Economic Policy Department
WORKING PAPER SERIES

MINISTRY FOR FINANCE
MALTA

Robustness in Foreign Exchange Rate Forecasting Models: Economics-based Modelling after the Financial Crisis *

Carlos Medel and Gilmour Camilleri†, Hsiang-Ling Hsu, Stefan Kania and Miltiadis Touloumtzoglou

November 2015

WP 04/2015

* The authors would like to thank Pablo Medel and Denise Debattista for their comments and suggestions.
Pablo Medel is a PhD Student at the University of Manchester (UK) while Denise Debattista is a Financial Analyst at the Central Bank of Malta.

† Corresponding authors: Carlos Medel and Gilmour Camilleri.
Carlos Medel is an economist at the Financial Stability Area, Central Bank of Chile.
Email: cmedel@bcentral.cl.
Gilmour Camilleri is an economics analyst at the Economic Policy Department, Ministry for Finance, Malta.
Email: gilmour.camilleri@gov.mt.

Robustness in Foreign Exchange Rate Forecasting Models: Economics-based Modelling After the Financial Crisis

Carlos A. Medel   Gilmour Camilleri   Hsiang-Ling Hsu   Stefan Kania
Miltiadis Touloumtzoglou

November 2015

Abstract

The aim of this article is to analyse the out-of-sample behaviour of a bunch of statistical and economics-based models when forecasting exchange rates (FX) for the UK, Japan, and the Euro Zone in relation to the US. A special focus is given to the commodity prices boom of 2007-8 and the financial crisis of 2008-9. We analyse the forecasting behaviour of six economic plus three statistical models when forecasting from one up to 60-steps-ahead, using a monthly dataset comprising from 1981.1 to 2014.6. We first analyse forecasting errors until mid-2006 to then compare to those obtained until mid-2014. Our six economics-based models can be classified in three groups: interest rate spreads, monetary fundamentals, and PPP with global measures. Our results indicate that there are indeed changes of the first best models when considering the different spans. Interest rate models tend to be better predicting using the short sample; also showing a better tracking when crisis hit. With the longer sample the models based on price differentials are more promising; however, with heterogeneous results across countries. These results are important since shed some light on what model specification use when facing different FX volatility.

1 Introduction

Modern macroeconomics relies hugely on foreign exchange rate (FX) dynamics. Several trade theory foundations give a key role for FX in terms of the informational content that it provides. FX typically measures structural misalignments anticipating future short-run dynamics of key macroeconomic variables aiming to correct those misalignments with or without external intervention. Some common models are the (un-)covered interest rate parity and the purchasing power parity, or *law of one price*. This kind of models–the former developed primarily for interest rate dynamics–has a long tradition as common wisdom in macroeconomics for both its tractability and modelling convenience.

As an example, take the case of an English agent that invests a certain amount of money in the US pursuing a rate of return of i^{US}. Her principal plus return after s periods in British Sterling pounds corresponds to $\mathbb{E}_t[1+e_{t+s}^{US/UK}]/(1+e_t^{US/UK})\times(1+i_t^{US})$, where $e_i^{US/UK}$ is the FX at period i, and $\mathbb{E}[\cdot]$ is the expectations operator. Obviously, this return must equate the return that she would receive in her home country, i_t^{UK}.[1] Hence,

$$(1+i_t^{UK}) = (1+i_t^{US}) \times \frac{\mathbb{E}_t[1+e_{t+s}^{US/UK}]}{(1+e_t^{US/UK})}, \tag{1}$$

corresponding approximately to $i^{UK} = i^{US} + \Delta\mathbb{E}_t(e_{t+s})/e_t$. While this simple model has been used as an interest rates model, it is rather useful to understand the foundations of a plethora of FX economics-based models. Note that Equation 1 can be rearranged as:

$$\frac{\mathbb{E}_t\left[e_{t+s}\right]}{e_t} = i^{UK} - i^{US}. \tag{2}$$

Several theories explain the mechanics behind interest rates spreads using twists to Equation 2, especially useful for policymaking. Many interest rates models inherit the economic fundamentals to the FX variable.[2] There are subsequent extensions to Equation 2 coming from the theory–including more determinants or in a multivariate ensemble–as well as methodological, such as cointegration and vector error correction modelling (VECM).

Several research papers make use of different versions of Equation 2 for forecasting purposes using a statistical evaluation criterion. Despite the quest for the true FX economic model, it is raised the question on the purpose of those exercises. While a macroeconomic answer relies on policymaking, a financial perspective takes the point of view of an investing problem–*i.e.* maximising utility as a function of wealth.

As the focus is changed to investor's utility, the evaluation of models also makes a shift in this direction (referred henceforth as "financial evaluation"; see Granger and Pesaran, 2000, for a discussion on this matter).[3] A common feature within these two kinds of analysis is the use of statistical models–mostly autoregressions–as candidate models.

The FX dynamics as well as its forecasts have a long-standing tradition in macroeconomics. Some selected surveys are Taylor (1995), Sarno and Taylor (2002), Engel, Mark, and West (2007), Della Corte, Sarno, and Tsiakas (2009), Williamson (2009), and more recently, Evans (2011).

[1] Or, if i_t^{UK} is lower than the equivalent return obtained outside, then more English agents will invest abroad pursuing a higher return. At last, i_t^{UK} will raise until equating *all* foreign returns following the typical no-arbitrage condition. See Dornbusch, Fischer, and Startz (2010) for details.

[2] See Chapter 10–eloquently entitled "Some Useful Models"–in Blanchard and Fischer (1989) for details.

[3] We will refer to model categories as "statistical" and "economic-based", and to the evaluation procedures–described later–as "statistical" and "financial".

The different statistical evaluation of either economical or statistical models, has been analysed and used since the beginning of the literature mostly associated with Meese and Rogoff (1983).[4]

Some other articles, such as Boothe and Glassman (1987), Leitch and Tanner (1991), Pesaran and Timmermann (1995), Kandel and Stambaugh (1996), focus on the out-of-sample evaluation of the investor problem. This is one of the avenues analysed in Garratt and Lee (2010) which also follows the approach of West, Edison, and Cho (1993), Barberis (2000), and Abhyankar, Sarno, and Valente (2005). More recently, Melvin, Prins, and Shand (2013) provide an overview as well as a complete exercise with after-crisis data for certain industrialised economies.

Garratt and Lee (2010; henceforth GL) analyses the forecasting behaviour of several economic and statistical models for the FX of Japan and the UK in relation to the US. A key feature is that the appropriateness of models depends on the evaluation criteria, either statistical or financial. In particular, they find that an autoregressive (AR) model outperforms economic models when point forecast is evaluated with a statistical criterion–*i.e.* root mean squared forecast error (RMSFE). However, financial evaluation suggests that economic models provide a higher rate of return. As economic models, GL use three restricted VECM of the type $\Delta\log(e_t^{i,j}) = \alpha + \lambda\nu_{t-1} + \beta' x_t + \varepsilon_t$,where ν_t and ε_t are white noises; ν_t coming from a restricted equation in levels, and $-1<\lambda<0$ is the long-run adjustment parameter. The variable x_t takes three different specifications, labelled *efficient market hypothesis* (EFH), *monetary fundamentals* (MF), and *purchasing power parity* (PPP). Multistep forecast with monthly variables are made for h={1,3,6,12,24,36,48}-steps-ahead with models estimated in a recursive sample scheme.

The financial evaluation is made assuming an investor with a portfolio of two assets, one returning in domestic (US) and the other in foreign terms (either from Japan or the UK). The portfolio weights are chosen with a sample-valued simulation of the obtained forecasts maximising investor utility. Results are reported as the ratio between the (risk-averse) utility obtained with the candidate forecasts and the random walk model (RW).

There are several ways to provide robustness of GL results. For instance, by allowing short sales (negative weights) to the investor, an asymmetrical utility function which penalises losses more severely, make available more assets in the portfolio, different portfolio-weighting strategy, among other financial set up. Furthermore, a rich assessment with more complex time series models, such as nonlinear models (Meese and Rose, 1991; Satchell and Timmermann, 1995), regime switching models (Engel and Hamilton, 1990; Cheung and Erlandsson, 2005), and neuronal networks (Andreou, Georgpoulos, Likothanassis, 2002), are of interest when major disruptions are experienced.[5]

In this article we analyse several extensions to GL analysis in the "first line". These are extensions to the candidate models and statistically evaluated. This is made with the genuine interest of provide some robustness to GL results, but also in regard to the out-of-sample behaviour of the different models during the commodity prices boom of 2007-8 and the financial crisis of 2008-9. Hence, as GL uses a sample covering from 1981.1 to 2006.6 (306 observations), we extend the analysis using the same dataset until 2014.6 (396 observations). We also include the FX of the Euro Zone with respect to the

[4]Some other articles using this methodology are MacDonald and Taylor (1994), Chinn and Meese (1995), Kim and Mo (1995), Mark (1995), Mark and Sul (2001), and Faust, Rogers, and Wright (2003), among others. Berkowitz and Giorgianni (2001) focus on the long-sample forecast accuracy; Clarida *et al.* (2003) on the information provided in the term-structure curve following the traditional Estrella and Mishkin (1996) argument; Kilian and Taylor (2003) on stressing the difficulties to beat the naive random walk forecast; and Cheung, Chinn, and García Pascual (2005) on the exploitation of the short-run adjustment to a long-run relationship in price indices.

[5]Also, a different statistical evaluation, such as the *direction of forecast*, *sign test*, or *hit rate*, is also worthwhile to analyse for the investor. See Cheung and Chinn (1998) for details.

US.[6] [7] Unlike the data availability of GL, we consider h=60 into analysis. Our modelling extensions consist mainly of the use of different global price indices instead of just one foreign measure (the US) as used by GL in their PPP model. These changes go in the avenue of the PPP model using core inflation, the Brent oil price, and the IMF's *Primary Commodity Prices Index*. Finally, we include the single exponential smoothing model in the statistical models family.

Our results suggest for the case of the UK that the MF outperform remaining candidates at every single horizon considering the GL sample span.[8] Also, the AR model performs as the second best alternative at longer horizons. When considering our whole evaluation sample (including 2006.7 onwards) the performance is more homogeneous across the models. Again, evidence is mixed between economical and statistical models across the considered horizons. In the short run ($h \leq 6$) the AR and oil-based models exhibit the best performance. When $h > 6$ there is virtually a tie between all the models except the ES and the EMH model outperforming at h=60. Obviously, the performance of all the models is spoiled and closer to the RW given the higher variance exhibited in the last part of the sample.

For the case of Japan with the GL sample, the EMH outperforms remaining models at any horizon. Japan has exhibited a particular FX dynamics historically characterised by values close to zero. When considering the whole sample, in the short run the proposed core inflation and oil price models, show the best forecasting performance from $h > 1$. In the long-run ($h > 12$) the remaining GL economics-based models also show a relatively better performance than the RW and proposed models. Statistical models seem less promising in this case.

Finally, for the Euro Zone, there are two models outperforming in the first evaluation sample, MF and AR. However, when considering the whole sample, it is the commodity price model that only outperforms the RW (for $h > 6$ onwards).

These results are in line with GL when comparable. Note that economics-based models are at least as good and sometimes superior than statistical models using a larger sample span. This is observed in the long run and when the variance of the dependent variable is increased. Remarkably, when considering the extended sample there is a transition of the best models towards PPP-implied models.

The rest of the article proceeds as follows. In Section 2 we fully describe the econometric setup of the forecasting exercise: the models, evaluation procedure, and dataset. In Section 3 we analyse our in-sample results with diagnostics checking as well as the out-of-sample performance. A special focus on the behaviour of forecasting errors is provided in order to assess the two aforementioned inflationary episodes. Finally, we conclude in Section 4.

2 Econometric setup

2.1 Competing models and extensions

We make use of nine forecasting models of which five comes from GL analysis. There are six economics models; three coming from GL. We also make use of a driftless RW as a benchmark. The baseline

[6] We will refer to the Euro Zone as a country since we consider the European Monetary Union sharing a common currency, and hence, the same FX. Research on this matter is provided by Rosemberg (2000), Owen (2001), Dal Bianco, Camacho, and Pérez-Quirós (2012), Kirikos (2013), among others. Moreover, Brzeszczynski and Melvin (2006) warn on the importance of the Euro and US Dollar as they act as numeraries for many countries affecting key policy decisions.

[7] Melvin and Taylor (2009) as well as Molodtsova and Papell (2009) focus also on the out-of-sample behaviour of FX models during the crisis.

[8] There are many reasons why GL figures are not exactly recovered in a replication exercise like this. These includes: different dataset vintages, different software algorithms (despite using the same estimation procedures at user level), and different decimal places–sensitive for log-likelihood function computation with breaks and data transformations.

specification of the economic models can be summarised in the following VECM:

$$\Delta \mathbf{z}_t = \mathbf{a} + \sum_{i=1}^{p} \mathbf{\Gamma} \Delta \mathbf{z}_{t-i} + \boldsymbol{\alpha}\boldsymbol{\beta}' \mathbf{z}_{t-1} + \mathbf{u}_t, \tag{3}$$

where $\mathbf{z}_t$ is a vector of order two containing the $\log(e_t^{i,j})$ and $\log(x_t)$ variables. The variable x_t is changing according to each model. The cointegrating vector $\boldsymbol{\beta}$ is restricted to $\boldsymbol{\beta\prime} = (1, -1)$.[9] $\mathbf{u}_t$ is a white noise; $\mathbf{a}$, $\mathbf{\Gamma}$, and $\boldsymbol{\alpha}$ are vector parameters to be estimated with OLS. The lag-length p is determined with the likelihood ratio sequential test using the estimation sample.

The used models are:

- **Efficient Market Hypothesis (EMH)**. In this model, the variable x_t is defined as $x_t = f_t$, where f_t is the log of the forward nominal bilateral exchange rate (end-of-period). The main reason for the use of this model relies on the information provided by the interest rate term structure curve as pointed out in Taylor (1989), Clarida and Taylor (1997), and more recently Møller (2014).

- **Monetary Fundamentals Model (MF).** In this model, the variable x_t takes the form $x_t = (m_t - m_t^*) - (y_t - y_t^*)$, where m_t is the log-level of M1 and y_t is a measure of activity, *i.e.* industrial production. Variables with "*" indicate foreign precedence, in this case, the US. This model has a long-standing tradition in economics as a fundamental FX determination model (see Frenkel, 1976, 1979; Dornbusch, 1976; Mussa, 1976; and Hooper and Morton, 1982).

- **Purchasing Power Parity (PPP; Core; P(Oil); Cmdty)**. This model makes use of $x_t = p_t - p_t^*$, where p_t is the log of CPI. This model constitutes the law of one price, suggesting that a no-arbitrage condition holds across countries. In this case, we label the model as "PPP".

  - We also consider $x_t = \widetilde{p}_t - \widetilde{p}_t^*$, where $\widetilde{p}_t$ is the so-called *core inflation* measure, *i.e.* the whole CPI excluding the components of food and energy. As this model concerns a measure more prone to policy decisions (Goodfriend, 2008), the FX trajectory should accommodate to those decisions becoming sensitive to the inner movements of the CPI. While Garratt *et al.* (2006) find evidence that PPP is fulfilled at the long run, Hakkio (2009) finds that inflationary shocks are spilled over countries even at core level. This fact, and since the PPP theory is not specific on which price level measure should be used, we opt to explore this avenue for our purposes. This model is labelled as "Core".
  - Lastly, we make use of two global prices measures in order to include a price gap beyond pairwise comparisons. These twists are made to assess the detrimental effects of two major disruptions occurred outside the sample span used in GL, which are the commodity prices boom of 2007-8 and the financial crisis of 2008-9. Hence, we define $x_t = p_t - p_t^{oil}$ and $x_t = p_t - p_t^{cmdty}$, where p_t^{oil} is the log-level of the Brent oil price and p_t^{cmdty} is the log-level of IMF's *Primary Commodity Prices* index (detailed later). These models are labelled "P(Oil)" and "Cmdty".

- **Stationary Autoregressive Model (AR)**. This model corresponds to the traditional AR time series model of order p, where p is chosen according to the Akaike Information Criteria (AIC). The model assumes that $z_t = \log(e_t^{i,j})$ and $\boldsymbol{\beta}' = 0$. Almost all empirical analysis and forecasting exercises of FX includes this model as either candidate or benchmark given its accurate results. This model alongside the (benchmark) RW and the exponential smoothing (ES) constitutes the set of statistical models.

[9] It is analysed later to what extent this restriction find empirical support.

- **Exponential Smoothing Forecast (ES).** GL makes use of the RW as a benchmark forecast. Nevertheless, and as suggested in Hyndman *et al.* (2008), the single ES model could provide at least similar results than the RW. Also, the simple version used in this article enriches the exercise in a tractable manner, as it is also used in Fat and Deszi (2011). If $y_t = \Delta \log(e_t^{i,j})$ then the single univariate ES forecast ($\widehat{y}_t$) is defined as $\widehat{y}_t = \varphi y_{t-1} + (1-\varphi)\widehat{y}_{t-1}$, with $0 < \varphi \leq 1$ (the smaller is the φ, smoother is the forecast series). Note that this one-step-ahead forecast corresponds also to the multihorizon forecast as is used with the RW (if $\varphi = 1$ we obtain exactly the RW). The model can be written as a recursion depending on $\widehat{y}_t$. Hence, it is needed an initial value of $\widehat{y}_t$ to estimate φ. This value corresponds to the average of the first $(T+1)/2$ values of y_t, where T is the number of observations.

2.2 Model evaluation and comparison

The statistical measure used to evaluate the accuracy of point forecast is the RMSFE:

$$\mathsf{RMSFE}_h = \left[\frac{1}{T}\sum_{t=1}^{T}(y_{t+h} - \widehat{y}_t^h)^2\right]^{\frac{1}{2}}, \tag{4}$$

where $\widehat{y}_t^h$ is the h-step-ahead forecast of y_{t+h} made at period t. Note that this statistic is computed given a forecasting horizon h, and hence, the difference $T-t$ is variable depending on h–*i.e.* $T = T(h)$. To make more plausible comparison with the RW, the analysed statistic corresponds to the RMSFE Ratio defined as:

$$\mathsf{RMSFE\ Ratio} = \frac{\mathsf{RMSFE}_h^{\mathcal{M}}}{\mathsf{RMSFE}_h^{RW}}, \tag{5}$$

where $\mathcal{M}$={EMH,MF,PPP,Core,P(Oil),Cmdty,AR,ES}. Hence, as the RW acts as a pivot, values greater than unity imply a worse performance of the competing model. Figures below unity represent a "predictive gain" of (1-RMSFE Ratio)% compared to the RW.

To investigate to what extent the predictive gains are statistically significant, we make use of the unconditional t-type test of Giacomini and White (2006) providing the advantage of comparing *forecasting methods* instead of *forecasting models.* As the null hypothesis (*NH*) is defined as *the competing model has a superior predictive ability compared to the* RW, there is used a one-side t-type GW statistic accordingly.

Formally, it is tested the *NH*: $\mathbb{E}_t(d_h) \leq 0$, against the alternative *AH*: $\mathbb{E}_t(d_h) > 0$, where:

$$d_h = (y_{t+h} - \widehat{y}_t^{RW})^2 - (y_{t+h} - \widehat{y}_t^{\mathcal{M}})^2, \tag{6}$$

using the Newey and West (1987) HAC estimator of the standard deviation of d_h. The *NH* is rejected if the subsequent t-statistic is greater than $t_{\alpha\%}$; this last term corresponding to the tabulated value of a normal distribution with probability $\alpha\%$.

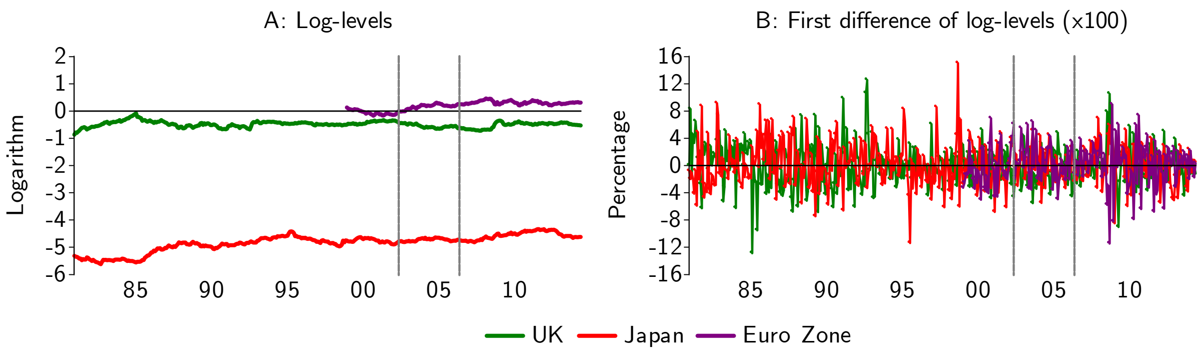
2.3 Data

The dataset comprises monthly variables from 1981.1 to 2014.6 (396 observations). Note that the GL sample cover from 1981.1 to 2006.6 (306 observations). We divide the sample in three branches: estimation sample, evaluation sample I (ES.I), and evaluation sample II (ES.II). The estimation sample and ES.I coincide with the division made in GL: 1981.1–2002.6 (255 observations) and 2002.7–2006.6 (51 observations), respectively. We then extend the analysis to 2014.6 (90 new observations), becoming ES.II. Our out-of-sample results are presented for these two samples to ease a direct comparison, noticing that ES.II includes ES.I. Given this scheme, we compute 144 forecasts at one-step-ahead until 85 for 60-steps-ahead. We estimate the models in a recursive manner adding one observation every time a new forecast is made.

The sources are IMF's *International Financial Statistics* (IFS) and OECD's *StatsExtracts* databases. In Figure 1, there are presented the three dependent variables analysed; in log-levels (Panel A) and the first difference of the log-levels (Panel B; $100 \times \Delta \log(e_t^{i,j})$). Note that for the Euro Zone, the sample starts in 1999.1 (186 observations), circumscribing the analysis for this region to this sample span. Note that in Panel A, Japan exhibits always negative values indicating values smaller than unity in its FX at least over the two last decades.

In Annex A, we fully describe the sources and the specific IFS code (when corresponding) of the dataset. Also, in Annex B we present different descriptive statistics for three different samples as well as the unit root testing results. It is observed major differences in mean, median, and variance between the two evaluation samples; obviously, increasing its volatility in the last part. The Augmented Dickey-Fuller test (ADF) is applied using the full sample to investigate the presence of a unit root, which also shed some light on the reliability of the restriction imposed in the VECM. The results suggest that in log-levels terms two variables are I(1) at 10% level of confidence, and I(0) in first differences. These are for the UK and Euro Zone, going in the direction that supports the restrictions. For Japan, the results suggest that the log-level is already I(0) at 10% confidence. The test is conducted assuming a constant in the cointegration equation and using the Bayesian Information Criterion (BIC) as a lag-length criteria ($p^{\max}$=24).

Figure 1: Foreign Exchange Rate Time Series (*)



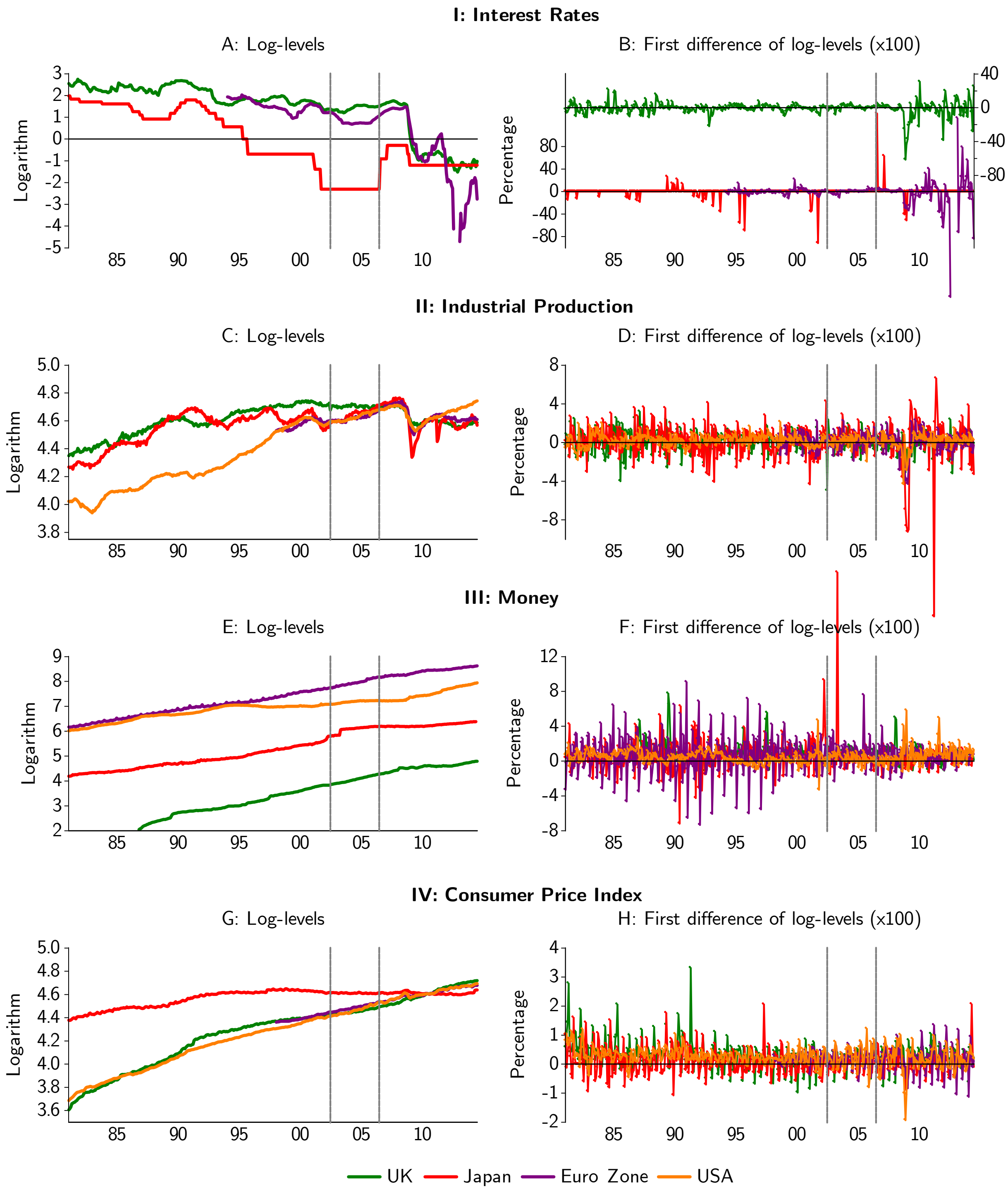
(*) Vertical lines: Evaluation Sample I and Evaluation Sample II. Source: IFS.

In Figure 2 there are presented the covariates (x_t candidates) used for the baseline GL models; in the same order that models were described in Subsection 2.1. It is observed in almost all the cases major disturbances especially in ES.II, which make the forecasting exercise more challenging.

Panel I shows the interest rates series used for the EMH model (see Table A1 in the Appendix). The behaviour of the UK rate seems to follow a different regime since the half of ES.II, considering that log-levels now exhibit negative values. Also, Panel B reveals a noticeable higher variance in ES.II. A big jump is adverted for Japan at the beginning of ES.II, for sure affecting the accuracy especially in models with a greater persistence. For the Euro Zone, there are four missing observations given negative values, which do not receive any special treatment. Major disturbances are also noticed especially in the second half of ES.II.

Panels II and III present the variables used for the MF model: money (M1) and industrial production. Money variables seem smooth with no major shifts for the whole sample, except of one outlier for Japan in ES.I. The industrial production series exhibit a strong correlation in the middle of ES.II, specifically corresponding to the financial crisis of 2008-9. Log-levels exhibit a typical V-shape dynamic except for Japan showing a W-shape series generating major disturbances in the first differences series accordingly.

Figure 2: Baseline Models Covariates Time Series (*)



(*) See notes in Figure 1. Source: IFS.

Panel IV shows the CPI variable used for the PPP model. No major disturbances are noticed except a minor hump in ES.II for all countries which does not generate atypical observations in the first difference transformation of the series.

Overall, we have that for ES.II the covariates are more volatile, showing outliers and a regime different from the previous ES.I span (and GL analysis).

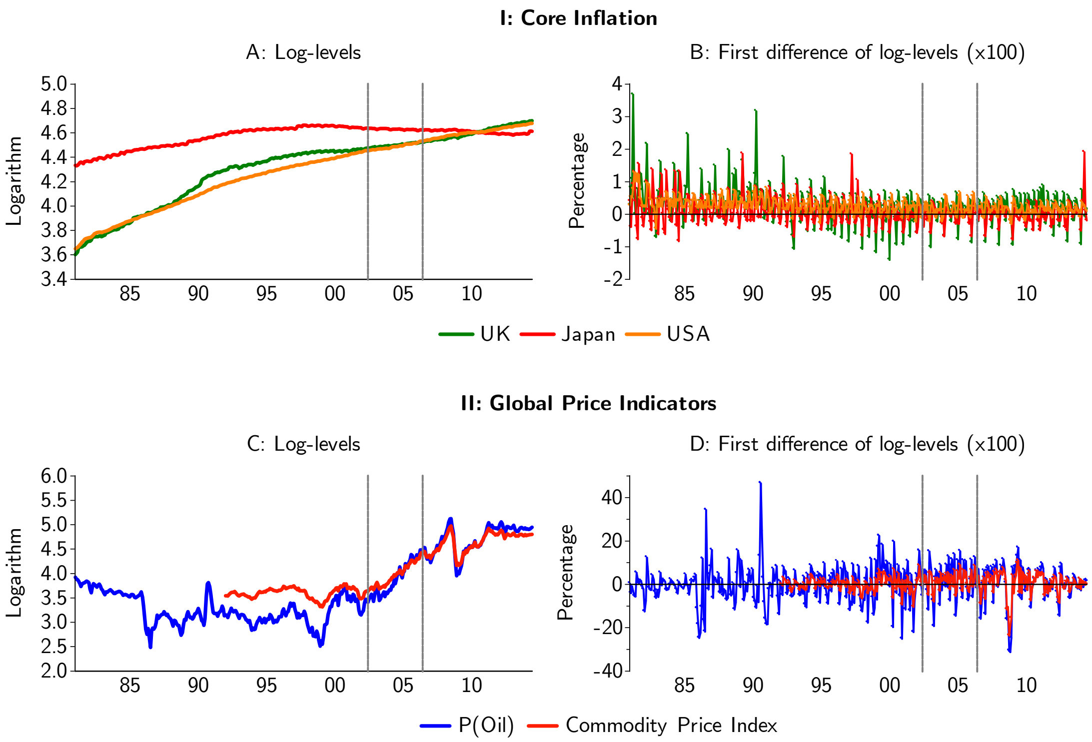
In Figure 3, we present the three variables used in the economics-based extensions: core inflation, oil price, and commodity price index.

Panel I shows core inflation for three economies: UK, Japan, and the US. There is no available an official core inflation for the Euro Zone; hence we leave this model for further research. Note that as core inflation excludes the most volatile components (food and energy) the series exhibit a smooth behaviour with both transformations.

Panel II shows the oil price and the commodity price index for P(Oil) and Cmdty. The oil price corresponds to the Brent oil price defined as USD per barrel. The log-level starts to grow precisely when ES.I begins. Also, in ES.II exhibits a V-shape dynamic but with a growth rate close to zero in the aftermath.

The commodity price index (IMF *Primary Commodity Prices*) consists in a weighted average of food, beverages, agricultural and raw-materials, metals, and petroleum index prices using the weights of an average commodity basket estimated for 2004-9. It is available since 1992.1 onwards. Note that its dynamics follows closely that of the oil price especially since mid-2002.

Figure 3: Extension Models Covariates Time Series (*)



(*) See notes in Figure 1. Source: IFS and OECD.

3 Results

This section analyses the different in- and out-of-sample results across the countries. All the results are obtained using an ad-hoc Eviews 8 program making use of the `VARForecast` add-in.

3.1 In-sample results

In Table 1, there are presented some diagnostic statistics for the case where the model is estimated using the restricted VECM. These results are reported for two spans: full sample (1981.1-2014.6) and estimation sample (1981.1-2002.5).

For the UK, there are just two models that exhibit a greater R^2 when considering the whole sample: the AR and P(Oil). Note that in the case of the AR, the R^2 level as well as the increment are marginal: from 0.003 to 0.008 (0.017 to 0.025 for P(Oil) model). The Cmdty model exhibit for both samples the best adjustment according to this measure. When considering jointly the fit to data given by the log-likelihood function and the number of coefficients of the model (*i.e.* the information criteria), the best adjustment is achieved with the PPP model for both spans. Hence, UK FX dynamics seems to be commanded mostly by external global price indices.

For Japan, all models exhibit a lower explanatory power when considering the last part of the sample according to the R^2 statistic. In this case, the proposed three economics models exhibit a better adjustment than the baseline GL models. When considering information criteria, the Core model exhibits the best in-sample fit outperforming within economics models.

For the Euro Zone (in Table 2), all the models also show a decline in their fit when considering the last part of the sample. In this case, the proposed economics model shows a narrow fit according to the R^2. There is not a single model showing the best results with both samples. However, the MF, PPP, and AR seem promising. Finally, according to information criteria, the best model is PPP.

Table 1: Restricted Estimation Diagnostics (*)

	UK					**Japan**				
	LLhd.	R^2	$\widehat{\sigma}^2_\varepsilon$	AIC	BIC	LLhd.	R^2	$\widehat{\sigma}^2_\varepsilon$	AIC	BIC
	EMH: Efficient markets hypothesis									
FS	859.808	0.074	0.029	-6.572	-6.075	829.427	0.052	0.031	-5.596	-5.298
ES	542.545	0.134	0.031	-7.265	-6.574	512.819	0.075	0.034	-5.714	-5.300
	MF: Monetary fundamentals									
FS	717.060	0.034	0.028	-9.584	-9.469	831.135	0.072	0.031	-8.294	-7.836
ES	392.744	0.043	0.029	-9.737	-9.563	513.817	0.083	0.034	-8.528	-7.893
	PPP: Purchasing power parity									
FS	855.498	0.053	0.030	-12.470	-11.973	836.603	0.086	0.031	-12.210	-11.713
ES	534.802	0.080	0.032	-12.346	-11.656	516.807	0.104	0.034	-11.987	-11.296
	AR: Autoregressive model									
FS	846.154	0.008	0.030	-4.205	-4.195	819.357	0.004	0.032	-4.071	-4.061
ES	525.869	0.003	0.032	-4.069	-4.055	505.329	0.006	0.034	-3.910	-3.896
	Core: Core inflation model									
FS	855.836	0.055	0.030	-12.450	-11.953	837.274	0.089	0.031	-12.789	-12.292
ES	534.826	0.080	0.032	-11.990	-11.300	517.888	0.111	0.034	-12.356	-11.666
	P(Oil): Oil price model									
FS	849.621	0.025	0.029	-6.358	-6.219	840.898	0.105	0.031	-6.170	-5.713
ES	526.263	0.017	0.032	-6.168	-5.975	522.663	0.144	0.033	-5.982	-5.347
	Cmdty: Commodity price index model									
FS	630.062	0.176	0.022	-8.299	-7.779	551.064	0.116	0.030	-7.605	-6.972
ES	298.158	0.216	0.020	-8.834	-7.927	232.951	0.226	0.035	-7.820	-6.709

(*) FS: Full Sample (1981.1–2014.6). ES: Estimation Sample (1981.1–2002.5). "LLhd." stands for log-likelihood function. AIC and BIC stand for Akaike and Bayesian Information Criteria. Source: Authors' elaboration.

Table 2: Restricted Estimation Diagnostics (*)

	Euro Zone				
	LLhd.	R^2	$\widehat{\sigma}_{\varepsilon}^2$	AIC	BIC
	EMH: Efficient markets hypothesis				
FS	368.666	0.011	0.030	-4.394	-4.287
ES	85.842	0.050	0.028	-7.294	-7.038
	MF: Monetary fundamentals				
FS	385.474	0.016	0.030	-9.282	-9.177
ES	85.942	0.055	0.028	-9.543	-9.287
	PPP: Purchasing power parity				
FS	384.595	0.007	0.030	-12.189	-12.084
ES	86.256	0.070	0.028	-13.037	-12.781
	AR: Autoregressive model				
FS	384.045	0.001	0.030	-4.153	-4.118
ES	86.843	0.092	0.028	-4.242	-4.158
	Core: Core inflation model				
FS					
ES					
	P(Oil): Oil price model				
FS	384.277	0.003	0.030	-6.242	-6.137
ES	85.982	0.056	0.028	-5.897	-5.641
	Cmdty: Commodity price index model				
FS	384.523	0.006	0.030	-7.561	-7.456
ES	85.860	0.051	0.028	-7.777	-7.521

(*) See notes in Table 1. In this case, FS spans from 1999.1 to 2014.6.
Source: Authors' elaboration.

Overall, we have that economic models based on price differentials provide a better in-sample fit, barely superior than the AR model. Interestingly, the latter model is also superior to the EMH which contains a forward-looking variable as the forward exchange rate interest rate is. It is suitable at this point to remark that given the presence of breaks and regime changes that the in-sample performance is not necessarily extrapolated to out-of-sample accuracy (see Hansen, 2009, for a formal discussion). Obviously, it is a necessity to have an estimated model closest to the true model to generate the forecasts.

We analyse next the appropriateness of the imposed restriction on the cointegrating vector. Same as before, this analysis should be read to complement the modelling mechanism behind the results, but in spite of their acceptance or economical motivation, it should be judged by the forecasting accuracy that provides.

In Table 3, we present the results of the Engle-Granger cointegration test. Note that we present the result for two samples just for robustness, because it is desirable for consistency and testing-power-enhancing the use of a *big* sample. Hence, our results under the column "FS" will determine to what extent the log-levels are cointegrated for each model.

Our estimates consist of a residual-based testing of cointegration. If the series are not cointegrated, all linear combinations between the independent and dependent variables deliver nonstationary residuals. Hence, the *NH* of this test is *no cointegration.* The equation used for residual unit root testing is:

$$\Delta v_{1,t} = (\rho - 1)v_{1,t-1} + \sum_{j=1}^{p} \delta_j \Delta v_{1,t-j} + \xi_t, \tag{7}$$

Table 3: Engle-Granger Cointegration Analysis (*)

Dependent variable: $\log(e_t^{i,j})$		UK FS	UK ES	Japan FS	Japan ES	Euro Zone FS	Euro Zone ES
EMH	τ-statistic	-3.797	-3.565	-1.614	-1.334	-1.349	-2.508
	p-value	0.015	0.029	0.717	0.821	0.816	0.284
	z-statistic	-20.290	-17.182	-4.963	-3.429	-4.067	-17.478
	p-value	0.050	0.094	0.734	0.850	0.803	0.075
	$\widehat{\rho}-1$	-0.051	-0.056	-0.012	-0.011	-0.023	-0.144
	$\widehat{\sigma}_{\widehat{\rho}}$	0.013	0.016	0.008	0.008	0.017	0.057
MF	τ-statistic	-3.457	-2.911	-1.682	-1.482	-2.947	-2.228
	p-value	0.039	0.137	0.687	0.770	0.128	0.414
	z-statistic	-23.113	-15.436	-4.236	-3.116	-14.927	-6.093
	p-value	0.027	0.131	0.791	0.870	0.142	0.634
	$\widehat{\rho}-1$	-0.059	-0.065	-0.011	-0.010	-0.081	-0.068
	$\widehat{\sigma}_{\widehat{\rho}}$	0.017	0.022	0.006	0.007	0.027	0.031
PPP	τ-statistic	-3.621	-3.510	-1.897	-1.262	-3.361	-1.570
	p-value	0.025	0.034	0.582	0.842	0.051	0.736
	z-statistic	-20.089	-17.033	-7.619	-3.948	-19.939	-3.932
	p-value	0.052	0.096	0.522	0.813	0.050	0.812
	$\widehat{\rho}-1$	-0.050	-0.056	-0.017	-0.013	-0.108	-0.044
	$\widehat{\sigma}_{\widehat{\rho}}$	0.014	0.016	0.009	0.010	0.032	0.028
Core	τ-statistic	-3.825	-3.753	-1.667	-1.375	-2.460	-3.651
	p-value	0.014	0.017	0.694	0.808	0.301	0.027
	z-statistic	-20.411	-17.911	-6.115	-4.452	-12.211	-23.133
	p-value	0.049	0.081	0.641	0.774	0.239	0.020
	$\widehat{\rho}-1$	-0.051	-0.059	-0.015	-0.015	-0.066	-0.260
	$\widehat{\sigma}_{\widehat{\rho}}$	0.013	0.016	0.009	0.011	0.027	0.071
P(Oil)	τ-statistic	-3.732	-3.488	-1.933	-1.203	-3.281	-2.292
	p-value	0.018	0.036	0.563	0.858	0.062	0.383
	z-statistic	-20.153	-16.982	-4.820	-3.755	-15.179	-6.100
	p-value	0.052	0.097	0.746	0.827	0.135	0.633
	$\widehat{\rho}-1$	-0.050	-0.056	-0.011	-0.010	-0.082	-0.069
	$\widehat{\sigma}_{\widehat{\rho}}$	0.013	0.016	0.006	0.009	0.025	0.030
Cmdty	τ-statistic	-2.570	-2.781	-2.744	-2.191	-2.849	-2.106
	p-value	0.253	0.177	0.188	0.431	0.156	0.476
	z-statistic	-12.115	-13.680	-14.892	-9.240	-13.070	-5.694
	p-value	0.247	0.181	0.146	0.401	0.204	0.668
	$\widehat{\rho}-1$	-0.045	-0.079	-0.047	-0.053	-0.071	-0.064
	$\widehat{\sigma}_{\widehat{\rho}}$	0.018	0.028	0.017	0.024	0.025	0.030

(*) FS: Full Sample. ES: Estimation Sample. *NH*: Series are not cointegrated. Cointegration equation deterministics: Constant. MacKinnon (1996) one-sided p-values. Highlighted cells: p-value>10%. Source: Authors' elaboration.

where $\Delta v_{1,t}$ are the first row of residuals of the log-level regression $\mathbf{z}_t = \boldsymbol{\varpi}_\mathbf{0} + \boldsymbol{\varpi}_\mathbf{1}'\mathbf{z}_{t-1} + \mathbf{v}_t$; ρ and δ_j are parameters to be estimated, and ξ_t is assumed a white noise. There are considered two statistics: a t-type statistic based on the hypothesis of nonstationarity (ρ=1), and the other on the normalised autocorrelation coefficient $\widehat{\rho}-1$. These statistics are labelled $\widehat{\tau}$ and $\widehat{z}$ and corresponds to:

$$\widehat{\tau} = \frac{\widehat{\rho}-1}{\widehat{\sigma}_{\widehat{\rho}}}, \text{ and } \widehat{z} = \frac{T(\widehat{\rho}-1)}{1-\sum_j \widehat{\delta}_j}. \tag{8}$$

The results for the case of the UK suggest that for all the models, except Cmdty, the variables are cointegrated at a confidence level of 5%. When cointegrated, just the MF model does it and with the full sample only–or at the non-conventional 13.7% level of confidence with the short sample. The rest of models are cointegrated even with the short sample. Note that the estimation of ρ seems to be surrounding 0.95 with all the models including Cmdty.

For the case of Japan, the result of no cointegration should not be surprising since previous unit root results suggest (Annex 2) that variables are already I(0). Hence, in this case the assumption of a restricted cointegration vector would not play any role since VECM's adjustment coefficient is (and should be) zero, and the model is driven by short-run dynamics. This apparently "no-result", however, is plainly endorsing the econometrics behind the forecasting exercise.

For the case of the Euro Zone, we find evidence of cointegration with the PPP and P(Oil) models, and for Core model just with the estimation sample. For both cases, PPP and P(Oil), the estimation of ρ is around 0.90. In these two cases the restriction on $\boldsymbol{\beta}$ will have a role as in both ensembles log-levels cointegrate.

Overall, we find a strong evidence of cointegration with the UK dataset, a weaker cointegration with Euro Zone models, and no cointegration with Japanese time series.

3.2 Out-of-sample results

These results comprise two kinds of analysis for each country: the RMSFE Ratio with statistical inference (for the two sample spans, ES.I-II), and the behaviour of forecast errors across time.

3.2.1 United Kingdom

The RMSFE results for the UK are presented in Table 4. It shows that the best model in ES.I is MF model for all the horizons. The rest of the models are not superior to the benchmark. The MF model shows predictive gains ranging from 1.2 at h=1 to 30.2% at h=60. The benchmark could, in principle, be improved for h=24 as the RMSFE (last column of Table 4) is greater for this horizon rather than the next one (11.799 versus 11.711), as evidence of forecast inefficiency.

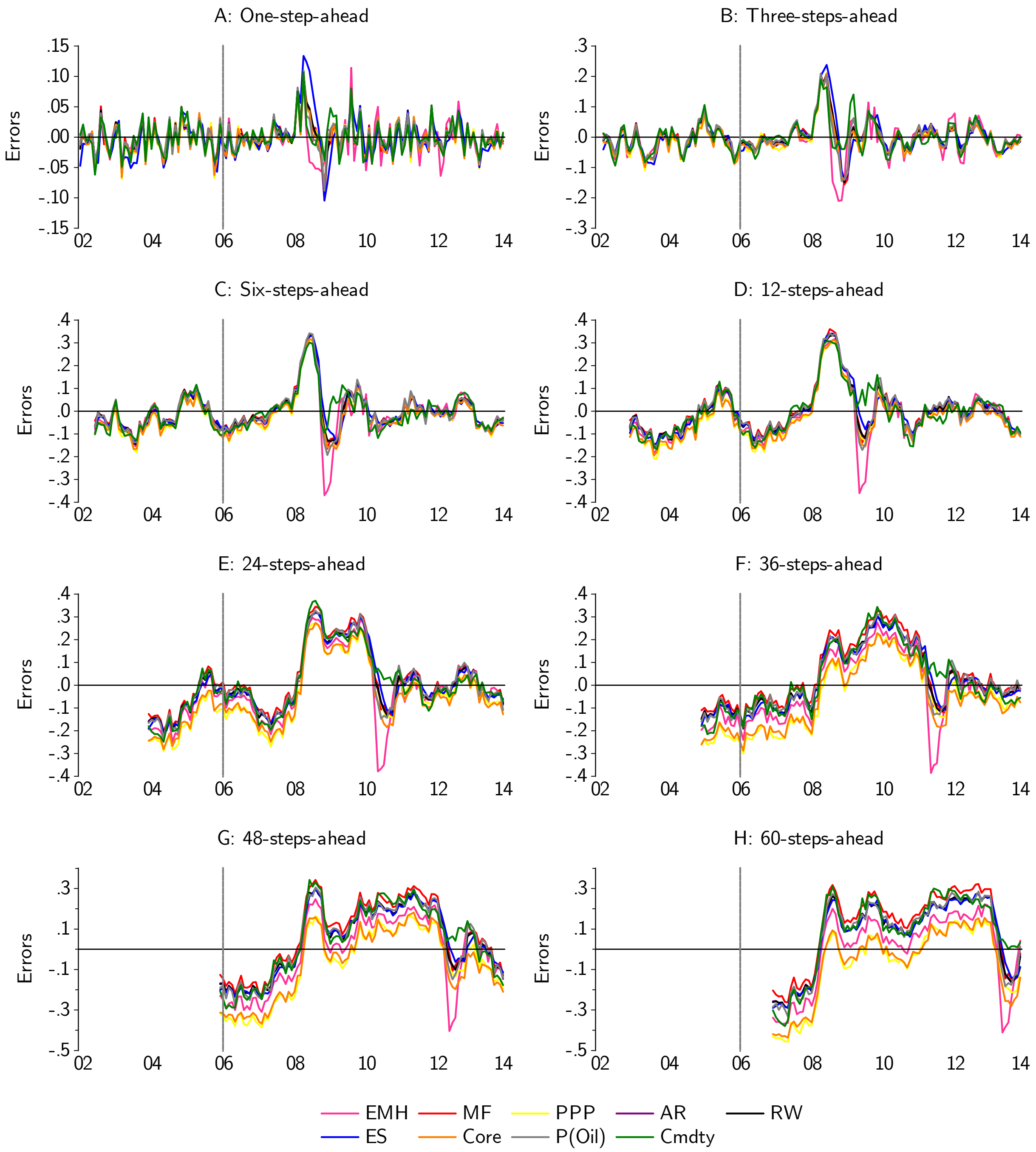
Table 4: UK. RMSFE Ratio Estimates (*)

	EMH	**MF**	**PPP**	**AR**	**ES**	**Core**	**P(Oil)**	**Cmdty**	**RW**
				Evaluation Sample I: 2002.6–2006.6					
h=1	1.000	0.988	1.086	1.004	1.140	1.084†	1.003	1.068	2.407
h=3	1.036	0.955	1.120†	1.006	1.052	1.128†	1.003	1.143†	3.945
h=6	1.062†	0.977	1.236†	1.002	1.019	1.202†	1.016	1.171†	6.273
h=12	1.104†	0.973	1.367†	1.000	1.036	1.295†	1.009	1.105	8.008
h=24	1.235†	0.934†	1.620†	0.997	1.043†	1.520†	1.051†	1.189†	11.799
h=36	1.434†	0.847†	2.010†	0.996	1.068†	1.888†	1.090	1.312†	11.711
h=48	1.403†	0.798	1.886	0.983	1.152	1.843†	1.080	1.323	17.102
				Evaluation Sample II: 2002.6–2014.6					
h=1	1.136	1.007	1.050	0.993	1.277	1.038	0.990	1.000	2.574
h=3	1.202	0.999	1.021	0.992	1.114	1.024	0.994	0.975	4.985
h=6	1.172	1.007	1.046	1.001	1.011	1.053	1.043	0.970	8.404
h=12	1.127	1.014	1.066	1.002	1.000	1.068	1.041	1.006	10.161
h=24	1.106	1.029	1.119	1.001	1.000	1.012	1.024	1.042	13.489
h=36	1.081	1.067	1.134	1.001	1.003	1.117	1.028	1.052	13.728
h=48	1.045	1.088	1.106	1.001	0.998	1.084	1.040	1.085	17.093
h=60	0.991	1.138	1.012	1.000	0.999	1.000	1.052	1.102	19.058

(*) Figures below unity imply a worst RW performance. For RW it is presented the RMSFE. (†) GW-test null hypothesis rejected at 10% of confidence level. Shaded cells indicate RMSFE Ratio<1. Source: Authors' elaboration.

When considering ES.II, there is no unique model outperforming at every single horizon. Instead, the AR and P(Oil), plus MF and Cmdty are the best at h=1 and 2, and h=2 and 3, respectively. Then, just since $h \geq 48$, one of the competing model beats the RW, the ES, for h=48 and 60, and the EMH at h=60 only. Statistical inference favouring a competing model is found just for MF at h=24 and 36 using ES.I.

Figure 4: UK. Forecasting Errors across Evaluation Sample (*)



(*) Vertical line: ES.I end and ES.II start point. Source: Authors' evaluation.

The forecasting errors across time for all the models and horizons are depicted in Figure 4. It is easy to notice that the FX disruption observed in 2008 was unpredictable for the UK. The forecast accuracy of all models seems similar from h=1 to 24, except for the EMH model when adapting to the two known breaks in the sample; same for the ES model but at h=1 only.

For the remaining horizons, the same performance of EMH is noticed. But more importantly, the PPP model shows the best performance during the disruption time, in a scenario in which all the models underpredict the FX. The PPP behaviour is particularly better at h=60, when a few observations even overpredict the FX during the crisis. This fact, while may not be desirable from the perspective of an investor, it is an important feature in terms of accuracy, unbiasedness, and variance. As the Core model follows closely the PPP model, the same performance is noticed for the former, outperforming the remaining models during the crisis.

3.2.2 Japan

The RMSFE results for Japan are presented in Table 5. When analysing the ES.I, the results suggest an astonishing performance of the EMF model, showing important predictive gains greater than 35% at h=36 and 60. For shorter horizons, the predictive gains oscillate between 1 and 9%. Note that in this case, the RMSFE profile of the RW model is specially inefficient at 24-months-ahead, describing clearly an nonadaptative forecasting behaviour of a hump. At h=12, the PPP, Core, and P(Oil) are also superior than the benchmark. Finally, at h=48 and 60 the AR model and specially the Core model show important predictive gains.

When considering ES.II, many of the models are superior to the RW since $h \geq 3$. Overall horizons, the best model is Core, showing predictive gains since $h \geq 3$, but beaten by the EMH and MF at h=60. Since $h \geq 6$, the P(Oil) model shows also important reductions in the RMSFE. Then, from $h \geq 24$ the EMH and MF models begin to outperform the RW. Both statistical models, AR and ES, are always close to the RW with ratios above or below unity.

Table 5: Japan. RMSFE Ratio Estimates (*)

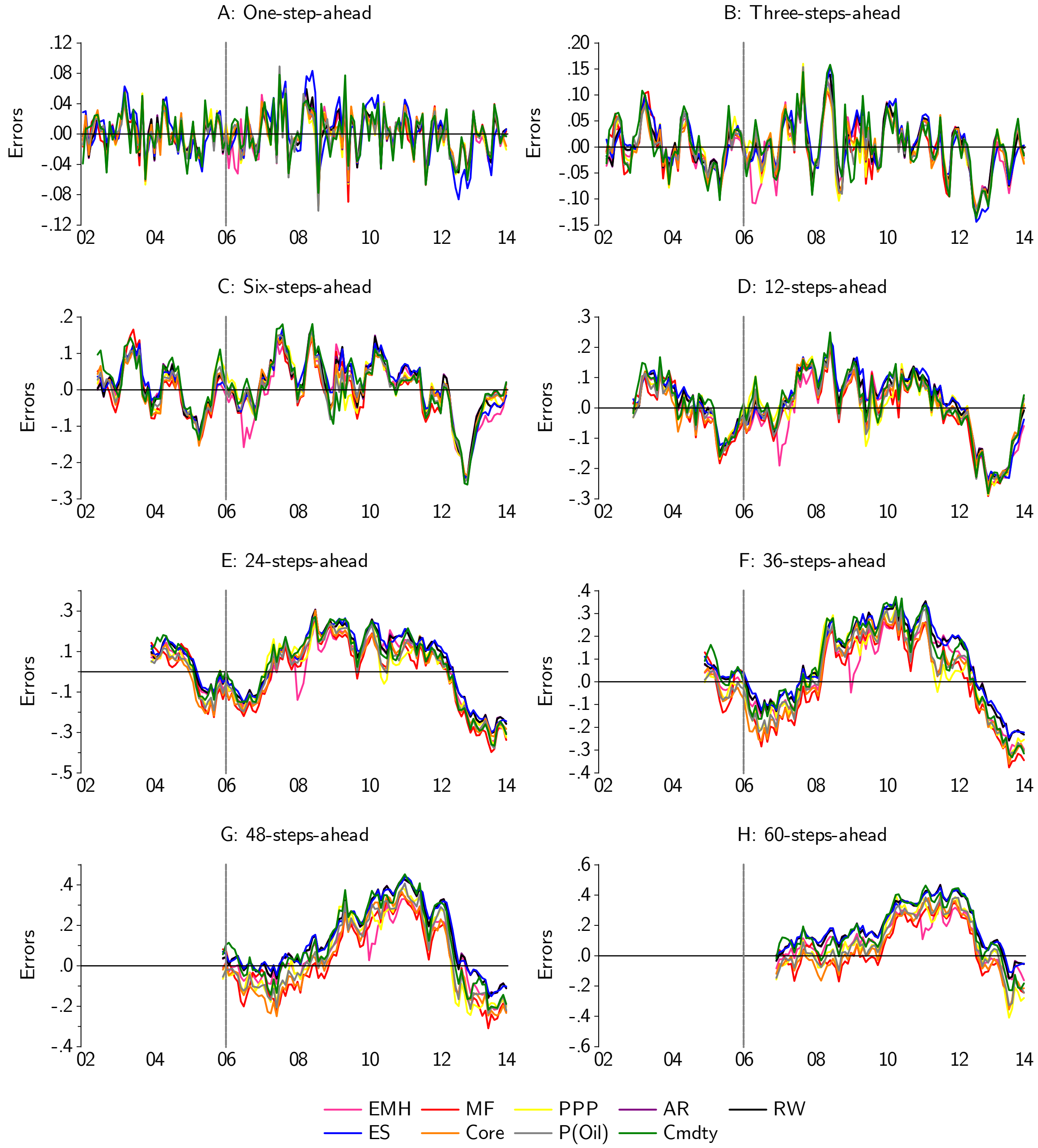
	EMH	**MF**	**PPP**	**AR**	**ES**	**Core**	**P(Oil)**	**Cmdty**	**RW**
					Evaluation Sample I: 2002.6–2006.6				
h=1	0.992	1.072	1.026	1.016	1.135	1.044†	1.012	1.221	2.200
h=3	0.983	1.209	1.067†	1.000	1.071	1.048†	1.031	1.308	3.644
h=6	0.995†	1.229	1.082†	1.005	1.000	1.039†	1.012	1.218†	5.908
h=12	0.995†	1.102	0.980†	1.004	1.000	0.977†	0.957	1.102	7.543
h=24	0.913†	1.252†	1.047†	1.001	1.008	1.210†	1.033†	1.194†	9.493
h=36	0.650†	1.783†	0.687†	0.995	1.127†	1.146†	0.943†	1.657†	4.814
h=48	0.320†	1.933	1.134†	0.956	1.571†	0.413†	1.002	2.060	4.079
					Evaluation Sample II: 2002.6–2014.6				
h=1	1.045	1.031	1.019	1.002	1.203†	1.018	1.046	1.118†	2.627
h=3	1.048	1.028	1.013	0.999	1.067	0.969	1.001	1.134†	4.843
h=6	1.053	1.021	1.003	1.002	1.018	0.937†	0.977	1.066	7.474
h=12	1.026	1.000	1.012	0.999	1.039†	0.944†	0.977	1.036	10.137
h=24	0.951	1.026	0.971	0.999	1.013	0.985	0.985	1.061	15.641
h=36	0.861†	0.962	0.966	1.000	1.005	0.973	0.957	1.051	18.299
h=48	0.754†	0.847†	0.934	1.000	1.005	0.893	0.896†	1.021	21.156
h=60	0.682†	0.736†	0.862†	1.000	0.999	0.789†	0.819†	0.998	24.219

(*) See notes in Table 4. Source: Authors' elaboration.

Statistical inference suggests that with ES.I, there are significant gains with EMH model at long-run horizons; same with Core model. When considering ES.II, at long horizons we find also significant differences favouring economics models. At intermediate horizons, only the Core model exhibit statistically significant gains at 10%.

The forecasting errors across the time are depicted in Figure 5. From h=1 to 12-steps-ahead the errors seems following a white noise behaviour with all models, reflecting a desirable efficiency characteristic. At h=1 the ES model tends to exaggerate the dynamics of the series delivering errors greater than the remaining models. At h=6 and 12, the EMH model overpredict the Japanese FX dynamics during 2007, and at the end of the sample. At short-run, there is no identifiable best or worst model. When predicting at long run, there is a noticeable positive error bias. Two models exhibit an overprediction of FX at h=24 and 36: EMH and PPP. For h=48 and 60, the EMH (again) plus the MF model captures in several observations best FX dynamics during the crisis.

Figure 5: Japan. Forecasting Errors across Evaluation Sample (*)



(*) See notes in Figure 4. Source: Authors' evaluation.

3.2.3 Euro Zone

The RMSFE results for the Euro Zone are exhibited in Table 6. With ES.I there are two models that outperform the RW: MF and AR. For both this behaviour is observed for $h>3$. At $h=12$ the EMH and PPP at $h=48$ also show a superior performance compared to the RW. When considering ES.II the AR model shows more accurate predictions at $h=24$, 48, and 60. However, the P(Oil) model is consistently superior than AR (and the RW) since $h \geq 12$. Note that in only one case (ES.I and $h=48$: PPP) the predictive gains are statistically significant in favour of the candidate model. No significant gains are obtained with the ES.II.

The forecast errors across time are presented in Figure 6. In the short run, there is an overprediction outlier of FX in mid-2008 that shared across all the models. At $h=6$ and 12, there is a W-shape dynamics due in part to an expected weak European economy in 2010 that finished with the ECB announcement of a raise in the interest rate in October 2011. The results for longer horizons exhibit an unclear bias, reflecting the lower variability obtained for statistical inference. It is remarkable, however, that the PPP model shows more volatile errors and exaggerating the dynamics of FX, whereas EMH model follows FX closely. The ES, P(Oil), and Cmdty models appear as a good alternative when predicting at the long-run.

Table 6: Euro Zone. RMSFE Ratio Estimates (*)

	EMH	**MF**	**PPP**	**AR**	**ES**	**Core**	**P(Oil)**	**Cmdty**	**RW**
	Evaluation Sample I: 2002.6–2006.6								
h=1	1.035	1.040	1.064	1.032	1.230†		1.039	1.042	2.626
h=3	1.023	1.017	1.081†	1.018	1.087†		1.033	1.032	4.961
h=6	1.009	0.992	1.111	0.995	1.046		1.023	1.011	7.954
h=12	0.998	0.996	1.139†	0.982	1.069		1.018	1.002	9.870
h=24	1.041	0.980	1.105	0.979	1.066†		1.061	1.022	14.845
h=36	1.106	0.938	1.066	0.970†	1.097†		1.113	1.064	15.375
h=48	1.201	0.960	0.834†	0.962	1.112		1.133	1.176	25.466
	Evaluation Sample II: 2002.6–2014.6								
h=1	1.032	1.025†	1.044†	1.020	1.144†		1.031	1.033	3.042
h=3	1.034	1.010	1.064†	1.005	1.079†		1.010	1.011	5.112
h=6	1.043	1.009	1.106†	1.003	1.009		1.005	1.003	8.283
h=12	1.052	1.016	1.178†	1.003	0.987		1.004	0.999	9.689
h=24	1.081	1.073	1.320†	0.997	1.016		1.022	0.989	11.515
h=36	1.090	1.175	1.583†	1.002	1.003		1.020	0.976	11.102
h=48	1.131	1.188	1.613†	0.996	1.015		1.007	0.957	14.842
h=60	1.173†	1.228	1.642†	0.991	1.030†		1.022	0.969	17.259

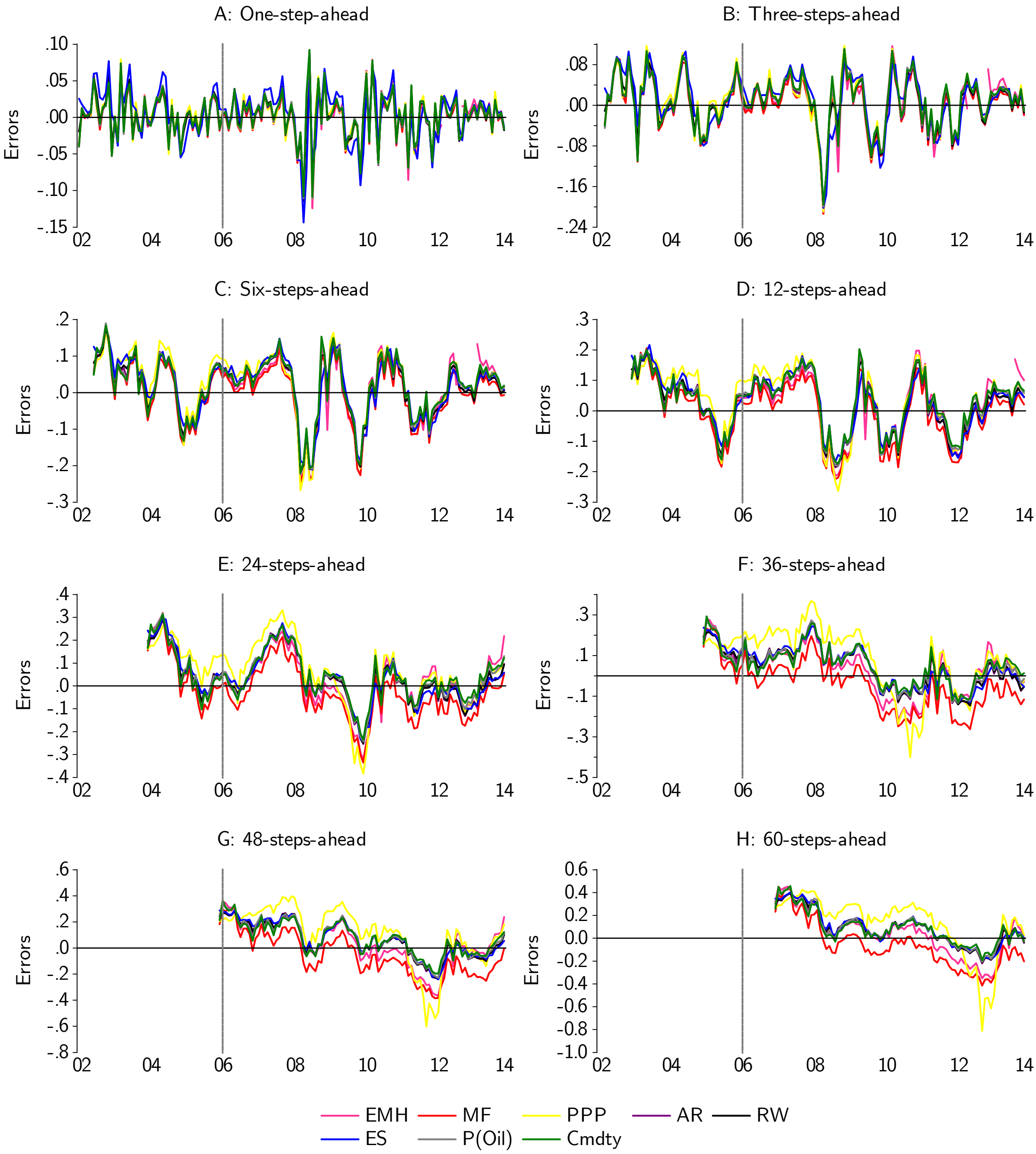
(*) See notes in Table 4. Source: Authors' elaboration.

4 Summary and concluding remarks

The aim of this article is to analyse the out-of-sample behaviour of a bunch of statistical and economics-based models when forecasting FX rates for certain countries especially during the commodity prices boom of 2007-8 and the financial crisis of 2008-9.

We rely on the GL analysis which comprises three economic plus two statistical models evaluated with both financial and statistical criteria. The GL FX dataset includes UK and Japan in relation to the US. We propose and analyse several modelling extensions to the GL article beyond just a sample increment. In particular, we include the Euro Zone, a forecast horizon of 60-steps-ahead, and three economic plus one modelling extension into analysis. Our modelling extensions are made through one of the three economic models of GL, the PPP, which is analysed along with EMH and MF models. As the PPP model stands for domestic/foreign prices gap, we include a version based in core inflation. Also, we use the Brent price of oil and IMF's commodity price index as a global price measures. We evaluate the models statistically regardless of the financial evaluation proposed in GL.

Figure 6: Euro Zone. Forecasting Errors across Evaluation Sample (*)



(*) See notes in Figure 4. Source: Authors' evaluation.

The sample extension covers from 2006.6 until 2014.6; a period characterised of high volatility and breaks. Hence, it is expected for forecasting accuracy to suffer changes and to that end we also analyse the forecasting errors across the time.

Our results indicate that there are indeed changes of the first best models when considering the longer sample span. For the three countries–the UK, Japan, and the Euro Zone–none of the models that outperform in the GL sample remain as the best after the two aforementioned disruptions.

The results are mixed between statistical and economic models. Our results for the UK case indicate that the MF and AR outperform considering the GL sample span. When considering the whole evaluation sample, again the AR and now the P(Oil) model exhibit the best performance in the short run. Then, all the models exhibit similar accuracy with ES and EMH being the best alternatives at 60-steps-ahead. During the period of higher volatility, the PPP model predicts better the FX dynamics.

For the case of Japan, the EMH model is the best alternative previous to 2007. When including the last part of the sample, the Core model as well as the remaining economic models shows the higher accuracy. The best model with higher FX volatility results is the EMH.

Finally, for the Euro Zone there are two models outperforming in the first evaluation sample, MF and AR. When considering the whole sample, the Cmdty model outperforms the RW benchmark at six-steps-ahead onwards. The PPP and P(Oil) models exhibit promising results when predicting at long-run. As in the previous case, the best model when FX volatility is high is the EMH.

These results are important since they reveal the accuracy and confidence magnitudes when forecasting an important variable for policymaking as the FX is. Also, they provide robustness and insights for a closer replication and extensions. This would be the case for the financial evaluation proposed in GL, suggested for further research.

References

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

2. Andreou, A.S., E.F. Georgpoulus, and S.D. Likothanassis (2002). "Exchange-Rates Forecasting: A Hybrid Algorithm Based on Genetically Optimized Adaptative Neural Networks." *Computational Economics* **20**(3): 191-210.

3. Barberis, N. (2000). "Investing for the Long Run When Returns are Predictable." *Journal of Finance* **LV**(1): 225-264.

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

5. Blanchard, O.J. and S. Fischer (1989). *Lectures on Macroeconomics*. MIT Press, USA.

6. Boothe, P. and D. Glassman (1987). "Comparing Exchange Rate Forecasting Models: Accuracy versus Profitability." *International Journal of Forecasting* **3**(1): 65-79.

7. Brzeszcynski, J. and M. Melvin (2006). "Explaining Trading Volume in the Euro." *International Journal of Finance and Economics* **11**(1): 25-34.

8. Cheung, Y.-W. and M.D. Chinn (1998). "Integration, Cointegration, and the Forecast Consistency of Structural Exchange Rate Models." *Journal of International Money and Finance* **17**: 813-830.

9. Cheung, Y.-W., M.D. Chinn, and A. García Pascual (2005). "An Economic Evaluation of Empirical Exchange Rate Models." *Review of Financial Studies* **22**(7): 3491-3530.

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

11. Chinn, M.D. and R.A. Meese (1995). "Banking on Currency Forecasts: How Predictable is Change in Money?" *Journal of International Economics* **38**(1-2): 161-178.

12. Clarida, R.H. and M.P. Taylor (1997). "The Term Structure of Forward Premiums and the Forecastability of Spot Exchange Rates." *Review of Economics and Statistics* **89**: 353-361.

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

14. Dal Bianco, M., M. Camacho, and G. Pérez-Quirós (2012). "*Short-Run Forecasting of the Euro-Dollar Exchange Rate with Economic Fundamentals*." Documento de Trabajo 1203, Banco de España.

15. Della Corte, P., L. Sarno, and I. Tsiakas (2009). "An Economic Evaluation of Empirical Exchange Rate Models." *Review of Financial Studies* **22**: 3491-3530.

16. Dornbusch, R. (1976). "Expectations and Exchange Rate Dynamics." *Journal of Political Economy* **84**(6): 1161-1176.

17. Dornbusch, R., S. Fischer, and R. Startz (2010). *Macroeconomics*. Eleventh Edition. McGraw-Hill/Irwin, USA.

18. Engel, C. and J. Hamilton (1990). "Long Swings in the Exchange Rate: Are They in the Data and Do Markets Know It?" *American Economic Review* **80**(4): 689-713.

19. Engel, C., N.C. Mark, and K.D. West (2007). *Exchange Rate Models are Not as Bad as You Think*. *NBER Macroeconomics Annual*: 381-441.

20. Estrella, A. and F.S. Mishkin (1996). "The Yield Curve as a Predictor of US Recessions." *Current Issues in Economics and Finance* **2**(7).

21. Evans, M.D.D. (2011). *Exchange Rate Dynamics*. Princeton University Press, Princeton, USA.

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

23. Fat, C.M. and E. Deszi (2011). "Exchange-Rates Forecasting: Exponential Smoothing Techniques and ARIMA Models." *Annals of Faculty of Economics - Babes-Bolyai University* **1**(1): 499-508.

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

25. Frenkel, J. (1979). "On the Mark: A Theory of Floating Exchange Rates Based on Real Interest Differentials." *American Economic Review* **69**(4): 610-622.

26. Garratt, A., K.C. Lee, M.H. Pesaran, and Y. Shin (2006). *Global and National Macroeconometric Modelling: A Long-Run Structural Approach*. Oxford University Press, Oxford, UK.

27. Garratt, T. and K.C. Lee (2010). "Investing under Model Uncertainty: Decision based Evaluation of Exchange Rate Forecasts in the US, UK and Japan." *Journal of International Money and Finance* **29**(3): 403-422.

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

29. Goodfriend, M. (2008). "The Case for Price Stability with a Flexible Exchange Rate in the New Neoclassical Synthesis." *Cato Journal* **28**(2): 247-254.

30. Granger, C.W.J. and M.H. Pesaran (2000). "Economic and Statistical Measures of Forecast Accuracy." *Journal of Forecasting* **19**: 537-560.

31. Hakkio, C.S. (2009). "*Global Inflation Dynamics*." Working Paper 09-01, Federal Reserve Bank of Kansas City.

32. Hansen, P.R. (2009). "*In-Sample Fit and Out-of-Sample Fit: Their Joint Distribution and its Implications for Model Selection*." Preliminary Version of 19 June, 2009. Department of Economics, Stanford University, USA.

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

34. Hyndman, R.J., A.B. Koehler, J.K. Ord, and R.D. Snyder (2008). *Forecasting with Exponential Smoothing. The State Space Approach*. Springer Series in Statistics, USA.

35. Kandel, S. and R.F. Stanbaugh (1996). "On the Predictability of Stock Returns: An Asset-Allocation Perspective." *Journal of Finance* **LI**(2): 385-424.

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

37. Kim, B.J.C. and S. Mo (1995). "Cointegration and the Long-Run Forecast of Exchange-Rates." *Economics Letters* **48**(3-4): 353-359.

38. Kirikos, D.G. (2013). "How Did the Financial Crisis Affect the Forecasting Performance of Time Series Exchange Rate Models? Evidence from Euro Rates," *International Journal of Economics and Finance* **5**(2): 28-34.

39. Leitch, G. and J.E. Tanner (1991). "Economic Forecasts Evaluation: Profits versus Conventional Measures." *American Economic Review* **81**(3): 580-590.

40. MacDonald, R. and M.P. Taylor (1994). "The Monetary Model of the Exchange Rate: Long-Run Relationships, Short-Run Dynamics and How to Beat a Random Walk." *Journal of International Money and Finance* **13**(3): 276-290.

41. MacKinnon, J.G. (1996). "Numerical Distribution Functions for Unit Root and Cointegration Tests." *Journal of Applied Econometrics* **11**(6): 601-618.

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

43. Mark, N.C. and D. Sul (2001). "Nominal Exchange Rates and Monetary Fundamental: Evidence from a Small Post-Bretton Woods Panel." *Journal of International Economics* **53**(1): 29-52.

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

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

46. Melvin, M. and M.P. Taylor (2009). "The Crisis in the Foreign Exchange Market." *Journal of International Money and Finance* **28**(8): 1317-1330.

47. Melvin, M., J. Prins, and D. Shand (2013). *Forecasting Exchange Rates: An Investor Perspective*. In G. Elliot and A. Timmermann (2013). *Handbook of Economic Forecasting. Volume 2B*. Elsevier, North-Holland.

48. Møller, S.V. (2014). "GDP Growth and the Yield Curvature." *Finance Research Letters* **11**(1): 1-7.

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

50. Mussa, M. (1976). "The Exchange Rate, the Balance of Payments, and the Monetary and Fiscal Policy under a Regime of Controlled Floating." *Scandinavian Journal of Economics* **78**: 229-248.

51. Newey, W.K. and K.D. West (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* **55**(3): 703-708.

52. Owen, D. (2001). "*Importance of Productivity Trends for the Euro. European Economics for Investors.*" Desdner Kleinwort Wasserstein.

53. Pesaran, M.H. and A. Timmermann (1995). "Predictability of Stock Returns: Robustness and Economic Significance." *Journal of Finance* **L**(4): 1201-1228.

54. Rosemberg, M. (2000). "*The Euro's Long-Term Struggle.*" *FX Research Special Report Series 2*. Deustche Bank, London.

55. Sarno, L. and M.P. Taylor (2002). *The Economics of Exchange Rates*. Cambridge University Press, Cambridge, UK.

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

57. Taylor, M.P. (1995). "The Economics of Exchange Rates." *Journal of Economic Literature* **33**(1): 13-47.

58. Taylor, M.P. (1989). "Covered Interest Arbitrage and Market Turbulence." *Economic Journal* **99**(396): 376-391.

59. West, K.D., H.J. Edison, and D. Cho (1993). "A Utility-Based Comparison of Some Models of Exchange Rate Volatility." *Journal of International Economics* **35**: 23-45.

60. Williamson, J. (2009). "Exchange Rate Economics." *Open Economies Review* **20**(1): 123-146.

A Data description and sources

In this Annex it is described the dataset in terms of its sources for further replication/checking purposes.

Table A1: Variable Description (*)

Variable	Country	Unity	Scale	Descriptor	IFS Code
Foreign exchange rate	UK	Nat. curr. per USD	None	Market rate, eop	112..AG.ZF...
	JPN	Nat. curr. per USD	None	Market rate	158..AE.ZF...
	EZ	Nat. curr. per USD	None	Market rate, eop	163..AE.ZF...
Interest rates	US	Percent per annum	None	Treasury Bill Rate	11160C..ZF...
	UK	Percent per annum	None	Treasury Bill Rate	11260C..ZF...
	JPN	Percent per annum	None	Discount rate, eop	15860.A.ZF...
	EZ	Percent per annum	None	Interbank rate (3 mths.)	16360B..ZF...
Industrial production	US	Index number	2010=100	Industrial prod. sa	11166..CZF...
	UK	Index number	2010=100	Industrial prod. sa	11266..CZF...
	JPN	Index number	2010=100	Industrial prod. sa	15866..CZF...
	EZ	Index number	2010=100	Industrial prod. sa	16366..CZF...
Consumer price index	US	Index number	2010=100	CPI all items city ave.	11164...ZF...
	UK	Index number	2010=100	CPI: all items	11264...ZF...
	JPN	Index number	2010=100	CPI: all JPN-588 items	15864...ZF...
	EZ	Index number	2010=100	Consumer prices	16364H..ZF...
Money (M1)	US	USD	Billions	M1 sa	11159MACZF...
	UK	Index number	2010=100	Narrow money, sa	*OECD*
	JPN	National currency	Trillions	M1, sa	15859MACZF...
	EZ	Billions	Blls. EUR	M1	16359MAUZF...
Core inflation	US	Index number	2010=100	CPI non food and engy.	*OECD*
	UK	Index number	2010=100	CPI non food and engy.	*OECD*
	JPN	Index number	2010=100	CPI non food and engy.	*OECD*
Miscellaneous		Index number	None	UK Brent oil price	11276AADZF...
		Index number	None	All commodities index	00176ACDZF...

(*) "eop" stands for end-of-period. "sa" stands for seasonally adjusted. Source: Authors' elaboration.

B Descriptive statistics: different samples

Table A2: Descriptive Statistics of Foreign Exchange Rates (*)

	Mean	Median	St. dev.	Min.	Max.	ADF-Stat.	*p*-value
				Full Sample: 1981.1–2014.6			
UK							
Log-level	-0.490	-0.476	0.110	-0.871	-0.086	-1.250	0.195
Δlog	0.085	0.000	2.965	-12.893	12.554	-21.024	0.000
Japan							
Log-level	-4.839	-4.779	0.310	-5.625	-4.335	-1.806	0.068
Δlog	0.173	-0.131	3.162	-11.392	15.009	-21.201	0.000
Euro Zone							
Log-level	0.190	0.242	0.162	-0.172	0.458	-0.648	0.435
Δlog	0.098	0.028	3.013	-11.439	8.938	-13.171	0.000
				Evaluation Sample I: 2002.6–2006.6			
UK							
Log-level	-0.551	-0.571	0.067	-0.658	-0.432		
Δlog	-0.447	-0.557	2.471	-5.540	4.827		
Japan							
Log-level	-4.728	-4.721	0.051	-4.808	-4.636		
Δlog	0.161	0.123	2.293	-5.540	5.127		
Euro Zone							
Log-level	0.164	0.186	0.087	-0.022	0.309		
Δlog	0.619	0.274	2.713	-4.952	6.424		
				Evaluation Sample II: 2006.7–2014.6			
UK							
Log-level	-0.519	-0.482	0.105	-0.728	-0.351		
Δlog	0.077	0.000	2.650	-9.038	10.536		
Japan							
Log-level	-4.555	-4.561	0.140	-4.814	-4.335		
Δlog	0.132	0.426	2.806	-8.501	6.937		
Euro Zone							
Log-level	0.310	0.301	0.059	0.205	0.458		
Δlog	0.075	0.405	3.224	-11.439	8.938		

(*) ADF-test assuming a constant in the cointegrating equation. Lag-length criteria: BIC. Source: Authors' elaboration using data from IFS and OECD.