Economics Working Paper Series

2016/022

Forecasting the nominal Exchange rate movements in a Changing world. The case of the U.S. and the U.K.

Pantelis Promponas and David Peel

The Department of Economics
Lancaster University Management School
Lancaster LA1 4YX
UK

LUMS home page: http://www.lancaster.ac.uk/lums/

Forecasting the nominal Exchange rate movements in a Changing world. The case of the U.S. and the U.K.*

Pantelis Promponas[a] and David Peel[b]

December 13, 2016

Abstract

Exchange rate forecasting has become an arena for many researchers the last decades while predictability depends heavily on several factors such as the choice of the fundamentals, the econometric model and the data form. The aim of this paper is to assess whether modelling time-variation and other forms of instabilities may improve the forecasting performance of the models. Paper begins with a brief critical review of the recently developed exchange rate forecasting models and continues with a real-time forecasting race between our fundamentals-based models, a DSGE model, estimated with Bayesian techniques and the benchmark random walk model without drift. Results suggest that models accounting for non-linearities may generate poor forecasts relative to more parsimonious and linear models.

Keywords: Forecasting exchange rate, Exchange rate literature, Instability, Taylor rule, PPP, UIP, Money supply, Real-time estimation, Time-Varying models, DSGE model, Bayesian methods.

JEL Classification: C53, E51, E52, F31, F37, G17.

---

* Authors would like to thank Mike Tsionas and Efthymios Pavlidis for their helpful suggestions. We would also like to thank participants of the ESRC PhD Conference in Lancaster for their useful comments.
[a] Corresponding author at: Department of Economics, Lancaster University Management School, Lancaster University, UK. Email address: p.promponas@lancaster.ac.uk.
[b] Department of Economics, Lancaster University Management School, Lancaster University, UK.

1. Introduction

This empirical paper revisits one of the long-standing puzzles in international economics stemming from the findings of Meese and Rogoff (1983a, 1983b), that macroeconomic fundamentals are weak predictors of the exchange rate movements especially at the short-horizon. In fact, subsequent literature suggests that the a-theoretical random walk model without drift appears to be the most successful model in forecasting out-of-sample the nominal exchange rates. This exchange rate disconnect puzzle has been researched by many researchers and practitioners for the last three decades, and many macroeconomic models and econometric techniques have been developed in an attempt to outperform this naïve model. A critical survey by Rossi (2013a) supports the view that exchange rate predictability depends on several factors such as the choice of the predictors, the forecasting model (linear or non-linear), the econometric techniques, the forecasts horizon, the estimation scheme, the forecasts evaluation methods and finally whether we are dealing with in-sample or out-of-sample predictability.[1] So what we observe in the literature is different studies, each of them focusing on a different set of the above factors, while the goal is always the same, to outperform the driftless random walk model.

In this study we examine the potential causes of the currency disconnect puzzle, employing a real-time out-of-sample forecasting race between several fundamentals-based models proposed in the literature and the benchmark random walk without drift. The predictors that we employ are motivated by fundamental –based models, while a number of these models are well known from earlier research. These are the UIP model, the PPP model, the Monetary model and the Term Structure Forward Premium model (e.g. see Clark and West, 2006; Clarida and Taylor, 1997 and Cheung *et al.*, 2005). In the last decade, Taylor-rule fundamentals have also been used as predictors for the future exchange rate changes (e.g. see Engel and West, 2005, 2006 and Molodtsova and Papell, 2009). This forecasting equation follows the monetary policy's principles as were set in Taylor (1993).We place a great emphasis on the possible non-linearities of the exchange rate forecasting models caused by the time-varying relationship between exchange rates and fundamentals, as well as on the relevance of the predictors which

[1] As Rossi (2013b) mentions, the empirical evidence of the in-sample predictability does not necessarily entails the out-of-sample predictability. The usual method for testing the in-sample predictability is by estimating the model in hand and then conducting a traditional Granger-causality test checking the significance of the estimated parameters using a simple t-test. On the other hand, the out-of-sample predictability is tested by splitting the sample into two parts, while the first part is used for estimations and generating forecasts and the second one is used for evaluating and comparing the forecasts with the actual data of that part.

may potentially change over time. Time-variation, as a special form of non-linearities (Rossi, 2013a) and parameters' instability have drawn the attention of many studies (see Sarantis, 2006; Clarida *et al*., 2003; Baillie and Kilic, 2006; Mark and Moh, 2002 and Byrne *et al.*, 2014, 2016) mainly due to the unstable macroeconomic conditions, the monetary policy shifts, asymmetric preferences and the weak rational expectation where agents are not fully informed about the economy and the monetary authorities' intervention in the exchange rate targeting policy, especially at the short-horizon (Mark and Moh, 2002). This unstable and sometimes weak connection between currencies and fundamentals is also explained in Bacchetta and van Wincoop (2004, 2013) using the "scapegoat theory", where observed variables are assigned with more weight (and become scapegoats) when exchange rate fluctuations are mainly driven by the unobserved macroeconomic shocks. Fratzscher *et al.* (2015) refers to it as a "rational confusion" of FX market's agents who interpret the true parameters of the model conditioning only on the observed predictors, at times when the exchange rate fluctuates in response to the unobservables. Hence accounting for the scapegoat fundamentals and the time-varying weights assigned to them may be helpful in an out-of-sample FX forecasting exercise.

In order to investigate whether the forecasting performance of our models are improved or not when considering these potential instabilities, we employ both linear and non-linear econometric "vehicles" in our exercise. We use a Bayesian Vector Autoregressive (BVAR) model, a Homoscedastic Time-varying parameter BVAR model which allows for the coefficients to change over time, a Heteroscedastic TVP-BVAR model accounting for time-variation in both parameters and the covariance matrix and finally Bayesian Dynamic Model Averaging and Selection (DMA, DMS) models which not only allow for the parameters and the covariance to change over time, but also allow for the entire set of predictors to switch over time. TVP models similar to these have been recently used in the exchange rate forecasting literature exhibiting a relevant out-of-sample success (see Byrne *et al.* 2014, 2016; Sarantis, 2006) and other studies using non-linear smooth transition regressions (STR) and regime-switching models (see Sarno *et al.*, 2006; Clarida *et al.*, 2003). Bayesian and time-varying approaches has become topical in the forecasting literature and we believe that is a good chance, informative and useful to empirically test the predictive performance of these econometric models in this challenging research area.

Apart from the fundamentals-based models discussed above we also novelly include an open-economy New Keynesian Dynamic Stochastic General Equilibrium (DSGE) model in our forecasting exercise. The nominal exchange rate forecasting literature has not reflected these structural models. DSGE models have become in the last decades an empirical tool for the central banks and other policy making institutions as they rely heavily on them for forecasting inflation, output gap and other macroeconomic variables since they may help their decision-making process. We therefore use a richly specified DSGE model following Gali and Monacelli (2005), Justiniano and Preston (2010) and Alpanda *et al.* (2011), estimated with Bayesian likelihood methods as used in An and Schorfheide (2007a) to compete the naïve benchmark in out-of-sample predictability.

We also use vintage data for the variables which suffer from consecutive revisions (GDP, Price level, Money stock, etc.), since we have evidence from the literature that forecasting performance of models is increased when data which are available at the time that agents are making forecasts are used (see Orphanides, 2001, Croushore and Stark, 2001 and Croushore, 2006).[2] Also Clements (2012) pinpoints the fact that findings about predictors' content and usefulness may be misleading when fully revised data are used instead. Hence we carry out a real-time forecasting race which uses variables' observations that were available to the forecaster at that specific vintage dates. Our study considers both iterated and direct forecasts for the 1-, 2- and 3-quarters ahead period, although Wright (2008) argues that both methods deliver very similar results. We use recursive estimation scheme in our forecasting exercise, while forecasts are evaluated using the Theil's-U statistic and the Clark and West (2006, 2007) (hereafter CW06) one-sided test of predictive superiority. Throughout the paper we consider the U.S. as the home country and the U.K. as the foreign one. We selected these countries mainly due to the availability of complete and well-structured real-time data bases, given the data requirements and the large number of variables that we employ in this study.

This paper is organized as follows: Section 2 provides a brief critical review of the literature for both theoretical and empirical exchange rate models and their characteristic findings for the sake of completeness. Section 3 presents the specifications of the time-varying and non-time-varying econometric models and the DSGE model. Section 4 summarises the forecasting models and discusses the data details, forecasts implementation and evaluation methods. Section 5 analyses the out-of-sample forecasting race results, provides a discussion for the forecasting performance of each model and the importance of the possible instabilities. Also a sensitivity analysis is conducted for robustness purposes. Section 6 concludes outlining the main empirical findings.

[2] More recently Molodtsova *et al.* (2008), Molodtsova and Papell (2009) and Nikolsko-Rzhevskyy (2011) refer to the importance of using real-time (vintage) data in forecasting exercises using several exchange rate models. Real-time data in the forecasting literature has become increasingly important and crucial and therefore it is necessary for our study, although lack of real-time data bases and data limitations are deterrents.

2. Exchange rate models and predictors

The most commonly used predictors in the exchange rate forecasting literature are the interest rate, real output, output gap, price level, money supply and forward premia. In this section we briefly present the relevant models that we use in this paper, along with a discussion about their usefulness and successfulness with a critical point of view. This may offer to the reader a wider picture of the literature as well as an understanding of what we have learned about the exchange rate forecasting so far.

Model 1. Driftless Random walk (RW)

A naïve a-theoretical random walk model without drift which represents the benchmark model in this empirical work. If the natural log of the exchange rate is denoted by s_t (measured as the home price for a unit of the foreign currency), E_t (.) the expectation at time *t* and *h* the horizon, then model predicts:

$$E_t(s_{t+h} - s_t) = 0 \ . \tag{1}$$

The vast majority of studies in the literature compare the forecasting models with the above specification, although studies like Engel *et al.* (2008) and Engel and Hamilton (1990) have tested both driftless and random walk with drift and found that random walk without drift delivers better results.

Model 2. Uncovered Interest rate Parity (UIP)

The UIP model due to Fisher (1896), where the gain from holding a currency should counterbalance the opportunity cost and risk of holding money in this currency. This can be written as:

$$E_t(s_{t+h} - s_t) = a + \beta(i_t - i_t^*) + u_{t+h} \quad , \tag{2}$$

where s_t is the logarithm of the spot exchange rate, i_t is the nominal interest rate, also $\alpha = 0$ and $\beta = 1$. Empirical findings for the studies which use the interest differential as predictors are not very positive. Clark and West (2006) reports predictability only for one out of four countries that considered and only for the short-horizon (one-month ahead). Somewhat moderate results from Molodtsova and Papell (2009) where UIP model was estimated without sign restrictions, they found predictability only for Australia and Canada out of twelve countries when a constant is included in the regression, and for Australia, Canada, Japan and Switzerland when it's not.[3]

[3] Cheung *et al.* (2005) contributes to longer-horizon predictability who found that there is an empirical support for the very long-horizons (20-quarters ahead) compared to the disappointed results for the short-horizons. Similar

Model 3. Forward Premium Term Structure (FPTS)

There is a consensus within the literature that the Risk-Neutral Efficient Market Hypothesis has been rejected (see Hodrick, 1987; Taylor, 1995; Chinn and Meredith, 2004 and Chinn, 2006), while the most common empirical way for testing this hypothesis is by estimating the Fama (1984) equation, assuming that Covered Interest rate Parity $(f_t^h - s_t = i_{t,h} - i_{t,h}^*)$ holds, where f_t^h is the forward exchange rate maturing in *h* periods ahead. Hence, assuming that UIP conditions hold, the Fama equation can be written as:

$$\Delta s_{t+h} = \alpha + \beta(f_t^h - s_t) + u_{t+h} \quad , \tag{3}$$

where $\alpha = 0$, $\beta = 1$ and $(f_t^h - s_t)$ is the forward premium; difference between the forward and the spot exchange rate. The vast majority of the studies estimating the above unrestricted equation, have found that constant α is different from zero and slope coefficient β is statistically significantly different from zero and actually very close to minus one (e.g., see Bilson, 1981; Froot and Thaler, 1990 and Bekaert and Hodrick, 1993). The opposite sign of β, which has become a stylised fact, is also referred as the "forward bias puzzle". The fact that β is significantly different from zero implies that forward premiums may contain enough predictive content for the depreciation rate and this information can be extracted and exploited in an out-of-sample forecasting exercise.

A seminal work which exploits the predictive content of the forward premia is that of Clarida and Taylor (1997), which uses the spot rate and the forward exchange rates at different maturities as dependent variables in a linear Vector Error Correction model (VECM), allowing for the *term structure of forward premia* to represent the long-run co-integrating vectors. What they found is that the term structure of forward premiums have statistically significant in-sample predictability but more importantly out-of-sample predictability, outperforming the random-walk forecasts in most of the cases. An extension of this work is that of Clarida *et al.* (2003) who examined the improvement in the predictive performance of the above model by considering for possible non-linearities, using a Markov-switching VECM allowing for regime shifts only in the intercept and the variance covariance matrix.[4] Empirical findings are quite

findings for the long-horizon predictability come from Alquist and Chinn (2008), while strong long-horizon in-sample predictability can also be found in Chinn and Meredith (2004).

[4] Although authors mention that regime-shifts may be allowed to the parameters as well, they eventually used the specification as described above.

promising since the non-linear specification is able to improve upon the linear VECM and the random walk, in both short and long-horizons.[5]

Model 4. Purchasing Power Parity (PPP)

According to Dornbusch (1985), the strong or absolute version of the PPP model which introduced by Cassel (1918), states that the real price of a common and identical basket of goods in two countries should be the same at all times (law of one price). Hence, taking logs the relative PPP model can be written as:

$$s_t = \alpha + \beta(p_t - p_t^*) + u_t \quad , \tag{4}$$

where p_t is the logarithm of price levels, $\alpha = 0$ and $\beta = 1$. The empirical findings for this model are disappointing. Cheung *et al.* (2005) report that forecasts generated by the PPP model are discouraging especially for the short horizons (1- and 4-quarters ahead) while for much longer horizon (20 quarters) the forecasting performance of the model is improved since it outperforms the driftless random walk in most of the cases considered. Similar pessimistic results are obtained from Molodtsova and Papell (2009) who found predictability of the exchange rate only for 1 out of 12 countries for the 1-quarter ahead horizon.[6]

Model 5. Monetary model

The monetary model which is due to Frenkel (1976), Mussa (1976) and Bilson (1978) can be derived using the conventional real money demand as a function of output and interest rate. The demand for money functions of two countries are:

$$\text{Home:} \quad m_t - p_t = a_1 y_t - a_2 i_t \,, \tag{5}$$

$$\text{Foreign:} \quad m_t^* - p_t^* = a_1 y_t^* - a_2 i \,, \tag{6}$$

where m_t is the logarithm of nominal money supply, y_t is the logarithm of real output and * denotes the foreign country. By subtracting the foreign demand function from the home one, bringing the money supply to the right hand side and assuming that the PPP holds at every point in time, we can derive the estimable monetary model by adding a constant term, a slope parameter for the relative money supply and an error term. Hence, we have:

[5] Nucci (2003) focuses on the forward premiums' predictive content as well, based on the evidence that there is a co-movement between the excess returns of cross-currency investments. He therefore investigated if forward premia of different currencies and maturities could have enough predictive content for the future spot FX rate of the home country. Although the in-sample evidence is empirically very supportive, the out-of-sample does not seem positive since he finds predictability only for 1 out of 3 currencies using a VECM.

[6] As regards the Molodtsova and Papell (2009), authors derive a specification for the PPP model as described in Mark (1995).

$$s_t = a + \beta(m_t - m_t^*) - \gamma(y_t - y_t^*) + \delta(i_t - i_t^*) + u_t \ . \tag{7}$$

If we consider the presence of sticky price adjustment (where prices in the goods' market adjust much slower than in the financial market), then the inflation rates (or price levels) enter into the monetary model as predictors. We should note that some studies use the inflation differential as predictor (see Cheung *et al.*, (2005), while others use the price levels instead (see Engel and West, 2005). Empirical evidence for these models' out-of-sample predictability is negative. Starting from Meese and Rogoff (1983a,b), random walk dominates both monetary models, while Cheung *et al.*(2005) cannot find any predictability even in 20-quarters ahead horizon among the five currencies they examine. Similar pessimistic results are reported by Chinn and Meese (1995) where both models cannot outperform the naïve benchmark at 1-month and 12-months horizon.[7] An exception is that of Chen and Mark (1996) which find quite positive results where flexible monetary model does predict well out-of-sample four exchange rate changes at either 3- and 4-year horizon.

Model 6. Taylor-rule fundamentals

As mentioned before, Taylor-rule predictors have been used the last decade for exchange rate forecasting purposes. This rule was formulated by Taylor (1993) and describes how central banks set the short-term nominal interest rates as a function of the inflation, the deviation of the inflation from its target level, the output gap and the equilibrium real interest rate. Also following Clarida *et al.* (1998) and Woodford (2003) we can assume that the nominal interest rate adjusts gradually to its target level and following Clarida *et al.* (1998) and Molodtsova and Papell (2009) we assume that monetary policy targets the real exchange rate making the PPP to hold at all time. Hence, by taking the difference of the two countries' policy rules and assuming that the UIP holds, then one derives similar specifications as used in Molodtsova *et al.* (2008), Molodtsova and Papell (2009) and Molodtsova *et al.* (2011):

$$\Delta s_{t+1} = \omega + a_{h1}\pi_t - \alpha_{f1}\pi_t^{\ f} + \alpha_{h2}y_t^{gap} - \alpha_{f2}y_t^{fgap} - a_{f3}q_t^{\ f} + \rho_h i_{t-1} - \rho_f i^f{}_{t-1} + \eta_{t+1} \tag{8}$$

where π_t is the inflation rate, $y_t{}^{gap}$ is the output gap measured as the natural logarithm of actual real GDP minus its estimated potential, $q_t{}^f$ is the real exchange rate of the foreign country, i_t is the short term nominal interest rate and f denotes the foreign country.[8] Out-of-sample empirical

[7] Engel *et al.* (2008) used panel regressions for 18 currencies and find that flexible model improves significantly upon the driftless RW in 5 cases only at 1-quarter horizon and in 11 out of 18 cases at 16-quarter ahead.

[8] Molodtsova and Papell (2009) consider several specifications of the above model; i) a *symmetric* or *asymmetric* model, depending on whether both countries' rules include the same fundamentals or not, respectively, ii) a model with or without interest rate *smoothing*, iii) a *homogeneous* or *heterogeneous* model where coefficients of inflation

findings are in favour of this model, while most of them are highly sample-dependent. Molodtsova *et al.* (2008), Molodtsova and Papell (2009), Molodtsova *et al.* (2011) and Inoue and Rossi (2012) are some of the studies which report the impressive forecasting performance of this model under different specifications, while Rogoff and Stavrakeva (2008) criticizes the robustness of some of the aforementioned studies, focusing on the forecasts evaluation using different test statistics and different date for the rolling regressions. Engel *et al.* (2008) uses a calibrated Taylor-rule forecasting model and results are not positive when forecasts are compared with the naïve benchmark model.

3. Linear and Non-linear models

According to Rossi's (2013b) critical survey the predictive power of many macroeconomic variables has been found unstable over time making the forecasting task less reliable, while the most challenging issue is to identify the source of these instabilities and time-variations (as a special form of non-linearities (Rossi, 2013a)) and to choose the best model which will lead to reliable forecasts. She also mentions studies such as Rogoff and Stavrakeva (2008), Giacomini and Rossi (2010), Sarno and Valente (2009), Bacchetta and van Wincoop (2013) and others, which report the relative predictive content of the macro fundamentals which appears to exist in an ephemeral manner, increasing the estimation errors of the parameters.[9] Hence, we are testing several linear and non-linear models, each of them incorporating different kind of instabilities, time-variations and flexibilities, in an attempt to make significant inferences on which model generates more accurate forecasts and beats the driftless random walk model.

BVAR(p) model

If we keep Korobilis's (2013) notations then we can write the reduced form VAR specification as:

$$y_t = c_t + B_1 y_{t-1} + B_2 y_{t-2} + B_3 y_{t-3} + \dots\dots\dots + B_p y_{t-p} + \varepsilon_t \ , \tag{9}$$

rates, output gaps and lagged interest rates are set equal or not, respectively. For our exchange rate forecasting race we use a heterogeneous, asymmetric (including the real exchange rate of the home country) forecasting model, with an interest rate smoothing process and a constant included.

[9] Rossi (2013b) lists additional cases and topics other than the exchange rate literature, where macro predictors' unstable predictive ability is documented. One of them is in finance when one is trying to forecast the stock returns or when output growth predictability is examined (see for instance Goyal and Welch, 2003; Paye and Timmermann, 2006 and Giacomini and Rossi, 2006).

where p is the number of lags, y_t is an $m \times 1$ vector of $t = 1, \ldots, T$ observations collecting the explanatory variables, errors $\varepsilon_t \sim N_m(0, \Sigma)$ where Σ is the covariance matrix of $m \times m$ dimension and m is the number of variables.[10] We can rewrite (9) in seemingly unrelated (SUR) form as:

$$y_t = z_t \beta + \varepsilon_t \ , \tag{10}$$

where $\beta = vec(B)$ with $B = [c, B_1', \ldots, B_p']$ being an $m \times k$ matrix collecting both intercepts and parameters , the error terms $\varepsilon_t \sim N_m(0, \Sigma)$, $z_t = I_m \otimes x_t = I_m \otimes (1, y_{t-1}, \ldots, y_{t-p})'$ is an $m \times n$ matrix, $n = mk$ and $k = mp + 1$. Although VAR models (due to the seminal work of Sims (1980)) are very popular, they do entail the danger of over-parameterization since most of the empirical studies require the inclusion of more than four or five dependent variables. Hence, the need for more flexible specifications caused the development of shrinkage methods such as restricting the parameters' space or imposing prior information and beliefs, reducing the dimensionality of the empirical models (see Koop and Korobilis (2010)). The linear BVAR model that we employ in this paper is accompanied by the *Minnesota prior* specification (see Litterman (1986) and Doan *et al.* (1984)). This specification assumes a value of 1 for the prior mean ($\underline{b}$) for parameters of the first own lag of each variable and a zero value for the rest. As regards the prior variance of the parameters, a diagonal matrix $(\underline{V})$ is assumed, with a prior of $100\, s_i^2$ for the intercepts, $1/r^2$ for their own lag coefficients and $\lambda s_i^2 / r^2 s_l^2$ otherwise, where $r = 1, \ldots, p$, $i = 1, \ldots, m$, $j = 1, \ldots, k$, s_i^2 is the residual variance from the unrestricted univariate AR(p) and $\lambda = 0.1$ is a hyperparameter which controls the shrinkage level. A non-informative prior for Σ has been assigned $(\underline{S} = \underline{v} = 0)$.The posterior inference can be found in the Appendix A.

<u>*Homoscedastic TVP- BVAR(p) model*</u>

Following Korobilis's (2013) notations, the reduced form homoscedastic TVP-BVAR model can be written as before, adding the time-variation notations:

$$y_t = c_t + B_{1,t} y_{t-1} + B_{2,t} y_{t-2} + B_{3,t} y_{t-3} + \ldots + B_{p,t} y_{t-p} + \varepsilon_t \ , \tag{11}$$

This model can be written once again in a SUR form, as:

[10] Korobilis (2013) examines the case of variable selection on a variety of different BVAR specifications. In our study we use Korobilis's (2013) BVAR models, modifying and changing notations according to our needs removing his variable selection techiques.

$$y_t = z_t \beta_t + \varepsilon_t, \tag{12}$$

$$\beta_t = \beta_{t-1} + u \tag{13}$$

where y_t is an m×1 vector of t =1,….,T observables' time series, β_t is an n×1 state vector $\left[c_t', vec(B_{1,t}'),...,vec(B_{p,t}')\right]'$ of the intercepts and the parameters, the error terms $\varepsilon_t \sim N_m(0,\Sigma)$, with Σ being an m×m constant covariance matrix, $z_t = I_m \otimes x_t = I_m \otimes (1, y_{t-1},..., y_{t-p})'$ is an m×n matrix, $n = mk$ and $k = mp+1$. Also, equation (13) describes the evolution of the parameters (as a driftless random walk process) and the n×n covariance matrix Q for the error term $u_t \sim \mathrm{N}(0,Q)$. Both equations (12) and (13) represent a state-space model with eq. (12) being the measurement equation and eq. (13) the state one. As regards the priors, we can follow the spirit of Primiceri (2005) and we use a training sample of size of τ = 40 observations to calibrate the parameters' priors. A time-invariant parameter VAR(1) model is estimated with OLS and the estimates are used as initial conditions for the Kalman filter.

Priors

The priors for the parameters are obtained as described above:

$$\beta_0 \sim N(\beta_{OLS}, 4 \cdot V(\beta_{OLS})),$$

The priors for the covariance matrices Σ and Q are:

$$\Sigma \sim IW(\underline{S}^{\Sigma}, \underline{V}^{\Sigma}) \quad \text{with} \quad \underline{S}^{\Sigma} = I_m \quad \text{and} \quad \underline{V}^{\Sigma} = m+1.$$

$$Q \sim IW(\underline{S}^{Q}, \underline{V}^{Q}) \quad \text{with} \quad \underline{S}^{Q} = 0.0001 \cdot \tau \cdot V(\beta_{OLS}) \quad \text{and} \quad \underline{V}^{Q} = \tau .$$

Posteriors

We estimate the parameters by sampling sequentially from the following conditional distributions:

a) Sample β_t conditioning on the data using the Carter and Kohn (1994) algorithm along with the Kalman filter and the smoother procedure. A detailed description of the algorithm can be found in Kim and Nelson's (1999) textbook (page 191).

b) Sampling Q^{-1} from the conditional density:

$$Q \mid \beta, \Sigma, z, y \sim IW(\overline{S}^{Q^{-1}}, \overline{V}^{Q}),$$

where $\overline{S}^{Q^{-1}} = \left(\underline{S}^{Q} + \sum_{t=1}^{T} (\beta_t - \beta_{t-1})(\beta_t - \beta_{t-1})' \right)^{-1}$ and $\overline{V}^{Q} = t + \underline{V}^{Q}$.

c) Sampling Σ^{-1} from the conditional density:

$$\Sigma | \beta, Q, z, y \sim IW(\overline{S}^{\Sigma^{-1}}, \overline{V}^{\Sigma}),$$

where $\overline{S}^{\Sigma^{-1}} = \left(\underline{S}^{\Sigma} + \sum_{t=1}^{T} (y_t - z_t \beta_t)(y_t - z_t \beta_t)' \right)^{-1}$ and $\overline{V}^{\Sigma} = t + m + 1$.

Heteroscedastic TVP- BVAR(p) model

This model assumes that both parameters and innovations are time-varying. The importance of possible non-linearities and instabilities, in the form of multivariate stochastic volatility, has been discussed in D'Agostino *et al.* (2013), Primiceri (2005) and Koop and Korobilis (2010), underlying the fact that capturing possible shocks' heteroscedasticity may be proved crucial in generating good forecasts for the macroeconomic variables. Hence we believe that this model may contribute in the exchange rate forecasting literature. We closely follow Primiceri's (2005) model. The reduced form model can be written as:

$$y_t = c_t + B_{1,t} y_{t-1} + B_{2,t} y_{t-2} + B_{3,t} y_{t-3} + ... + B_{p,t} y_{t-p} + u_t, \tag{14}$$

where y_t is an $m\times1$ vector of the observed variables, $B_{i,t}$ collects the parameters with $m\times m$ dimensions, $u_t \sim N(0, \Omega_t)$ with the time-varying covariance matrix Ω_t. The covariance matrix can be decomposed as follows:

$$A_t \Omega_t A_t' = \Sigma_t \Sigma_t', \tag{15}$$

where A_t is the lower triangular matrix with ones on its diagonal, summarising the relationships between the variables and Σ_t is the diagonal matrix with the standard deviations of the structural innovations as its elements. The aforementioned matrices are depicted below:

$$A_t = \begin{bmatrix} 1 & 0 & . & . & 0 \\ \alpha_{21,t} & 1 & . & . & . \\ . & . & 1 & . & . \\ . & . & . & 1 & 0 \\ \alpha_{n1,t} & . & . & \alpha_{nn-1,t} & 1 \end{bmatrix}, \qquad \Sigma_t = \begin{bmatrix} \sigma_{1,t} & 0 & . & . & 0 \\ 0 & \sigma_{2,t} & . & . & . \\ . & . & . & . & . \\ . & . & . & \sigma_{n-1,t} & 0 \\ 0 & . & . & 0 & \sigma_{n,t} \end{bmatrix}.$$

The above model can be rewritten as a linear and Gaussian state space representation, where the measurement equation is given by:

$$y_t = z_t \beta_t + A_t^{-1} \Sigma_t \varepsilon_t. \tag{16}$$

Stacking all the parameters and intercepts in a vector; $B_t = [c_t, B_{1,t}, B_{2,t}, \ldots, B_{p,t}]$ and $\beta_t = vec(B_t)$, $z_t = I_m \otimes x_t = I_m \otimes (1, y_{t-1}, \ldots, y_{t-p})'$ is an $m \times n$ matrix, $n = mk$ and $k = mp + 1$. The dynamics of the time varying parameters of the reduced form model, which represent the transition equations, follow driftless random walk processes as:

$$\beta_t = \beta_{t-1} + v_t \tag{17}$$

$$a_t = a_{t-1} + \zeta_t \tag{18}$$

$$\log \sigma_t = \log \sigma_{t-1} + \eta_t \tag{19}$$

where eq. (19) follows a geometric random walk. The structural error terms are assumed to be normally distributed and are independent to the parameters, such as:

$$V = \begin{bmatrix} \varepsilon_t \\ v_t \\ \zeta_t \\ \eta_t \end{bmatrix} \overset{iid}{\sim} N\left(0, \begin{pmatrix} I_m & 0 & 0 & 0 \\ 0 & Q & 0 & 0 \\ 0 & 0 & S & 0 \\ 0 & 0 & 0 & W \end{pmatrix}\right). \tag{20}$$

Note that $E(\varepsilon_t \varepsilon_t') = \Sigma_\varepsilon = I_m$, $\otimes$ is the Kronecker product and the matrix S is a block diagonal matrix (with the two blocks S_1 and S_2), enhancing the independency of the parameters' evolution among the equations. As regards the parameters' priors, we follow the ones proposed in Primiceri (2005) and Cogley and Sargent (2005), which seems to perform empirically well. Given that we have set our priors, the Gibbs sampler (MCMC algorithm) is used to simulate the conditional densities for β^T, A^T, Σ^T and V. A detailed description of the priors and the sequential sampling for the posterior inference can be found in Appendix A.

Bayesian Dynamic Model Averaging and Selection

The advantage of this model is not only to allow for both coefficients and covariance matrices to change over time but it also allows for the entire set of predictors to switch over time, depending on the relevance and importance of each predictor. Structural breaks in macroeconomic variables, structural parameters' instability and uncertainty, changes in the monetary policy and consequences from the "scapegoat theory" necessitate the use of a flexible econometric model which is able to pick the most relevant predictors-fundamentals, based on a posterior probability-weight. Bacchetta and van Wincoop (2013) provide an empirical survey describing how financial market's participants place more weight on macro fundamentals which they do not actually deserve it, while such a common practise has been found to shift across the fundamentals over time. The second feature of the DMA model is that it manages a large number of predictors and this allows us to use a wide range of them coming from all the exchange rate models that we consider in this paper as the theoretical drivers of the exchange rate changes, and examine which fundamental is more relevant to the FX future movements.[11]

We closely follow the DMA model as developed by Raftery *et al.* (2010) and Koop and Korobilis (2012, 2013) in a Heteroscedastic TVP-AR specification. The state space model can be written as:[12]

$$y_t = z_t^{(k)}\theta_t^{(k)} + \varepsilon_t^{(k)} \tag{21}$$

$$\theta_t^{(k)} = \theta_{t-1}^{(k)} + \eta_t \tag{22}$$

where y_t is the log of exchange rate change, $k = 1$,….., K is the number of models, each using different set of predictors, $z_t^{(k)}$ is a matrix of predictors that each of these K models uses and $\theta_t^{(k)}$ collects the corresponding coefficients. Also $\varepsilon_t^{(k)} \sim N(0, H_t^{(k)})$ and $\eta_t^{(k)} \sim N(0, Q_t^{(k)})$. Following Korobilis's (2012) notations, let $\Theta_t = (\theta_t^{(1)'}, \ldots\ldots, \theta_t^{(k)'})'$, $y_t = (y_t, \ldots\ldots, y_t)'$ and $L_t \in \{1, 2, \ldots, K\}$ indexing which individual model applies at every time period. The DMA model is accompanied by a probability calculated in each time period, that indicates which model should be used more (which predictors are more relevant) in the forecasting exercise. So given this individual-model probability, DMA will simply average the *h*-period ahead forecasts across these models, while DMS will pick the individual model, with the corresponding relevant predictors, with the highest probability to forecast the exchange rate movements.[13] Hence, given the priors for the unobserved parameters (initial conditions for the Kalman filter) and a prior model probability, the Bayesian inference can be easily achieved using the Kalman filter. The advantage of this model is that it uses some forgetting factors which allow us to avoid the usual MCMC simulation methods which would have been computationally unaffordable, by replacing the Kalman filter's components which require simulation. Another feature of the forgetting factors is that they control the weight assigned to the past observations

[11] Assuming that *m* predictors (including the intercept) are included in the DMA model, then 2^m forecasting models will be examined. When *m* is very large (more than 18 predictors) a forecasting exercise can be computationally demanding and sometimes infeasible.

[12] This is actually a TVP-ARX model (as in Ljung, 1987) which allows for both lags of the independent and exogenous variables to predict.

[13] Wright (2008) describes the Bayesian Model Averaging model as a "judicious pooling" of the predictive content from the whole set of numerous predictors and uses the most relevant variables-predictors to forecast the dependent variable.

and hence ruling the evolution speed of the coefficients. When these factors are set equal to 1, then there is neither forgetting nor time-variation in our parameters and the DMA converges to a standard recursive Bayesian Model Average (BMA) model.[14] More details for the priors, posterior inference, the model probabilities and the forgetting factors can be found in the Appendix A.

New Keynesian DSGE model

New Keynesian DSGE models estimated using Bayesian likelihood methods, have become a standard tool for the monetary policy authorities and other policy making institutions around the world, for macroeconomic analysis, examining business cycle dynamics and forecasting purposes (Smets and Wouters, 2007). Over the last decade these models have been extended embodying open-economy characteristics and allowing for more observables to enter into the model (such as the nominal exchange rates) as well as a richer set of disturbances.

The model that we use is an open-economy DSGE model for the U.S., following closely Gali and Monacelli (2005), Justiniano and Preston (2010), Steinbach *et al.* (2009) and Alpanda *et al.* (2011). In the goods market, monopolistically competitive firms set the prices, and households provide their labour services and set their wages. External habit formation in the households' consumption, staggered prices and wages as in Calvo (1983), indexing wages following Rabanal and Rubio-Ramirez (2005) and goods' prices according to Smets and Wouters (2002) to the previous period's inflation, are some of the model's features. Model also assumes that domestic retail firms import goods from abroad and sell them domestically paying the exporters of the foreign country in terms of the home currency using the exchange rate. So far the law of one price holds, but when the domestic retailer will set the imported product's price he will face his own optimal mark-up problem and the price that he will charge may not be the same as the one he paid to the exporter. This will lead to incomplete exchange rate pass-through in the short run, while the deviations from the law of one price will be eliminated only in the very long-run (Monacelli, 2005). To close the model, the UIP condition as in Adolfson *et al.* (2008) is used, which deals with several components of the country's risk premium and a Taylor-rule with interest rate smoothing. The rich set of disturbances includes the home productivity shock, consumption demand shock, cost-push shocks for both home and foreign

[14] A standard BMA model has been used in Wright (2008, 2009) for FX and inflation forecasting purposes. In both studies model exhibit sufficient forecasting performance outperforming the benchmark models in most of the cases in short and long horizon.

country, wage cost-push shock, home country risk-premium shock, monetary policy shock, as well as shocks for the foreign output, the inflation and the interest rate.

The equations which characterize the equilibrium of the model, after the variables are log-linearized around their steady-state, as well as its micro-foundation and the estimation details can be found in the Appendix A.

4. Empirical section

This section summarizes the fundamentals-based forecasting models, a description and discussion of the data used along with the transformations suggested from the literature, details for the forecasts implementation and evaluation methods.

<u>Exchange rate forecasting models</u>

For the purpose of estimation and forecasting, the corresponding Y_t vector of dependent variables, for the BVAR(*p*) and TVP-BVAR(*p*) (both Homoscedastic and Heteroscedastic) for each fundamentals-based model, is displayed as follows:

UIP predictors: $Y_t = \left[\Delta s_t, i_t, i_t^*\right]'$, (23)

FPTS predictors: $Y_t = \left[\Delta s_t, \left(f_t^1 - s_t\right), \left(f_t^3 - s_t\right), \left(f_t^6 - s_t\right), \left(f_t^{12} - s_t\right)\right]'$, (24)

Taylor-rule predictors: $Y_t = \left[\Delta s_t, y_t^{gap}, y_t^{gap*}, \pi_t, \pi_t^*, q_t, i_{t-1}, i_{t-1}^*\right]'$, (25)

where Δs_t is the nominal exchange depreciation rate, i_t is the nominal interest rate, $\left(f_t - s_t\right)$ is the forward exchange premium in different monthly maturities, y_t^{gap} is the output gap, π_t is the inflation rate, q_t is the real exchange rate and * denotes the variables of the foreign economy.

For the rest of the fundamental-based models we follow Mark (1995), Molodtsova and Papell (2009), Engel *et al.* (2008) and Byrne et al. (2016), modelling the nominal exchange rate change as a function of its deviation from its fundamental-value:

$$\Delta s_{t+h} = a_t + \beta_t \left(\Omega_t - s_t\right) + u_{t+h} , \tag{26}$$

where Ω_t is the fundamental implied value.[15] Hence for the rest forecasting models the vector of predictors will contain both the fundamentals and the current nominal exchange rate:

PPP predictors: $Y_t = \left[\Delta s_t, p_t, p_t^*, s_t\right]'$, (27)

Monetary model (flexible) predictors: $Y_t = \left[\Delta s_t, y_t, y_t^*, m_t, m_t^*, i_t, i_t^*, s_t\right]'$, (28)

where p_t is the price level, y_t is the real GDP and m_t denoting the money supply. We generate 45,000 draws and discard the first 5,000 for every parameter. We also thin the chain by keeping only the every 10th draw in order to mitigate the autocorrelation in the Markov chain. At the end we obtain the mean of the marginal conditional posterior distribution as the point estimate.

As regards the DMA and DMS models which follow a Heteroscedastic TVP-AR(p) specification, we opt to include a complete set of predictors coming from the exchange rate models that we examine in the literature section and thought to be fundamental by theory and empirical studies. We therefore consider the following variables as predictors:

1. 1-month Forward (USD/GBP) exchange rate premium.
2. 3-month Forward (USD/GBP) exchange rate premium.
3. 6-month Forward (USD/GBP) exchange rate premium.
4. 12-month Forward (USD/GBP) exchange rate premium.
5. U.S. Real GDP (seasonally adjusted).
6. U.K. Real GDP (seasonally adjusted).
7. U.S. Output gap (HP-filtered).
8. U.K. Output gap (HP-filtered).
9. U.S. Money supply (M1).
10. U.K. Money supply (M4).
11. Real (USD/GBP) exchange rate.
12. U.S. Price Inflation (annualised).
13. U.K. Price Inflation (annualised).
14. U.S. 10-year maturity Government Bond rates.
15. U.K. 10-year maturity Government Bond rates.

For this kind of econometric model, literature suggests transforming the data into stationary following Koop and Korobilis (2012, 2013) and Byrne *et al.* (2014). More details follow in the data section.

As regards the DSGE model we use the following 10 observables: the real output growth (Δy_t), the labour productivity growth (Δz_t), the nominal exchange rate depreciation rate (d_t),

[15] One of the papers using this prediction model in a panel data framework is the Engel *et al.* (2008), while Engel *et. al* (2015) uses similar model adding an extra term which describes the deviation of factors, generated from a cross-section exchange rates, from the s_t.

the consumer price inflation (π_t), the GDP deflator inflation $(\pi_{h,t})$, the nominal wage inflation $(\pi_{w,t})$, the nominal interest rate (i_t), the foreign real output growth (Δy_t^*), the foreign GDP deflator inflation $(\pi_{h,t}^*)$ and the foreign nominal interest rate (i_t^*), as in Alpanda *et al.* (2011).

Data description

As mentioned before, we conduct a real-time forecasting study which is heavily based on data vintages that were available to the forecaster at the time the predictions were made, instead of fully revised data that most of the pseudo out-of-sample empirical studies use in the literature. As Clements (2012, 2015) and Clements and Galvao (2013) mention, a real-time forecasting exercise should mimic the conditions, environment and information framework that the forecaster was facing at the time he was making the predictions, and hence use data from the vintages that were published at that time.[16]

For the BVAR, TVP-BVAR and the DMS-DMA models we use 18 in total, major quarterly macroeconomic variables (for both U.S. and U.K.) spanning from 1979Q1 to 2012Q3. Starting with the vintage data, real GDP (seasonally adjusted) for the U.S. extracted from the FED of Philadelphia real-time database and for the real GDP of U.K. data extracted from the Office of National Statistics (ONS). For both countries Output gap is measured as *log (Actual real GDP) – log (Potential GDP)* while the potential one is obtained by applying the Hodrick-Prescott filter.[17] As regards the price level and the inflation rate, we use the GDP deflator (seasonally adjusted) and we compute the inflation rate as the rate of inflation over the previous four quarters, $\pi_t = deflator_t - deflator_{t-4}$ (GDP deflators in natural logs). Price index for GDP is taken from the Federal Reserve Bank of Philadelphia for U.S. and for the U.K. from Bank of England. For data that is not revised, we use Pacific Exchange Rate Service website for the nominal USD/GBP exchange rate.[18] We calculate the real USD/GBP exchange rate as, the nominal USD/GBP exchange rate plus the log of the UK price level minus the log of US price level. The bilateral USD/GBP exchange rate is defined as the domestic price for a unit of British pound. The USD/GBP forward exchange rates are from the Bank of England website.[19] We use

[16] We should note that in the real-time datasets there is always one period lag between the vintage date and the last (most recent) observation of that vintage. So, assuming for example the 2000:Q1 vintage, the last observation of that vintage is at 1999:Q4.

[17] Following Clausen and Meier (2005), we backcast and forecast our dataset by 12 quarterly datapoints, with an AR(4) model, in order to correct for the end-of-sample problem that filters like the HP present.

[18] Pacific Exchange rate Service's website can be found in: http://fx.sauder.ubc.ca/data.html.

[19] Forward rates have the following codes: XUDLDS1, XUDLDS3, XUDLDS6, XUDLDSY.

the M4 as a money supply proxy for the U.K. as in Chinn and Meese (1995) and Byrne *et al.* (2016) and M1 for the U.S. as in Chen and Mark (1996).[20]

Due to our dataset span, it would not be wise to use the Federal fund rates and the Official Bank rates as a proxy for the nominal interest rates since both figures hit the zero lower bound at the end of 2008. Instead, there are studies which suggest the long-term interest rates as an alternative monetary policy instrument. McCough *et al.* (2005) characteristically mention that the long-term rates might be a physical substitute as they are highly related to the future expected path of the short-term interest rates, while Jones and Kulish (2013) show that long-term rates are good instruments for conducting monetary policy and sometimes performing better than the standard Taylor-rules. Also Chinn and Meredith (2004) provide empirical evidence showing that testing UIP model using interest rates on longer-maturity bonds, lead to better in-sample results consistent to the UIP theory. Boughton (1988) and Sarantis (1995) examined UIP specifications with interest rates at different maturities finding different behaviour each time, making clear that the appropriateness of the rates' maturity is not an easy decision. We therefore use the 10-year Treasury note rates as a proxy for the nominal interest rates throughout this paper. As for the variables transformation for the DMA-DMS models, a table with the respective transformations can be found in the Appendix A, Table A2.

As regards the DSGE model we use the CPI (seasonally adjusted) from the FED of Philadelphia for the consumer price inflation, the Employment Cost Index as a proxy for the nominal wages and the Output per hour index as a proxy for the labour productivity.[21] Before we estimate the model we demean the data since zero inflation, growth and depreciation rate are assumed at the steady state, while the sample mean is added back to the generated forecasts before we evaluate them. All variables are transformed into natural logarithms while interest rates are divided by 100.

Forecasts Implementation and Evaluation

Our full sample runs from 1979Q1 to 2012Q3, while the out-of-sample period used for the forecasts evaluation runs from 2006Q3 to 2012Q3. The number of lags (based on the BIC) for the VAR models has been set to 1, while we use a TVP-AR(2) and 1 lag length for the predictors for the DMA-DMS models (as in Koop and Korobilis, 2012). We opt to run recursive

[20] We use revised data for the money stock since vintage data for this time-series is not available for the U.K.

[21] We would like to thank Dr. Thomas Stark of the Federal Reserve Bank of Philadelphia for supplying us with his real-time data for the employment cost indexes.

estimations rather than rolling ones since regressions under the latter scheme has the potential gain of lessening the parameter instability effects over time (Cheung *et al.*, 2005). Hence, recursive estimation scheme will allow us to examine the predictive content of the fundamentals and the performance of our models, taking into account these time-variation effects. In addition, we do not apply any sign restrictions throughout the estimations as well as we assume heterogeneous coefficients for each single predictor. As regards the generated forecasts, we use iterated forecasts for the VAR models by bringing our models in the following form as in D'Agostino *et al.* (2013) and Korobilis (2013):

$$y_t = c_t + B_t y_{t-1} + \varepsilon_t \quad , \tag{29}$$

where $y_t = \left[y_t', \ldots.., y_{t-p+1}'\right]'$, $c_t = \left[c_t', 0, \ldots., 0\right]'$, $\varepsilon_t = \left[\varepsilon_t', 0, \ldots., 0\right]'$ and $B_t = \begin{bmatrix} B_{1,t} \ldots B_{p-1,t} & B_{p,t} \\ I_{m(p-1)} & 0_{m(p-1)\times m} \end{bmatrix}$.

Then iterated forecasts for h =1-, 2- and 3-quarters ahead are generated according to the following formula:

$$E(y_{t+h|t}) = \sum_{j=1}^{h} \hat{B}_t^{j-1} \hat{c}_t + \hat{B}_t^h y_t \,. \tag{30}$$

For the DMA models we generate direct forecasts as in Koop and Korobilis (2012), as a weighted average of the model-specific predictions with $\pi_{t|t-1,k}$ denoting the single-model's weight:

$$E(y_{t|t-1}) = \sum_{k=1}^{K} \pi_{t|t-1,k} z_t^{(k)} \hat{\theta}_{t-1}^{(k)} \,, \tag{31}$$

while the DMS selects the single model with the highest probability (weight) and uses this one for generating the forecasts.[22]

As regards the forecasts evaluation we use the relative RMSFE and the CW06 test of predictive ability testing the hypothesis that Random walk and our forecasting model predict the same, against the alternative that our model outperforms the benchmark RW model in predictive accuracy. This test is appropriate for nested models and predictions generated with recursive estimations as well.

[22] Koop and Korobilis (2012) argue that iterated forecasts with DMA-DMS models require predictive simulations, where given the large number of predictors will make this task computationally infeasible. Nevertheless, forecasts are still comparable since according to Wright (2008) BMA models deliver quite similar results under both methods.

5. Empirical results and discussion

In this section we present our empirical results and discuss our findings comparing them with the relevant literature. Firstly we should refer to Rossi's (2013a) conclusions about the most successful predictors and econometric methods in the FX forecasting literature. What she characterizes as a "negative" stylized fact is the relevant failure of the PPP and Monetary fundamentals to predict the exchange rate movements at the short horizon, and the limited successfulness of the non-linear modes. Also literature does not agree on whether UIP fundamentals predict well at short horizons and monetary predictors at long horizons. On the other hand, Taylor-rule fundamentals present a significant predictive ability especially at short horizon, while the Bayesian Model Averaging model seems to perform well in the out-of-sample exchange rate forecasting exercises. Also data handling and transformations play a crucial role on the empirical results since robustness with different samples has been pivotal for the final conclusions. In addition we should not forget that comparing findings among different studies is not an easy task, since we have already mentioned that different empirical factors (data, predictors' selection, econometric method, etc.) will definitely lead to different results (Rossi, 2013a). Hence, there is inherently a uniqueness in every forecasting study, and any "bad" result (failure to outperform the benchmark random walk model) does not entail the absence of contribution. Nevertheless we can carefully discuss our results, given the difficulty mentioned in the literature to predict the FX rates, and find the points where we agree or not with the literature.

Table 1. Relative RMSFE of the BVAR(1) model *vs* the RW model for *h* = 1-, 2- and 3-quarters ahead.

	Δs_t	FPTS	UIP	PPP	Monetary	Taylor-rule
BVAR(1)	*t+1*	**0.991***	1.009	**0.950***	**0.985***	**0.965***
	t+2	1.054	1.051	**0.986**	**0.995**	**0.990***
	t+3	1.016	1.032	**0.981***	**0.984**	**0.995**

Notes: This table shows the root mean square forecast errors using the FPTS, UIP, PPP, Monetary and Taylor-rule fundamentals as described in section 4. Values in bold denote the metrics which are below one. Also asterisks indicate the cases where the null hypothesis of equal predictive accuracy (one-sided CW06 test) is rejected against the alternative of outperforming the benchmark RW model at 1% (***), 5% (**) and 10% (*) significance level.

The first results are shown in Table 1 where forecasts are generated from the linear BVAR(1) model using predictors from the FPTS, UIP, PPP, Monetary and Taylor-rule model, as discussed in section 4. Overall in 10 out of 15 cases our models outperform the benchmark driftless RW model, according to the relative RMSFE which is less than one (in bold). The majority of these results are confirmed by the CW06 test at 10% significance level. More specifically, the UIP fundamentals seem to have not enough out-of-sample predictive power for the depreciation rate especially for longer horizons. This is in line with Cheung *et al.* (2005) who find more positive evidence for the shorter horizons, while Clark and West (2006) and Molodtsova and Papell (2009) do not find any predictability of the USD/GBP depreciation rate for 1-month ahead horizon using linear regressions. The term structure of forward premia seem to be significantly (at 10% level) good predictors but only for the 1-quarter ahead horizon. This is in contrast to the Clarida and Taylor (1997) whose VECM performs much better in longer horizons, while they use both the forward premia and the forward rates as predictors. Results are way more positive when the PPP, Monetary and Taylor-rule fundamentals are used as predictors for both short and long horizon. As regards the Taylor-rule fundamentals our findings are in line with the majority of the literature, where the RW has been dominated in Molodtsova and Papell (2009) using a similar to ours theoretical model, while Engel *et al.* (2008) couldn't significantly predict the depreciation rate using a restricted asymmetric model with a constant, no interest rate smoothing and homogeneous coefficients. Similar results we obtain using the monetary fundamentals where we outperform the benchmark martingale difference model at all horizons, significantly only for the 1-quarter ahead though. This finding is in contrast to the majority of the literature which confess the poor predictability of the monetary fundamentals. Cheung *et al.* (2005) does not find any predictability at the short and very long horizons, Engel *et al.* (2008) find some predictability using an error-correction framework while Molodtsova and Papell (2009) do not find any predictive content of the monetary fundamentals, testing different income elasticities, for the short-horizon. We end the linear analysis with the PPP fundamentals where we significantly outperform the naïve model at all horizons. Cheung *et al.* (2005) find predictability for the very long horizon only, Engel *et al.* (2008) for 1- and 16-quarter ahead horizon but only when a drift is included in their PPP model, while Molodtsova and Papell (2009) fail to significantly generate better forecasts than the RW at 1-month ahead horizon.

Table 2. Relative RMSFE of the Homoscedastic TVP- BVAR(1) model *vs* the RW model for h = 1-, 2- and 3-quarters ahead.

	Δs_t	**FPTS**	**UIP**	**PPP**	**Monetary**	**Taylor-rule**
	t+1	**0.969**	1.015	1.029	1.191	1.036
Homoscedastic	*t+2*	1.044	1.059	1.191	3.374	3.406
TVP-BVAR(1)	*t+3*	1.023	1.049	1.128	4.338	6.523

Notes: See Table's 1 notes for details.

Table 3. Relative RMSFE of the Heteroscedastic TVP- BVAR(1) model *vs* the RW model for h = 1-, 2- and 3-quarters ahead.

	Δs_t	FPTS	UIP	PPP	Monetary	Taylor-rule
	$t+1$	**0.986***	1.017	**0.967***	1.237	1.028
Heteroscedastic	$t+2$	1.076	1.076	1.006	1.461	1.232
TVP-BVAR(1)	$t+3$	1.097	1.044	**0.995**	1.234	1.037

Notes: See Table's 1 notes for details.

Moving to the results generated from the TVP models, at a first glance we can see the lack of the forecasting improvement. Although we were expecting for the TVP models to forecast better than the time-invariant ones, due to the reasons mentioned before, literature seems to support the opposite view. As Rossi (2013a) mentions, the empirical evidence does not seem to be in favour of the non-linear models. They actually fit better in-sample than forecasting out-of-sample (see Terasvirta, 2006; Chin, 1991 and Chinn and Meese, 1995). So, even if we may not be surprised by the moderate forecasting performance of the TVP exchange rate models, we believe that it's useful and a good opportunity to make this empirical analysis, expose their power, compare and discuss their results. According to our findings and the relative RMSFE metric, the only noteworthy case which displays a slight improvement are the FPTS fundamentals for the 1-quarter ahead horizon, which generates better forecasts by 2.2% when the Homoscedastic-TVP is used, and a 0.5% improvement when the Heteroscedastic-TVP specification is used compared to the BVAR respectively. It is apparent that for none of the rest cases (fundamentals) the non-linear models exhibit any forecasting improvement, although the PPP predictors outperform the naïve model at the 1- and 3-quarters ahead horizon when both parameters and the covariance matrix are allowed to evolve over time.

The poor forecasting ability of the proposed TVP models may be due to several reasons which have been discussed in the literature, while some possible solutions have been proposed. The first reason is that TVP-BVAR models are dealing with quite many parameters (estimating the parameters for all time periods in each recursion) with probably short sample period (due to the fact that we sacrifice a sufficient sample for calibrating our data-based priors) which may lead to poor and imprecise in-sample parameter estimates (Koop and Korobilis, 2010). Another inherent drawback of this class of models is the fact that they use the same set of predictors in every time period until the sample exhausts, assuming that *all* the explanatory variables are more or less relevant for forecasting the LHS variable. Also results from the Table 2 indicate that as the number of the explanatory variables increases (Monetary and Taylor-rule predictors) the predictive power of the model decreases probably due to the in-sample overfit (Koop and Korobilis, 2012 and Clements *et al.*, 2004).[23,24,25]

We therefore follow two potential solutions in this study. First we conduct a sensitivity analysis in order to investigate the change of the TVP-models' out-of-sample forecasting performance by imposing non-informative normal priors $\beta_0 \sim \mathrm{N}(0,10^2)$ for all the predictors' parameters instead of the data-based priors.[26,27] Details about the priors for the Heteroscedastic TVP model can be found in the Appendix A. We believe that exposing the models on different priors may be crucial and helpful for their performance. Doing so we also "release" the training sample that we used for the data-based priors calibration, and we include it in our estimation sample. The second solution is the usage of the DMA and DMS models which take into account the relevance of explanatory variables' predictive content at each point in time, potentially reducing the set of the predictors in every time period.

Sensitivity analysis

Results in Table 4 verify the improvement of the forecasting performance of the fundamentals-based models with respect to different prior specification. At a first glance the non-informative priors deliver better forecasts especially for the PPP, Monetary and Taylor-rule models outperforming the RW only for the short horizon (1-quarter ahead). So, although it is obvious that training sample priors deliver worse results, still improving upon the benchmark model is a hard task since cases with relative ratio below 1 for the Homoscedastic TVP model are only

---

[23] Koop and Korobilis (2012) is a US inflation forecasting study which documents the predictive failure of the TVP models to outperform the benchmark models, while the proposed DMA and DMS models were found to forecast, out-of-sample, much better at both short and long horizons. They also find that these models' shrinkage has a great contribution, in terms of good predictive performance, in their forecasting exercise.

[24] Another sensitivity check would be that of discarding the explosive draws obtained during the Gibbs sampling process as in D'Agostino *et al.* (2013). However we keep all the remaining draws after the burn-in period and the every 10th draw.

[25] Clements *et al.* (2004) is a critical survey which compares the linear with the non-linear forecasting models like the Markov-switching and the smooth-transition models, from other studies. Their conclusion centred on the relative poor forecasting performance of the non-linear models and their inability to mimic the dynamics of the economy. They also argued that the parsimony and simplicity of the linear models may be proved sometimes more useful, while the large number of parameters and the in-sample overfitting are their main drawbacks.

[26] As regards the prior for the Homoscedastic TVP-BVAR(1) covariance matrix $Q \sim IW(\underline{S}^Q, \underline{V}^Q)$ where $\underline{S}^Q = 0.0001 * diag(\underline{V}_{\beta_0})$ and $\underline{V}^Q = (1+n)2$, as in Korobilis (2013), while the posterior sampling remains the same.

[27] The Minnesota prior has also been tested but out-of-sample results are not better than using the non-informative $N(0, 10^2)$ priors.

3 out of 15 (also verified by the CW06 test), compared to the 1 out of 15 cases using the informative priors.[28]

Table 4. Relative RMSFE of the Homoscedastic TVP- BVAR(1) model *vs* the RW model for h = 1-, 2- and 3-quarters ahead.

	Δs_t	FPTS	UIP	PPP	Monetary	Taylor-rule
	t+1	1.045	1.073	**0.974***	**0.978***	**0.981*****
Homoscedastic	*t+2*	1.155	1.178	1.009	1.074	1.011
TVP-BVAR(1)	*t+3*	1.453	1.124	1.186	2.172	1.091

Notes: These results are obtained using the non-informative priors as presented before. See Table's 1 notes for details.

Table 5. Relative RMSFE of the Heteroscedastic TVP- BVAR(1) model *vs* the RW model for h = 1-, 2- and 3-quarters ahead.

	Δs_t	FPTS	UIP	PPP	Monetary	Taylor-rule
	t+1	1.015	1.026	**0.940***	**0.981***	**0.995****
Heteroscedastic	*t+2*	1.073	1.053	**0.963****	1.135	1.103
TVP-BVAR(1)	*t+3*	1.063	1.054	**0.994***	1.005	**0.993**

Notes: These results are obtained using the non-informative priors as presented before. See Table's 1 notes for details.

As regards the Heteroscedastic TVP model, the difference between the results that corresponding priors deliver, is more striking. First of all, the majority of the ratios have decreased indicating the forecasting improvement of the models, while the ratios below one have increased to 6 out of 15 cases. The PPP fundamentals' predictive content appears to be quite sufficient in forecasting the FX movements, outperforming the RW for all horizons, the monetary model improves upon the benchmark only for the 1-quarter ahead horizon, while the Taylor-rule fundamentals are able to forecast well for the 1- and 3-quarters ahead horizon (Table 5). Overall, the PPP, Monetary and the Taylor-rule fundamentals exhibit a substantial improvement, while the FPTS and the UIP predictors seem to predict slightly worse when we switch to the non-informative priors.

Results from the Bayesian DMA and DMS models are mixed. Both models significantly beat the driftless random walk model but only for the 1-quarter ahead horizon while for the rest horizons both models' forecasts are almost the same with the benchmark's forecasts (see Table 6). Similar model without incorporating the time-variation in the parameters and innovations

[28] Byrne et al. (2016) uses similar to ours econometric vehicle (Homoscedastic TVP-BVAR) generating, inter alia, the GBP/USD change forecasts using the Taylor-rule fundamentals and three forecasts samples, finding mixed evidence of predictability, while the most successful monetary policy specification appears to be a homogeneous rule with interest rate smoothing targeting the real FX rate for the home country.

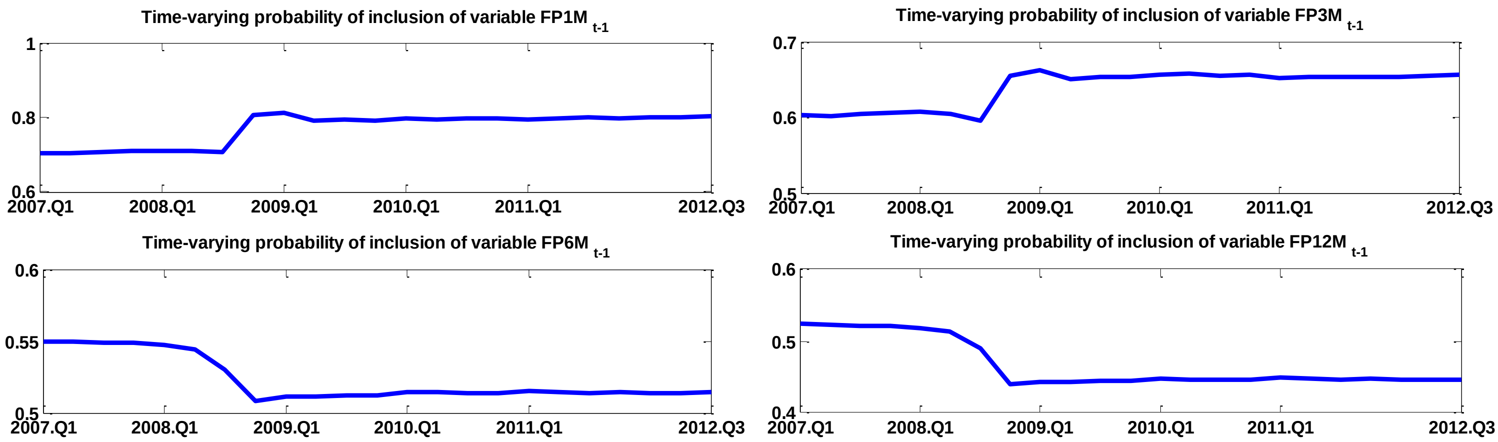
has been used in Wright (2008) for FX movements forecasting purposes. His results are mixed, finding significant predictability for the four bilateral FX rates that he examines except for the USD/GBP rate, using a set of 15 predictors with financial and macroeconomic variables. His inability to significantly outperform the benchmark model may give credit to the corresponding dynamic models that we use and the choice of our predictors which appear to sufficiently capture the time-variations, compute the posterior probability and forecast well especially at short horizon.[29]

Table 6. Relative RMSFE of Bayesian DMA-DMS models *vs* the RW model for h = 1-, 2- and 3-quarters ahead.

	Δs_t	**15 Major Predictors**		Δs_t	**15 Major Predictors**
	t+1	**0.970***		*t+1*	**0.979****
DMA	*t+2*	1.055	**DMS**	*t+2*	1.018
	t+3	1.062		*t+3*	1.042

Notes: See Table's 1 notes for details. Also 15 theory-based predictors (as described in section 4), two lags of the dependent variable, one lag of the fundamentals and a intercept are used in both models.

The probability (weight) that DMA model assigns to each predictor at each point in time is of great interest and importance. We therefore plot the time-varying posterior probabilities of inclusion of the predictors throughout the forecasting sample, indicating which predictor has the most relevant predictive content over time. We focus at the 1-qurter ahead horizon where we find significant predictability.



[29] Byrne *et al.* (2014) examining the sources of the FX rate changes predictability uncertainty, uses BMA and BMS incorporating parameters' time-variation, finding predictability for most of the cases (assuming the US as the foreign country) for horizon greater than one month. They also find that uncertainty lies on the estimation errors and inability to capture the correct degree of coefficients' time-variation at the 1-month horizon.

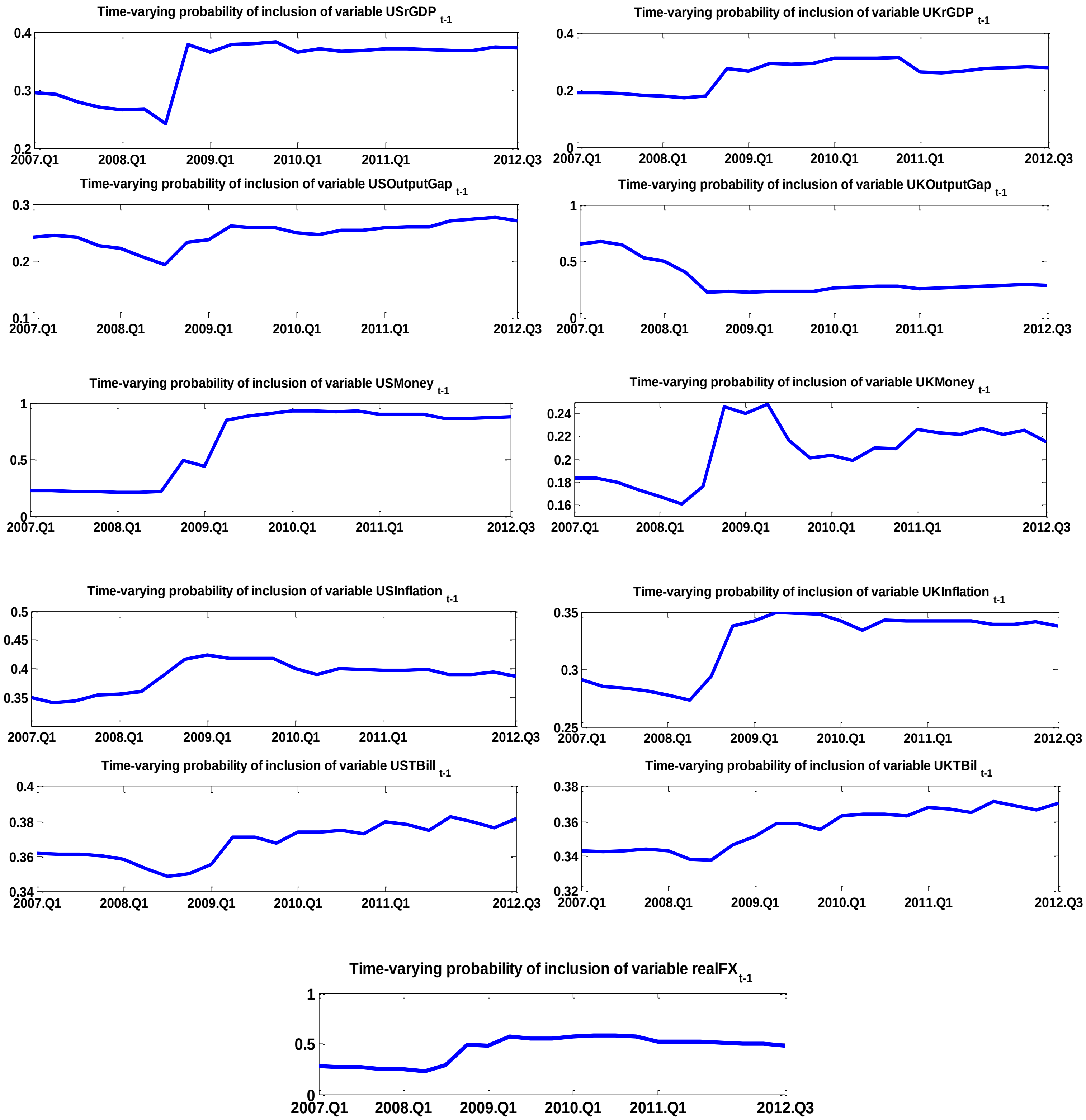


Figure 1. Time-varying posterior probability of inclusion of predictors for the 1-quarter ahead horizon.

It is obvious from the graphs that not all the predictors are useful in forecasting since probabilities of inclusion above 0.5 throughout the forecasting sample have been achieved only for 5 predictors. These are the forward premiums and especially the 1-month and 3-month forward premia, and the U.S. money stock which forecasting importance is quite high from 2009:Q2-2012:Q3. The rest of the predictors do not present any impressive predictive information, while probabilities do not appear to switch abruptly during that period.

As mentioned before, DSGE models have not been used extensively in the nominal exchange rate forecasting literature, and therefore we believe that such model's exposure would make this study more complete. According to our results the structural model cannot outperform the naïve benchmark although the forecasts generated are very similar to the RW model especially for the 1- and 3-quarters ahead horizon. An analytical table with both prior and posterior densities can be found in the Appendix Table A1.

Table 7. Relative RMSFE of the DSGE model *vs* the RW model for h = 1-, 2- and 3-quarters ahead.

	Δs_t	
	t+1	1.282
DSGE	*t+2*	1.060
	t+3	1.034

Notes: See Table's 1 notes for details.

To the best of our knowledge, Zorzi *et al.* (2016) and Alpanda *et al.* (2011) are some of the very few studies focusing on the nominal FX forecasting, by setting-up rich DSGE. Both studies are based on similar theoretical framework and references (Justiniano and Preston, 2010 and Gali and Monacelli, 2005) while their results are mixed, with those of Alpanda *et al.* (2011) being much more positive at both short and long horizon.

6. Conclusions

We revisited the well-known Meese and Rogoff puzzle in an attempt to find the suitable econometric model and macroeconomic fundamentals with the adequate predictive content and the conditions under which we can forecast the nominal USD/GBP exchange rate changes and significantly outperform the driftless random walk model. We also focus on the highly topical issue of whether time-variation, as a special form of non-linearity, in both parameters and innovations can be crucial in the forecasting performance of our models.

Empirical evidence suggest that the two most successful models are the BVAR(1) model with the Minnesota prior and the Heteroscedastic TVP-BVAR(1) model with the non-informative priors as described in the main text. This first finding lead us to mixed conclusions where we cannot safely infer about the usefulness and the forecasting improvement of the models when time-variation is taken into account. Our findings can merely join Clements *et al.* (2004) conclusion and Koop and Korobilis (2012) finding that forecasting with TVP models or with non-linear models more generally, sometimes may lead to poor out-of-sample results, while their "Achilles' heel" seems to be the large number of the variables that these models are dealing with and the over-fitting problem that is caused. On the other hand, parsimonious specifications and linearity can be proved more useful leading to better estimates and forecasts. Also the choice of the priors for the TVP models has been proved of great importance, since the behaviour and the forecasting results of the models do change significantly and are improved when we switch from the training sample priors to the non-informative priors. As regards the choice of the fundamentals, the most successful variables in terms of out-of-sample predictive content are the price levels, since the PPP-based model beats the benchmark random walk model in most of the cases with better performance in the 1-quarter ahead horizon. The next best predictors are the Taylor-rule fundamentals and the Monetary model predictors especially for the short horizon. The 1-quarter ahead horizon in our study has been proved the most predictable, when PPP, Taylor-rule and Monetary fundamentals are used, although literature do not agree on the existence of predictability at this horizon when Monetary predictors are used (Rossi, 2013a).

Results from the Bayesian DMA and DMS models are on slightly different direction. Although models agree on the 1-quarter ahead predictability of the FX movements, they are actually pointing to different fundamentals, in terms of predictive content, such as the 1- and 3-month forward premiums as well as the U.S. money stock, while results from the DSGE model are clearer finding no predictability for the FX changes for any horizon. We should also mention the fact that forecasting particularly the nominal USD/GBP changes at both short and long horizon, has been proved by the literature one of the most difficult exercises among other bilateral exchange rates. Some characteristic papers are Chinn and Meese (1995), Chen and Mark (1996) and Engel, Mark and West (2008), where for the vast majority of the cases they consider, predictability for the nominal USD/GBP change couldn't be achieved by using the standard PPP, Monetary and Taylor-rule model fundamentals.

Empirical findings and results reported in this study denote the difficulty in outperforming the driftless random walk model in forecasting accuracy, while there is still progress to be made in convincingly resolve the Meese and Rogoff puzzle and bring the nominal exchange rates closer to the macroeconomic variables. Nevertheless, the BVAR model with the Minnesota priors was proved very competitive model and we reinforce the view of Carriero *et al.* (2009) that this model should be established as a benchmark exchange rate forecasting model. Hence we think that the nominal FX forecasting literature should choose an orientation to less sophisticated models exploiting the random walk behaviour of the nominal foreign exchange rates.

Appendix A

BVAR(p) model

The BVAR model that we consider, takes the form:

$$y_t = z_t \beta + \varepsilon_t \tag{A.1}$$

where y_t is an $m\times 1$ vector of $t = 1,\ldots,T$ time-series observations on the dependent variables, β is an $n\times 1$ state vector $\left[c', vec(B_1'),\ldots,vec(B_p')\right]'$ of parameters, $\varepsilon_t \sim N_m(0,\Sigma)$ where Σ is the covariance matrix of $m\times m$ dimensions and m is the number of variables. $z_t = I_m \otimes x_t = I_m \otimes (1, y_{t-1},\ldots, y_{t-p})'$ is an $m\times n$ matrix, $n = mk$ and $k = mp+1$. The conditional posterior for β can be obtained using the Normal distribution only, like:

$$\beta|\Sigma, y \sim N(\overline{B}, \overline{V}), \tag{A.2}$$

where the posterior variance $\overline{V} = (\underline{V}^{-1} + z'Vz)^{-1}$, with $V = \hat{\Sigma}^{-1} \otimes \mathrm{I}_m$ and $\hat{\Sigma}^{-1}$ is the OLS estimate of Σ. And the posterior mean for the parameters are given from $\overline{B} = \overline{V}(\underline{V}^{-1}\underline{B} + z'Vy)$. The posterior for the covariance of the VAR are obtained using the Inverse Wishart density, as:

$$\Sigma^{-1} \mid y, \beta \sim Wishart\left(\overline{v}, \overline{S}^{-1}\right) \tag{A.3}$$

where $\overline{S} = \underline{S} + (y - x\beta)'(y - x\beta)$ and $\overline{v} = T + \underline{v}$ with $\underline{S} = \underline{v} = 0$. Bayesian inference is obtained using the Gibbs sampler as an MCMC method.

Heteroscedastic TVP- BVAR(p) model

We set our priors as in Primiceri (2005), where a time-invariant VAR model of size τ = 40 is used to calibrate them. A normal prior for the coefficients and the log σ_t and the inverse Wishart and Gamma for the hyperparameteres Q, W, S_1 and S_2. Note that model assumes a block diagonal matrix for S to ensure the independency of the variables' parameters evolution.

Priors

$$\beta_0 \sim N(\hat{\beta}_{OLS}, 4\cdot V(\hat{\beta}_{OLS})),$$

$$A_0 \sim N(\hat{A}_{OLS}, 4\cdot V(\hat{A}_{OLS})),$$

$$\log \sigma_0 \sim N(\log \hat{\sigma}_{OLS}, I_m),$$

$$Q \sim IW(0.01^2 \cdot 40 \cdot V(\hat{\beta}_{OLS}), 40),$$

$$W \sim IG(0.001, 8),$$

$$S_1 \sim IW(0.1^2 \cdot 2 \cdot V(\hat{A}_{1,OLS}), 2),$$
$$S_2 \sim IW(0.1^2 \cdot 3 \cdot V(\hat{A}_{2,OLS}), 3),$$

where the variance of β_0 and A_0 are 4 times the variance of the OLS estimates and the log of the OLS estimates for the σ_0. The degrees of freedom of the inverse Wishart densities is equal to one plus the dimension of the matrices and 40 (size of the training sample) for the Q.

Posteriors

The Bayesian inference can be obtained sequentially using the Gibbs sampler as:

a) Draw samples for the β^T conditional on y^T, A^T, Σ^T, V using the Carter and Kohn (1994) algorithm which employs the Kalman filter along with a smoothing process, using the initial conditions as described above.

b) Sample A^T from the conditional density $p(A^T | y^T, \beta^T, \Sigma^T, V)$ using the Carter and Kohn (1994) algorithm which employs the Kalman filter along with a smoothing process. Further transformations are need since the posterior of A^T is Normal but non-linear. (see Primiceri's (2005) Appendix for more details).

c) In order to draw samples for Σ^T, further modifications are needed since the posterior now is non-Gaussian and non-linear. The innovations of the measurement equations are distributed as a log χ^2 and a mixture of their normal approximation is used as in Kim *et al.* (1998). Defining and sampling s^T which is a matrix indicator variables that governs the Gaussian approximations, the system now becomes normal and linear, and the conditional posterior of Σ^T is obtained.

d) Finally the posterior density for the diagonals of V conditional on y^T, A^T, Σ^T, B^T can be drawn from the Inverse Wishart and Inverse Gamma distributions. For draws from the $IW(\overline{V}, \overline{S})$ density:

$$\overline{S}^{Q^{-1}} = \left(\underline{S}^Q + \sum_{t=1}^{T} (\beta_t - \beta_{t-1})(\beta_t - \beta_{t-1})' \right)^{-1} \text{ and } \overline{V}^Q = t + \underline{V}^Q,$$
$$\overline{S}^{S_{1,2}^{-1}} = \left(\underline{S}^{S_{1,2}} + \sum_{t=1}^{T} (A_t - A_{t-1})(A_t - A_{t-1})' \right)^{-1} \text{ and } \overline{V}^{S_{1,2}} = t + 1 + size(S_{blocks}).$$

For draws for the W matrix from the $IG\left(\frac{\overline{u}}{2}, \frac{2}{\overline{\delta}}\right)$ density:

$$\bar{u} = \underline{u} + t - p - 1 \text{ and } \bar{\delta} = \underline{\delta} + \sum_{t=1}^{T} (\Delta \ln \sigma_t)^2 .$$

Priors for the sensitivity analysis

For the purposes of the sensitivity analysis we set non-informative priors as we did for the Homoscedastic TVP case. Hence the following priors are set:

$$\beta_0 \sim N(0, I_K),$$
$$A_0 \sim N(0, I_g),$$
$$\log \sigma_0 \sim N(0, I_m),$$
$$Q \sim IW(0.01^2 \cdot V(\beta_0), 40),$$
$$W \sim IG\left(\frac{0.01^2}{2}, \frac{1}{2}\right),$$
$$S_1 \sim IW(0.1^2 \cdot 2 \cdot V(A_0), 2),$$
$$S_2 \sim IW(0.1^2 \cdot 3 \cdot V(A_0), 3),$$

where $K = m + pm^2$ and $g = m(m-1)/2$. The sampling process and specifications remain the same except for W which is drawn from the $IG\left(\frac{\bar{u}}{2}, \frac{\bar{\delta}}{2}\right)$ density with $\bar{u} = \underline{u} + T - p - 1$ and $\bar{\delta} = \underline{\delta} + \sum_{t=1}^{T} (\varDelta \ln \sigma_t)^2$.

Dynamic Model Averaging and Selection

We closely follow the DMA model as developed by Raftery *et al.* (2010) and Koop and Korobilis (2012, 2013) in a Heteroscedastic TVP-AR specification. The state space model can be written as:[30]

$$y_t = z_t^{(k)} \theta_t^{(k)} + \varepsilon_t^{(k)} \tag{A.4}$$
$$\theta_t^{(k)} = \theta_{t-1}^{(k)} + \eta_t$$

where y_t is the log of exchange rate change, k = 1 ,….., K is the number of models, each using different set of predictors, $z_t^{(k)}$ is a matrix of predictors where each of these K models uses and $\theta_t^{(k)}$ collects the corresponding coefficients. Also $\varepsilon_t^{(k)} \sim N(0, H_t^{(k)})$ and $\eta_t^{(k)} \sim N(0, Q_t^{(k)})$.

[30] This is actually a TVP-ARX model (as in Ljung, 1987) which allows for both lags of the independent and exogenous variables to predict.

Following Korobilis' (2012) notations, let $\Theta_t = (\theta_t^{(1)'},....., \theta_t^{(k)'})'$, $y_t = (y_t ,....., y_t)'$ and $L_t \in \{1, 2,......, K\}$ indexing which model applies at every time period. First we need to specify the prior mean and variance for the parameters, which are based on the data (as in Raftery *et al.*, 2010). The prior mean has been set equal to zero $\left(\hat{\theta}_0^{(k)} = 0\right)$ and prior variance: $\Sigma_0^{(k)} = \text{diag}$ $(s_1^{2(k)},...., s_K^{2(k)})$, where $s_j^{2(k)} = \text{Var}(y_t) / \text{Var}(z_{t,j}^{(k)})$ and j = 2,…, *K*. As regards the prior of the single-model probability $\pi_{0|0} = 1/K$ implying that all models are initially equally weighted.

Posterior inference –Kalman filter

Below we present the modified Kalman filter taking into account the multi-model case that we face in the DMA, DMS models and the fact that we replace some components with their estimates. So given the priors (initial conditions) for the coefficients, the filter's prediction procedure, conditional on information up to *t* – 1, begins by specifying:

$$\theta_t^{(k)} \big| L_t = k, y^{t-1} \sim N(\hat{\theta}_{t-1}^{(k)}, \Sigma_{t|t-1}^{(k)}), \tag{A.5}$$

where $\Sigma_{t|t-1}^{(k)} = \Sigma_{t-1|t-1}^{(k)} + Q$. Raftery *et al.* (2010) employs the forgetting factor λ in order to approximate directly:

$$\Sigma_{t|t-1}^{(k)} = \frac{1}{\lambda} \Sigma_{t-1|t-1}^{(k)}$$

where $0 < \lambda \le 1$ where a value of 0.99 (using quarterly data) implies that the five-years ago observations will bear an 80% as much weight as the last quarter's observation. This factor also implies a smooth evolution of the parameters. Hence there is no need to simulate *Q* and the computation time is reduced significantly. What follows is the standard equations for the prediction errors and their conditional variance. And finally the updating equations for the coefficients and their covariance matrix conditional on information up to *t*. To be more specific:

$$\theta_t^{(k)} \big| L_t = k, y^t \sim N(\hat{\theta}_{t|t}^{(k)}, \Sigma_{t|t}^{(k)}) \tag{A.6}$$

where

$$\hat{\theta}_{t|t}^{(k)} = \hat{\theta}_{t|t-1} + KG * pe \tag{A.7}$$

and

$$\Sigma_{t|t}^{(k)} = \Sigma_{t|t-1}^{(k)} - KG * z_t * \Sigma_{t|t-1}^{(k)} \tag{A.8}$$

where *KG* is the Kalman gain and *pe* is the one-period ahead prediction error. When introducing the single-model probability in the Kalman filter, then the probability density of (A.5) is written:

$$p(\theta_t^{(k)}, L_t) = \sum_{k=1}^{K} p(\theta_t^{(k)} | L_t = k) \Pr(L_t = k) \quad \text{(A.9)}$$

where $\Pr(L_t = k) = \pi_{t|t-1,k}$ is the probability (weight) indicates the relevance-importance of the individual model k. This probability is calculated recursively using the forgetting factor α, such as:

$$\pi_{t|t-1,k} = \frac{\pi_{t-1|t-1,k}^{a}}{\sum_{1}^{K} \pi_{t-1|t-1,l}^{a}} . \quad \text{(A.10)}$$

Forgetting factor α has the same interpretation as λ. As regards the covariance matrix $H_t^{(k)}$, Koop and Korobilis (2012) uses an Exponentially Weighted Moving Average (EWMA) estimate:

$$\hat{H}_t^{(k)} = \sqrt{(1-\kappa)\sum_{j=1}^{t} \kappa^{j-1} * pe^2} . \quad \text{(A.11)}$$

<u>DSGE log-linearized model</u> [31]

The partially forward-looking New Keynesian IS curve is derived by log-linearizing eq. A.30,

$$\hat{c}_t = \frac{1}{1+\zeta} E_t \hat{c}_{t+1} + \frac{\zeta}{1+\zeta} \hat{c}_{t-1} - \frac{1-\zeta}{\sigma(1+\zeta)} (\hat{i}_t - E_t \hat{\pi}_{t+1}) + \mu_t^c \quad \text{(A.12)}$$

where c_t is the domestic consumption, π_t is the inflation rate, ζ is the external habit formation coefficient, and σ is the risk aversion; μ_t^c is the consumption demand shock (assets risk premium) which follows an AR(1) process $\mu_t^c = \rho_c \mu_{t-1}^c + \varepsilon_t^c$ and $\varepsilon_t^c \sim i.i.d.\, N(0, \sigma_c^2)$ while $\hat{i}_t - E_t \hat{\pi}_{t+1}$ is the real interest rate. This equation links the current domestic consumption with the expected consumption and inflation and the one-period lagged consumption, while E_t denotes the expectations of a given variable formed at time *t*.

The equation which relates the domestic output with consumption comes from goods markets clearing condition and is given by log-linearizing approximately the goods market clearing conditions:

[31] The hat above each variable denotes the log-deviation of this variables from its steady-state value, while the bar over them denotes their steady-state. For instance, $\hat{y}_t = \log y_t - \log \bar{y}_t$.

$$\hat{y}_t = (1-\gamma)\hat{c}_t + \eta\gamma(2-\gamma)\hat{s}_t + \gamma\hat{y}_t^* + \eta\gamma\hat{\psi}_t \tag{A.13}$$

where γ is the import share $(0 \le \gamma < 1)$, η is the intertemporal elasticity of substitution between foreign and domestic products. In addition, $\hat{y}_t$ is the domestic output, * denotes the foreign county's variables and foreign output is assumed to follow an AR(1) as $\hat{y}_t^* = \rho_{y*}\hat{y}_{t-1}^* + \varepsilon_t^{y*}$ and $\varepsilon_t^{y*} \sim i.i.d.\,N(0,\sigma_{y*}^2)$. Also $\hat{s}_t$ is the terms of trade with $\hat{s}_t = \hat{p}_{f,t} - \hat{p}_{h,t}$ and $\hat{\psi}_t = \hat{e}_t + \hat{p}_t^* - \hat{p}_{f,t}$ denotes the deviation from the law of one price in the short-run, while $\hat{p}_t^*$ is the world price of the imported goods, $\hat{p}_{f,t}$ is the home currency price of the imports and $\hat{e}_t$ is the nominal exchange rate. Both equations can be derived by log-linearizing equations A.49a and A.43.

The partially forward-looking domestic inflation Phillips-curve is given by combining equations A.39 and A.42:

$$\hat{\pi}_{h,t} = \frac{\beta}{1+\beta\phi_h}E_t\hat{\pi}_{h,t+1} + \frac{\phi_h}{1+\beta\phi_h}\hat{\pi}_{h,t-1} + \frac{(1-\theta_h)(1-\beta\theta_h)}{\theta_h(1-\beta\phi_h)}\hat{k}_t + \mu_t^h \tag{A.14}$$

where $\hat{\pi}_{h,t}$ is the home goods price inflation, β is the discount factor, ϕ_h describes the degree to which prices are indexed to the previous period's price inflation, θ_h is the Calvo-type probability describing producers that do not adjust their prices, $\hat{k}_t$ is the firm's marginal cost and defined as $\hat{k}_t = \hat{w}_t - \hat{z}_t + \gamma\hat{s}_t$ (derived from the log-linear eq. A.36 in terms of labour productivity and terms of trade) where $\hat{w}_t$ is the real wage rate, $\hat{z}_t$ is the labour productivity which follows an AR(1) process as $\hat{z}_t = \rho_z\hat{z}_{t-1} + \varepsilon_t^z$ and $\varepsilon_t^z \sim i.i.d.\,N(0,\sigma_z^2)$.[32] Same AR(1) process is assumed for the cost-push shock μ_t^h. The UK economy is modelled as a closed-form version of the domestic economy. The foreign goods inflation is similar to the domestic producers Phillips curve:

$$\hat{\pi}_{f,t} = \frac{\beta}{1+\beta\phi_f}E_t\hat{\pi}_{h,t+1} + \frac{\phi_f}{1+\beta\phi_f}\hat{\pi}_{f,t-1} + \frac{(1-\theta_f)(1-\beta\theta_f)}{\theta_f(1-\beta\phi_f)}\hat{\psi}_t + \mu_t^f \tag{A.15}$$

where $\hat{\pi}_{f,t}$ is the domestic goods price inflation, ϕ_f has the same interpretation as ϕ_h, and μ_t^f is an exogenous cost-push shock following an AR(1) process, added in the Phillips curve (as in

[32] It should be noted that from log-linearized equations A.25 and A.49a, it can be derived: $p_t = p_{h,t} = \gamma s_t$.

Justiniano and Preston (2010) p.101), capturing the mark-up fluctuations. The wage-inflation Phillips-curve type equation is given by combining equations (A.31) and (A.32):

$$\hat{\pi}_{w,t} - \phi_w \hat{\pi}_{w,t-1} = \beta E_t \hat{\pi}_{w,t+1} - \phi_w \beta \hat{\pi}_{w,t} + \frac{(1-\theta_w)(1-\theta_w\beta)}{\theta_w(1+\theta\Xi)}(\hat{m}_t - \hat{w}_t) + \mu_t^w \tag{A.16}$$

where $\hat{\pi}_{w,t}$ is the nominal wage-inflation and equals to $w_t - w_{t-1} + \pi_t$, π_t is the weighted sum of prices for both domestically produced and foreign goods and defined as $\pi_t = (1-\gamma)\pi_{h,t} + \gamma\pi_{f,t}$. Also ϕ_w describes the degree to which the nominal wage inflation is indexed to the price inflation, θ_w is the Calvo-type probability describing households that do not adjust their wage, θ is the inverse of labour supply elasticity, Ξ is the elasticity of substitution between households' labour services, $\hat{m}_t$ is the marginal rate of substitution which is defined as $\hat{m}_t = \sigma(\hat{c}_t - \zeta\hat{c}_{t-1})/(1-\zeta) + \theta(y_t - z_t)$ and μ_t^w is the exogenous mark-up shock which follows an AR(1) process as well. Three more characteristic equations need to specify so we close the model, the UIP modified conditions, the real exchange rate and the monetary policy rule. The UIP as modified by Adolfson *et al.* (2008) is used, which takes into account the forward premium puzzle allowing for negative correlation between the expected depreciation rate and the risk premium. The log-linearized UIP is given by:

$$i_t - i_t^* = (1-\phi)E_t d_{t+1} - \phi d_t - \chi a_t + \mu_t^d \tag{A.17}$$

where i_t is the nominal interest rate, foreign interest rate i_t^* follows the univariate AR(1) process, $E_t \hat{d}_{t+1}$ is the expected depreciation rate, χ and ϕ are elasticity parameters, the nominal depreciation rate is defined as $\hat{d}_t = \hat{e}_t - \hat{e}_{t-1}$, a_t is the U.S. net asset position and defined as $\hat{a}_t = (1/\beta)\hat{a}_{t-1} + \hat{y}_t - \hat{c}_t - \gamma(\hat{s}_t + \hat{\psi}_t)$ (Schmitt-Grohe and Uribe, 2003) and the last component is the time-varying shock to the risk premium μ_t^d which follows the AR(1) process. The real exchange rate is defined as $\hat{q}_t = \hat{e}_t + \hat{p}_t^* - \hat{p}_t$, and if we time differentiate it we can obtain $\hat{q}_t - \hat{q}_{t-1} = \hat{d}_t + \hat{\pi}_t^* - \hat{\pi}_t$, where $\hat{\pi}_t^*$ follows an AR(1) process . Also an equation which relates the real exchange rate with the terms of trade and the deviation from the law of one price is given by $\hat{q}_t = (1-\gamma)\hat{s}_t + \hat{\psi}_t$. The last log-linearized equation is the Taylor-rule which is given by:

$$\hat{i}_t = \rho\hat{i}_{t-1} + (1+\rho)(\lambda_\pi E\hat{\pi}_{t+1} + \lambda_y \Delta\hat{y}_t + \lambda_d \hat{d}_t) + \mu_t^i \tag{A.18}$$

where ρ is the smoothing parameter, λ_π , λ_y , λ_d are the relative weights of the expected inflation, real output and depreciation rate respectively, while μ_t^i is the monetary policy shock which follows an AR(1) process as well.

DSGE estimation

Following the recent literature, we use Bayesian methods with prior assumptions from the literature and allowing for the data likelihood to estimate the parameters of the DSGE system. Many empirical works of the last decade have focus on this kind of methods, (see for example, Smets and Wouters, 2004; An and Schorfheide, 2007a,b; Justiniano and Preston, 2010 and Marcellino and Rychalovska, 2014), taking the observed data as given and treating the unknown parameters as random variables. Following Villemot (2011) the linear rational expectations model can be written as:

$$E_t\left\{f(y_{t+1}, y_t, y_{t-1}, u_t)\right\} = 0, \qquad \text{(A.19)}$$

or

$$A(E_t y_{t+1}) + By_t + Gy_{t-1} + u_t = 0, \qquad \text{(A.20)}$$

where y_t is the vector of our endogenous variables, $u_t \sim i.i.d.\ N(0, H)$ collects al the exogenous stochastic shocks *A*, *B* and *G* collect all the deep parameters of the DSGE system. We can also define the vector Ψ which contain all the parameters and shocks (*A*, *B*, *G* and *H*). The solution to the system is given by the policy function, using the Blanchard and Kahn (1980) method, which relates the current state of the variables with the past one and the current shocks, such as:

$$y_t = Ry_{t-1} + Qu_t\ . \qquad \text{(A.21)}$$

The above equation can be used as the transition equation while the measurement equation can take the following form:

$$y_t^* = M\bar{y} + M\hat{y} + \varepsilon_t, \qquad \text{(A.22)}$$

where y_t^* are the observables, $\bar{y}$ is the steady state vector, $\hat{y}$ is a vector containing the deviations of the variables from their steady state and ε_t an error term. Both equations (A.21) and (A.22) represent the state space form of our DSGE model and the likelihood function can be obtained using the Kalman filter. The posterior kernel of the structural parameters can be obtained by combining the likelihood function with the prior distributions. Still the posterior is non-linear and a complicated density while an MCMC simulation method such as the Metropolis-Hasting algorithm is required. For this empirical work we use the Dynare version 4.4.3 software which implements the above estimation procedure.[33] As regards the priors, we use the ones that Smets and Wouters (2007) use examining the U.S. economy, and also from Justiniano and Preston (2010). We calibrate only three parameters, the discount factor β which is set equal to 0.99 implying a 4% riskless annual interest rate at the steady state, γ is set equal to 0.10 which is the average imports-to-GDP ratio over our sample period for the U.S., and Ξ equals to 6 as in Alpanda *et al.* (2011).[34]

DSGE model's micro-foundations

The simple open economy model is based on Monacelli (2005), Gali and Monacelli (2005), Justiniano and Preston (2010), Steinbach *et al.* (2009) and Alpanda *et al.* (2011). The U.S. represents the domestic economy and U.K. the foreign one.

Households and optimal wage

Domestic economy consists of infinitely-lived households (followed by an index *i*, where $i \in [0,1]$) and consuming both domestically produced $(C_{h,t})$ and imported goods $(C_{f,t})$, where the composite consumption index is given by:

$$C_t \equiv \left[(1-\gamma)^{\frac{1}{\eta}} C_{h,t}^{\frac{\eta-1}{\eta}} + \gamma^{\frac{1}{\eta}} C_{f,t}^{\frac{\eta-1}{\eta}} \right]^{\frac{\eta}{\eta-1}}, \qquad (A.23)$$

where γ is the imports share, taking values [0,1), and $\eta > 0$ measuring the intertemporal elasticity of substitution between domestic and foreign imported goods. Households allocate their expenditures optimally between these goods according to:

$$C_{h,t} = (1-\gamma)\left[\frac{P_{h,t}}{P_t}\right]^{-\eta} C_t \quad \text{and} \quad C_{f,t} = \gamma\left[\frac{P_{f,t}}{P_t}\right]^{-\eta} C_t \qquad (A.24)$$

where $P_{h,t}$ and $P_{f,t}$ are the prices for the home and foreign products respectively, while the consumer price index P_t is given by:

[33] For more details, see the manual (Adjemian *et al.*, 2011) following this link (http://www.dynare.org/documentation-and-support) and An and Schorfheide (2007a).
[34] We would like to thank Dr. Sami Alpanda for sharing Alpanda's *et al.* (2011) Dynare code.

$$P_t = \left[(1-\gamma)P_{h,t}^{1-\eta} + \gamma P_{f,t}^{1-\eta}\right]^{\frac{1}{1-\eta}} \quad . \tag{A.25}$$

Monopolistically competitive households, supply the economy-wide labour market, while the labour demand function is given by:

$$N_t(i) = \left[\frac{W_t(i)}{W_t}\right]^{-\Xi} N_t \tag{A.26}$$

where Ξ is the labour demand elasticity and greater than one and constant across workers, while N_t is the per capita employment and aggregate wage index, W_t is given by the following equation:

$$W_t = \left[\int_0^1 W_t(i)^{1-\Xi} di\right]^{\frac{1}{1-\Xi}} . \tag{A.27}$$

Each household in every period maximises the following utility function:

$$E_0 \sum_{t=0}^{\infty} \beta^t \left[\frac{(C_t - H_t)^{1-\sigma}}{1-\sigma} - \frac{N_t(i)^{1+\theta}}{1+\theta}\right] , \tag{A.28}$$

where β is the discount factor, H_t is the external habit formation assuming that consumption in every period is affected by the previous period consumption and given by $H_t = \zeta C_{t-1}$, σ is the inverse elasticity of intertemporal substitution, while θ denotes the labour supply elasticity. Optimization is obtained subject to the period budget constraint:

$$P_{h,t}C_{h,t} + P_{f,t}C_{f,t} + E_t\{Q_{t,t+1} D_{t+1}\} \leq D_t + W_t N_t \quad , \tag{A.29}$$

where D_{t+1} is the portfolio of assets maturing in a period ahead, $Q_{t,t+1}$ is the discount factor and W_t is the nominal wage. Maximizing the utility function subject to the budget constraint, we obtain the standard consumption Euler equation:

$$Q_{t,t+1} = \beta E_t \left\{\frac{U_{c,t+1}}{U_{c,t}} \frac{P_t}{P_{t+1}}\right\}, \tag{A.30}$$

where $U_{c,t} = (C_t - \zeta C_{t-1})^{-\sigma}$. Following Smets and Wouters (2007), $Q_{t,t+1} = (\mu_t^d I_t)^{-1}$ where μ_t^d is the households' assets risk premium and I_t is the assets' nominal rate of return. Steinbach *et al.* (2009) derive the optimal wage-setting rule by assuming; i) workers have the right to set their wages in a Calvo (1983) style, where θ_w represents those who do not reset their wage

(Erceg *et al.*, 2000); ii) those who do not eventually reset their wage in the current period, they can index it to the previous period's price inflation Π_{t-1} (Rabanal and Rubio-Ramirez, 2005). Hence they derive the f.o.c. for the labour supply of the households as:

$$E_t \sum_{k=0}^{\infty} (\beta\theta_w)^k \left\{ \left[\frac{\widetilde{W}_t}{P_{t+k}} \Pi_{t+k-1}^{\phi_w} (C_{t+k} - H_{t+k})^{-\sigma} - (1+\mu^w) N_{t+k}^{\theta} \right] N_{t+k} \right\} = 0, \qquad \text{(A.31)}$$

where $\widetilde{W}_t$ being the optimal reset wage, α controls the indexation degree to the lagged inflation, and $(1+\mu^w)$ is the wage markup. Hence combining equation (A.27) with (A.31) and applying the law of large numbers, they obtain:

$$W_t = \left[\theta_w (W_{t-1}\Pi_{t-1}^{\phi_w})^{1-\Xi} + (1-\theta_w)\widetilde{W}_t^{1-\Xi} \right]^{\frac{1}{1-\Xi}}. \qquad \text{(A.32)}$$

Steinbach *et al.* (2009) describe the domestic production process is two stages. The first stage assumes monopolistically competitive firms indexed by j where $j \in [0,1]$ producing intermediate differentiated goods and setting prices is Calvo-style (Gali and Monacelli, 2005). At the second stage the perfectly competitive final producer will combine the differentiated foods and produce the final good.

Intermediate goods Producers, Technology and Price

Each domestic firm produces $Y_t(j)$ goods with a production function:

$$Y_t(j) = Z_t N_t(j) \qquad \text{(A.33)}$$

where $z_t = \log(Z_t)$ and follows an AR(1) process. The labour input for each *j* firm is given by the composite function:

$$N_t(j) = \left[\int_0^1 N_t(i)^{1-\frac{1}{\Xi}} di \right]^{\frac{\Xi}{\Xi-1}}, \qquad \text{(A.34)}$$

and the total nominal cost function:

$$TC_t^n = W_t N_t(j). \qquad \text{(A.35)}$$

Combining equations (A.33) with (A.35) yields the marginal cost function in terms of real wage, as:

$$K_t = \frac{W}{Z_t P_{h,t}} \quad . \tag{A.36}$$

As mentioned before, intermediate firms set their prices as in Calvo (1983) with θ_h being the probability for each firm which does not reset its price. In addition, is assumed that prices for the home country are indexed to the last period's inflation (Smets and Wouters, 2002). According to Justiniano and Preston (2010) firms will select the optimal reset price $\widetilde{P}_{h,t}$ by solving their profit maximization problem given by the following expected present discounted profits:

$$E_t \sum_{k=0}^{\infty} \theta_h^k Q_{t,t+k} Y_{t+k}(j) \left[\widetilde{P}_{h,t} \Pi_{h,t+k-1}^{\delta} - K_{t+k} P_{h,t+k} \right], \tag{A.37}$$

where $P_{h,t}$ is the Dixit-Stiglitz aggregate price index. Subject to the demand curve for intermediate goods:

$$Y_t(j) = \left[\frac{P_{h,t}(j) \Pi_{h,t-1}^{\phi_h}}{P_{h,t}} \right]^{-\xi_h} Y_t, \tag{A.38}$$

where δ is the indexation degree to the past inflation and Y_t is the market clearing condition. Thus, maximizing equation (A.37) implies the f.o.c.:

$$E_t \sum_{k=0}^{\infty} \theta_h^k Q_{t,t+k} Y_{t+k}(j) \left[\widetilde{P}_{h,t} \Pi_{h,t+k-1}^{\phi_h} - \left(\frac{\theta_h}{\theta_h - 1} \right) K_{t+k} P_{h,t+k} \right] = 0 \, . \tag{A.39}$$

See Justiniano and Preston (2010) p.98 for more details.

Final goods producers and prices

Producers use the intermediate goods as input and compose the final ones, while their technology production function is given by:

$$Y_t = \left[\int_0^1 Y_t(j)^{\frac{\xi_h - 1}{\xi_h}} dj \right]^{\frac{\xi_h}{\xi_h - 1}}, \tag{A.40}$$

and the price index:

$$P_{h,t} = \left[\int_0^1 P_{h,t}(j)^{1-\xi_h} dj \right]^{\frac{1}{1-\xi_h}} . \tag{A.41}$$

Bringing the Calvo-style price setting and the price indexation behaviour into equation (A.41), the following Dixit-Stiglitz aggregate price index is derived as in Justiniano and Preston (2010):

$$P_{h,t} = \left[\theta_h (P_{h,t-1}\Pi_{h,t}^{\phi_h})^{1-\xi_w} + (1-\theta_h)\tilde{P}_{h,t}^{1-\xi_h}\right]^{\frac{1}{1-\xi_h}}, \tag{A.42}$$

where $\tilde{P}_{h,t}$ is the optimal reset price.

International trade and Incomplete exchange rate pass-through

As discussed and explained before, the existence of a deviation from the law of one price (l.o.p.) in the short-run and the achievement of the complete exchange rate pass-through in the long run, can be assumed and remains to model it. Hence, the deviation from the (l.o.p.) is defined by:

$$\Psi_{f,t} = \varepsilon_t \frac{P_t^*}{P_{f,t}} \tag{A.43}$$

where $\Psi_{f,t}$ captures the deviation, ε_t is the current nominal exchange rate (home price for a unit of a foreign currency), P_t^* is the world-market price paid by the importer and $P_{f,t}$ is the home currency price that retailers charge domestically for the imported goods. Similarly, the retailers now face their own profit maximization problem and need to find the optimal price $\tilde{P}_{f,t}(j)$, assuming a Calvo-type behaviour once again. They seek to maximize the following objective:

$$E_t \sum_{k=0}^{\infty} \theta_f{}^k Q_{t,t+k} C_{f,t+k}(j) \left[\tilde{P}_{f,t}(j)\Pi_{f,t+k-1}^{\delta} - \varepsilon_{t+k} P_{t+k}^*\right] \tag{A.44}$$

Subject to the demand curve that they face: $$C_{f,t}(j) = \left[\frac{\tilde{P}_{f,t}(j)}{P_{f,t}}\right]^{-\xi_f} C_{f,t} \tag{A.45}$$

Hence the optimal solution to their problem is given by the f.o.c.:

$$E_t \sum_{k=0}^{\infty} \theta_f{}^k Q_{t,t+k} \left[\tilde{P}_{f,t}(j)\Pi_{f,t+k-1}^{\delta} - \left(\frac{\theta_h}{\theta_h - 1}\right)\varepsilon_{t+k} P_{t+k}^*\right] = 0. \tag{A.46}$$

The price index for the imported goods taking into account the price-setting behaviour:

$$P_{f,t} = \left[\int_0^1 P_{f,t}(j)^{1-\xi_f} dj\right]^{\frac{1}{1-\xi_f}}, \tag{A.47}$$

and applying the law of large numbers the overall price index:

$$P_{f,t} = \left[\theta_f P_{f,t-1}^{1-\xi_f} + (1-\theta_f)\widetilde{P}_{f,t}^{1-\xi_f}\right]^{\frac{1}{1-\xi_f}} . \tag{A. 48}$$

Note that retailers who do not reset their price they do not index them to the lagged inflation as well. Next the terms of trade and the real exchange rate are defined respectively as:

$$S_t = \frac{P_{f,t}}{P_{h,t}} \qquad \text{and} \qquad Q_t = \varepsilon_t \frac{P_t^*}{P_t} . \tag{A.49a, b}$$

The goods market clearing conditions implies: $Y_t = C_{h,t} + C_{h,t}^*$ where $C_{h,t}^*$ denotes the exports of the domestically produced goods. The UIP condition for the nominal interest rates is given by its log-linearized version as in Adolfson *et al.* (2008), with the risk premium components capturing the forward premium puzzle.

Table A1. Prior selection and estimated posterior means of the DSGE parameters.

Structural Parameters	Prior density	Posterior mean
ζ Habit in consumption	B (0.7 , 0.15)	0.5861
σ Inverse of intertemporal substitution elasticity	G (1.5 , 0.37)	1.5507
θ Inverse of labour supply elasticity	G (2 , 0.75)	1.5576
η Substitution elasticity between home and foreign	G (1.5 , 0.75)	1.0084
χ Debt elasticity of risk premium	N (0.01 , 0.001)	0.0103
φ UIP parameter	B (0.1 , 0.2)	0.0240
θ_h Calvo probability: home good price	B (0.5 , 0.15)	0.9069
θ_f Calvo probability: foreign good price	B (0.5 , 0.15)	0.4942
θ_w Calvo probability: wage	B (0.5 , 0.15)	0.6308
φ_h Indexation: home good price	B (0.7 , 0.15)	0.6793
φ_f Indexation: foreign good price	B (0.7 , 0.15)	0.5414
φ_w Indexation: wage	B (0.7 , 0.15)	0.8194
ρ Taylor rules: smoothing	B (0.7 , 0.15)	0.6288
λ_π Taylor rule: inflation	G (0.5 , 0.25)	1.7135
λ_y Taylor rule: output growth	G (0.25 , 0.1)	0.1542
λ_d Taylor rule: depreciation	G (0.12 , 0.05)	0.1997

Persistence parameters	Prior density	Posterior mean
Productivity shock	B (0.5 , 0.2)	0.9242
Consumption demand shock	B (0.5 , 0.2)	0.8110
Home good cost-push shock	B (0.5 , 0.2)	0.0224
Foreign good cost-push shock	B (0.5 , 0.2)	0.8988

Wage cost push shock	B (0.5 , 0.2)	0.7250
Depreciation shock	B (0.5 , 0.2)	0.4425
Monetary policy shock	B (0.5 , 0.2)	0.3503
Foreign output shock	B (0.5 , 0.2)	0.7120
Foreign inflation shock	B (0.5 , 0.2)	0.7331
Foreign interest rate shock	B (0.5 , 0.2)	0.3979
Standard deviations of shocks	**Prior density**	**Posterior mean**
Productivity shock	IG (0.5%, inf)	0.73 %
Consumption demand shock	IG (0.5%, inf)	0.24 %
Home good cost-push shock	IG (0.5%, inf)	0.21 %
Foreign good cost-push shock	IG (0.5%, inf)	5.45 %
Wage cost push shock	IG (0.5%, inf)	0.30 %
Depreciation shock	IG (0.5%, inf)	0.40 %
Monetary policy shock	IG (0.5%, inf)	1.06 %
Foreign output shock	IG (0.5%, inf)	1.09 %
Foreign inflation shock	IG (0.5%, inf)	0.92 %
Foreign interest rate shock	IG (0.5%, inf)	2.03 %

Notes: These posterior estimates have been obtained using the most recent vintage of our data set and Bayesian methods as explained above. Also, priors are mainly from Smets and Wouters (2007).

Table A2. Macroeconomic variables' transformation for the DMA, DMS analysis.

	Macroeconomic variables	*Transformation code*
1.	1-month Forward (USD/GBP) exchange rate premium.	1
2.	3-month Forward (USD/GBP) exchange rate premium.	1
3.	6-month Forward (USD/GBP) exchange rate premium.	1
4.	12-month Forward (USD/GBP) exchange rate premium.	1
5.	U.S. Real GDP (seasonally adjusted).	2
6.	U.K. Real GDP (seasonally adjusted).	2
7.	U.S. Output gap (HP-filtered).	1
8.	U.K. Output gap (HP-filtered).	1
9.	U.S. Money supply (M1).	2
10.	U.K. Money supply (M4).	2
11.	Real (USD/GBP) exchange rate.	2
12.	U.S. Price Inflation (annualised).	1
13.	U.K. Price Inflation (annualised).	1
14.	U.S. 10-year maturity Government Bond rates.	1
15.	U.K. 10-year maturity Government Bond rates.	1

Notes: Transformation codes are as follow: (1)-variable in logarithm, (2)-second difference of the variable in logarithm.

References

1. Adjemian, S., Bastani, H., Juillard, M., Karamé, F., Mihoubi, F., Perendia, G., Pfeifer, J., Ratto M. and Villemot, S. (2011). Dynare: Reference Manual, Version 4. *Dynare Working Papers*, 1, CEPREMAP.
2. Adolfson, M. Laséen, S., Lindé, J. and Villani, M. (2008). Evaluating an estimated new Keynesian small open economy model. *Journal of Economic Dynamics and Control*, 32(8):2690-2721.
3. Alpanda, S., Kotzé, K., and Woglom, G. (2011). Forecasting Performance of an Estimated DSGE modes for the South African Economy. *South African Journal of Economics,* 79(1):50-67.
4. Alquist, R. and Chinn, M.D. (2008). Conventional and unconventional approaches to exchange rate modelling and assessment. *International Journal of Finance & Economics,* 13(1): 2-13.
5. An, S., and Schorfheide, F. (2007a). Bayesian Analysis of DSGE Models. *Econometric Reviews,* 26(2-4):113-172.
6. An, S., and Schorfheide, F. (2007b). Bayesian Analysis of DSGE Models - Rejoinder. *Econometric Reviews*, 26(2-4):211-219.
7. Bacchetta, and Van Wincoop. (2013). On the unstable relationship between exchange rates and macroeconomic fundamentals. *Journal of International Economics,* 91(1):18-26.
8. Bacchetta, P. and Van Wincoop, E. (2004). A scapegoat model of exchange-rate fluctuations. *American Economic Review,* 94(2):114-118.
9. Baillie, R.T. and Kiliç, R. (2006). Do asymmetric and nonlinear adjustments explain the forward premium anomaly? *Journal of International Money and Finance*, *25*(1), 22-47.
10. Bekaert, G. and Hodrick, R.J. (1993). On biases in the measurement of foreign exchange risk premiums. *Journal of International Money and Finance,* 12(2):115-138.
11. Bilson, J. (1978). The monetary approach to the exchange rate-some empirical evidence. *IMF Staff Papers,* 25:48-75.
12. Bilson, J.F. (1981). The "Speculative Efficiency" Hypothesis. *The Journal of Business (pre-1986),* 54(3), 435.
13. Blanchard, O. and Kahn, C. (1980). The Solution of Linear Difference Models under Rational Expectations. *Econometrica,* 48(5):1305-1311.
14. Boughton, J.M. (1988). Exchange Rates and the Term Structure of Interest Rates. *Staff Papers - International Monetary Fund*, *35*(1):36.
15. Byrne, J.P., Korobilis, D. and Ribeiro, P.J. (2014). On the Sources of Uncertainty in Exchange Rate Predictability. *MPRA Paper 58956*, University Library of Munich, Germany.
16. 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.
17. Calvo, G. (1983). Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics*, 12(3):383-398.
18. Carriero, A., Kapetanios, G. and Marcellino, M. (2009). Forecasting exchange rates with a large Bayesian VAR. *International Journal of Forecasting,* 25(2):400-417.

19. Carter, C. and Kohn R. (1994). On a Gibbs sampling for state space models. *Biometrica*, 81:541-553.
20. Cassel, G. (1918). Abnormal deviations in international exchanges. *The Economic Journal*, 28:413-415.
21. Chen, J. and Mark, N.C. (1996). Alternative long-horizon exchange-rate predictors. *International Journal of Finance and Economics*, 1:229–250.
22. Cheung, Y.W., Chinn, M.D. and Pascual, A.G. (2005). Empirical exchange rate models of the nineties: Are any fit to survive?. *Journal of International Money and Finance,* 24(7):1150-1175.
23. Chinn, M. (1991). Some linear and nonlinear thoughts on exchange rates. *Journal of International Money and Finance,* 10(2):214-230.
24. Chinn, M.D, and Meredith, G., (2004). Monetary Policy and Long-Horizon Uncovered Interest Parity. *IMF Staff Papers,* 51(3):409-430.
25. Chinn, M.D. (2006). The (partial) rehabilitation of interest rate parity in the floating rate era: Longer horizons, alternative expectations, and emerging markets. *Journal of International Money and Finance,* 25(1):7-21.
26. Chinn, M.D. and Meese, R.A. (1995). Banking on currency forecasts: How predictable is change in money?. *Journal of International Economics,* 38(1):161-178.
27. Clarida, R., Gali, J. and Gertler, M. (1998). Monetary rules in Practice: Some International evidence, *European Economic Review*, 42:1033-1067.
28. Clarida, R.H., and Taylor, M.P. (1997). The Term Structure of Forward Exchange Premiums and the Forecastability of Spot Exchange Rates: Correcting the Errors. *Review of Economics and Statistics*, 79(3):353-361.
29. Clarida, R.H., Sarno, L., Taylor, M.P. and Valente, G. (2003). The out-of-sample success of term structure models as exchange rate predictors: A step beyond. *Journal of International Economics,* 60(1):61-83.
30. Clark, T.E., and West, K.D. (2006). Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis. *Journal of Econometrics*, 135(1):155-186.
31. Clark, T.E., and West, K.D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1):291-311.
32. Clausen, J. and Meier, C.-P. (2005). Did the Bundesbank Follow a Taylor Rule? An Analysis Based on Real-Time Data. *Swiss Journal of Economics and Statistics*. 127:213-246.
33. Clements, M.P and Galvao, A.B. (2013). Real-time forecasting of inflation and output growth with autoregressive models in the presence of data revisions. *Journal of Applied Econometrics*, 28(3):458-477.
34. Clements, M.P, (2012). Forecasting U.S. Output Growth with Non-Linear Models in the Presence of Data Uncertainty. *Studies in Nonlinear Dynamics & Econometrics,* 16(1):1-25.
35. Clements, M.P, (2015). Assessing Macro Uncertainty In Real-Time When Data Are Subject To Revision. *ICMA Centre Discussion Papers in Finance*, Henley Business School, University of Reading.

36. Clements, M.P, Franses, P.H. and Swanson, N.R. (2004). Forecasting economic and financial time-series with non-linear models. *International Journal of Forecasting*, 20:169-183.
37. Cogley, T. and Sargent, T. (2005). Drifts and volatilities: Monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, 8(2):262-302.
38. Croushore D. and Stark T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics*, 105:111-130.
39. Croushore, D. (2006). Forecasting with real-time macroeconomic data. In G. Elliott, C. Granger and A. Timmerman neds., *Handbook of Economic Forecasting, Volume I*, Elsevier, Horth-Holland, 961:982.
40. D'Agostino, A., Gambetti, L., and Giannone, D. (2013). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics*, 28(1):82-101.
41. Doan, T., Litterman, R.B. and Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, 3:1-144.
42. Dornbusch, R. (1985). Purchasing Power Parity. *NBER Working Papers 1591*.
43. Engel, C. and Hamilton, J.D. (1990). Long Swings in the Dollar: Are They in the Data and Do Markets Know It?. *American Economic Review*, 80(4):689-713.
44. Engel, C. and West, K.D. (2005). Exchange Rate and Fundamentals. *Journal of Political Economy*, 113:485-517.
45. Engel, C., and West, K. (2006). Taylor Rules and the Deutschmark: Dollar Real Exchange Rate. *Journal of Money, Credit and Banking*, 38(5):1175-1194.
46. Engel, C., Mark, N.C. and West, K.D. (2008). Exchange Rate Models Are Not As Bad As You Think. *NBER Macroeconomics Annual 2007*, 22:381-441.
47. Fama, E. (1984). Forward and spot exchange rates. *Journal of Monetary Economics*, 14(3):319-338.
48. Fisher, I. (1896). Appreciation and Interest. *Publications of the American Economic Association*, First Series, 11(4): 1–110 [331– 442], and as Appreciation and Interest, New York: Macmillan, 1896.
49. Frankel, J.A. (1984). Tests of Monetary and Portfolio Balance Models of Exchange Rate Determination. NBER Chapters, in: Exchange Rate Theory and Practice, 239:260.
50. Fratzscher, M., Rime D., Sarno, L. and Zinna, G. (2015). The scapegoat theory of exchange rates: The first tests. *Journal of Monetary Economics*, 70:1-21.
51. Frenkel, J. (1976). A Monetary Approach to the Exchange Rate: Doctrinal Aspects and Empirical Evidence. *The Scandinavian Journal of Economics*,78(2):200-224.
52. Froot, K. and Thaler, R. (1990). Anomalies: Foreign Exchange. *The Journal of Economic Perspectives*, 4(3):179-192.
53. Galí, J. and Monacelli, T. (2005). Monetary Policy and Exchange Rate Volatility in a Small Open Economy. *Review of Economic Studies*, 72(3):707-734.
54. Giacomini, R., and Rossi, B. (2006). How Stable is the Forecasting Performance of the Yield Curve for Output Growth?. *Oxford Bulletin of Economics and Statistics*, 68:783-795.
55. Giacomini, R., and Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, 25(4):595-620.

56. Goyal, A., and I. Welch. (2003). Predicting the Equity Premium with Dividend Ratios. Management *Science*, 49:639-654.
57. Hodrick, R. (1987). The empirical evidence on the efficiency of forward and futures foreign exchange markets. *Fundamentals of pure and applied economics*, 24,Chur, Switzerland New York: Harwood Academic.
58. Inoue, A. and Rossi, B. (2012). Out-of-Sample Forecast Tests Robust to the Choice of Window Size. *Journal of Business & Economic Statistics,* 30(3):432-453.
59. Jones, C. and Kulish, M. (2013). Long-term interest rates, risk premia and unconventional monetary policy. *Journal of Economic Dynamics and Control,*37(12):2547-2561.
60. Justiniano, A. and Preston, B. (2010). Monetary policy and uncertainty in an empirical small open-economy model. *Journal of Applied Econometrics, 25*(1):93-128.
61. Kim, C.-J. and Nelson, C.R. (1999). State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications, MIT Press: Cambridge.
62. Koop, G. and Korobilis, D. (2010). Bayesian Multivariate Time Series Methods for Empirical Macroeconomics. *Foundations and Trends(R) in Econometrics*, 3(4):267-358.
63. Koop, G., and Korobilis, D. (2012). Forecasting Inflation using Dynamic Model Averaging. *International Economic Review,* 53(3):867-886.
64. Koop, G., and Korobilis, D.. (2013). Large time-varying parameter VARs. *Journal of Econometrics,* 177(2):185-198.
65. Korobilis, D., (2013). VAR Forecasting Using Bayesian Variable Selection. *Journal of Applied Econometrics,* 28(2):204-230.
66. Litterman, R. (1986). Forecasting with Bayesian vector autoregressions — Five years of experience: Robert B. Litterman, *Journal of Business and Economic Statistics,* 4:25–38.
67. Ljung, L. (1987). *System identification : Theory for the user*. Prentice-Hall information and system sciences series. Englewood Cliffs, NJ: Prentice-Hall.
68. Marcellino, M. and Rychalovska, Y. (2014). Forecasting with a DSGE Model of a Small Open Economy within the Monetary Union. *Journal of Forecasting*, 33(5):315-338.
69. Mark, N. (1995). Exchange-rates and Fundametnals – Evidence on Long-Horizon Predictability. *American Economic Review,* 85(1):201-218.
70. Mark, N.C. and Moh, Y.-K. (2002). Continuous-Time Market Dynamics, ARCH Effects, and the Forward Premium Anomaly, unpublished working paper, University of Notre Dame and Tulane University.
71. Mcgough, B., Rudebusch, G.D. and Williams, J.C. (2005). Using a long-term interest rate as the monetary policy instrument. *Journal of Monetary Economics,* 52(5):855-879.
72. Meese, R.A., and Rogoff, K. (1983a). Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?, *Journal of International Economics,*14:3-24.
73. Meese, R.A., and Rogoff, K. (1983b). *The Out-of-Sample Failure of Empirical Exchange Rate Models: Sampling Error or Misspecification?,* Exchange Rates and International Macroeconomics. Jacob Frenkel, eds., Chicago: NBER and University of Chicago Press.

74. Molodtsova, T. and Papel, D.H. (2009). Out-of-sample exchange rate predictability with Taylor rule fundamentals. *Journal of International Economics*, 77(2):167-180.
75. Molodtsova, T., Nikolsko-Rzhevskyy, A. and Papell, D. (2011). Taylor rules and the euro. *Journal of Money, Credit and Banking,* 43(2-3): 535-552.
76. Molodtsova, T., Nikolsko-Rzhevskyy, A. and Papell, D.H. (2008). Taylor rules with real-time data: A tale of two countries and one exchange rate. *Journal of Monetary Economics,* 55:63-79.
77. Monacelli, T. (2005). Monetary Policy in a Low Pass-Through Environment. *Journal of Money, Credit, and Banking,* 37(6):1047-1066.
78. Mussa, M. (1976). A Model of Exchange Rate Dynamics. *Scandinavian Journal of Economics,* 78:229-248.
79. Nikolsko-Rzhevskyy, A. (2011). Monetary Policy Estimation in Real Time: Forward-Looking Taylor Rules without Forward-Looking Data. *Journal of Money, Credit and Banking,* 43(5):871-897.
80. Nucci, F. (2003). Cross-currency, cross-maturity forward exchange premiums as predictors of spot rate changes: Theory and evidence. *Journal of Banking and Finance,* 27(2):183-200.
81. Orphanides, A. (2001). Monetary Policy Rules Based on Real-Time Data. *American Economic Review*, 91(4):694-985.
82. Paye, B.S. and Timmermann, A. (2006). Instability of return prediction models. *Journal of Empirical Finance*, 13(3):274-315.
83. Primiceri, G.E. (2005). Time Varying Structural Vector Autoregressions and Monetary Policy, *Review of Economic Studies*, 72(3):821-852.
84. Rabanal, P. and Rubio-Ramírez. J.F. (2005). Comparing New Keynesian models of the business cycle: A Bayesian approach. *Journal of Monetary Economics*, 52(6):1151-1166.
85. Raftery, A., Kárný, M., and Ettler, P. (2010). Online Prediction Under Model Uncertainty via Dynamic Model Averaging: Application to a Cold Rolling Mill. *Technometrics,* 52(1):52-66.
86. Rogoff K. and Stavrakeva V. (2008). The Continuing Puzzle of Short Horizon Exchange Rate Forecasting. *National Bureau of Economics Research*, Working Paper Series, No. 14071.
87. Rossi, B. (2013a). Exchange Rate Predictability. *Journal of Economic Literature*, 51(4):1063-1119.
88. Rossi, B. (2013b). Advances in Forecasting under Instability, in: *Handbook of Economic Forecasting*, Volume 2B, Elsevier Publications, 1203-1324.
89. Sarantis, N. (1995). Modelling sterling exchange rates and interest rate differentials. *Applied Financial Economics,* 5(3-4):345-356.
90. Sarantis, N. (2006). On the short-term predictability of exchange rates: A BVAR time-varying parameters approach. *Journal of Banking and Finance,* 30(8):2257-2279.
91. Sarno, L., and Valente, G. (2009). Exchange Rates and Fundamentals: Footloose or Evolving Relationship?. *Journal of the European Economic Association*, 7(4):786-830.

92. Sarno, L., Valente, G., and Leon, H. (2006). Nonlinearity in Deviations from Uncovered Interest Parity: An Explanation of the Forward Bias Puzzle. *Review of Finance*, 10(3):443-482.
93. Schmitt-Grohé, S. and Uribe, M. (2003). Closing small open economy models. *Journal of International Economics*, 61(1):163-185.
94. Sims, C. (1980). Macroeconomics and reality. *Econometrica*, 48:1-48.
95. Smets, F. and Wouters, R. (2004). Forecasting with a Bayesian DSGE Model: An Application to the Euro Area. *Journal of Common Market Studies*, 42(4):841-867.
96. Smets, F., & Wouters, R. (2002). Openness, imperfect exchange rate pass-through and monetary policy. *Journal of Monetary Economics*, *49*(5):947-981.
97. Smets, F., & Wouters, R. (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review*, *97*(3):586-606.
98. Steinbach, M., Mathuloe, P., and Smit, B. (2009). An Open Economy New Keynesian DSGE Model of the South African Economy. *South African Journal of Economics*, *77*(2):207-227.
99. Taylor, J.B. (1993). Discretion Versus Policy Rules in Practice. *Carnegie-Rochester Conference Series on Public Policy*, 39:195-214.
100. Taylor, M. (1995). The Economics of Exchange-rates. *Journal of Economic Literature*, 33(1):13-47.
101. Terasvirta, T. (2006). Forecasting economic variables with nonlinear models," Handbook of Economic Forecasting, Elsevier.
102. Villemot, S. (2011). Solving rational expectations models at first order: what Dynare does. *Dynare Working Papers 2*, CEPREMAP.
103. Woodford, M. (2003). Optimal Interest-Rate Smoothing. *Review of Economic Studies*, 70(4):861-886.
104. Wright, J. (2008). Bayesian Model Averaging and exchange rate forecasts. *Journal of Econometrics*, 146(2):329-341.
105. Wright, J. (2009). Forecasting US inflation by Bayesian model averaging. *Journal of Forecasting*, 28(2):131-144.