
| *No. U's <1* | 2 | 3 | 3 | 3 | 3 | 4 | 2 | 3 |
| *No. DMW** | 0 | 0 | 0 | 0 | 2 | 1 | 1 | 1 |

**Notes:** See notes to Table 2.2. The prediction sample is 1999Q1+h - 2013Q1.

Table 2.4: Theil's U and DMW test, 2007Q1+h - 2013Q1

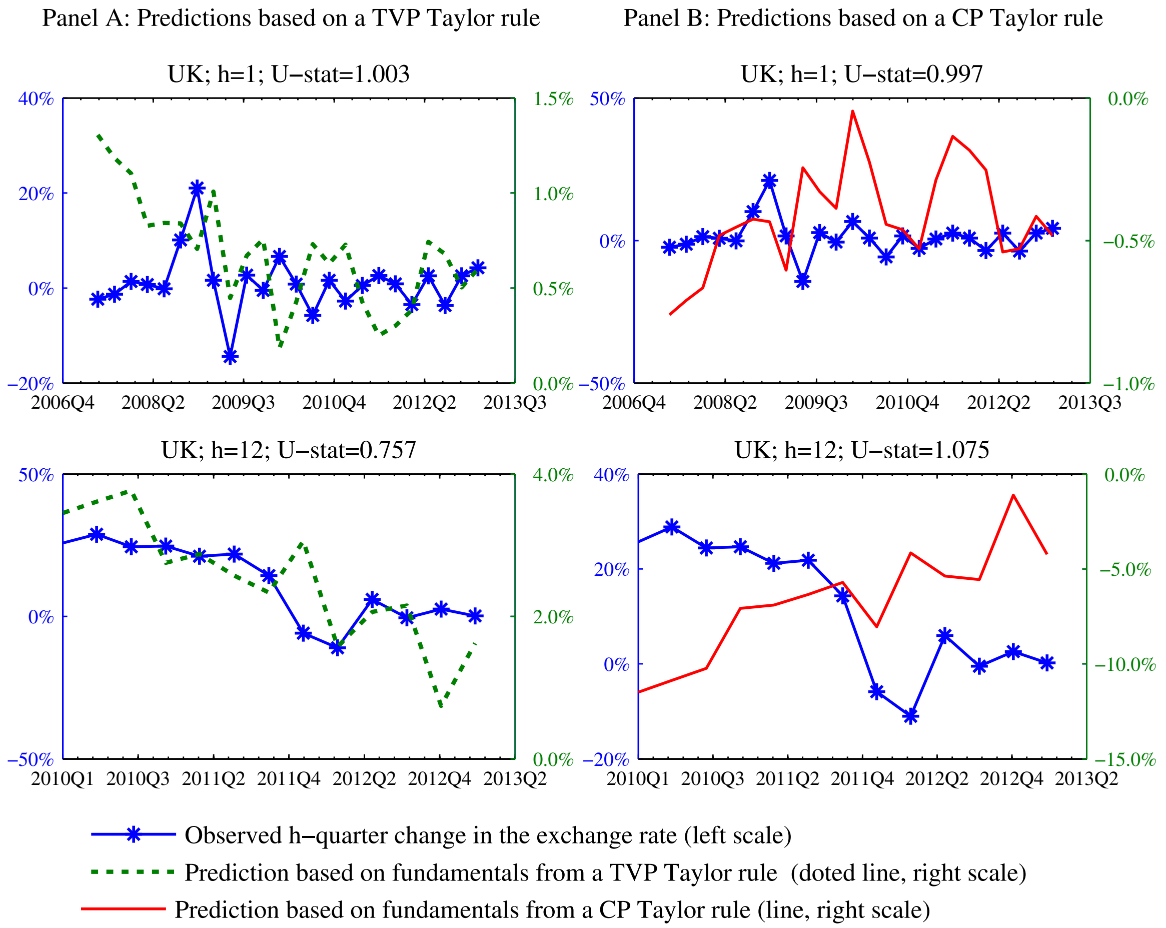
| | TVP Regression | | | | Fixed-Effect Panel Regression | | | |
|---|---|---|---|---|---|---|---|---|
| | U(1) | U(4) | U(8) | U(12) | U(1) | U(4) | U(8) | U(12) |
| | TRon: Homogenous rule, without interest rate smoothing | | | | | | | |
| Australia | 1.019 | 1.103 | 1.311 | 1.548 | 1.007 | 1.018 | 1.061 | 1.218 |
| Canada | 1.007 | 1.019 | 1.050 | 1.678 | 1.002 | 1.001 | **0.971** | **0.952** |
| Denmark | 1.003 | **0.992** | **0.988** | 1.218 | 1.005 | **0.993** | **0.997** | 1.060 |
| UK | 1.010 | 1.005 | **0.941** | **0.754** | 1.000 | **0.993** | **0.978** | **0.985** |
| Japan | **0.999** | **0.918** | **0.815** | **0.736** | **0.997** | **0.890*** | **0.777*** | **0.703*** |
| Korea | **0.995** | **0.958** | **0.933** | **0.985** | **0.995** | **0.974** | **0.952** | **0.936** |
| Norway | 1.009 | 1.026 | **0.996** | 1.062 | 1.004 | 1.006 | **0.978** | **0.953** |
| Sweden | 1.004 | 1.011 | 1.047 | 1.128 | 1.007 | 1.021 | 1.023 | 1.303 |
| Switzerland | **0.991** | **0.940** | **0.752*** | **0.495*** | **0.991** | **0.954** | **0.796** | **0.639*** |
| Euro | 1.008 | **0.999** | **0.894*** | **0.613*** | 1.004 | 1.001 | **0.996** | **0.759** |
| *No. U's <1* | 3 | 5 | 7 | 5 | 3 | 5 | 8 | 7 |
| *No. DMW** | 0 | 0 | 2 | 2 | 0 | 1 | 1 | 2 |
| | TRos: Homogenous rule, with interest rate smoothing | | | | | | | |
| Australia | 1.015 | 1.007 | 1.098 | 1.325 | 1.012 | 1.003 | 1.018 | 1.142 |
| Canada | 1.009 | 1.020 | 1.117 | 1.755 | 1.008 | 1.000 | **0.952** | 1.010 |
| Denmark | 1.014 | 1.055 | 1.097 | 1.321 | 1.011 | 1.034 | 1.094 | 1.279 |
| UK | **1.000** | **0.979** | **0.903** | **0.840** | 1.003 | **0.979** | **0.916** | **0.868** |
| Japan | 1.007 | **0.967** | **0.828** | **0.773** | 1.000 | **0.911** | **0.797** | **0.739** |
| Korea | **0.992** | **0.968** | **0.949** | 1.016 | **0.995** | **0.949** | **0.892** | **0.844** |
| Norway | 1.012 | **0.999** | **0.955** | **0.794*** | 1.013 | 1.007 | **0.934** | **0.860** |
| Sweden | 1.002 | 1.016 | 1.063 | 1.211 | 1.017 | 1.032 | 1.024 | 1.374 |
| Switzerland | **0.997** | **0.954** | **0.675*** | **0.388*** | **0.993** | **0.941** | **0.703** | **0.480**** |
| Euro | 1.005 | 1.004 | **0.896*** | **0.686*** | 1.012 | 1.028 | 1.051 | 1.038 |
| *No. U's <1* | 3 | 5 | 6 | 5 | 2 | 4 | 6 | 5 |
| *No. DMW** | 0 | 0 | 2 | 3 | 0 | 0 | 0 | 1 |
| | TRen: Heterogeneous rule, without interest rate smoothing | | | | | | | |
| Australia | 1.026 | 1.116 | 1.357 | 1.571 | 1.009 | 1.022 | 1.086 | 1.274 |
| Canada | 1.007 | 1.019 | 1.048 | 1.751 | 1.004 | 1.002 | **0.989** | 1.021 |
| Denmark | 1.006 | 1.018 | 1.062 | 1.508 | 1.011 | 1.002 | 1.112 | 1.633 |
| UK | 1.003 | **0.973** | **0.909** | **0.757** | **0.997** | 1.006 | 1.030 | 1.075 |
| Japan | **0.994** | **0.853*** | **0.795** | **0.854** | **0.995** | **0.869*** | **0.769** | **0.770** |
| Korea | 1.007 | **0.984** | **0.980** | 1.081 | **0.991** | **0.953** | **0.910** | **0.908** |
| Norway | **0.968** | **0.942** | **0.896** | **0.740** | **0.974**** | **0.944** | **0.742** | **0.643** |
| Sweden | **0.996** | **0.969** | 1.050 | **0.886*** | 1.001 | **0.995** | **0.966** | 1.188 |
| Switzerland | **0.993** | **0.923** | **0.690** | **0.449*** | **0.983** | **0.903** | **0.648** | **0.680*** |
| Euro | 1.005 | **0.985** | **0.897*** | **0.794*** | 1.015 | 1.009 | 1.161 | 1.248 |
| *No. U's <1* | 4 | 7 | 6 | 6 | 5 | 5 | 6 | 4 |
| *No. DMW** | 0 | 1 | 1 | 3 | 1 | 1 | 0 | 1 |

**Notes:** See notes to Table 2.2. The prediction sample is 2007Q1+h - 2013Q1.

Constant Parameter (CP) Taylor rule, along with the observed $h$-quarter change in the exchange rate. Recall that the former fundamentals are employed with the TVP forecasting regression, and the latter with the FE panel forecasting regression. The example is based on the UK, for the last forecast sample, at $h$=1 and $h = 12$, and the Taylor rule specification with heterogeneous coefficients and no smoothing (TRen). The U-statistics are 1.003 ($h = 1$) and 0.757 ($h = 12$) for the TVP regression, and 0.997 ($h = 1$) and 1.075 ($h = 12$) for the FE panel regression.

Panel A shows the case of the forecasting regression with TVP Taylor rule fundamentals. At one-quarter horizon the regression fails to improve upon the RW; and as

Figure 2.1: Predicted Change in Exchange Rate: TVP Taylor Rules vs. Constant Parameter Taylor Rules




Notes: Predicted change in exchange rate based on fundamentals from a TVP Taylor rule vs. a Constant Parameter (CP) Taylor rule, along with the observed h-quarter change in the exchange rate. The Taylor rule specification assumes heterogeneous coefficients and no smoothing (TRen). The fundamentals, or more precisely the interest rate differentials, are estimated recursively to nest the forecasting method. The out-of-sample period is 2007Q1+h - 2013Q1.

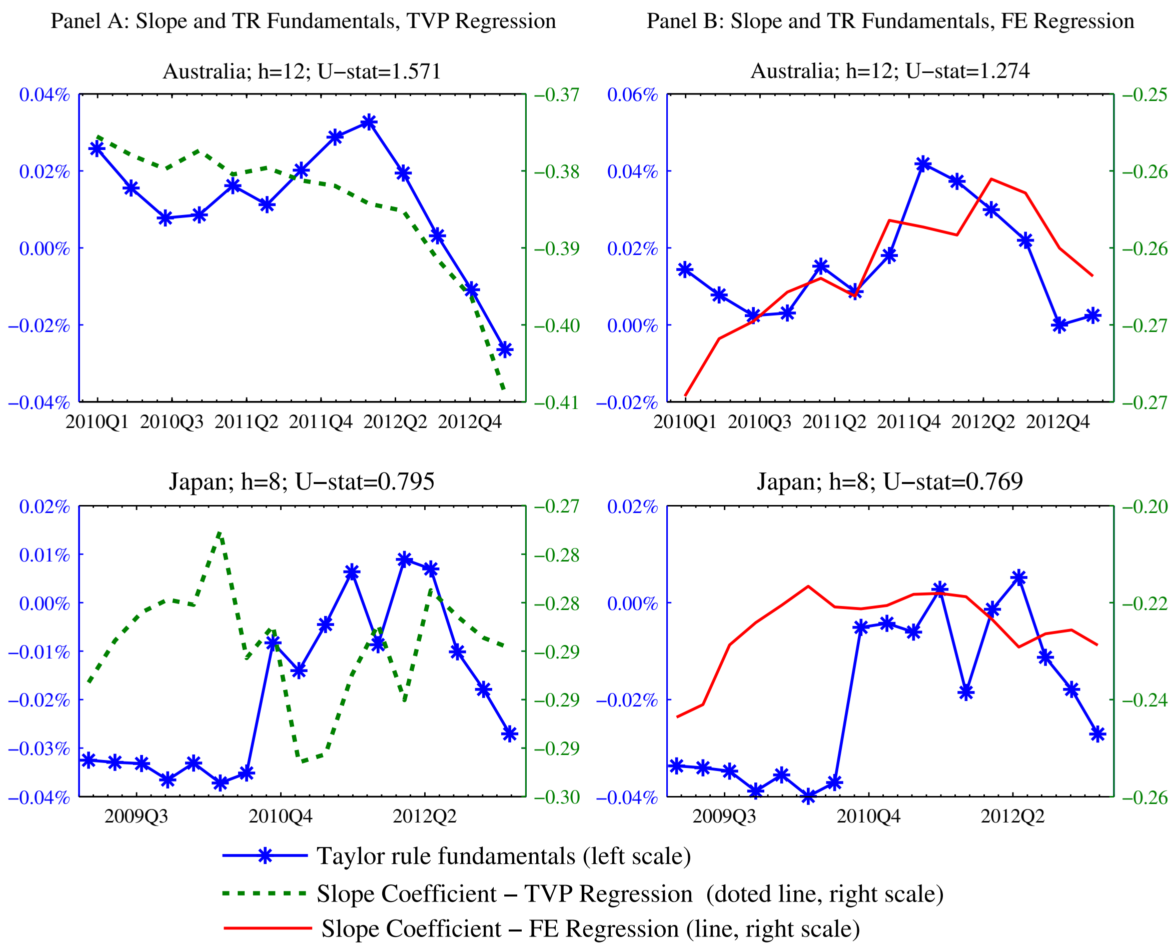
depicted, this might be explained by its failure to predict the path of the subsequent change in the Pound sterling/USD exchange rate in several periods of the forecast sample, resulting in a U-statistic above one. For instance, while it predicts a fall in the Pound sterling from 2007Q2 up to 2008Q4, the data shows an opposite path. In the following periods the regression predicts the correct movements until 2009Q4, failing subsequently until 2010Q3. In the remaining periods it does reasonably well, except between 2011Q1 and 2011Q3. In contrast, at the twelve-quarter-ahead forecast horizon it predicts almost all the subsequent movements in the exchange rate, yielding a U-statistic significantly less than one (U=0.757).

Panel B shows predictions based on the FE panel regression with fundamentals from the constant parameter Taylor rule. At one-quarter forecast horizon, the regression is able to accurately signal the subsequent change in the Pound sterling exchange rate for the most part of the forecast sample. However, since there are also some periods in which it fails, for example between 2009Q1- 2009Q3 and 2011Q3-2012Q2, the gains in terms of reduction in the RMSFE are small (0.3%). By contrast, at twelve-quarter horizon, it correctly signals the changes in the Pound sterling exchange rate only in a few cases, resulting in a U-statistic greater than one (1.075).

To shed more light on the sources of differences in forecasting performance, Figure 2.2 shows the slope coefficients we use to forecast, along with the conditioning information. Panel A shows the coefficients from the TVP forecasting approach, and Panel B the FE panel regression. Once more, we consider the Taylor specification with heterogeneous coefficients, the last forecast sample, and cases of Australia ($h = 12$) and Japan ($h = 8$). The countries and forecast horizons were chosen to illustrate situations where both regressions fail to improve upon the benchmark and situations where they succeed.

As the Figure shows, the slope coefficients are higher in absolute value in the TVP regression than in the FE panel, regardless of forecasting performance. This implies that, in the TVP approach, the larger the magnitude of deviation of the exchange rate from the fundamental, the higher is the speed of correction towards its fundamental

Figure 2.2: Slope Coefficients and Taylor Rule Fundamentals




Notes: Panel A - Slope coefficients from the TVP forecasting regression, along with Taylor rule fundamentals estimated via a TVP regression. Panel B - Slope coefficients from the FE panel regression, along with Taylor rule fundamentals estimated via a constant-parameter OLS regression. The Taylor rule specification assumes heterogeneous coefficients and no smoothing (TRen). The coefficients and the fundamentals are estimated recursively in an expanding window of data to nest the forecasting approach. The out-of-sample period is 2007Q1+h - 2013Q1.

implied level. However, whether this yields a better forecasting performance depends on how well this speed of adjustment accounts for the path of the exchange rate $h$-periods in the future. Since exchange rate deviations from the level implied by fundamentals are frequent, the speed of adjustment (which is estimated in-sample) may fail to account for the path of exchange rate out-of-sample. For example, mean reversion may not be fast enough to correct short-term deviations which are large. Similarly, the relatively small (absolute) value of the slope coefficient for the FE panel regression does not necessarily imply poor forecasting performance. As the graphs illustrate, either a small or a high degree of adjustment may be consistent with satisfactory or unsatisfactory forecasting performance.

The performance of the FE panel regression in our sample is partially similar to the results in Engel et al. (2008). Using a FE panel regression that includes a time effect and a fixed effects, they find that the driftless RW outperforms the Taylor-rule based regression at their short ($h = 1$ quarter) and long ($h = 16$ quarters) forecast horizons. Here, while the findings for the short-run forecasts are similar, for long-run forecasts ($h = 8$ and $h = 12$) we find improvement upon the RW benchmark. We note, however, that there are a number of differences between the analysis in Engel et al. (2008) and ours. Probably the most significant are: (i) the differences in the forecast samples considered and the sample span,[14] and (ii) their use of a Taylor rule specification with posited coefficients, whereas here we estimate the coefficients.

## 2.6 Summary Results and Robustness Checks

Table 2.5 sums up our empirical results. It provides the answer to the following question: "Based on RMSFE, does the forecasting regression conditioned on each of the Taylor rule fundamentals we consider outperform the RW for at least half of the currencies in the sample? If *Yes*, for how many currencies?" It turns out that the FE panel regression accumulates relatively many "Yes" answers in the first forecast sample; and mostly for $h$ =4, $h$ =8, and $h$ =12. The TVP regression aggregates positive answers for a relatively large number of currencies in the last forecast sample, and similar forecasting horizons. The highest improvement occurs when we allow for Taylor rules with heterogeneous coefficients across countries. Thus, our forecasting approach appears to be useful in the recent periods, where significant shifts in fundamentals occurred and exchange rate volatility has been markedly high (see, e.g., Mumtaz and Sunder-Plassmann, 2013).[15]

To verify how robust our results are, we examined different scenarios. These in-

[14] Engel's et al. (2008) sample covers the period 1973Q1-2005Q4, while our sample extends for an extra eight years from 2005Q4.

[15] In unreported results, we further experimented comparing directly the forecasts from the TVP regression with those from the FE panel regression, instead of normalizing forecasts from both approached to the RW. Essentially, we found similar results regarding the better performance of the FE panel relative to the TVP regression in the early forecasting sample and the opposite in the last forecast sample. As well, these differences were largely insignificant using the DMW test.

Table 2.5: Based on RMSFE, Does the Forecasting Regression Outperform the RW for at Least Half of the Currencies in the Sample?

| Fundamentals from: | TVP Regression | | | | FE Panel Regression | | | |
|---|---|---|---|---|---|---|---|---|
| | h=1 | h=4 | h=8 | h=12 | h=1 | h=4 | h=8 | h=12 |
| | Forecast Sample: 1992Q4+h - 1998Q4+h; N=17 | | | | | | | |
| TRon | No | No | Yes (11) | Yes (9) | No | No | Yes (11) | Yes (10) |
| TRos | No | No | No | Yes (9) | No | Yes (9) | Yes (14) | Yes (11) |
| TRen | No | No | No | No | No | Yes (10) | Yes (10) | No |
| | Forecast Sample: 1999Q1+h - 2013Q1; N=10 | | | | | | | |
| TRon | No | No | No | No | No | No | No | No |
| TRos | No | No | No | No | No | No | No | No |
| TRen | No | No | No | No | No | No | No | No |
| | Forecast Sample: 2007Q1+h - 2013Q1; N=10 | | | | | | | |
| TRon | No | Yes (5) | Yes (7) | Yes (5) | No | Yes (5) | Yes (8) | Yes (7) |
| TRos | No | Yes (5) | Yes (6) | Yes (5) | No | No | Yes (6) | Yes (5) |
| TRen | No | Yes (7) | Yes (6) | Yes (6) | Yes (5) | Yes (5) | Yes (6) | No |

**Notes:** Summary of the overall forecasting performance of the TVP regression and the FE panel regression, conditioned on alternative Taylor rule fundamentals - see Table 2.1. The table provides the answer to the question: "Based on RMSFE, does the regression outperform the driftless RW for at least half of the currencies in the sample?". When the answer is "Yes", we indicate the corresponding number of currencies in brackets. For each forecast sample, N is the total number of currencies in that sample.

cluded: (i) forecasting using a linear regression in rolling windows; (ii) changing the base currency from the U.S. dollar to the Pound sterling; (iii) using unemployment gap rather than output gap in the Taylor rule specifications; and (iv) using monthly data, instead of quarterly data. In essence, as we summarize in Table 2.6 and elaborate next, the results from the TVP forecasting approach stand out. Repeatedly, the TVP regression conditioned on Taylor rules with heterogeneous coefficients delivers the highest performance. That is, for horizons greater than one quarter and regardless of the scenario under consideration, it outperforms the driftless RW for at least half of the currencies in the late sample. Regarding the FE panel, its ability to forecast better than the RW for a large number of currencies in the sample is relatively less robust.

### 2.6.1 Forecasting with a Linear Regression in Rolling Windows

Our main forecasting approach allows for time-varying coefficients in the regression used to estimate Taylor rule fundamentals and in the forecasting model. In addition, all the parameters are estimated recursively. Molodtsova and Papell (2009, 2013) and Rossi (2013), estimate Taylor rule fundamentals via OLS in a single-equation constant-

Table 2.6: In Robustness Checks: Based on RMSFE, Does the Forecasting Regression Outperform the RW for at Least Half of the Currencies in the Sample?

| | TVP Regression | | | | Linear Regression | | | |
|---|---|---|---|---|---|---|---|---|
| | h=1 | h=4 | h=8 | h=12 | h=1 | h=4 | h=8 | h=12 |
| | Forecast Sample: 1992Q4+h - 1998Q4+h; N=17 | | | | | | | |
| TRon | No | No | Yes(11) | Yes(9) | No | No | No | No |
| TRos | No | No | No | Yes(9) | No | No | No | No |
| TRen | No | No | No | No | No | No | No | No |
| | Forecast Sample: 1999Q1+h - 2013Q1; N=10 | | | | | | | |
| TRon | No | No | No | No | No | No | No | No |
| TRos | No | No | No | No | No | No | No | No |
| TRen | No | No | No | No | No | No | No | No |
| | Forecast Sample: 2007Q1+h - 2013Q1; N=10 | | | | | | | |
| TRon | No | Yes(5) | Yes(7) | Yes(5) | No | No | Yes(6) | No |
| TRos | No | Yes(5) | Yes(6) | Yes(5) | No | Yes(5) | Yes(5) | Yes(5) |
| TRen | No | Yes(7) | Yes(6) | Yes(6) | No | Yes(6) | Yes(6) | Yes(5) |

| | TVP Regression | | | | FE Panel Regression | | | |
|---|---|---|---|---|---|---|---|---|
| | Change in Base Currency, Forecast Sample: 2007Q1+h - 2013Q1; N=10 | | | | | | | |
| | h=1 | h=4 | h=8 | h=12 | h=1 | h=4 | h=8 | h=12 |
| TRon | Yes (5) | Yes (6) | Yes (6) | Yes (7) | No | No | Yes (5) | Yes (5) |
| TRos | No | Yes (6) | Yes (5) | Yes (8) | No | Yes (5) | Yes (5) | Yes (6) |
| TRen | Yes (6) | Yes (6) | Yes (5) | Yes (5) | No | No | Yes (5) | No |
| | Unemployment Gap, Forecast Sample: 2007Q1+h - 2013Q1; N=9 | | | | | | | |
| | h=1 | h=4 | h=8 | h=12 | h=1 | h=4 | h=8 | h=12 |
| TRon | No | No | No | Yes (5) | No | No | Yes (5) | No |
| TRos | No | No | No | Yes (5) | No | Yes (5) | Yes (6) | No |
| TRen | No | Yes (5) | Yes (5) | Yes (6) | No | No | No | No |
| | Monthly Data, Forecast Sample: 2007Q1+h - 2013Q1; N=10 | | | | | | | |
| | h=3M | h=12M | h=24M | h=36M | h=3M | h=12M | h=24M | h=36M |
| TRon | No | No | Yes (6) | Yes (5) | No | Yes (5) | Yes (7) | Yes (7) |
| TRos | No | Yes (5) | Yes (6) | Yes (6) | No | No | Yes (7) | Yes (5) |
| TRen | Yes (6) | Yes (5) | Yes (7) | Yes (5) | Yes (6) | No | Yes (5) | No |

**Notes:** Summary of the overall forecasting performance of the regressions under different robustness checks scenarios - see Section 2.6. The table provides the answer to the question: "Based on RMSFE, does the regression outperform the driftless RW for at least half of the currencies in the sample?". When the answer is "Yes", we indicate the corresponding number of currencies in brackets. For each forecast sample, N is the total number of currencies in that sample, see also Table 2.1.

Table 2.7: Theil's U and DMW test, Rolling Windows Forecasting Approach

| | TRon | | | | TRen | | | |
|---|---|---|---|---|---|---|---|---|
| | U(1) | U(4) | U(8) | U(12) | U(1) | U(4) | U(8) | U(12) |
| | Forecast Sample: 1992Q4+h - 1998Q4+h | | | | | | | |
| Australia | **0.977** | **0.998** | 1.041 | **0.996** | **0.988** | 1.075 | 1.057 | 1.027 |
| Canada | 1.031 | **0.981** | **0.913** | **0.962** | **0.993** | 1.082 | 1.441 | 2.237 |
| Denmark | 1.070 | 1.227 | 1.446 | 1.602 | 1.061 | 1.169 | 1.264 | 1.660 |
| UK | 1.029 | 1.022 | 1.104 | 1.178 | 1.048 | 1.363 | 1.568 | 1.578 |
| Japan | 1.043 | 1.203 | 1.315 | 1.868 | 1.040 | 1.090 | 1.285 | 1.781 |
| Korea | 1.004 | 1.017 | 1.047 | 1.033 | 1.013 | 1.028 | 1.046 | 1.029 |
| Norway | 1.002 | 1.123 | 1.168 | 1.171 | **0.991** | 1.017 | 1.156 | 1.360 |
| Sweden | 1.043 | 1.050 | 1.101 | 1.395 | 1.024 | 1.263 | 1.186 | 1.583 |
| Switzerland | 1.037 | 1.130 | 1.238 | 1.680 | 1.038 | 1.141 | 1.237 | 1.696 |
| Austria | 1.055 | 1.240 | 1.373 | 1.578 | 1.088 | 1.304 | 1.361 | 1.603 |
| Belgium | 1.062 | 1.312 | 1.515 | 1.556 | **0.985** | 1.225 | 1.407 | 1.802 |
| France | 1.127 | 1.255 | 1.443 | 1.561 | 1.012 | 1.325 | 1.418 | 1.742 |
| Germany | 1.064 | 1.239 | 1.472 | 1.620 | 1.092 | 1.281 | 1.357 | 1.636 |
| Spain | 1.022 | **0.914*** | 1.016 | 1.209 | 1.018 | 1.288 | 1.562 | 1.859 |
| Italy | 1.010 | **0.997** | **0.704** | **0.534*** | 1.022 | 1.195 | 1.266 | 1.062 |
| Finland | 1.042 | 1.024 | **0.916** | 1.066 | 1.075 | 1.121 | **0.818** | 1.134 |
| The Netherlands | 1.049 | 1.130 | 1.226 | 1.343 | 1.035 | 1.228 | 1.268 | 1.511 |
| *No. U's <1* | 1 | 4 | 3 | 3 | 4 | 0 | 1 | 0 |
| *No. DMW** | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| | Forecast Sample: 2007Q1+h - 2013Q1 | | | | | | | |
| Australia | 1.002 | 1.022 | 1.050 | 1.204 | 1.010 | 1.005 | 1.047 | 1.155 |
| Canada | **0.995** | 1.000 | **0.965** | **0.944** | 1.001 | **0.997** | **0.916** | **0.789** |
| Denmark | 1.051 | 1.033 | 1.217 | 1.478 | 1.060 | 1.077 | 1.268 | 1.803 |
| UK | 1.004 | 1.001 | **0.920** | **0.802** | 1.004 | **0.994** | **0.990** | **0.868** |
| Japan | 1.019 | 1.060 | **0.991** | **0.818** | 1.019 | **0.993** | **0.905** | **0.972** |
| Korea | 1.000 | **0.985** | **0.959** | **0.991** | **0.981*** | **0.906*** | **0.861*** | **0.846** |
| Norway | **0.972** | **0.935** | **0.801** | 1.159 | **0.919*** | **0.895** | **0.591** | **0.878** |
| Sweden | 1.001 | 1.010 | **0.936** | 1.052 | 1.003 | **0.952** | **0.885** | 1.046 |
| Switzerland | 1.012 | 1.094 | 1.359 | 1.004 | 1.021 | 1.146 | 1.618 | 1.077 |
| Euro | 1.010 | 1.059 | 1.236 | 1.590 | 1.017 | 1.044 | 1.184 | 1.652 |
| *No. U's <1* | 2 | 2 | 6 | 4 | 2 | 6 | 6 | 5 |
| *No. DMW** | 0 | 0 | 0 | 0 | 2 | 1 | 1 | 0 |

**Notes:** Results obtained by first estimating interest rate differentials (via OLS) with a single-equation constant-parameter model (SECP). The estimates are then employed as conditioning information in a second SECP forecasting regression. The forecasts are generated in rolling windows of (i) 64 quarters for the first forecast sample (1992Q4+h - 1998Q4+h) and 112 quarters for the last forecast sample (2007Q1+h - 2013Q1). These rolling windows were defined such that the number of forecasts generated in this approach matches the number of forecasts in the recursive forecasting approach. The interpretation is similar to Table 2.2.

parameter (SECP) model. These fundamentals are then employed as conditioning information in subsequent SECP forecasting regression, in a rolling window forecasting approach. Accordingly, we explored their methodology. In particular, we defined the rolling windows such that the number of forecasts generated using this method matches with the number of forecasts in the recursive forecasting method. We focused in the first and last forecast samples and two Taylor rule specifications, TRon and TRen.

Results from this experiment are reported in Table 2.7. As shown, this forecasting approach improves upon the RW for at least half of the currencies in the last forecast sample and TRen, although the differences in RMSFE are not significant. In this case, it yields better forecasts for six ($h = 4$ and $h = 8$) and five ($h = 12$) out of the 10 currencies considered. We note though that in this sample and for the same Taylor specification, the TVP regression outperformed the RW for more currencies: seven ($h = 4$) and six ($h = 8$ and $h = 12$), see Table 2.4 in the Empirical Section.

In comparison with other studies that employ a SECP forecasting regression conditioned on Taylor rule fundamentals estimated with a SECP, our results are different. For example, focusing on monthly data up to June 2006, Molodtsova and Papell (2009) find improvement upon the RW benchmark for as many as 10 out of 12 OECD currencies at one-month-ahead forecast horizon. Rossi (2013) uses monthly data up to 2011 and finds improvement over the RW for seven out of 17 currencies at one-month forecast horizon, but for none of the currencies at long horizons. While there are, potentially, several reasons why our results differ from those in the above studies, the most obvious aspects are the differences in the data-frequency, sample period and forecast samples.

### 2.6.2 Change in Base Currency

Chen et al. (2010) and Engel et al. (2015) stress the importance of verifying the sensitiveness of the model's forecasting performance to a different base numeraire. Accordingly, we replaced the U.S. dollar base currency by the Pound sterling (GBP), defined all the home country variables relative to the United Kingdom (UK), and repeated the forecasting exercise for the last forecast sample. In this setting, the TVP regression generated the strongest performance: it significantly outperformed the RW in almost all forecast horizons, regardless of the Taylor rule specification, and up to eight out of 10 currencies - see Table 2.8. The FE panel regression generated more accurate forecasts mostly at longer horizons and for no more than six currencies.

### 2.6.3 Taylor Rules with Unemployment Gap

Monetary policy rules can focus on the unemployment gap rather than the output gap. Molodtsova and Papell (2013) find that Taylor rules with the unemployment gap outperform specifications with the output gap. We therefore replaced the output gap by the unemployment gap and proceeded with the forecasting exercise, focusing on the last forecast sample (2007Q1+h - 2013Q1). Due to unavailability of data on unemployment gap for all the countries in the sample, we forecasted nine exchange rates. As Table 2.9 shows, the TVP regression conditioned on the Taylor rule with heterogeneous coefficients (TRen) delivered the most prominent results. At horizons greater than one quarter, it generated more precise forecasts than the RW for five to six currencies out of the nine considered.[16]

[16] However, in constrast with Molodtsova and Papell (2013), the performance of our TVP regression, as well as of the FE panel regression with either Taylor rule specification, was generally inferior to regressions based on the output gap.

### 2.6.4 Forecasting with Monthly Data

Finally, to verify how results would vary in the frequency of data used we experimented with monthly data. We concentrated on the last forecast sample and four monthly forecast horizons ($h$ =3; $h$ =12; $h$ =24 and $h$ =36) which are comparable to our quarterly horizons. Overall, results in Table 2.10 confirm our main findings: the TVP forecasting approach yields significantly more accurate forecasts in the more recent periods of our dataset. As well, the Taylor rule with heterogeneous coefficients (TRen) has the strongest predictive content. It outperforms the RW for at least 50% of the currencies

Table 2.8: GBP base currency: Theil's U and DMW test, 2007Q1+h - 2013Q1

| | TVP Regression | | | | Fixed-Effect Panel Regression | | | |
|---|---|---|---|---|---|---|---|---|
| | U(1) | U(4) | U(8) | U(12) | U(1) | U(4) | U(8) | U(12) |
| | TRon: Homogenous rule, without interest rate smoothing | | | | | | | |
| Australia | **0.993** | 1.003 | 1.054 | 1.077 | 1.006 | 1.018 | 1.039 | 1.075 |
| Canada | **0.991** | **0.993** | **0.906** | **0.847**** | **0.994** | **0.935*** | **0.917*** | **0.913*** |
| Denmark | 1.008 | **0.941** | **0.942** | **0.973** | 1.008 | 1.021 | 1.010 | **0.995** |
| UK | 1.011 | 1.009 | **0.963** | **0.775** | **0.999** | **0.994** | **0.969** | **0.943** |
| Japan | **0.988** | **0.927** | **0.808**** | **0.752*** | **0.994** | **0.932*** | **0.858** | **0.815** |
| Korea | 1.026 | 1.066 | 1.066 | 1.388 | 1.016 | 1.062 | 1.180 | 1.368 |
| Norway | 1.007 | **0.979**** | **0.968**** | **0.980** | 1.007 | 1.010 | **0.998** | 1.010 |
| Sweden | **0.993** | 1.070 | 1.158 | 1.122 | 1.028 | 1.102 | 1.167 | 1.203 |
| Switzerland | **0.982*** | **0.838**** | **0.779** | **0.466*** | **0.988** | **0.937*** | **0.856**** | **0.834** |
| Euro | 1.000 | **0.998** | 1.011 | **0.934** | 1.008 | 1.026 | 1.022 | 1.010 |
| *No. U's <1* | 5 | 6 | 6 | 7 | 4 | 4 | 5 | 5 |
| *No. DMW** | 1 | 2 | 2 | 3 | 0 | 3 | 2 | 1 |
| | TRos: Homogenous rule, with interest rate smoothing | | | | | | | |
| Australia | 1.030 | 1.074 | 1.107 | 1.110 | 1.000 | 1.003 | 1.028 | 1.083 |
| Canada | 1.006 | **0.921**** | **0.913**** | **0.889*** | **0.990*** | **0.932*** | **0.913*** | **0.912*** |
| Denmark | **0.995** | 1.005 | 1.027 | 1.006 | 1.005 | 1.014 | 1.007 | **0.983** |
| UK | **0.999** | **0.975** | **0.917** | **0.839** | 1.002 | **0.990** | **0.958** | **0.950** |
| Japan | 1.003 | **0.954** | **0.850** | **0.791** | **0.994** | **0.931*** | **0.856** | **0.817** |
| Korea | 1.010 | 1.065 | 1.025 | **0.885** | 1.010 | 1.051 | 1.160 | 1.394 |
| Norway | 1.005 | **0.988** | **0.996** | **0.970** | 1.003 | **0.996** | **0.988** | 1.019 |
| Sweden | 1.005 | 1.070 | 1.154 | **0.923** | 1.020 | 1.094 | 1.170 | 1.204 |
| Switzerland | **0.970*** | **0.884** | **0.765**** | **0.441*** | **0.982*** | **0.920*** | **0.834*** | **0.854** |
| Euro | **0.999** | **0.992** | 1.007 | **0.940** | 1.005 | 1.019 | 1.019 | **0.998** |
| *No. U's <1* | 4 | 6 | 5 | 8 | 3 | 5 | 5 | 6 |
| *No. DMW** | 1 | 1 | 2 | 2 | 2 | 3 | 2 | 1 |
| | TRen: Heterogeneous rule, without interest rate smoothing | | | | | | | |
| Australia | 1.020 | 1.020 | 1.032 | 1.066 | 1.009 | 1.023 | 1.038 | 1.077 |
| Canada | 1.003 | **0.941**** | **0.887**** | **0.896*** | **0.997** | **0.952** | **0.931*** | **0.920*** |
| Denmark | **0.989*** | 1.043 | 1.132 | 1.155 | 1.010 | 1.032 | 1.024 | 1.014 |
| UK | **0.992** | **0.982** | **0.922** | **0.777** | **0.998** | 1.001 | **0.983** | **0.951** |
| Japan | **0.995** | **0.956** | **0.891**** | **0.920** | **0.994** | **0.934*** | **0.864** | **0.818** |
| Korea | 1.002 | **1.082** | 1.134 | 1.532 | 1.014 | 1.037 | 1.159 | 1.366 |
| Norway | 1.018 | **0.987** | **0.999** | 1.012 | 1.001 | 1.003 | **0.976** | 1.003 |
| Sweden | **0.994** | 1.058 | 1.139 | 1.121 | 1.018 | 1.079 | 1.142 | 1.199 |
| Switzerland | **0.979*** | **0.876** | **0.777**** | **0.301*** | **0.986** | **0.920*** | **0.831*** | **0.825*** |
| Euro | **0.999** | **0.991** | 1.018 | **0.947** | 1.010 | 1.045 | 1.041 | 1.027 |
| *No. U's <1* | 6 | 6 | 5 | 5 | 4 | 3 | 5 | 4 |
| *No. DMW** | 2 | 1 | 3 | 2 | 0 | 2 | 2 | 2 |

**Notes:** For interpretation of the entries in the Table, see notes to Table 2.2. Here the base currency is the Pound Sterling (GBP), rather than the U.S dollar. The forecasting sample is 2007Q1+h - 2013Q1.

in the sample at all forecasting horizons. We interpret these results as an endorsement of our approach to allow for time-evolving fundamentals, and more generally, time-changing dynamics in the interaction between exchange rates and fundamentals.

Table 2.9: Theil's U and DMW test: Taylor Rule Fundamentals based on Unemployment Gap, 2007Q1+h - 2013Q1

| | TVP Regression | | | | Fixed-Effect Panel Regression | | | |
|---|---|---|---|---|---|---|---|---|
| | U(1) | U(4) | U(8) | U(12) | U(1) | U(4) | U(8) | U(12) |
| | TRon: Homogenous rule, without interest rate smoothing | | | | | | | |
| Australia | 1.014 | 1.065 | 1.281 | 1.530 | 1.004 | 1.012 | 1.061 | 1.235 |
| Canada | **0.999** | 1.006 | 1.050 | 1.738 | 1.000 | **0.993** | **0.965*** | **0.965** |
| Denmark | **0.998** | 1.013 | **0.930*** | **0.817**** | 1.008 | 1.001 | 1.029 | 1.200 |
| UK | 1.003 | 1.019 | 1.048 | **0.826** | 1.001 | **0.988** | **0.956**** | **0.938**** |
| Japan | **0.994** | **0.874**** | **0.768**** | **0.753**** | **0.994** | **0.871**** | **0.763**** | **0.715**** |
| Norway | 1.024 | 1.059 | 1.101 | 1.178 | 1.007 | 1.015 | 1.016 | 1.253 |
| Sweden | 1.009 | 1.015 | 1.064 | 1.109 | 1.005 | 1.006 | **0.983** | 1.043 |
| Switzerland | 1.005 | **0.952** | **0.816**** | **0.415**** | **0.988** | **0.926** | **0.695**** | **0.288**** |
| Euro | 1.014 | 1.013 | **0.963*** | **0.776*** | 1.015 | 1.070 | 1.295 | 2.048 |
| *No. U's <1* | 3 | 2 | 4 | 5 | 3 | 4 | 5 | 4 |
| *No. DMW** | 0 | 1 | 4 | 4 | 0 | 1 | 4 | 3 |
| | TRos: Homogenous rule, with interest rate smoothing | | | | | | | |
| Australia | 1.015 | 1.004 | 1.093 | 1.329 | 1.012 | **0.990** | **0.992** | 1.127 |
| Canada | 1.009 | 1.020 | 1.148 | 1.787 | 1.010 | **0.996** | **0.951**** | 1.031 |
| Denmark | 1.014 | 1.060 | 1.113 | 1.350 | 1.015 | 1.053 | 1.134 | 1.355 |
| UK | **0.999** | **0.978** | **0.910*** | **0.827*** | 1.005 | **0.973** | **0.903*** | **0.850**** |
| Japan | 1.004 | **0.956** | **0.824*** | **0.783**** | 1.001 | **0.919** | **0.802**** | **0.758**** |
| Norway | 1.013 | 1.003 | **0.961*** | **0.811** | 1.008 | 1.021 | 1.050 | 1.268 |
| Sweden | **1.000** | 1.015 | 1.065 | 1.194 | 1.020 | 1.022 | **0.939** | 1.069 |
| Switzerland | **0.994** | **0.952** | **0.677*** | **0.400**** | **0.996** | **0.958** | **0.678*** | **0.329**** |
| Euro | 1.005 | 1.029 | 1.016 | **0.821*** | 1.024 | 1.105 | 1.399 | 1.906 |
| *No. U's <1* | 3 | 3 | 4 | 5 | 1 | 5 | 6 | 3 |
| *No. DMW** | 0 | 0 | 4 | 4 | 0 | 0 | 4 | 3 |
| | TRen: Heterogeneous rule, without interest rate smoothing | | | | | | | |
| Australia | 1.014 | 1.065 | 1.279 | 1.550 | 1.009 | 1.031 | 1.101 | 1.303 |
| Canada | 1.009 | 1.034 | 1.137 | 1.650 | 1.006 | 1.008 | 1.000 | 1.049 |
| Denmark | 1.011 | 1.017 | 1.110 | 1.293 | 1.018 | 1.041 | 1.167 | 1.699 |
| UK | **0.999** | **0.984** | **0.935** | **0.839*** | 1.002 | **0.996** | **0.982** | **0.990** |
| Japan | **0.991** | **0.864**** | **0.773***** | **0.772**** | **0.994** | **0.860**** | **0.739**** | **0.685**** |
| Norway | **0.983*** | **0.965**** | **0.936** | **0.820** | 1.011 | 1.025 | 1.058 | 1.357 |
| Sweden | 1.006 | **0.991** | 1.072 | **0.927*** | **0.980*** | **0.946*** | **0.843**** | **0.768*** |
| Switzerland | **0.991** | **0.915** | **0.698**** | **0.439**** | **0.992** | **0.930** | **0.686** | **0.586**** |
| Euro | 1.012 | 1.014 | **0.931*** | **0.755*** | 1.028 | 1.149 | 1.522 | 2.447 |
| *No. U's <1* | 4 | 5 | 5 | 6 | 3 | 4 | 4 | 4 |
| *No. DMW** | 1 | 2 | 3 | 5 | 1 | 2 | 2 | 3 |

**Notes:** For interpretation of the entries in the Table, see notes to Table 2.2. Here the fundamentals are from Taylor rules estimated with Unemployment gap, rather than Output gap. The forecasting sample is 2007Q1+h - 2013Q1.

Table 2.10: Monthly data: Theil's U and DMW, 2007M1+h - 2013M5

| | TVP Regression | | | | Fixed-Effect Panel Regression | | | |
|---|---|---|---|---|---|---|---|---|
| | U(3M) | U(12M) | U(24M) | U(36M) | U(3M) | U(12M) | U(24M) | U(36M) |
| | TRon: Homogenous rule, without interest rate smoothing | | | | | | | |
| Canada | 1.005 | 1.032 | 1.114 | 1.664 | 1.004 | **1.000** | **0.969**** | **0.962** |
| Denmark | **1.000** | 1.028 | 1.036 | 1.168 | 1.004 | 1.003 | 1.014 | 1.090 |
| UK | 1.012 | 1.008 | **0.948** | **0.763**** | 1.002 | **0.994** | **0.967**** | **0.944***** |
| Japan | **0.998** | **0.908*** | **0.779***** | **0.751***** | **0.996** | **0.911**** | **0.801***** | **0.728***** |
| Korea | 0.998 | 0.986 | 0.968 | 1.034 | 1.000 | 0.988 | 0.968 | 0.973 |
| Norway | 1.015 | 1.008 | **0.984*** | **0.970***** | 1.005 | 1.004 | 0.981 | 0.981 |
| Sweden | 1.004 | 1.013 | 1.025 | 1.185 | 1.007 | 1.013 | 1.007 | 1.183 |
| Switzerland | **0.993** | **0.947** | **0.755***** | **0.517***** | **0.992** | **0.960** | **0.833***** | **0.701***** |
| Euro | 1.011 | 1.002 | **0.920** | **0.797**** | 1.002 | 1.002 | 0.995 | 0.923* |
| *No. U's <1* | 4 | 3 | 6 | 5 | 2 | 5 | 7 | 7 |
| *No. DMW** | 0 | 1 | 3 | 5 | 0 | 1 | 4 | 4 |
| | TRos: Homogenous rule, with interest rate smoothing | | | | | | | |
| Canada | 1.009 | 1.014 | 1.137 | 1.659 | 1.010 | 1.009 | **0.967**** | 1.029 |
| Denmark | 1.015 | 1.037 | 1.105 | 1.259 | 1.013 | 1.028 | 1.112 | 1.251 |
| UK | 1.009 | **0.986** | **0.898*** | **0.778**** | 1.008 | **0.995** | **0.938*** | **0.886**** |
| Japan | 1.018 | **0.984** | **0.854*** | **0.795***** | 1.002 | **0.935** | **0.829**** | **0.763***** |
| Korea | 0.998 | 0.976 | **0.941** | **0.999** | 1.001 | 0.972 | 0.911 | 0.871 |
| Norway | 1.013 | 1.015 | **0.999** | **0.894** | 1.015 | 1.019 | 0.991 | 0.966 |
| Sweden | 0.993 | 1.005 | 1.037 | 1.212 | 1.011 | 1.020 | 0.998 | 1.225 |
| Switzerland | **0.995** | **0.958** | **0.708**** | **0.337***** | **0.994** | **0.952** | **0.751***** | **0.495***** |
| Euro | **0.999** | 0.997 | **0.928** | **0.825*** | 1.013 | 1.025 | 1.074 | 1.062 |
| *No. U's <1* | 4 | 5 | 6 | 6 | 1 | 4 | 7 | 5 |
| *No. DMW** | 0 | 0 | 3 | 4 | 0 | 0 | 4 | 3 |
| | TRen: Heterogeneous rule, without interest rate smoothing | | | | | | | |
| Canada | 1.009 | 1.023 | 1.056 | 1.666 | 1.007 | 1.012 | 1.018 | 1.090 |
| Denmark | 0.998 | 1.039 | 1.023 | 1.532 | 1.007 | 1.012 | 1.107 | 1.548 |
| UK | 1.024 | **0.982** | **0.934** | **0.735*** | 0.991 | 1.010 | 1.029 | 1.073 |
| Japan | **0.994** | **0.885**** | **0.791***** | **0.816***** | **0.997** | **0.910*** | **0.780***** | **0.760***** |
| Korea | **1.000** | 1.002 | 0.993 | 1.057 | 0.990 | 0.961 | 0.910 | 0.911 |
| Norway | **0.978***** | **0.943***** | 0.895* | 0.802 | 0.979** | **0.942***** | 0.750*** | 0.773 |
| Sweden | 0.991 | 1.005 | 0.986 | 1.125 | 0.995 | 1.004 | 0.945 | 1.165 |
| Switzerland | **0.990** | **0.932** | **0.695***** | **0.504***** | **0.975** | **0.917*** | **0.681***** | **0.630***** |
| Euro | 1.011 | 0.999 | **0.940** | **0.841** | 1.000 | 1.009 | 1.108 | 1.345 |
| *No. U's <1* | 6 | 5 | 7 | 5 | 6 | 4 | 5 | 4 |
| *No. DMW** | 1 | 2 | 3 | 3 | 1 | 3 | 3 | 2 |

**Notes:** For interpretation of the entries in the Table, see notes to Table 2.2. In the Table, the forecasts are based on a monthly-frequency data. Thus, the forecast horizon is defined in months. The forecast sample is 2007M1+h - 2013M5, and comparable results for U(3M), U(12M), U(24M) and U(36M) in the quarterly-frequency data, 2007Q1+h - 2013Q1, are reported in Table 2.4.

## 2.7 Conclusion

An expanding literature articulates the view that Taylor rules are helpful in predicting exchange rates in the sense that structural exchange rate models incorporating Taylor rule fundamentals exhibit predictive content for exchange rates. See, for example, Engel and West (2005) and Molodtsova and Papell (2009). At the same time, an established literature documents time-evolving macroeconomic conditions and relationships among macroeconomic variables (e.g., Stock and Watson, 1996). Taken together, these observations raise the possibility that accounting for time-evolving dynamics may be fundamental to improve exchange rate models' forecasting ability.

To explore this possibility, we estimate Taylor rule fundamentals with Time Varying Parameters (TVP) models and examine their predictive content for exchange rates in a framework that also allows for the parameters of the forecasting regression to change over time. We focus in three alternative forecast samples and four quarterly forecast horizons. In the more recent parts of our dataset and horizons beyond 1-quarter, our approach yields a lower root mean squared forecast error than the driftless random walk for at least half of the currencies in the sample, reaching as many as seven out of 10. Results are especially strong when the TVPs of the Taylor rule are allowed to differ between countries. We interpret this support for heterogeneity as reflecting the varying degree at which country-specific fundamentals altered during the recent financial turmoil.

When we experiment with the usual approach in the literature, whereby constant-parameter models are used to compute Taylor rule fundamentals and forecast, we find a slightly limited performance in the recent turbulent periods. However, these constant parameter models do well in the earlier parts of our dataset. Our TVP regressions also perform only marginally better than standard linear regressions employed in a rolling window forecasting approach. Our results are robust to a number of situations, including to the use of an alternative forecasting approach (rolling windows), to changing the base currency, to using monthly data, and to using unemployment gap in the Taylor rule. Hence, we remain optimistic about the forecasting approach we pursue.

# Chapter 3

# Data-Based Evidence on Instabilities in Exchange Rate Predictability

## 3.1 Introduction

Thirty years on since Meese and Rogoff (1983) identified that exchange rate fluctuations are difficult to predict using standard economic models, the search for a satisfactory rationalization of their findings remains active. As Rossi (2013) notes, the task of finding such answer is by no means an easy endeavour. Decisions regarding the choice of the predictor, forecast horizon, forecasting model, and methods for forecast evaluation, all exert influence in uncovering exchange rate predictability. Ultimately, the predictive power appears to be specific to some countries in certain periods, signalling the presence of instability in the models' forecasting performance (Rogoff and Stavrakeva, 2008; Rossi, 2013). The issue of instability was also put forward by Meese and Rogoff (1983) and is echoed in other recent papers including, Bacchetta and van Wincoop (2004, 2013), Bacchetta et al. (2010), Sarno and Valente (2009), among others. However, as Rossi (2013) points out, models that take into account these instabilities, by allowing for time-variation in the coefficients for instance, do not greatly succeed in outperforming a random walk benchmark in an out-of-sample forecasting exercise.

In this chapter, we employ a framework that allows us to pin down several sources of instability that might affect the out-of-sample forecasting performance of exchange rate models. The starting point of our analysis is the exact conjecture by Meese and Rogoff (1983), that time-variation in parameters may play a significant role in explaining the predictive power of these models. However, unlike prior attempts to explain this conjecture, we do not assume ex-ante that coefficients in the forecasting regressions change in the same fashion over time (e.g., Rossi, 2006). Instead, we allow for a range of possible degrees of time-variation in coefficients, encompassing moderate to sudden changes, and even no-change in coefficients. We then use a likelihood-based approach to identify what degree of time-variation in coefficients is consistent with the data. In this framework we can infer, for example, whether allowing for sudden changes in coefficients leads to a better forecasting performance, relative to situations where coefficients change gradually or remain constant over time.

In light of the hypotheses advanced in recent papers, not only are coefficients in an exchange rate model likely to change over time, but the relevant set of exchange rate determinants may also differ at each point in time. See for example the scapegoat theory of exchange rates of Bacchetta and van Wincoop (2004, 2013), as well as the empirical evidence in Berge (2013), Fratzscher et al. (2015), Markiewicz (2012), and Sarno and Valente (2009). Hence in our setting, in addition to allowing for varying degrees of coefficients adaptivity over time, we also entertain the possibility that a different fundamental may be relevant at each point in time. In this unified framework, we can examine whether models with a certain configuration, characterized by a specific degree of time-variation in coefficients and choice of predictor, can forecast well.

Our key contribution in this chapter goes much further than merely establishing whether our models outperform a random walk benchmark. As the evidence on time-varying forecasting performance suggests, the possibility that a model with a specific configuration can forecast well in a certain period and country, and not in another setting, introduces uncertainty regarding the ex-ante choice of the model. In this context, our unified approach provides the ideal framework to analyze the sources of exchange rate models' prediction uncertainty. Within it, we can distinguish between (i) model uncertainty due to errors when estimating the coefficients, (ii) model uncertainty originating from time-variation in coefficients, (iii) model uncertainty due to a time-varying set of exogenous predictors, and (iv) model uncertainty due to random or unpredictable fluctuations in the data. To the best of our knowledge, this is the first exchange rate prediction study that seeks to accomplish these goals. We investigate, for example, how relevant is the issue of time-variation in coefficients relative to the choice of fundamentals when forecasting out-of-sample.

We use dynamic linear models of the sort considered in Dangl and Halling (2012) when examining stock returns, which we also extend to allow for a time-changing volatility. These models not only allow for a time-varying relationship between exchange rates and fundamentals, but also facilitate assigning posterior probability weights to specifications that differ in the selected fundamental and in the degree of time-variation in coefficients, in light of the relevant evidence. We can then find the specification supported by the data at each point in time, based on these weights. The methodology is also flexible enough in that it enables us to decompose the prediction variance of the exchange rate into its constituent components, highlighting the origins of prediction uncertainty.

Our predictive regressions employ information sets from Taylor rules and classic fundamentals. Engel and West (2005) use an exchange rate model based on Taylor (1993) rules as an example of models that are consistent with the present-value asset pricing framework. Molodtsova and Papell (2009) and Molodtsova et al. (2011)

examine the out-of-sample predictive content of different Taylor rule specifications. Molodtsova and Papell (2009) find evidence of predictability for most currencies at short-horizons. Nevertheless, consistent with the hypothesis that the relevant set of predictors may change over time, in their results predictability differs for different specifications across countries and periods. In Molodtsova and Papell (2009) the strongest support is from Taylor rule specifications with heterogeneous coefficients and interest rate smoothing. In contrast, in Molodtsova et al. (2011), the most successful Taylor rules impose equality in coefficients across countries, and do not incorporate interest rate smoothing.

In terms of the empirical design, our dataset consists of monthly data spanning 1977M1 - 2013M5 on seven OECD currency exchange rates relative to the US dollar. We use a direct method to forecast recursively, the period-ahead change in the exchange rate at 1-, 3-, and 12-months horizons. The forecasts from our fundamentals-based models are compared to those of the toughest benchmark – the driftless random walk (RW) (Rossi, 2013). We compute the ratio of the Root Mean Squared Forecast Error of our models relative to that of the RW. To evaluate the statistical significance of the differences in the forecasts we use the Diebold and Mariano (1995) and West (1996) tests. In order to take account of concerns about data-mining in light of our search over multiple predictors, we employ critical values computed using a data-mining robust bootstrap technique proposed in Inoue and Kilian (2005) and implemented, for example, in Rapach and Wohar (2006). An additional measure of relative forecast accuracy is based on predictive likelihoods (Geweke and Amisano, 2010).

While measures of statistical significance reveal the degree of forecast accuracy, they fall short of measuring whether an investor conditioning on the models' forecasts would be able to obtain tangible economic gains. Hence, following recent studies, including Della Corte et al. (2012) and Li et al. (2015), we also examine the ability of our models to generate economic value in stylized asset portfolio management setting. Set in a mean-variance analysis, we compute indicators such as, Sharpe ratios, maximum performance fees, excess premium returns, and break-even transaction costs that render an investor indifferent between using our models and the RW.

Apart from the research on the role of instabilities in obstructing model forecasting performance, this study is related to the literature on forecast combinations. Among articles studying the importance of instabilities in an exchange rate setting, we compare our analysis to Rossi and Sekhposyan (2011), Bacchetta et al. (2010), and Giannone (2010). Rossi and Sekhposyan (2011) decompose measures of out-of-sample forecasting performance into components of relative predictive ability. Their results point to a lack of predictive content and time-variation in forecasting performance as the main obstacles to models' forecasting ability. However, while they mention that time-variation in parameters of the models might cause time-variation in forecasting performance, they do not explicitly examine the influence of the former in the latter. Thus, our study complements theirs, as time-variation in parameters is an integral part of our scrutiny.[1]

Among papers focusing in pooling exchange rate forecasts, we note contributions by Wright (2008), Sarno and Valente (2009), Beckmann and Schuessler (2015), and Li et al. (2015). The main difference with our contribution is their emphasis on finding whether combined forecasts from several models with a certain configuration are superior to those from univariate models and to the random walk benchmark. Instead, we extend the analysis in the above papers to examine the origins of model prediction uncertainty. An additional difference is our use of a data-mining robust bootstrap procedure when evaluating our models' forecasting performance.[2]

To preview our results, we find that models which allow for the relevant set of predictors to change over time and with varying degrees of coefficients adaptivity forecast well. For the majority of the currencies we examine, these models significantly outperform the benchmark at all, but 1-month forecasting horizon. At horizons greater than 1-month, regressions with a high degree of time-variation in coefficients dominate regressions with constant and moderately time-varying coefficients. They also provide economic gains greater than those accruing from strategies based on the RW and the constant-coefficients models. Importantly, we identify that uncertainty in coefficient estimation and uncertainty regarding the correct level of time-variation in coefficients are key obstructions to exchange rate prediction. When the models successfully embed these sources of uncertainty, they yield a satisfactory out-of-sample forecasting performance and economic value. In this sense, our findings are consistent with the simulation-based results of Giannone (2010) and they provide supportive evidence for Rossi and Sekhposyan's (2011) conjectures on the causes of time-variation in the models' predictive ability.

The chapter is organized as follows. In the next Section we lay out our econometric methodology. Section 3.3 covers data description and forecasting mechanics. We cover the range of statistical and economic criteria for evaluation in Section 3.4. Results are reported in Section 3.5, followed by robustness checks in Section 3.6. Section 3.7 concludes.

[1]Bacchetta et al. (2010) use a theoretical reduced-form model of exchange rate calibrated to match the moments of the data to examine whether parameter instability could rationalize the Meese-Rogoff puzzle. They conclude that it is not time-variation in parameters, but small sample estimation error that explains the puzzle. However, Giannone (2010) disputes these findings and points out that both, time-variation in parameters and estimation uncertainty, are important in accounting for the puzzle. As we noted above, we extend the analysis to consider other sources of instabilities, quantify their relative importance, and our approach is entirely data-based.

[2]Sarno and Valente (2009) use a Reality Check procedure to account for data-mining.

## 3.2 Econometric Methodology

### 3.2.1 Predictive Regression

In line with the majority of the studies in exchange rate forecasting we model the exchange rate as a function of its deviation from its fundamental's implied value.[3] As advanced by Mark (1995), this fits with the notion that in the short-run, exchange rates frequently deviate from their long-run fundamental's implied level. More precisely, let $s_{t+h} - s_t \equiv \Delta s_{t+h}$ be the $h$-step-ahead change in the log of the exchange rate, and $\Omega_t$ a set of exchange rate fundamentals. Then, we consider predictive regressions of the following state space form:

$$\Delta s_{t+h} = X_t'\theta_t + v_{t+h}, \quad v_{t+h} \sim N(0, V_t), \text{ (observation equation)}; \tag{3.1}$$

$$\theta_t = \theta_{t-1} + \varpi_t, \quad \varpi_t \sim N(0, W_t), \text{ (transition equation)}; \tag{3.2}$$

where

$$X_t = [1, z_t], \text{ and } \theta_t = [\theta_{0t}; \theta_{1t}]; \tag{3.3}$$

$$z_t = \Omega_t - s_t. \tag{3.4}$$

As Eq. (3.4) indicates, $z_t$ measures the disequilibrium between the exchange rate's spot value and the level of the fundamentals. When the spot exchange rate is higher than its fundamental's implied level, then the spot rate is expected to decrease, as long as the coefficient attached to $z_t$ in Eq. (3.1) is less than one. In the next Subsection we discuss what spans our set of fundamentals contained in $\Omega_t$. In this Section we note that the predictive regression given by the system of equations (3.1) and (3.2) allows the coefficient linked to the disequilibrium term $z_t$, and to the constant to change over time. In fact, as Eq. (3.2) suggests, we assume a random walk process for the parameter $\theta_t$, following Wolff (1987), Rossi (2006), Mumtaz and Sunder-Plassmann (2013), among others. We further assume that the disturbance terms, $v_{t+h}$ and $\varpi_t$, are uncorrelated and normally distributed with mean zero and time-time-varying matrices $V_t$ and $W_t$, respectively.

The variance of the error term in the transition equation, $W_t$, is crucial in determining the degree of time-variation in the regression's coefficient. Setting this matrix to zero implies that the coefficients are constant over time, and therefore Eq. (3.1) nests a constant-coefficients predictive regression. In contrast, if the variance increases, the shocks to the coefficients also increase. While this renders more flexibility to the model, the increased variability of the coefficients translates into high prediction variance, which increases the prediction error. In light of this, a common practice is to impose some structure on $W_t$; see, for example, Dangl and Halling (2012), Koop and Korobilis (2012), Raftery et al. (2010) and West and Harrison (1997, Ch. 4). We define

[3] See, for example, Cheung et al. (2005), Engel et al. (2008), Mark (1995), Molodtsova and Papell (2009), and Rossi (2013).

this structure together with the description of the estimation methodology below.

We use Bayesian methods to estimate the parameters of our dynamic linear model, following Dangl and Halling (2012) and West and Harrison (1997 Ch. 3&4). The estimation is based on full conjugate Bayesian analysis, implying that when prior information on the unknown parameters is combined with the likelihood function, we obtain a posterior with the same distribution as the prior, hence no simulation algorithms are required. Specifically, let the prior for the coefficients vector $\theta_t$ be normally distributed, and the prior for the observational variance $V_t$ come from an inverse-gamma distribution. In a conjugate analysis, the posteriors are jointly normally/inverse-gamma distributed. In Appendix D.1 at the end of the thesis we provide details on the updating scheme of the system of equations at some arbitrary time $t+1$, given the information available at time $t$ ($D_t$). This information set contains the exchange rate variations, the predictors, and the prior parameters at time-zero. i.e., $D_t = [\Delta s_t, \Delta s_{t-h}, ..., \Delta s_1, X_t, X_{t-h}, ..., X_1, Priors_{t=0}]$. We note that for the prior parameters at $t = 0$, we follow Fernandez et al. (2001) and use a benchmark conjugate g-prior specification:

$$\theta_0 | D_0, V_0 \sim N\left(0, H_0 \left[gX'X\right]^{-1}\right), \tag{3.5}$$

$$V_0 | D_0 \sim IG\left[\frac{1}{2}, \frac{1}{2} H_0\right], \tag{3.6}$$

where

$$H_0 = \frac{1}{N-1} \Delta s' (I - X(X'X)^{-1} X') \Delta s. \tag{3.7}$$

The prior for the coefficient vector in expression (3.5) is a diffuse prior centered around the null-hypothesis of no predictability, with $g$ as the scaling factor that conveys the confidence assigned to this hypothesis. The coefficients' variance-covariance matrix is a multiple of the OLS estimate of the variance in coefficients, $H_0$. The fact that this matrix is multiplied by a large scalar translates into an uninformative prior, implying that the estimation procedure adapts quickly to the empirical pattern. This is consistent with our objective of examining which instabilities are supported by the data. In our empirical exercise, we adopt a common procedure and set our g-prior based on estimates from the entire sample.[4] Following the recommendations in Fernandez et al. (2001), we set $g = 1/T$ for the main results and examine cases of $g = [0.5, 1]$, but find similarities in the results, hence we do not report results based on the other values of $g$.

The other crucial element in the methodology we employ is the predictive density. This is obtained by integrating the conditional density of $\Delta s_{t+h}$ over the space spanned by $\theta_t$ and $V_t$. West and Harrison (1997 Ch. 4) show that it is a Student $t-$distribution with $n_t$ degrees-of-freedom, mean $\widehat{\Delta s}_{t+h}$, variance $Q_{t+h}$, evaluated at $\Delta s_{t+h}$ (for details,

[4]See, for instance, Wright (2008) and Dangl and Halling (2012).

see Appendix D.1):

$$f(\Delta s_{t+h}|D_t) = \mathbf{t}_{n_t}(\Delta s_{t+h}; \widehat{\Delta s}_{t+h}, Q_{t+h}). \tag{3.8}$$

Using this predictive distribution we can recursively forecast $\Delta s_{t+h}$.

As we pointed out, the degree of time-variation in the regressions' coefficients is determined by the matrix $W_t$. Given that the coefficients are exposed to random shocks that follow a normal distribution with mean zero and variance $W_t$, when the variance is low, the estimation error shrinks towards zero as more data becomes available. In contrast, in periods of high variance the estimation error increases, affecting the prediction. To capture this direct relationship between the coefficients' estimation error and the variance, we follow Dangl and Halling (2012) and let $W_t$ be proportional to the variance of the coefficients at time $t$:

$$W_t = \frac{1-\delta}{\delta} H_t C_t^*, \quad 0 < \delta \leq 1; \tag{3.9}$$

where $H_t$ is the estimate of the variance of the error term in the observation equation, $C_t^*$ is the estimated conditional covariance matrix of $\theta_{t-1}$, and $\delta$ is a discount factor that controls the degree of time-variation in coefficients.

Effectively, setting $\delta = 1$ implies that $W_t = 0$, and therefore the coefficients are assumed constant over-time. By contrast, specifying $0 < \delta < 1$ is consistent with time-varying coefficients, with the underlying variability determined by the magnitude of increase in the variance by a ratio of $1/\delta$. For instance, with $\delta = 0.98$ the variance increases by 50% within 20 months. Reducing $\delta$ to 0.96, translates into 50% increase in 10 months, suggesting very abrupt changes in coefficients. Thus, in our empirical work we consider $\delta = [0.96, 0.97, 0.98, 0.99, 1.00]$ as the possible support points for time-variation in coefficients. We then examine empirically which support point is consistent with the data in a Bayesian model averaging approach, which we discuss in the next Section.

### 3.2.2 Dynamic Model Averaging and Selection

While allowing for time-varying coefficients addresses one potential source of instability in predictive ability, the literature on exchange rate predictability also points out that the relevant set of predictors appears to change over time (Bacchetta and van Wincoop, 2004; Rossi, 2013; and Sarno and Valente, 2009). To address this latter source of instability, we allow for the possibility that from a set of $k$ potential predictors, one applies at each time period. If we let $d$ be the number of possible discrete support points for time-variation in coefficients as defined by each $\delta$, then our range of possible models is $d.k$.

The range of predictors we consider follows from the recent literature exploiting the information content from Taylor (1993) rules and classic fundamentals.[5] See for example, Engel and West (2005), Engel et al. (2008), Mark (2009), Molodtsova et al. (2011), Molodtsova and Papell (2009), and Rossi (2013). The premise in this literature is that exchange rates and fundamentals are related in a manner that is coherent with asset pricing models. Therefore, the exchange rate can be expressed as a present-value of a linear combination of fundamentals and random noise - see Appendix A for details. When combined with rational expectations and a random walk process for the fundamentals, the spot exchange rate becomes a function of current observable fundamentals and unexpected noise (Engel and West, 2005).

In our empirical work we consider observable fundamentals from:

- A symmetric (TRsy) and an asymmetric (TRasy) Taylor rule:

$$\begin{aligned}\Omega_{t,TRsy} &= 1.5(\pi_t - \pi_t^*) + 0.5(\overline{y}_t - \overline{y}_t^*) + e_t,\\ \Omega_{t,TRasy} &= 1.5(\pi_t - \pi_t^*) + 0.1(\overline{y}_t - \overline{y}_t^*) + 0.1(e_t + p_t^* - p_t) + e_t,\end{aligned}$$

  where $\pi_t$ is the inflation rate, $\overline{y}_t$ the output gap in the home country, $p_t$ is the log of the domestic price level, and $e_t$ the log exchange rate. Asterisks denote identical variables for the foreign country. Fundamentals of this type are considered, for example, in Engel et al. (2008, 2015).

- The Monetary Model (MM): $\Omega_{t,MM} = (m_t - m_t^*) - (y_t - y_t^*)$, where $m_t$ is the log of money supply and $y_t$ is the log of income;[6]

- The Purchasing Power Parity (PPP) condition: $\Omega_{t,PPP} = (p_t - p_t^*)$; and

- The Uncovered Interest Rate Parity (UIRP) condition: $\Omega_{t,UIRP} = (i_t - i_t^*) + e_t$, where $i_t$ denotes the short-term interest rate.

Selecting one specific model characterized by a certain fundamental and degree of time-variation in coefficients, and using it to forecast at time $t$, requires a method. Bayesian model selection is a methodical approach that tests the validity of all $d.k$ models against the observed data. The approach involves assigning prior probabilities to each candidate predictor (fundamental), as well as prior probability to each possible support point for time-variation in parameters. Then based on the realized likelihood of the model's prediction, the posterior probability of each of the $d.k$ models is updated according to Bayes rule. In Appendix D.2 we provide details on the exact formulae. Note that we elicit diffuse conditional prior probability for each predictor $M_i$; and

[5]The Taylor (1993) rule postulates that monetary authorities should set the policy interest rate considering the recent inflation path, inflation deviation from its target, output deviation from its potential level, and the equilibrium real interest rate. Then, it follows that they increase the short-term interest rate when inflation is above the target and/or output is above its potential level. Note that the Taylor principle presupposes an increase in the nominal policy rate more than the rise in inflation rate to stabilize the economy.

[6]Note that we have assumed an income elasticity of one in the monetary model, following Mark (1995) and Engel and West (2005).

equally, an uninformative prior for the range of support points for the degree of time-variation in coefficients. In notation, the prior probabilities are $P(M_i|\delta_j, D_0) = 1/k$ and $P(\delta_j|D_0) = 1/d$, respectively. Hence, at the beginning of the forecast window, each predictor and model setting has the same chance of becoming probable.

The overall model's predictive density is the posterior probability weighted average predictive density of all $k.d$ models. In this sense, we perform Bayesian Model Averaging (BMA) in a setting with varying degrees of time-variation in coefficients. The flexibility of the approach implies, for instance, that we can implement Bayesian Model Selection (BMS), thus selecting the single model with the highest probability at each point and using it to forecast. We can further let $\delta = 1$, such that all the models exhibit constant coefficients and then average over models with this characteristic (BMA excluding time-varying coefficients). We can alternatively keep $\delta = 1$, but select the best model at each time-period (BMS excluding time-varying coefficients).[7] Furthermore, the approach permits us to track all sources of uncertainty with respect to the prediction in a variance decomposition framework. We elaborate on this framework in what follows.

### 3.2.3 Variance Decomposition and Sources of Instability

We use the law of total variance to decompose the variance of the random variable ($\Delta s$) into its constituent parts. Following Dangl and Halling (2012), we begin with the decomposition with respect to different values of $\delta$:

$$Var(\Delta s) = E_\delta(Var(\Delta s|\delta)) + Var_\delta(E(\Delta s|\delta)), \tag{3.10}$$

where, $E_\delta$ and $Var_\delta$ indicate the expected value and the variance with regards to $\delta$. Since the expected value of the variance of $\delta$ is conditional on specific choice of model $M$, it can be also decomposed as follows:

$$Var(\Delta s|\delta) = E_M(Var(\Delta s|M, \delta)) + Var_M(E(\Delta s|M, \delta)). \tag{3.11}$$

Using Eq. (3.11) to substitute for the expected value of the variance of $\delta$ in Eq. (3.10), and employing the corresponding expressions of these variances detailed in appendix

[7] In fact, as we show in the empirical section, we can analyze several other cases depending on model specifications and choice of degree of time-variation, including cases of BMA over time-varying coefficients with single predictors.

D, we obtain (Dangl and Halling, 2012):

$$
\begin{aligned}
Var(\Delta s_{t+h}) \;=\; & \sum_j \left[ \sum_i (H_t|M_i, \delta_j, D_t) P(M_i|\delta_j, D_t) \right] P(\delta_j|D_t) \\
& + \sum_j \left[ \sum_i (X_t' R_t X_t|M_i, \delta_j, D_t) P(M_i|\delta_j, D_t) \right] P(\delta_j|D_t) \\
& + \sum_j \left[ \sum_i (\widehat{\Delta s}^j_{t+h,i} - \widehat{\Delta s}^j_{t+h})^2 P(M_i|\delta_j, D_t) \right] P(\delta_j|D_t) \\
& + \sum_j (\widehat{\Delta s}^j_{t+h} - \widehat{\Delta s}_{t+h})^2 P(\delta_j|D_t). \qquad (3.12)
\end{aligned}
$$

The four individual terms in Eq. (3.12) highlight the sources of uncertainty in the prediction. The first term is the expected variance of the disturbance term in the observation equation, with $(H_t|M_i, \delta_j, D_t)$ measuring the time $t$ estimate of the variance $V_t$, given the choice of the predictor and degree of time-variation in coefficients. This provides a measure of random fluctuations in the data relative to the predicted trend component. The second term captures the expected variance from errors in the estimation of the coefficients. It can be referred to as estimation uncertainty. The third term characterizes model uncertainty with respect to the choice of the predictor. The last term also characterizes model uncertainty, but with respect to time-variability of the coefficients. Hence both, the third and fourth term, capture model uncertainty. Implementing this four-fold variance decomposition represents an innovation in the exchange rate literature. We now turn to our data and forecasting mechanics.

## 3.3 Data and Forecasting Mechanics

We use monthly data from 1977M1 to 2013M5 for seven countries: Canada, Euro-area(Germany), Japan, Norway, Sweden, Switzerland, and the United Kingdom (UK). The home country is taken as the United States. The data is obtained from the IMF's International Financial Statistics (IFS), supplemented by national central banks sources. The exchange rate is defined as the end-of-month value of the U.S. dollar (USD) price of a unit of national currency. We measure the money supply by the aggregate M1.[8]

Computation of information sets from Taylor rules necessitates data on the short-run central bank nominal interest rate, the inflation rate, and the output gap. We employ the central bank's policy rate when available for the entire sample period and, alternatively, the discount rate or the money market rate. The inflation rate is calculated as the change in the (log of) monthly consumer price index (CPI). We use monthly

[8] In cases where the M1 aggregate is unavailable, we use a broader aggregate. This is M3 for Sweden and Belgium; and M4 for the UK. Note too that data limitations prevent us from using real-time data for the countries we consider. For extra details on Data, see Appendix G.

industrial production (IP) as the proxy for output. Following a common practice in the literature, the output gap is obtained by applying the Hodrick and Prescott (1997) filter recursively to the output series. We equally correct for the uncertainty about these estimates at our recursive sample end-points by following Watson's (2007) method. This entails estimating bivariate VAR($\ell$) models that include the first difference of inflation and the change in the log IP, with $\ell$ determined by Akaike Information criterion. These VARs are used to forecast and backcast three years of monthly data-points of IP, and the HP filter is applied to the resulting extended series.[9] The data on money supply, IP, and CPI were seasonally adjusted by taking the mean over twelve months following Engel et al. (2015).

We use a direct, rather than an iterative, method to forecast the $h$-month-ahead change in the exchange rate for $h = [1, 3, 12]$. As Wright (2008) notes, both methods lead to qualitatively similar conclusions. The forecasting exercise is based on a recursive approach using data available up to the time the forecast is made. For example, a 3-months ahead forecast of the change in exchange rate for 1995M1 is made using data available up to 1994M10. Our forecasting window begins in 1987M12+$h$ for all regressions, except for one of the competing forecast combination method we consider in the empirical results Section, which requires a holdout out-of-sample period.

## 3.4 Forecast Evaluation Methods

### 3.4.1 Statistical Evaluation Criteria

We first apply statistical measures of forecasting performance. We compute the ratio of the Root Mean Squared Forecast Error (RMSFE) of our models relative to RMSFE of the driftless random walk (RW). Models that perform better than the RW benchmark have a value of this ratio, also known as the Theil's U, less than one. According to Rossi (2013), the RW is the most appropriate benchmark.

To assess the statistical significance of the differences in the forecasts, many studies employ the Diebold and Mariano (1995) and West (1996) tests (hereafter DMW), and/or the Clark and West (2006, 2007) test (hereafter CW). The DMW tests whether two competing forecasts are identical under general conditions (Diebold, 2015). The CW tests whether the benchmark model is equivalent to the competing model in population. However, Clark and West (2006) show that when comparing nested models, the DMW test is undersized, hence, the RMSFE differential should be adjusted by a term that accounts for the bias introduced by the larger model. On the other hand, Rogoff and Stavrakeva (2008) make the case for using the bootstrapped DMW test, rather than the CW test, arguing that the latter does not always test for minimum

[9] We have also experimented with estimating an AR($\ell$) model for $\Delta \ln(IP_t)$ instead of a VAR($\ell$) model. The resulting output gap series were similar to those based on the VAR forecasts, suggesting small differences in the forecast precision between the two models. In the HP filter, we use the standard smoothing parameter for monthly data frequency (i.e., 14400).

mean squared forecast error. Additionally, they recall that the asymptotics of the CW test are well-defined when forecasting in a rolling, rather than recursive framework. Accordingly, we construct bootstrapped critical values for the DMW test in the spirit of Rogoff and Stavrakeva (2008). In light of our search over several predictors, however, we modify their procedure to account for data-mining, following Inoue and Kilian (2005) and Rapach and Wohar (2006). See Appendix K for full details on the procedure. Bootstrapping also accounts for the fact that for horizons beyond 1-month, the forecast errors are likely to be serially correlated.[10]

### 3.4.2 Economic Evaluation Criteria

A limitation of the statistical measures of forecasting performance is their inability to convey the economic gains associated with better forecasting performance. To address this limitation we follow Della Corte et al. (2012) and Li et al. (2015) and use economic evaluation criteria. We consider a stylized dynamic asset allocation strategy in which a US investor rebalances her portfolio by allocating her assets between a US bond and seven alternative bonds ($B7$). She can rebalance her portfolio either monthly, in three-months period or in twelve-months period, which correspond to the horizons for which she generates exchange rates forecasts ($h = 1, 3, 12$). At every period, the $B7$ bonds carry a risk-free return in local currency but a risky return $r_{t+h}$ in US dollars. The yields on the bonds consist of the Eurodeposit rates at the respective adjustment periods. When pursuing an investment strategy in the $B7$ bonds, she expects a return of $r_{t+h|t} = i_t^* + \Delta s_{t+h|t}$, where, $r_{t+h|t} = E_t[r_{t+h}]$ is the conditional expectation of $r_{t+h}$; $i_t^*$ denotes the nominal interest rate in the corresponding countries; and $\Delta s_{t+h|t} = E_t[\Delta s_{t+h}]$ is the conditional expectation of $\Delta s_{t+h}$. Since the interest rate is known at time $t$, the return that the investor projects from time $t$ to $t + h$ is only exposed to the exchange rate risk.

In her dynamic asset allocation process the investor wishes to minimize the foreign exchange risk exposure by finding the optimal portfolio weights. At each time period and rebalancing horizon, she uses our models' forecasts to rebalance her portfolio by calculating new optimal weights on each bond, taking into account the portfolio's mean return and variance. Essentially, her problem is to find the optimal portfolio weights subject to a target volatility of the portfolio returns. The solution to this problem assigns the following weights on the risky bonds (see, Della Corte et al., 2012):

$$w_t^g = \frac{\sigma_p^*}{\sqrt{C_t^g}} \Sigma_{t+h|t}^{-1} (u_{t+h|t}^g - \iota r_f); \tag{3.13}$$

[10] The Diebold and Mariano (1995) and West (1996) test is computed as: $DMW = \overline{f}\sqrt{P}/[\text{sample variance of } \widehat{f}_{t+h} - \overline{f}]^{1/2}$; where $P$ is the number of out-of-sample forecasts, $\widehat{f}_{t+h} = \widehat{fe}_{1,t+h}^2 - \widehat{fe}_{2,t+h}^2$, with $\widehat{fe}_{1,t+h}$ denoting the h-step-ahead forecast error of the RW, $\widehat{fe}_{2,t+h}$ the corresponding forecast error of the alternative model, and $\overline{f}$ is the mean of $\widehat{f}_{t+h}$. Note that in computing the tests, including in our bootstrap, we use HAC standard errors with a lag truncation parameter of $int\{Sample^{0.25}\}$, following Rossi (2013).

where $C_t^g = (u_{t+h|t}^g - \iota r_f)'\Sigma_{t+h|t}^{-1}(u_{t+h|t}^g - \iota r_f)$; $\sigma_p^*$ is the target volatility; $\Sigma_{t+h|t} = [(r_{t+h} - u_{t+h|t}^g)(r_{t+h} - u_{t+h|t}^g)']$ is the conditional covariance matrix; $r_{t+h}$ is the $(B7 \times 1)$ vector of risky asset returns; and $u_{t+h|t}^g = E_t[r_{t+h}]$ defines the conditional expectation of $r_{t+h}$. The weight on the risk-free asset is $(1 - w_t'\iota)$.

The gross return on the portfolio is given by:

$$R_{p,t+h} = 1 + w_t^{g\prime} r_{t+h} + (1 - w_t^{g\prime}\iota) r_f = R_f + w_t^{g\prime}(R_t^g - \iota R_f); \tag{3.14}$$

where, $R_t^g$ is the $(B7 \times 1)$ vector of gross returns on risky bonds and $R_f$ are the gross returns on the risk-free bond. In line with Della Corte et al. (2012) and Li et al. (2015), the investor replaces the unconditional covariance matrix with the conditional one when solving her portfolio allocation problem: $\Sigma_{t+h|t} = \overline{\Sigma}$. This ensures that the optimal weights will vary only to the extent of the differences in our models' forecasting ability.

We can employ this framework to examine our models' ability to produce tangible economic benefits, compared to a strategy that uses forecasts from the driftless random walk (RW). We compute the following indicators of economic value:

- Sharpe ratio ($SR$): Defined as the ratio of the average realized excess portfolio returns relative to the portfolio returns standard deviation. We compute Sharpe ratios associated with our models' forecasts and the forecasts from the RW. Higher Sharpe ratios are preferred to lower ones. To assess if differences in Sharpe ratios are statistically significant, we apply the bootstrap method propounded by Ledoit and Wolf (2008).

- Maximum performance fee ($pf$): Proposed by Fleming et al. (2001), this is the fee that a risk-averse investor with quadratic utility is willing to pay to use our models rather than the RW. The measure is founded on the principle that forecasts from a specific model are superior to those from an alternative one, if investment decisions that rely on the specific model yield greater average realized utility than the alternative. The starting point is the assumption that a portfolio based on the RW for example, generates the same average utility as compared to a portfolio based on an alternative model that is subject to expenses $pf$, at $h$-month(s) horizon. Since the investor would be neutral between the two strategies, $pf$ is interpreted as the maximum performance fee that she is ready to pay to swap from the RW to the alternative model. It is often expressed as a fraction of the wealth invested and obtained by solving:

$$\begin{aligned} &\sum_{t=0}^{T-h} \left\{ (R_{p,t+h}^* - pf) - \frac{RRA}{2(1+RRA)} (R_{p,t+h}^* - pf)^2 \right\} \\ = &\sum_{t=0}^{T-h} \left\{ R_{p,t+h} - \frac{RRA}{2(1+RRA)} R_{p,t+h}^2 \right\}; \end{aligned} \tag{3.15}$$

where, $R^*_{p,t+h}$ is the gross return from using our models, $R_{p,t+h}$ is the gross return from the benchmark RW, and $RRA$ is the investor's constant degree of relative risk aversion. Larger values of $pf$ suggest that the investor wishes to pay more to swap from the RW to our models (we report $pf$ in annualized basis points - *bps*).

- Excess premium ($pr$): Based on the manipulation-proof performance measure of Goetzmann et al. (2007), this indicator captures the portfolio's premium return after adjusting for risk. To compute the premium, first obtain the manipulation-proof performance measure:

$$M(R_p) = \frac{1}{1-RRA} \ln \left\{ \frac{1}{T} \sum_{t=0}^{T-h} \left( \frac{R_{p,t+h}}{R_f} \right)^{1-RRA} \right\}; \tag{3.16}$$

where $M(R_p)$ is the risk adjusted portfolio's premium return from the RW, while a similar measure for our models is $M(R^*_p)$. The excess premium return from our models relative to the RW is therefore:

$$pr = M(R^*_p) - M(R_p). \tag{3.17}$$

Higher values of $pr$ are indicative of greater premium returns of our models relative to the RW after accounting for risk. We equally present $pr$ in annualized *bps*.

- Break-even transaction costs: These are the proportional transaction costs that eliminate any positive performance fee obtained by conditioning on our models. When the investor reaches this point, she becomes indifferent between using the RW and our models. To compute the cost, we follow Han (2006) and Della Corte et al. (2012), and assume that transaction costs constitute a fixed fraction ($T_r$) of the value traded in each bond. Therefore, the costs are: $T_r \left| w_t - w_{t-h}(R^g_t/R_p) \right|$. In cases where the investor's transaction costs are below the break-even transaction cost level, $T^{be}$, she will continue to prefer using our models; alternatively she would keep the RW. The value of $T^{be}$ is reported in monthly *bps* for $h = 1$, quarterly *bps* for $h = 3$, and annual *bps* for $h = 12$.

## 3.5 Empirical Results

We begin by examining the statistical and economic performance of the models that allow predictors and coefficients to change over time, as well as their restricted versions. Specifically, we report results from:

- BMA including Time-varying Coefficients (BMA incl. TVar-Coeffs): Based on Bayesian model averaging over individual models and with varying degrees of coefficient adaptivity.

- BMA excluding Time-varying Coefficients (BMA excl. TVar-Coeffs): This is a restricted version of the above, as it is based on BMA over individual models that exclude time-variation in coefficients. It corresponds to conventional BMA.

- BMS including Time-varying Coefficients (BMS incl. TVar-Coeffs): This is determined by the individual models that receive the highest posterior probability, among all individual models and with varying degrees of coefficient variation.

- BMS excluding Time-varying Coefficients (BMS excl. TVar-Coeffs): This specification is nested in the BMS including TVar-Coeffs model. It includes the individual models that receive the highest posterior probability, among all individual models excluding time-variation in coefficients.

- Single Predictor including Time-varying Coefficients (Single Predictor and BMA incl. TVar-Coeffs): These models consider only a single predictor at a time, but average over the range of all degrees of time-variation in coefficients.

- Single Predictor excluding Time-varying Coefficients (Single Predictor excl. TVar-Coeffs): This is a restricted version of the Single Predictor including TVar-Coeffs model. It includes only one predictor at a time in a setting excluding time-variation in coefficients.

While the above models are based on Bayesian methods, we also consider forecast combination methods based on frequentist approaches. In this case the forecast combination of $\Delta s_{t+h}$ made at time $t$, is a weighted average of the $k$ individual models' forecast based on OLS estimates of simple linear regressions, excluding time-varying coefficients. That is,

$$\widehat{\Delta s}^{c}_{t+h} = \sum_{i=1}^{k} \omega_{i,t} \widehat{\Delta s}^{i}_{t+h}, \tag{3.18}$$

where $\{\omega_{i,t}\}_{i=1}^{k}$ are the ex-ante combining weights formed at time $t$. Following Stock and Watson (2004) and Rapach et al. (2010) we consider the following combination methods:

- Mean Combination: The combined forecasts are obtained by using the following constant weighting scheme: $\omega_{i,t} = 1/k$, for $i = 1, ..., k$ in Eq. (3.18).

- Median Combination: The median combination forecasts is the median of $\{\widehat{\Delta s}^{i}_{t+h}\}_{i=1}^{k}$.

- Trimmed Mean Combination: The combined forecasts are obtained by setting $\omega_{i,t} = 0$ for the smallest and largest individual forecasts, and $\omega_{i,t} = 1/(k-2)$ for the remaining forecasts in Eq. (3.18). As in the median combination and the DMSPE combination method below, the weights change over time.

- DMSPE Combination: In this method, the weights are related to the historical forecasting performance of the individual models in a holdout-out-of-sample period. The discount mean squared prediction error (DMSPE) method uses the following weights: $\omega_{i,t} = \Phi_{i,t}^{-1} / \sum_{i=1}^{k} \Phi_{i,t}^{-1}$, where $\Phi_{i,t} = \sum_{so}^{t-1} \vartheta^{t-1-so} (\Delta s_{so+h} - \widehat{\Delta s}_{so+h}^{i})^2$ and $so$ is the end of the in-sample portion. The parameter $\vartheta$ denotes the discount factor applied to the mean squared prediction error. Based on results in Rapach et al. (2010) and Stock and Watson (2004), we set its value to 0.9.[11] This is consistent with attaching greater weight to the individual models that performed better in the holdout-out-of-sample period. We set this holdout-out-of-sample period to five years, implying that for this combination method, the forecast evaluation period starts five years later relative to that of the other models.

The reason we consider the restricted versions of the BMA including time-varying coefficients is to differentiate the influence of repeated updating of dynamic coefficients within a model, from the influence of the BMA that changes weights between models depending on past performance (Dangl and Halling, 2012). In contrast to BMA, BMS fixes a specific choice of predictors and degree of time-variation in coefficients. In addition, examining combined forecasts from models which exclude time-variation in coefficients, allows us to further check the sources of differences in performance and better understand the importance of time-variation in coefficients.

After establishing the main results in terms of the statistical and economic performance of the methods we consider, we then proceed and study in detail the characteristics of the BMA including time-varying coefficients. In this respect, we analyze the sources of prediction uncertainty, the degree of time-variation in coefficients consistent with the data, and which macroeconomic fundamentals are highly informative about exchange rates movements.

### 3.5.1 Out-of-Sample Statistical Evaluation

Table 3.1 summarizes in three panels the results from the predictive regressions that allow predictors and coefficients to change over time, and all the restricted versions that take into account multiple predictors. Comparing the panels, regressions which allow for time-changing sets of predictors and varying degrees of coefficients adaptivity improve upon the RW at all, but 1-month forecasting horizon. In fact, as shown in Panel A, at 3- and 12-months horizons the relative RMSFE are below one for all currencies and they are mostly significant. In the case of the BMA including TVar-Coeffs, for example, the improvement in terms of the reduction in the RMSFE is of at least 0.7% and a maximum of 7.1% at $h = 3$. As the forecast horizon increases to 12-months, the gains raise to a minimum of 9.5% and a maximum of 24.2%. And for over half of the currencies, the differences in the RMSFE are statistically significant. None of the other competing models achieve these magnitudes of improvements at these horizons. At best, methods based on constant-coefficients in Panel B and C tend to predominantly

[11] In Rapach et al. (2010) and Stock and Watson (2004) the best forecasting performance is achieved with a discount factor of 0.9, in a set that includes 0.95 and 1.0.

Table 3.1: Statistical Evaluation of BMA (BMS) Including or Excluding TVar-Coeffs and Simple Forecast Combinations Methods

Panel A: Models that allow predictors and coefficients to change over time

| | h=1 | h=3 | h=12 | h=1 | h=3 | h=12 |
|---|---|---|---|---|---|---|
| | BMA incl. TVar-Coeffs | | | BMS incl. TVar-Coeffs | | |
| CAD | 1.010 | **0.968**** | **0.839**** | 1.006 | **0.964**** | **0.836**** |
| GBP | 1.011 | **0.929** | **0.763**** | **0.983*** | **0.891**** | **0.751**** |
| YEN | 1.018 | **0.993** | **0.887** | 1.008 | **0.987** | **0.883** |
| NOK | 1.009 | **0.956**** | **0.758***** | 1.001 | **0.952**** | **0.753***** |
| SEK | 1.007 | **0.963**** | **0.797*** | **0.996** | **0.960**** | **0.789*** |
| CHF | 1.016 | **0.969** | **0.818**** | 1.010 | **0.955** | **0.795**** |
| EUR | 1.004 | **0.942**** | **0.905** | 1.001 | **0.937***** | **0.897** |

Panel B: Models that allow predictors to change over time, excl. TVar-Coeffs

| | BMA excl. TVar-Coeffs | | | BMS excl. TVar-Coeffs | | |
|---|---|---|---|---|---|---|
| CAD | 1.009 | 1.006 | 1.030 | 1.009 | 1.004 | 1.021 |
| GBP | 1.010 | **0.968** | **0.838** | **0.988*** | **0.949** | **0.836** |
| YEN | 1.014 | 1.018 | **0.946** | 1.008 | 1.018 | **0.948** |
| NOK | 1.009 | 1.004 | **0.936** | 1.007 | 1.005 | **0.936** |
| SEK | 1.007 | **0.991** | **0.980** | 1.002 | **0.990** | **0.980** |
| CHF | 1.025 | 1.005 | **0.962** | 1.025 | **0.986** | **0.958** |
| EUR | 1.008 | **0.982** | 1.019 | 1.008 | **0.980** | 1.018 |

Panel C: Combined forecasts, excl. TVar-Coeffs

| | OLS-Mean Combination | | | OLS-Median Combination | | |
|---|---|---|---|---|---|---|
| CAD | 1.002 | 1.007 | 1.019 | 1.003 | 1.007 | 1.016 |
| GBP | 1.000 | **0.999** | **0.962** | 1.002 | 1.005 | **0.975** |
| YEN | 1.000 | **0.998** | **0.953** | 1.000 | 1.000 | **0.974** |
| NOK | 1.003 | 1.010 | 1.025 | 1.002 | 1.005 | 1.019 |
| SEK | 1.003 | 1.014 | 1.012 | 1.000 | **0.999** | **0.973** |
| CHF | 1.000 | 1.000 | **0.964**** | 1.000 | 1.000 | **0.965*** |
| EUR | **0.999** | **0.999** | **0.974*** | **0.996***** | **0.990***** | **0.954**** |
| | OLS-Trimmed Mean | | | OLS-DMSPE Combination | | |
| CAD | 1.002 | 1.006 | 1.015 | 1.002 | 1.004 | **0.989** |
| GBP | **0.999** | **0.997** | **0.949** | 1.002 | **0.998** | **0.948** |
| YEN | 1.000 | **0.997** | **0.961** | 1.000 | **0.995** | **0.924*** |
| NOK | 1.002 | 1.006 | 1.016 | 1.002 | 1.007 | **0.999** |
| SEK | 1.000 | 1.002 | **0.979** | 1.000 | **0.994** | **0.939** |
| CHF | **0.999** | **0.997** | **0.963*** | **0.997**** | **0.988***** | **0.929***** |
| EUR | **0.997**** | **0.994** | **0.964**** | **0.997**** | **0.991***** | **0.926***** |

**Notes:** Relative RMSFE of the BMA (BMS) including or excluding time-varying coefficients and simple forecast combinations methods. For all methods, the driftless Random Walk (RW) constitutes the benchmark model. Hence, values below one indicate that the method under scrutiny generates a lower RMSFE than RW. The Table also reports the DMW test-statistic, with p-values based on a data-mining robust semi-parametric bootstrap. Asterisks (*10%, **5%, ***1%) indicate the level of significance at which the null hypothesis of equal RMSFE is rejected, favouring the alternative that the fundamental-based method yields a lower RMSFE than the RW. Currency codes denote: CAD - Canadian Dollar; GBP - Pound Sterling; YEN - Japanese Yen; NOK - Norwegian Krone; SEK - Swedish Krona; CHF - Swiss Franc; EUR - Euro. The forecast evaluation period begins in 1987M12+h in all, but the OLS-DMSPE Combination case (1992M12+h).

outperform the RW at the longest horizon; but the reduction in the RMSFE never exceeds 8% in general. This is the case, for instance, for the BMA and BMS excluding TVar-Coeffs and for the forecast combination methods. Furthermore, the differences in RMSFEs are rarely significant at the critical values implied by our data-mining robust semi-parametric bootstrap. Finally, we note that at 1-month horizon, these methods generally fail to outperform the RW benchmark.[12,13]

Table 3.2 presents results for Bayesian models based on Single Predictors including TVar-Coeffs as well as Single Predictors excluding TVar-Coeffs. The Table reveals that the above reported performance of the regressions that allow predictors and coefficients to change over time, stem from the individual models with time-varying coefficients. In Panel A, and at $h = [3, 12]$, virtually all the individual models with time-varying coefficients generate a smaller RMSFE than the RW for all currencies - though not always significant. By contrast, in Panel B not all the individual models excluding TVar-Coeffs improve upon the RW at these horizons; and this appears to be impacting upon the average forecasting performance of the BMA and BMS excluding TVar-Coeffs. Furthermore, the magnitude of reductions in the forecast error relative to the RW is smaller in the constant-coefficient models than in the models with time-varying coefficients. Overall, the statistical evidence points to the beneficial effects of the flexibility of models with varying degrees of time-variation in coefficients in terms of improving the out-of-sample forecasting performance of exchange rate models. We next examine whether these statistical gains yield concrete economic gains.

### 3.5.2 Out-of-Sample Economic Evaluation

Following our exposition, the economic evaluation of our models builds upon the maximum expected return strategy, which is often used in active currency management (Della Corte et al., 2012 and Li et al., 2015). We recall that we focus on four measures, namely, the Sharpe ratio ($SR$), the performance fee ($pf$), the excess premium return ($pr$), and the break-even transaction cost ($T^{be}$). In line with results in Li et al. (2015) our investor targets an annualized volatility of $\sigma_p^* = 10\%$ and her degree of relative risk aversion is $RRA = 6$.[14]

Table 4.3 presents results from portfolio allocation schemes based on forecasts from each of the methods we examine and the RW. At a glance, results are congruent with those from the statistical evaluation. At horizons greater than 1-month, strategies

[12]We also experimented other forecasting horizons (i.e., h=6, 24 and 36 months) and found that models with time-varying coefficients still forecast better than the RW for all exchange rates. We do not report these results to save space.

[13]Our results on the performance of BMA excluding TVar-Coeffs are coherent with those from Wright (2008). He finds that in a setting of tighter priors and shrinkage towards the null of no predictability, BMA excluding TVar-coeffs improves upon the RW, although the improvement in terms of reduction in the RMSFE is small. However, with loose priors and less shrinkage, it fails to improve upon the RW.

[14]Li et al. (2015) also experiment with different values of $\sigma_p^*$ and $RRA$, and find that results are insensitive to changes in these parameters.

Table 3.2: Statistical Evaluation of Single-Predictor Models Including or Excluding TVar-Coeffs

Panel A: Single Predictor and BMA incl.TVar-Coeffs

| | CAD | GBP | YEN | NOK | SEK | CHF | EUR |
|---|---|---|---|---|---|---|---|
| | | | | h=1 | | | |
| TRsy | 1.009 | 1.011 | 1.018 | 1.009 | 1.007 | 1.004 | 1.003 |
| TRasy | 1.015 | 1.025 | 1.023 | 1.013 | 1.012 | 1.019 | 1.013 |
| PPP | 1.010 | 1.000 | 1.011 | 1.003 | 1.008 | 1.008 | 1.008 |
| UIRP | 1.010 | 1.017 | 1.016 | 1.014 | 1.003 | 1.019 | 1.022 |
| MM | 1.012 | 1.029 | 1.020 | 1.016 | 1.014 | 1.017 | 1.019 |
| | | | | h=3 | | | |
| TRsy | **0.968**** | **0.935** | **0.991** | **0.955**** | **0.961**** | **0.948**** | **0.932**** |
| TRasy | **0.990** | **0.981** | **0.999** | **0.997** | **0.992** | 1.005 | 1.000 |
| PPP | **0.980** | **0.960** | **0.987** | **0.957**** | **0.975**** | **0.988***** | **0.953**** |
| UIRP | **0.984** | **0.991** | **0.990** | **0.989** | **0.985** | 1.007 | 1.006 |
| MM | 1.002 | 1.009 | **0.993** | **0.980** | **0.987** | **0.972** | **0.979** |
| | | | | h=12 | | | |
| TRsy | **0.759**** | **0.745**** | **0.797***** | **0.762***** | **0.706**** | **0.761***** | **0.769***** |
| TRasy | **0.895** | **0.979** | **0.931** | **0.959** | **0.932** | **0.945** | **0.905** |
| PPP | **0.835**** | **0.746**** | **0.797***** | **0.758***** | **0.845**** | **0.904***** | **0.756***** |
| UIRP | **0.886** | **0.937** | **0.814**** | **0.899** | 1.313 | **0.890** | **0.869** |
| MM | **0.868** | **0.842** | **0.881** | **0.908** | **0.797*** | **0.811**** | **0.868*** |

Panel B: Single Predictor excl. TVar-Coeffs

| | CAD | GBP | YEN | NOK | SEK | CHF | EUR |
|---|---|---|---|---|---|---|---|
| | | | | h=1 | | | |
| TRsy | 1.007 | 1.010 | 1.014 | 1.008 | 1.005 | 1.002 | 1.006 |
| TRasy | 1.006 | 1.026 | 1.019 | 1.011 | 1.008 | 1.016 | 1.009 |
| PPP | 1.008 | 1.000 | 1.010 | 1.002 | 1.006 | 1.007 | 1.007 |
| UIRP | 1.009 | 1.016 | 1.013 | 1.012 | 1.006 | 1.017 | 1.021 |
| MM | 1.012 | 1.027 | 1.018 | 1.015 | 1.012 | 1.025 | 1.021 |
| | | | | h=3 | | | |
| TRsy | 1.006 | **0.956** | **0.995** | **0.997** | **0.989** | **0.984** | **0.980** |
| TRasy | 1.001 | **0.992** | 1.003 | 1.009 | **0.996** | 1.009 | 1.019 |
| PPP | 1.004 | **0.968** | **0.991** | **0.991** | **0.983** | **0.991***** | 1.004 |
| UIRP | 1.010 | 1.039 | **0.992** | 1.012 | **0.996** | 1.014 | 1.016 |
| MM | 1.012 | 1.016 | 1.018 | 1.021 | 1.011 | 1.026 | 1.015 |
| | | | | h=12 | | | |
| TRsy | **0.952** | **0.817** | **0.901** | **0.922** | **0.822** | **0.879*** | **0.879**** |
| TRasy | **0.984** | **0.990** | 1.010 | 1.002 | **0.978** | **0.966** | 1.018 |
| PPP | 1.002 | **0.822** | **0.910** | **0.936** | **0.885***** | **0.928***** | **0.887**** |
| UIRP | **0.976** | 1.066 | **0.906** | 1.003 | 1.140 | **0.960** | **0.995** |
| MM | 1.021 | **0.985** | **0.976** | **0.973** | **0.980** | **0.964** | **0.978** |

**Notes:** RMSFE of the single-predictor models including or excluding time-varying coefficients, relative to the RMSFE of the driftless Random Walk (RW). Values below one indicate that the model under under scrutiny generates a lower RMSFE than the RW. The description of the predictors is as follows: TRsy and TRasy correspond to the symmetric and asymmetric Taylor rules, respectively; MM- fundamentals from the Monetary Model, PPP - Purchasing Power Parity; and UIRP- Uncovered Interest Rate Parity. Asterisks (*10%, **5%, ***1%) denote the level of significance of the DMW test based on a semi-parametric bootstrap, for the null hypothesis of equality in the RMSFE. The forecast evaluation period begins in 1987M12+h.

Table 3.3: Economic Evaluation of BMA (BMS) Including or Excluding TVar-Coeffs and Simple Forecast Combinations Methods

| Exchange rate forecast based on: | $\overline{Rt}(\%)$ | Vol (%) | SR | $pf(bps)$ | $pr(bps)$ | $T^{be}(bps)$ |
|---|---|---|---|---|---|---|
| | monthly rebalancing period (i.e., h=1) | | | | | |
| Random Walk | 11.02 | 9.77 | 0.72 | | | |
| BMA incl. TVar-Coeff. | 3.14 | 9.95 | -0.09 | -799 | -798 | - |
| BMS incl. TVar-Coeff. | 9.04 | 9.93 | 0.50 | -208 | -201 | - |
| BMA excl. TVar-Coeff. | 2.82 | 8.84 | -0.14 | -765 | -772 | - |
| BMS excl. TVar-Coeff. | 4.96 | 8.45 | 0.11 | -530 | -531 | - |
| OLS-Mean Combination | 9.24 | 10.25 | 0.51 | -208 | -196 | - |
| OLS-Median Combination | 9.72 | 9.95 | 0.57 | -141 | -127 | - |
| OLS-Trimmed Mean | 9.42 | 10.26 | 0.53 | -191 | -177 | - |
| OLS-DMSPE Combination | 7.99 | 11.12 | 0.42 | -357 | -358 | - |
| | 3-months rebalacing period (i.e., h=3) | | | | | |
| Random Walk | 12.27 | 10.11 | 0.80 | | | |
| BMA incl. TVar-Coeff. | 12.83 | 10.50 | 0.83 | 26 | 51 | 4 |
| BMS incl. TVar-Coeff. | 13.50 | 10.43 | 0.90* | 99 | 120 | 15 |
| BMA excl. TVar-Coeff. | 9.60 | 9.40 | 0.58 | -216 | -214 | - |
| BMS excl. TVar-Coeff. | 10.01 | 9.48 | 0.62 | -181 | -177 | - |
| OLS-Mean Combination | 9.59 | 10.97 | 0.50 | -334 | -332 | - |
| OLS-Median Combination | 11.23 | 10.33 | 0.68 | -120 | -96 | - |
| OLS-Trimmed Mean | 10.56 | 10.44 | 0.61 | -196 | -168 | - |
| OLS-DMSPE Combination | 8.83 | 11.56 | 0.47 | -343 | -324 | - |
| | 12-months rebalancing period (i.e., h=12) | | | | | |
| Random Walk | 15.45 | 12.57 | 0.86 | | | |
| BMA incl. TVar-Coeff. | 20.65 | 15.80 | 1.02* | 399 | 304 | 64 |
| BMS incl. TVar-Coeff. | 21.29 | 15.92 | 1.05 | 462 | 372 | 73 |
| BMA excl. TVar-Coeff. | 11.96 | 19.06 | 0.39 | -471 | -785 | - |
| BMS excl. TVar-Coeff. | 12.23 | 19.10 | 0.40 | -444 | -767 | - |
| OLS-Mean Combination | 13.48 | 13.60 | 0.65 | -319 | -608 | - |
| OLS-Median Combination | 14.57 | 13.63 | 0.73 | -209 | -110 | - |
| OLS-Trimmed Mean | 14.37 | 13.03 | 0.75 | -230 | -84 | - |
| OLS-DMSPE Combination | 15.82 | 14.76 | 0.81 | -84 | -40 | - |

**Notes:** Economic gains generated by BMA (BMS) including or excluding time-varying coefficients and simple forecast combinations methods. Using forecasts from these methods an investor constructs a strategy of maximum expected return, conditional on portfolio volatility target of Vol = 10%. Every $h$-month (s) period, she dynamically adjusts her portfolio investing in US bonds and seven foreign bonds. The Table shows the gains obtained by the investor by conditioning on the forecasts from each method and the driftless Randow Walk (RW) at every rebalancing period. The gains are gauged based on: (i) the annualized mean return- $\overline{Rt}$ (ii) Vol - the annualized volatility, (iii) SR - annualized Sharpe ratio, (iv) $pf$, the maximum performance fee a risk-averse investor with quadratic utility would be willing to pay to use the corresponding method instead of the RW - in annualized *bps* (v) *pr* - the excess premium return (in annualized *bps*) and (vi) $T^{be}$ - the break-even proportional transation costs that offset any positive performance fee obtained by using the method under consideration - in $h$-month(s) period *bps*. Asterisk denotes statistically significant differences in the SR in favor of the method in question relative to the driftless RW at the 10% level of significance based on the bootstrap procedure of Ledoit and Wolf (2008). The evaluation period is 1987M12+h to 2013M5.

conditioning on forecasts from BMA or BMS including TVar-Coeffs yield economic gains above those accruing from the RW, regardless of the specific economic indicator. For instance, the Sharpe ratio implied by BMA including TVar-Coeffs is 0.83 at $h = 3$, slightly higher than that of the RW at 0.80. As the rebalancing horizon extends to 12-months period, the gains become salient and significant, scoring a Sharpe ratio of 1.02, against 0.86 of the RW benchmark. None of the constant-coefficients forecasting models provide gains superior to the ones obtained from using the RW at these horizons. (Note too that at 1-month rebalancing period, the RW delivers the best outcomes).

Other indicators of economic value also convey the beneficial effects of conditioning on the more flexible models at longer rebalancing periods. At $h = 12$ for example, the performance fee is $pf = 462bps$ when using BMS including TVar-Coeffs, which implies that a risk-averse investor is ready to pay a fee of 4.62% annually to use this forecasting approach rather than the RW. He also obtains an excess premium return of 3.72% annually, whilst the break-even transaction costs are $T^{be} = 73bps$. If the investor's proportional transaction costs are higher than this magnitude of $T^{be}$, he will continue using the RW. However, as Della Corte et al. (2012) point out this is unlikely, as transaction costs in the currency markets are very low.

### 3.5.3 Characterization of BMA with Time-Varying Coefficients

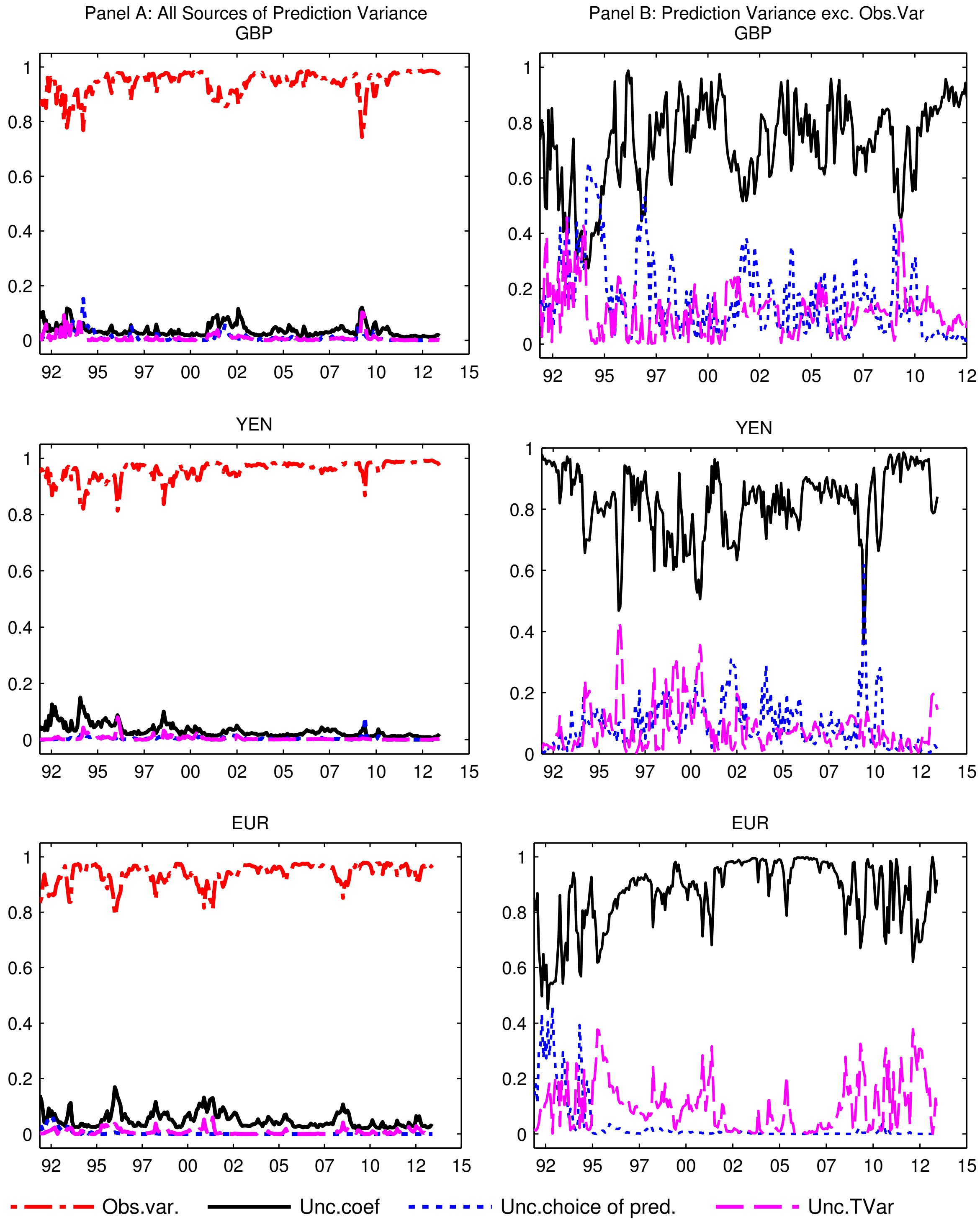
Our results thus far imply that Bayesian averaging or selection over individual models with varying degrees of coefficients evolution lead to forecast improvements and economic gains, particularly at horizons beyond 1-month. Since this model emanates from a complex combination of many individual models, understanding its characteristics is useful in explaining the sources of difference in forecasting performance relative to other competing models and across forecasting horizons. This constitutes a key contribution of this chapter.

#### Sources of Prediction Uncertainty

We begin by analyzing the sources of prediction uncertainty through a variance decomposition process. As noted in our methodological section, we decompose the total variance into observational variance, variance due to errors in the estimation of the coefficients, variance due to model uncertainty with respect to the choice of the predictor and variance due to model uncertainty with respect to the choice of degree of time-variation in coefficients.

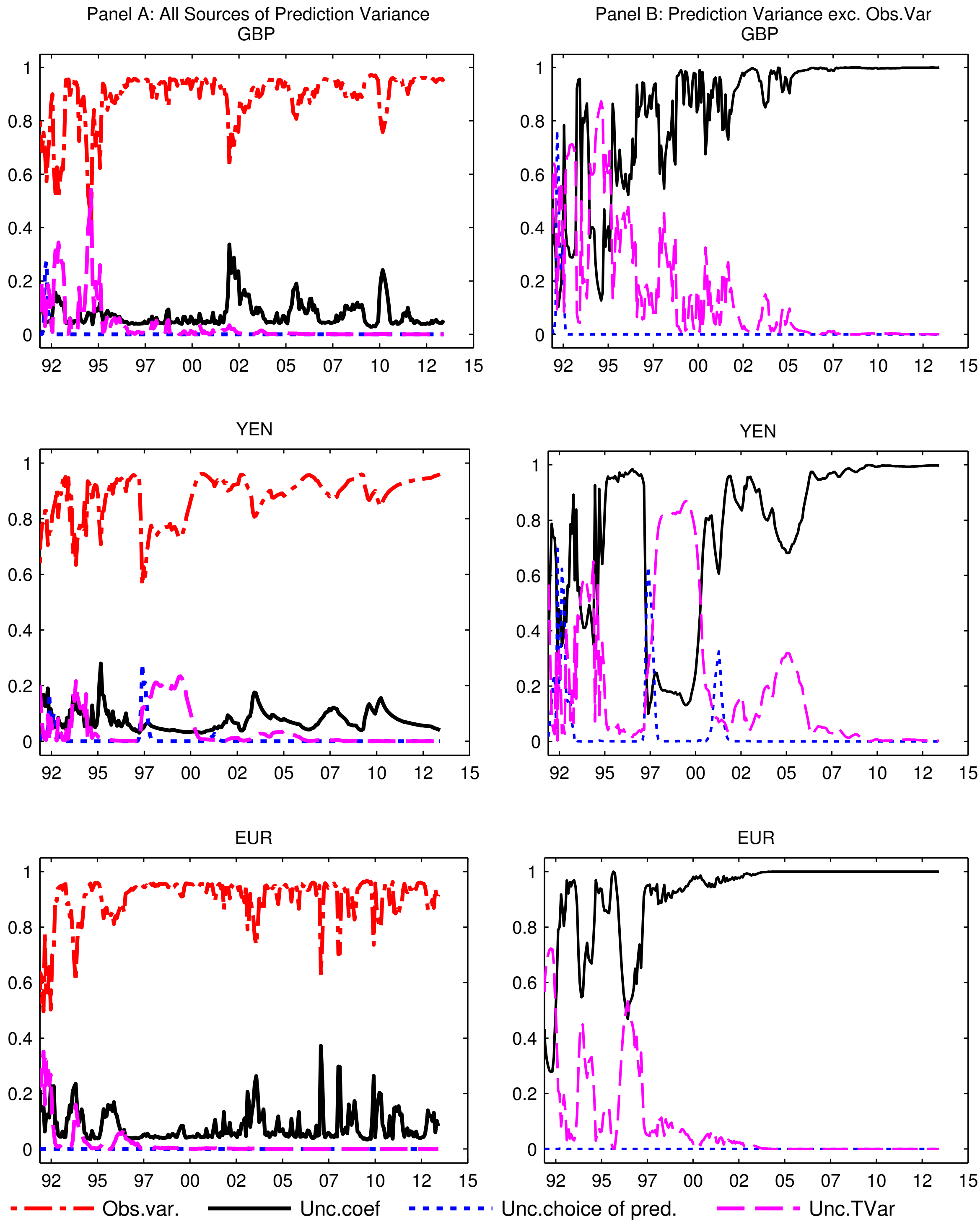
Focusing on a representative selection of three countries, Figures 3.1 and 3.2 depict this decomposition for the 1-month and 12-months horizons, respectively. The figure for three-months horizon resembles the pattern shown for twelve months horizon in Figure 3.2, hence, is omitted to conserve space. In both figures, Panel A illustrates the relative weight of each of the components of prediction variance in the total variance. In all cases, the predominant source of uncertainty is observational variance. As Dangl

Figure 3.1: Sources of Prediction Variance at 1-Month Forecasting Horizon




Notes: Decomposition of the prediction variance into its constituent parts at 1-month forecasting horizon. Panel A shows all sources of prediction variance: (i) the variance caused by random fluctuations in the data (Obs.var.); (ii) variance due to errors in the estimation of the coefficients (Unc.coef); (iii) variance due to model uncertainty with respect to the choice of the predictor (Unc.choice of pred); and (iv) variance due to model uncertainty with respect to the choice of degree of time-variation in coefficients (Unc.TVar). The Panel shows relative proportions of these variances. Panel B excludes the variance due to random fluctuations in the data (Obs.var.) and shows the relative weights of the remaining sources of prediction variance, and hence also sum to one.

Figure 3.2: Sources of Prediction Variance at 12-Months Forecasting Horizon




Notes: As in Figure 3.1, except that here the forecasting horizon is 12-months.

and Halling (2012) point out this is normal for asset prices, as they frequently fluctuate randomly over their expected values. These fluctuations are expected to be noticeable for the horizons we are considering and to dominate the predicted trend component.

In Panel B of the same figures we exclude the observational variance allowing us to focus upon the relative weights of the remaining sources of prediction uncertainty. At both forecasting horizons, the variance from errors in the estimation of the coefficients is now the dominant source of prediction uncertainty. In this sense, estimation uncertainty is one of the main factors hindering model forecasting performance.

Between the two forecasting horizons however, there are differences with respect to the remaining sources of prediction variance. At 1-month horizon, the uncertainty about which predictor is more informative about changes in exchange rates and, most notably, the uncertainty regarding the correct degree of time-variation in the regressions coefficients are detectable throughout the forecast sample. For example, in the case of GBP and in various periods, each of these sources of uncertainty represents over one tenth of the total variance excluding observational variance, peaking in periods around financial stresses, such as the 2008 financial crisis. In contrast, at 12-months horizon they are clustered at the beginning of the out-of-sample period, which corresponds to the initial data-points in the expanding window of the forecasting procedure. As more evidence is accumulated, they remain low for the most part of the forecasting window. We note though that while the uncertainty about the choice of the relevant predictor is apparently non-negligible in the cases of the GBP and the YEN at 1-month horizon, this is not critically influencing forecasting performance. As we will illustrate when examining the importance of individual predictors, the reason is that for the most part of the forecast window, effectively one predictor is selected and used in the regression. The switches in the predictors which are influencing the variance occur between a maximum of two regressors and they are largely infrequent.

We interpret these findings as suggesting that although estimation uncertainty plays a role at both horizons, at 1-month horizon our flexible models fail to improve upon the RW because of the additional uncertainty regarding the precise level of time-variation in coefficients, necessary to capture instabilities present in the data. Put differently, there is no certainty about the exact degree of time-variation in coefficients to embed in the model, in order to offset the loss in forecasting performance emanating from estimation uncertainty. By contrast, as the forecast horizon increases our models successfully embed the level of time-variation in coefficients present in the data. Consequently, they consistently outperform the RW by counterbalancing the loss in the precision in coefficient estimation, with increased variability in the coefficients. This signifies that both, estimation uncertainty and coefficient instability obstruct model forecasting performance, and our BMA and BMS including TVar-Coeffs adapt to the pattern in the data at longer forecasting horizons.

We relate these findings to Bacchetta et al. (2010) and Giannone (2010). Bacchetta et al. (2010) calibrate a theoretical reduced-form model of the exchange rate on actual data to examine whether parameter instability rationalizes the Meese and Rogoff (1983) result of exchange rate unpredictability. They find that estimation uncertainty is the main factor that hinders exchange model's forecasting performance and not time-variation in coefficients. But Giannone (2010) disputes these findings and shows that both, estimation uncertainty and parameter instability, are relevant in explaining the Meese-Rogoff puzzle.

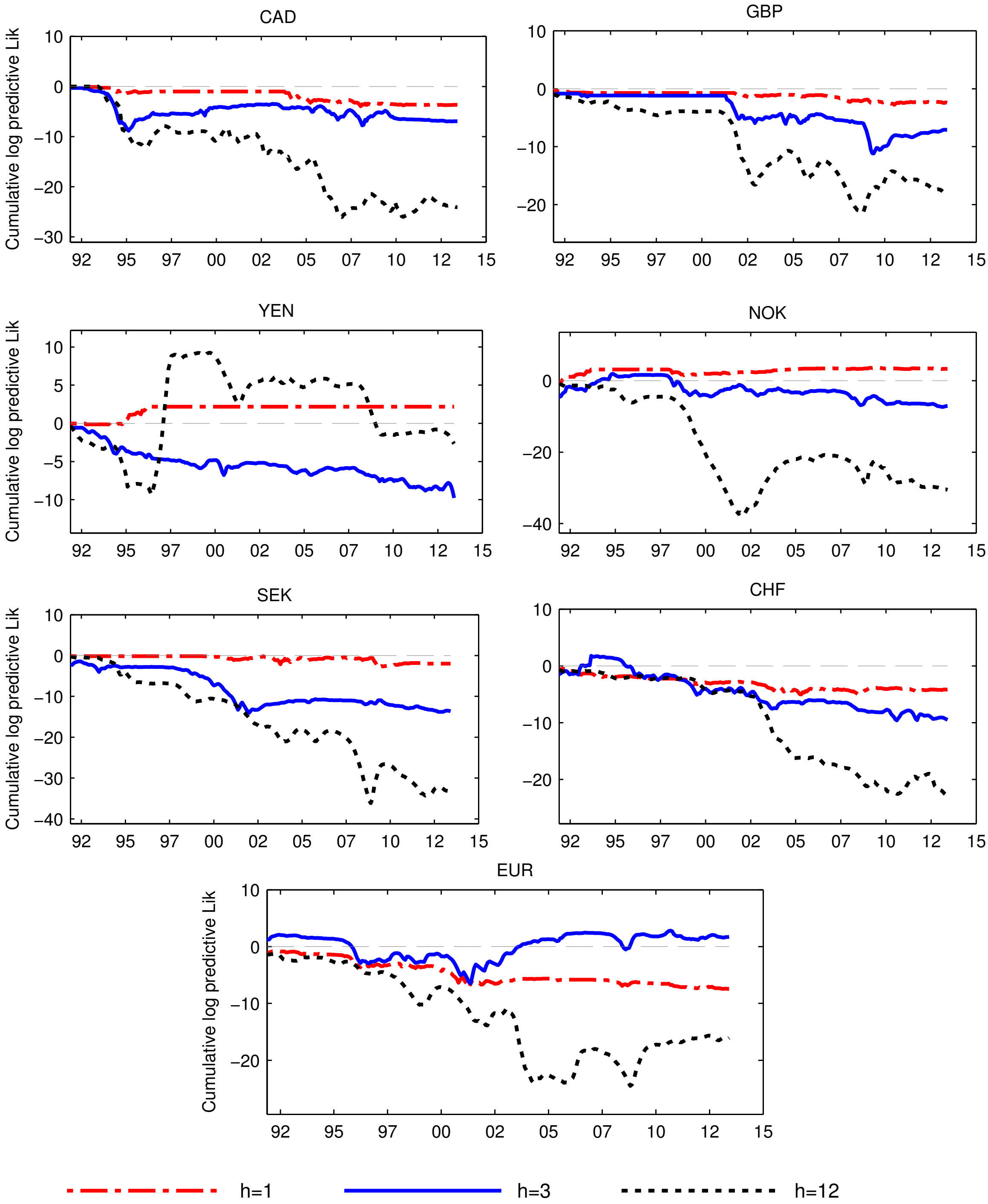
Giannone (2010) also examines the trade-offs between estimation error and parameter uncertainty. He observes three stages when forecasting in an expanding window of data. The first stage is characterized by a high forecast error with the first few observations. In the second stage, the forecast accuracy increases as the estimation window is expanded beyond these few initial observations, signalling reduction in the coefficients' estimation error. In the third stage, however, further increasing the window deteriorates model-based forecasting performance relative to the RW, since gains from reduced estimation error are compensated by losses due to the presence of structural instabilities. Thus, the recursive ratio of the relative RMSFE exhibits a $U$ shape. To conserve space we omit the figures of our recursive relative RMSFE. The main fact is our confirmation of the Giannone's (2010) observations for the first and second stages, but not the third stage. In our case, further increasing the window does not significantly deteriorate model-based forecasting performance relative to the RW. In fact, for the most part of the forecast period and horizons greater than 1-month the relative RMSFE is below one, favouring our flexible models. When read in conjunction with the main sources of instabilities we detect, this reinforces our conclusions that BMA and BMS including TVar-Coeffs successfully capture the degree of time-variation in parameters necessary to offset the loss in forecast accuracy due to estimation uncertainty.

To add up to these results we focus in another measure of forecast accuracy that underlies our Bayesian approach, namely the predictive likelihoods, see Geweke and Amisano (2010). Figure 3.3 depicts the cumulative log predictive likelihoods for models with constant-coefficients relative to the ones with time-varying coefficients. A value of zero corresponds to equal marginal support for both models; negative values are in support of models with time-varying coefficients; and positive values are in favour of models with constant-coefficients.

Two main results are apparent in Figure 3.3. First, it confirms that models with time-varying coefficients are empirically plausible, especially at 3- and 12-months horizons. At these horizons and with the exception of the YEN and the EUR, the cumulative log predictive likelihoods become negative after a number of out-of-sample data-points have been accumulated.[15] These cumulative predictive likelihoods show a downward trend, consistent with additive evidence favouring models with time-varying coefficients. Second, observations around the 2008 financial crisis, where significant

[15]Geweke and Amisano (2010) point out that it is customary for the results to be sensitive at the beginning of the out-of-sample period, as this reflects sensitiveness to the prior density. As they emphasize, nonetheless, after a number of observations have been accumulated the results become invariant to substantial changes in the prior density distribution.

Figure 3.3: Cumulative Log Predictive Likelihoods: BMS excl. TVar-Coeffs/BMS incl. TVar-Coeffs




Notes: Cumulative log predictive likelihoods of the BMS excl. TVar-Coeffs relative to the BMS incl. TVar-Coeffs. A value of zero corresponds to equal marginal support for both models; negative values are in support of models with time-varying coefficients and positive values are in favour of models with constant coefficients.

shifts in economic conditions occurred, contribute highly to the evidence in favor of the time-varying coefficients models. Overall, our findings remain invariant to this measure of forecast accuracy and they further rule out the $U$ pattern reported in Giannone (2010).

### Analysis of the Degree of Time-Variation in Coefficients

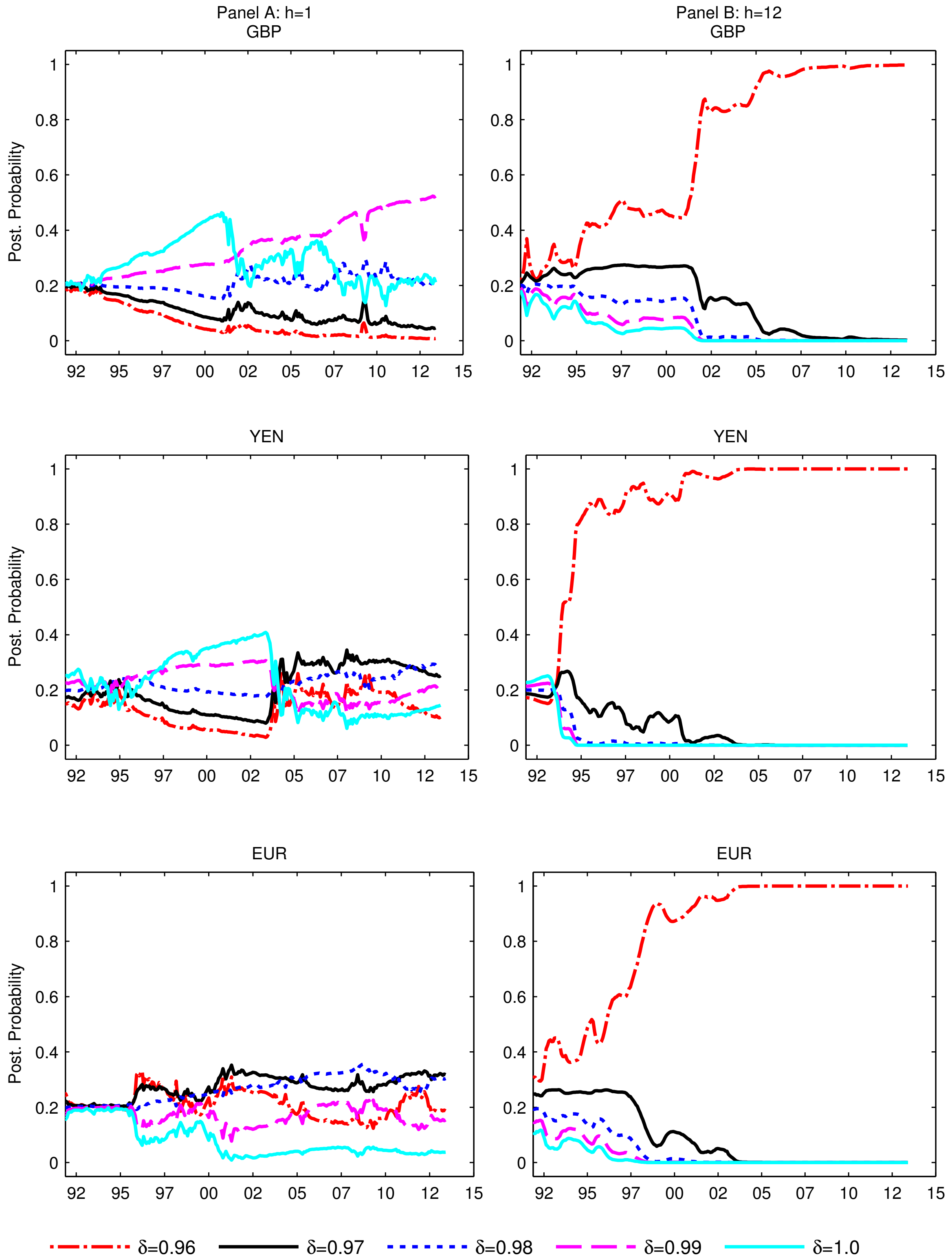
The preceding results indicate that the uncertainty regarding the degree of time-variation in parameters is not trivial at 1-month horizon, while at longer forecasting horizons it becomes low as more evidence is gathered. But the precise amount of time-variation was not referred. Figure 3.4 provides this information focusing on the same representative set of currencies. It depicts the total posterior probability of each of the support points for time-variation in coefficients, $\delta$. At 1-month horizon in Panel A, most of the support points for time-variation in coefficients are fairly likely, as reflected in the magnitude of their weights over the out-of-sample window. In the case of the EUR and the YEN, for example, models with moderate degree of time-variations in coefficients ($\delta = [0.99, 0.98]$) are, on average, as likely as the ones with high degree of time-variation ($\delta = [0.96, 0.97]$), both with probability varying around 20%. For the GBP, the weight vary between models with constant ($\delta = 1$) to moderate degree of time-variation in coefficients. This switch between models that are all equally supported by the data is reflected in the relatively high uncertainty about the correct degree of time-variation in coefficients at 1-month forecasting horizon.

In contrast, at 12-months horizon in Panel B, very dynamic models with $\delta = 0.96$, exhibit the highest weight (posterior probability), while constant-coefficient models typically receive the lowest. The preponderance of $\delta = 0.96$ over the other support points, is mirrored in the low uncertainty with respect to the degree of time-variation in coefficients at this horizon. Interestingly, Giannone (2010) finds that to match the pattern of exchange rate unpredictability present in the data, a significant amount of time-variation in coefficients were necessary in his simulations. Beckmann and Schuessler (2015) show in a Monte Carlo Simulation that a time-varying parameter model like ours, i.e., which allows for gradual to high degree of time-variation in coefficients, is well suited to recover the patterns in the data. As well, in an application to equity returns, Dangl and Halling (2012) find that models with moderate ($\delta = 0.98$) to high ($\delta = 0.96$) degree of time-variation are empirically supported.

### Analysis of the Importance of Individual Predictors

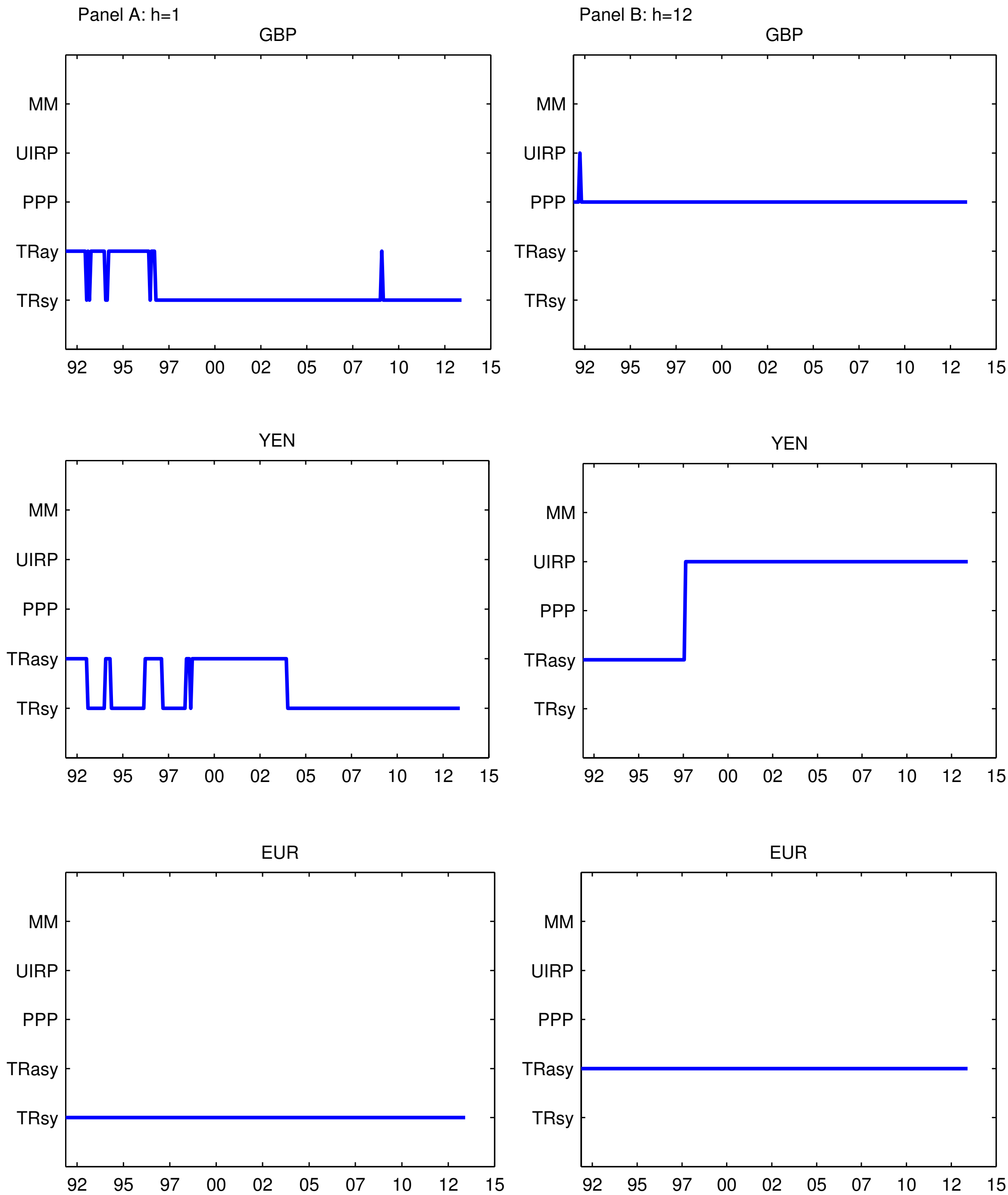
Another characteristic to explore in our flexible models is the importance of individual predictors. Figure 3.5 shows which predictors accumulate the highest probability at each point in time. Within each horizon-currency, there is only one or two predictors that are highly informative about movements in that currency exchange rate. But for the same currency, the relevant predictors vary over different horizons. In the case

Figure 3.4: Posterior Probabilities of Degrees of Time-variation in Coefficients




Notes: Posterior probabilities of values of $\delta$ (support points for time-variation in coefficients) for a representative selection of countries. Panel A is for 1-month forecasting horizon and Panel B for 12-months. These are the weights employed to produce the average forecasts in the BMA including TVar-Coeffs.

Figure 3.5: Predictors with the Largest Posterior Probability at Each Period




Notes: Predictors with the highest probability at each point in time for a representative selection of countries. Panel A is for 1-month forecasting horizon and Panel B for 12-months. The forecasts from the BMS including TVar-Coeffs are based on these predictors.

of the GBP for instance, while fundamentals from the symmetric Taylor rule (TRsy) are dominant at 1-month horizon, those based on PPP are more relevant in explaining variations in the USD/GBP rate at 12-months horizon. The narrow number of empirically plausible predictors within each horizon-currency implies smaller sensitivity of our flexible models to change in the predictors. Moreover, the swings among the smaller set of selected predictors are mostly few and confined to certain periods.

An additional fact from observing Figure 3.5 pertains to the most likely predictors across horizons and countries. At 1-month horizon fundamentals implied by Taylor rules appear to be more important in accounting for the path of the exchange rates. As shown, in all cases either the symmetric (TRsy) or the asymmetric (TRasy) rules attract the largest posterior probabilities. As the forecast horizon increases, parity conditions such as PPP and UIRP become pertinent, whilst the monetary model rarely receives the highest weight.

## 3.6 Robustness Checks

We verified the robustness of the empirical findings in the previous Section in three dimensions. First, although the cumulative log predictive likelihoods allowed us to examine the forecast accuracy over the entire path of the forecast window, we equally experimented with changing the beginning of the forecast window.[16,17] We considered an evaluation period starting in more recent periods (2004M12+h), and examined the statistical and economic performance of BMA including and excluding time-varying coefficients. In Panel A of Tables 3.4 and 3.5 we present the results. Overall the choice of the forecast evaluation period does not materially affect our conclusion that models with time-varying coefficients yield larger forecast improvements and economic gains at horizons beyond 1-month. In the BMA including TVar-Coeffs, the relative RMSFE are below one in six of the seven currencies; and the magnitudes of the reductions in the forecast errors are greater than the ones from the BMA excluding TVar-Coeffs. In terms of economic value, BMA including TVar-Coeffs outperforms a strategy based on the RW at the longest horizon, though at 1- and 3-months is still preferable than BMA excluding TVar-Coeffs.

Second, we changed the base numeraire (home country) to the GBP (UK) following Chen et al. (2010). Accordingly, we redefined our predictors taking the UK as the home country. Focusing on the same models and measures of forecasting performance as above, results in Panel B of Tables 3.4 and 3.5 are congruent to our early findings. The corresponding analysis of prediction variance, based on the example of the EUR in Figure 3.6, also reveals the pattern we documented: estimation uncertainty and uncertainty with regards to the exact degree of time-variation are the main obstacles to models forecasting performance.

Third, we estimated directly the degree of time-variation in coefficients following the approach in Koop and Korobilis (2013), instead of inferring from model's posterior probability. In this case the estimated $\delta$ is: $\widehat{\delta} = \delta_{Min} + (1 - \delta_{Min})exp(L_g \times v_{t+h}^2)$, where

[16]Giacomini and Rossi (2010) formalize the issue of forecast robustness over different windows in the presence of instabilities by developing appropriate test-statistics. The asymptotics of their tests require a use of a rolling or fixed estimation window approach when producing the forecasts, rather than our recursive scheme.

[17]Due to computational costs of implementing the bootstrap for the $d.k.$ models and each sensitivity analysis we consider, in this Section we evaluate our forecasts solely on the basis of the relative RMSFE.

Table 3.4: Statistical Evaluation of Forecasting Performance in a Sensitivity Analysis

| | BMA incl. TVar-Coeffs | | | BMA excl. TVar-Coeffs | | |
|---|---|---|---|---|---|---|
| | h=1 | h=3 | h=12 | h=1 | h=3 | h=12 |
| Panel A: Change of the beginning of the forecast window to 2004M12+h | | | | | | |
| CAD | 1.005 | **0.967** | **0.719** | 1.004 | **0.981** | **0.811** |
| GBP | 1.035 | **0.842** | **0.801** | 1.032 | **0.854** | **0.935** |
| YEN | 1.032 | **0.990** | **0.893** | 1.033 | 1.014 | **0.976** |
| NOK | **0.999** | **0.931** | **0.667** | **0.999** | **0.935** | **0.727** |
| SEK | 1.012 | **0.943** | **0.817** | 1.006 | **0.941** | **0.897** |
| CHF | 1.025 | 1.003 | **0.896** | 1.017 | 1.008 | **0.936** |
| EUR | **0.982** | **0.952** | 1.001 | **0.982** | **0.965** | 1.036 |
| Panel B: Change in base currency to GBP | | | | | | |
| CAD | 1.021 | **0.922** | **0.728** | 1.020 | 1.003 | **0.954** |
| USD | 1.011 | **0.929** | **0.763** | 1.010 | **0.968** | **0.838** |
| YEN | 1.012 | **0.972** | **0.862** | 1.008 | 1.001 | **0.964** |
| NOK | 1.024 | **0.984** | **0.838** | 1.020 | 1.016 | 1.088 |
| SEK | 1.065 | **0.997** | **0.906** | 1.040 | 1.009 | **0.980** |
| CHF | **0.999** | **0.944** | **0.715** | **0.999** | **0.962** | **0.821** |
| EUR | 1.024 | **0.985** | **0.906** | 1.018 | 1.015 | 1.010 |

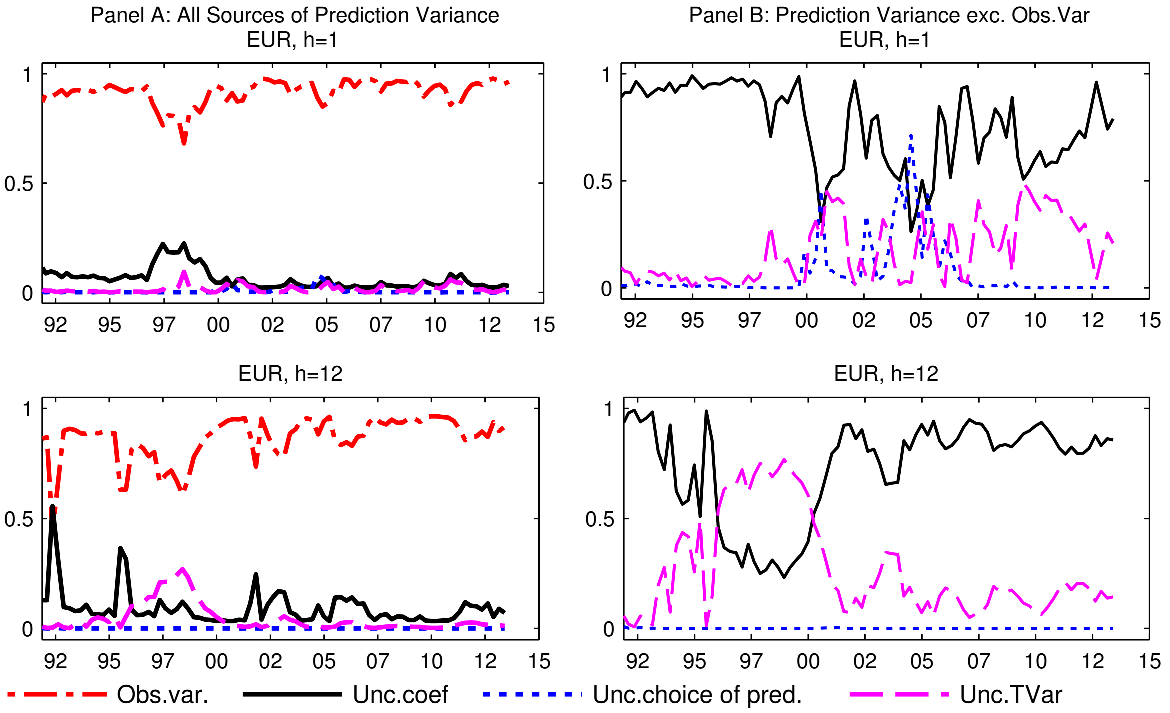
**Notes:** The entries in the Table are the RMSFE of BMA including or excluding TVar-Coeffs relative to the RMSFE of the driftless Random Walk (RW). Values below one indicate that the method under consideration generates a lower RMSFE than RW. Panel A reports the sensitivity to changing the beginning of the forecast evaluation period to 2004M12+h. Panel B reports the sensitivity to change in the base currency to the GBP.

Table 3.5: Economic Evaluation of Forecasting Performance in a Sensitivity Analysis

| | Random Walk | | | BMA incl. TVar-Coeffs | | | BMA excl. TVar-Coeffs | | |
|---|---|---|---|---|---|---|---|---|---|
| | h=1 | h=3 | h=12 | h=1 | h=3 | h=12 | h=1 | h=3 | h=12 |
| Panel A: Change of the beginning of the forecast window to 2004M12+h | | | | | | | | | |
| Rt (%) | 14.66 | 16.05 | 18.59 | 2.35 | 13.77 | 18.08 | 1.88 | 12.38 | 13.78 |
| Vol (%) | 10.90 | 9.42 | 11.20 | 11.93 | 11.37 | 10.61 | 11.45 | 10.28 | 9.63 |
| SR | 1.15 | 1.46 | 1.40 | 0.02 | 1.01 | 1.43 | -0.02 | 0.98 | 1.13 |
| pf (bps) | - | - | - | -1306 | -389 | 547 | -1317 | -433 | -313 |
| pr (bps) | - | - | - | -1290 | -305 | -25 | -1307 | -395 | -373 |
| Tbe (bps) | - | - | - | - | - | 111 | - | - | |
| Panel B: Change in base currency to GBP | | | | | | | | | |
| Rt (%) | 10.82 | 11.77 | 14.73 | 5.15 | 14.77 | 17.95 | 6.22 | 10.53 | 12.54 |
| Vol (%) | 10.38 | 11.00 | 11.19 | 10.84 | 11.67 | 12.70 | 10.52 | 10.10 | 12.79 |
| SR | 0.65 | 0.69 | 0.91 | 0.10 | 0.91* | 1.05* | 0.21 | 0.63 | 0.62 |
| pf (bps) | - | - | - | -598 | 244 | 128 | -470 | -55 | -412 |
| pr (bps) | - | - | - | -614 | 286 | 240 | -472 | -62 | -254 |
| Tbe (bps) | - | - | - | - | 43 | 41 | - | - | - |

**Notes:** The Table shows: (i) the annualized mean return- $\overline{Rt}$ (ii) Vol - the annualized volatility, (iii) SR - annualized Sharpe ratio, (iv) $pf$, the maximum performance fee a risk-averse investor with quadratic utility would be willing to pay to use the corresponding method instead of the RW - in annualized $bps$ (v) $pr$ - the excess premium return (in annualized $bps$) and (vi) $T^{be}$ - the break-even proportional transation costs that offset any positive performance fee obtained by using the method under consideration - in $h$-month period $bps$. Panel A reports the sensitivity to changing the beginning of the forecast evaluation period to 2004M12+h. Panel B reports the sensitivity to change in the base currency to the GBP.

Figure 3.6: Sources of Prediction Variance: Sensitivity to Change in the base Currency to the GBP




Notes: As in Figures 3.1 and 3.2, except that here the exchange rate is defined relative to the Pound Sterling, and hence all the data employed in the predictive regression are redefined relative to the UK.

Table 3.6: Degree of Time-Variation in Coefficients Inferred from the Predictive Regression

| | Quartlies | CAD | GBP | YEN | NOK | SEK | CHF | EUR |
|---|---|---|---|---|---|---|---|---|
| h=1 | First | 0.99 | 0.98 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |
| | Second | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| | Third | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| h=3 | First | 0.97 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 |
| | Second | 0.99 | 0.98 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |
| | Third | 1.00 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| h=12 | First | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 |
| | Second | 0.97 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 |
| | Third | 1.00 | 0.98 | 0.97 | 0.98 | 0.97 | 0.97 | 0.97 |

**Notes:** Estimates of the degree of time-variation in coefficients ($\delta$) based on Koop and Korobilis's (2013) method. For each country we first estimate $\hat{\delta}$ as we run the predictive model excluding time-variation in coefficients for each $t$. We then average the estimates over predictors and obtain a series of $\hat{\delta}$. From this series we compute the first, second and third quartiles. The estimates are obtained from the period begining in 1978M12+h, in line with the mechanics of our forecasting exercise.

$\delta_{Min}$ is the minimum value of support points for time-variation in coefficients considered (we set $\delta_{Min} = 0.96$), $exp$ is the exponential function, $L_g$ is a constant scaling parameter and $v_{t+h}$ is the predictive regression's error. Essentially, results are coherent with the degree of time-variation in coefficients ensuing from our flexible Bayesian approach. As shown in Table 3.6, at 1-month horizon the median ($2^{nd}$ quartile) estimate of $\widehat{\delta}$ is

0.99, and for over three quarters of the out-of-sample data-points its value is above 0.97 in most cases. This indicates that models with constant to moderate degree of time-variation in coefficients are empirically plausible at this horizon. As the forecasting horizon increases, the value of $\widehat{\delta}$ is consistent with more time-variation in coefficients, with a median value of $\widehat{\delta} = 0.97$ for the majority of currencies at 3-months horizon, and $\widehat{\delta} = 0.96$ at the longest horizon.

## 3.7 Conclusion

The exchange rate literature points out that the out-of-sample predictive power of fundamental-based exchange rate models is erratic. Models that forecast well for certain currencies and periods, often fail when applied to other exchange rates and samples (Rogoff and Stavrakeva, 2008; Rossi, 2013). While this signals presence of instabilities, attempts to account for them, for example by considering regressions with time-varying coefficients, have not yet produced overwhelming results (Rossi, 2013). In this chapter we employ a systematic approach to properly take into account time-variation in the coefficients of exchange rate forecasting regressions. The approach also incorporates the idea that the relevant set of regressors or fundamentals may change at each point in time; as articulated, for example, by Bacchetta and van Wincoop (2004, 2013), Berge (2013), and Sarno and Valente (2009). Inspired by recent advances in Bayesian methods, we further employ our systematic framework to investigate all sources of uncertainty in the predictive models, through a variance decomposition procedure.

Using statistical and economic evaluation criteria, we find that fundamentals-based models significantly outperform the driftless random walk benchmark for most currencies at all the forecasting horizons we consider, except at the 1-month horizon. The key to improving upon the benchmark is forecasting with predictive regressions that capture both, the possibly changing set of explanatory variables, and most importantly, the appropriate time-varying weights associated with these variables. At horizons beyond 1-month, i.e., $h = 3$ and $h = 12$, our regressions successfully embed these characteristics. Models which allow for switching sets of regressors and sudden, rather than smooth, changes in the time-varying weights of these regressors are empirically plausible. By contrast, at 1-month forecasting horizon our predictive regressions fail to successfully capture the suitable time-varying weights associated with the regressors, yielding poor statistical and economic performance.

We then proceed and track the sources of uncertainty in the regressions, in an effort to pin down the origins of the weak performance. In this regard we find that the uncertainty in the estimation of the models' coefficients, and the uncertainty about the level of time-variation in coefficients to allow in the model, are the main factors hindering models' predictive ability. When the uncertainty emanating from these sources is low or is successfully embedded in the model, the out-of-sample forecasting performance of the models is satisfactory leading to economic gains. In further characterization of our models, we find that the set of variables that are more informative about exchange rate movements generally differ between forecasting horizons and between countries. But within a specific country-horizon often few variables matter. We view our results as providing direct evidence towards the prevalent conjectures or simulation based evidence that time-variation in parameters of the models might cause time-variation in the models' forecasting performance (Giannone, 2010; Meese and Rogoff, 1983; Rossi, 2013; Rossi and Sekhposyan, 2011).

# Chapter 4

# Revealing Exchange Rate Fundamentals by Bootstrap

## 4.1 Introduction

Identifying macroeconomic fundamentals that can reliably predict fluctuations in exchange rates has long been a catalyst for intense research efforts in exchange rate economics. At the core of this research is the evidence, first put forward by Meese and Rogoff (1983), that a random walk model often provides more accurate forecasts than fully-fledged models of exchange rate determination (e.g., monetary model). Meese and Rogoff (1983) attributed their findings to small sample estimation biases, model misspecification - including unexplained nonlinearities, and parameter instability. Since then, the Meese-Rogoff results have attracted numerous studies, yet mounting evidence suggests that their findings have not yet been fully upturned; see, for example, Rogoff and Stavrakeva (2008) or Rossi (2013), for recent accounts. One main unresolved issue relates to the time-varying predictive content of economic fundamentals, which manifests itself in accurate exchange forecasts in specific periods, but not in others (Cheung et al., 2005; Rossi, 2013).[1]

This chapter employs bootstrap-based methods to reveal the set of exchange rate fundamentals that apply at each point in time. Bootstrapping is a technique for randomly drawing with replacement multiple samples from the existing data. The bootstrap, therefore, allows us to potentially make sharper inferences about model attributes and other quantities of interest, by examining these quantities across sample replications. Exploiting this feature, we can infer for instance how well our estimation

[1]Cheung and Chinn (2001) report survey evidence suggesting that this time-evolving relationship reflects the market participants' changing views on the factors driving exchange rate movements. Macdonald and Marsh (1996) demonstrate that forecasters use common information in different ways when forming their expectations. Bacchetta and van Wincoop (2004, 2013) formalize the idea in a scapegoat model of exchange rate determination. Fratzscher et al. (2015) find mixed support for out-of-sample forecasting ability of the scapegoat model, with the RW producing more accurate forecasts by the mean-squared prediction error metric, but failing on the basis of the direction of change metric. Berge (2013) also considers an idea implied by the scapegoat model, but fails to uncover predictive ability at 1-month horizon.

method fits multiple areas of an exchange rate model, defined according to the fundamental it includes. The strength of the inference is generally strong when the observed sample is a good approximation of the true unobserved population.

In light of the documented time-varying predictive ability of exchange rate fundamentals and concerns about model misspecification, bootstrapping per se may not provide a complete econometric solution to pin down informative exchange rate fundamentals. In order to deal with these issues, we embed a model selection and validation procedure in our bootstrap methods. We begin by generating a large number of bootstrap samples corresponding to our set of forecasting models and applying a pretest model selection rule to each sample. One variant of the bootstrap method we employ takes the models selected in each bootstrap sample to forecast, and subsequently averages the forecasts across sample replications. Breiman (1996) introduced and called this approach bootstrap aggregation or bagging. An alternative bootstrap method we adopt, employs the single best model revealed and trained across bootstrap samples to forecast. We term it bumping, following its initial advocates - Tibshirani and Knight (1999).

Despite their mutual foundation in the bootstrap, the potential of these two methods to improve forecast accuracy hinges upon different aspects. The bagging technique is designed to improve the performance of unstable forecasting procedures, defined as ones in which forecasts differ substantially across sample replications (Breiman, 1996 and Bühlmann and Yu, 2002). Bumping, on the other hand, is intended for procedures that yield many local optima for a specific target criterion (Tibshirani and Knight, 1999). Under a minimum squared prediction error target, for instance, bumping can lead to improvements if the prediction errors from many exchange rate models differ by a narrow margin across samples. To the best of our knowledge, we are the first to apply these techniques in exchange rate economics. We are aware, however, that bagging has been applied to forecasting inflation, unemployment, stock returns, and hedge funds; see, Inoue and Kilian (2008), Jin et al. (2014), Panopoulou and Vrontos (2015), Rapach and Strauss (2010, 2012), and Vivian et al. (2015).[2]

In our application, we juxtapose the performance of our bootstrap-based methods to a set of other competing methods. In this set, all the methods rely on the single sample historical realization to forecast or select and combine forecasts. These include: (i) simple linear regressions conditioned on each fundamental, (ii) combination methods based on the mean, the median, the trimmed mean, and the discounted mean squared prediction error (DMSPE), (iii) Bayesian model averaging (BMA) or selection (BMS) of the sort considered in Wright (2008), and (iv) the kitchen-sink regression of Welch and Goyal (2008). To facilitate our comparison, forecasts from all methods are normalized to those of the toughest benchmark to beat in the exchange rate literature, namely the driftless random walk (Rossi, 2013).

[2] Applications of the bumping method are more common in machine learning, statistics, and medical science.

Focusing on five OECD bilateral currency rates against the U.S. dollar, and a monthly dataset spanning January 1989 to May 2013, we use our methods to forecast recursively at 1-month horizon. As Rossi's (2013) survey reveals, exchange rate predictability has been challenging to detect at this specific horizon. Our study employs standard and more recently propounded exchange rate fundamentals, including from (i) the Taylor (1993) rule (Engel and West, 2005 and Molodtsova and Papell, 2009), (ii) the Nelson-Siegel (1987) relative factors from yield curves (Chen and Tsang, 2013), and (iii) the Engel et al. (2015) exchange rate factors. We evaluate the statistical performance of our methods relative to the RW using the Welch and Goyal's (2008) out-of-sample $R^2_{oos}$, complemented with the Cheung et al.'s (2005) direction of change statistic. Inference on statistical significance is based on the Clark and West (2006) test for the $R^2_{oos}$, and a studentized version of the Diebold and Mariano (1995) test for the direction metric.

In this chapter we further assess whether our methods generate economically significant gains in a stylized dynamic asset allocation strategy. Inspired by Della Corte et al. (2012) and Li et al. (2015), we compute the fee that a risk-averse investor with quadratic utility would be willing to pay to use our methods instead of the RW. In addition, we implement the performance measure of Goetzmann et al. (2007) and Sharpe ratios. To examine whether differences in Sharpe ratios are significant we employ the bootstrap method of Ledoit and Wolf (2008). Finally, we consider Han's (2006) break-even transaction costs that renders an investor indifferent between using our methods and the RW.

In our main findings, bumping reveals sets of economic fundamentals with a strong and significant predictive power for exchange rates. From a statistical perspective, it is the only method that improves upon the RW for a minimum of three exchange rates, irrespective of the measure used. Moreover, the improvement is not ephemeral; rather it is detectable throughout the forecast sample. In terms of economic gains, bumping leads to a Sharpe ratio as high as 0.79, which is significantly different from that of the RW (0.37). The gains associated with bumping are also verifiable over the out-of-sample period. None of the other methods surpasses the performance of bumping, although BMA and the Nelson-Siegel relative curvature factor tend to outperform the RW by the direction of change metric and economic measures. But for BMA, the performance weakens after the 2008 financial crisis. The bagging method improves upon the RW for at most two currencies, but only insignificantly so. Inspection of the performance underlying our bootstrap-based approaches shows that although they both benefit from the bootstrap's attribute to sharpen the inference, bagging tends to over-fit. Bumping, on the contrary, robustly reveals parsimonious models with good out-of-sample predictive power.

The next section sets out our econometric methodology focusing on (i) our bootstrap-based methods, (ii) the competing approaches we consider, (iii) the set of fundamentals or predictors, (iv) the data and forecasting mechanics, (v) the metrics for statistical evaluation, and (vi) the criteria for economic evaluation. Section 4.3 reports the empirical results, including the main features underlying bumping's performance. We conduct robustness checks in Section 4.4 and conclude in Section 4.5.

## 4.2 Methodology

### 4.2.1 Predictive Regression

In our empirical study we are interested in predicting monthly fluctuations in exchange rates using regressors based on empirical exchange rate models. We start by defining a predictive regression that conditions on all potential regressors, known as the kitchen-sink (KS) regression (Welch and Goyal, 2008):

$$\Delta s_{t+1} = \alpha_0 + \sum_{j=1}^{M} \theta_j z_{j,t} + v_{t+1}, \tag{4.1}$$

where $\Delta s_{t+1} = s_{t+1} - s_t$, is the period-ahead change in the log of the spot exchange rate, $\alpha_0$ and $\theta = \{\theta_j\}$ are coefficients to be estimated, $z_{j,t}$ is the regressor $j \leq M$, and $v_{t+1}$ is a normally distributed fitting error.

Although exploiting concomitantly the predictive content of many variables may seem appealing, the success of such strategy in predicting exchange rate fluctuations has been abysmal (see, e.g., Li et al., 2015). An alternative and common approach consists in forecasting with each of the regressors. But the empirical success of such an approach has also been elusive. Often, it is found that while a certain regressor may exhibit predictive power in some periods, in others it fails; see Rossi (2013) for a thorough survey. In our analysis we exploit bootstrap-based methods to select the regressors (macroeconomic fundamentals) that apply at each point in time. By wandering in multiple areas of the model space, these methods may yield superior forecasting performance relative to strategies based on a single sample realization.

### 4.2.2 Bootstrap-based Approaches to Forecasting Exchange Rates

#### Bootstrap Aggregation

We build upon the bootstrap aggregation (bagging) method for models with possibly correlated predictors of Inoue and Kilian (2008), see also Breiman (1996). Bagging involves generating a large number of bootstrap samples from the original data, applying a model pre-selection rule to each sample, and subsequently averaging the forecasts generated from the selected models across bootstrap samples. Instead of using a single sample realization to select a forecasting model, bagging stabilizes the decision rule by repeating the pre-testing on a large number of pseudo-data replications. The method proceeds as follows:

1. Collect $\Delta s_t$ and $z_{j,t}$ for $j = (1, ..., M)$ in a matrix $L_B$ of size $(T_0 \times (M+1))$, and construct $B$ sample replications of this matrix, i.e., the left and right-hand side variables, by randomly drawing with replacement blocks of size $m$. Note that $T_0$ is the portion of the data available to the forecaster at time $t$, i.e., the in-sample period;

2. For each pseudo-data sample, estimate via Ordinary Least Squares (OLS) the unrestricted model in Eq. (4.1), and conduct a two-sided $t$-test on each parameter estimate $\theta_j$ based on a pre-specified critical value $c^*$. The predictors $z_{j,t}$ with $t$-statistics less than $c^*$ in absolute value are dropped from Eq. (4.1). If all the predictors and the constant are insignificant, then a full restricted model is obtained, since $\Delta s_{t+1} = 0$. In computing the $t$-statistics we use Newey and West's (1987) heteroscesdasticity and autocorrelation consistent (HAC) standard errors with a lag truncation parameter of $int\{T_0\}^{1/4}$;

3. Re-estimate the model in the pseudo-data using the selected regressors, and use the parameter estimates with the actual values of these regressors to generate the forecast;

4. Obtain the bagging forecast as the average of the $B$ forecasts across bootstrap samples.

We use $B = 100$, $m = 1$, and $c^* = 2.807$, as in Inoue and Kilian (2008).[3]

Breiman (1996) shows that the bagging method's ability to improve forecast accuracy depends on the instability of the forecasting procedure. If forecasts from multiple areas of an exchange rate model tend to be similar, the bagging forecast will be centered around the forecast from the observed sample, often without improving upon it. The improvement typically occurs when small alterations in the bootstrapped samples lead to large changes in the resulting forecast. In this case the gains in mean squared forecast error stem from reduction in variance.[4]

More generally, due to its foundation in the bootstrap, bagging may yield forecast improvements by fitting the regression in many areas of the model space. For instance, if few observations or outliers in the historical sample are causing the regression to fit poorly, the fit may improve in sample replications that exclude those observations. Since the probability of an historical observation to appear in a bootstrap sample is $1 - (1 - 1/T_0)^{T_0} \approx 1 - e^{-1} = 0.632$, there is a 0.368 probability that any outlier will be washed away from a sample replication (Tibshirani and Knight, 1999).

[3]Inoue and Kilian (2008) employ similar values for $B$ and $m$, but use a grid of critical values and report results based on the one that produces the best results. Our results are qualitatively insensitive to using alternative values of $c^*$, such as 1.960 or 2.576.

[4]Plainly, bagging does not improve forecast accuracy if the predictors do not exhibit predictive content or they are at the limit of accuracy attainable in the data.

**Model Search by Bootstrap Sampling**

Because bagging can get stuck in poor solutions, we also use bumping or model search by bootstrap (Tibshirani and Knight, 1999). In bumping, we pick as our forecasting model the one that best satisfies a target criterion in a training period of the bootstrap replications. In this manner, if various predictors tend to yield a similar forecasting performance, bumping allows us to approximate the most satisfactory ones. In this sense, the method relies on the bootstrap to robustly identify these predictors.

In terms of implementation, we also begin by drawing bootstrap replications of the $L_B$ matrix as in the first step of bagging. If we split the $T_0$ in-sample observations into $T_e$ and $T_s$ with $T_0 = T_e + T_s$, then we use the first $T_e$ observations from each bootstrap sample to execute the second step (pre-testing). Subsequently, we employ the $T_s$ observations to re-estimate the model and forecast as in the third step, and calculate the mean squared forecast error (MSFE) in the original data. We proceed and select the model from the bootstrap sample $\widehat{b}$, if it produces the lowest MSFE. In notation,

$$\widehat{b} = \arg\min_{b} \frac{1}{T_s} \sum_{t=T_e+1}^{T_s} \left( \Delta s_{t+1} - \widehat{f}_t^{*b} \right)^2, \tag{4.2}$$

with $\widehat{f}_t^{*b}$ denoting the model's forecast in the bootstrap replication $b$. Finally, the forecasts from bumping are generated by using the parameters estimated from this model in the original data. Note that by convention, the original data-matrix is also included in the set of bootstrap replications to allow the method to pick the model based on it, in case it yields the lowest training forecast error.[5]

### 4.2.3 Other Competing Approaches

In bagging and bumping we forecast or select forecasting models using pseudo-data. We now turn to examining if other schemes to forecast combinations or model selection based on the sample realization lead to forecast improvements. These methods entail first generating forecasts from predictive models based on a single regressor. The combined forecast made at time $t$, denoted $\Delta s_{c,t+1}$, is a weighted average of the $M$ individual models' forecasts:

$$\Delta s_{c,t+1} = \sum_{j=1}^{M} \omega_{j,t} \Delta \widehat{s}_{j,t+1}, \tag{4.3}$$

where $\{\omega_{j,t}\}_{j=1}^{M}$ are the combining weights computed using one of the approaches discussed below.

[5]Note too that unlike bagging, bumping preserves the original data-structure when generating the forecast.

### Bayesian Model Averaging or Selection

We first consider Bayesian model averaging (BMA) and selection (BMS). In BMA the combining weights are derived from the posterior probabilities of each model. The starting point is to assign a prior probability, $P(Md_j)$, which is the belief that among the $M$ models, the j$^{th}$ model is the true one. And then updating the prior belief by computing the model's posterior probability as new information becomes available. This posterior probability takes the following form:

$$P(Md_j|D_t) = \omega_{j,t} = \frac{P(D_t|Md_j)P(Md_i)}{\sum_{i=1}^{M} P(D_t|Md_i)P(Md_j)}, \tag{4.4}$$

where $P(D_t|Md_j)$=$\int P(D_t|\theta_j, Md_j)P(\theta_j|Md_j)\mathrm{d}\theta_j$ is the marginal likelihood of the $Md_j$ model, $P(\theta_j|Md_j)$ is the prior density of the parameter vector $\theta_j$ which includes the constant $\alpha_0$, $P(D_t|\theta_j, Md_j)$ is the likelihood, and $D_t$ is the conditioning information set comprising the response and explanatory variables and the prior parameters at time $t$ (see Wright, 2008). BMS uses the weights derived from Eq. (4.4) to pick the single most likely model, defined as the one with the largest posterior probability, and using it to forecast at time $t$.

Apart from knowing the set of models under consideration in Eq. (4.4), we need to specify the prior probability for each model and the prior and posterior distributions of the parameters. We set an equal initial probability for each model: $P(Md_j) = 1/M$. On the parameter vector we employ a natural conjugate g-prior, which ensures that the posterior distributions belong to the same family as the prior distributions. We follow Wright (2008) and define these prior distributions as:

$$V_t|D_0 \sim IG\left[\frac{1}{2}, \frac{1}{2}S_0\right], \tag{4.5}$$

$$\theta_0|D_0, V_t \sim N\left[0_{n\times 1}, S_0(gz'z)^{-1}\right], \tag{4.6}$$

$$S_0 = \frac{1}{n-1}\Delta s'(I - z(z'z)^{-1}z')\Delta s, \tag{4.7}$$

where $\theta_0|(\cdot)$ is the prior for the coefficients vector, $D_0$ denotes the information set at $t_0$. $V_t$ is the time-varying variance of the error term in the predictive regression, $(n-1)$ are the degrees of freedom, $S_0$ is the OLS estimate of the variance in coefficients, and $g$ is a scaling factor that characterizes the confidence assigned to the prior for the coefficients in (4.6). Our results are based on a typical benchmark prior of $g = 1/T$, following Fernandez et al. (2001). For extra details on the exact updating scheme of the prior beliefs in this BMA approach see Dangl and Halling (2012) or Wright (2008).[6]

[6] Note that here we estimate our models allowing for stochastic volatility for the error term in Eq. (4.1), i.e., $v_{t+1} \sim N(0, V_t)$. We do so following results in Byrne et al. (2014) regarding the importance of allowing for time-changing volatility to improve exchange rate forecasts when using Bayesian methods.

**Classic Forecast Combinations**

Apart from BMA, other researchers have found simple weighting schemes to be optimal (e.g., Rapach et al., 2010). In this category we focus on four methods:

- Mean combination. The combined forecasts are obtained by applying a constant weight of: $\omega_{j,t} = 1/M$ in Eq. (4.3).
- Median combination. The combined forecast is the median of $\{\Delta\widehat{s}_{j,t+1}\}_{j=1}^{M}$.
- Trimmed mean combination. The combined forecasts are obtained by setting $\omega_{j,t} = 0$ for the smallest and largest individual forecasts, and $\omega_{j,t} = 1/(M-2)$ for the remaining forecasts in Eq. (4.3).
- DMSPE combination. The combining weights in the discount mean squared prediction error (DMSPE) method are related to the historical forecasting performance of the individual models in an holdout-out-of-sample period. The method uses the following time-varying weights:

$$\omega_{j,t} = \Phi_{j,t}^{-1} / \sum_{i=1}^{M} \Phi_{i,t}^{-1}, \qquad \Phi_{i,t} = \sum_{t=T_e+1}^{T-1} \vartheta^{T-1-t} (\Delta s_{j,t+1} - \Delta\widehat{s}_{j,t+1})^2, \tag{4.8}$$

  where $T_e$ is the end of the in-sample portion on which we condition to generate the initial forecasts, $\vartheta$ denotes a discount factor. Based on investigations from Rapach et al. (2010) we use $\vartheta$=0.9. This corresponds to attaching larger weight to the most recent forecasts in the holdout-out-of-sample period.

### 4.2.4 Menu of Macroeconomic Fundamentals

Our empirical application employs information sets derived from empirical exchange rate models. In this sense, our framework is broadly consistent with the Engel and West (2005) asset pricing setting. The setting links fluctuations in the spot exchange rate to current observable fundamentals and unobservable noise. In our case, each information set defines a regressor $z_{j,t}$, for $j = 1, ...M$ and $M = 8$.

**Taylor rule.** The first information set is based on the asymmetric Taylor (1993) rule (TRasy):

$$z_{1,t} = 1.5(\pi_t - \pi_t^*) + 0.1(\overline{y}_t - \overline{y}_t^*) + 0.1(s_t + p_t^* - p_t), \tag{4.9}$$

where $\pi_t$ is the domestic inflation rate, $\overline{y}_t$ the domestic output gap, $p_t$ is the log of the domestic price level, and asterisks denote identical variables for the foreign country. The parameters on inflation differentials (1.5), output differentials (0.1), and the real exchange rate (0.1) are inspired by Li et al. (2015).

**Monetary fundamentals.** The second regressor stems from the monetary model (MM) and is computed as:

$$z_{2,t} = (m_t - m_t^*) - k_y(y_t - y_t^*) - s_t, \tag{4.10}$$

with $m_t$ ($m_t^*$) denoting the log of domestic (foreign) money supply and $y_t$ ($y_t^*$) the log of domestic (foreign) income. Following Mark (1995), we assume an income elasticity of one in the output differentials; i.e., $k_y = 1$.

**Purchasing power parity (PPP) and uncovered interest rate parity (UIP)**. Our third and fourth regressors are derived from the PPP and UIP conditions:

$$z_{3,t} = p_t - p_t^* - s_t, \tag{4.11}$$

$$z_{4,t} = i_t - i_t^*, \tag{4.12}$$

with $i_t$ and $i_t^*$ defining the domestic and foreign short-term nominal interest rate, respectively.

**Nelson-Siegel relative factors**. An extra set of three regressors builds on the recently developed view that the Nelson-Siegel (1987) relative factors can explain future movements in exchange rates (Chen and Tsang, 2013). Therefore:

$$z_{5,t} = L_t^{NS}, \qquad z_{6,t} = S_t^{NS}, \qquad z_{7,t} = C_t^{NS}, \tag{4.13}$$

where $L_t^{NS}$ is the relative level factor, $S_t^{NS}$ the relative slope factor, and $C_t^{NS}$ the relative curvature factor. Following Chen and Tsang (2013) we obtain the relative factors for each period from OLS estimates of the following equation:

$$i_t^m - i_t^{m*} = L_t^{NS} + S_t^{NS}\left(\frac{1-e^{-\lambda m}}{\lambda m}\right) + C_t^{NS}\left(\frac{1-e^{-\lambda m}}{\lambda m} - e^{-\lambda m}\right) + \varepsilon_t^m \tag{4.14}$$

with $i_t^m$ denoting the continuously compounded zero-coupon nominal domestic yield on an $m$-month bond, $i_t^{m*}$ is its foreign counterpart, $\lambda$ is a parameter set to 0.0609 as typical in the literature, and $\varepsilon_t^m$ is an estimation error. Note that we employ each relative factor as a regressor following results in Chen and Tsang (2013), regarding their distinct predictive ability. Further, the relative factors tend to be highly correlated.[7]

**Co-movement or factors extracted from exchange rates**. While the above information sets are based on observable fundamentals, Engel and West (2015) show that factors extracted from exchange rates may embed non-observable information related

[7]The level factor $L_t^{NS}$ has a constant loading in the yield curve and picks up aspects that shift the relative yield curve (e.g., inflation expectations). The slope factor $S_t^{NS}$ captures the short end of the yield curve and is connected with dynamics in monetary policy. The curvature factor $C_t^{NS}$ has its largest effect in the middle of the yield curve, and a zero loading at maturity $m = 0$ as well as at the extreme maturities (see also Berge, 2013)

to observable fundamentals that is useful for predicting exchange rates. Accordingly, we extract one factor from the five exchange rates in our sample and define our last predictor as:

$$z_{8,t} = F_t^s = \widehat{\delta}_n \widehat{f}_t - s_{n,t}, \tag{4.15}$$

where $\widehat{f}_t$ is a factor estimated by principal components analysis, and $\widehat{\delta}_n$ is the loading for the currency $n$, $n = 1, ..., 5$.[8]

### 4.2.5 Data and Forecasting Procedure

We use monthly figures from January 1989 to May 2013 for a panel of six countries (currencies): Canada (CAD), Germany/Euro area (EUR), United Kingdom (GBP), Japan (JPY), Sweden (SEK), and the United States (USD). Exchange rates are end-of-month values of the U.S. dollar (USD) price of a unit of foreign currency. Each regressor is therefore defined from the perspective of the United States (home country) relative to the foreign country.[9]

The main data source is the IMF's International Financial Statistics (IFS), supplemented by national official sources, see Appendix H. The money supply is measured by the M1 aggregate.[10] The price level consists of the consumer price index (CPI) and the inflation rate is defined as the (log) CPI monthly change. We proxy output using monthly industrial production (IP). We compute the output gap by applying the Hodrick and Prescott (1997) filter recursively to the output series while correcting for the uncertainty about the HP estimates at the recursive sample end-points, see Watson (2007).[11] In computing Taylor rule fundamentals we follow Molodtsova and Papell (2009) and use central bank's policy rate when available for the entire sample period or alternatively the money market rate. The data on money supply, IP, and CPI were transformed by taking logs and adjusted for seasonality by computing the mean over twelve months as in Engel et al. (2015).

The Nelson-Siegel relative factors are constructed from zero-coupon bonds yields for various maturities, typically of 3, 6, 12, 24, 36, 60, 72, 84, 96, 108, and 120 months. Bond yields for Sweden are available for only four maturities, hence we exclude the Nelson-Siegel relative factors when forecasting the SEK-U.S. dollar exchange rate.

In terms of forecasting mechanics, we generate forecasts recursively from February 2000 by re-estimating our model parameters every time an additional observation is

[8] We limit our number of factors to one, due to the small number of currencies in our sample.

[9] Prior to the introduction of the EUR in January 1999, we employ the U.S. dollar–Deutsche mark exchange rate scaled by the official conversion factor. We also construct fundamentals using Germany data.

[10] Due to unavailability of the M1 aggregate for Sweden and the UK we use the M3 and M4 aggregate, respectively.

[11] We estimate bivariate VAR($\ell$) models that include the first difference of inflation and the change in the log IP, with $\ell$ determined by Akaike Information criterion. These models are then used to forecast and backcast three years of monthly data-points of IP, and the HP filter is applied to the resulting extended series. In the HP filter, we use the standard smoothing parameter for monthly data (i.e., 14400).

added to the sample (see, e.g., Engel et al. 2015). In detail, after data transformations we employ data from January 1990 to January 2000 to estimate the exchange rate factors and loadings. We then use data on our regressors from January 1990 to December 1999 to estimate parameters of our forecasting regression. These parameters are used with January 2000 right hand-side data to generate the first forecast. We add one extra data point and repeat the exercise until May 2013, providing us with $P = 160$ out-of-sample (OOS) forecasts. For methods that require an hold-out-of sample period, such as bumping and DMSPE, we set this period to $T_s = 60$, implying that for these methods the evaluation period begins in February 2005, for a total of $P = 100$ OOS forecasts.

### 4.2.6 Statistical Evaluation Criteria

To evaluate the statistical forecasting accuracy of each method, which we henceforth denote Fundamentals-based Forecasting Method (FbFM), we use two main criteria: the Welch and Goyal's (2008) out-of-sample $R^2$ and the Cheung et al.'s (2005) direction of change statistic. In both cases the driftless random walk (RW) is the benchmark model. The out-of sample $R^2$, hereafter $R^2_{oos}$, is based on the Mean Squared Forecast Error (MSFE) and is computed as:

$$R^2_{oos} = 1 - (MSFE_{FbFM}/MSFE_{Rw}). \tag{4.16}$$

A positive $R^2_{oos}$ means that on average, the FbFM generates forecasts with smaller MSFE than the RW. Inference on the statistical significance of the $R^2_{oos}$ is based on the usual asymptotic Clark and West (2006) test. This is a test of the null hypothesis of equal MSFE between our unrestricted FbFM and the restricted RW, after adjusting the MSFE of the FbFM to account for the noise it introduces in the forecasting process (Clark and West, 2006).

The direction of change statistic, also known as the success or hit rate, captures the ability of the FbFM to time the market, and we use it as a complement to the $R^2_{oos}$ metric. It is calculated as the average number of correct predictions of the direction of change of the exchange rate. A value greater (smaller) than 0.5 implies that the FbFM is able (unable) to predict the correct sign of exchange rate change. We also test the null of equal timing ability between the FbFM and a naive benchmark that predicts that the exchange rate might go down or up with the same probability. Cheung et al. (2005) propose using a studentized version of the Diebold and Mariano (1995) test:

$$DM = (\overline{d} - 0.5)/\sqrt{0.25/P}, \tag{4.17}$$

where $\overline{d}$ is the mean of the series $d_t$, which in turn contains ones for OOS observations where the FbFM predicts the correct direction of change and zero otherwise. In large samples, critical values from the standard Normal distribution can be readily used with the test.[12]

### 4.2.7 Economic Evaluation Criteria

Our models' forecasting performance is also examined based on their ability to generate significant economic gains in a stylized dynamic asset allocation strategy. Following Della Corte et al. (2012), we consider a U.S. investor who dynamically rebalances her portfolio by allocating her wealth between a U.S. bond and the bonds of the five countries in our sample ($N_5$). The yields of the bonds are taken to be the end-of month money market rates. At each period, $t+1$, the $N_5$ bonds are associated with a risk-free return in local currency but a risky return $r_{t+1}$ in U.S. dollars. Thus, when investing in any of the $N_5$ bonds at time $t$, the U.S. investor expects a return of $r_{t+1|t} = i_t^* + \Delta s_{t+1|t}$, where here, $r_{t+1|t} = E_t[r_{t+1}]$ is the conditional expectation of $r_{t+1}$; $i_t^*$ is the nominal interest rate in the foreign country; and $\Delta s_{t+1|t} = E_t[\Delta s_{t+1}]$ is the conditional expectation of $\Delta s_{t+1}$. Given that at time $t$ the interest rate $i_t^*$ is known, the return that the U.S. investor expects from time $t$ to $t+1$ by investing in an $N_5$ bond is only exposed to the foreign exchange risk. To the extent that she can accurately forecast $\Delta s_{t+1|t}$, she can minimize this risk.

To implement the dynamic asset allocation strategy the investor proceeds in two stages. First, she uses each of the forecasting method to predict monthly fluctuations in the exchange rates. Second, using each model's forecast she rebalances her portfolio by calculating new optimal weights on each bond and time $t$, taking into account the portfolio's mean return and the variance. If we define $r_{t+1}$, as the vector of ($N_5 \times 1$) risky asset returns; $u_{t+1|t} = E_t[r_{t+1}]$, as the conditional expectation of $r_{t+1}$; and $\Sigma_{t+1|t} = [(r_{t+1} - u_{t+1|t})(r_{t+1} - u_{t+1|t})']$, as the conditional covariance matrix; then at each period $t$, the investor wishes to find the optimal portfolio weight subject to a target volatility of the portfolio returns. The solution to the investor's problem yields the following weights on the risky assets (Della Corte et al., 2012):

$$w_t = \frac{\sigma_p^*}{\sqrt{C_t}} \Sigma_{t+1|t}^{-1}(u_{t+1|t} - \iota r_f), \tag{4.18}$$

where $\sigma_p^*$ is the portfolio return target volatility and $C_t = (u_{t+1|t} - \iota r_f)' \Sigma_{t+1|t}^{-1}(u_{t+1|t} - \iota r_f)$. The weight on the risk-free asset is $(1 - w_t'\iota)$.

The gross portfolio return at time $t+1$ is:

$$R_{p,t+1} = 1 + w_t' r_{t+1} + (1 - w_t'\iota) r_f = R_f + w_t'(R_t - \iota R_f), \tag{4.19}$$

where $R_t$ is the ($N_5 \times 1$) vector of gross risky returns and $R_f$ is the gross return on the risk-free assets. Note that as in Della Corte et al. (2012), the investor uses the unconditional covariance matrix $\overline{\Sigma}$, rather than the conditional one ($\Sigma_{t+1|t}$), i.e.,

[12] In computing the CW and DW tests we use heteroscedasticity and autocorrelation-consistent standard errors, with a lag truncation parameter of $int\{T_0\}^{1/4}$ following Rossi (2013).

$\Sigma_{t+1|t} = \overline{\Sigma}$, hence, the optimal weights will change only to the extent of the differences in model's forecasting performance.

Given that each method's forecast implies different portfolio weights and hence distinct gross portfolio returns, we can evaluate the methods' ability to deliver higher returns relative to a strategy based on the RW. Inspired by Della Corte et al. (2012), we focus on the following metrics:

- Sharpe ratio ($SR$). Calculated as the ratio of the average realized excess portfolio returns relative to the standard deviation of the portfolio returns. This is computed for each FbFM and for the RW benchmark. The method that generates higher Sharpe ratios is preferred to the one with lower $SR$ scores. To examine whether differences in Sharpe ratios are significant we employ the bootstrap procedure of Ledoit and Wolf (2008).

- Maximum performance fee. This is the fee a risk-averse investor with quadratic utility would be willing to pay to use the FbFM, instead of the RW. It is based on Fleming et al.'s (2001) notion that forecasts from one model are better than those from a second model, if investment decisions based on the first model lead to higher average realized utility than the second one. To compute the fee consider, for instance, that holding a portfolio based on the RW yields the same average utility as holding a portfolio based on a FbFM that is subject to expenses $pf$ at monthly frequency. Given that the investor would be indifferent between the two alternatives, $pf$ is interpreted as the maximum performance fee that she would be willing to pay to switch from the RW model to the FbFM. The $pf$, expressed as a proportion of the assets invested, is estimated as a value that satisfies the following equality:

$$\begin{aligned} & \sum_{t=0}^{T-1} \left\{ (R^*_{p,t+1} - pf) - \frac{RRA}{2(1+RRA)} (R^*_{p,t+1} - pf)^2 \right\} \\ = & \sum_{t=0}^{T-1} \left\{ R_{p,t+1} - \frac{RRA}{2(1+RRA)} R^2_{p,t+1} \right\}, \end{aligned} \tag{4.20}$$

  where $R^*_{p,t+1}$ is the gross return from using the FbFM, $R_{p,t+1}$ is the gross return from using the RW, and $RRA$ is the investor's constant degree of relative risk aversion. We report the estimate of $pf$ in annualized basis points ($bps$). The higher the value of $pf$, the more the investor wishes to pay to switch from the RW to the FbFM.

- Excess premium of the FbFM relative to the RW. This is based on the manipulation-proof performance measure of Goetzmann et al. (2007), which captures the portfolio's premium return after accounting for risk. The manipulation-proof performance measure is defined as:

$$M(R_p) = \frac{1}{1 - RRA} \ln \left\{ \frac{1}{T} \sum_{t=0}^{T-1} \left( \frac{R_{p,t+1}}{R_f} \right)^{1-RRA} \right\}, \tag{4.21}$$

where $M(R_p)$ is the risk adjusted portfolio's premium return based on the RW model. The corresponding measure for the FbFM is denoted by $M(R_p^*)$. Hence, the excess premium return of the FbFM relative to the RW, $pr$, is computed as:

$$pr = M(R_p^*) - M(R_p). \tag{4.22}$$

The higher the value of $pr$, the greater is the portfolio's premium return based on the FbFM relative to the benchmark RW after adjusting for risk. We also report the excess premium return in annualized *bps*.

- Break-even transaction cost of Han (2006). It is the fractional transaction cost that offsets the positive performance fee obtained by using the FbFM. At this break-even point, the investor is indifferent between using the RW and the FbFM. To compute the cost, we follow Han (2006) and Della Corte et al. (2012), and assume that transaction costs equal a fixed fraction ($T_r$) of the value traded in each bond. Hence, the costs are: $T_r \left| w_t - w_{t-h}(R_t/R_p) \right|$. Whenever the investor's transaction costs are below the break-even transaction cost level, denoted $T^{be}$, she will continue to prefer using the FbFM; otherwise she would use the RW benchmark. Given that $T^{be}$ is a proportional cost paid at every period at which the portfolio is rebalanced, $T^{be}$ is reported in monthly *bps*.

# 4.3 Empirical Results

## 4.3.1 Statistical Evaluation of Forecasting Performance

We begin by reporting the performance of our FbFM relative to the RW using the $R^2_{oos}$ and the direction of change metric. Recall that a positive $R^2_{oos}$ means that, on average, the FbFM generates a lower MSFE and therefore forecasts better than the RW. For the direction of change statistic, a value above (below) 0.5 implies that the forecasts from the FbFM are on average consistent (inconsistent) with the direction of the exchange rate change and therefore are better (worse) than a random directional forecast. The statistical significance of the $R^2_{oos}$ is examined using the Clark and West (2006) test, while for the direction of change statistic we use a Diebold and Mariano (1995) type-test.[13]

[13] Note that although we discuss exclusively the success rates of the FbFM relative to the RW, an important feature of the direction of change metric is the usefulness of the information contained in the proportion of cases where the FbFM wrongly predicts the direction of the exchange rate change. For trading purposes this information can be used to derive a potentially profitable trading rule by going against the model's prediction (Cheung et al., 2005).

Table 4.1: Statistical Evaluation Based on $R^2_{oos}$ (%)

| | CAD | EUR | GBP | JPY | SEK |
|---|---|---|---|---|---|
| Bumping | $0.815^a$ | $3.477^a$ | $2.485^b$ | -0.637 | -1.919 |
| Bagging | 0.738 | -0.536 | 0.279 | -1.704 | -0.505 |
| | | | | | |
| Kitchen-sink regression | -3.947 | -7.902 | -6.986 | -14.623 | -1.939 |
| | | | | | |
| Bayesian model averaging | -0.545 | 0.023 | -0.330 | -0.111 | 0.497 |
| Bayesian model selection | -2.100 | 0.276 | -0.943 | 0.328 | -0.190 |
| | | | | | |
| Mean combination | -0.341 | -0.434 | 0.279 | 0.248 | 0.129 |
| Median combination | -0.579 | 0.231 | -0.584 | 0.019 | 0.492 |
| Trimmed mean combination | -0.483 | -0.109 | 0.263 | 0.144 | 0.356 |
| DMSPE combination | -0.666 | -0.018 | 0.741 | 0.323 | 0.872 |
| | | | | | |
| Taylor rule | -0.414 | $1.247^a$ | -4.165 | -1.065 | 0.493 |
| Monetary fundamentals | -2.886 | -3.378 | -1.941 | -1.580 | -2.294 |
| Purchasing power parity | -0.800 | 0.460 | -2.358 | -1.488 | $0.676^a$ |
| Uncovered interest parity | -0.916 | -0.856 | -0.538 | -1.138 | -0.707 |
| NS relative level factor | -0.977 | -1.368 | -0.883 | 0.798 | - |
| NS relative slope factor | -1.124 | -1.042 | -0.954 | -0.032 | - |
| NS relative curvature factor | $-0.240^a$ | -1.009 | -0.856 | -5.441 | - |
| Factor from exchange rates | -1.610 | -0.708 | -1.927 | -1.069 | 0.331 |

**Notes:** The table reports the out-of-sample $R^2_{oos}$ in percent for each of the Fundamentals-based Forecasting Method (FbFM) in the first column relative the driftless random walk (RW). Positive values mean that the FbFM generates a lower MSFE and therefore forecasts better than the RW. The superscripts $a$, $b$, and $c$ denote statistical significance at the 10%, 5%, and 1% level, respectively using the Clark and West (2006) one-sided $t$-statistic. The forecast evaluation period is February 2000 to May 2013 for all methods, except the bumping method and DMSPE combination method (February 2005 to May 2005). Currency codes refer to the Canadian dollar - CAD, euro - EUR, British pound - GBP, Japanese yen - JPY, and the Swedish krona - SEK.

Table 4.1 presents the $R^2_{oos}$ statistics. The main findings are as follows. First, bumping is the only method that generates positive and significant $R^2_{oos}$ for up to three of the five exchange rates we focus upon. The magnitude of the $R^2_{oos}$ range from 0.815 for the Canadian dollar to 3.477 for the euro. Second, although the $R^2_{oos}$ statistics for classic combination methods are also positive for three currencies, they tend to be statistically insignificant. This is the case for the mean, median, trimmed mean, and the DMSPE combination methods. The BMA, BMS, and bagging methods also fall behind the RW, as they produce a positive but insignificant $R^2_{oos}$ for two currencies. Third, our results confirm the abysmal forecasting performance of the kitchen-sink regression, with a negative $R^2_{oos}$ in all cases and the worst magnitudes across all methods for four currencies. Finally, our findings also highlight the usual poor performance of the individual empirical models relative to the RW, especially at the horizon we are considering. In this category, the $R^2_{oos}$ are mostly negative, with the Taylor rule model and the PPP condition outperforming the RW for up to two currencies but for only

Table 4.2: Statistical Evaluation Based on the Direction of Change Statistic

| | CAD | EUR | GBP | JPY | SEK |
|---|---|---|---|---|---|
| Bumping | $0.580^a$ | $0.610^b$ | $0.640^c$ | $0.640^c$ | 0.460 |
| Bagging | $0.594^c$ | 0.494 | 0.525 | 0.494 | 0.456 |
| Kitchen-sink regression | 0.544 | 0.513 | 0.550 | 0.506 | 0.544 |
| Bayesian model averaging | $0.575^b$ | $0.575^b$ | 0.538 | $0.556^a$ | $0.569^b$ |
| Bayesian model selection | $0.563^a$ | $0.613^c$ | 0.538 | 0.500 | 0.525 |
| Mean combination | $0.556^a$ | 0.475 | 0.519 | 0.513 | 0.500 |
| Median combination | 0.494 | 0.525 | 0.494 | 0.500 | 0.544 |
| Trimmed mean combination | 0.519 | 0.506 | 0.506 | 0.506 | 0.513 |
| DMSPE combination | 0.530 | 0.470 | 0.510 | 0.540 | 0.560 |
| Taylor rule | 0.463 | 0.544 | 0.506 | 0.513 | 0.550 |
| Monetary fundamentals | 0.550 | 0.456 | 0.494 | 0.444 | 0.525 |
| Purchasing power parity | 0.444 | 0.519 | 0.488 | 0.525 | $0.569^b$ |
| Uncovered interest parity | 0.488 | 0.519 | 0.488 | 0.544 | 0.494 |
| NS relative level factor | $0.556^a$ | 0.538 | 0.494 | 0.525 | - |
| NS relative slope factor | 0.500 | 0.519 | 0.500 | 0.538 | - |
| NS relative curvature factor | $0.600^c$ | $0.569^b$ | 0.513 | 0.544 | - |
| Factor from exchange rates | 0.431 | 0.500 | 0.538 | 0.538 | 0.481 |

**Notes:** The table shows the proportion of forecasts from each Fundamentals-based Forecasting Method (FbFM) listed in the first column that correctly predict the direction of the exchange rate change. Values above 0.5 indicate that the FbFM is able to predict the direction of exchange rate change, while values below 0.5 suggest that the method tends to predict the wrong direction of change. The superscripts $a$, $b$, and $c$ denote statistical significance at the 10%, 5%, and 1% level, respectively using a studentized version of the Diebold and Mariano (1995) one-sided $t$-statistic. The forecast evaluation period is February 2000 to May 2013 for all methods, except the bumping method and DMSPE combination method (February 2005 to May 2005). Currency codes refer to the Canadian dollar - CAD, euro - EUR, British pound - GBP, Japanese yen - JPY, and the Swedish krona - SEK.

one, significantly so.

In Table 4.2 we focus on the direction of change metric. At a glance, forecasts from bumping are significantly consistent with the direction of the exchange rate change for four exchange rates, with the success rate ranging between 58% to 64% of the total forecasts. Next to bumping is BMA, with significant success rates revolving around 57% also for four exchange rates. Finally, the classic forecast combinations methods are able to correctly predict the direction of change in about 51% of the forecasts for most currencies. However, these hit rates are typically insignificant, a finding also common to the kitchen-sink regression, the bagging method, and most of the single predictor models. The exception in the single predictor case is the regression conditioned on the Nelson-Siegel relative curvature factor. It exhibits significant success rates of 60% and 57% for the Canadian dollar and the euro, respectively. Overall, the statistical evaluation reveals the ability of the bumping method to improve upon the RW and all the competing methods, regardless of the statistical metric. BMA and the Nelson-

Siegel relative curvature factor line up next after bumping when using the direction of change metric.

### 4.3.2 Economic Evaluation of Forecasting Performance

Our statistical results are not fully informative to an investor, as they do not convey the magnitude of economic value associated with better predictive ability. In effect, Campbell and Thompson (2008) note that seemingly small statistical improvements in forecasting performance can still yield significant economic gains in dynamic trading strategies. Our analysis focus on four measures of economic performance: the Sharpe ratio ($SR$), the performance fee ($pf$), the excess premium return ($pr$), and the break-even transaction cost ($T^{be}$). We set the annualized target volatility to $\sigma_p^* = 10\%$ and the relative risk aversion factor to $RRA = 6$, following Della Corte et al. (2012) and Li et al. (2015).

Table 4.3 reports the performance associated with each method or forecasting strategy. Clearly, bumping delivers the highest economic value irrespective of the economic measure used. It exhibits a $SR$ of 0.74 which is statistically different from the $SR$ of the RW (0.37) at the 5% level of significance. Moreover, its performance fee is 1105 annual basis points ($bps$), implying that a risk-averse investor is willing to pay an annual fee of 11% to use bumping instead of the RW. The excess premium return is of similar magnitude (1125$bps$) and the break-even transaction costs remain at 61 $bps$. None of the other strategies surpasses the performance of bumping. At most, several are only able to outperform the RW, including bagging, BMA, the mean and trimmed mean combination, and the Nelson-Siegel relative curvature factor. Note however that in these cases, indicators like Sharpe ratios are statistically insignificant using the Ledoit and Wolf (2008) bootstrapped test-statistic.

The rest of the forecasting strategies fall behind the RW. The kitchen-sink regression for example, yields a $SR$ of 0.08 and a negative performance fee ($-329bps$). Similarly, BMS, and all the single predictor models excluding the Nelson-Siegel relative curvature factor also deliver negative performance fees and premium returns. In conclusion, the statistical predictive ability of the bumping method leads to significant economic gains, unmatched by any of the competing methods including the benchmark RW.

### 4.3.3 Forecasting Performance Over Time

Our findings thus far are based on global performance over the entire OOS period. In this sense they mask the performance of our methods over time. To inspect this performance in Figure 4.1 we compute the recursive or local $R^2_{oos}$ metric for bumping and five extra methods. Three aspects are salient in the Figure. First, for all the three currencies (CAD, EUR, and GBP) for which the global or average performance of bumping is better than the RW, the recursive $R^2_{oos}$ is generally positive at each OOS data-point. This means that its statistical performance is consistent over time.

Table 4.3: Economic Evaluation of Forecasting Performance

| Strategy | $\overline{Rt}(\%)$ | vol (%) | $SR$ | $pf(bps)$ | $pr(bps)$ | $T^{be}(bps)$ |
|---|---|---|---|---|---|---|
| Random Walk | 6.42 | 10.93 | 0.37 | | | |
| Bumping | 9.81 | 10.40 | $0.74^b$ | 1105 | 1128 | 61 |
| Bagging | 10.23 | 13.96 | 0.56 | 147 | 160 | 52 |
| Kitchen-sink regression | 3.33 | 11.23 | 0.08 | -329 | -309 | - |
| Bayesian model averaging | 7.96 | 11.10 | 0.50 | 142 | 136 | 41 |
| Bayesian model selection | 6.14 | 12.02 | 0.31 | -106 | -119 | - |
| Mean combination | 7.58 | 12.14 | 0.43 | 30 | 34 | 19 |
| Median combination | 6.90 | 12.28 | 0.37 | -49 | -55 | - |
| Trimmed mean combination | 7.75 | 11.97 | 0.45 | 59 | 61 | 32 |
| DMSPE combination | 2.25 | 12.25 | 0.01 | 225 | 230 | -324 |
| Taylor rule | 3.38 | 11.33 | 0.09 | -331 | -341 | - |
| Monetary fundamentals | 3.37 | 12.23 | 0.08 | -398 | -395 | - |
| Purchasing power parity | 2.67 | 11.22 | 0.02 | -396 | -406 | - |
| Uncovered interest parity | 2.77 | 11.11 | 0.03 | -377 | -361 | - |
| NS relative level factor | 7.43 | 13.36 | 0.38 | -82 | -122 | - |
| NS relative slope factor | 3.01 | 10.79 | 0.06 | -269 | -258 | - |
| NS relative curvature factor | 9.73 | 13.25 | 0.55 | 220 | 245 | 35 |
| Factor from exchange rates | 2.85 | 13.33 | 0.03 | -473 | -445 | - |

**Notes:** The table reports the out-of-sample economic value generated by the Fundamentals-based Forecasting Method (FbFM) and the driftless random walk (RW). Using forecasts generated by each FbFM, a U.S. investor builds a maximum expected return strategy subject to a target portfolio volatility of vol = 10%. The strategy is based on dynamically rebalancing her portfolio on a monthly basis by investing in U.S. bonds and five foreign bonds. For each forecasting method, the table shows (i) the annualized mean return, $\overline{Rt}$, (ii) the annualized volatility (vol), (iii) the annualized Sharpe ratio ($SR$), (iv) the maximum performance fee ($pf$ in annualized $bps$) a risk-averse investor with quadratic utility would be willing to pay to use the corresponding method instead of the RW, (v) the excess premium return ($pr$ in annualized $bps$), and (vi) the break-even proportional transaction cost ($T^{be}$ in monthly $bps$) that offsets any positive performance fee obtained by using the FbFM. The superscripts $a$, $b$, and $c$ denote statistical significance at the 10%, 5%, and 1% level, respectively, based on the Ledoit and Wolf (2008) bootstrapped test-statistic of whether the Sharpe ratio of the FbFM is different from that of the RW. The forecasts are generated recursively for February 2000 to May 2013 for all methods, except the bumping method and DMSPE combination method (February 2005 to May 2013).

Second, the bagging method, the Nelson-Siegel relative factor, and to some extent the DMSPE combination method also display a recursive $R^2_{oos}$ that explains their corresponding global statistical result. To be precise, for a certain currency, each of these methods consistently yields either a negative local $R^2_{oos}$ for most of the OOS period or a positive local $R^2_{oos}$. Third, by contrast, the overall performance of BMA and BMS is influenced by observations before and after the 2008 financial crisis. Prior to the crisis, the recursive $R^2_{oos}$ is mostly positive for the CAD and the EUR in both methods, but becomes negative for almost all currencies following the crisis. The exception is perhaps the SEK for which BMA recovers the loss in forecast precision after January 2010.

To examine similar aspects for economic performance we compute the cumulative wealth that an investor would obtain at each OOS period. We focus on bumping versus the RW, juxtaposed to other strategies whose average Sharpe ratios are larger than the $SR$ of the RW and with positive performance fees. These include bagging, BMA, mean combination, Trimmed mean combination, and the Nelson-Siegel relative curvature factor. In all cases we set the initial wealth to \$10, which increases or decreases at the monthly portfolio return implied by the forecasting strategy. For comparability across methods we assume that the initial investment is made in February 2005, for a period of 100 months.

The cumulative gains are displayed in Figure 4.2. The plots confirm that by using bumping, the investor obtains greater economic value than the RW at each point in time. As well, the Nelson-Siegel relative curvature factor lead to higher wealth than the RW, an aspect somewhat masked by the average metrics reported in Table 4.3. Note, though, that the gains are more volatile in comparison with bumping and they decline steeply towards the end of the OOS period.

Regarding the other strategies, although they clearly outperform the RW over the OOS period, they fall behind bumping and the Nelson-Siegel relative curvature factor. Among them, BMA delivers the smallest gains over the OOS period. Putting together all the local, as well as the global statistical and economic evidence, we can conclude that the predictive ability of the bumping method is not ephemeral; rather, it is stable and robustly prevails over time.

### 4.3.4 Examining Bumping

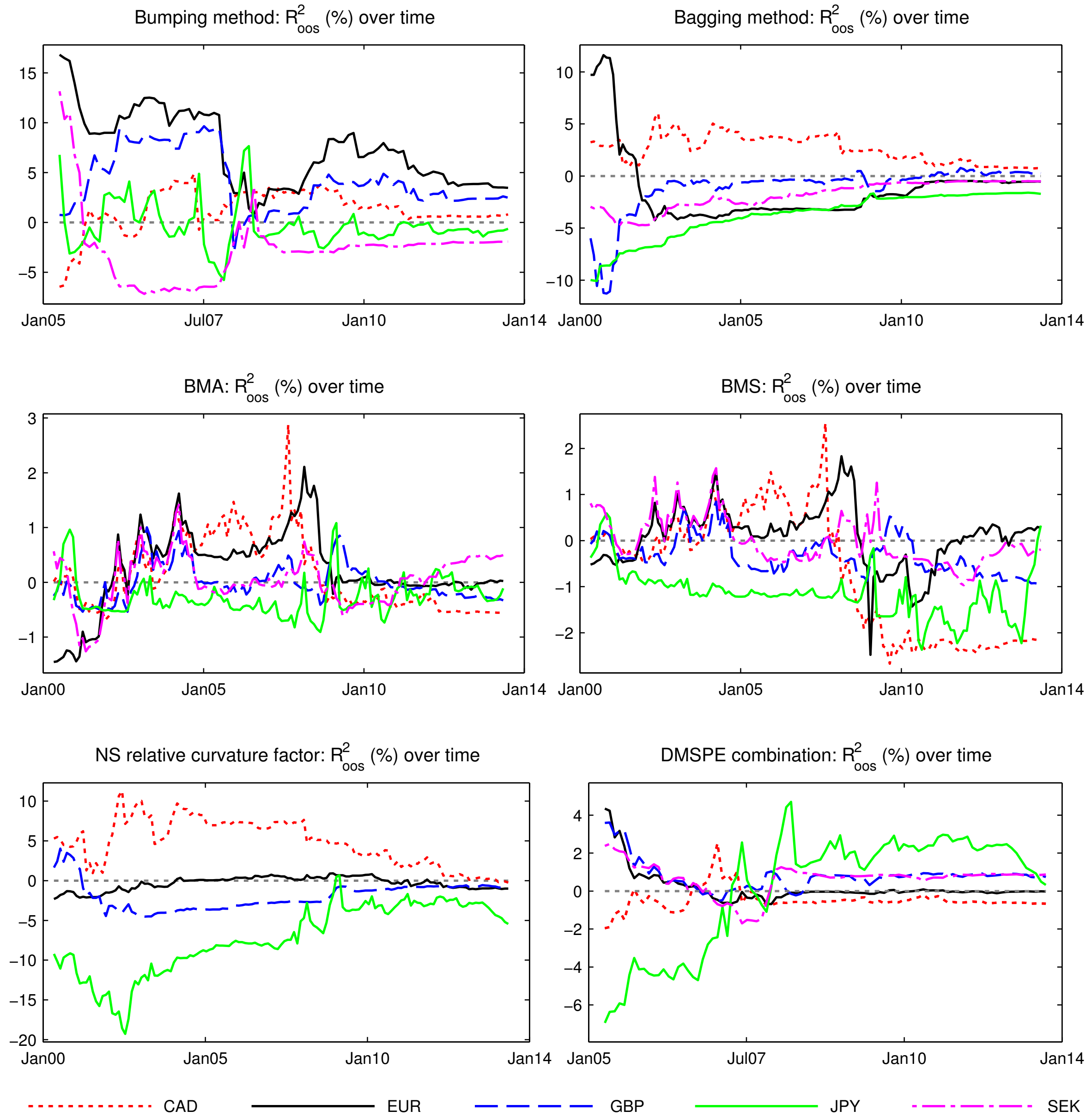
We now turn to examining why bumping yield better outcomes than the alternative approaches. In particular, we juxtapose bumping to bagging due to their mutual foundation in the bootstrap but distinct predictive ability.

#### Bootstrap and Unmodelled Data Dynamics

We start by analyzing the bootstrap's ability to wash away some noise in the data. Focusing on the GBP and the JPY as representative cases, Figure 4.3 compares box plots of the original data with those from bootstrap samples. For each box plot, the central mark is the median, the edges of the box are the $25^{th}$ and $75^{th}$ percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually. Points are drawn as outliers if they are larger than $qrt3 + w(qrt3 - qrt1)$ or smaller than $qrt1 - w(qrt3 - qrt1)$, where $qrt1$ and $qrt3$ are the $25^{th}$ and $75^{th}$ percentiles, respectively. We set $w = 1.5$, corresponding to approximately $+/-2.7\sigma$ and 99.3% data coverage assuming normality.

As evident in the Figure, the pseudo-data replications exclude some of the outliers

Figure 4.1: Forecasting Performance over Time



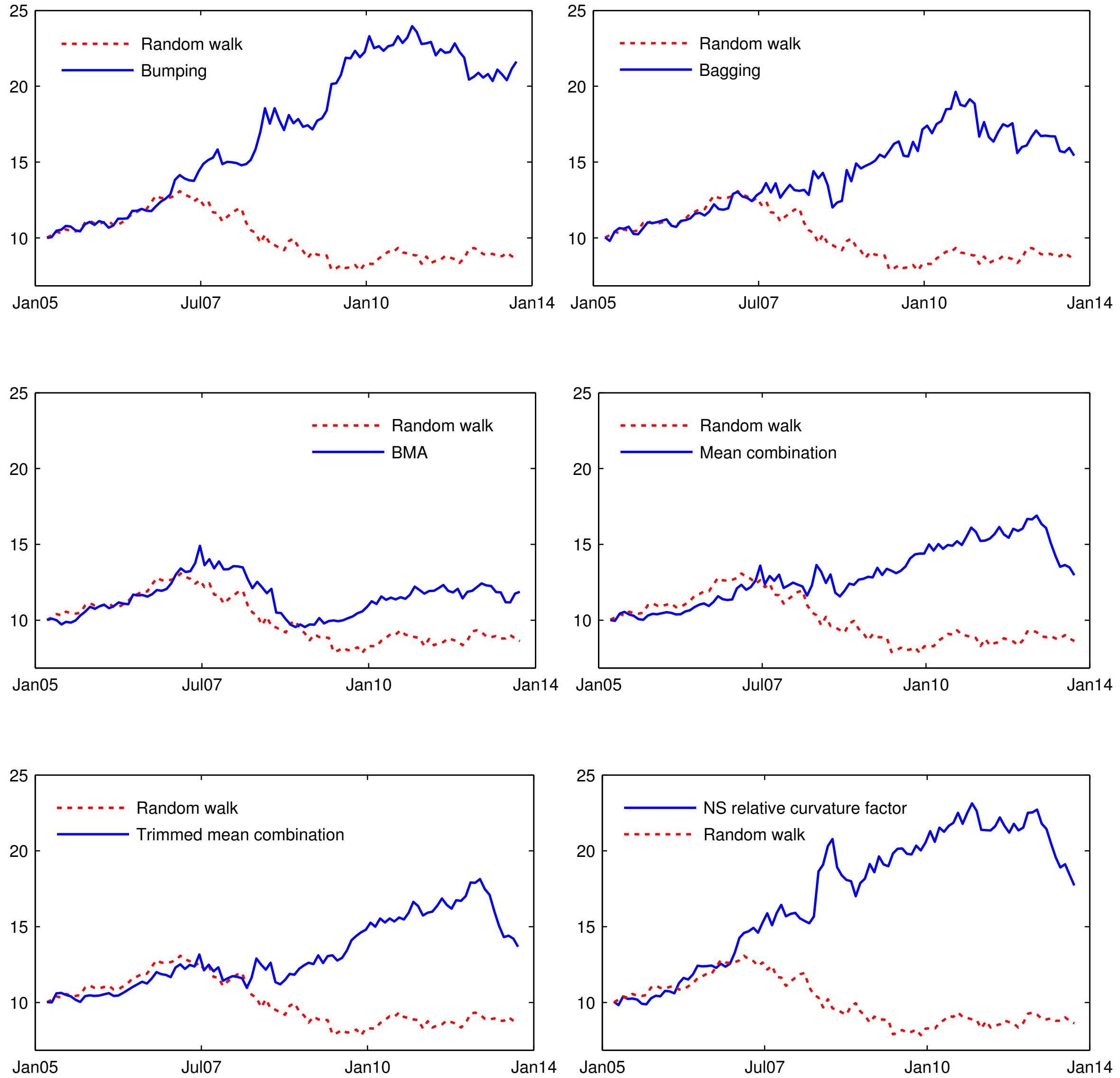

Notes: The figure shows the $R^2_{oos}$ metric computed recursively over time for selected forecasting methods. Positive (negative) values mean that the method forecasts better (worse) than the RW up to that point in time. The forecast evaluation period is February 2005 to May 2013 for bumping and DMSPE combination method. For the remaining methods is February 2000 to May 2013.

from the original sample. The data on exchange rate variations, fundamentals from the asymmetric Taylor rule (TRasy) and PPP, all exhibit fewer or no outliers when compared to the original sample. While this evidence supports the bootstrap's attribute to account for some unusual swings in the data, it falls short of explaining the differences in forecasting ability between bumping and bagging. As we noted, they both benefit from this attribute but, empirically, have a distinct performance.

### Bias and Variance of the Forecast Error

To further scrutinize the differences in performance we decompose the prediction error into bias and variance. For bagging, bias is the difference between the expected predic-

Figure 4.2: Cumulative Wealth Based on Returns Implied by Selected Forecasting Methods



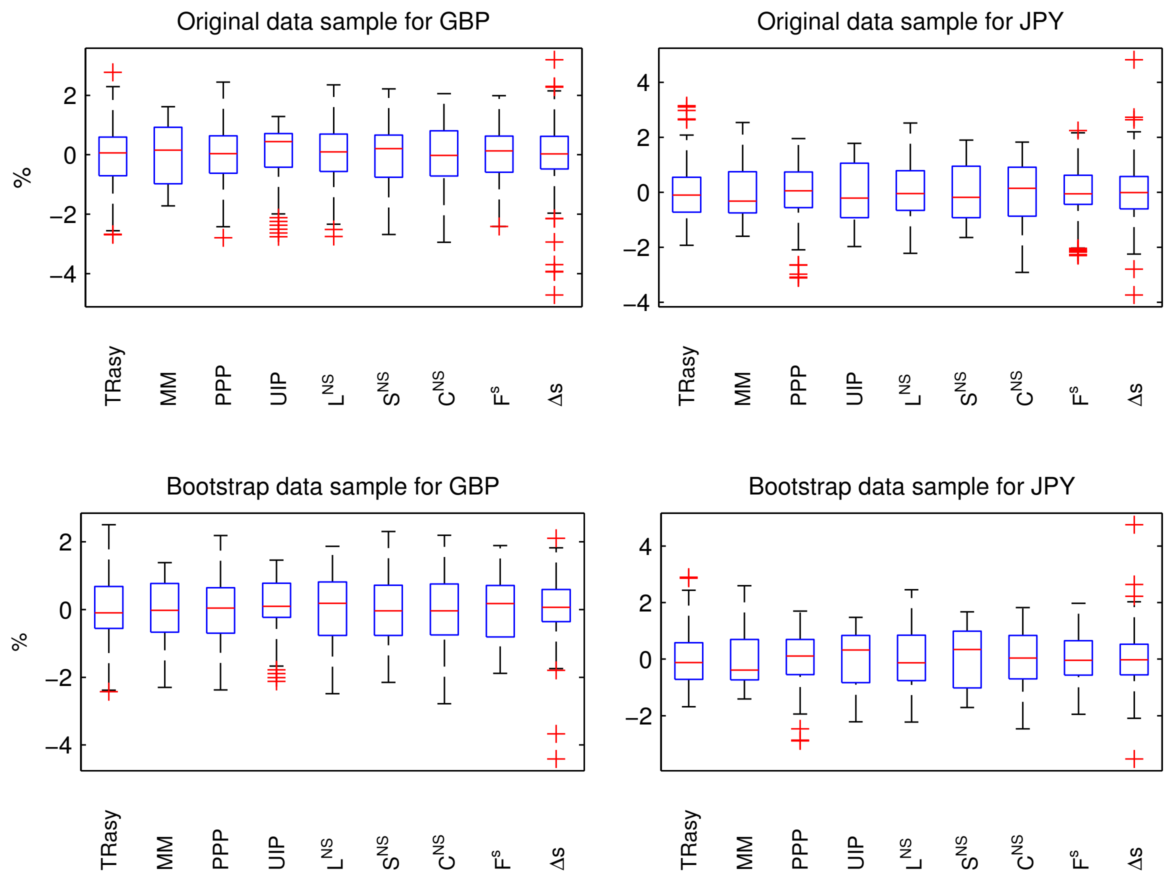

Notes: The figure shows the cumulative wealth associated with monthly returns implied by selected forecasting methods. The initial wealth at the beginning of the investment period (OOS) is set to \$10, which increases or decreases at the monthly portfolio return. To facilitate comparison across methods, the initial investment is made in February 2005.

tion (i.e., average across bootstrap replications) and the observed value of the exchange rate change. Variance is the variability of the model's forecast over different bootstrap replications and across the out-of-sample period.[14] For bumping, bias is proxied by the averaged difference between the observed value of the exchange rate variation and the forecast over the out-of-sample period.

Ideally, the best prediction method is one which reduces bias and variance to zero.

[14]Algebraically, given the true function $\Delta s = f(z) + error| \sim N(0, \sigma^2)$, the forecast error in predicting $\Delta s$ at a new data point $z^*$ with observed value $\Delta s^* = f(z^*) + error$, can be decomposed into bias and variance as: $Err(z^*) = \left(E[\widehat{f}(z^*)] - f(z^*)\right)^2 + E\left[\widehat{f}(z^*) - E[\widehat{f}(z^*)]\right]^2 + E\left[(\Delta s^* - f(z^*)^2)\right]$, where $Err(z^*)$ is the forecast error. The first term in the expression above is the squared bias; the second term is the variance; and the last term is the noise or irreducible error (Hastie et al., 2009).

Figure 4.3: Median, Percentile Ranges, and Outliers in the Original Data versus Bootstrap Samples




Notes: The figure shows box plots from the original sample *versus* box plots from bootstrap samples. On each box plot, the central mark is the median, the edges of the box are the $25^{th}$ and $75^{th}$ percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually beyond the whiskers. Points are drawn as outliers if they are larger than $qrt3 + w(qrt3 - qrt1)$ or smaller than $qrt1 - w(qrt3 - qrt1)$, where, $qrt1$ and $qrt3$ are the $25^{th}$ and $75^{th}$ percentiles respectively. The value of $w$ is set to 1.5 which corresponds to approximately $+/-2.7\sigma$ and 99.3% data coverage assuming normality. The samples for the box plots in the bottom plots correspond to a pseudo-data for a random data-point in the estimation window.

In practice, however, there is a trade-off between minimizing concomitantly bias and variance. In general, the variance tends to increase with the inclusion of additional variables in the model (additional model complexity), while the (squared) bias decreases with additional model complexity. The opposite occurs when model complexity is reduced. Hence, realistically, the method's ability to produce a minimum forecast error hinges upon achieving an adequate balance of model complexity by trading bias off with variance (Zou and Hastie, 2005).

Table 4.4 reports estimates of bias and variance, and repeats for convenience, the $R^2_{oos}$ statistics. The estimates are normalized to those of the RW such that for bias, values between $[-1, +1]$ indicate that the FbFM has less average bias than the RW. For variance, values below one mean that the method is superior to the RW in terms of accuracy in that the variation around the mean is narrow.

We notice immediately that for either method, the $R^2_{oos}$ is positive when the trade-off balance is attained or when the variance is sufficiently below one. Bumping, for

Table 4.4: Bias and Variance of the Forecast Error for Selected FbFM Relative to the RW

| | | CAD | EUR | GBP | JPY | SEK |
|---|---|---|---|---|---|---|
| Bias | Bumping | -0.220 | 16.077 | 1.650 | -8.959 | 4.578 |
| | Bagging | 0.383 | 2.023 | 4.560 | -0.499 | 0.679 |
| Variance | Bumping | 0.995 | 0.963 | 0.964 | 0.998 | 1.015 |
| | Bagging | 0.998 | 0.999 | 0.988 | 0.993 | 0.988 |
| $R^2_{oos}$ (%) | Bumping | $0.815^a$ | $3.477^a$ | $2.485^b$ | -0.637 | -1.919 |
| | Bagging | 0.738 | -0.536 | 0.279 | -1.704 | -0.505 |

**Notes:** Bias is the difference between the expected prediction (i.e., average across bootstrap replications for bagging) and the observed value of the exchange rate change. Values between -1 and +1 indicate that the Fundamental-based Forecasting Method (FbFM) has less average bias than the random walk (RW) benchmark. Variance measures the variability of the model's forecast (over different bootstrap replications for bagging). Values below +1 are consistent with a better average performance of the FbFM against the RW. For bumping, bias is proxied by the averaged difference between the observed value of the exchange rate variation and the forecast over the out-of-sample period. The Table also repeats for convenience, the $R^2_{oos}$ statistics for bagging and bumping from Table 4.1. The forecast evaluation period is February 2000 (February) to May 2013 for bagging (bumping).

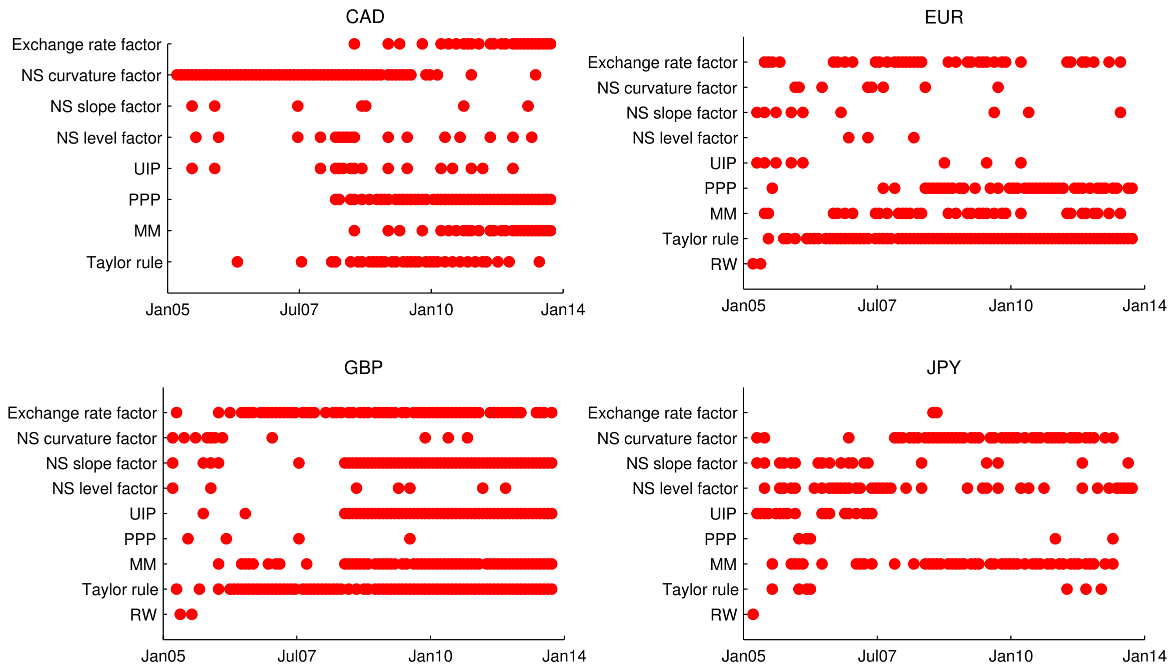
example, achieves the smallest magnitudes of variance for the three currencies it forecasts well. In some cases, such as for the EUR and the GBP, the reduction in variance occurs at the expense of high bias. Bagging, on the other hand, produces magnitudes of variance slightly below one for all currencies, but it leads to somewhat more complex models as reflected in the low bias. This suggests that the bumping's forecasts are generated from relatively parsimonious models, whereas bagging tends to over-fit the regressions. This evidence hints at the importance of the bumping's model validation stage, in light of the widespread evidence regarding the good in-sample fit of empirical exchange rate models which does not translate into out-of-sample forecasting power. Next we look at the severity of the model complexity in both methods.

### Fundamentals Selected in Bumping and Bagging

Our final check focus on the predictors selected in our bootstrap-based methods. In Figure 4.4 we plot the sets of fundamentals used to generate forecasts from bumping at each OOS period, whereas Figure 4.5 shows the frequency of selection of each fundamental in the bagging method. The illustrations concentrate on four currencies: the CAD, EUR, GBP, and JPY.

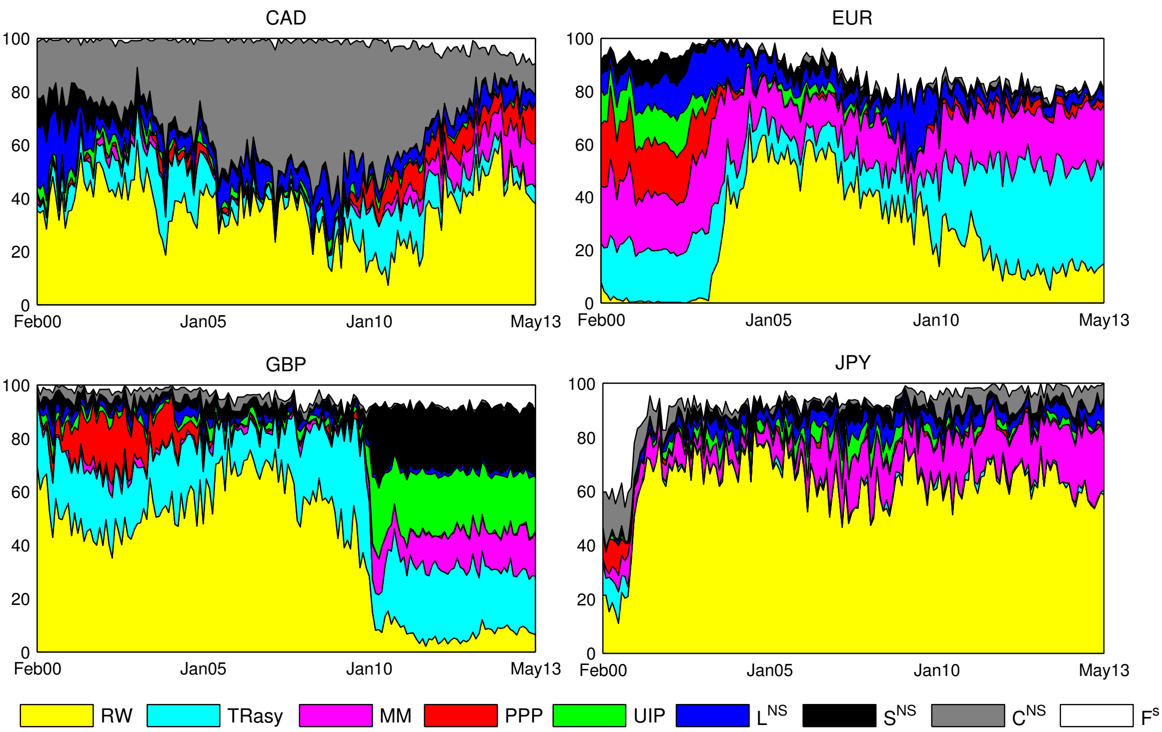
For all currencies, we observe three key aspects. First, bagging exhibits a high degree of time-variation in the predictors' selection frequency or, equivalently, frequency of inclusion in the forecasting model. For example, while monetary and PPP fundamentals are rarely included in the CAD bagging forecast before January 2010, they are picked by the method after this period. The bumping forecasts are also mostly generated from models that employ a combination of different sets of fundamentals at each point in time. The cases of the CAD, EUR, and JPY show this time-changing

Figure 4.4: Predictors Selected in the Bumping Method




Notes: The figure shows the predictors selected in the training step of the bumping method. The forecasts from bumping are generated from regressions based on these predictors. For example, the last forecast for the GBP is generated using parameters estimated from a regression that uses the Taylor rule, monetary fundamentals (MM), UIP fundamentals, the NS slope factor, and the exchange rate factor as regressors. The out-of-sample period runs from February 2000 to May 2013.

Figure 4.5: Frequency of Selection of Each Predictor in the Bagging Method




Notes: The figure shows the time-varying frequency of selection of each regressor in the bagging method over bootstrap samples. For each out-of-sample (OOS) period, it is computed as the number of times the regressor was pre-selected (and hence used to forecast) across 100 bootstrap replications. The OOS period runs from February 2000 to May 2013.

combinations, while for the GBP, there is less variability from June 2009. Note too that in both methods, there is one or more currency-specific fundamentals which are consistently included in the forecasting model.

Second, the RW is rarely selected in the bumping method, while in bagging it is prominent for all OOS observations and across bootstrap samples. For instance, the JPY bagging forecasts for each period are generated as averages of 100 forecasts of which more than half are from the RW. On the contrary, for bumping and the same currency only one forecast is generated from the RW.

Third, the forecasts from bumping are generated from relatively less complex models than bagging. In the cases of the CAD, EUR, and the JPY for example, the forecasting models from bumping often include up to three fundamentals at a time. This number increases to at most five for the GBP, but this is still below the average of six predictors included in bagging for most currencies. Our inference, therefore, is that bagging fails to deliver improvement in forecasting accuracy because the inclusion of many predictors at a time leads to a fairly stable forecasting procedure (see also Breiman, 1996). Bumping on the other hand benefits from the model validation in the training period, allowing it to robustly identify relatively parsimonious models with OOS predictive power.

## 4.4 Robustness Checks

We examine the robustness of our results in five dimensions, see Table 4.5. In all cases we concentrate on statistical measures for forecast evaluation, with significance levels assessed using the Clark and West (2006) test. First, we apply our pre-selection rules and model validation to the kitchen-sink (KS) regression using the single sample realization. In panel A, this defines a (i) KS regression with pre-selection akin to bagging, except that no averaging takes place, and a (ii) KS regression with model validation which is counterpart to bumping. Results indicate that using only the sample realization to pre-select and validate models does not lead to forecast improvements. In nearly all cases the RW dominates these regressions. This is a further piece of evidence supporting the benefits of bootstrapping in revealing informative fundamentals.

Table 4.5: Robustness Checks

| | CAD | EUR | GBP | JPY | SEK |
|---|---|---|---|---|---|
| Panel A: $R^2_{oos}$ (%), forecasts from pre-selected models in the sample realization | | | | | |
| KS reg. with model validation | -3.480 | -0.215 | -0.735 | -0.573 | -4.027 |
| KS reg. with pre-selection | -0.269 | 0.091 | 0.094 | -0.317 | -4.607 |
| Panel B: Bootstrap methods with cross-validation | | | | | |
| Bumping with cross-validation | -2.261 | $1.690^b$ | $0.498^a$ | 1.398 | 0.619 |
| Bagging with cross-validation | -0.460 | -0.543 | 0.043 | -0.167 | 0.055 |
| Panel C: Forecast evaluation based on relative RMSFE | | | | | |
| Bumping | $0.996^a$ | $0.982^a$ | $0.987^b$ | 1.003 | 1.010 |

**Table 4.5 Continued**

| | CAD | EUR | GBP | JPY | SEK |
|---|---|---|---|---|---|
| Bagging | 0.996 | 1.003 | 0.999 | 1.008 | 1.003 |
| Panel D: $R^2_{oos}$ (%), forecasting with rolling regressions | | | | | |
| Bumping | -5.654$^a$ | 3.460$^a$ | -2.763 | 1.514 | 0.416 |
| Bagging | 2.409$^b$ | -4.590 | -3.701 | -1.923 | -2.802 |
| Kitchen-sink regression | -10.934$^b$ | -12.763 | -16.009 | -29.139 | -9.962 |
| Mean combination | -0.558 | -1.526 | -1.024 | 0.338 | -0.790 |
| Median combination | -1.039 | -0.791 | -1.227 | -0.226 | -0.146 |
| Trimmed mean combination | -0.768 | -1.116 | -1.053 | 0.015 | -0.298 |
| DMSPE combination | -1.170 | -1.584 | -1.174 | 0.941 | -0.336 |
| Taylor rule | -1.273 | -0.335 | -3.872 | -3.249 | 0.193 |
| Monetary fundamentals | -2.692 | -3.902 | -2.157 | -3.928 | -3.373 |
| Purchasing power parity | -1.651 | -0.786 | -2.672 | -3.752 | 0.771$^a$ |
| Uncovered interest parity | -1.380 | -3.941 | -3.526 | -0.993 | -5.531 |
| NS relative level factor | -1.218 | -3.412 | -1.574 | 2.257$^b$ | - |
| NS relative slope factor | 0.045 | -4.859 | -3.409 | -0.596 | - |
| NS relative curvature factor | -1.892 | -1.733 | -2.356 | -5.046 | - |
| Factor from exchange rates | -2.188 | -1.552 | -2.939 | -1.971 | 0.213 |
| Panel E: $R^2_{oos}$, 3-months ahead forecasting horizon | | | | | |
| Bumping | -6.998 | -11.043 | -21.421 | -8.260 | -12.630 |
| Bagging | 1.393$^b$ | -23.859 | -10.258 | -18.890 | -5.826 |
| Kitchen-sink regression | -10.342 | -17.950 | -34.932 | -37.010 | -8.102 |
| Bayesian model averaging | 5.341$^c$ | 7.626$^c$ | 9.248$^b$ | 7.452$^c$ | 11.678$^c$ |
| Bayesian model selection | 7.071$^b$ | 8.853$^c$ | 9.480$^b$ | 7.155$^b$ | 11.753$^c$ |
| Mean combination | -1.428 | -1.178 | 1.219 | 1.316 | 0.049 |
| Median combination | -2.061 | -0.066 | -2.171 | 0.676 | 2.257$^a$ |
| Trimmed mean combination | -1.828 | -0.175 | 1.005 | 1.236 | 1.387 |
| DMSPE combination | -1.475 | -0.468 | 3.695$^a$ | 1.757 | 3.592$^b$ |
| Taylor rule | -1.412 | 2.728$^a$ | -12.364$^a$ | -3.906 | 0.443 |
| Monetary fundamentals | -9.789 | -11.775 | -8.140 | -4.949 | -11.215 |
| Purchasing power parity | -2.491 | 0.856 | -7.151 | -5.301 | 1.856 |
| Uncovered interest parity | -3.001 | -2.838 | -1.713 | -2.829 | -2.427 |
| NS relative level factor | -3.417 | -3.231 | -3.760 | 0.909 | - |
| NS relative slope factor | -3.686 | -3.391 | -2.563 | -1.294$^a$ | - |
| NS relative curvature factor | -0.208$^a$ | -2.177 | -8.734 | -6.312 | - |
| Factor from exchange rates | -5.038 | -4.217 | -6.003 | -2.274 | 1.220 |

**Notes:**: The table presents results for four robustness checks. Panel A shows the $R^2_{oos}$ for a forecasting approach that consists in pre-selecting models from the sample realization instead of bootstrap samples. Positive values mean that the FbFM generates a lower MSFE and therefore forecasts better than the RW. Panel B reports $R^2_{oos}$ for bumping and bagging with cross-validation. Panel C focus on evaluating forecasting performance using the relative RMSFE, i.e., $RMSFE_{FbFM}/RMSFE_{RW}$. Values below one imply better forecasting of the FbFM relative to the RW. Panel D shows the $R^2_{oos}$ for forecasts generated with rolling regressions using a window of 10 years rather than recursive regressions. Panel E reports forecasting performance based on the $R^2_{oos}$ at 3-months horizon. In all panels, the superscripts $a$, $b$, and $c$ denote statistical significance at the 10%, 5%, and 1% level, respectively using the Clark and West (2006) one-sided $t$-statistic.

Second, we check if our results from bumping and bagging are driven by our choice of a pre-selection rule that relies on in-sample results, like our $t$-statistic. Sarno and

Valente (2009), for instance, argue that such selection criteria are unable to pick models which optimally use the information in the fundamentals. In particular we replace the $t$-statistic rule in the pre-selection step by a K-fold cross-validation procedure. For each model, this consists in splitting the $T_0$ in-sample observations into $K = 6$ roughly equal-sized parts, and fitting the model to each of the $K - 1$ part while using the left out part to compute the training error. We then use the model with the smallest training error to generate bagging forecasts or further validate the models across replications before generating bumping forecasts. The results from these cross-validated procedures are reported in Panel B. As shown, the results remain unfavorable to bagging. The findings from bumping robustly prevail, yielding positive $R^2_{oos}$ for four of the five currencies, of which two are statistically significant.

Third, we report the forecasting performance of bumping and bagging using the usual relative RMSFE (see, e.g, Byrne et al. 2016). That is we construct the ratio $\text{RMSFE}_{FbFM}/\text{RMSFE}_{RW}$. Values below one are consistent with better forecasting performance of the FbFM relative to the RW. As expected, the results in Panel C mimic those from the $R^2_{oos}$. The relative RMSFE are significantly below one for three currencies in bumping, whereas in bagging are insignificantly below one for two currencies.

Fourth, we experiment with generating forecasts with rolling regressions in a window of 10 years. Our findings in Panel D indicate that bumping is the by far the best forecasting method. It improves upon the RW for four of the five currencies, with the improvement being statistically significant for two currencies. The remaining methods are dominated by the RW.

Finally, we examine the forecasting performance at 3-months horizon. As shown in Panel E, none of our bootstrap-based methods outperforms the RW. Our conjecture for this unfavorable result is related to the behavior of asset prices like exchange rates. For this class of prices, random fluctuations in the data are expected to be heightened at short-horizons; see Dangl and Halling (2012). The bootstrap, therefore, seems to be useful at the 1-month horizon, as it accounts for these abnormal fluctuations in the data. Bayesian methods, however, deliver an outstanding performance at 3-months horizon, with positive and significant $R^2_{oos}$, of magnitudes above 5% in all cases. More generally, most of the forecast combinations methods yield positive $R^2_{oos}$, even though they are statistically significant only in few cases. These latter findings are common in the exchange rate literature (see, for example, Byrne at al. 2014 or Rossi, 2013).

All in all, we uncover that bumping has strong predictive power for exchange rates, especially at horizons for which predictability has been difficult to establish, such as the 1-month horizon. The evidence appears to be robust to using alternative metrics for forecast evaluation, different pre-selection rules, and distinct forecasting approaches.

## 4.5 Conclusion

The empirical evidence shows that exchange rate fundamentals have time varying predictive ability (Cheung et al., 2005; Rossi, 2013). This chapter applies bootstrap-based methods to robustly reveal fundamentals that carry predictive information at each point in time. We consider the bagging and bumping procedures, which despite their common foundation in the bootstrap, are designed to handle prediction problems of distinct nature. Bagging works best for unstable forecasting procedures, whereas bumping has the potential to produce better outcomes in problems for which finding the optimum forecasting model is challenging.

Using a variety of metrics, our results show the usefulness of bumping in uncovering economic fundamentals with predictive power for exchange rates. Conditioning on these fundamentals leads to significant forecasts improvements upon the driftless random walk (RW) benchmark and several competing approaches. These include: (i) simple linear regressions conditioned on each of the fundamentals we consider, (ii) combination methods based on the mean, the median, the trimmed mean, and the discounted mean squared prediction error, (iii) Bayesian model averaging or selection, and (iv) the kitchen-sink regression of Welch and Goyal (2008). In addition, an investor who uses the bumping method's predictions to optimally rebalance his portfolio can obtain economic gains above those provided by a strategy based on the RW.

When examining the characteristics underlying the performance of bumping, we find that it hinges upon the ability to pin down parsimonious models. Bagging, on the contrary, tends to reveal complex models which often over-fit in-sample but with weak out-of-sample forecasting power. We also find that in our bootstrap-based methods, while there is one or more currency-specific fundamentals which are consistently included in the forecasting model across the out-of-sample period, the exact combination of fundamentals changes slightly over time.

# Chapter 5

# Forecasting Commodity Currencies: The Role of Fundamentals with Short-Lived Predictive Content

## 5.1 Introduction

The analyses in the previous chapters reveal some of main challenges involved in predicting exchange rates with macroeconomic fundamentals. The challenges are particularly hard to overcome at short-horizons as there are a host of issues, including for instance, estimation uncertainty, parameter instability, identification of the relevant country-specific fundamentals, among others.

In a recent paper, however, Ferraro et al. (2015) state that the frequency of the data used in the predictive regressions for exchange rates may be important in pinning down forecasting ability. Based on empirical and Monte Carlo evidence they assert that the most probable reason for the failure to uncover predictive power in fundamentals such as commodity prices, is the reliance on data sampled at low frequency. At this frequency, the predictive content of this sort of fundamental is transitory. In fact, using monthly and quarterly data on oil price changes to predict fluctuations in exchange rates at similar frequencies, Ferraro et al. (2015) hardly detect predictive content (see also Chen et al. 2010 for congruent results). When instead they regress the contemporaneous daily change in the exchange rate on the current daily fluctuation on oil prices, they find a significant and consistent relationship. The relationship is short-lived, in the sense that it can mainly be detected using high frequency data and it washes away quickly.[1]

In this chapter we employ a systematic approach to exploit the short-lived effect of commodity prices on exchange rates, in a pseudo out-of-sample context. In a major break with the existing exchange rate studies, we allow the effect of daily fluctuations in

[1] In a Monte Carlo simulation, Ferraro et al. (2015) show that if exogenous commodity price spikes are occasional events, and exchange rates are contemporaneously related to them, one may be able to pin down out-of-sample predictability in daily but not monthly or quarterly data.

commodity prices to carry on to the end-of-month change in the exchange rate. Using the so-called MIxed DAta Sampling (MIDAS) framework, each daily observation on price fluctuations can have a different weight or impact on the end-of-month observation on the exchange rate change. In this way, we can fully take advantage of the predictors' sampling properties. The MIDAS regression is a simple, parsimonious, and flexible modeling approach that allows the variables entering a time series regression to be sampled at different frequencies (Ghysels, 2007). For instance, fluctuations at the end of the month can have more predictive power than fluctuations further back. With our approach we can attribute more importance to these observations that are closer in time, while the literature typically would aggregate them to the lowest frequency with equal weights. Aggregating, therefore, dampens down the short-lived effects, whereas our MIDAS approach can potentially pin them down.

Further, the empirical literature also suggests that the predictive content of either commodity prices or the standard macroeconomic fundamentals is momentary. Ferraro et al. (2015), for example, find that lagged daily fluctuations in oil prices were better predictors of daily changes in the Canadian-U.S. dollar exchange rate around 2006-2007, while at the monthly and quarterly frequencies predictive ability is never found. Using fundamentals derived from uncovered interest rate parity, Giacomini and Rossi (2010) detect predictability for the British pound-U.S. dollar exchange rate in the late Eighties but not in the Nineties. To account for these issues, our MIDAS predictive models allow for changing sets of high and/or low frequency regressors at each period in time.[2]

In this setting, we make two additional contributions. First, we use a likelihood-based approach to shed light on whether regressors sampled at high frequency are more informative about monthly changes in exchange rates than predictors sampled at low frequency. Second, we equally employ the likelihood information to account for potential time-variation in the predictive content of our predictors (i.e., commodity prices and standard macroeconomic fundamentals). Hence, we can forecast with the predictors with the highest support from the data at each period in time. Alternatively, we can compute the forecast as a weighted average of each model's forecast. In this methodical manner, we can analyze if accounting for the time-changing predictive ability improves forecast accuracy. More generally, in our framework, we examine if the predictive content of fundamentals sampled at high and low frequency should be regarded as complementary, rather than substitute of one another.

All our models are estimated with Bayesian methods and we introduce the random walk Metropolis-Hastings (MH) algorithm as a tool for estimating MIDAS regressions. The main advantage of using Bayesian methods over the classical methods is their provision of a systematic framework to incorporate model and parameter uncertainty, by focusing on the full predictive density; see, for example, Pettenuzzo et al. (2015).

[2]See also an explanation for the time-variation in predictive content based on a scapegoat theory of exchange rates of Bacchetta and van Wincoop (2004, 2013) or the empirical evidence in Fratzscher et al. (2015).

They also allow us to systematically achieve our goals of (i) examining the degree of informativeness of predictors sampled at different frequencies and (ii) accounting for time-variation in forecasting performance.

We focus on commodity prices/currency pairs of three major commodity exporting countries: (a) Australia with emphasis on gold and copper prices; (b) Canada, concentrating on oil and copper prices, and (c) Norway with oil and gas prices. In addition, following the indication in Chen et al. (2010) that commodity price movements may induce exchange rates fluctuations for large commodity importers, we examine the case of Japan - focusing on oil prices and a commodity price index, as an example of this category of countries. Overall, while our left-hand side variable is always sampled at monthly frequency, on the right-hand-side our regressions allow for commodity prices sampled at daily or monthly frequency, or standard macroeconomic predictors sampled at monthly frequency, such as interest rate differentials, money supply differentials, and price differentials.

Using monthly and the corresponding daily data from 1986M9 to 2014M3, we forecast recursively the period-ahead change in the exchange rate at one and three-months horizons. Our forecasts are compared to those of the driftless Random Walk, which according to Rossi (2013), is the most appropriate benchmark in the exchange rate literature. To assess our methods' forecasting performance we employ the root mean squared forecast error (RMSFE) for point forecasts, and log-score differentials for density forecasts. We examine the statistical significance of our forecasts using the Clark and West (2006, 2007) test-statistic.

In our findings, exploiting the properties of daily commodity prices in a MIDAS setting leads to forecast improvements. In terms of point forecasts, for example, MIDAS regressions with daily copper prices improve upon the RW benchmark for the Australian dollar and the Canadian dollar, especially at the 1-month forecasting horizon. In contrast, and consistent with the existing evidence, standard regressions with commodity prices sampled at low frequency hardly improve upon our benchmark.

Regarding density forecasts, our results suggest that once we account for the full forecast distribution, the RW never forecasts better than the commodity or fundamentals-based regressions. As well, when we account for time-variation in forecasting ability by combining the forecasts from our MIDAS models and from regressions based on standard macroeconomic fundamentals we also improve upon the RW. Inspection of the data-driven weights underlying our forecast combinations approaches reveals that daily commodity prices are relatively more informative about changes in exchange rates at 1-month horizon than monthly commodity prices or the typical macroeconomic variables. As the forecast horizon increases to 3-months ahead, the role of macroeconomic fundamentals becomes more prominent.

The chapter proceeds as follows. In the next Section we use a small simulation example to illustrate why a MIDAS setup might be appropriate to pin down the forecasting ability of daily commodity prices. In Section 5.3 we detail our methodology, including our new contribution in terms of estimation of MIDAS regression using the random walk Metropolis-Hastings algorithm. Section 5.4 reports our main results, while in Section 5.5 we conduct robustness checks. Section 5.6 concludes.

## 5.2 A Small Simulation Example

To visualize the importance of considering our approach, Figure 5.1 shows a simulation exercise based on the moments (mean and standard deviation) taken from the data we consider in our empirical section. The econometric procedure underlying the simulation is detailed in the methodological section - Section 5.3.

In Panel A we assume a Data Generating Process (DGP) in which monthly fluctuations in the exchange rate, $\Delta s_t$, are driven by a daily-frequency variable, denoted $x_{td}$, where each observation is allowed to have a different effect on $\Delta s_t$. The simulated DGP is:[3]

$$\Delta s_t = -0.001 + 0.50 \times [f(\theta_1, \theta_2)(x_{td21} + x_{td20} + ... + x_{d1})] + \varepsilon_t, \tag{5.1}$$

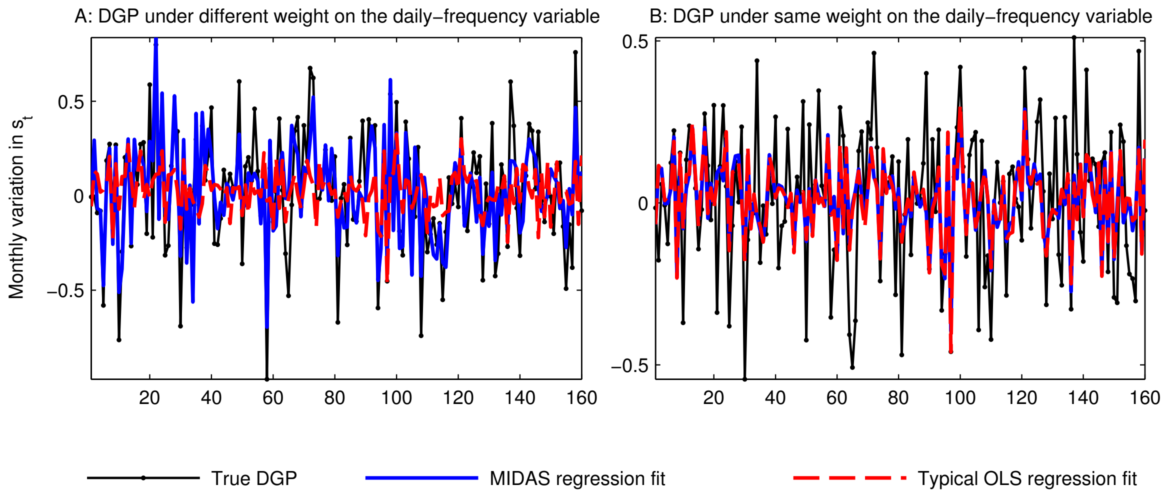
where $\varepsilon_t$ is an i.i.d. error term with $\mathrm{var}(\varepsilon_t) = 0.11$. The subscript $d()$ attached to $x_t$ indicates the occurrence of the daily observation in a month. Essentially, we consider 22 working days within a month. Further, we assume that the previous 21 daily observations affect the value of $\Delta s_t$. The function, $f(\theta_1, \theta_2)$, is a polynomial that allows us to smooth the past daily observations on the basis of the two parameters. We set these parameters to $\theta_1 = 0.3$ and $\theta_2 = -0.1$, implying that for this DGP, observations close to the end of the month have higher impact on $\Delta s_t$ than those at the beginning of the month. Subsequently, we simulate 160 monthly data points corresponding to 3520 daily observations and fit a MIDAS regression and a typical linear regression. Note that the latter regression imposes equal weights on the daily observations.

The panel illustrates how well we can fit the true DGP if we use the usual constant weighting scheme, as opposed to the MIDAS approach. As depicted, imposing equal weights on the effects of the daily variable results in a relatively poor fit. There are larger mismatches between the true DGP and the line fitted with the constant weighting approach, especially in the high and low spikes. In contrast, the line fitted with the MIDAS approach recovers reasonably well the true DGP. Accordingly, the adjusted R-squared is 0.63 in the MIDAS setup, whereas in the usual weighting scheme is 0.14.

In Panel B we consider a DGP in which the monthly fluctuations in the exchange rate are driven by a daily-frequency variable where each observation has the same weight on $\Delta s_t$. Consistent with this assumption, we set $\theta_1 = \theta_2 = 0$ in Eq. (5.1), and keep the remaining features unchanged. In the panel we inspect how well we can fit the true DGP if we use the MIDAS approach instead of the typical constant weighting scheme (linear regression). The graph shows that the MIDAS approach is as well suited

[3]The DGP is a MIDAS model based on the exponential Almon lag polynomial with the parameters that determine the weights defined by $\theta_1$and $\theta_2$. A complete description of this type of model is given in Section 5.3.

Figure 5.1: Fitting a MIDAS and a Typical OLS Regression to Daiy-frequency Variables




Notes: The Figure shows to what extent we can recover a simulated Data Generating Process (DGP) using a MIDAS model as opposed to a typical linear regression. In Panel A we assume a DGP in which monthly fluctuations in the exchange rate, $\Delta s_t$, are driven by a daily-frequency variable whose weight on $\Delta s_t$ is allowed to differ. The parameters that determine the weights (see Eq. (5.1)) are set to $\theta_1 = 0.3$ and $\theta_2 = -0.1$, implying that, for the DGP in Panel A, observations close to the end of the month have higher impact on $\Delta s_t$ than those at the beginning of the month. We fit the data from this DGP to a MIDAS regression and a typical linear regression (which imposes equal weight on the daily observations). In Panel B, we equally fit the two regressions, but assuming a DGP in which the monthly fluctuations in the exchange rate are driven by a daily-frequency variable where each observation has the same weight on $\Delta s_t$. Hence, under this DGP, $\theta_1 = \theta_2 = 0$. The Figure shows that the MIDAS regression is well-suited to capture the properties of the simulated DGPs, whereas the linear regression fits poorly when the DGP posits different weights on the daily frequency variable (Panel A).

as the same weighting scheme to recover the true DGP; the two lines fitted to the data overlap, and the adjusted R-squared is approximately 0.27 in the two setups. As it will become clearer in the next section, the line fitted under the MIDAS approach coincides with the one based on the linear regression because the weights on the daily observations are also estimated from the data. And in this case, under the MIDAS approach the estimates of the parameters that determine these weights happened to be zero - consistent with equal weighting scheme and hence the true DGP.

All in all, our simulation suggests that, at least in-sample, the MIDAS approach is well-suited to capture the properties of the unknown true DGP. Whether the better in-sample fit will translate into a better out-of-sample forecast depends on the existence of predictive content of daily commodity prices on monthly exchange rates. Our first aim is to use the MIDAS approach to pin down the properties of commodity prices at daily-frequency and examine their predictive power for commodity currency exchange rates at monthly frequency.

## 5.3 Methodology

### 5.3.1 Predictive MIDAS Model

In our empirical analysis we are firstly interested in forecasting the $h$-month-ahead change in spot exchange rate using a predictor sampled daily. The usual procedure in this case would be to aggregate the daily-frequency data to match the frequency of the low-sampled variable. As shown in Section 5.2, this aggregation might result in a poor fit, as well as loss of the properties of the data and econometric estimation issues related to inconsistent estimators (see, Andreou et al. 2010). To potentially avoid these issues, a MIDAS regression allows mixing variables sampled at different frequencies. A simple MIDAS regression for our forecasting problem is:

$$\Delta s_{t+h} = \beta_0 + \beta_1 B_1(L^{1/m}; \theta_1) x_t^{(m)} + \varepsilon_{t+h}; \quad \varepsilon_{t+h} \sim N(0, \sigma^2), \tag{5.2}$$

where

$$\beta_1 B_1(L^{1/m}; \theta_1) \equiv B(L^{1/m}; \theta) = \sum_{k=0}^{K-1} B(k; \theta) L^{k/m}, \tag{5.3}$$

for $t = 1, ..., T-h$, and $h = 1, 3$. In Eq. (5.2), $\Delta s_{t+h}$ is the period-ahead change in the spot exchange rate at monthly frequency. Our daily regressor, denoted $x_t^{(m)}$, is sampled $m$ times between $t$ and $t{+}1$, and $m = 22$ assuming that there are always 22 observations within a month.[4] The key ingredient in the MIDAS model is the polynomial function, $B(L^{1/m}; \theta)$, which allows to smooth $K$ past observations of $x_t^{(m)}$ on the basis of a few number of parameters $\theta = (\theta_0, \theta_1, ..., \theta_p)$, where $p + 1 << K$. In this function, $L^{k/m}$ is a lag operator such that $L^{1/m} x_t^{(m)} = x_{t-1/m}^{(m)}$, i.e., we denote lags of $x_t^{(m)}$ by $x_{t-j/m}^{(m)}$. Once the parameters of this function are obtained, the effect of past values of $x_t^{(m)}$ on $\Delta s_{t+h}$ is captured by $\beta_1$. Models of this type are also used by Pettenuzzo et al. (2015) and Ghysels et al. (2007).

To gain insights on these concepts, consider for instance that a time $t$ monthly change in the exchange rate is affected by the previous 21 daily observations of $x_t^{(m)}$. Without using a smoothing function or restricting the parameters in $B(L^{1/m}; \theta)$, we would have to include $K = 21$ daily lags in Eq. (5.2) and estimate $21 + 2$ parameters. Instead, in a MIDAS regression the smoothing function, $B(L^{1/m}; \theta)$, uses fewer parameters (two in our application).

We can extend the model in Eq. (5.2) to include $n$ other regressors, $\mathbf{z}_t = (z_{1t}, ..., z_{nt})'$, sampled at the same frequency as $\Delta s_{t+h}$:

$$\Delta s_{t+h} = \beta_0 + \beta_1 B(L^{1/m}; \theta_1) x_t^{(m)} + \gamma' \mathbf{z}_t + \varepsilon_{t+h}, \tag{5.4}$$

[4] To create balanced monthly observations we assume the following. First, for months with less than 22 observations, we consider that the observation in the last working day of the previous month extends to one day before the first working day of the current month. If this does not complete 22 days, we further posit that the last observation of the current month is valid for one extra day. Second, for months with more than 22 days, typically 23, we average the first two daily observations.

where $\gamma$ is a vector of $n$ coefficients associated with $\mathbf{z}_t$. The model in Eq.(5.4) nests two specifications that we consider in our empirical work: (i) a MIDAS model if we exclude the predictors in $\mathbf{z}_t$ and (ii) a typical linear regression, if we exclude the daily ($x_t^{(m)}$) variables and forecast the monthly change in the exchange only with commodity prices or macroeconomic fundamentals sampled at the same frequency as $\Delta s_{t+h}$.

To complete the specification of the MIDAS regression we need to define the functional form of the polynomial $B(L^{1/m};\theta)$. While several alternatives exist, and the adoption of any particular depends on the application at hand, we employ the exponential Almon lag polynomial following Ghysels et al. (2007):

$$B(k;\theta) = \frac{e^{(\theta_1 k+\theta_2 k^2)}}{\sum_{i=1}^{K} e^{\theta_1 i+\theta_2 i^2}}, \tag{5.5}$$

with $\theta = (\theta_1, \theta_2)$. This polynomial is flexible enough to take various shapes for different values of its parameters, $(\theta_1, \theta_2)$, and Ghysels et al. (2005) have found it to work well in practice.[5] If we consider that only the past 21 trading days affect the value of $\Delta s_{t+h}$, then under this polynomial Eq. (5.2) is a compact representation of:

$$\Delta s_{t+h} = \beta_0+\beta_1 \left( \frac{e^{(\theta_1\times 1+\theta_2\times 1^2)}}{\sum_{i=1}^{K} e^{\theta_1 i+\theta_2 i^2}} x_{td21} + \frac{e^{(\theta_1\times 2+\theta_2\times 2^2)}}{\sum_{i=1}^{K} e^{\theta_1 i+\theta_2 i^2}} x_{td20} + \ldots + \frac{e^{(\theta_1\times 21+\theta_2\times 21^2)}}{\sum_{i=1}^{K} e^{\theta_1 i+\theta_2 i^2}} x_{td1} \right) + \varepsilon_{t+h}. \tag{5.6}$$

This MIDAS regression is non-linear, requiring non-linear methods for estimation. We focus on an appropriate algorithm to implement in the next subsection.

### 5.3.2 Bayesian Estimation and Forecasting

We use Bayesian methods to estimate the parameters of our regressions. The major advantage of Bayesian techniques over the typical frequentist methods is the possibility of accounting for model and parameter uncertainty. This is achieved by obtaining the full predictive density, rather than a point forecast underlying the frequentist approach. As we elaborate next, in a Bayesian setup we can also combine forecasts in a more systematic fashion.

To describe the mechanics of our novel MIDAS estimation techniques with a simple notation, first express Eq. (5.6) in the following functional form:

$$S = f(X,\gamma)+\varepsilon, \qquad \varepsilon \sim N(0,\frac{1}{\eta}), \text{ and } \qquad \frac{1}{\eta} = \sigma^2; \tag{5.7}$$

where we have suppressed, for notational simplicity, the dependence on the forecast horizon $h$ and time $t$. Moreover, $f(.)$ indicates that our function of interest depends on

[5] In our empirical exercise we also experimented with the unrestricted MIDAS approach of Foroni et al. (2013). In unreported results we find that forecasts based on this approach were generally less precise than the RW benchmark. A possible explanation for this weak performance might be the loss of precision in parameter estimates, since in this approach and given our daily-frequency predictors, a relatively large number of parameters have to be estimated.

the data ($X$) and parameters in $\gamma$, where $X$ contains our daily predictors ($x_{td}$), and $\gamma$ includes the parameters $\beta_0, \beta_1, \theta_1, \theta_2$.

As usual in a Bayesian framework, estimation involves definition of prior distributions, the likelihood function, and the posterior distribution. We use independent Normal-Gamma priors. As such, the prior for $\gamma$ is independent of the prior for $\eta$ and is defined as:

$$\gamma \sim N(\underline{\gamma}, \underline{V}). \tag{5.8}$$

For the error precision, $\eta$, the prior is:

$$\eta \sim G(\underline{s}^{-2}, \underline{\nu}). \tag{5.9}$$

We set $\underline{\gamma} = (0,0,0,0)'$, $\underline{V} = 0.35I$, $\underline{\nu} = 1$, and $\underline{s}^{-2}$ is based on OLS estimate of Eq. (5.2) assuming that the data is aggregated to the monthly frequency under the constant weighting scheme. All these choice of priors are sensible but relatively diffuse. For instance, the elements of the prior mean in $\underline{\gamma}$ incorporate the view that the driftless Random Walk model provides better exchange rate forecasts. At the same time, the prior variance, $\underline{V}$, allows the coefficients estimates to wander in the region $[-1.2, 1.2]$ with 95% prior probability assuming normality. We further note that only data available up to the beginning of our first forecast are used to estimate any data-based quantity such as $\underline{s}^{-2}$.

If we combine these priors with the likelihood we obtain the following conditional posterior for $\eta$:

$$p(\eta|S,\gamma) \sim G(\overline{s}^{-2},\overline{\nu}), \tag{5.10}$$

where $\overline{s}^2 = \frac{[S-f(X,\gamma)]'[S-f(X,\gamma)]+\underline{\nu}\underline{s}^2}{\overline{\nu}}$ and $\overline{\nu} = \underline{\nu} + T$; see Appendix E for details. As shown in Koop (2003, Ch. 5), the conditional posterior distribution of $\gamma$ is:

$$\begin{aligned} p(\gamma|S,\eta) \propto \exp\left[-\frac{\eta}{2}\{S - f(X,\gamma)\}'\{S - f(X,\gamma)\}\right] \\ \exp\left[(-\frac{1}{2}(\gamma-\underline{\gamma})'\underline{V}^{-1}(\gamma-\underline{\gamma}))\right]. \end{aligned} \tag{5.11}$$

This latter conditional posterior, $p(\gamma|S,\eta)$, does not match any known density from which to directly sample from. We propose the random walk chain Metropolis-Hasting (RW-MH) posterior simulator to sequentially draw parameters from a suitable candidate generating density, in the spirit of Koop (2003, Ch. 5). Essentially, candidate draws of $\gamma$, denoted by $\gamma^*$, are generated according to a random walk. Following a typical procedure, we choose the multivariate Normal distribution as the candidate generating density:

$$q(\gamma^{(dr-1)}, \gamma) \sim fN(\gamma|\gamma^{(dr-1)}, \Sigma), \tag{5.12}$$

where $\gamma^{(dr-1)}$ denotes the last accepted draw of $\gamma$, and $\Sigma$ is a pre-selected covariance matrix which guarantees that the acceptance probability is within a reasonable range, typically $[0.2, 0.5]$. Using data available up to the beginning of our first forecast we set this covariance matrix to the maximum likelihood variance estimate, $\Sigma = var(\widehat{\gamma}_{ML})$. The acceptance probability of the candidate draw is calculated as:

$$a(\gamma^{(dr-1)}, \gamma^*) = \min\left[\frac{p(\gamma = \gamma^*|S, \eta)}{p(\gamma = \gamma^{(dr-1)}|S, \eta)}, 1\right]. \tag{5.13}$$

with $p()$ at the current and previous draw evaluated using Eq. (5.11).

The RW-MH algorithm simulates draws for $p(\gamma|S, \eta)$, but we also require draws from $p(\eta|S,\gamma)$. Since we know the form of this density - see Eq. (5.10) - we can easily combine the RW-MH step with the Gibbs sampler. Such Metropolis-within-Gibbs algorithm allows us to sequentially draw $\eta$ conditional on $\gamma$. In Appendix E we provide further details and exact steps.

To forecast with our model we need the predictive density. This is given by:

$$p(S^*|S, \gamma) = t(S^*|f(X^*, \gamma), \overline{s}^2 I_T, T), \tag{5.14}$$

where $\overline{s}^2 = (S - f)'(S - f)/T$. Using the Gibbs sampler we can obtain draws from this predictive density, from which we can compute point and density forecasts. In our empirical exercise, we generate 31000 draws from which we discard the first 1000 and keep every third draw for inference. Details about the convergence measures are relegated to Appendix J.[6]

### 5.3.3 Bayesian Model Averaging or Selection and the Optimal Predictive Pool

So far we have focused on estimating and forecasting with a model defined according to the predictors it includes. Since we can estimate and obtain predictive densities for several alternative models at each point in time, we can optimally exploit the predictive content of each predictor. For example, we can compute forecast combinations based on each model's relative importance over time. Alternatively, we can forecast with the model that yields the highest weight (i.e., probability) at each point in time or compute the optimal predictive pool of Geweke and Amisano (2011). The first two approaches assume that the true model is in the model set and the selection or combination converges asymptotically to it. The optimal predictive pool, on the contrary, allows for model incompleteness, meaning the true model might not be present in the model set, see Mitchell and Hall (2007) and Geweke and Amisano (2011).

To visualize these weighting and forecasting schemes, let $M_i$ identify a specific model from the set of $M^N$ models, such that the predictive density in Equation (5.14) is now also model-specific, $p(S^*|S, \gamma, M_i)$. Bayesian Model Selection (BMS) uses weights derived from the realized likelihood of the model's prediction to select a single model.

[6] We checked the convergence and adequacy of the number of draws using standard procedures, such as Geweke's (1992) numerical standard errors (NSE) and acceptance rates in the RW-MH algorithm. Overall, results indicate an acceptable degree of efficiency of the algorithm.

Bayesian Model Averaging (BMA) employs the weights to average results over all models.

The starting point is to assign prior probabilities to each model, and subsequently obtaining posterior probabilities (weights) based on the model's realized likelihood. We assume a priori that each model has the same chance of becoming probable, hence, the prior is: $\Pr(M_i) = 1/M^N$. The posterior probability of model $i$ at time $t$, defined by $\Pr(M_i|D)$, is given by:

$$\Pr(M_i|D) = \frac{\Pr(D|M_i)\Pr(M_i)}{\sum_{j=1}^{M^N}\Pr(D|M_j)\Pr(M_j)}, \tag{5.15}$$

where $\Pr(D|M_i)$ is the marginal likelihood of the i$^{th}$ model. We compute this likelihood using the method of Gelfand and Dey (1994), see Appendix E for details. Note that the posterior model probability also allows us to infer about which predictor receive more support from the data.

The forecasts from BMA are computed by weighting each model's forecast by the model's posterior probability:

$$p(S^*|S,\gamma) = \sum_{i=1}^{M^N}\Pr(M_i|D)p(S^*|S,\gamma,M_i). \tag{5.16}$$

In BMS, instead, the forecasts are based on the model with the highest posterior probability. Finally, the optimal predictive pool combines the forecasts of the $M^N$ models according to weights related to the model's past predictive performance:

$$p(S^*|S,\gamma) = \sum_{i=1}^{M^N}\mathbf{w}_i^*(S^*|S,\gamma,M_i), \tag{5.17}$$

with $\mathbf{w}_i^*$ denoting an $(M^N \times 1)$ vector of weights obtained by solving a maximization problem conditional on information available at the time the forecast is made:

$$\mathbf{w}_i^* = \arg\max_w \log\left[\sum_{i=1}^{M^N} w_i^* \times \exp(LS_i)\right]. \tag{5.18}$$

where $\mathbf{w} \in [\mathbf{0},\mathbf{1}]$ and $LS_i$ is the log score for model $i$ computed using information available up to time $t$. In the next section we describe our predictors, and hence the set of models contained in $M^N$.

### 5.3.4 Choice of Regressors

While our left-hand side variable is always sampled at monthly frequency, on the right-hand-side our regressions allow for commodity prices sampled at daily or monthly frequency, or standard macroeconomic predictors at monthly frequency. The menu of commodity-related regressors includes oil, gold, gas, and copper prices and a commodity price index. These choices reflect the commodities exported by the countries we focus upon and are in line with recent studies on the commodity price - exchange rate relationship, such as Chen et al. (2010) and Ferraro et al. (2015).

The selection of the macroeconomic variables is guided by the standard models of exchange rate determination. See, among others, Engel and West (2005), Molodtsova and Papell (2009), and Rossi (2013). Thus, in addition to commodity prices changes at monthly frequency, $\mathbf{z}_t$ can also be a predictor derived from:

- The Monetary Model (MM):

$$\mathbf{z}_{t,MM} = (m_t - m_t^*) - (y_t - y_t^*) - s_t, \tag{5.19}$$

  where $m_t$ is the log of money supply, $y_t$ is the log of income, and asterisks denote foreign country variables;[7]

- Purchasing Power Parity (PPP) condition:

$$\mathbf{z}_{t,PPP} = p_t - p_t^* - s_t, \tag{5.20}$$

where $p_t$ is the log of price level;

- Uncovered Interest Rate Parity (UIP) condition:

$$\mathbf{z}_{t,UIP} = i_t - i_t^*, \tag{5.21}$$

  with $i_t$ denoting the short-term nominal interest rate;

- A symmetric and an asymmetric Taylor rule (TRsy and TRasy, respectively):

$$\mathbf{z}_{t,TRsy} = 1.5(\pi_t - \pi_t^*) + 0.5(\overline{y}_t - \overline{y}_t^*), \tag{5.22}$$
$$\mathbf{z}_{t,TRasy} = 1.5(\pi_t - \pi_t^*) + 0.1(\overline{y}_t - \overline{y}_t^*) + 0.1(s_t + p_t^* - p), \tag{5.23}$$

  where $\pi_t$ is the inflation rate, and $\overline{y}_t$ the output gap.[8]

[7]Note that we have assumed an income elasticity of one in the monetary model ($\mathbf{z}_{t,MM}$), following Mark (1995) and Engel and West (2005).

[8]The proxy for the output is monthly industrial production (IP). In line with the standard practice in exchange rate economics, the output gap is obtained by applying the Hodrick and Prescott (1997) filter recursively to the output series. We also use the conventional smoothing parameter for monthly data - 14400. To correct for the uncertainty about these estimates at the recursive sample end-points, we follow Watson's (2007) method. We estimate bivariate VAR($\ell$) regressions on the first difference of inflation and the change in the log IP, with the lag length in the VAR determined by AIC. These regressions are then used to forecast and backcast three years worth of monthly data on IP, and the filter is applied to the resulting extended series.

### 5.3.5 Data and Forecasting Mechanics

The data consists of exchange rates of the following (home) countries relative to the US dollar: Australia (AUD), Canada (CAD), Norway (NOK), and Japan (YEN). The first three countries can be currently categorized as net commodity exporters, while Japan is a net oil importer. The exchange rate is the end-of-month value of the national currency per U.S. dollar. Our effective sample period runs from 1986M9 to 2014M3 for all countries, except Norway. Due to unavailability of data on daily gas prices fluctuations, the sample period for Norway comprises 1997M1 - 2014M3. Further details on exact data sources, definitions, and descriptive statistics are provided in Appendix I.

We employ a recursive forecasting scheme, while generating direct forecasts at 1- and 3-months horizons.[9] In diagram (5.1) we exemplify the mechanics of our forecasting procedure with our MIDAS regression for 1-month horizon. We use data from 1986:M9:D22 to 1998:M9:D22 to estimate parameters of our MIDAS regression - as in Eq. (5.2). Data from 1998:M10:D1 to 1998:M10:D21 is used to forecast the period ahead change in the exchange rate ($s_{1998:M12} - s_{1998:M11}$). In this sense, we use information up to one day before the end-of-the month to generate the forecast. We then add one month worth of daily data and repeat, until the end of the sample, providing us with a long series of $P$ out-of-sample forecasts.[10]

Diagram 5.1: Example of Data Timming Scheme in the Forecasting Regression, $h = 1$

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ... | 1998:M9:D22 | 1998:M10:D1 | 1998:M10:D2 | ... | 1998:M10:D21 | 1998:M10:D22 | ... | 1998:M11:D22 |
| ... | 1998:M10 | | | | | 1998:M11 | | 1998:M12 |

### 5.3.6 Measures of Forecasting Performance

We employ the root mean squared forecast error (RMSFE) as a statistical measure of out-of-sample point forecast accuracy. The benchmark model is the driftless random walk (RW).[11] To be precise, we compute the ratio of the RMSFE of our commodity or fundamentals-based models relative to the RMSFE of the RW:

$$\text{Relative RMSFE} = \frac{\sqrt{\frac{1}{P}\sum_{\overline{p}=1}^{P} fe_{i,\overline{p}}^2}}{\sqrt{\frac{1}{P}\sum_{\overline{p}=1}^{P} fe_{RW,\overline{p}}^2}}, \tag{5.24}$$

where $P$ is the number of out-of-sample forecasts, $fe_i^2$ and $fe_{RW}^2$ are the squared forecast errors of our model $i$ and the RW, respectively. Values of the relative RMSFE

[9] According to Wright (2008), direct and iterated forecasting approaches yield qualitatively similar conclusions.

[10] For all currencies, except the NOK, $P = 167$ at $h = 1$ and $P = 163$ at $h = 3$. For the NOK, $P = 102$ and 98 at $h = 1$ and 3, respectively.

[11] According to Rossi (2013), the forecasts from this naive benchmark are the hardest to improve upon.

below one are consistent with a more accurate point forecast of model $i$ against the RW. To evaluate whether the differences in the RMSFE between our models and the RW are significant we use the Clark and West (2006, 2007) test, hereafter CW-test. To examine the forecasting performance of our models over time in terms of point forecast, we compute the relative RMSFE recursively over the out-of-sample period.

Our use of Bayesian methods allow us to fully exploit the information in the predictive density, rather than focusing exclusively on point forecast. In this regard, we first compute the mean log-score differentials (MLSD):

$$MLSD = P^{-1} \sum_{\overline{p}=1}^{P} (LS_{i,\overline{p}} - LS_{RW,\overline{p}}), \tag{5.25}$$

where $LS_{i,\overline{p}}$ and $LS_{RW,\overline{p}}$ are the log-scores of our model $i$ and the RW, respectively. Positive values of $MLSD$ are consistent with more accurate density forecasts of model $i$ relative to the RW. Finally, we calculate the cumulative log-score differentials (CLSD) of our regressions relative to those of the RW over the out-of-sample period. Positive values of the CLSD indicate that our commodity or fundamentals-based regressions produce more accurate density forecasts than the RW benchmark.

## 5.4 Empirical Results

### 5.4.1 Evaluation of Forecasting Performance

#### Forecasts from Regressions with Individual Predictors

In Tables 5.1 and 5.2 we assess the forecasting performance of models conditioned on each of the regressors we consider. In Table 5.1 we focus on the relative RMSFE to examine performance in terms of point forecast, and in Table 5.2 we look at log-score differentials to inspect density forecast improvements. Table 5.1 conveys three key findings. First, models conditioned on daily regressors, i.e. the MIDAS models, yield a lower RMSFE than the RW benchmark for some commodity-currency pairs, especially at $h = 1$. This is the case for the Australian and Canadian dollar MIDAS regressions with copper prices. For instance, for the Australian dollar and changes in copper prices the MIDAS regression reduces the RMSFE by 1.1% relative to the benchmark. An improvement of the same magnitude is also apparent in the commodity currency we examine - the Japanese yen with daily oil prices. While these reductions in the RMSFE are seemingly low, our tests of equal predictive ability suggest that the differences in the RMSFE we detect are statistically significant for the AUD and the Yen. Later, we will examine other metrics to gain more insights on the consistence of these gains over the out-of-sample period.

Second, regressions with monthly commodity prices fail to forecast better than the RW regardless of the forecasting horizon. That is, the relative RMSFE are above one

Table 5.1: Relative RMSFE and CW-test for Models with Single Predictor

| | AUD | CAD | NOK | YEN | AUD | CAD | NOK | YEN |
|---|---|---|---|---|---|---|---|---|
| | h=1 | | | | h=3 | | | |
| Daily regressors (MIDAS model) | | | | | | | | |
| ΔOil_Dp | - | 1.006 | 1.021 | **0.989**** | - | 1.013 | 1.003 | 1.003 |
| ΔGold_Dp | 1.026 | - | - | - | 1.022 | - | - | - |
| ΔCopper_Dp | **0.989*** | **0.993** | - | - | **0.995** | 1.016 | - | - |
| ΔGas_Dp | - | - | 1.003 | - | - | - | 1.020 | - |
| ΔDP_index | - | - | - | 1.000 | - | - | - | 1.008 |
| Monthly regressors | | | | | | | | |
| ΔOil_Mp | - | 1.004 | 1.008 | 1.011 | - | 1.004 | 1.009 | 1.010 |
| ΔGold_Mp | 1.005 | - | - | - | 1.014 | - | - | - |
| ΔCopper_Mp | 1.001 | 1.000 | - | - | 1.004 | 1.003 | - | - |
| ΔGas_Mp | - | - | 1.003 | - | - | - | 1.017 | - |
| ΔMP_index | - | - | - | 1.013 | - | - | - | 1.021 |
| Monthly regressors | | | | | | | | |
| TRsy | 1.000 | **0.987**** | 1.009 | 1.016 | 1.001 | **0.969**** | 1.027 | 1.018 |
| TRasy | 1.002 | **0.988**** | 1.000 | 1.003 | **0.997** | **0.968**** | 1.003 | 1.005 |
| MM | 1.015 | 1.008 | 1.018 | 1.008 | 1.042 | 1.041 | 1.058 | 1.029 |
| PPP | 1.007 | 1.001 | 1.005 | 1.008 | 1.017 | 1.007 | 1.023 | 1.022 |
| UIP | 1.005 | 1.003 | 1.022 | 1.004 | 1.015 | 1.014 | 1.046 | 1.014 |

**Notes:** The Table shows the Root Mean Squared Forecast Error (RMSFE) of the commodity or fundamental-based forecasting model relative to the RMSFE of the driftless Random Walk (RW). Values less than 1 (one) indicate that the commodity or fundamental-based model generates a lower RMSFE than the RW, hence, it forecasts better than the RW. The Table also reports the CW-test with asterisks (*10%, **5%, ***1%) denoting the level of significance at which the null hypothesis of equal RMSFE is rejected, favouring the alternative that the commodity or fundamental-based model has a lower RMSFE. The commodity or fundamentals-based forecasting model uses the relevant country-commodity or fundamental listed in the first column and grouped in terms of daily and monthly regressors. In all models, the forecasts are generated recursively for $h$-month(s)-ahead change in the exchange rate. When only daily regressors are used, the forecasts are from the MIDAS model. The list of daily regressors include, change in daily prices (Dp) of oil, gold, copper, gas, and a daily commodity price index (ΔMP_index). In the monthly regressors group we have a similar set of commodities, but also fundamentals from the symmetric Taylor rule - TRsy, the asymmetric Taylor rule - TRasy the Monetary Model - MM, Purchasing Power Parity - PPP, and Uncovered Interest Rate Parity - UIP. The currency codes in the first row denote the Australian dollar (AUD), the Canadian dollar (CAD), the Norwegian Kroner (NOK), and the Japanese YEN. The forecast evaluation period is 1998M11+$h$ to 2014M3 for all currencies, except the NOK (2005M7 +$h$ to 2014M3).

in all cases. Hence, in line with Ferraro et al. (2015) and Chen et al. (2010), we affirm the lack of predictive content of commodity prices sampled at low-frequency for monthly variations in exchange rates.

Finally, our results in Table 5.1 also support the prevalent view in the literature regarding the predictive ability of fundamentals derived from Taylor rules. See, for instance, Molodtsova and Papell (2009) and Rossi (2013). As shown, among the standard macroeconomic fundamentals we use, those from the Taylor rule display a significant predictive content for the Canadian dollar at the $h = 1$ and $h = 3$. In contrast, fundamentals from the Monetary Model (MM), PPP, and UIP yield a relative RMSFE above one, with MM exhibiting the weakest performance for most currency pairs and horizons.

Table 5.2: Average Log-Score Differentials for Models with Single Predictors

| | AUD | CAD | NOK | YEN | AUD | CAD | NOK | YEN |
|---|---|---|---|---|---|---|---|---|
| | h=1 | | | | h=3 | | | |
| Daily regressors (MIDAS model) | | | | | | | | |
| $\Delta$Oil_Dp | - | 2.109 | **2.657** | 2.227 | - | 2.419 | 2.711 | 2.454 |
| $\Delta$Gold_Dp | **2.932** | - | - | - | 2.442 | - | - | - |
| $\Delta$Copper_Dp | 2.665 | **2.852** | - | - | 2.686 | 2.783 | - | - |
| $\Delta$Gas_Dp | - | - | 1.729 | - | - | - | 2.590 | - |
| $\Delta$DP_index | - | - | - | 1.510 | - | - | - | 1.882 |
| | | | | | | | | |
| $\Delta$oil_Mp | - | 2.097 | 1.718 | 2.793 | - | 2.456 | 2.645 | 2.448 |
| $\Delta$Gold_Mp | 2.199 | - | - | - | 1.999 | - | - | - |
| $\Delta$Copper_Mp | 2.273 | 2.330 | - | - | 2.180 | 2.710 | - | - |
| $\Delta$Gas_Mp | - | - | 1.615 | - | - | - | 2.515 | - |
| $\Delta$MP_index | - | - | - | 1.902 | - | - | - | 2.282 |
| | | | | | | | | |
| TRsy | 1.922 | 2.500 | 1.867 | 1.582 | 2.679 | 2.810 | 2.603 | 1.937 |
| TRasy | 2.721 | 2.757 | 1.744 | 2.271 | **3.217** | 3.078 | 2.699 | 2.784 |
| MM | 2.403 | 2.839 | 2.273 | 1.962 | 2.778 | **3.611** | 2.270 | 2.598 |
| PPP | 2.049 | 1.491 | 2.380 | 2.696 | 2.934 | 2.440 | 2.674 | 3.291 |
| UIP | 2.058 | 1.834 | 1.703 | **3.069** | 2.255 | 2.652 | **2.714** | **3.653** |

**Notes:** The Table reports the average log-score differentials between the commodity or fundamental-based forecasting model and the driftless Random Walk (RW). Positive values indicate that the commodity or fundamental-based model improves upon the RW in terms of density forecasts. We also highlight in bold the model with the largest log-score differential for each currency/horizon. The commodity or fundamentals-based forecasting model uses the relevant country-commodity or fundamental listed in the first column and grouped and grouped in terms of daily and monthly regressors. In all models, the forecasts are generated recursively for $h$-month(s)-ahead change in the exchange rate. When only daily regressors are used, the forecasts are from the MIDAS model. The list of daily regressors include, change in daily prices (Dp) of oil, gold, copper, gas, and a daily commodity price index ($\Delta$MP_index). In the monthly regressors group we have a similar set of commodities, but also fundamentals from the symmetric Taylor rule - TRsy, the asymmetric Taylor rule - TRasy, the Monetary Model - MM, Purchasing Power Parity - PPP, and Uncovered Interest Rate Parity - UIP. The currency codes in the first row denote the Australian dollar (AUD), the Canadian dollar (CAD), the Norwegian Kroner (NOK), and the Japanese YEN. The forecast evaluation period is 1998M11+$h$ to 2014M3 for all currencies, except the NOK (2005M7 +$h$ to 2014M3).

Turning to density forecasts, results in Table 5.2 reveal that once we account for the entire forecast distribution, the RW never outperforms our commodity or fundamentals-based regressions. In all cases, the log-score differentials are positive, with MIDAS models on some commodity-currency pairs exhibiting the largest values at $h = 1$. For example, the MIDAS model with daily copper prices changes displays the largest log-score differentials among all the forecasting models for the Australian dollar. A similar assertion holds for oil prices and the Norwegian Kroner, as well as for gold prices and the Australian dollar. In contrast, at $h = 3$ the largest log-score differentials occur among models with standard macroeconomic fundamentals, such as those derived from the asymmetric Taylor rule for the CAD and UIP for the NOK.

On balance, we find that when we exploit the full predictive density, all the commodity or fundamentals-based models provide more accurate forecasts than the RW benchmark. In this sense, our results complement the findings in Wang and Wu (2012) who uncover that exchange rate models generate tighter forecast intervals than RW when employing standard macro fundamentals. In terms of point forecasts, daily commodity prices are useful in predicting the Australian dollar and the Japanese yen at $h = 1$, and less significantly so the Canadian dollar.

**Forecast Combinations**

The results in the previous section are based on individual model performance and therefore do not exploit the possibility that one regressor might have forecasted well in parts of the out-of-sample period and poorly in other parts. To exploit this possibility and account for time-variation in forecasting performance, we now turn to forecast combinations methods.

Table 5.3 reports results for forecast combinations under BMA, BMS, and the optimal predictive pool. We notice immediately the benefits of forecast combinations, since in most cases we improve upon the benchmark. In the case of the Canadian dollar for instance, while with individual daily commodity prices the relative RMSFE were all above one at $h = 3$, when the forecasts are combined the resulting figure drops to a value slightly below one. At this horizon, the highest improvement relative to the RW occurs when both, forecasts based on daily and monthly predictors are combined. In contrast, at $h = 1$ combinations methods using solely monthly regressors deliver larger improvements for this currency. For the Australian dollar and at either horizon, the gains from combining only daily predictors' forecasts are greater than those from monthly regressors; but they are almost of the same magnitude as the combination from both, daily and monthly, regressors' forecasts. In the case of the Norwegian krone either combination method is unable to improve upon the RW, in line with results from the single predictor forecast evaluation.

Our results further suggest that there is no clear ranking in terms of the forecast combination method that performs consistently better across horizons and/or model combinations. For example, when the forecasts from monthly regressors are combined,

Table 5.3: Relative RMSFE and CW-test for Forecast Combinations

| | Daily regressors - Commodity Prices (CmdtyP) | | | Monthly regressors (CmdtyP and macro fundamentals) | | | Daily and monthly regressors (CmdtyP and macro fundamentals) | | |
|---|---|---|---|---|---|---|---|---|---|
| | BMA | BMS | OptPool | BMA | BMS | OptPool | BMA | BMS | OptPool |
| | | | | | h=1 | | | | |
| AUD | **0.992** | **0.987*** | **0.990*** | 1.002 | 1.000 | 1.013 | **0.993** | **0.987*** | **0.990*** |
| CAD | 1.000 | **0.999** | 1.002 | **0.991**** | **0.995*** | **0.990**** | **0.996** | **0.998** | **0.996** |
| NOK | 1.014 | 1.016 | 1.019 | 1.009 | 1.005 | 1.029 | 1.019 | 1.017 | 1.036 |
| YEN | **0.989**** | **0.989**** | **0.990**** | 1.004 | 1.003 | 1.011 | **0.998** | **0.999** | 1.009 |
| | | | | | h=3 | | | | |
| AUD | **0.998** | **0.996** | **0.998** | **0.999** | **0.998** | 1.026 | **0.997** | **0.997** | 1.024 |
| CAD | **0.995** | **0.995** | **0.991** | **0.975**** | **0.987**** | **0.980*** | **0.975**** | **0.980**** | **0.976**** |
| NOK | 1.007 | 1.001 | 1.005 | 1.025 | 1.013 | 1.042 | 1.021 | 1.006 | 1.040 |
| YEN | **0.996** | **0.994** | **0.999** | 1.014 | 1.018 | 1.041 | 1.012 | 1.016 | 1.047 |

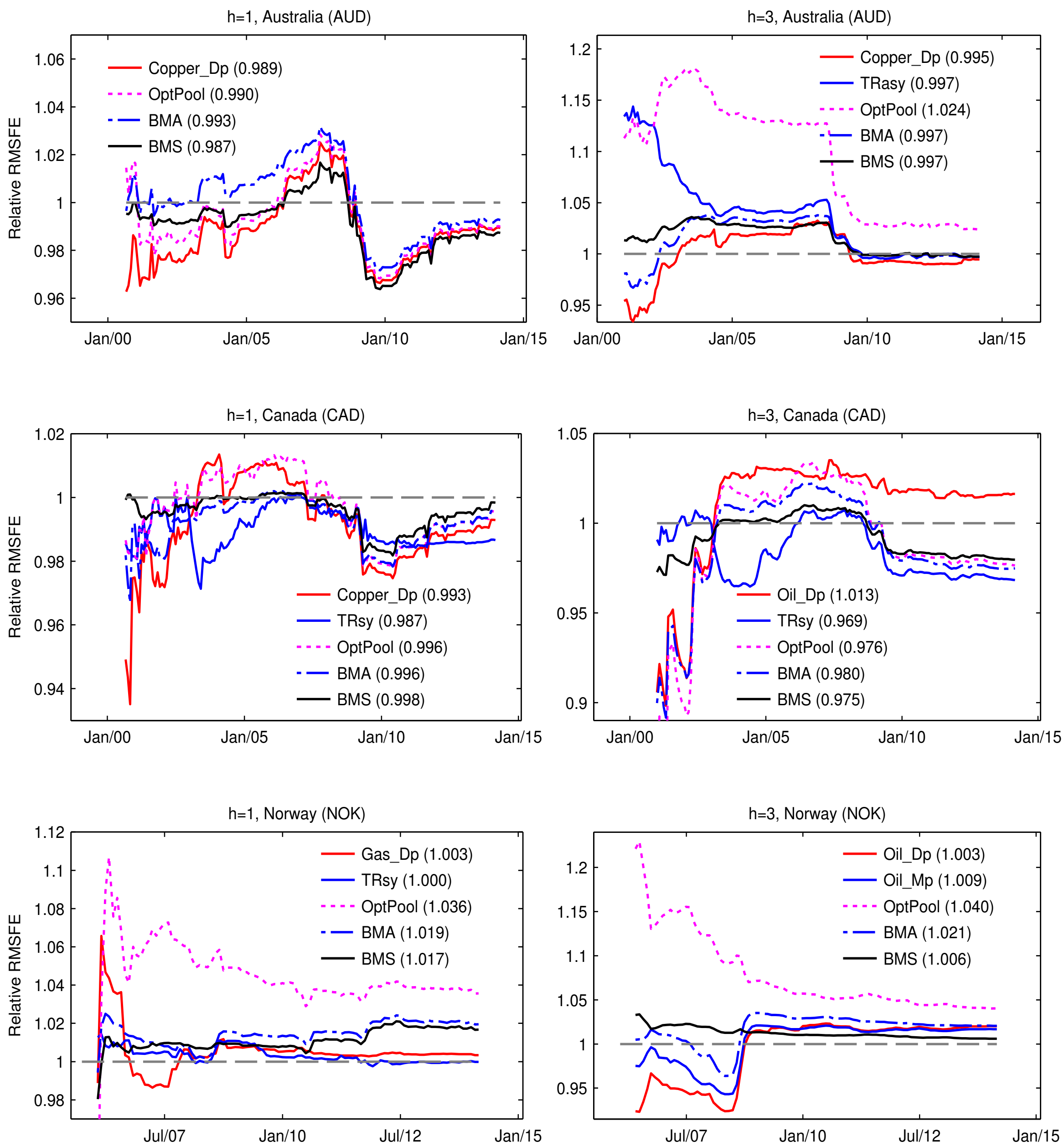
**Notes:** The Table reports the Root Mean Squared Forecast Error (RMSFE) for forecast combination methods relative to the RMSFE of the driftless Random Walk (RW). The methods include, Bayesian Model Averaging (BMA), Bayesian Model Selection, (BMS), and the Optimal Predictive Pool (OptPool) of Geweke and Amisano (2011). Values less than 1 (one) indicate that the combination method generates a lower RMSFE than the RW, hence, it forecasts better than the RW. The Table also reports the CW-test with asterisks ( *10%, **5%, ***1%) denoting the level of significance at which the null hypothesis of equal RMSFE is rejected, favouring the alternative that the combination method has a lower RMSFE. The forecast combinations are based on the relevant commodity-currency and standard macroeconomic fundamentals. For the Australian dollar (AUD) the relevant commodities are gold and copper; for the Canadian dollar (CAD) - oil and copper; and for the Norwegian Kroner (NOK) these include oil and gas. When only daily regressors are used the combination is based on forecasts from the MIDAS models - reported in column [2-4]. In column [5-7] the combination is based on forecasts from monthly regressors, while the last three columns report results from combining daily and monthly regressors. In all cases, the forecasts are generated recursively for $h$-month(s)-ahead change in the exchange rate. In the group of monthly regressors we have a set of commodity pairs similar to the daily group, but also fundamentals from the symmetric Taylor rule - TRsy, the asymmetric Taylor rule - TRasy, the Monetary Model - MM, Purchasing Power Parity - PPP, and Uncovered Interest Rate Parity - UIP. The forecast evaluation period is 1998M11+$h$ to 2014M3 for all currencies, except the NOK (2005M7 +$h$ to 2014M3).

BMA delivers the largest reduction in the relative RMSFE for the Canadian dollar at $h = 3$ (around, 2.5%). But for the same currency and at $h = 1$, the optimal predictive pool achieves the best performance. In some instances, such as for the Australian dollar with daily regressors and $h = 1$, it is the BMS that exhibits the best performance.

### Forecasting Performance Over-time

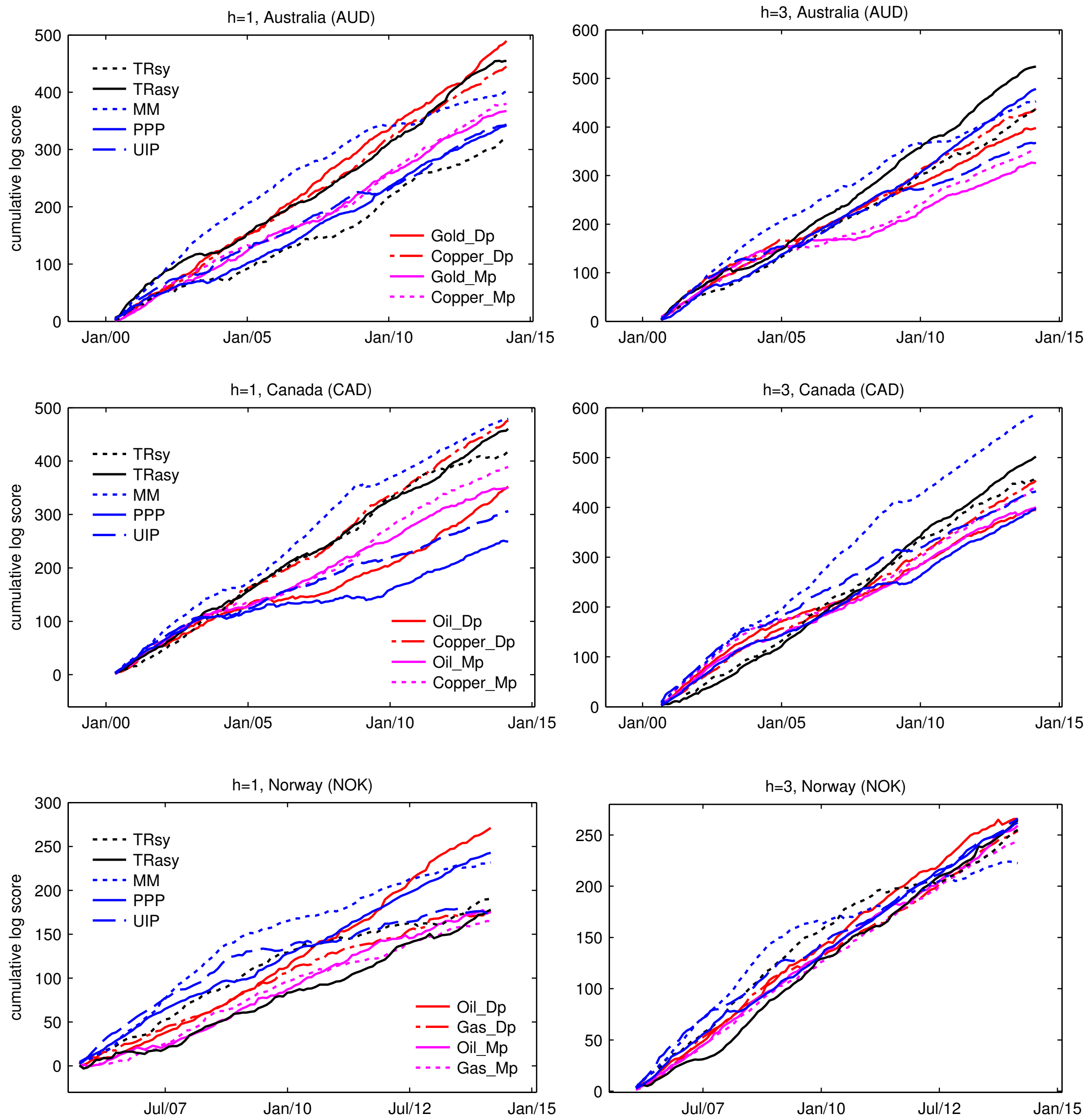
All our results so far are based on measures of global performance since they are based on averages over the out-of-sample (OOS) period. These metrics leave open the question of whether they are influenced by a few data-points in the OOS period and if the performance we obtain is consistent over the entire OOS period. To shed light on these questions, we next examine metrics of local relative performance, namely the recursive relative RMSFE and the cumulative log-score differentials.

Figure 5.2: Recursive Relative RMSFE for Selected Predictors and Forecast Combination Methods




Notes: The figure shows the recursive relative Root Mean Squared Forecast Error (RMSFE) for selected commodity or fundamental-based models and forecast combinations methods. In all cases the benchmark is the driftless Random Walk (RW), so that values less than 1 (one) mean that the commodity/fundamental-based model or the combination method improves upon the RW at that point in time. The numbers in each plot's legend (in brackets) are the relative RMSFE at the last recursion, which coincide with the relative RMSFE reported in Table 5.1 for the respective regressor-currency pair. In the legend, the suffixes Dp and Mp attached to the commodity prices denote daily and monthly prices respectively. The other monthly regressors include fundamentals from the symmetric (TRsy) and asymmetric Taylor rules (TRasy). When the regressor is sampled daily, the recursive relative RMSFE is generated from the MIDAS model. The forecast combinations methods, namely the Optimal Predictive Pool (OptPool), Bayesian Model Averaging (BMA) or Selection (BMS) are based on regressors sampled daily and monthly.

Figure 5.3: Statistical Evaluation based on Cumulative Log-Scores for Models with Single Predictor




Notes: The figure presents the cumulative sum of log predictive scores of the commodity or fundamental-based forecasting model, computed relative to cumulative sum of log predictive scores of the Random Walk. Positive values indicate that the commodity/fundamental-based model outperforms the RW at that point in time, while negative values suggest the opposite. In the plot's legend, the suffixes Dp and Mp attached to the commodity prices denote daily and monthly prices respectively. The other monthly regressors include fundamentals from the symmetric (TRsy) and asymmetric Taylor rules (TRasy), the Monetary Model - MM, Purchasing Power Parity - PPP, and Uncovered Interest Rate Parity - UIP. When the regressor corresponds to a daily commodity price, the recursive cumulative log-score differentials are generated from MIDAS models.

Figure 5.2 depicts the recursive relative RMSFE for a representative selection of single-predictor regressions and all forecast combination methods. Examination of the figure suggests that the improvements we obtain are consistent for the most part of the OOS period. In the case of the Australian dollar and copper prices, for example, the relative RMSFE is below one for the most part of the forecast window except around the 2008 financial crisis. A similar pattern holds for the recursive relative RMSFE of the Canadian dollar and copper prices, excluding the period between 2003 and 2007. As anticipated, the consistency is stronger with forecast combination methods, particularly the Bayesian Model Selection (BMS). We further note that the poor point forecasting performance of our regressions for the Norwegian Kroner is essentially a phenomenon of the entire OOS forecast window.

Figure 5.3 shows cumulative log-score differentials from regressions with individual predictors. A key observation from the graphs is that all the commodity or fundamentals-based models improve upon the RW over the OOS period. However, among them, there are generally variations over-time in terms of the model with the best forecasting performance. Looking at the Australian dollar case at $h = 1$, regressions with fundamentals from MM provided the best density forecasts up to 2010 and among all the regressions considered. From this period onwards, MIDAS models based on gold price changes turned to be the best. An analogous shift occurred between models with MM fundamentals and daily oil prices for Norway at $h = 1$. Whilst at $h = 1$ the shift in the best models occur mostly between forecasting models with daily commodity prices and standard macroeconomic fundamentals, at $h = 3$ the switch often involves models employing standard macroeconomic fundamentals. Examples of these cases include regressions for the Australian and Canadian dollars.

Overall, our metrics of local relative performance indicate that our results are not influenced by a few data-points in the OOS. Rather, they prevail over the entire path of the forecasting period. In the following section we take a closer look at some of the characteristics of the Bayesian combination methods.

### 5.4.2 Characterization of MIDAS Models and Forecast Combination Methods

The previous results hint at the usefulness of the MIDAS regression in forecasting commodity currencies, especially at 1−month forecasting horizon. They also reveal the benefits of Bayesian approaches to forecast combinations. Since the forecasts from the BMA and BMS methods emanate from a combination of several individual models, here we study some of their embedded characteristics, in an effort to pin down the degree of informativeness of predictors sampled at the different frequencies. We also examine the weights on daily commodity prices ensuing from our MIDAS regressions.

### Weights on Daily Observations

The key element in our MIDAS regression is the exponential Almon polynomial function. Parameterized on two coefficients, $(\theta_1, \theta_2)$, the function allow us to smooth past daily observations on commodity prices changes. Values of $\theta_1 = \theta_2 = 0$ are consistent with equal weights on each daily observation and the MIDAS regression forecasts might be similar to those of the standard linear regression.

Figure 5.4 illustrates the typical weights estimated in our MIDAS regressions. The example is based on the coefficients' estimates for our last out-of-sample forecast. Clearly, several non-constant weighting patterns are patent, regardless of the corresponding forecasting performance. For example, for the Australian and Canadian dollars at 1-month forecasting horizon, observations close to the end of the month have a higher weight than those at the beginning of the month. And we recall that with these weights, the corresponding MIDAS model improved upon the driftless RW. For the Norwegian Kroner and Canadian dollar at $h = 3$ only the first three daily observations at the beginning of the month carry most of the weight.

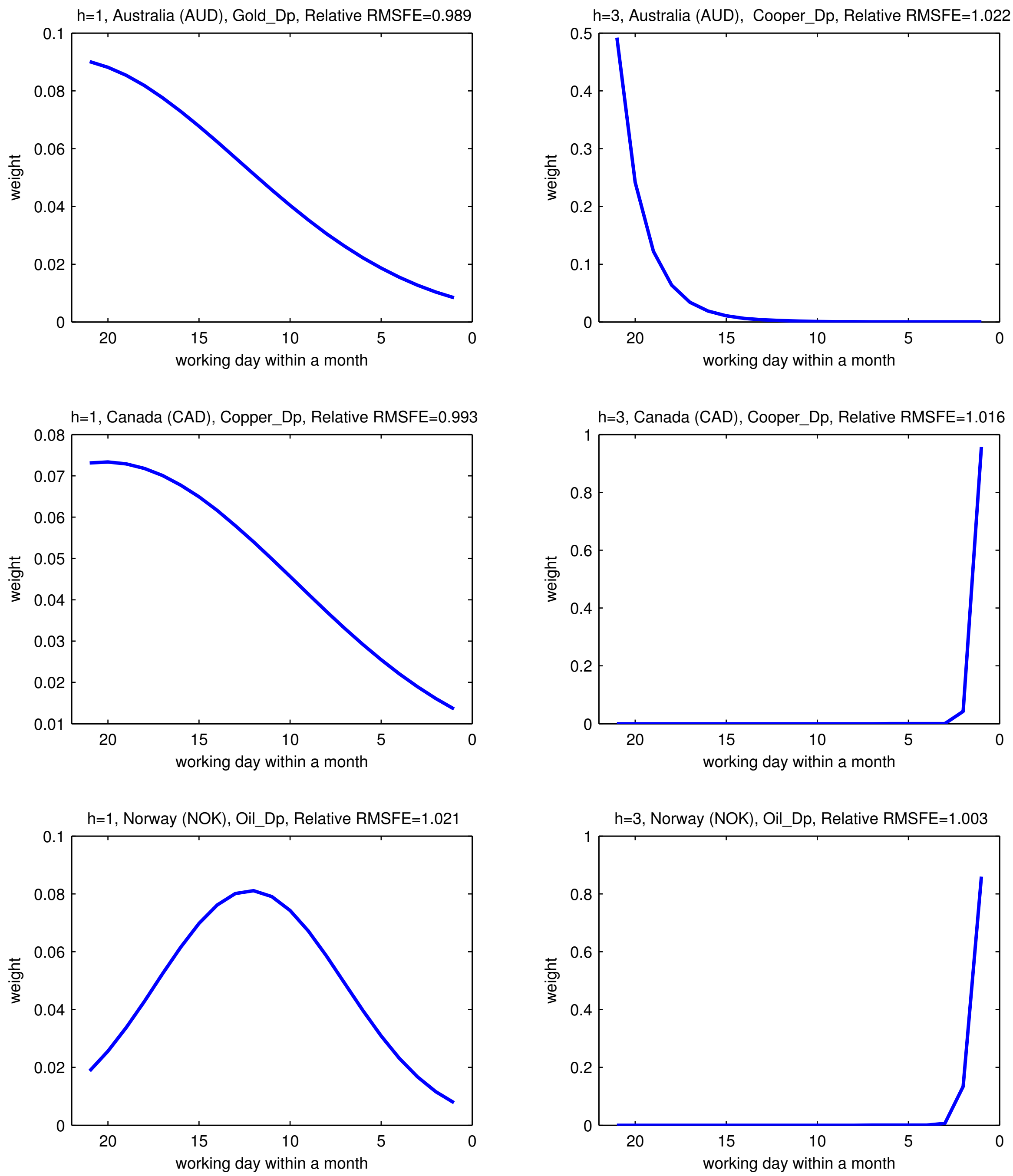
### Predictors' Weights in the BMA Method

The forecasts from the BMA method in Table 5.3 are generated by weighting each model's forecast by the model's posterior probability. This implies that the larger the model posterior probability, the greater the weight attached to model's forecast in the BMA method. Figure 5.5 plots the posterior probabilities associated with each model. At a glance, MIDAS models with daily commodity prices exhibit the largest posterior probabilities at 1-month forecasting horizon, regardless of their ability to improve upon the RW. This is the case for MIDAS regressions for (i) Australia with copper and gold prices; (ii) Canada and copper prices; and (iii) Norway with oil prices. By contrast, at $h = 3$ regressions with standard macroeconomic fundamentals sampled at low frequency display large posterior probabilities. In particular, models employing fundamentals from the symmetric and asymmetric Taylor rule are largely supported by the data for Australia and Canada. For Norway, although the asymmetric Taylor rule is empirically plausible, the weights on daily oil price fluctuations and fundamentals from PPP in the BMA forecast are relatively high.

### Predictors' with the Largest Weights

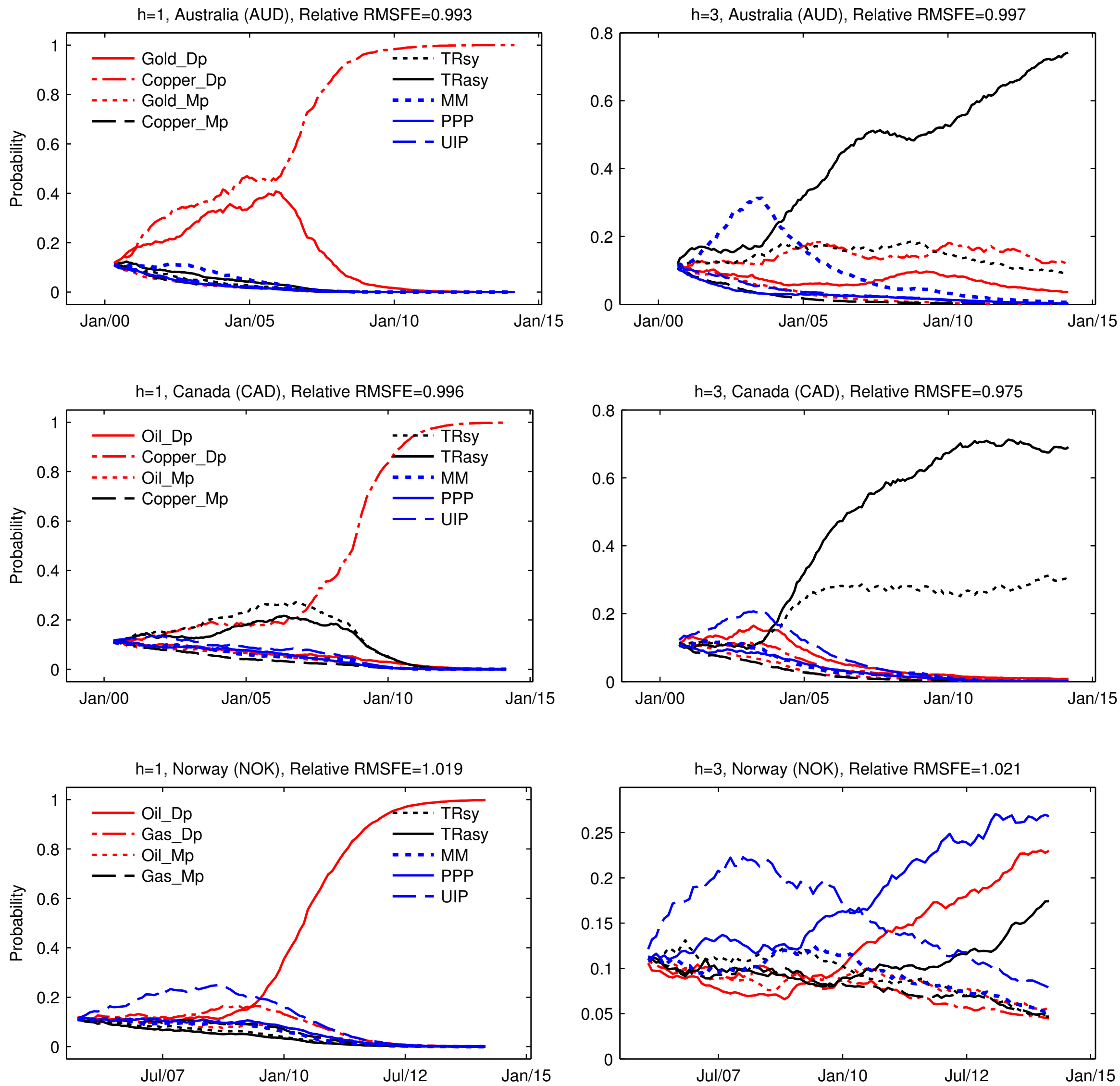
Contrary to the BMA method, where forecasts from all models are averaged, in BMS only the forecasts from the model with the highest posterior probability are considered. Figure 5.6 shows which model exhibits the highest posterior probability over the OOS period. Clearly, at $h = 3$ models with macroeconomic fundamentals display the largest weight. In contrast, at $h = 1$ the pattern is mixed for Canada and Norway, with models featuring commodity prices and macroeconomic variables having the largest posterior probability in parts of the OOS period. In the case of Japan, the switches

Figure 5.4: Weights Associated with Daily Observations on Commodity Prices Fluctuations



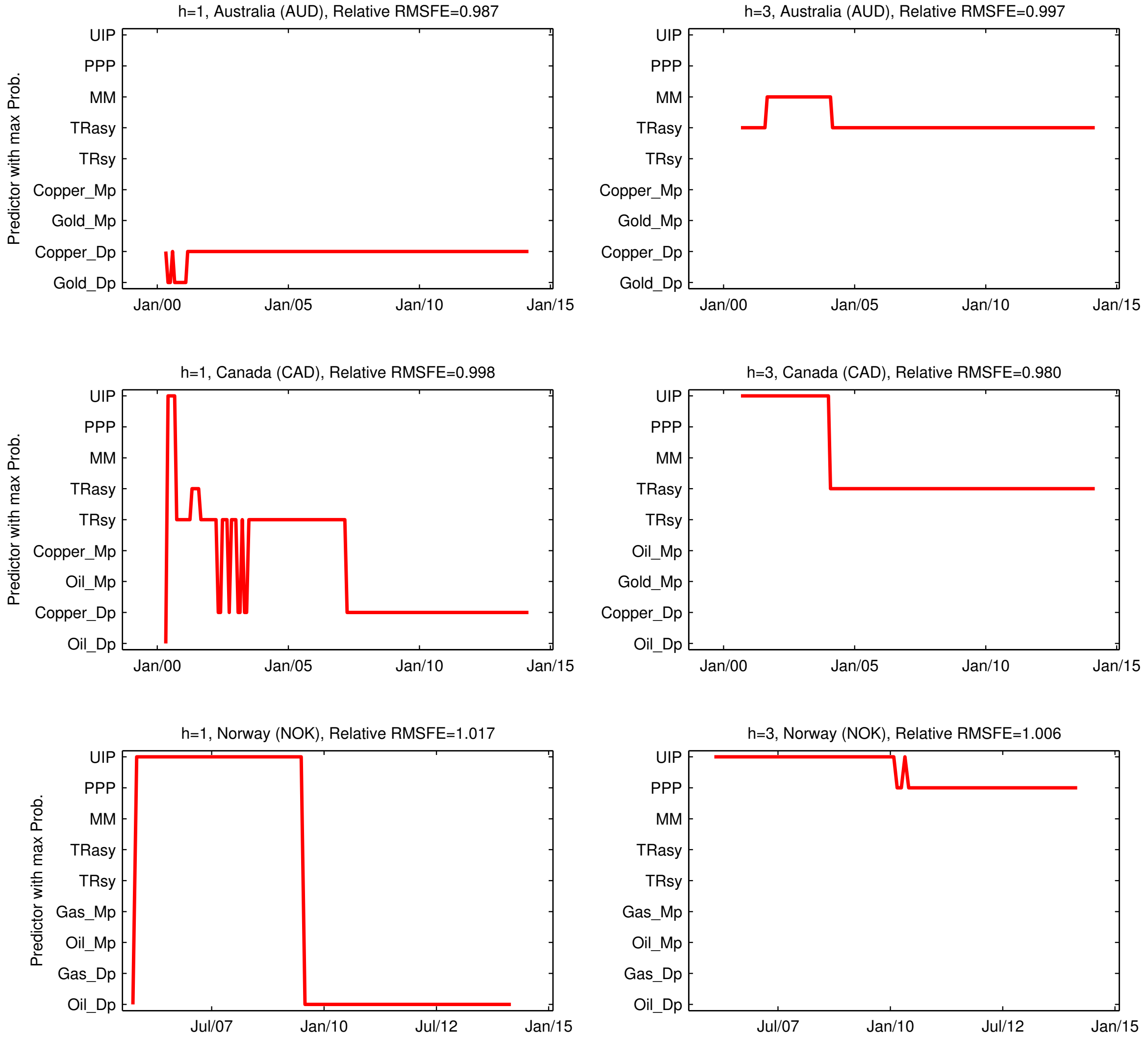

Notes: The Figure displays the weights associated with daily fluctuations on commodity prices within a month. The illustration is based on the MIDAS regression parameter estimates for the last out-of-sample forecast and a representative selection of commodity-currency pairs. To compute the weights, we fit a MIDAS regression with the exponential Almon polynomial function using the relevant country-specific daily commodity price as a regressor. In the regression we allow for the previous 21 daily observations on the daily regressor to affect the end-of-month change in the exchange rate, with each daily observation carrying its specific weight. In all MIDAS regressions we assume 22 working days within a month, so that the $22^{nd}$ day corresponds to the end-of-month. Once the parameters are estimated we then compute the weights using the polynomial function. In each plot's heading, we indicate the relative RMSFE obtained over the entire-out-of sample period using the MIDAS regression on the specified daily commodity price.

Figure 5.5: Weights Associated with Each Predictor in the BMA with Daily and Monthly Predictors




Notes: The figure presents the weights associated with each predictor in the Bayesian Model Averaging (BMA) method with daily and monthly regressors. In the plot's legend, the suffixes Dp and Mp attached to the commodity prices denote daily and monthly prices respectively. The other monthly regressors include fundamentals from the symmetric (TRsy) and asymmetric Taylor rules (TRasy), the Monetary Model - MM; Purchasing Power Parity - PPP, and Uncovered Interest Rate Parity - UIP. When the regressor corresponds to a daily commodity price, the posterior probability corresponds to the MIDAS model based on that specific commodity. In each plot's heading, we indicate the relative RMSFE corresponding to the entire-out-of sample period obtained from the BMA with all predictors - identical to the numbers in the eighth column of Table 5.3.

Figure 5.6: Predictors with the Highest Weight at Each Point in Time (Among Daily and Monthly)




Notes: The figure shows the models (defined according to the predictor they include) with the largest posterior probability at each point in time. The forecasts from the Bayesian Model Selection (BMS) method are based on these models. In the graph's vertical axis, the suffixes Dp and Mp attached to the commodity prices denote daily and monthly prices respectively. The other monthly regressors include fundamentals from the symmetric (TRsy) and asymmetric Taylor rules (TRasy), the Monetary Model - MM; Purchasing Power Parity - PPP, and Uncovered Interest Rate Parity - UIP. When the regressor corresponds to a daily commodity price, the model with the largest posterior probability corresponds to the MIDAS regression. In each plot's heading, we indicate the relative RMSFE corresponding to the entire-out-of sample period obtained from the BMS among all predictors - identical to the numbers in the ninth column of Table 5.3.

occur between models with daily oil prices and monthly oil prices. For the Australian dollar, the MIDAS model with copper prices displays the highest posterior probability. And we recall that BMS produced the largest reduction in the relative RMSFE for this currency (1.3%).

Overall, given that the weights in the BMA and BMS are computed from the model's realized likelihood, our results suggest that daily commodity prices are relatively more informative about monthly changes in exchange rates at $h = 1$. Conversely, traditional macroeconomic fundamentals, such as those from the Taylor rules, MM, PPP and UIP, exhibit higher predictive content for exchange rates at $h = 3$. These latter findings are typical in the exchange rate literature - see Rossi (2013) for recent accounts. Our former findings, however, are novel and open up new venue for improving exchange rate forecasts.

## 5.5 Robustness Checks

We verified the robustness of our findings to two situations. First, to the choice of priors in our Bayesian estimation methods. Second, we inspect if our results are due to the specific polynomial function that we employ in the MIDAS models. In essence, as we elaborate next, our previous results hold up strongly.

### 5.5.1 Sensitivity to Change in Priors

Our baseline priors are sensible but relatively diffuse. One may question the role of these priors in driving the results we obtain. To address this potential concern, we focus on a setting that assigns equal weights to the prior and the data in the posterior covariance matrix. In particular, we redefine the prior for the coefficients vector ($\gamma$) to conform with a g-prior type:

$$\gamma \sim N\left(0_{(n\times 1)}, \eta\left[gX'X\right]^{-1}\right). \tag{5.26}$$

Consistent with our objective we set $g = 1$ (see Koop, 2003, Ch. 11). Point forecast results from models estimated with this prior setting are presented in Table 5.4. If anything, the baseline line results for the MIDAS models at 1-month forecasting horizon are stronger. At this horizon, we detect statistically significant improvements over the RW for the Australian and Canadian dollars with copper prices; and for the Japanese yen and oil prices. As before, regressions conditioned on Taylor rule fundamentals deliver the best forecasting performance at 3-months forecasting horizons. And similar to our baseline findings, when we combine both, daily and monthly regressors, our results highlight the gains from such combinations in terms of relative forecast accuracy improvements.

### 5.5.2 Sensitivity to the Choice of the Polynomial Function

The MIDAS regressions that generate our main results employ the exponential Almon lag polynomial. To examine whether the results are sensitive to this choice we experiment with the non-normalized Almon lag polynomial (with three degrees):

$$B(k;\theta) = \sum_{i=0}^{3} \theta_i k^i, \tag{5.27}$$

Using this polynomial, we can rewrite (5.2) as:

$$\Delta s_{t+h} = \beta_0 + \sum_{k=0}^{K-1} \sum_{i=0}^{p} \theta_i k^i L^{k/m} x_t^{(m)} + \varepsilon_{t+h}; \tag{5.28}$$

Table 5.4: Forecast Evaluation: Sensitivity to the Choice of the Prior

| | Relative RMSFE and CW-test | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | AUD | CAD | NOK | YEN | AUD | CAD | NOK | YEN |
| | h=1 | | | | h=3 | | | |
| Daily regressors | | | | | | | | |
| ΔOil_Dp | - | 1.011 | 1.026 | **0.983**** | - | 1.005 | 1.003 | 1.001 |
| ΔGold_Dp | 1.007 | - | - | - | 1.010 | - | - | - |
| ΔCopper_Dp | **0.987*** | **0.988*** | - | - | **0.999** | 1.018 | - | - |
| ΔGas_Dp | - | - | 1.002 | - | - | - | 1.015 | - |
| ΔDP_index | - | - | - | 1.003 | - | - | - | 1.008 |
| Monthly regressors | | | | | | | | |
| ΔOil_Mp | - | 1.001 | 1.002 | 1.002 | - | 1.000 | 1.006 | 1.005 |
| ΔGold_Mp | 1.003 | - | - | - | 1.007 | - | - | - |
| ΔCopper_Mp | 1.001 | **0.999** | - | - | 1.002 | **0.998** | - | - |
| ΔGas_Mp | - | - | 1.003 | - | - | - | 1.006 | - |
| ΔMP_index | - | - | - | 1.002 | - | - | - | 1.010 |
| TRsy | 1.000 | **0.993**** | 1.007 | 1.006 | **0.998** | **0.981**** | 1.014 | 1.009 |
| TRasy | **0.999** | **0.993**** | 1.000 | 1.001 | **0.995** | **0.983**** | **0.999** | **0.997** |
| MM | 1.005 | 1.003 | 1.006 | 1.003 | 1.019 | 1.015 | 1.025 | 1.009 |
| PPP | 1.004 | 1.001 | 1.001 | 1.000 | 1.006 | 1.003 | 1.013 | 0.999 |
| UIP | 1.001 | 1.003 | 1.012 | **0.998** | 1.006 | 1.006 | 1.026 | **0.997** |
| Forecast combination (daily and monthly regressors) | | | | | | | | |
| BMA | **0.991** | **0.993** | 1.015 | **0.991**** | **0.997** | **0.988**** | 1.011 | 1.002 |
| BMS | **0.989*** | **0.994** | 1.013 | **0.991**** | **0.998** | **0.992**** | 1.005 | 1.005 |
| OptPool | **0.986*** | **0.990*** | 1.030 | **0.992**** | 1.001 | **0.989*** | 1.018 | 1.017 |

**Notes:** The Table presents the Root Mean Squared Forecast Error (RMSFE) of the commodity or fundamental-based forecasting model relative to the RMSFE of the driftless Random Walk (RW). Here, we use a g-type prior for the coefficients vector in our Bayesian estimation method. The last three rows show results for forecast combination methods. In all cases, values less than 1 (one) indicate that the commodity or fundamental-based model generates a lower RMSFE than the RW, hence, it forecasts better than the RW. The Table also reports the CW-test with asterisks ( *10%, **5%, ***1%) denoting the level of significance at which the null hypothesis of equal RMSFE is rejected, favouring the alternative that the commodity or fundamental-based model has a lower RMSFE. In all models, the forecasts are generated recursively for $h$-month(s)-ahead change in the exchange rate. The list of daily regressors include, change in daily prices (Dp) of oil, gold, cooper, gas, and a daily commodity price index (ΔMP_index). In the monthly regressors group we have a similar set of commodities, but also fundamentals from the symmetric Taylor rule - TRsy, the asymmetric Taylor rule - TRasy, the Monetary Model - MM, Purchasing Power Parity - PPP, and Uncovered Interest Rate Parity - UIP. The currency codes in the first row denote the Australian dollar (AUD), the Canadian dollar (CAD), the Norwegian Kroner (NOK), and the Japanese YEN. The forecast evaluation period is 1998M11+$h$ to 2014M3 for all currencies, except the NOK (2005M7 +$h$ to 2014M3).

Table 5.5: Forecast Evaluation: Sensitivity to the Choice of the Polynomial Function

| | Relative RMSFE and CW-test | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | AUD | CAD | NOK | YEN | AUD | CAD | NOK | YEN |
| | h=1 | | | | h=3 | | | |
| Daily regressors | | | | | | | | |
| ΔOil_Dp | - | 1.005 | 1.012 | **0.991*** | - | 1.006 | 1.003 | 1.000 |
| ΔGold_Dp | 1.022 | - | - | - | 1.008 | - | - | - |
| ΔCopper_Dp | **0.994** | **0.993** | - | - | **0.981*** | **0.986** | - | - |
| ΔGas_Dp | - | - | **0.995** | - | - | - | 1.016 | - |
| ΔDP_index | - | - | - | **0.994*** | - | - | - | 1.016 |
| Monthly regressors | | | | | | | | |
| ΔOil_Mp | - | 1.001 | 1.004 | 1.002 | - | **0.998** | 1.006 | 1.004 |
| ΔGold_Mp | 1.004 | - | - | - | 1.008 | - | - | - |
| ΔCopper_Mp | 1.000 | 1.000 | - | - | 1.002 | 1.001 | - | - |
| ΔGas_Mp | - | - | 1.001 | - | - | - | 1.007 | - |
| ΔMP_index | - | - | - | 1.001 | - | - | - | 1.012 |
| TRsy | **0.998** | **0.991**** | 1.005 | 1.007 | 1.000 | **0.980**** | 1.011 | 1.010 |
| TRasy | 1.000 | **0.992**** | 1.001 | **0.997** | **0.997** | **0.983**** | 1.000 | **0.996** |
| MM | 1.007 | 1.003 | 1.009 | 1.000 | 1.016 | 1.015 | 1.023 | 1.010 |
| PPP | 1.003 | 1.001 | 1.002 | 1.001 | 1.008 | 1.002 | 1.012 | 1.000 |
| UIP | 1.003 | 1.002 | 1.011 | 1.002 | 1.008 | 1.007 | 1.022 | 1.000 |
| Forecast combination (daily and monthly regressors) | | | | | | | | |
| BMA | **0.992** | **0.996** | 1.019 | **0.994*** | **0.993** | 1.000 | 1.011 | 1.002 |
| BMS | **0.990** | **0.996** | 1.017 | **0.994** | **0.991*** | 1.003 | 1.007 | 1.005 |
| OptPool | **0.994** | **0.993** | 1.012 | **0.992** | **0.993** | 1.005 | 1.030 | 1.014 |

**Notes:** The Table presents the Root Mean Squared Forecast Error (RMSFE) of the commodity or fundamental-based forecasting model relative to the RMSFE of the driftless Random Walk (RW). Here, models with daily commodity prices (MIDAS models) are estimated with the Almon lag polynomial. The interpretation of the entries in the Table is similar to Table 5.4.

and estimate the model via a two-blocks Gibbs Sampling as in Pettenuzzo et al. (2015). We use the same g-type prior as in the previous robustness check. As shown in Table 5.5, the baseline results only change to somewhat poor forecasting performance for the Canadian dollar in our forecast combination methods at $h = 3$. Our conclusion that MIDAS models with daily commodity prices help in forecast accuracy remain largely unaffected.

## 5.6 Conclusion

In this chapter we exploit the properties of daily commodity price changes to predict commodity currencies. Using MIDAS models in a Bayesian setting, we regress monthly changes in exchange rates on daily fluctuations of commodity prices. The smoothing function that underlies a MIDAS model allows for each daily fluctuation to affect the end-of-month exchange rate change with a possibly different weight. We put forward the random walk Metropolis-Hasting algorithm, as a new technique to estimate MIDAS models with the particular smoothing function we employ - the exponential Almon lag polynomial. Our use of Bayesian methods also allow us to account for potential instabilities in forecasting performance and examine the degree of informativeness of the daily commodity price changes, as opposed to the monthly commodity prices and standard macroeconomic fundamentals.

Focusing on data for Australia, Canada, Norway, and Japan we first find evidence favouring daily commodity prices fluctuations in terms of providing more accurate forecasts than the naive no-change benchmark. In particular, daily fluctuations in copper prices yield point forecast improvements for the Australian and Canadian dollar at 1-month horizon. We also detect significant predictive content of daily oil prices for the non-commodity currency we examine, the Japanese yen. In contrast and as reported in other studies, we identify rare instances in which monthly commodity prices changes lead to systematic point forecast improvements. However, consistent with the existing evidence, macroeconomic fundamentals derived from Taylor rules do exhibit predictive power for some commodity currencies, especially at long (3-month) forecasting horizons.

We then proceed and combine forecasts from regressions based on daily and monthly commodity prices and monthly traditional macroeconomic fundamentals, in an effort to account for time-variation in forecasting ability of our predictors. Here our empirical findings reveal the usefulness of such combinations in terms of forecast accuracy improvements relative to our benchmark. Our results also point at the importance of accounting for the full forecast distribution, since in terms of density forecasts, our predictions are always better than those from the RW. Finally, when we inspect the weights underlying our forecast combinations approaches we find that daily commodity prices are relatively more informative about 1-month changes in the exchange rate than monthly commodity prices or the typical macroeconomic variables. As the forecast horizon increases to 3-months ahead, the role of macroeconomic fundamentals becomes salient. Overall, our results endorse the importance of exploiting the properties of high-frequency information in pinning down exchange rate predictability.

# Chapter 6

# Concluding Remarks

Many recent exchange rate studies suggest that macroeconomic fundamentals exhibit an erratic forecasting power for exchange rates.[1] And one potential reason for such erratic predictive ability is the presence of instabilities in the relationship between exchange rates and fundamentals (Bacchetta and van Wincoop, 2013 and Rossi, 2013).

In this thesis we tackle the problem of forecasting exchange rates in the presence of instabilities from several perspectives. First, we consider the idea that the process that governs the fundamentals themselves, and the relationship between fundamentals and exchange rates, change rapidly over time. Following this idea, we employ Time-Varying Parameters (TVP) models to estimate exchange rate fundamentals and to forecast. Second, we exploit a more general and flexible approach to incorporate both, time-variation in the parameters of the forecasting regression, and time-changing sets of exchange rate fundamentals. Within this flexible TVP approach, the thesis identifies some of the major sources of instability that affect the performance of the empirical exchange rate models. Third, the thesis advances two bootstrap-based approaches to uncover the time-specific conditioning information for predicting exchange rates. Forth, we consider mixed-frequency dynamics between commodity-currency exchange rates and commodity prices.

With the exception of the material based on the bootstrap, in most of the perspectives we rely heavily on Bayesian methods. In this respect, we also introduce the random walk Metropolis-Hasting algorithm, as a new tool to estimate a certain class of MIDAS models. Juxtaposed with frequentist methods, Bayesian techniques have the advantage of providing a framework to account for model and parameter uncertainty, a key aspect of our analyses.

Overall, what we find can be summarized in the following four points. First, allowing for time-variation in parameters appears to be useful mostly at long horizons, typically beyond 1-month. In particular, our flexible TVP approach generates more accurate forecasts than the random walk (RW) benchmark at the 3- and 12- forecasting

[1] A short list of these studies includes Berge (2013), Bacchetta and van Wincoop (2004, 2013), Cheung et al. (2005), Fratzscher at al.(2015), Rogoff and Stavrakeva (2008), Rossi (2013), and Sarno and Valente (2009).

horizons, but not at the 1-month horizon. Using TVP models to account for instabilities in the fundamentals and in the forecasting regressions has advantages for forecasting at 4-quarters ahead and beyond, and mainly in the more recent turbulent times. In similar forecasting settings, our experiments with the usual constant parameter models, including in a panel setting, yielded weaker results than our TVP regressions. Hence, we remain more optimistic about our TVP models than constant-parameter models in improving exchange rate models forecasting ability at long horizons.

Second, at short horizons (e.g., 1-month) our results suggest that exploiting the properties of high-frequency data, such as daily commodity prices, is a more fruitful route to forecast improvements. In this respect, the Mixed-Data Sampling (MIDAS) framework is a particularly useful tool, as it enables to pin down the often short-lived effects of commodity prices on exchange rates. We also affirm that using commodity prices sampled at low-frequencies is fruitless, unless if applied in a Bayesian forecast combination setting.

Third, at short horizons, identifying the country-specific macro fundamentals with robust predictive content for exchange rates is challenging. Often, models overfit in sample but exhibit poor out-of-sample forecasting performance. The problem is cumbersome one, as we are faced with samples of finite length. Our bootstrap-based forecasting methods appeared effective in revealing fundamentals with predictive power, leading to forecasting gains at 1-month horizon.

Fourth, in terms of the obstacles to exchange rate regressions' predictive ability, we identify the uncertainty in coefficients' estimation and the uncertainty about the precise degree of coefficients variability to incorporate in the regressions, as the main limiting factors at short-horizons. In this sense, time-variation in parameters is an important source of instability in exchange rate models. The uncertainty from observational variance, which provides a measure of random fluctuations in the data relative to the predicted trend component, and the uncertainty with respect to the choice of the predictor appear to be minor obstructing factors. Or at least, they can be successfully embedded in our TVP flexible approach. Thus, our finding of low uncertainty with respect to selection of predictor means that our methods are able to pin down the specific country-horizon fundamentals that are relevant for the corresponding exchange rate. In general, these specific country-horizon fundamentals differ between forecasting horizons and between countries.

# Appendices

## A Present Value Models

Present value models formalize the notion that exchange rates are relative prices of different assets, and as such, they are determined by equilibrium in asset markets. According to this asset market approach view, the spot rate is determined by the expected future exchange rate change and other factors (fundamentals) affecting the demand and supply of a foreign currency (Engel and West, 2005). In notation:

$$s_t = \lambda E_t \Delta s_{t+1} + \Omega_t, \qquad \lambda > 0, \tag{A.1}$$

or equivalently,

$$s_t = (1-b)\Omega_t + bE_t s_{t+1}, \qquad b = \frac{\lambda}{1+\lambda}, \qquad 0 < b < 1, \tag{A.2}$$

where $s_t$ is the log exchange rate, $E_t$ is the expectation operator, $b$ is a discount factor, $\Omega_t$ represents the exchange rate fundamentals or factors influencing the demand for and supply of foreign currency. The fundamentals in $\Omega_t$ may consist of observable and/or non-observable factors, or linear or non-linear combination of several factors. Further, the non-observable component may include omitted variables that in principle could be measured (e.g., measurement error of the observable fundamentals).

To derive the present value relation, we follow Engel and West (2005) and iterate Eq. (A.2) forward up to time $T$ using the Law of Iterated Expectations (i.e., $E_t E_{t+i} s_{t+i+1} = E_t s_{t+i+1}$ for all $i > 0$). For example, the first iteration for $t = 1$ yields the following forward solution:

$$\begin{aligned} E_t s_{t+1} &= (1-b)E_t\Omega_{t+1} + bE_t E_{t+1} s_{t+2} \\ &= (1-b)E_t\Omega_{t+1} + bE_t s_{t+2}, \end{aligned} \tag{A.3}$$

which substituted for $E_t s_{t+1}$ in Eq. (A.2) produces:

$$s_t = (1-b)\Omega_t + b(1-b)E_t\Omega_{t+1} + b^2 E_t s_{t+2}. \tag{A.4}$$

Proceeding in this fashion for the remaining periods, $t = 2...T$, gives:

$$s_t = (1-b)E_t \sum_{i=0}^{T-1} b^i \Omega_{t+i} + b^T E_t s_{t+T}. \tag{A.5}$$

According to Eq. (A.5), the spot exchange rate is a function of the expectation of the exchange rate in the distant future and the discounted stream of future fundamentals. Assuming that expectations about future asset prices are always fulfilled, i.e., there are no-bubbles and therefore the term $b^T E_t s_{t+T}$ goes to zero as $T \to \infty$, then:

$$s_t = (1-b)E_t \sum_{i=0}^{T-1} b^i \Omega_{t+i}. \tag{A.6}$$

which links the exchange rate only to the discounted future fundamentals.

The dynamic implications of this present value relationship can be easily derived under the assumption that fundamentals follow a particular process. Positing, for instance, that fundamentals follows a random walk process, i.e., $\Omega_t = \Omega_{t-1} + \varepsilon_t^f$, and computing $E_t\Omega_{t+i}$ from this process, Eq. (A.6) becomes:

$$s_t = \Omega_t + \varepsilon_t^f,$$

which relates the exchange rate to current fundamentals and some unexpected disturbances in the market. The several empirical exchange rate models we consider are, therefore, founded in the present-value representation.

# B Interest Rate Differentials Implied by Taylor Rules

In this Appendix we derive interest rate differentials implied by Taylor rules. Taylor (1993) suggested the following rule for monetary policy:

$$i_t^T = \pi_t + \tau_1(\pi_t - \pi^T) + \tau_2 \overline{y}_t + r^T, \tag{B.1}$$

where $i_t^T$ is the target for the nominal short-term interest rate set by the central bank, $\pi_t$ is the inflation rate, $\pi^T$ is the target inflation rate, $\overline{y}_t = (y_t - y_t^p)$ is the output gap measured as deviation of actual GDP level ($y_t$) from its potential ($y_t^p$), and $r^T$ is the equilibrium real interest rate.[2] In Eq. (B.1) the central bank rises the short-term interest rate when inflation is above the target and/or output is above its potential level. In Taylor's (1993) formulation, $\tau_1 = 0.5$, $\pi^T = 2\%$, $\tau_2 = 0.5$, and $r^T = 2\%$.

Rearranging Eq. (B.1) by aggregating the constant parameters, $r^T$and $\pi^T$, and

[2]Under the assumption that the target for the nominal interest rate is always attained, there is no difference between the actual and the target interest rate (Molodtsova and Papell, 2009).

collecting the $\pi_t$ terms we obtain:

$$i_t^T = \delta_0 + \delta_1 \pi_t + \delta_2 \overline{y}_t, \tag{B.2}$$

where $\delta_0 = r^T - \tau_1 \pi^T$, $\delta_1 = (1+\tau_1)$, and $\delta_2 = \tau_2$. In Eq. (B.2) an increase in inflation, for instance by 1%, calls for more than 1% augment in the short-term nominal interest rate by the central bank, since $\delta_1 = (1+\tau_1)$. The Taylor principle is therefore maintained.

Following Clarida et al. (1998) it is typical to assume that most countries, apart from the U.S., target the real exchange rate $q_t$. Hence, Eq. (B.2) becomes:

$$i_t^T = \delta_0 + \delta_1 \pi_t + \delta_2 \overline{y}_t + \delta_3 q_t, \tag{B.3}$$

where $q_t = s_t + p_t^* - p_t$, $s_t$ is the log exchange rate (home price of foreign currency), $p_t$ is the log price level, and asterisk (*) indicates the foreign country. Including $q_t$ in the rule is consistent with the idea that the monetary authority is concerned with exchange rate deviation from the level implied by Purchasing Power Parity (PPP); with an increase in $q_t$ being associated with a rise in $i_t^T$ (see Engel and West, 2005).

We can further extend Eq. (B.3) by assuming that central banks set monetary policy at each period by adjusting the actual rate to eliminate a fraction $(1-\theta)$ of the gap between the current interest rate target and its recent past level (Clarida et al., 1998):

$$i_t = (1-\theta) i_t^T + \theta i_{t-1} + \varepsilon_t. \tag{B.4}$$

Using Eq. (B.3) in (B.4) yields:

$$i_t = \phi_c + \phi_1 \pi_t + \phi_2 \overline{y}_t + \phi_3 q_t + \phi_4 i_{t-1} + \varepsilon_t \tag{B.5}$$

where $\phi_c = (1-\theta)\delta_0$, $\phi_1 = (1-\theta)\delta_1$, $\phi_2 = (1-\theta)\delta_2$, and $\phi_3 = (1-\theta)\delta_3$; $\phi_4 = \theta$.

In contrast with the immediate adjustment in the policy interest rate implied by equations (B.2) and (B.3), in Eq. (B.5) the change in the interest rate is gradual. In response to an inflation rate that is above the target, the central bank increases the interest rate by $(1-\theta^p)\delta_1$ at each $p$ period, with $p = \{1 \ldots . P\}$. As $p$ increases, the maximum change in policy interest rate converges to $\delta_1$, also satisfying the Taylor principle (see Molodtsova and Papell, 2009).

Let Eq. (B.5) denote the home country's Taylor rule. The foreign country is the U.S., and following Clarida et al. (1998) and Engel and West (2005), it is assumed that the Federal Reserve Bank does not target the real exchange rate. Hence, its Taylor rule is similar to the expression in Eq. (B.5), except that the real exchange rate is excluded. Subtracting from the home country's Taylor rule the foreign counterpart yields the following interest rate differentials equation:

$$
\begin{aligned}
i_t - i_t^* = {} & \phi_0 + (\phi_1 \pi_t + \phi_2 \overline{y}_t + \phi_3 q_t + \phi_4 i_{t-1}) \\
& - (\phi_1^* \pi_t^* + \phi_2^* \overline{y}_t^* + \phi_4^* i_{t-1}^*) + \mu_t,
\end{aligned} \tag{B.6}
$$

where the term $\phi_0 = \phi_c - \phi_c^*$, and $\mu_t = \varepsilon_t - \varepsilon_t^*$.

Table B.1: Interest Rate Differentials Implied by Taylor Rules

| **Assumption** | **Interest rate differentials specification** |
|---|---|
| *TRon: Homogeneous rule, asymmetric and without interest rate smoothing.* (i) The equilibrium real interest rate and the inflation target of the home and foreign country are identical, hence their difference is zero; (ii) The coefficients on inflation and the output gap are equal between home and foreign country; (iii) Central banks do not smooth interest rate; (v) The home central bank targets the real exchange rate. In eq. (B.6): $\phi_0 = 0$; $\alpha_1 = \phi_1 = \phi_1^*$; $\alpha_2 = \phi_2 = \phi_2^*$; $\phi_4 = \phi_4^* = 0$. | $\begin{aligned} i_t - i_t^* &= \alpha_1(\pi_t - \pi_t^*) \\ &+ \alpha_2(\overline{y}_t - \overline{y}_t^*) \\ &+ \phi_3 q_t + \mu_t \end{aligned}$ |
| *TRos: Homogeneous rule, asymmetric and with interest rate smoothing.* (i) The equilibrium real interest rate and the inflation target of the home and foreign country are identical, hence their difference is zero; (ii) The coefficients on inflation, the output gap and the interest rate smoothing are equal between home and foreign country; and (iii) The home central bank targets the real exchange rate. In eq. (B.6): $\phi_0 = 0$; $\alpha_1 = \phi_1 = \phi_1^*$; $\alpha_2 = \phi_2 = \phi_2^*$; $\alpha_3 = \phi_4 = \phi_4^*$. | $\begin{aligned} i_t - i_t^* &= \alpha_1(\pi_t - \pi_t^*) \\ &+ \alpha_2(\overline{y}_t - \overline{y}_t^*) \\ &+ \alpha_3(i_{t-1} - i_{t-1}^*) \\ &+ \phi_3 q_t + \mu_t \end{aligned}$ |
| *TRen: Heterogeneous rule, asymmetric and without interest rate smoothing.* (i) The equilibrium real interest rate and the inflation target of the home and foreign country are identical, hence their difference is zero; (ii) The coefficients on inflation and the output gap are allowed to differ between home and foreign country; (iii) Central banks do not smooth interest rate; and (iv) The home central bank targets the real exchange rate. In eq. (B.6): $\phi_0 = 0$; $\phi_4 = \phi_4^* = 0$. | $\begin{aligned} i_t - i_t^* &= \phi_1 \pi_t - \phi_1^* \pi_t^* \\ &+ \phi_2 \overline{y}_t - \phi_2^* \overline{y}_t^* \\ &+ \phi_3 q_t + \mu_t \end{aligned}$ |

Notes: All the assumptions are relative to equation (B.6) in this Appendix. That is, $i_t - i_t^* = \phi_0 + (\phi_1 \pi_t + \phi_2 \overline{y}_t + \phi_3 q_t + \phi_4 i_{t-1}) - (\phi_1^* \pi_t^* + \phi_2^* \overline{y}_t^* + \phi_4^* i_{t-1}^*) + \mu_t$. The alternative specifications are then derived in line with the assumptions in the first column of the Table.

In Eq. (B.6) the constant parameter $\phi_0$ allows for the equilibrium real interest rates and inflation targets to differ across home and foreign countries. By contrast, if we assume that the equilibrium real interest rate and the inflation target of the home and foreign country are identical, then the constant is excluded. Also in Eq. (B.6), all coefficients are heterogeneous, only the home central bank targets the real exchange rate and both countries limit volatilities in interest rates. In terms of the parameters of Eq. (B.6) we have: $\phi_0 \neq 0; \phi_1 \neq \phi_1^*; \phi_2 \neq \phi_2^*; \phi_4 \neq \phi_4^*; \phi_3 \neq 0$. In Table (B.1) we relax some of the assumptions in Eq. (B.6) to derive three alternative specifications that we use in our empirical exercise in Chapter 2. These variants are inspired by Engel and West (2005), Molodtsova and Papell (2009), and Engel et al. (2008).[3]

# C Bayesian Estimation of Time-Varying Parameter Models

This Appendix describes the Bayesian approach we pursue to estimate our Time-Varying Parameter (TVP) models in Chapter 2. We present the prior hyperparameters, the conditional posterior distributions, and the steps or algorithm used to draw from these conditional distributions. Our exposition draws mainly from Kim and Nelson (1999, Ch. 3 & 8) and Blake and Mumtaz (2012, Ch. 3).

Our TVP models have the following general state-space representation:

$$\begin{aligned} y_t &= H_t\beta_t + Az_t + e_t, \quad \text{observation equation;} && \text{(C.1)} \\ \beta_t &= \mu + F\beta_{t-1} + \upsilon_t, \quad \text{transition equation;} && \text{(C.2)} \end{aligned}$$

where $e_t \sim i.i.d.(0, R)$, $\upsilon_t \sim i.i.d.(0, Q)$, and $Cov(e_t, \upsilon_t) = 0$. Further, $y_t$ is a $(T \times 1)$ vector of observations on our regressand; $\beta_t$ is a $(k \times 1)$ vector of unobserved state variables (e.g. the time-varying coefficients); $H_t$ is an $(n \times k)$ matrix with elements that are not fixed or given as data; $z_t$ is an $(r \times 1)$ vector of exogenous variables with time-invariant coefficients $A$. In terms of our precise TVP specifications in Sections (2.2) and (2.3), $y_t \equiv \Delta s_t$ and $y_t \equiv i_t - i_t^*$, $H_t$ contains the respective explanatory variables, $Az_t = 0$, $\mu = 0$, and $F$ is an identity matrix ($I_k$), refer to Eq. (2.1) and Eq. (2.5) in Chapter 2.

**Priors hyperparameters and initial conditions**

The form of our TVP models suggests that we need priors for the variance $R$ of the measurement or observation equation and the variance-covariance matrix $Q$ of the transition equation. In addition, to recover the unobserved state variable $\beta_t$ we need initial conditions or starting values for the Kalman filter (i.e., the initial state, $\beta_{0|0}$, and its initial variance $P_{0|0}$). See Box C.1 for details of the Kalman filter.

[3]Engel and West (2005) derive a Taylor rule specification similar to the one denoted TRos in Table (B.1). Molodtsova and Papell (2009) consider 16 alternative specifications, including the three included in our Table. Engel et al. (2008) consider the specification denoted TRon in the Table, with posited coefficients as follows: $\alpha_1 = 1.5, \alpha_2 = 0.1$ and $\alpha_3 = 0.1$.

To parameterize the prior distributions and initial conditions we use pre-sample information. Specifically, we use a training sample of $T_0 = 20$ observations to estimate via OLS estimator a fixed-coefficient model which is a counterpart to Eq. (C.1). The estimated coefficients and their corresponding covariance matrix are set as initial conditions for the Kalman filter. In notation:

$$\beta_{0|0} \equiv \beta_{OLS} = (H_{0t}^{'}H_{0t})^{-1}(H_{0t}^{'}y_{0t}), \tag{C.3}$$

$$P_{0|0} \equiv P_{OLS} = \Sigma_0 \otimes (H_{0t}^{'}H_{0t})^{-1}, \tag{C.4}$$

where $\beta_{OLS}$ and $P_{OLS}$ are, respectively, the coefficients' vector and covariance matrix from an OLS regression, and $\Sigma_0 = (y_{0t} - H_{0t}\beta_0)^{'}(y_{0t} - H_{0t}\beta_0)/(T_0 - k)$.

The prior for $Q$ is inverse Wishart, with $T_0$ degrees of freedom and $Q_0$ scale matrix, i.e., $P(Q) \sim IW(Q_0, T_0)$. This prior influences the amount of time-variation in the coefficients. A large value for the scale matrix $Q_0$ is consistent with more fluctuation in the coefficients. We set $Q_0 = P_{0|0} \times T_0 \times \tau$, where $\tau$ is a scaling factor that reflects our beliefs about the preciseness of $P_{0|0}$. Since our training sample $T_0$ is small, we consider that the estimate of $P_{0|0}$ is very imprecise and set $\tau = 3.510^{-6}$ for all models.[4] This reasoning accords with Blake and Mumtaz (2012, Ch. 3).

---

**Box C.1: The Kalman Filter**

Consider our state-space model given by the system of equations (C.1) and (C.2). The Kalman filter is a recursive algorithm for computing the optimal estimate of $\beta_t$ given an appropriate information set and knowledge of the other parameters of the state-space. Let $H, A, R, \mu, F, Q$ be the known parameters. The algorithm consists in the steps summarized in Diagram C.1.

The first step is to define initial conditions. For a stationary state vector, the unconditional mean and its associated covariance matrix may be employed as initial conditions. For non-stationary processes, unconditional means and covariance matrixes do not exist (Kim and Nelson, 1999 Ch. 8). In this case the initial condition for the state variable $\beta_{0|0}$ may be defined arbitrarily. However, to indicate a high uncertainty surrounding this arbitrary defined value, we must set the diagonal elements of the covariance matrix $P_{0|0}$ to a very large number. For more details on initial conditions see Kim and Nelson, (1999 Ch. 3).

In the second step, i.e., for period $t = 1$ we can now form an optimal prediction of $y_{1|0}$ after computing $\beta_{1|0}$ and its associated covariance matrix, $P_{1|0} = FP_{0|0}F^{'} + Q$. Note that the subscripts make it clear that we are conditioning on the information set at $t = 0$, i.e., contained in our prior initial conditions, $\beta_0$ and $P_0$.

In the third step, we use the observed value of $y_t$ at $t = 1$ to compute the prediction error, $n_{1|0} = y_1 - y_{1|0}$ and its covariance matrix $f_{1|0} = HP_{t|t-1}H^{'} + R$. The information contained in the prediction error can be used to improve the initial inference about $\beta_t$. Thus in the fourth

[4] Note also that the training sample size reduces with the forecast horizon. For example, the size of the training sample used to parameterize the prior for the forecasting regression at 12-quarters-ahead is $T_0 = 20 - h$. With two coefficients ($k = 2$) to be estimated we have six degrees of freedom.

and last step, we can compute $\beta_{1|1} = \beta_{1|0} + K_t n_{1|0}$; where $K_t = P_{1|0}H' f_{1|0}^{-1}$ is the Kalman gain, which indicates the weight attributed to new information. It constitutes the ratio of the prediction error variance associated with uncertainty about $\beta_{1|0}$ and the prediction error variance of the error term $e_t$ in Eq. (C.1). A high uncertainty about $\beta_{1|0}$ implies that more weight is attributed to new information in the prediction error.

The second, third, and fourth steps are then repeated for $t = 2, 3, 4 \ldots, T$. The filter provides an optimal estimate of the state variable at each point in time.

**Diagram C.1: State-Space Model and the Kalman Filter Algorithm**

| | |
|---|---|
| Step 1: Define starting values for the state ($\beta_{t-1\|t-1}$) and its covariance matrix ($P_{t-1\|t-1}$ ) at $t = (t-1)$, i.e. Define initial conditions. | $\beta_{t-1\|t-1}$<br>$P_{t-1\|t-1}$ |
| Step 2: At $t = 1$, predict the state vector and its associated covariance matrix. | $\beta_{t\|t-1} = \mu + F\beta_{t-1\|t-1}$<br>$P_{t\|t-1} = FP_{t-1\|t-1}F' + Q$ |
| Step 3: Calculate prediction error ($n_{t\|t-1}$) and its covariance matrix ($f_{t\|t-1}$). | $y_{t\|t-1} = H\beta_{t\|t-1} - Az_t$<br>$n_{t\|t-1} = y_t - y_{t\|t-1}$<br>$f_{t\|t-1} = HP_{t\|t-1}H' + R$ |
| Step 4: Refine inference about ($\beta_{t\|t}$) via Kalman gain. | $K_t = P_{t\|t-1}H'^{-1}_{t\|t-1}$<br>$\beta_{t\|t} = \beta_{t\|t-1} + K_t n_{t\|t-1}$<br>$P_{t\|t} = P_{t\|t-1} - K_tHP_{t\|t-1}$ |
| Step 5: Repeat steps two to four for $t = 2, 3, \ldots, T$. | |

Notes: This Diagram illustrates the Kalman filtering process when the state vectors are the only unknowns. The first step involves defining the initial conditions for the recursions. In the second step the predicted state and its covariance matrix are computed. In the third step, one calculates the prediction error and the associated covariance matrix. The variances computed in the second and third steps are then used to calculate the Kalman gain, which is then employed to update the state vector. This procedure is repeated for each observation in the sample.

The prior for the variance of the measurement equation is $P(R) \sim IG(R_0, T_0 - k)$, where $R_0 = \Sigma_{OLS}$ is the scale parameter, and $(T_0 - k)$ is the prior degree of freedom. To initialize the first step of the Gibbs sampling we need starting values for $R$ and $Q$. We set them to $R_0 = \Sigma_{OLS}$ and $Q_0 = P_{0|0} \times T_0 \times \tau$.

### Conditional posterior distribution

In addition to priors and initial conditions our methods necessitate the forms of the conditional posterior distributions. The conditional posterior distribution for the state variable ($\widetilde{\beta}_T$) given the other parameters of our TVP model is given by:

$$H(\widetilde{\beta}_T|\widetilde{y}_T, R, Q) = H(\beta_T|\widetilde{y}_T) \prod_{t=1}^{T-1} H(\beta_t|\beta_{t+1}, \widetilde{y}_t), \tag{C.5}$$

where $\widetilde{\beta}_T = [\beta_1, \beta_2, ..., \beta_T]$ and $\widetilde{y}_T = [y_1, y_2, ..., y_T]$. The conditional posterior distribution of $R$ given a draw of the state variable $\beta_t$ and the other parameters is given by:

$$H(R|\beta_t, y_t, Q) \sim \Gamma^{-1}(\frac{T_0 - k + T}{2}, \frac{\theta_1}{2}), \tag{C.6}$$

where $\theta_1 = R_0 + (y_t - \beta_t H)^{'}(y_t - \beta_t H)$. The conditional posterior distribution of $Q$ given a draw of the state variable $\beta_t$ and the other parameters is:

$$H(Q|\beta_t, y_t, R) \sim IW(\overline{Q}, T + T_0), \tag{C.7}$$

where $\overline{Q} = Q_0 + (\beta_t - \beta_{t-1})'(\beta_t - \beta_{t-1})$.

**Sampling from the conditional posterior distribution**

To draw samples from the conditional posterior distributions we use the Carter and Kohn (1994) algorithm with the Gibbs sampler. The Carter and Kohn algorithm provide us with the draws of the state variable $\widetilde{\beta}_T = [\beta_1, \beta_2, ..., \beta_T]$ from its conditional posterior distribution. The key updating equations are:

$$\beta_{t|t,\beta_{t+1}} = \beta_{t|t} + K^* \times (\beta_{t+1} - \mu + F\beta_{t|t}), \tag{C.8}$$
$$P_{t|t,\beta_{t+1}} = P_{t|t} - K^* \times H^* \times P_{t|t} \tag{C.9}$$

where $\beta_{t|t}$ and $P_{t|t}$ are obtained from the Kalman filter and $K^* = P_{t|t} \times H^{*'} \times f^{-1}_{t+1|t}$. Equations (C.8) and (C.9) are substituted backwards from $T-1$, and iterating backwards to period 1. In fact, this algorithm constitutes an integral part of the Gibbs sampling framework, which comprises the following steps:

- Step 1: Conditional on $R$ and $Q$, draw $\beta_t$ from its conditional posterior distribution given in Eq. (C.5) using the Kalman filter and the Carter and Kohn algorithm. More in detail:
  - 1.1: Run the Kalman filter from $t = 1 \dots T$ to obtain the mean $\beta_{T|T}$ and the variance $P_{T|T}$ of the distribution $H(\beta_T|\widetilde{y}_T)$. Also obtain $\beta_{t|t}$ and $P_{t|t}$ for $t = 1 \dots T$.
  - 1.2: Draw $\beta_T$ from the normal distribution with mean $\beta_{T|T}$ and variance $P_{T|T}$. Denote it $\widehat{B}_T$.
  - 1.3: At time $t = (T-1)$, use (C.8) to obtain $\beta_{T-1|T-1,\beta_T} = \beta_{T-1|T-1} + K^* \times (\widehat{B}_T - \mu + F\beta_{T-1|T-1})$. Note that $\beta_{T-1|T-1}$ is the Kalman filter estimate

of the state variable at time $(T-1)$, whereas $\widehat{B}_T$ is a draw from $N \sim (\beta_{T|T}, P_{T|T})$ at time $T$ (both from step 1.1). Use also Eq. (C.9) to calculate $P_{T-1|T-1,\beta_T} = P_{T-1|T-1} - K^* \times H^* \times P_{T-1|T-1}$, where $P_{T-1|T-1}$ is obtained from step 1.1 for $(T-1)$.

  - 1.4: Repeat above step for $t = T-2, T-3, \ldots 1$, to complete the backward recursions. This provides the first sample of $\beta_t$ from $t = 1...T$. Denote it $\beta^1_{1T}$.

- Step 2: Conditional on $\beta_t$ sample $R$ from its conditional posterior distributions given in Eq. (C.6). That is, use the draw of $\beta_t$, i.e. $\beta^1_{1T}$, to compute the elements necessary to sample from the inverse Gamma distribution. These are the scale matrix given by $scale = (y_t - \beta^1_{1T}H)'(y_t - \beta^1_{1T}H)$ and the posterior degrees of freedom defined as $T_1 = T_0 - k + T$. This provides one draw of $R$ from the inverse Gamma distribution with $\theta_1 = R_0 + scale$ as a scale parameter and $T_1$ degrees of freedom.

- Step 3: Conditional on $\beta_t$ sample $Q$ from its conditional posterior distribution given by Eq. (C.7). The draw obtained in step 1, $\beta^1_{1T}$, also allows to sample $Q$. To do so, compute the elements necessary to draw $Q$ from the inverse Wishart distribution. That is, compute the scale matrix $(\beta_t - \beta_{t-1})'(\beta_t - \beta_{t-1})$ and add the prior scale parameter $Q_0$ to obtain the posterior scale matrix $\overline{Q}$. Then, use $\overline{Q}$ and $T_1 = T + T_0$ to draw $Q$ from the inverse Wishart distribution.

- Step 4: Repeat steps 1-3 a sufficient number of times until convergence is detected. We use Geweke's convergence test and the relative numerical efficiency measure to assess the convergence of the algorithm, and find that 1700 draws are sufficient. We then discard the first 300 draws and save the last 1400 draws for inference. We then use the mean of the marginal posterior distribution of $\beta_t$, as the coefficient's point estimate.

# D Bayesian Dynamic Linear Models: Estimation and Forecasting

This Appendix provides details on the methods used to forecast with the dynamic linear model defined by equations (3.1) and (3.2) in Chapter 3, as well as the averaging approach. We draw from expositions in West and Harrison (1997), Dangl and Halling (2012), and Koop and Korobilis (2012).

## D.1 Bayesian Estimation

For convenience we begin by transcribing the system of equations from Section 3.2.1 in Chapter 3:

$$\Delta s_{t+h} = X_t'\theta_t + v_{t+h}, \quad v_{t+h} \sim N(0, V_t), \text{ (observation equation)}; \tag{D.1}$$

$$\theta_t = \theta_{t-1} + \varpi_t, \quad \varpi_t \sim N(0, W_t), \text{ (transition equation)}; \tag{D.2}$$

The essential components of the Bayesian approach we employ are the priors for $V_t$ and $\theta_t$, along with a method to estimate $W_t$; the joint or conditional posterior distribution of $V_t$ and $\theta_t$; and in the context of our predictive regression, the predictive density. Finally, we also require an updating scheme for the priors after observing the data.

The approach involves a full conjugate Bayesian analysis. The starting point is the natural conjugate $g-$prior specification set at $t = 0$:

$$V_0|D_0 \sim IG\left[\frac{1}{2}, \frac{1}{2}H_0\right], \tag{D.3}$$

$$\theta_0|D_0, V_0 \sim N[0, H_0(gX'X)^{-1}], \tag{D.4}$$

where

$$H_0 = \frac{1}{N-1}\Delta s'(I - X(X'X)^{-1}X')\Delta s, \tag{D.5}$$

and $D_0$ indicates the conditioning information at $t = 0$. In general, at any arbitrary subsequent period, $D_t = [\Delta s_t, \Delta s_{t-h}, ..., X_t, X_{t-h}, ..., Priors_{t=0}]$. That is, $D_t$ contains the exchange rate variations, the predictors, and the prior parameters. At this arbitrary period we can form a posterior belief about the unobservable coefficient $\theta_{t-1}|D_t$, and the variance of the observation equation error term (observational variance $V_t|D_t$). The use of a natural conjugate prior implies that the posterior distributions are from the same family as the priors. Specifically, the posteriors are also jointly normally-inverse gamma distributed:

$$V_t|D_t \sim IG\left[\frac{n_t}{2}, \frac{n_t H_t}{2}\right], \tag{D.6}$$

$$\theta_{t-1}|D_t, V_t \sim N(m_t, VC_t^*), \tag{D.7}$$

where $H_t$ is the mean of the estimate of the observational variance at time $t$, with $n_t$ as the corresponding number of degrees-of-freedom; $m_t$ is the estimate of coefficient vector $\theta_{t-1}$ conditional on $D_t$ and $V_t$; and $C_t^*$ corresponds to the conditional variance matrix of $\theta_{t-1}$, normalized by the observational variance. Integrating out the distribution given by (D.7) with respect to $V_t$, yields a multivariate $t-$distribution for the coefficients' posterior:

$$\theta_{t-1}|D_t \sim T_{n_t}(m_t, H_t C_t^*). \tag{D.8}$$

The form of the transition equation given by (D.2) suggests that, when updating the coefficients vector, the posterior distribution of $\theta_{t-1}|D_t$ represented by (D.8) does not necessarily become the prior for $\theta_t|D_t$. The equation indicates that the transition process is exposed to normally distributed random shocks, which widens the variance but maintains the mean:

$$\theta_t|D_t \sim T_{n_t}(m_t, H_t C_t^* + W_t). \tag{D.9}$$

The predictive density of the $h-$step-ahead change in the exchange rate, $\Delta s_{t+h}$, is obtained by integrating the conditional density of $\Delta s_{t+h}$ over the space spanned by $\theta$ and $V_t$. To derive it, let $\varphi(x;\mu,\sigma^2)$ be the density of a normal distribution evaluated at $x$, and $ig(V_t;a,b)$ be the density of an $IG[a,b]$ distributed variable evaluated at $V_t$. Then, the predictive density is:

$$\begin{aligned}
f(\Delta s_{t+h}|D_t) &= \int_0^\infty \left[\int_\theta \varphi(\Delta s_t; X_t'\theta, V_t)\varphi(\theta; m_t, V_t C_t^* + W_t)d\theta\right] \\
&\quad \times ig\left(\frac{n_t}{2}, \frac{n_t H_t}{2}\right) dV_t \\
&= \int_0^\infty \varphi(\Delta s_t; X_t' m_t, X_t'(V_t C_t^* + W_t)X_t + V_t) \\
&\quad \times ig\left(\frac{n_t}{2}, \frac{n_t H_t}{2}\right) dV_t \\
&= \mathbf{t}_{n_t}(\Delta s_{t+h}; \widehat{\Delta s}_{t+h}, Q_{t+h}), \qquad (\text{D.10})
\end{aligned}$$

where $\mathbf{t}(\Delta s_{t+h}; \widehat{\Delta s}_{t+h}, Q_{t+h})$ denotes the density of a $t-$distribution with $n_t$ degrees-of-freedom, mean $\widehat{\Delta s}_{t+h}$, variance $Q_{t+h}$, evaluated at $\Delta s_{t+h}$. The mean of the predictive distribution is computed as:

$$\widehat{\Delta s}_{t+h} = X_t' m_t, \tag{D.11}$$

and the total unconditional variance of the same distribution is given by:

$$Q_{t+h} = X_t' R_t X_t + H_t, \tag{D.12}$$

with

$$R_t = H_t C_t^* + W_t, \tag{D.13}$$

where $R_t$ is the unconditional variance of the coefficient vector $\theta_t$ at time $t$. The first term in Eq. (D.12) captures the variance arising from uncertainty in the estimation of the coefficient vector $\theta_t$. The last term $H_t$ denotes the estimate of the variance of the disturbance term of the observation equation.

After observing the exchange rate change at $t+h$, the priors on $\theta_t$ and $V_t$ are updated as described in equations (D.14)-(D.18) below. The first element is the prediction error:

$$\varepsilon_{t+h} = \Delta s_{t+h} - \widehat{\Delta s}_{t+h}, \text{ (prediction error)}, \tag{D.14}$$

which is useful in the estimate of the observational variance:[5]

$$H_{t+h} = \kappa H_t + (1-\kappa)\varepsilon_{t+h}^2, \text{ (estimator of observational variance).} \tag{D.15}$$

where $\kappa$ $(0 < \kappa < 1)$ is a (decay) factor that governs the responsiveness of the estimator to the most recent data. Setting $\kappa = 1$ implies that all the observations receive the same weight in the estimate and, in fact, the estimate of the observational variance remains constant. Smaller values of $\kappa$ induce more variability of the estimate and hence in the coefficients. We set $\kappa = 0.97$ following the study of Koop and Korobilis (2012).

An additional element that induces changes in the coefficients is the adaptive vector:

$$A_{t+h} = \frac{R_t X_t}{Q_{t+h}}, \text{ (adaptive vector).} \tag{D.16}$$

It characterizes the degree to which the posterior of the coefficient vector $\theta_t$ changes to new observation. The numerator of Eq. (D.16) conveys the information content of the current observation, and the denominator measures the precision of the estimated coefficients. With the above elements, we can update the coefficients' point estimate $m_t$ and the covariance matrix $C_t^*$:

$$m_{t+h} = m_t + A_{t+h}\varepsilon_{t+h}, \text{ (expected coeff. vector estimator),} \tag{D.17}$$

$$C_{t+h}^* = \frac{1}{H_t}\left(R_t - A_{t+h}A'_{t+h}Q_{t+h}\right), \text{ (variance of the coeff. vector estimator).} \tag{D.18}$$

The exposition so far does not include a method to estimate $W_t$. As we noted in Section 3.2.1, to capture the relationship between the coefficients' estimation error and the variance, we let $W_t$ be proportional to the estimation variance $H_t C_t^*$ of the coefficients $\theta_t | D_t$ at time $t$. That is:

$$W_t = \frac{1-\delta}{\delta} H_t C_t^*, \quad \delta \in \{\delta_1, \delta_2, ..., \delta_d\},\ 0 < \delta_j \leq 1. \tag{D.19}$$

Therefore, the variance of the predicted coefficient vector expressed in Eq. (D.13) simplifies to:

$$R_t = H_t C_t^* + \frac{1-\delta}{\delta} H_t C_t^* = \frac{1}{\delta} H_t C_t^*. \tag{D.20}$$

This completes the requisites for forecasting with one model. The approach we pursue, however, allows for $k$ candidate predictors and $d$ possible support points for time-variation in coefficients and therefore $k.d$ models. We deal with these possibilities in a Bayesian model selection and averaging approach that we outline next.

[5]This estimator is known as the Exponentially Weighted Moving Average (EWMA), frequently used to model stochastic volatility in financial aplications.

## D.2 Dynamic Model Averaging

Let $M_i$ constitute a specific selection of a predictor from a set of $k$ candidates, and $\delta_j$ a specific choice of degree of time-variation in coefficients from the space $\{\delta_1, \delta_2, ..., \delta_d\}$. The mean of the predictive distribution computed above, see Eq. (D.11), is influenced by these specific choices. Hence, the point estimate of $\Delta s_{t+h}$ also becomes conditional on $M_i$ and $\delta_j$:

$$\widehat{\Delta s}^{j}_{t+h,i} = E(\Delta s_{t+h}|M_i, \delta_j, D_t) = X_t' m_t | M_i, \delta_j, D_t. \tag{D.21}$$

The starting point in examining which model setting turns out to be important empirically, is to assign prior weights to each individual predictor $M_i$ and each support point $\delta_j$. We begin with a prior that allows each predictor and each support point to have the same chance of becoming probable. That is, for each $M_i$ and $\delta_j$ we set uninformative priors:

$$P(M_i|\delta_j, D_0) = 1/k, \tag{D.22}$$

$$P(\delta_j|D_0) = 1/d. \tag{D.23}$$

At time $t$, the posterior probabilities are updated using Bayes's rule. We first update the posterior probability of a certain model, given a value of $\delta_j$:

$$P(M_i|\delta_j, D_t) = \frac{f(\Delta s_t|M_i, \delta_j, D_{t-h})P(M_i|\delta_j, D_{t-h})}{f(\Delta s_t|\delta_j, D_{t-h})}, \tag{D.24}$$

where

$$f(\Delta s_t|\delta_j, D_{t-h}) = \sum_M f(\Delta s_t|M_i, \delta_j, D_{t-h})P(M_i|\delta_j, D_{t-h}). \tag{D.25}$$

The key ingredient is the conditional density:

$$f(\Delta s_t|M_i, \delta_j, D_{t-h}) \sim \frac{1}{\sqrt{Q^j_{t,i}}} \mathbf{t}_{n_{t-1}} \left( \frac{\Delta s_t - \Delta s^j_{t,i}}{\sqrt{Q^j_{t,i}}} \right), \tag{D.26}$$

where $\mathbf{t}_{n_{t-1}}$ is the density of a Student$-t-$distribution and $\Delta s^j_{t,i}$ and $Q^j_{t,i}$ are the corresponding point estimates and variance of the predictive distribution of model $M_i$, given $\delta = \delta_j$ - refer to Eq. (D.10). The prediction of the average model for each of the specific value of $\delta = \delta_j$ is given by:

$$\widehat{\Delta s}^{j}_{t+h} = \sum_{i=1}^{k} P(M_i|\delta_j, D_t)\widehat{\Delta s}^{j}_{t+h,i}. \tag{D.27}$$

Essentially, for each specific $\delta$, it is the sum of the forecasts of each of the $k$ models weighted by their posterior probability. If there was only one support point for time-variation in coefficients, such that $d = 1$, then Eq. (D.27) would complete the averaging approach. However, since we are considering several possibilities for $\delta$, we also perform

Bayesian averaging over these values.

Starting with the prior probability in Eq. (D.23), the posterior probability of a specific $\delta$ is:

$$P(\delta_j|D_t) = \frac{f(\Delta s_t|\delta_j, D_{t-h})P(\delta_j|D_{t-h})}{\sum_\delta f(\Delta s_t|\delta, D_{t-h})P(\delta|D_{t-h})}. \tag{D.28}$$

We note that using this probability, we can infer the degree of time-variation in coefficients supported by the data; see also Eq. (3.12) in Chapter 3.

We can now find the total posterior probability of a model determined by a specific selection of predictor $M_i$ and degree of coefficient variation $\delta_j$,

$$P(M_i, \delta_j|D_t) = P(M_i|\delta_j, D_t)P(\delta_j|D_t), \tag{D.29}$$

and the unconditional average prediction of the average model,

$$\widehat{\Delta s}_{t+h} = \sum_{j=1}^{d} P(\delta_j|D_t)\widehat{\Delta s}^{j}_{t+h}. \tag{D.30}$$

Thus, $\widehat{\Delta s}_{t+h}$ is obtained by averaging over the average models' prediction, over degrees of time-variation in coefficients.

# E MIDAS Model with the Exponential Almon Polynomial: Bayesian Estimation and Forecasting

## E.1 MIDAS model

In this Appendix we provide further details of the Bayesian approach we pursue to estimate and forecast with our MIDAS models in Chapter 5. Our ideas are inspired by Koop (2003 Ch. 5 & 11).

We begin by transcribing the MIDAS model we consider in Chapter 5:

$$\Delta s_{t+h} = \beta_0 + \beta_1 B(L^{1/m};\theta_1)x_t^{(m)} + \varepsilon_{t+h}, \qquad \varepsilon_{t+h} \sim N(0,\sigma^2); \tag{E.1}$$

where we use the exponential Almon polynomial to characterize the weight of each high frequency (daily) observation. This polynomial has the following form:

$$B(k;\theta) = \frac{e^{(\theta_1 k + \theta_2 k^2)}}{\sum_{i=1}^{K} e^{\theta_1 i + \theta_2 i^2}}. \tag{E.2}$$

If we consider, for example, that only the past 21 trading days affect the value of $\Delta s_{t+h}$, then Eq. (E.1) is a compact representation of:

$$\Delta s_{t+h} = \beta_0 + \beta_1 \left( \frac{e^{(\theta_1 \times 1 + \theta_2 \times 1^2)}}{\sum_{i=1}^{K} e^{\theta_1 i + \theta_2 i^2}} x_{td21} + \frac{e^{(\theta_1 \times 2 + \theta_2 \times 2^2)}}{\sum_{i=1}^{K} e^{\theta_1 i + \theta_2 i^2}} x_{td20} + ... + \frac{e^{(\theta_1 \times 21 + \theta_2 \times 21^2)}}{\sum_{i=1}^{K} e^{\theta_1 i + \theta_2 i^2}} x_{td1} \right) + \varepsilon_{t+h}. \tag{E.3}$$

where $\beta_0$ is the coefficient associated with the constant, $\beta_1$ captures the overall impact of all past values of daily observations on $\Delta s_{t+h}$, and $\theta_1$ and $\theta_2$ are the polynomials' parameters. Eq. (E.3) is a non-linear regression equation in the following unknown parameters to be estimated: $\beta_0, \beta_1, \theta_1, \theta_2, \sigma^2$.

## E.2 Estimation

To estimate the model we use the random walk chain Metropolis–Hastings within Gibbs algorithm. To simplify notation we express Eq. (E.3) in the following functional form:

$$S = f(X, \gamma) + \varepsilon, \qquad \varepsilon \sim N(0, \frac{1}{\eta}), \text{ and } \qquad \frac{1}{\eta} = \sigma^2. \tag{E.4}$$

where $f(.)$ indicates that our function of interest depends on the data $X$ (containing $x_{td}$) and the parameters $\gamma$, which include $\beta_0, \beta_1, \theta_1, \theta_2$.

In a Bayesian setup, estimation involves definition of prior distributions, the likelihood function, and the posterior distributions. We use independent Normal-Gamma priors. Thus, the prior for $\gamma$ is independent of the prior for $\eta$ and is defined as:

$$\gamma \sim N(\underline{\gamma}, \underline{V}). \tag{E.5}$$

We set $\underline{\gamma} = (0, 0, 0, 0)'$ and $\underline{V} = 0.35I$. For $\eta$ the prior is:

$$\eta \sim G(\underline{s}^{-2}, \underline{\nu}), \tag{E.6}$$

where $\underline{\nu} = 1$ and $\underline{s}^{-2}$ is based on OLS estimate under equal weighting assumption.

Using the definition of the multivariate Normal density, the likelihood function has the following form (see Koop 2003, Ch. 5):

$$p(S|\gamma,\eta) = \frac{\eta^{\frac{T}{2}}}{(2\pi)^{\frac{T}{2}}} \left\{ \exp\left[ -\frac{\eta}{2} \left\{ S - f(X,\gamma) \right\}' \left\{ S - f(X,\gamma) \right\} \right] \right\}. \tag{E.7}$$

Combining the prior with this likelihood yields the following conditional posterior for $\eta$:

$$p(\eta|S,\gamma) \sim G(\overline{s}^{-2}, \overline{\nu}), \tag{E.8}$$

$$\overline{s}^2 = \frac{[S - f(X,\gamma)]'[S - f(X,\gamma)] + \underline{\nu}\underline{s}^2}{\overline{\nu}}, \tag{E.9}$$

$$\overline{\nu} = \underline{\nu} + T; \tag{E.10}$$

while the conditional posterior distribution of $\gamma$ is:

$$p(\gamma|S,\eta) \propto \exp\left[-\frac{\eta}{2}\left\{S - f(X,\gamma)\right\}'\left\{S - f(X,\gamma)\right\}\right] \tag{E.11}$$
$$\exp\left[(-\frac{1}{2}(\gamma-\underline{\gamma})'\underline{V}^{-1}(\gamma-\underline{\gamma}))\right].$$

The form of this conditional density $p(\gamma|S,\eta)$ does not suggest any density from which to draw upon. Therefore, we employ the random walk chain Metropolis–Hastings (RW-MH) within Gibbs algorithm to sequentially draw $\eta$ conditional on $\gamma$. The RW-MH algorithm consists of the following steps:

1. Choose starting values for $\eta$ and $\gamma^{(0)}$:

   We use data available up to the beginning of our first forecast to fix these values. Precisely, we set $\eta = 1/s^2$, where $s^2$ is based on OLS estimates assuming that $\theta_1 = \theta_2 = 0$. Using the same data, we maximize the likelihood function in (E.7) and set $\gamma^{(0)} = \widehat{\gamma}_{ML}$; i.e., to maximum likelihood estimates under $\theta_1 = \theta_2 = 0$;

2. Draw $\eta$ from its conditional posterior as given by Eq. (E.8);

3. Conditional on $\eta$ take a candidate draw, $\gamma^{(*)}$, from the candidate candidate generating density:

   $\gamma^{(*)} \sim N(\gamma^{(0)}, \Sigma)$,

   where $\Sigma = var(\widehat{\gamma}_{ML})$ is the covariance matrix of the maximum likelihood estimator obtained in step 1;

4. Calculate acceptance probability:

   $\alpha(\gamma^{(dr-1)}, \gamma^{(*)}) = \min\left[\frac{p(\gamma=\gamma^{(*)}|\mathbf{S})}{p(\gamma=\gamma^{(dr-1)}|\mathbf{S})}, 1\right]$,

   where $p()$ is evaluated at the current, $\gamma^{(*)}$, and previous, $\gamma^{(dr-1)}$, draw using Eq. (E.11);

5. Set $\gamma^{(dr)} = \gamma^{(*)}$ with probability $\alpha(\gamma^{(dr-1)}, \gamma^{(*)})$ else $\gamma^{(dr)} = \gamma^{(dr-1)}$;

6. Repeat above steps many times (e.g. 31000) and after discarding the first draws (e.g. 1000), thin the chains by keeping every third draw. Point estimates of the parameters are obtained as average of the retained draws.

## E.3 Prediction

Prediction is based on the predictive density:

$p(S^*|S,\gamma) = t(S^*|f(X^*,\gamma), \overline{s}^2 I_T, T)$; where, $\overline{s}^2 = (S-f)'(S-f)/T$.

# F Data for the Application in Chapter 2

This Appendix describes the data used in the empirical estimation in Chapter 2. The sample period is 1973Q1:2013Q1, for 18 OECD countries - Table F.1.

Table F.1: Data for the Empirical Aplication in Chapter 2

| Country | Nominal exchange rate | Industrial production index, NSA, 2005=100 | Money supply, NSA |
|---|---|---|---|
| Australia | IFS, 193..AE.ZF | IFS, 19366..CZF | M3, OECD, MEI |
| Canada | IFS, 156..AE.ZF | IFS, 15666..CZF | M1, OECD, MEI |
| Denmark | IFS, 128..AE.ZF | IFS, 12866..BZF | M1, OECD, MEI |
| UK | IFS, 112..AE.ZF | IFS, 11266..CZF | M4, Bank of England |
| Japan | IFS, 158..AE.ZF | IFS, 15866..CZF | M1, OECD, MEI |
| Korea | IFS, 542..AE.ZF | IFS, 54266..CZF | M1, OECD, MEI |
| Norway | IFS, 142..AE.ZF | IFS, 14266..CZF | M2, OECD, MEI |
| Sweden | IFS, 144..AE.ZF | OECD MEI | M3, OECD, MEI |
| Switzerland | IFS, 146..AE.ZF | IFS, 14666..BZF | M1, OECD, MEI |
| Austria+ | IFS, 122..AE.ZF | IFS, 12266..BZF | M2=34A.NZF + 34B.NZF+35; IFS |
| Belgium+ | IFS, 124..AE.ZF | IFS, 12466..CZF | M1=34A.NZF + 34B.NZF; IFS |
| France+ | IFS, 132..AE.ZF | 13266..CZF | M1=34A.NZF + 34B.NZF; IFS |
| Germany+ | IFS, 134..AE.ZF | IFS, 13466..CZF | M1=34A.NZF + 34B.NZF; IFS |
| Spain+ | IFS, 184..AE.ZF | IFS, 18466..CZF | M2=34A.NZF + 34B.NZF+35;IFS |
| Italy+ | IFS, 136..AE.ZF | IFS, 13666..CZF | M2=34A.NZF + 34B.NZF+35; IFS |
| Finland+ | IFS, 172..AE.ZF | IFS, 17266..CZF | M2=34A.NZF + 34B.NZF+35; IFS |
| The Netherlands+ | IFS, 138..AE.ZF | IFS, 13866..CZF | M2=34A.NZF + 34B.NZF+35; IFS |
| United States | | IFS, 11166..CZF | M2, OECD, MEI |

| Country | Short-term nominal interest rate (annual) | Consumer price index NSA, 2005=100 | Unemployment rate, NSA (last month of quarter) |
|---|---|---|---|
| Australia | IFS, 19360...ZF | IFS, 64...ZF | OECD, MEI |
| Canada | IFS, 15660B..ZF | IFS, 15664...ZF | OECD, MEI |
| Denmark | IFS, 12860...ZF | IFS, 12864...ZF | OECD, MEI |
| UK | IFS, 11260...ZF | IFS, 11264B..ZF | OECD, MEI |
| Japan | IFS, 15860B..ZF | IFS, 15864...ZF | OECD, MEI |
| Korea | 54260B..ZF | IFS, 54264...ZF | n.a |
| Norway | IFS, 14260...ZF | IFS, 14264...ZF | OECD, MEI |
| Sweden | IFS, 14460B..ZF | IFS, 14464...ZF | OECD, MEI |
| Switzerland | IFS, 14660...ZF | IFS, 14664...ZF | OECD, MEI |
| Austria+ | IFS, 12260B..ZF | IFS, 12264...ZF | OECD, MEI |
| Belgium+ | IFS, 12460B..ZF | IFS, 12464...ZF | OECD, MEI |
| France+ | IFS, 13260B..ZF | IFS, 13264...ZF | OECD, MEI |
| Germany+ | IFS, 13460B..ZF | Bundesbank | OECD, MEI |
| Spain+ | IFS, 18460B..ZF | IFS, 18464...ZF | n.a |
| Italy+ | IFS, 13660B..ZF | IFS, 13664...ZF | n.a |
| Finland+ | Central bank rate, MEI | IFS, 17263EY.ZF | n.a |
| The Netherlands+ | IFS, 13860B..ZF | IFS, 13864...ZF | n.a |
| United States | IFS, 11160B..ZF | IFS, 11164...ZF | OECD, MEI |

Notes: The "+" symbol indicates an Euro Area country. The exchange rate is defined as national currency per U.S. dollar at the end of the quarter. To generate the exchange rate series for the eight Euro Area countries from 1999 onwards, the irrevocable conversion factors adopted by each country on the 1st of January 1999 were employed. These conversion factors are from the IMF International Financial Statistics (IFS) database. For example, the Mark/U.S. dollar exchange rate is obtained by multiplying the conversion factor 1.95583/EUR by the EUR/U.S. dollar exchange rate at each post 1998Q4 date. The conversion factors for the other countries are: Austria - 13.760, Belgium - 40.340, Finland - 5.94573, France - 6.560, Italy - 1936.270, The Netherlands - 2.204, and Spain - 166.386. OECD, MEI denotes the OECD's Main Economic Indicators database. NSA stands for non-seasonally adjusted and "n.a" indicates that the series is not available for the entire sample period.

# G Data for the Application in Chapter 3

This Appendix describes the data used in Chapter 3. The sample period is from 1979M1 to 2013M5, for eight countries. Data on the Eurodeposit rates were obtained from Datastream. Table G.1 describes the exchange rate data, as well as the data used to compute the various sets of fundamentals. For each country in the first column, the Table indicates the source of information for each variable in the subsequent columns.

Table G.1: Data for the Application in Chapter 3

| Country | Nominal exchange rate (USD/National currency) | Industrial prod. index, NSA, 2005=100 | Money supply, NSA, National currency (10ˆ9) |
|---|---|---|---|
| Canada | IFS, 156..AE.ZF | IFS, 15666..CZF | M1, OECD, MEI |
| Germany/Eur | IFS, 134..AE.ZF | IFS, 13466..CZF | M1; Bundesbank |
| Japan | IFS, 158..AE.ZF | IFS, 15866..CZF | M1, OECD, MEI |
| Norway | IFS, 142..AE.ZF | IFS, 14266..CZF | M2, OECD, MEI |
| Sweden | IFS, 144..AE.ZF | OECD MEI | M3, OECD, MEI |
| Switzerland | IFS, 146..AE.ZF | IFS, 14666..BZF | M1, OECD, MEI |
| UK | IFS, 112..AE.ZF | IFS, 11266..CZF | M4, Bank of England |
| US | | IFS, 11166..CZF | M1, FED |

| | Short-term nominal interest rate (%) | Consumer price index NSA, 2005=100 |
|---|---|---|
| Canada | IFS, 15660B..ZF | IFS, 15664...ZF |
| Germany/Eur | IFS, 13460B..ZF | Bundesbank |
| Japan | IFS, 15860B..ZF | IFS, 15864...ZF |
| Norway | IFS, 14260...ZF | IFS, 14264...ZF |
| Sweden | IFS, 14460B..ZF | IFS, 14464...ZF |
| Switzerland | IFS, 14660...ZF | IFS, 14664...ZF |
| UK | IFS, 11260...ZF | IFS, 11264B..ZF |
| US | IFS, 11160B..ZF | IFS, 11164...ZF |

**Notes:** The exchange rate is defined as the end-of-month value of the U.S. dollar (USD) price of a unit of national currency. IFS denotes International Financial Statistics as published by the IMF. OECD, MEI denotes the OECD's Main Economic Indicators database. NSA stands for non-seasonally adjusted

# H Data for the Application in Chapter 4

This Appendix describes the data used in our empirical exercise in Chapter 4. The sample period is 1989M1:2013M5, for six countries. Table H.1 indicates the sources for exchange rate data and all variables used to compute the various sets of fundamentals.

Table H.1: Data for the Application in Chapter 4

| Country | Nominal exchange rate (USD/National currency) | Industrial prod. index, NSA, 2005=100 | Money supply, NSA, National currency (10ˆ9) |
|---|---|---|---|
| Canada | IFS, 156..AE.ZF | IFS, 15666..CZF | M1, OECD, MEI |
| Germany/Eur | IFS, 134..AE.ZF | IFS, 13466..CZF | M1; Bundesbank |
| UK | IFS, 112..AE.ZF | IFS, 11266..CZF | M4, Bank of England |
| Japan | IFS, 158..AE.ZF | IFS, 15866..CZF | M1, OECD, MEI |
| Sweden | IFS, 144..AE.ZF | OECD MEI | M3, OECD, MEI |
| | **Central bank interest rate (%)** | **Consumer price index NSA, 2005=100** | **Bond yields (maturities in months)** |
| Canada | IFS, 15660B..ZF | IFS, 15664...ZF | Bank of Canada (3,6,12,24, 36,48, 60, 72, 84, 96, 108, 120, 132, 144, 156, 168, 180, 192, 204, 216, 228, 240, 252, 264, 276, 288, 300) |
| Germany/Eur | IFS, 13460B..ZF | Bundesbank | Bundesbank (3, 6, 12, 60, 96, 120, 180, 360) |
| UK | IFS, 11260...ZF | IFS, 11264B..ZF | Bank of England (12, 24, 36, 48, 60, 72, 84, 96, 108, 120, 132) |
| Japan | IFS, 15860B..ZF | IFS, 15864...ZF | Japanese Ministry of Finance (12, 24, 36, 48, 60, 72, 84, 96, 108, 120) |
| Sweden | IFS, 14460B..ZF | IFS, 14464...ZF | - |
| US | IFS, 11160B..ZF | IFS, 11164...ZF | US Federal Reserve (3, 6, 12, 24, 36, 48, 60, 72, 84, 96, 108, 120, 132, 144,156, 168, 180, 192, 204, 216, 228, 240, 252, 264, 276, 288, 300, 312, 324,336, 348, 360) |

**Notes:** The exchange rate is defined as the end-of-month value of the U.S. dollar (USD) price of a unit of national currency. IFS denotes International Financial Statistics as published by the IMF. OECD, MEI denotes the OECD's Main Economic Indicators database. NSA stands for non-seasonally adjusted.

# I Data for the Application in Chapter 5

This Appendix describes the sources (Table I.1) and descriptive statistics (Tables I.3) and I.2) for the data used in Chapter 5. In Table I.1, for each country in the first column we indicate the source of information for each variable in the subsequent columns. The sample period is 1986M9:2014M3 for Australia, Canada, Japan and the US. Due to unavailability of data on daily gas prices fluctuations, the sample period for Norway comprises 1997M1:2014M3.

Data on commodity prices were obtained from Datastream. The oil price is the Crude Oil-WTI Spot Cushing, USD/BBL (Mnemoic: S71926). Gold price corresponds to the Gold Bullion London Bullion Market price, USD/Troy Ounce (S20665). Copper price is the London Metal Exchange Copper Grade A Cash price, USD/Metric Tonne (S76871). Gas price corresponds to the Henry Hub Natural Gas Spot Price USD/MMBTU (S214W9). The commodity price index is as compiled by the Commodity Research Bureau under BLS Spot Index. The Index measures price movements of 22 commodities (see the Commodity Research Bureau webpage for more details).

Table I.1: Data for the Application in Chapter 5

| Country | Nominal exchange rate (national currency/USD) | Short-term nominal interest rate (%) | Consumer price index SA (2010=100) |
|---|---|---|---|
| Australia | IFS,193..AE-ZF | IFS,19360...ZF | OECD MEI[a] |
| Canada | IFS,156..AE-ZF | IFS,15660B..ZF | OECD MEI |
| Norway | IFS,142..AE-ZF | IFS,14260...ZF | OECD MEI |
| Japan | IFS,158..AE-ZF | IFS,15860B..ZF | OECD MEI |
| US | IFS,111..AE-ZF | FRED | OECD MEI |
| | Consumer price index SA (2010=100) | Money supply, SA, (national currency, 10ˆ6) | |
| Australia | IFS, 19366..CZF[a] | OECD MEI, M1 | |
| Canada | IFS, 15666..CZF | OECD MEI, M1 | |
| Norway | IFS,14266..CZF | Norges Bank, M2 | |
| Japan | IFS, 15866..CZF | OECD MEI, M1 | |
| US | IFS,11166..CZF | OECD MEI, M1 | |

**Notes:** The exchange rate is the end-of-month value of the national currency per U.S. dollar. IFS denotes International Financial Statistics as published by the IMF. OECD MEI denotes the OECD's Main Economic Indicators database. FRED indicates Federal Reserve Economic Data database. SA stands for non-seasonally adjusted and the superscript ([a]) denotes monthly data obtained via quadratic-match-average interpolation method from quartely data.

Table I.2: Commodity Prices Data - Descriptive Statistics and Pairwise Correlations

| | Daily data | | | | |
|---|---|---|---|---|---|
| | $\Delta$oil_Dp | $\Delta$gold_Dp | $\Delta$copper_Dp | $\Delta$gas_Dp | $\Delta$DP_index |
| Mean | 0.0003 | 0.0002 | 0.0002 | 0.0002 | 0.0001 |
| Std | 0.0239 | 0.0098 | 0.0175 | 0.0360 | 0.0040 |
| Skew | -0.8619 | -0.4049 | 0.4683 | 0.6452 | -0.5036 |
| Kurt | 20.5525 | 11.0925 | 17.5955 | 17.6805 | 9.5617 |
| | Pairwise Correlation | | | | |
| $\Delta$oil_Dp | 1.000 | | | | |
| $\Delta$gold_Dp | 0.155 | 1.000 | | | |
| $\Delta$copper_Dp | 0.160 | 0.209 | 1.000 | | |
| $\Delta$gas_Dp | 0.053 | 0.044 | 0.024 | 1.000 | |
| $\Delta$DP_index | 0.176 | 0.177 | 0.318 | 0.053 | 1.000 |

| | Monthly data | | | | |
|---|---|---|---|---|---|
| | $\Delta$oil_Mp | $\Delta$gold_Mp | $\Delta$copper_Mp | $\Delta$gas_Mp | $\Delta$MP_index |
| Mean | 0.0059 | 0.0034 | 0.0049 | 0.0045 | 0.0025 |
| Std | 0.0906 | 0.0442 | 0.0812 | 0.1753 | 0.0274 |
| Skew | -0.1251 | -0.0625 | -0.4110 | -0.0812 | -1.8304 |
| Kurt | 4.8736 | 4.2976 | 7.6427 | 5.6248 | 18.0528 |
| | Pairwise Correlation | | | | |
| $\Delta$oil_Mp | 1.000 | | | | |
| $\Delta$gold_Mp | 0.236 | 1.000 | | | |
| $\Delta$copper_Mp | 0.265 | 0.301 | 1.000 | | |
| $\Delta$gas_Mp | 0.230 | 0.079 | 0.009 | 1.000 | |
| $\Delta$MP_index | 0.337 | 0.282 | 0.494 | 0.015 | 1.000 |
| | | | h=1 | | |
| $\Delta$s (AUD) | -0.327 | -0.380 | -0.469 | -0.131 | -0.496 |
| $\Delta$s (CAD) | -0.329 | -0.326 | -0.375 | -0.099 | -0.420 |
| $\Delta$s (NOK) | -0.332 | -0.335 | -0.322 | -0.130 | -0.401 |
| $\Delta$s (JPY) | -0.220 | -0.314 | -0.267 | -0.128 | -0.289 |
| | | | h=3 | | |
| $\Delta$s (AUD) | -0.317 | -0.171 | -0.337 | -0.068 | -0.453 |
| $\Delta$s (CAD) | -0.268 | -0.162 | -0.289 | -0.075 | -0.337 |
| $\Delta$s (NOK) | -0.262 | -0.163 | -0.269 | -0.068 | -0.328 |
| $\Delta$s (JPY) | -0.150 | -0.181 | -0.213 | -0.037 | -0.215 |

**Notes:** The table shows the mean, standard deviation, skewness and kurtosis for commodity prices data (at daily and monthly frequency). It also shows the pairwise correlations among commodity prices, as well as between exchange rates variations ($\Delta$s ) at h-month(s)-horizon and monthly commodity price changes. The suffixes Dp and Mp attached to the commodity prices denote daily and monthly prices respectively. The currency codes are AUD - for the Australian dollar, CAD - Canadian dollar, NOK - Norwegian Kroner, and YEN for the Japanese currency.

Table I.3: Exchange Rates and Economic Fundamentals Data - Descriptive Statistics

| | | Δs | TRsy | TRasy | MM | PPP | UIP |
|---|---|---|---|---|---|---|---|
| AUD | Mean | -0,001 | 0,001 | 0,033 | -2,629 | -0,319 | 0,031 |
| | Std | 0,032 | 0,035 | 0,037 | 0,498 | 0,192 | 0,025 |
| | Skew | 0,766 | -1,373 | -0,485 | -0,197 | 0,036 | 0,917 |
| | Kurt | 5,718 | 7,092 | 4,385 | 1,974 | 2,731 | 3,983 |
| CAD | Mean | -0,001 | -0,001 | 0,015 | -1,713 | -0,165 | 0,008 |
| | Std | 0,022 | 0,037 | 0,039 | 0,335 | 0,133 | 0,015 |
| | Skew | 0,688 | 4,002 | 3,364 | 0,297 | -0,350 | 0,568 |
| | Kurt | 9,359 | 20,236 | 16,326 | 1,761 | 1,919 | 3,638 |
| NOK | Mean | -0,001 | -0,005 | 0,185 | -2,549 | -1,870 | 0,029 |
| | Std | 0,031 | 0,031 | 0,031 | 0,447 | 0,119 | 0,024 |
| | Skew | 0,471 | 1,467 | 0,999 | 0,504 | -0,872 | 0,848 |
| | Kurt | 4,277 | 6,579 | 4,583 | 1,780 | 3,447 | 3,431 |
| YEN | Mean | -0,001 | -0,007 | 0,442 | 0,586 | -4,478 | -0,023 |
| | Std | 0,031 | 0,153 | 0,155 | 0,548 | 0,137 | 0,023 |
| | Skew | -0,231 | 5,075 | 5,032 | -0,340 | 0,390 | 0,012 |
| | Kurt | 4,556 | 27,197 | 26,950 | 1,513 | 3,010 | 1,704 |

**Notes:** Descriptive statistics for monthly economic fundamentals and monthly changes in the log exchange rate (Δs). Monthly fundamentals include those from the symmetric Taylor rule - TRsy, the asymmetric Taylor rule - TRasy, the Monetary Model - MM, Purchasing Power Parity - PPP, and Uncovered Interest Rate Parity - UIP. The currency codes in the first column denote the Australian dollar (AUD), the Canadian dollar (CAD), the Norwegian Kroner (NOK), and the Japanese YEN.

Table I.4: Commodity Average Exports as a Percentage of Total Merchandise Exports

| Australia (1988-2014) | | Canada (1997-2014) | | Norway (1997-2014) | |
|---|---|---|---|---|---|
| Gold | Copper | Oil | Cooper | Oil | Gas |
| 6 | 17 | 6 | 1 | 39 | 16 |

**Notes:** Commodity Export Share in the Country's Total Merchandise Export. Compiled by the authors based on countries' official statistics as published by (i) the Australian Bureau of Statistics (Table on International Trade in Goods and Services, Australia, May 2015); (ii) Statistics Canada (Table on Merchandise imports and exports between Canada and World, by Harmonized System section, customs basis, May 2015); and (iii) Statistics Norway (STATBANK, Table on External Trade in Goods).

# J Chapter 5 Convergence Diagnostics

Our forecasting models in Chapter 5 are estimated using algorithms pertaining to Markov Chain Monte Carlo Methods (MCMC), which rely on drawing samples from candidate generating densities. We recall that we generated 31000 draws from which we discarded the first 1000 and used every third draw in the estimates. In this Appendix we evaluate the convergence of our algorithms.

In Table J.1 we report the average acceptance probability in the random walk Metropolis-Hastings (RW-MH) component of the algorithm. Values in the region 0.2 to 0.5 are regarded as satisfactory. The averages are computed over estimates across all data points in the recursive estimations.

In Table J.2 we look at the convergence of the overall RW-MH within Gibbs algorithm. In particular, we focus on Geweke's (1992) measure of numerical standard error

(NSE). Smaller values of NSE relative to the posterior standard deviations convey an acceptable degree of approximation error. The NSE is based on 4% tapered window in the estimate of the spectral density at zero frequency. To manage space, we also average the estimates obtained over all data points in our recursive estimations. Overall, results indicate an acceptable degree of efficiency of the algorithm.

Table J.1: Mean Acceptance Rates in the RW-MH Algorithm

| | AUD | CAD | NOK | YEN | AUD | CAD | NOK | YEN |
|---|---|---|---|---|---|---|---|---|
| | h=1 | | | | h=3 | | | |
| ΔOil_Dp | - | 0.2 | 0.3 | 0.3 | - | 0.3 | 0.2 | 0.3 |
| ΔGold_Dp | 0.3 | - | - | - | 0.2 | - | - | - |
| ΔCopper_Dp | 0.3 | 0.3 | - | - | 0.3 | 0.3 | - | - |
| ΔGas_Dp | - | - | 0.3 | - | - | - | 0.2 | - |
| ΔDP_index | - | - | - | 0.1 | - | - | - | 0.1 |

**Notes:** The Table reports the average acceptance probability in the random walk Metropolis-Hastings algorithm. The averages are computed over estimates across all data points in the recursive estimations. Values in the region 0.2 to 0.5 are regarded as satisfactory.

Table J.2: Convergence Diagnostics for the RW-MH within Gibbs Algorithm

| | h=1 | | | h=3 | | |
|---|---|---|---|---|---|---|
| | postMean | postStdv | NSE | postMean | postStdv | NSE |
| Australia (AUD) | | | MIDAS with $\Delta$Gold_Dp | | | |
| $\beta_0$ | 0.000 | 0.002 | 0.000 | 0.001 | 0.003 | 0.000 |
| $\beta_1$ | -0.907 | 0.348 | 0.016 | -0.621 | 0.357 | 0.009 |
| $\theta_1$ | 0.210 | 0.474 | 0.036 | 0.016 | 0.500 | 0.027 |
| $\theta_2$ | 0.034 | 0.033 | 0.005 | 0.360 | 0.048 | 0.008 |
| $\eta$ | 0.001 | 0.000 | 0.000 | 0.003 | 0.000 | 0.000 |
| | | | MIDAS with $\Delta$Copper _Dp | | | |
| $\beta_0$ | 0.000 | 0.002 | 0.000 | 0.001 | 0.003 | 0.000 |
| $\beta_1$ | -1.430 | 0.339 | 0.019 | -0.593 | 0.286 | 0.012 |
| $\theta_1$ | 0.168 | 0.400 | 0.024 | 0.037 | 0.489 | 0.034 |
| $\theta_2$ | 0.039 | 0.019 | 0.002 | -0.048 | 0.041 | 0.007 |
| $\eta$ | 0.001 | 0.000 | 0.000 | 0.003 | 0.000 | 0.000 |
| Canada (CAD) | | | MIDAS with $\Delta$Oil_Dp | | | |
| $\beta_0$ | 0.000 | 0.001 | 0.000 | 0.000 | 0.002 | 0.000 |
| $\beta_1$ | -0.368 | 0.162 | 0.012 | -0.319 | 0.226 | 0.019 |
| $\theta_1$ | 0.207 | 0.409 | 0.012 | 0.092 | 0.491 | 0.018 |
| $\theta_2$ | 0.060 | 0.061 | 0.011 | -0.153 | 0.115 | 0.022 |
| $\eta$ | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 |
| | | | MIDAS with $\Delta$Copper _Dp | | | |
| $\beta_0$ | 0.000 | 0.001 | 0.000 | 0.000 | 0.002 | 0.000 |
| $\beta_1$ | -0.745 | 0.250 | 0.017 | -0.320 | 0.175 | 0.007 |
| $\theta_1$ | 0.150 | 0.381 | 0.012 | -0.046 | 0.485 | 0.017 |
| $\theta_2$ | 0.022 | 0.039 | 0.006 | -0.300 | 0.078 | 0.015 |
| $\eta$ | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 |
| Norway (NOK) | | | MIDAS with $\Delta$Oil_Dp | | | |
| $\beta_0$ | -0.001 | 0.003 | 0.000 | -0.004 | 0.005 | 0.000 |
| $\beta_1$ | -0.802 | 0.293 | 0.012 | -0.405 | 0.288 | 0.014 |
| $\theta_1$ | 0.261 | 0.388 | 0.013 | -0.020 | 0.485 | 0.013 |
| $\theta_2$ | 0.015 | 0.059 | 0.010 | -0.289 | 0.249 | 0.046 |
| $\eta$ | 0.001 | 0.000 | 0.000 | 0.003 | 0.000 | 0.000 |
| | | | MIDAS with $\Delta$Gas_Dp | | | |
| $\beta_0$ | -0.001 | 0.003 | 0.000 | -0.004 | 0.005 | 0.000 |
| $\beta_1$ | -0.071 | 0.128 | 0.011 | 0.053 | 0.181 | 0.010 |
| $\theta_1$ | 0.072 | 0.495 | 0.018 | 0.051 | 0.503 | 0.013 |
| $\theta_2$ | -0.202 | 0.108 | 0.020 | 0.048 | 0.304 | 0.057 |
| $\eta$ | 0.001 | 0.000 | 0.000 | 0.003 | 0.000 | 0.000 |
| Japan (YEN) | | | MIDAS with $\Delta$Oil_Dp | | | |
| $\beta_0$ | -0.001 | 0.002 | 0.000 | -0.002 | 0.004 | 0.000 |
| $\beta_1$ | -0.267 | 0.191 | 0.011 | -0.340 | 0.258 | 0.010 |
| $\theta_1$ | 0.105 | 0.472 | 0.036 | -0.048 | 0.489 | 0.027 |
| $\theta_2$ | -0.269 | 0.086 | 0.016 | -0.293 | 0.077 | 0.014 |
| $\eta$ | 0.001 | 0.000 | 0.000 | 0.004 | 0.000 | 0.000 |
| | | | MIDAS with $\Delta$DP_index | | | |
| $\beta_0$ | -0.001 | 0.002 | 0.000 | -0.003 | 0.004 | 0.000 |
| $\beta_1$ | 0.056 | 0.472 | 0.028 | 0.095 | 0.528 | 0.015 |
| $\theta_1$ | 0.006 | 0.500 | 0.013 | -0.002 | 0.498 | 0.014 |
| $\theta_2$ | -0.021 | 0.436 | 0.078 | -0.025 | 0.240 | 0.046 |
| $\eta$ | 0.001 | 0.000 | 0.000 | 0.004 | 0.000 | 0.000 |

**Notes:** The Table presents convergence diagnostics for the RW-MH within Gibbs algorithm, namely the average numerical standard error (NSE). These are averages from the estimates obtained over all data points in the recursive estimations of the forecasting procedure. Smaller values of NSE relative to the posterior standard deviations (postStdv) convey an acceptable degree of approximation error. The NSE are based on 4% tapered window in the estimate of the spectral density at zero frequency.

# K Bootstrap Techniques for Forecast Evaluation

## K.1 A Standard Bootstrap Procedure

Our bootstrap procedures to construct $p$-values or $t$-statistics used to assess statistical differences in forecasting performance build upon Kilian (1999) and Rogoff and Stavrakeva (2008). They use a semi-parametric bootstrap with the data generating process (DGP) for the fundamentals specified in an error correction form. They also assume cointegration between the exchange rate and fundamentals.

Similarly, in Chapter 2 we begin by postulating the following DGP under the null of no predictability (the country subscript $i$, is omitted for simplicity):

$$\Delta s_t = v_t^e, \tag{K.1}$$

$$\Delta z_t = c_0 + t + \Upsilon z_{t-1} + \sum_{\ell=1}^{\ell e} B_\ell^e \Delta s_{t-\ell} + \sum_{\ell=1}^{\ell z} B_\ell^z \Delta z_{t-\ell} + v_t^z, \tag{K.2}$$

where, $\Delta s_t = s_t - s_{t-1}$; $\Delta z_t = z_t - z_{t-1}$; $c_0$ is a constant, $t$ is a trend, and $v_t^e$ and $v_t^z$ are $i.i.d$ error terms. We first estimate equations (K.1) and (K.2) via OLS, with lag orders $\ell e$ and $\ell z$ selected using Akaike's Information Criterion (AIC). The AIC also allows us to determine the inclusion or exclusion of the constant, the trend or both.[6] Subsequently, we re-sample with replacement the residuals matrix $(v_t^e, v_t^z)$ in tandem to preserve the contemporaneous correlation in the original sample. We then use the re-sampled residuals to recursively generate pseudo-samples of the variables $s_t$ and $z_t$ with length of $T + 100$. The first 100 observations are discarded to avoid potential bias due to using the sample averages as initial values for the recursions. Each of the predictive models (the FE panel regression and the OLS regression) is in turn used to forecast using the pseudo-samples, and calculate the DMW test-statistic.

To compute the DMW test, first obtain the squared forecast error differences:

$$\widehat{f}_{t+h} = \widehat{fe}^2_{1,t+h} - \widehat{fe}^2_{2,t+h}, \tag{K.3}$$

with $\widehat{fe}_{1,t+h}$ denoting the h-step-ahead forecast error of the RW, and $\widehat{fe}_{2,t+h}$ the corresponding forecast error of the fundamental-based model. Under the null of equal predictive ability, $E(\widehat{f}_{t+h}) = 0$; and the DMW test is: $DMW = \overline{f}\sqrt{P}/[\text{sample variance of } \widehat{f}_{t+h} - \overline{f}]^{1/2}$; where $P$ is the number of out-of-sample forecasts, and $\overline{f}$ is the mean of $\widehat{f}_{t+h}$. We repeat this process of computing the DMW test 1000 times, providing us with an empirical distribution of the statistic. The $p$-value is the proportion of the bootstrap statistics that are above the test-statistic calculated using observed data.

Implementing this bootstrap with the MCMC methods that we use to estimate the TVP regression in Chapter 2 would be computationally very demanding. Thus, to evaluate the forecasts from the TVP regressions we exploit our MCMC methods in

[6] In equation (K.2) the sum of the coefficients of the lags of $\Delta z_t$ is restricted to one to avoid exploding simulated pseudo data (Rogoff and Stavrakeva, 2008).

a procedure that leads to results equivalent to those in the bootstrap above (see also Garratt et al., 2009 and Korobilis, 2013). Since for each draw of the Gibbs sampler we can compute the DMW test, we can obtain an entire empirical distribution of the DMW test. From this distribution we can equally compute the $p$-value in the same manner as above.

## K.2 Bootstrap Procedure in a Data-Mining Environment

In Chapter 3 our Bayesian approach allows for search over many predictors, and the bootstrap procedure just described assumes that each predictor is analyzed in isolation. To take into account concerns about data-mining we extend the procedure above to a data-mining environment, following Inoue and Kilian (2005). See also Rapach and Wohar (2006), for an application to stock returns. The extension involves assuming that under the null of no predictability the DGP comprises:

$$\Delta s_t = v_t^e, \tag{K.4}$$

$$\begin{aligned} \Delta z_{1,t} &= c_{1,0} + t + \Upsilon_1 z_{1,t-1} + \sum_{\ell=1}^{\ell e} B_{1,\ell}^e \Delta s_{t-\ell} + \sum_{\ell=1}^{\ell z} B_{1,\ell}^z \Delta z_{1,t-\ell} + v_{1,t}^z \\ &\vdots \\ \Delta z_{k,t} &= c_{k,0} + t + \Upsilon_k z_{k,t-1} + \sum_{\ell=1}^{\ell e} B_{k,\ell}^e \Delta s_{t-\ell} + \sum_{\ell=1}^{\ell z} B_{k,\ell}^z \Delta z_{k,t-\ell} + v_{k,t}^z, \end{aligned} \tag{K.5}$$

where $v_t^e, v_{1,t}^z, \ldots, v_{k,t}^z$, are $i.i.d$ error terms. As the system of equations suggests, we are now considering $k$ candidate predictors. Each of the equation is also estimated via OLS. We next repeat the same steps as above: re-sample with replacement the residuals matrix, use the re-sampled residuals to recursively generate pseudo-samples of all the variables and discard the initial 100. We then employ each of the predictive model (Single Predictor including or excluding TVar-Coeffs, and the simple forecast combination methods) to forecast using the pseudo-samples, and calculate the DMW test statistic. We repeat this process 1000 times and obtain an empirical distribution of the statistic. The $p$-value is the proportion of the bootstrap statistics that are above the test-statistic calculated using observed data. For BMA and BMS we also generate BMA or BMS forecasts in pseudo-samples, but for each bootstrap we store the maximal DMW statistic, providing us with an empirical distribution of the maximal statistic. After ordering the empirical distribution of maximal statistics the $900^{th}$, $950^{th}$, and $990^{th}$ values constitute the 10%, 5% and 1% critical values, respectively.

# Bibliography


Andreou, E., E. Ghysels, and A. Kourtellos (2010). Regression models with mixed sampling frequencies. *Journal of Econometrics 158*(2), 246–261.

Bacchetta, P. and E. van Wincoop (2004). A scapegoat model of exchange-rate fluctuations. *American Economic Review 94*(2), 114–118.

Bacchetta, P. and E. van Wincoop (2013). On the unstable relationship between exchange rates and macroeconomic fundamentals. *Journal of International Economics 91*(1), 18–26.

Bacchetta, P., E. van Wincoop, and T. Beutler (2010). Can parameter instability explain the Meese-Rogoff puzzle? In *NBER International Seminar on Macroeconomics 2009*, pp. 125–173. National Bureau of Economic Research: University of Chicago Press.

Balke, N. S., J. Ma, and M. E. Wohar (2013). The contribution of economic fundamentals to movements in exchange rates. *Journal of International Economics 90*(1), 1–16.

Beckmann, J. and R. Schussler (2015). Forecasting exchange rates under parameter and model uncertainty. *Journal of International Money and Finance 60*, 267–288.

Berge, T. J. (2013). Forecasting disconnected exchange rates. *Journal of Applied Econometrics 29*(5), 713–735.

Blake, A. P. and H. Mumtaz (2012). Applied Bayesian econometrics for central bankers. Technical Handbook 4, Bank of England.

Boivin, J. (2006). Has US monetary policy changed? Evidence from drifting coefficients and real-time data. *Journal of Money, Credit and Banking 38*(5), 1149–1173.

Breiman, L. (1996). Bagging predictors. *Machine learning 24*(2), 123–140.

Buhlmann, P. and B. Yu (2002). Analyzing bagging. *The Annals of Statistics 30*(4), 927–961.

Byrne, J. P., D. Korobilis, and P. J. Ribeiro (2014). On the sources of uncertainty in exchange rate predictability. Working Paper 2014–16, Adam Smith Business School.

Byrne, J. P., D. Korobilis, and P. J. Ribeiro (2016). Exchange rate predictability in a changing world. *Journal of International Money and Finance 62*, 1–24.

Campbell, J. Y. and S. B. Thompson (2008). Predicting excess stock returns out of sample: can anything beat the historical average? *Review of Financial Studies 21*(4), 1509–1531.

Carter, C. K. and R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika 81*(3), 541–553.

Chen, Y.-C., K. S. Rogoff, and B. Rossi (2010). Can exchange rates forecast commodity prices? *The Quarterly Journal of Economics 125*(3), 1145–1194.

Chen, Y.-c. and K. P. Tsang (2013). What does the yield curve tell us about exchange rate predictability? *Review of Economics and Statistics 95*(1), 185–205.

Cheung, Y.-W. and M. D. Chinn (2001). Currency traders and exchange rate dynamics: a survey of the US market. *Journal of International Money and Finance 20*(4), 439 – 471.

Cheung, Y.-W., M. D. Chinn, and A. G. Pascual (2005). Empirical exchange rate models of the nineties: are any fit to survive? *Journal of International Money and Finance 24*(7), 1150–1175.

Clarida, R., J. Galı, and M. Gertler (1998). Monetary policy rules in practice: some international evidence. *European Economic Review 42*(6), 1033–1067.

Clark, T. E. and K. D. West (2006). Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis. *Journal of Econometrics 135*(1), 155–186.

Clark, T. E. and K. D. West (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics 138*(1), 291–311.

Cogley, T., G. E. Primiceri, and T. J. Sargent (2010). Inflation-gap persistence in the US. *American Economic Journal: Macroeconomics 2*(1), 43–69.

Dangl, T. and M. Halling (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics 106*(1), 157 – 181.

Della Corte, P., L. Sarno, and G. Sestieri (2012). The predictive information content of external imbalances for exchange rate returns: how much is it worth? *Review of Economics and Statistics 94*(1), 100–115.

Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: a personal perspective on the use and abuse of Diebold-Mariano tests. *Journal of Business and Economic Statistics 33*(1), 1–9.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics 13*(3), 253–63.

Draghi, M. (2014). Monetary policy communication in turbulent times. In *Conference De Nederlandsche Bank 200 years: Central banking in the next two decades*. European Central Bank: European Central Bank.

Engel, C. (1996). The forward discount anomaly and the risk premium: a survey of recent evidence. *Journal of Empirical Finance 3*(2), 123–192.

Engel, C., N. C. Mark, and K. D. West (2008). Exchange rate models are not as bad as you think. In *NBER Macroeconomics Annual 2007*, Volume 22, pp. 381–441. University of Chicago Press.

Engel, C., N. C. Mark, and K. D. West (2015). Factor model forecasts of exchange rates. *Econometric Reviews 34*(1-2), 32–55.

Engel, C. and K. D. West (2004). Accounting for exchange-rate variability in present-value models when the discount factor is near one. *American Economic Review 94*(2), 119–125.

Engel, C. and K. D. West (2005). Exchange rates and fundamentals. *Journal of Political Economy 113*(3), 485–517.

Engel, C. and K. D. West (2006). Taylor rules and the deutschmark–dollar real exchange rate. *Journal of Money, Credit and Banking 38*(5), 1175–1194.

Fernandez, C., E. Ley, and M. F. Steel (2001). Benchmark priors for Bayesian model averaging. *Journal of Econometrics 100*(2), 381–427.

Ferraro, D., K. Rogoff, and B. Rossi (2015). Can oil prices forecast exchange rates? An empirical analysis of the relationship between commodity prices and exchange rates. *Journal of International Money and Finance 54*(0), 116 – 141.

Fleming, J., C. Kirby, and B. Ostdiek (2001). The economic value of volatility timing. *The Journal of Finance 56*(1), 329–352.

Foroni, C., M. Marcellino, and C. Schumacher (2013). Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 178*(1), 57–82.

Foroni, C., F. Ravazzolo, and P. J. Ribeiro (2015). Forecasting commodity currencies: the role of fundamentals with short-lived predictive content. Norges Bank Working Paper 2015–14, Norges Bank.

Fratzscher, M., D. Rime, L. Sarno, and G. Zinna (2015). The scapegoat theory of exchange rates: the first tests. *Journal of Monetary Economics 70*(0), 1 – 21.

Garratt, A., G. Koop, E. Mise, and S. P. Vahey (2009). Real-time prediction with UK monetary aggregates in the presence of model uncertainty. *Journal of Business and Economic Statistics 27*(4), 480–491.

Gelfand, A. E. and D. K. Dey (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society. Series B (Methodological) 56*(3), 501–514.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics 4*, 169–194.

Geweke, J. and G. Amisano (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting 26*(2), 216–230.

Geweke, J. and G. Amisano (2011). Optimal prediction pools. *Journal of Econometrics 164*(1), 130 – 141. Annals Issue on Forecasting.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2005). There is a risk-return trade-off after all. *Journal of Financial Economics 76*(3), 509–548.

Ghysels, E., A. Sinko, and R. Valkanov (2007). MIDAS regressions: further results and new directions. *Econometric Reviews 26*(1), 53–90.

Giacomini, R. and B. Rossi (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics 25*(4), 595–620.

Giannone, D. (2010). *Comment on "Can Parameter Instability Explain the Meese-Rogoff Puzzle?"*, pp. 180–190. NBER International Seminar on Macroeconomics 2009. University of Chicago Press.

Goetzmann, W., J. Ingersoll, M. Spiegel, and I. Welch (2007). Portfolio performance manipulation and manipulation-proof performance measures. *Review of Financial Studies 20*(5), 1503–1546.

Gourinchas, P.-O. and A. Tornell (2004). Exchange rate puzzles and distorted beliefs. *Journal of International Economics 64*(2), 303 – 333.

Groen, J. J. (2000). The monetary exchange rate model as a long-run phenomenon. *Journal of International Economics 52*(2), 299–319.

Han, Y. (2006). Asset allocation with a high dimensional latent factor stochastic volatility model. *Review of Financial Studies 19*(1), 237–271.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). New York: Springer.

Hodrick, R. J. and E. C. Prescott (1997). Postwar US business cycles: an empirical investigation. *Journal of Money, Credit and Banking 29*(1), 1–16.

Ince, O. (2014). Forecasting exchange rates out-of-sample with panel methods and real-time data. *Journal of International Money and Finance 43*(0), 1 – 18.

Inoue, A. and L. Kilian (2005). In-sample or out-of-sample tests of predictability: which one should we use? *Econometric Reviews 23*(4), 371–402.

Inoue, A. and L. Kilian (2008). How useful is bagging in forecasting economic time series? A case study of US consumer price inflation. *Journal of the American Statistical Association 103*(482), 511–522.

Jin, S., L. Su, and A. Ullah (2014). Robustify financial time series forecasting with bagging. *Econometric Reviews 33*(5-6), 575–605.

Kilian, L. (1999). Exchange rates and monetary fundamentals: what do we learn from long-horizon regressions? *Journal of Applied Econometrics 14*(5), 491–510.

Kilian, L. and M. P. Taylor (2003, May). Why is it so difficult to beat the random walk forecast of exchange rates? *Journal of International Economics 60*(1), 85–107.

Kim, C.-J. and C. R. Nelson (1999). *State-space models with regime switching: classical and Gibbs-sampling approaches with applications.* London: The MIT Press.

Kim, C.-J. and C. R. Nelson (2006). Estimation of a forward-looking monetary policy rule: a time-varying parameter model using ex post data. *Journal of Monetary Economics 53*(8), 1949–1966.

Koop, G. (2003). *Bayesian econometrics.* London: John Wiley & Sons, Ltd.

Koop, G. and D. Korobilis (2012). Forecasting inflation using dynamic model averaging. *International Economic Review 53*(3), 867–886.

Koop, G. and D. Korobilis (2013). Large time-varying parameter VARs. *Journal of Econometrics 177*(2), 185–198.

Korobilis, D. (2013). VAR forecasting using Bayesian variable selection. *Journal of Applied Econometrics 28*(2), 204–230.

Ledoit, O. and M. Wolf (2008). Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance 15*(5), 850–859.

Li, J., I. Tsiakas, and W. Wang (2015). Predicting exchange rates out of sample: can economic fundamentals beat the random walk? *Journal of Financial Econometrics 13*(2), 293–341.

MacDonald, R. and I. W. Marsh (1996). Currency forecasters are heterogeneous: confirmation and consequences. *Journal of International Money and Finance 15*(5), 665–685.

MacDonald, R. and J. Nagayasu (2000). The long-run relationship between real exchange rates and real interest rate differentials: a panel study. *IMF Staff Papers 48*(1), 116–128.

Mark, N. C. (1995). Exchange rates and fundamentals: evidence on long-horizon predictability. *The American Economic Review 85*(1), 201–218.

Mark, N. C. (2009). Changing monetary policy rules, learning, and real exchange rate dynamics. *Journal of Money, Credit and Banking 41*(6), 1047–1070.

Mark, N. C. and D. Sul (2001). Nominal exchange rates and monetary fundamentals: evidence from a small post-bretton woods panel. *Journal of International Economics 53*(1), 29–52.

Markiewicz, A. (2012). Model uncertainty and exchange rate volatility. *International Economic Review 53*(3), 815–844.

Meese, R. A. and K. Rogoff (1983). Empirical exchange rate models of the seventies: do they fit out of sample? *Journal of International Economics 14*(1), 3–24.

Mitchell, J. and S. G. Hall (2005). Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESER "fan" charts of inflation. *Oxford Bulletin of Economics and Statistics 67*, 995–1033.

Molodtsova, T., A. Nikolsko-Rzhevskyy, and D. H. Papell (2011). Taylor rules and the Euro. *Journal of Money, Credit and Banking 43*(2-3), 535–552.

Molodtsova, T. and D. H. Papell (2009). Out-of-sample exchange rate predictability with Taylor rule fundamentals. *Journal of International Economics 77*(2), 167–180.

Molodtsova, T. and D. H. Papell (2013). Taylor rule exchange rate forecasting during the financial crisis. In *NBER International Seminar on Macroeconomics 2012*, pp. 55–97. National Bureau of Economic Research: University of Chicago Press.

Mumtaz, H. and L. Sunder-Plassmann (2013). Time-varying dynamics of the real exchange rate: an empirical analysis. *Journal of Applied Econometrics 28*(3), 498–525.

Nelson, C. R. and A. F. Siegel (1987). Parsimonious modeling of yield curves. *The Journal of Business 60*(4), 473–89.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–08.

Panopoulou, E. and S. Vrontos (2015). Hedge fund return predictability; To combine forecasts or combine information? *Journal of Banking and Finance 56*, 103 – 122.

Papell, D. H. (1997). Searching for stationarity: purchasing power parity under the current float. *Journal of International Economics 43*(3), 313 – 332.

Park, C. and S. Park (2013). Exchange rate predictability and a monetary model with time-varying cointegration coefficients. *Journal of International Money and Finance 37*(0), 394 – 410.

Pettenuzzo, D., A. G. Timmermann, and R. I. Valkanov (2015). A Bayesian MIDAS approach to modeling first and second moment dynamics. Working Papers 76, Brandeis University, Department of Economics and International Businesss School.

Raftery, A. E., M. Kárnỳ, and P. Ettler (2010). Online prediction under model uncertainty via dynamic model averaging: application to a cold rolling mill. *Technometrics 52*(1), 52–66.

Rapach, D. E. and J. K. Strauss (2010). Bagging or combining (or both)? An analysis based on forecasting US employment growth. *Econometric Reviews 29*(5-6), 511–533.

Rapach, D. E. and J. K. Strauss (2012). Forecasting US state-level employment growth: an amalgamation approach. *International Journal of Forecasting 28*(2), 315 – 327.

Rapach, D. E., J. K. Strauss, and G. Zhou (2010). Out-of-sample equity premium prediction: combination forecasts and links to the real economy. *Review of Financial Studies 23*(2), 821–862.

Rapach, D. E. and M. E. Wohar (2006). In-sample vs. out-of-sample tests of stock return predictability in the context of data mining. *Journal of Empirical Finance 13*(2), 231 – 247.

Rogoff, K. S. and V. Stavrakeva (2008). The continuing puzzle of short horizon exchange rate forecasting. NBER Working Paper 14071, National Bureau of Economic Research.

Rossi, B. (2006). Are exchange rates really random walks? Some evidence robust to parameter instability. *Macroeconomic dynamics 10*(01), 20–38.

Rossi, B. (2013). Exchange rate predictability. *Journal of Economic Literature 51*(4), 1063–1119.

Rossi, B. and T. Sekhposyan (2011). Understanding models forecasting performance. *Journal of Econometrics 164*(1), 158–172.

Sarno, L. and G. Valente (2009). Exchange rates and fundamentals: footloose or evolving relationship? *Journal of the European Economic Association 7*(4), 786–830.

Stock, J. H. and M. W. Watson (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics 14*(1), 11–30.

Stock, J. H. and M. W. Watson (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting 23*(6), 405–430.

Taylor, J. B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy 39*, 195 – 214.

Taylor, J. B. (2009). The financial crisis and the policy responses: an empirical analysis of what went wrong. NBER Working Paper 14631, National Bureau of Economic Research.

Tibshirani, R. and K. Knight (1999). Model search by bootstrap "bumping". *Journal of Computational and Graphical Statistics 8*(4), 671–686.

Trecroci, C. and M. Vassalli (2010). Monetary policy regime shifts: new evidence from time-varying interest rates. *Economic Inquiry 48*(4), 933–950.

Wang, W., J. and J. J. Wu (2012). The Taylor rule and forecast intervals for exchange rates. *Journal of Money, Credit and Banking 44*(1), 103–144.

Watson, M. W. (2007). How accurate are real-time estimates of output trends and gaps? *Economic Quarterly 93*(2), 143–161.

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies 21*(4), 1455–1508.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica 64*(5), 1067–1084.

West, M. and J. Harrison (1997). *Bayesian Forecasting and Dynamic Models*. New York, NY: Springer New York.

Wolff, C. C. P. (1987). Time-varying parameters and the out-of-sample forecasting performance of structural exchange rate models. *Journal of Business and Economic Statistics 5*(1), 87–97.

Wright, J. H. (2008). Bayesian model averaging and exchange rate forecasts. *Journal of Econometrics 146*(2), 329–341.

Zou, H. and T. Hastie (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67*(2), 301–320.