

Forecasting errors, directional accuracy and profitability of currency trading: The case of EUR/USD exchange rate

Costantini, Mauro; Crespo Cuaresma, Jesus; Hlouskova, Jaroslava

Published in:
Journal of Forecasting

DOI:
[10.1002/for.2398](https://doi.org/10.1002/for.2398)

Published: 01/01/2016

Document Version
Peer reviewed version

[Link to publication](#)

Citation for published version (APA):
Costantini, M., Crespo Cuaresma, J., & Hlouskova, J. (2016). Forecasting errors, directional accuracy and profitability of currency trading: The case of EUR/USD exchange rate. *Journal of Forecasting*, 35(7), 652 - 668. <https://doi.org/10.1002/for.2398>

Forecasting errors, directional accuracy and profitability of currency trading: The case of EUR/USD exchange rate*

Mauro Costantini[†]

Jesus Crespo Cuaresma[‡]

Jaroslava Hlouskova[§]

Abstract

We provide a comprehensive study of out-of-sample forecasts for the EUR/USD exchange rate based on multivariate macroeconomic models and forecast combinations. We use profit maximization measures based on directional accuracy and trading strategies in addition to standard loss minimization measures. When comparing predictive accuracy and profit measures, data snooping bias free tests are used. The results indicate that forecast combinations, in particular those based on principal components of forecasts, help to improve over benchmark trading strategies, although the excess return per unit of deviation is limited.

JEL codes: C53, F31, F37

Keywords: Exchange rate forecasting, forecast combination, multivariate time series models, profitability

*The research in this paper has been supported by the Anniversary Fund of the Austrian Central Bank (Project No. 15308). The authors would like to thank Robert Kunst and the participants of an internal seminar at the Vienna University of Economics and Business for thoughtful comments and remarks. An earlier version of this piece with a broader exchange rate coverage was circulated as Costantini, Mauro, Crespo Cuaresma, Jesus and Hlouskova, Jaroslava (2014) Can Macroeconomists Get Rich Forecasting Exchange Rates? *Department of Economics Working Paper Series, 176. WU Vienna University of Economics and Business, Vienna.*

[†]Department of Economics and Finance, Brunel University, London, UK

[‡]Department of Economics, Vienna University of Economics and Business; Wittgenstein Centre for Demography and Global Human Capital (WIC); World Population Program, International Institute of Applied Systems Analysis (IIASA) and Austrian Institute for Economic Research (WIFO), Vienna, Austria

[§]Department of Economics and Finance, Institute for Advanced Studies, Vienna, Austria, and Department of Economics, Thompson Rivers University, Kamloops, BC, Canada

1 Introduction

Forecasting exchange rates is a notoriously difficult task. Myriads of empirical studies (see for example the recent survey by James et al., 2012) document the challenges associated with specifying macro-econometric models aimed at exchange rate forecasting with satisfactory out-of-sample predictive performance, in particular for short-run forecasting horizons.

Since the seminal work by Meese and Rogoff (1983), which shows that specifications based on macroeconomic fundamentals are unable to outperform simple random walk forecasts, a large number of studies have proposed models aimed at providing accurate out-of-sample predictions of spot exchange rates (see MacDonald and Taylor, 1994; Mark, 1995; Chinn and Meese, 1995; Kilian, 1999; Mark and Sul, 2001; Berkowitz and Giorgianni, 2001; Cheung et al., 2005; Boudoukh et al., 2008, among others). In parallel, a literature has emerged which examines empirically the potential profitability of technical trading rules (see Menkhoff and Taylor, 2007, for a review). The analysis of profitability of technical trading rules can be thought of as a simple and robust test for the weak form of the efficient market hypothesis, which concludes that if the foreign exchange market is efficient, one should not be able to use publicly available information to correctly forecast changes in exchange rates and thus make a systematic profit.

The aim of this paper is to offer the most comprehensive analysis to date of the out-of-sample forecasting performance of macro-econometric models of the exchange rate between the euro (EUR) and the US dollar (USD) in terms of predictive error, directional accuracy and profitability of trading strategies. The EUR/USD exchange rate plays a central role in global financial markets and understanding its dynamics has been the goal of countless studies (see, for example, Dal Bianco et al., 2012 or King et al., 2012). Since both currencies act as an anchor or reference currency and are used as an international store of value (Dal Bianco et al., 2012), the EUR and USD are the most traded currencies worldwide by value.

Models for the EUR/USD exchange rate based on macroeconomic fundamentals are used in Jamaleh (2002), Karfais (2006), Molodtsova et al. (2011), and Dal Bianco et al. (2012), whereas Sosvilla-Rivero and Garcí (2005) and Hsing and Sergi (2009) apply frameworks based on the purchasing power parity (PPP) equilibrium and Heimonen (2009) considers an equity flow model.¹ In most of these works predictive accuracy is assessed using standard loss accuracy measures, encompassing testing procedures (Jamaleh, 2002; Dal Bianco et al., 2012), and metrics to detect the percentage of correct predictions of the direction of changes (Jamaleh, 2002; Sosvilla-Rivero and Garcí, 2005; Karfakis, 2006; Heimonen, 2009; Dal Bianco et al., 2012). While the findings of the literature show a tendency for the benchmark random walk model to be outperformed, the predictive superiority of alternative modelling frameworks is

¹Sosvilla-Rivero and Garcí (2005) use an expectation version of the relative PPP that integrates the parity conditions of both commodity and financial markets. Hsing and Sergi (2009) apply different versions of PPP along with a Mundell-Fleming model.

not properly assessed in terms of multiple hypothesis, which implies that the results may be generated by chance. A remarkable exception is Molodtsova et al. (2011), where the test of superior predictive ability (SPA) proposed by Hansen (2005) is employed to assess whether the predictive superiority of certain models is systematic and not due to luck.

In our analysis, we overcome some of the limitations of the existing literature on EUR/USD exchange rate modelling. First, in addition to loss and directional accuracy measures with respect to the benchmark random walk model, we also employ profitability measures by comparing the (risk adjusted) profits generated by forecast-based trading strategies to those using benchmark trading rules where forecasts are not used. Second, the statistical significance of differences in predictive accuracy and profit measures is assessed using the StepM-RC test by Romano and Wolf (2005) and the StepM-SPA test by Hsu et al. (2010), in addition to the RC test of White (2000) and the SPA test by Hansen (2005). One drawback of both RC and SPA tests is that they do not aim at explicitly identifying the models which outperforms the benchmark. The tests proposed by Romano and Wolf (2005) and Hsu et al. (2010) overcome such a problem. Third, we use recently developed directional forecast accuracy measures along with standard loss measures based on prediction errors. The directional accuracy measures account for both the realized directional changes in exchange rates as well as for their magnitudes (see Blaskowitz and Herwartz, 2009, 2011; Bergmeir et al., 2014). Such measures are robust to outliers and provide an economically interpretable loss/success functional framework in a decision-theoretical context, which is extremely relevant for traders and investors. Fourth, we exploit the potential of a large number of forecast combination methods available in the literature and propose a new method of combination based on the economic evaluation of directional forecasts. The other methods of combination used are the mean, median or trimmed mean, the ordinary least squares combining methods, combinations based, on principal components, on discounted mean square forecast errors, on hit rates and on Bayesian and frequentist model averaging techniques.

Exploiting monthly data ranging back to 1980 and considering forecast horizons ranging from one-month-ahead to one-year-ahead, our analysis delivers four main results. First, the use of methods of forecast combination, and in particular those based on principal components, improves the predictive performance of exchange rate models substantially over single multivariate time series specifications. Second, concerning the differences between specifications based on monetary and capital flows modelling frameworks, there is no clear-cut superiority of one of the approaches over the other. The forecast performance of specifications based on the capital flows model tends to be slightly better than those of the monetary model for shorter forecasting horizons, while specifications with variables based on the monetary model tend to be better for long term predictions. Third, our empirical results tend to favour macroeconomic models in comparison with their benchmark specifications when looking at directional accuracy and profitability measures, whereas this is not true for standard loss

accuracy measures. Fourth, findings based on the Sharpe ratios reveals that the margin for achieving systematic profits in the EUR/USD market exploiting the information contained by macroeconomic variables is rather limited.

The rest of the paper is organized as follows. Section 2 describes the analytical framework used, including the forecast combination approaches and the forecast accuracy measures, as well as the trading strategies they are based on. In section 3, the design of the empirical exercise and the testing procedures for data snooping biases are presented. The results are discussed in section 4 and section 5 concludes.

2 Analytical framework

2.1 The monetary model of exchange rates

The theoretical framework of the monetary model of exchange rate formation (for the original formulations, see Frenkel, 1976; Dornbusch, 1976; Hooper and Morton, 1982) has become the workhorse for constructing macroeconomic models aimed at exchange rate prediction. Let real money demand in the domestic and foreign economies be given by log-linear functions,

$$M_t^d - P_t^d = \beta^d Y_t^d - \gamma^d i_t^d, \quad (1)$$

$$M_t^f - P_t^f = \beta^f Y_t^f - \gamma^f i_t^f, \quad (2)$$

where M_t refers to (log) nominal money demand, P_t is (the log of) the price level, Y_t is (log) income and i_t is the interest rate. The superindices d and f identify the parameters and variables of the domestic and foreign economy, respectively. If the (long-run) equilibrium exchange rate is assumed to be given by purchasing power parity, then

$$s_t = P_t^d - P_t^f, \quad (3)$$

where s_t denotes the (log) nominal exchange rate; i.e., $s_t = \log(S_t)$ and thus S_t is the exchange rate of the domestic currency against the foreign currency. Combining equation (1) and (2) with the equilibrium condition given by equation (3) results in

$$s_t = M_t^d - M_t^f + \beta^f Y_t^f - \beta^d Y_t^d + \gamma^d i_t^d - \gamma^f i_t^f, \quad (4)$$

a specification that suggests a relationship between the exchange rate, the money stock, output and interest rates. The empirical literature on exchange rate modelling and forecasting based on the monetary model of exchange rate determination often combines these variables in the form of vector autoregressive (VAR) models, so that

$$x_t = \psi(0) + \sum_{l=1}^p \psi(l)x_{t-l} + \varepsilon_t, \quad \varepsilon_t \sim \mathbf{NID}(\mathbf{0}, \Sigma_\varepsilon), \quad (5)$$

where $\psi(l)$ ($l = 1, \dots, p$) are matrices of coefficients. The x_t vector in our model is composed by the corresponding exchange rate (s_t), an output measure for the domestic and foreign economy (Y_t^d and Y_t^f), money supply² in the domestic and foreign economy (M_t^d and M_t^f), as well as short and long-term interest rates in both countries ($i_t^{s,d}$, $i_t^{l,d}$, $i_t^{s,f}$ and $i_t^{l,f}$). If the variables of the model are linked by some cointegration relationship, the model in (5) can be written as a vector error correction (VEC) model

$$\Delta x_t = \delta(0) + \alpha\beta'x_{t-1} + \sum_{l=1}^p \delta(l)\Delta x_{t-l} + \varepsilon_t, \quad \varepsilon_t \sim \mathbf{NID}(\mathbf{0}, \Sigma_\varepsilon), \quad (6)$$

where the cointegration relationships are given by $\beta'x_t$ and α measures the speed of adjustment to the long run equilibrium. Alternatively, if the variables in x_t are unit-root nonstationary but no cointegration relationship exists among them, a VAR model in first differences (DVAR) would be the appropriate representation,

$$\Delta x_t = \psi(0) + \sum_{l=1}^{p-1} \psi(l)\Delta x_{t-l} + \varepsilon_t, \quad \varepsilon_t \sim \mathbf{NID}(\mathbf{0}, \Sigma_\varepsilon). \quad (7)$$

If the income and interest rate elasticities of money demand are assumed equal for the domestic and foreign economy, the multivariate models above can be rewritten using vectors of differences in the variables, so that $x_t = (s_t, m_t, y_t, i_t^s, i_t^l) = (s_t, M_t^d - M_t^f, Y_t^d - Y_t^f, i_t^{s,d} - i_t^{s,f}, i_t^{l,d} - i_t^{l,f})$. We refer to models containing these variables as *restricted* models, while the models based on separated domestic and foreign variables are labeled *unrestricted* models.

The monetary model rests on two important simplifying assumptions: (i) domestic and foreign assets are perfect substitutes (implying perfect capital mobility) and (ii) current account effects (surplus or deficit) are negligible. These assumptions can be relaxed if the role of capital flows in explaining exchange rate movements is taken into account (see Bailey et al., 2001; Aliber, 2000). Thus, it might be possible to tie together movements in the exchange rates, the real interest rate, equity prices and current account balance. Current account dynamics can be thought of as the result of changes in productivity. For instance, if a positive productivity shock raises expected future output in the domestic economy, capital inflows are induced for at least two reasons. On the one hand, if consumers in the home economy expect to be richer in the future, they will want to borrow from abroad to increase their consumption today (assuming they are sufficiently forward-looking to smooth their consumption over time). On the other hand, the expected increase in future productivity raises expected future profits, increasing equity prices, thereby stimulating investment demand; insofar the

²We consider the model in equilibrium, thus money demand equals to money supply.

additional demand for funds to finance such investment is not available domestically, which causes inflows of capital (foreign direct investment and portfolio investment). Such arguments call for the inclusion of capital flow variables or proxies thereof into the exchange rate models. In addition to the unrestricted and restricted monetary model specifications, we consider a class of models which substitutes the output and money supply variables in the monetary model by a leading indicator variable and a stock market index. These specifications are labelled *capital flows* models.

Finally, for the empirical implementation of the models in the form of VAR specifications, we consider both parametrizations which include all variables and their respective lags as well as specifications where insignificant lags are omitted (subset-VAR models).

2.2 Forecasts and combinations

The aim of our analysis is to assess the profitability of trading strategies based on out-of-sample predictions of individual VAR, VEC and DVAR models, as well as combinations of these. Let us denote $\hat{S}_{i,t+h|t}$ the exchange rate forecast obtained using model i , $i = 1, \dots, k$, for time $t+h$ conditional on the information available at time t (i.e., h is the forecast horizon). Pooled forecasts, $\hat{S}_{c,t+h|t}$, take the form of a linear combination of the predictions of individual specifications,

$$\hat{S}_{c,t+h|t} = w_{c,0t}^h + \sum_{i=1}^k w_{c,it}^h \hat{S}_{i,t+h|t}, \quad (8)$$

where c is the combination method, k is the number of individual forecasts and the weights are given by $\{w_{c,it}^h\}_{i=0}^k$.

Since several combination methods require statistics based on a hold-out sample, let us introduce here some notation on the subsample limits: T_0 is used to denote the first observation of the available sample, the interval (T_1, T_2) is used as a hold-out sample used to obtain weights for those methods where such a subsample is required and T_3 is the last available observation. The sample given by (T_2, T_3) is the proper out-of-sample period used to compare the different methods.

We consider a large number of combination methods proposed in the literature:

- (i) *Mean, trimmed mean, median.* With regard to the mean, $w_{\text{mean},0t}^h = 0$ and $w_{\text{mean},it}^h = \frac{1}{k}$ in equation (8). The trimmed mean uses $w_{\text{trim},0t}^h = 0$ and $w_{\text{trim},it}^h = 0$ for the individual models that generate the smallest and largest forecasts, while $w_{\text{trim},it}^h = \frac{1}{k-2}$ for the remaining individual models. For the median combination method, $\hat{S}_{c,t+h|t} = \text{median}\{\hat{S}_{i,t+h|t}\}_{i=1}^k$ is used (see Costantini and Pappalardo, 2010).
- (ii) *Ordinary least squares (OLS) combination* (see Granger and Ramanathan, 1984). The method estimates the parameters in equation (8) using recursive and rolling windows. In the recursive case, to compute the initial OLS combination forecast, for S_{T_2} , we regress

$\{S_{t+h}\}_{t=T_1-1}^{T_2-2h}$ on a constant and $\{\hat{S}_{i,t+h|t}\}_{t=T_1-1}^{T_2-2h}$, $i = 1, \dots, k$, and set the weights in equation (8), w_{OLS,i,T_2-h}^h , equal to the estimated OLS coefficients. To construct the second combination forecast, for S_{T_2+1} , the OLS coefficients are estimated by regressing $\{S_{t+h}\}_{t=T_1-1}^{T_2-2h+1}$ on a constant and $\{\hat{S}_{i,t+h|t}\}_{t=T_1-1}^{T_2-2h+1}$, $i = 1, \dots, k$, and the fitted OLS coefficients, $w_{\text{OLS},i,T_2-h+1}^h$, are used as weights for equation (8). This procedure is applied until the available out-of-sample period; i.e., the weights of the h -step ahead forecast for S_{T_3} are obtained by regressing $\{S_{t+h}\}_{t=T_1-1}^{T_3-2h}$ on a constant and $\{\hat{S}_{i,t+h|t}\}_{t=T_1-1}^{T_3-2h}$, $i = 1, \dots, k$. In the case of the rolling window, we proceed in a similar fashion but discard the first observations in each replication of the procedure, so that the time series are consistently of length $T_2 - T_1 - 2h + 2$. Thus, for the second combination forecast S_{T_2+1} , for instance, we regress $\{S_{t+h}\}_{t=T_1}^{T_2-2h+1}$ on a constant and $\{\hat{S}_{i,t+h|t}\}_{t=T_1}^{T_2-2h+1}$, $i = 1, \dots, k$ and for the last combination forecast S_{T_3} , we regress $\{S_{t+h}\}_{t=T_3-T_2+T_1-1}^{T_3-2h}$ on a constant and $\{\hat{S}_{i,t+h|t}\}_{t=T_3-T_2+T_1-1}^{T_3-2h}$, $i = 1, \dots, k$.

- (iii) *Combination based on principal components (PC)*. This method allows to overcome multicollinearity when having many forecasts by reducing them to a few principal components (factors). The method is identical to the OLS combining method by replacing forecasts by their principal components and thus equation (8) changes to

$$\hat{S}_{\text{PC},t+h|t} = w_{\text{PC},0t}^h + \sum_{i=1}^{k_{h,t-h}} w_{\text{PC},it}^h f_{it}^h, \quad (9)$$

where $1 \leq k_{h,t-h} \leq k$ is the number of principal components extracted based on the information available at $t - h$ and $f_{1t}^h, \dots, f_{k_{h,t-h}t}^h$ are the first $k_{h,t-h}$ principal components for $\hat{S}_{1t}^h, \dots, \hat{S}_{k_{h,t-h}t}^h$. In our application, we choose the number of principal components using the so-called variance proportion criterion, which selects the smallest number of principal components such that a certain fraction (α) of variance is explained. In our application we set $\alpha = 0.8$. Hlouskova and Wagner (2013), where the principal components augmented regressions were used in the context of the empirical analysis of economic growth differentials across countries, provide more details on the method.³

- (iv) *Combination based on the discount mean square forecast errors (DMSFE)*. Following Stock and Watson (2004), the weights in equation (8) depend inversely on the historical forecasting performance of the individual models

$$w_{\text{DMSFE},i,t}^h = \frac{m_{ith}^{-1}}{\sum_{l=1}^k m_{lth}^{-1}}, \quad (10)$$

³We are not aware of the existence of any study using this approach in the context of the exchange rate forecasts.

where

$$m_{ith} = \sum_{s=T_1-1+h}^t \theta^{T-h-s} \left(S_{s+h} - \hat{S}_{i,s+h|s}^h \right)^2, \quad (11)$$

for $t = T_2 - h, \dots, T_3 - h$, $i = 1, \dots, k$, $w_{\text{DMSFE},0,t} = 0$ and θ is a discount factor. When $\theta = 1$ there is no discounting, while if $\theta < 1$, greater importance is attributed to the recent forecast performance of the individual models. In the empirical application, we use alternatively $\theta = 0.95$.

- (v) *Combination based on hit/success rates (HR)*. The method uses the proportion of correctly predicted directions of exchange rate changes of model i to the number of all correctly predicted directions of exchange rate changes by the models entertained,

$$w_{\text{HR},it}^h = \frac{\sum_{j=T_1+h-1}^t DA_{jh}^i}{\sum_{l=1}^k \left(\sum_{j=T_1+h-1}^t DA_{jh}^l \right)} \quad (12)$$

where $t = T_2 - h, \dots, T_3 - h$ and the index of directional accuracy is given by $DA_{jh} = I\left(\text{sgn}(S_j - S_{j-h}) = \text{sgn}(\hat{S}_{j|j-h} - S_{j-h})\right)$, where $I(\cdot)$ is the indicator function.

- (vi) *Combination based on the exponential of hit/success rates (EHR)* (Bacchini et al., 2010). The weights in this method are obtained as

$$w_{\text{EHR},it}^h = \frac{\exp\left(\sum_{j=T_1+h-1}^t (DA_{jh}^i - 1)\right)}{\sum_{l=1}^k \exp\left(\sum_{j=T_1+h-1}^t (DA_{jh}^l - 1)\right)} \quad (13)$$

where $t = T_2 - h, \dots, T_3 - h$.

- (vii) *Combination based on the economic evaluation of directional forecasts (EEDF)*. It uses weights that capture the ability of models to predict the direction of change of the exchange rate taking into account the magnitude of the realized change,

$$w_{\text{EEDF},it}^h = \frac{\sum_{j=T_1+h-1}^t DV_{jh}^i}{\sum_{l=1}^k \left(\sum_{j=T_1+h-1}^t DV_{jh}^l \right)} \quad (14)$$

where $t = T_2 - h, \dots, T_3 - h$ and $DV_{th} = |S_t - S_{t-h}| DA_{th}$.

- (viii) *Combination based on predictive Bayesian model averaging (BMA)*. The weights used are based on the corresponding posterior model probabilities based on out-of-sample (rather than in-sample) fit. See for example Raftery et al. (1997), Carriero et al. (2009), Crespo Cuaresma (2007), Feldkircher (2012).

$$w_{\text{BMA},it}^h = P(M_i | \mathbf{S}_{T_1+h-1:t}) = \frac{P(\mathbf{S}_{T_1+h-1:t} | M_i) P(M_i)}{\sum_{l=1}^k P(\mathbf{S}_{T_1+h-1:t} | M_l) P(M_l)}, \quad (15)$$

where $P(M_i|\mathbf{S}_{T_1+h-1:t})$ is the posterior model probability of model i , $P(\mathbf{S}_{T_1+h-1:t}|M_i)$ is the marginal likelihood of the model and $t = T_2 - h, \dots, T_3 - h$. Using the predictive likelihood in order to address the out-of-sample fit of each model and assuming equal prior probability across models, $P(M_l)$, the weights can be approximated as

$$w_{\text{BMA},it}^h = \frac{(t - T_1 - h + 2)^{\frac{p_1 - p_i}{2}} \left(\frac{\sum_{j=T_1+h-1}^t MSE_{jh}^1}{\sum_{j=T_1+h-1}^t MSE_{jh}^i} \right)^{\frac{t - T_1 - h + 2}{2}}}{\sum_{l=1}^k (t - T_1 - h + 2)^{\frac{p_1 - p_l}{2}} \left(\frac{\sum_{j=T_1+h-1}^t MSE_{jh}^1}{\sum_{j=T_1+h-1}^t MSE_{jh}^l} \right)^{\frac{t - T_1 - h + 2}{2}}} \quad (16)$$

where MSE_{jh}^i is the mean squared error of model i , namely $MSE_{jh}^i = (\hat{S}_{i,j|j-h} - S_j)^2$.

- (ix) *Combinations based on frequentist model averaging (FMA)* (see Claeskens and Hjort, 2008, and Hjort and Claeskens, 2003). The weights are calculated as follows

$$w_{\text{FMA},it}^h = \frac{\exp\left(-\frac{1}{2}IC_t^i\right)}{\sum_{l=1}^k \exp\left(-\frac{1}{2}IC_t^l\right)} \quad (17)$$

where IC_t^i stands for an information criterion of model i and t is the last time point of the data over which are models estimated.

We use combinations of forecasts based on the Akaike criterion (AIC), Schwarz criterion (BIC) and Hannan-Quinn criterion (HQ). The weights corresponding to the BIC can be interpreted as an approximation to the posterior model probabilities in BMA (see Raftery et al., 1997; Sala-i-Martin et al., 2004).

2.3 Predictive accuracy: Loss and profit measures

We evaluate the exchange rate forecasts using performance measures based on both profit maximization and the loss minimization. The loss measures include the standard mean squared error, $MSE_{th} = (\hat{S}_{t|t-h} - S_t)^2$ and the mean absolute error, $MAE_{th} = |\hat{S}_{t|t-h} - S_t|$, which have been routinely used in most empirical assessments of exchange rate forecasting models. The former include the directional accuracy measure (DA), the directional value measure (DV), the annualized returns from two different trading strategies generated by our forecasts and risk adjusted performance measures given by the Sharpe ratios for both of the trading strategies.

The directional accuracy measure $DA_{th} = I\left(\text{sgn}(S_t - S_{t-h}) = \text{sgn}(\hat{S}_{t|t-h} - S_{t-h})\right)$, introduced already above, is a binary variable indicating whether the direction of the exchange rate change was correctly forecast at horizon h ($DA_{th} = 1$) or not ($DA_{th} = 0$). While the function DA_{th} is robust to outlying forecasts, it does not consider the size of the realized directional movements. The economic value of directional forecasts is better captured by

assigning to each correctly predicted change its magnitude (see Blaskowitz and Herwartz, 2011). The directional value (*DV*) statistic, defined as $DV_{th} = |S_t - S_{t-h}|DA_{th}$ is used for this purpose.

The performance of exchange rate forecasts based on their profitability is evaluated by constructing simple trading strategies based on the predictions. We start with a simple trading strategy as described in Gencay (1998), where the selling/buying signal is based on the current exchange rate, namely, forecast upward movements of the exchange rate with respect to the actual value (positive returns) are executed as long positions while the forecast downward movements (negative returns) are executed as short positions; i.e., the total return of the trading strategy over n periods is given by

$$R_h^S = \sum_{t=1}^n y_{t-h,h}^S r_{th} = \sum_{t=1}^n R_{th}^S \quad (18)$$

where $r_{th} = \log(S_t/S_{t-h})$, $t = 1, \dots, n$,

$$y_{t-h,h}^S = \begin{cases} -1, & \text{for selling signal (forecast downward movement for horizon } h) \\ & \hat{S}_{t|t-h} < S_{t-h} \\ 1, & \text{for buying signal (forecast upward movement for horizon } h) \\ & \hat{S}_{t|t-h} > S_{t-h} \end{cases}$$

and $R_{th}^S = y_{t-h,h}^S r_{th}$. We label this trading strategy TS^S . While this trading strategy is based on comparing current and predicted exchange rates, a comparison of the forecast with the forward rate would be a natural building block for an alternative trading strategy. The trading strategy used in Boothe (1983), for instance, generates signals based on the comparison of the forecast value to the current forward rate

$$R_h^F = \sum_{t=1}^n y_{t-h,h}^F r_{th} = \sum_{t=1}^n R_{th}^F \quad (19)$$

where

$$y_{t-h,h}^F = \begin{cases} -1, & \hat{S}_{t|t-h} < F_{t|t-h} \\ 1, & \hat{S}_{t|t-h} > F_{t|t-h} \end{cases} \quad (20)$$

$F_{t|t-h}$ is the forward rate for time t given at time $t-h$ and $R_{th}^F = y_{t-h,h}^F r_{th}$. We label this trading strategy TS^F . Returns generated by the trading strategy where the selling/buying signal is based on the current exchange rate, TS^S , are denoted by R^S , and the returns generated by the trading strategy where the selling/buying signal is based on the current

forward rate, TS^F , are denoted by R^F .

In addition to the profitability measures presented above, we also perform comparisons based on *Sharpe ratios* - the excess return per unit of deviation generated by a trading strategy; i.e., $SR = \frac{R}{\sigma}$, where R is the (annualized) return of a trading strategy and σ is its standard deviation. The natural benchmark return in the definition of the Sharpe ratio for our application appears to be a zero return, reflecting that the investor does not take any position in the foreign exchange market.

The different performance measures that can be computed based on the forecasts of our macro-econometric models need to be compared with a set of performance measures implied by reference models against which to benchmark the ability of the models entertained. The benchmark model for MAE and MSE measures is the random walk model, for DA and DV measures it is the random walk with an intercept and for trading strategies TS^S and TS^F the following benchmark trading strategies are used (for more details see Neely and Weller, 2013):

- The *buy-and-hold* strategy: $R^{BH} = \log(S_n/S_1)$; i.e., buying at period 1 and holding it at least till period n .
- Trading signals based on the *forward rate*; i.e., whether the forward exchange rate indicates appreciation or depreciation. I.e.,

$$R_h^{Fo} = \sum_{t=1}^n y_{t-h,h}^{Fo} r_{th} \quad (21)$$

where

$$y_{t-h,h}^{Fo} = \begin{cases} -1, & S_{t-h} > F_{t|t-h} \\ 1, & S_{t-h} < F_{t|t-h} \end{cases} \quad (22)$$

- *Moving average rules*, based on $MA_t(m, n) = \frac{1}{m} \sum_{i=1}^{m-1} S_{t-i} - \frac{1}{n} \sum_{i=1}^{n-1} S_{t-i}$ where $m < n$. If $MA_t(m, n) > 0$ then a buying signal is generated and if $MA_t(m, n) < 0$ then a selling signal is generated.⁴ The corresponding return is given by

$$R_h^{MA} = \sum_{t=1}^n y_{t-h,h}^{MA} r_{th} \quad (23)$$

⁴See for instance Harris and Yilmaz (2009).

where

$$y_{t-h,h}^{MA} = \begin{cases} -1, & MA_{t-h}(m, n) < 0 \\ 1, & MA_{t-h}(m, n) > 0 \end{cases} \quad (24)$$

For monthly exchange rates and one-step-ahead predictions, the most widely used MA rule in the fund management industry is $MA(1, 2)$. For a forecast horizon of h , we generalize the statistic to $MA(h, 2h)$ and build the signals based on this moving average statistic.

- *Filter rules*, where the buy signal is generated when the exchange rate has increased by more than a certain percent above its most recent low and the sell signal is generated when the exchange rate has fallen by more than the same percent from its most recent high. The resulting return is then given by

$$R_h^{Filter} = \sum_{t=1}^n y_{t-h,h}^{Filter} r_{th} \quad (25)$$

where

$$y_{t-h,h}^{Filter} = \begin{cases} -1, & S_{t-h} < S_{t-2h}(1 - x) \\ 1, & S_{t-h} > S_{t-2h}(1 + x) \end{cases} \quad (26)$$

where the filter size x is such that $0 < x < 1$. For our application, $x = 0.01, 0.02$ and 0.1 are used alternatively.

- *Carry trade rules* are based on borrowing in low interest rate currencies to fund investments in high-yield currencies (or target currencies), a strategy implied by the uncovered interest rate parity (see Ilut, 2012).⁵ The resulting return is given by

$$R_h^{CT} = \sum_{t=1}^n y_{t-h,h}^{CT} r_{th} \quad (27)$$

where

$$y_{t-h,h}^{CT} = \begin{cases} -1, & i_{t-h,h}^d < i_{t-h,h}^f \\ 1, & i_{t-h,h}^d > i_{t-h,h}^f \end{cases}$$

⁵Bekaert et al. (2007) and Krishnakumar and Neto (2012) point out the importance of the link between the interest rate parity and the hypothesis of the term structure for the determination of the exchange rate.

where $i_{t-h,h}^d$ is a domestic interest rate for h -steps ahead while $i_{t-h,h}^f$ is a foreign interest rate for h -steps ahead.

3 Estimation, prediction and testing for data snooping

3.1 Estimation details

We base our comparison on monthly data spanning the period from January 1980 until December 2013 for the EUR/USD exchange rate. The beginning of the sample is thus $T_0 =$ January 1980, the beginning of the hold-out forecasting sample for individual models used in order to obtain weights based on predictive accuracy is given by $T_1 =$ January 2007. The beginning of the actual out-of-sample forecasting sample is $T_2 =$ January 2010, and the end of the data sample is $T_3 =$ December 2013.⁶

The lag length of the VAR, VEC and DVAR specifications is selected using the AIC criterion for potential lag lengths ranging from 1 to 12 lags.⁷ For the VEC models, selection of the lag length and the number of cointegration relationships is carried out simultaneously using the AIC. Since VAR models are known to forecast poorly due to overfitting (see, e.g., Fair, 1979), we also estimate subset-VAR specifications, where individual parameters of the VAR specification are set equal to zero recursively using t-tests (see Kunst and Neusser, 1986, for a similar approach). While in the set of restricted specifications based on the monetary model which are mentioned in section 2 the parameters are constrained based on theoretical assumptions, in the case of subset-VARs the corresponding specification is estimated and insignificant lags of the endogenous variables are removed from the model specification. The restrictions are imposed by setting to zero those parameters for which we cannot reject that they are equal to zero using a one-sided t-test.

In addition to standard VAR, DVAR and VEC models, we also estimate Bayesian VARs. The standard Bayesian approach for estimating VAR models was mainly developed by Doan et al. (1984) and Litterman (1986), who suggest that assuming as a prior that the variables in the VAR follow a random walk would be sensible for economic variables (the Litterman/Minnesota prior). In the case of exchange rates, it would furthermore be consistent with the efficient market hypothesis. We thus estimate DVAR specifications using Bayesian methods, setting the mean of the prior for the estimated coefficients to zero. Regarding the specification of the prior variance-covariance matrix, V , of the coefficients of different lags of the endogenous variables of the model a typical element is set to

$$v_{ij,l} = \begin{cases} (\lambda/l^d)^2 & \text{for } i = j, \\ (\lambda\rho\sigma_i/l^d\sigma_j)^2 & \text{for } i \neq j, \end{cases} \quad (28)$$

⁶The sources for all variables used are given in the data appendix.

⁷Our results are however robust to model selection using BIC or the HQ criterion.

where $v_{ij,l}$ is the prior variance of the parameter corresponding to the l -th lag of variable j in equation i , $\lambda > 0$ is the ‘overall tightness’ parameter, d is the rate of decay, and $\rho \in (0, 1)$ allows for differences in the weight of own lags of the explained variable with respect to lags of other variables.⁸

We consider rolling-window estimation for our analysis; i.e., we keep the estimation sample size constant (equal to $T_1 - T_0$) as we re-estimate the models, thus moving the window that defines the sample used to estimate the model parameters. The performance measures for each model, as introduced in section 2.3, are calculated over the out-of-sample period for a given forecasting horizon and aggregated as follows

$$\begin{aligned}
MSE_h &= \sum_{j=0}^{T_3-T_2} MSE_{T_2+j,h} \\
MAE_h &= \sum_{j=0}^{T_3-T_2} MAE_{T_2+j,h} \\
DA_h &= \sum_{j=0}^{T_3-T_2} \frac{DA_{T_2+j,h}}{T_3 - T_2 + 1} \\
DV_h &= \frac{\sum_{j=0}^{T_3-T_2} DV_{T_2+j,h}}{\sum_{j=0}^{T_3-T_2} |S_{T_2+j} - S_{T_2+j-h}|} \\
&= \frac{\sum_{j=0}^{T_3-T_2} |\hat{S}_{T_2+j} - S_{T_2+j-h}| DA_{T_2+j,h}}{\sum_{j=0}^{T_3-T_2} |S_{T_2+j} - S_{T_2+j-h}|}
\end{aligned}$$

where $h = 1, \dots, 12$.

3.2 Data snooping bias free tests for equal predictive ability

In order to assess whether the predictive superiority of certain models is systematic and not due to luck, we also perform bootstrap tests for the comparison of predictive ability with respect to the benchmark models and trading strategies. In particular, we use the ‘reality check’ (RC) test by White (2000), the test for superior predictive ability (SPA) by Hansen (2005), the stepwise test of multiple check (stepM-RC) by Romano and Wolf (2005) and the stepwise multiple superior predictive ability test (stepM-SPA) by Hsu et al. (2010).

The following relative performance measures, d_{ith} , $i = 1, \dots, k$, $t = T_2, T_2 + 1, \dots, T_3$,

⁸For our estimation results, we set $\lambda = 0.1$, $\rho = 0.99$, and $d = 1$. See also Carriero et al. (2015) for Bayesian VARs and prediction.

$h = 1, \dots, 12$ are computed and the tests are defined based on them:

$$d_{ith} = \begin{cases} MSE_{RW,th} & - & MSE_{ith} \\ MAE_{RW,th} & - & MAE_{ith} \\ DA_{ith} & - & DA_{RW_{int},th} \\ DV_{ith} & - & DV_{RW_{int},th} \\ y_{ith}^S r_{th} & - & y_{ref,th} r_{th} \\ y_{ith}^F r_{th} & - & y_{ref,th} r_{th} \\ SR_{ith}^S & - & SR_{ith}^{ref} \\ SR_{ith}^F & - & SR_{ith}^{ref} \end{cases} \quad (29)$$

Index *ref* denotes the reference/benchmark trading rule, implying that we concentrate on relative returns. The benchmark trading strategies are defined by (21)–(28). Thus, $ref \in \{Fo, MA, Filter, CT\}$. SR^S stands for the Sharpe ratio implied by the trading strategy TS^S as defined in (18), SR^F stands for the Sharpe ratio implied by the trading strategy TS^F as defined in (19)⁹ and RW_{int} stands for the random walk with an intercept.

White’s (2000) bootstrap RC test is a comprehensive test across all models considered and directly quantifies the effect of data snooping by testing the null hypothesis that the performance of the best model is no better than the performance of the benchmark model.¹⁰ The null hypothesis of the test is

$$H_0 : \mathbb{E}(\mathbf{d}_t) \leq 0 \quad (30)$$

where $\mathbf{d}_t = (d_{1t}, \dots, d_{kt})$ is a k –dimensional vector of relative performance measures as defined in (29). Rejection of (30) implies that at least one model beats the benchmark. The RC test is constructed using the test statistic

$$T_n^{RC} = \max\{\sqrt{n}\bar{d}_1, \dots, \sqrt{n}\bar{d}_k\} \quad (31)$$

where n is the number of out-of-sample observations ($n = T_3 - T_2 + 1$) and $\bar{d}_i = \sum_{t=T_2}^{T_3} d_{it}$ for $i = 1, \dots, k$. Following White (2000), the bootstrap RC p –values are calculated using the stationary bootstrap method of Politis and Romano (1994), where the potential dependence in \mathbf{d}_t is taken into account. At first, the empirical distribution of T_n^{RC*} is obtained, where

$$T_n^{RC*}(b) = \max\{\sqrt{n}(\bar{d}_1(b) - \bar{d}_1), \dots, \sqrt{n}(\bar{d}_k(b) - \bar{d}_k)\} \quad (32)$$

for $b = 1, \dots, B$, where B is the number of bootstrap simulations. The p –values are obtained

⁹To ease the notation, we omit the index h that indicates the forecast horizon in the discussion below.

¹⁰The term ‘model’ is obviously used in a broad sense that includes forecasting rules and methods (like forecast combinations).

by comparing T_n^{RC} with the quantiles of the empirical distribution of T_n^{RC*} .¹¹

Hansen (2005) points out that the RC test of White (2000) is too conservative because its null distribution is obtained under the least favorable configuration to the alternative. The RC test may lose power when poor models are included in the group of models under consideration. To improve the power of the test, Hansen (2005) proposes the superior predictive ability (SPA) test. The null hypothesis of the SPA test is the same as in the in White's RC test, but Hansen (2005) uses the studentized test statistic to improve the power.¹² The test statistic for the SPA test is

$$T_n^{SPA} = \max \left[\max \left\{ \frac{\sqrt{n}\bar{d}_1}{\hat{s}_1}, \dots, \frac{\sqrt{n}\bar{d}_k}{\hat{s}_k} \right\}, 0 \right] \quad (33)$$

where \hat{s}_i is a consistent estimator of $\text{var}(\sqrt{n}\bar{d}_i)$, $i = 1, \dots, k$. The same bootstrap method of Politis and Romano (1994) is used to calculate the empirical distribution of the statistic under the null.

One drawback of both RC and SPA tests is that they do not aim at explicitly identifying the models which outperforms the benchmark. Romano and Wolf (2005) propose the stepM-RC test that can identify also those models for which $\mathbb{E}(d_{it}) > 0$ holds. For a given model i , ($i = 1, \dots$) the following individual testing problems are considered

$$H_0^i : \mathbb{E}(d_{it} \leq 0) \quad \text{vs} \quad H_A^i : \mathbb{E}(d_{it} > 0) \quad (34)$$

This multiple testing method yields a decision for each individual testing problem (by either rejecting H_0^i or not). The individual decisions are made such that the familywise error rate¹³ is asymptotically achieved at the significance level α which is achieved by constructing a joint confidence region with a nominal joint coverage probability of $1 - \alpha$. This stepwise procedure is implemented as follows. Without loss of generality we assume that $\{\bar{d}_i\}_{i=1}^k$ are arranged in a descending order. Top j_1 null hypotheses are rejected (i.e., top j_1 models outperform the benchmark) if $\sqrt{n}\bar{d}_l$, $l = 1, \dots, j_1$ is greater than the bootstrapped critical value computed from the bootstrap procedure as in the RC test. If none of the null hypotheses is rejected, the procedure terminates. Otherwise, d_{1t}, \dots, d_{j_1t} , $t = T_2, \dots, T_3$ are removed from the data and the bootstrap simulation is applied to the rest of the data to obtain the new critical value. If $\sqrt{n}\bar{d}_l$, $l = 1, \dots, j_2$ is greater than the new bootstrapped critical value then the following

¹¹This procedure involves choosing a dependence parameter q that serves to preserve possible time dependence (see White, 2000). We present in our empirical analysis the results for $q=0.9$, which corresponds to a plausibly low persistence level in exchange rate changes. Qualitatively similar results are found for $q=0.5$ and are not reported but are available from the authors upon request. Similar values for the smoothing parameter are used in Gonzalez-Rivera et al. (2004), Qi and Wu (2006) and Yang et al. (2008).

¹²The improvement of the power of the SPA test over the RC test is confirmed by simulations in Hansen (2005).

¹³The familywise error rate is defined as the probability of rejecting at least one true null hypothesis. For more details, see Romano and Wolf (2005).

j_2 null hypotheses are rejected. The procedure continues until no more null hypotheses are rejected. In our analysis we use significance levels of 5% and 10%.

Hsu et al. (2010) extend the SPA of Hansen (2005) to a stepwise SPA test in the way Romano and Wolf (2005) did it for the RC test. They show analytically that the stepM-SPA test is more powerful than stepM-RC test. The step wise procedure is the same as in the stepM-RC test but with RC test statistics replaced by PCA test statistics.

4 Results

Table 1 presents the acronyms of the models, forecast combination techniques and benchmark trading strategies entertained in the analysis. Table 2 shows the performance of specifications based on the variables of the monetary model and three different prediction horizons (one, six and twelve months ahead). Table 3 shows the same information for specifications based on the capital flow model. The tables are structured in three blocks, each one corresponding to a different forecasting horizon. Each block, in turn, is divided into three different parts. The top part of the block presents the results for those individual specifications which perform best according to the criteria described in section 2.2 and section 2.3. In the central part of the block, we present the results for all forecast combination methods used. The bottom part of each block presents the corresponding measures for the best-performing benchmark strategies. The forecasts are evaluated using the loss-based and profit-based measures described in section 2.3¹⁴ and the predictive superiority of the models which perform better than the benchmark is assessed by means of the bootstrap stepM-SPA test by Hsu et al. (2010).¹⁵

[Include Table 1 about here]

The random walk model is always beaten by the best single individual specification and the best combination of forecasts for 1-step and 6-steps ahead in terms of predictive ability as measured by MAE, MSE, DA and DV (except for the best single individual model for MAE and MSE and 6-steps ahead). The results are slightly different for measures based on 12-steps ahead forecasts. Here, the random walk prevails over the other models for MAE and MSE. However, according to the stepM-SPA test, only differences in forecasting ability in terms of DA and DV appear significant, while those measured by MAE and MSE measures are all insignificant. More specifically, we find that in case of the DA and DV performance measures, the benchmark random walk model is systematically beaten by the combination of forecasts based on the principal components for 6-steps and 12-steps ahead, which appears superior at the 5% significance level using the bootstrap-based stepM-SPA test. On the other

¹⁴The loss measures are based on currency units. Note that returns generated by trading strategies are calculated from the position of a foreign investor.

¹⁵We used all the tests described in section 3.2, but report only the results for the stepM-SPA test in the tables. Detailed results using the other tests are available from the authors upon request.

hand, some capital flows models perform significantly better than the random walk for 1-step ahead in terms of the DA and DV measures. This is the case for both a particular individual specification and for forecasting averaging techniques based on the exponential of hit rates (EHR) and on Bayesian model averaging weights (BMA).

Our results are in line with those in Dal Bianco et al. (2012), who use weekly data for the EUR/USD exchange rate and show that their model achieves large improvements in predictive ability when using direction of change measures. Our findings thus strengthen and extend the empirical evidence presented in Dal Bianco et al. (2012), as we use directional measures that take in account both the realized directional changes in exchange rates as well as for their magnitudes and are robust to outliers, and apply data snooping free tests for the evaluation of the performance.

[Include Tables 2 & 3 about here]

As for the returns of trading strategies based on exchange rate forecasts, the results show that only the returns implied by trading strategy TS^F (see equations (19) and (20)) based on the principal components forecasts combination method is significantly better than the best benchmark models at a 10% significance level. This occurs for 6-steps and 12-steps ahead in the case of the monetary model (see Table 2) and only for 6-steps ahead for the capital flows model. Looking at the Sharpe ratios based on returns generated by trading strategies TS^S and TS^F , a slightly stronger evidence of risk adjusted profitability is found (in some cases the results are significant at 5% level). Sharpe ratios based on forecasts combined using principal components are significantly better than those implied by the best benchmark trading strategies, carry trade and MA(12,24), for 6-steps and 12-steps ahead in both monetary and capital flows models, even though these ratios take values below the unity, as in other studies for the EUR/USD exchange rate (see for example Dunis et al., 2010). Although it is argued that market practitioners in the foreign exchange market may not be interested in currency investment strategies that yield a Sharpe ratio which is lower than one (see Sarno et al., 2006), it should be however noticed that the difference in performance of the other alternative forecasting models with respect to the benchmark model is insignificant.

Out of the results, several general conclusions can be drawn. First of all, as compared to individual specifications and benchmark strategies, the use of forecast combination methods leads to improvements in the performance of trading rules implied by these forecasts with respect to the benchmarks under consideration. In particular, forecast pooling based on principal components methods appears to be the most robust technique for the EUR/USD exchange rate as it significantly outperforms benchmark models (or trading strategies) in terms of the success rate (DA), the economic value of directional forecasts (DV), returns implied by trading strategy TS^F and Sharpe ratios for both monetary and capital flows

models.¹⁶ While the results concerning differences in Sharpe ratios of returns given by the trading strategies indicate that the margin for achieving systematic monetary profits in the foreign exchange market using macro-econometric models is relatively limited, it is important to point out that our empirical evidence on the performance of the forecast combination method based on principal components represents a novelty for the literature on forecasting exchange rates, as other simpler rules for combining forecasts are often preferred (see for example MacDonald and Marsh, 1994; Dunis and Huang, 2002; Altavilla and De Grauwe, 2010; Wu and Wang, 2013; Li et al., 2015).

Second, there is no evidence of a “*one size fits all*” approach to the specification of single multivariate time series models for exchange rate forecasting based on macroeconomic covariates or variables related to capital flows which leads to systematically good predictive ability in terms of the performance measures under consideration. Although for some predictive error measures individual specifications present decent forecasting accuracy characteristics, the use of single error correction specifications or Bayesian VAR models does not systematically ensure a lower loss or a higher profit, and there is no systematic relationship between the use of variables related to a particular theoretical setting (monetary or capital flows) and improvements in the predictive ability of the model as measured by our loss and profit measures. In the short term (one-month-ahead predictions) exchange rate models based on the capital flows framework tend to present slightly better forecasting ability in terms of directional change, but the predictive quality of both types of models converge as the forecasting horizon gets larger, with monetary models performing minimally better than capital flows models for the longest forecasting horizons.

We carried out robustness checks repeating the forecasting exercise with reduced samples that excluded the observations corresponding to the financial crisis. The predictive ability of the forecast averaging methods was significantly worse for the reduced sample, in particular for measures based on the performance of trading strategies.¹⁷ Such a result indicates that the macroeconomic volatility experienced during the recent crisis has been helpful to improve the estimation of the effects of the variables entertained on exchange rate developments, as well as to obtain more reliable model weights in forecast averaging exercise.

To the extent that predicting the direction of change in the EUR/USD exchange rate is important for policy makers and investors, our results strongly support the use of forecast averaging based on principal components. This is the only technique that provides systematic and sizeable improvements in predictive power in the medium and long run when measuring the error through directional accuracy indicators. Independently of whether the monetary model or the capital flows model is used as a basis for the identification of variables and

¹⁶Not in terms of asymptotic based tests but in terms of bootstrap-based data snooping bias free tests.

¹⁷Results based on subsamples of the available data are not reported but available from the authors upon request.

the selection of models, relying on individual specifications appears to be a less promising strategy for EUR/USD exchange rate prediction as compared to forecast combination and, in particular, averaging of model-specific predictions based on principal components.

5 Conclusions

Using a large battery of multivariate time series models and forecast combination methods, we offer the most comprehensive out-of-sample forecast analysis to date in terms of predictive error, directional accuracy and profitability of trading strategies for the EUR/USD exchange rate. Our study contributes to the empirical literature on forecasting EUR/USD exchange rates in several respects. First, we not only test for the predictability of the EUR/USD exchange rate based on both loss and directional accuracy measures but also compare both profit and risk adjusted profits generated by forecast-based trading strategies to those obtained using benchmark trading rules. Second, in order to assess whether the predictive superiority of certain models is systematic, we use the StepM-RC test by Romano and Wolf (2005) and StepM-SPA test by Hsu et al. (2010) in addition to the RC test of White (2000) and the SPA test by Hansen (2005), since the latter two tests are not explicitly designed to identify the models which outperforms the benchmark. Third, we use a recently developed directional forecast accuracy that account for both the realized directional changes in exchange rates as well as for their magnitudes. Fourth, we exploit the potential of a large number of forecast combination methods for both forecast accuracy evaluation and profitability, and propose a new method of combination based on the economic evaluation of directional forecasts.

Our empirical results emphasize the lack of superiority of single multivariate specifications based on macroeconomic fundamentals over different prediction horizons. Instead, forecast combinations, especially those building on the principal component method, appear particularly promising in improving profitability based performance. Future research will extend this study by considering optimal currency portfolios based on a variety of foreign exchange trading strategies and their impact on different (risk adjusted) profit measures.

References

- [1] R.Z. Aliber. 2000. Capital Flows, Exchange Rates, and the New International Financial Architecture: Six Financial Crises in Search of a Generic Explanation. *Open Economies Review*, 11, 43-61.
- [2] C. Altavilla, P. De Grauwe. Forecasting and combining competing models of exchange rate determination. *Applied Economics*, 42, 3455-3480.
- [3] F. Bacchini, A. Ciammola, R. Iannaccone, M. Marini. 2010. Combining forecasts for a flash estimate of Euro area GDP. *Contributi ISTAT*, No. 3.
- [4] A. Bailey, S. Millard, S. Wells. 2001. Capital flows and exchange rates. *Bank of England Quarterly Bulletin*, Autumn, 310-318.
- [5] G. Bekaert, R. Campbell, R. Harvey, C. Lundblad. 2007. Liquidity and expected returns: Lessons from emerging markets. *Review of Financial Studies*, 20, 1783-1831.
- [6] C. Bergmeir, M. Costantini, J.M. Benítez. 2014. On the usefulness of cross validation for directional forecast evaluation. *Computational Statistics and Data Analysis*, 76, 132-143.
- [7] J. Berkowitz, L. Giorgianni. 2001. Long-horizon exchange rate predictability? *Review of Economics and Statistics*, 83, 81-91.
- [8] O. Blaskowitz, H. Herwartz. 2009. Adaptive forecasting of the EURIBOR swap term structure. *Journal of Forecasting*, 28, 575-594.
- [9] O. Blaskowitz, H. Herwartz. 2011. On economic evaluation of directional forecasts. *International Journal of Forecasting*, 27, 1058-1065.f
- [10] P. Boothe. 1983. Speculative profit opportunities in the canadian foreign exchange market, 1974-78. *The Canadian Journal of Economics*, 16, 603-611.
- [11] J. Boudoukh, M. Richardson, R.F. Whitelaw. 2008. The myth of long-horizon predictability. *Review of Financial Studies*, 21, 1577-1605.
- [12] A. Carriero, G. Kapetanios, M. Marcellino. 2009. Forecasting exchange rates with a large Bayesian VAR. *International Journal of Forecasting*, 25, 400-417.
- [13] A. Carriero, , T.G. Clark and M. Marcellino. 2015. Bayesian VARs: specification choices and forecast accuracy, *Journal of Applied Econometrics*, 30, 46-73.
- [14] G. Claeskens, N.L. Hjort. 2008. Model Selection and Model Averaging. Cambridge: Cambridge University Press.

- [15] Y.W. Cheung, M.D. Chinn, A.G. Pascual. 2005. Empirical exchange rate models of the nineties: Are any fit to survive? *Journal of International Money and Finance*, 24, 1150-1175.
- [16] M.D. Chinn, R.A. Meese. 1995. Banking on currency forecasts: How predictable is change in money? *Journal of International Economics*, 38, 161-178.
- [17] M. Costantini, C. Pappalardo. 2010. A hierarchical procedure for the combination of forecasts. *International Journal of Forecasting*, 26, 725-743.
- [18] J. Crespo Cuaresma. 2007. Forecasting euro exchange rates: How much does model averaging help? Working paper of University of Innsbruck, 2007-24.
- [19] M. Dal Bianco, M. Camacho, G.P. Quiros. 2012. Short-run forecasting of the euro-dollar exchange rate with economic fundamental. *Journal of International Money and Finance*, 31, 377-396.
- [20] T. Doan, R. Litterman, C. Sims. 1984. Forecasting and conditional projections using realist prior distributions. *Econometric Review*, 3, 1-100.
- [21] R. Dornbusch. 1976. Expectations and exchange rate dynamics. *Journal of Political Economy*, 84, 1161-1176.
- [22] C.L. Dunis and X. Huang. 2002. Forecasting and Trading Currency Volatility: An Application of Recurrent Neural Regression and Model Combination. *Journal of Forecasting*, 21, 317-354.
- [23] Dunis, C.L., J. Laws, G. 2010. Modelling and trading the EUR/USD exchange rate at the ECB fixing. *The European Journal of Finance*, 16, 541-560.
- [24] R. Fair. 1979. An analysis of the accuracy of four macroeconomic models. *Journal of Political Economy*, 87, 701-18.
- [25] M. Feldkircher. 2012. Forecast combination and Bayesian model averaging: A prior sensitivity analysis. *Journal of Forecasting*, 31, 361-376.
- [26] J.A. Frenkel. 1976. A monetary approach to the exchange rate: Doctrinal aspects and empirical evidence. *Scandinavian Journal of Economics*, 78, 200-224.
- [27] R. Gencay. 1998. The predictability of security returns with simple technical trading rules. *Journal of Empirical Finance*, 5, 347-359.
- [28] G. Gonzalez-Rivera, T.H. Lee, S. Mishra. 2004. Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of forecasting*, 20, 629-645.

- [29] C.W.J. Granger, R. Ramanathan. 1984. Improved methods of combining forecasts. *Journal of Forecasting*, 3, 197-204.
- [30] P.R. Hansen. 2005. A test for superior predictive ability. *Journal of Business and Economic Statistics*, 23, 365-380.
- [31] R.D.F. Harris, F. Yilmaz. 2009. A momentum trading strategy based on the low frequency component of the exchange rate. *Journal of Banking and Finance*, 33, 1575-1585.
- [32] K. Heimonen. 2009. The eurodollar exchange rate and equity flows. *Review of Financial Economics*, 18, 202-209.
- [33] N.L. Hjort, G. Claeskens. 2003. Frequentist model average estimators. *Journal of American Statistical Association*, 98, 879-899.
- [34] J. Hlouskova, M. Wagner. 2013. The determinants of long-run economic growth: A conceptually and computationally simple approach. *Swiss Journal of Economics and Statistics*, 149, 445-492.
- [35] P. Hooper, J. Morton. 1982. Fluctuations in the dollar: A model of nominal and real exchange rate determination. *Journal of International Money and Finance*, 1, 39-56.
- [36] Y. Hsing, B.S. Sergi. 2009. The dollar/euro exchange rate and a comparison of major models. *Journal of Business Economics and Management*, 10, 199-205.
- [37] N.P. Hsu, Y. Hsu, C. Kuan. 2010. Testing the predictive ability of technical analysis using a new stepwise test without data snooping bias. *Journal of Empirical Finance*, 17, 471-484.
- [38] C. Ilut. 2012. Ambiguity aversion: Implications for the uncovered interest rate parity puzzle. *American Economic Journal: Macroeconomics*, 4, 33-65.
- [39] A. Jamaleh. 2002. Explaining and forecasting the euro/dollar exchange rate through a non-linear threshold model. *The European Journal of Finance*, 8, 422-448.
- [40] J. James, I. Marsh, L. Sarno. 2012. Handbook of Exchange Rates. John Wiley & Sons Inc.
- [41] C. Karfakis. 2006. Is there an empirical link between the dollar price of the euro and the monetary fundamentals? *Applied Financial Economics*, 16, 973-980
- [42] L. Kilian. 1999. Exchange rates and monetary fundamentals: What do we learn from long-horizon regressions? *Journal of Applied Econometrics*, 14, 491-510.

- [43] King, M.R., Osler, C., Rime, D. 2012. Foreign Exchange Market Structure, Players, and Evolution. In *Handbook of Exchange Rates*, James, J., Marsh, I., Sarno,L.(eds). John Wiley & Sons Inc.
- [44] J. Krishnakumar and D. Neto. 2012. Testing uncovered interest rate parity and term structure using a three-regime threshold unit root VECM: An application to the swiss isle of interest rates. *Oxford Bulletin of Economics and Statistics*, 74, 180-202.
- [45] R.M. Kunst, K. Neusser. 1986. A forecasting comparison of some VAR techniques. *International Journal of Forecasting*, 2, 447-456.
- [46] J. Li, I.Tsiakas, W.Wang. 2015 Predicting Exchange Rates Out of Sample: Can Economic Fundamentals Beat the Random Walk? *Journal of Financial Econometrics*, 13, 293-341
- [47] R.B. Litterman. 1986. Forecasting with Bayesian vector autoregressions. Five years of experience. *Journal of Business and Economic Statistics*, 4, 25-38.
- [48] R. MacDonald, M.P. Taylor. 1994. The monetary model of the exchange rate: Long-run relationships, short-run dynamics and how to beat a random walk. *Journal of International Money and Finance*, 13, 276-290.
- [49] R. MacDonald, I.W. Marsh. 1994. Combining exchange rate forecasts: What is the optimal consensus measure? *Journal of Forecasting*, 13, 313-332.
- [50] N.C. Mark. 1995. Exchange rates and fundamentals: evidence on long-horizon prediction. *American Economic Review*, 85, 201-218.
- [51] N.C. Mark, D. Sul. 2001. Nominal exchange rates and monetary fundamentals: Evidence from a small post-Bretton Woods panel. *Journal of International Economics*, 53, 29-52.
- [52] R.A. Meese, K.S. Rogoff. 1983. Empirical exchange rate models of the seventies. Do they fit out of sample? *Journal of International Economics*, 14, 3-24.
- [53] L. Menkhoff, M.P. Taylor. 2007. The obstinate passion of foreign exchange professionals: Technical analysis. *Journal of Economic Literature*, 45, 936-972.
- [54] T. Molodtsova, A. Nikolsko-Rzhevskyy, D. Papell. 2011. Taylor Rules and the Euro. *Journal of Money, Credit, and Banking*, 43, 535-552.
- [55] C.J. Neely, P.A. Weller. 2013. Lessons from the evolution of foreign exchange trading strategies. *Journal of Banking & Finance*, 37, 3783-3798.
- [56] D. Politis, J. Romano. 1994. The stationary bootstrap. *Journal of the American Statistical Association*, 89, 1303-1313.

- [57] M. Qi, Y. Wu. 2006. Technical trading-rule profitability, data snooping, and reality check: Evidence from the foreign exchange market. *Journal of Money Credit and Banking*, 38, 2135-2158.
- [58] A.E. Raftery, D. Madigan, J.A. Hoeting. 1997. Bayesian model averaging for regression models. *Journal of the American Statistical Association*, 92, 179-191.
- [59] J. Romano, M. Wolf. 2005. Stepwise multiple testing as formalized data snooping. *Econometrica*, 73, 1237-1282.
- [60] X. Sala-i-Martin, G. Doppelhofer, R.I. Miller. 2004. Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach. *American Economic Review*, 94, 813-835.
- [61] L. Sarno, G. Valente, H. Leon. 2006. Nonlinearity in deviations from uncovered interest parity: An explanation of the forward bias puzzle. *Review of Finance*, 10, 443-482.
- [62] S. Sosvilla-Rivero, E. Garcí. 2005. Forecasting the Dollar/Euro Exchange Rate: Are International Parities Useful? *Journal of Forecasting*, 24, 369-377.
- [63] J.H. Stock, M.W. Watson. 2004. Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23, 405-430.
- [64] H. White. 2000. A reality check for data snooping. *Econometrica*, 68, 1097-1127.
- [65] J.-L. Wu, Y.-C. 2013. Fundamentals, forecast combinations and nominal exchange-rate predictability. *International Review of Economics and Finance*, 25, 129-145.
- [66] J. Yang, X. Su, J.W. Kolari. 2008. Do Euro exchange rates follow a martingale? Some out-of-sample evidence? *Journal of Banking and Finance*, 32, 729-740.

Appendix: Data description and sources

All time series have monthly periodicity (January 1980 to December 2013), and have been extracted from Thomson Financial Datastream. The variables used for EU-11 and USA are:

- Money supply: M1 aggregate, indexed 1990:1=100. Seasonally unadjusted.
- Output: Industrial production index 1990:1=100.
- Short term interest rate: 3-month interbank offered rate.
- Long term interest rate: 10-year rate interest rate on government bonds
- leading indicator for Germany as a proxy for Europe: IFO index
- leading indicator for US: ISM index
- Stock market indices covering at least 80% of market capitalization in the respective country.

Table 1: Models, combination methods and benchmarks

Abbreviations	Description
	Individual models
VAR(p)	Vector autoregression in levels based on domestic and foreign variables with p lags
DVAR(p)	Vector autoregression in first differences based on domestic and foreign variables with p lags
VEC(c,p)	Vector error correction model based on domestic and foreign variables with c cointegration relationships
r-VAR(p)	Restricted VAR, based on differences between domestic and foreign variables
r-DVAR(p)	Restricted DVAR, based on differences between domestic and foreign variables
r-VEC(c,p)	Restricted VEC, based on differences between domestic and foreign variables with c cointegration relationships
s-VAR(p)	Subset vector autoregression in levels based on domestic and foreign variables with p lags
s-DVAR(p)	Subset vector autoregression in first differences based on domestic and foreign variables with p lags
rs-VAR(p)	Restricted subset VAR, based on differences between domestic and foreign variables
rs-DVAR(p)	Restricted subset DVAR, based on differences between domestic and foreign variables
BDVAR(p)	Bayesian vector autoregression in first differences based on domestic and foreign variables
r-BDVAR(p)	Bayesian vector autoregression in first differences based on differences between domestic and foreign variables
	Forecast combination methods
mean	Forecasting combination based on mean of individual predictions
tmean	Forecasting combination based on trimmed mean of individual predictions
median	Forecasting combination based on median of individual predictions
OLS	Forecasting combination based on pooling using OLS
PC	Forecasting combination based on principal components
DMSFE	Forecasting combination based on discounted mean square forecast errors
HR	Forecasting combination based on hit rates
EHR	Forecasting combination based on exponential of hit rates
EEDF	Forecasting combination based on the economic evaluation of directional forecasts
BMA	Forecasting combination based on Bayesian model averaging weights using the predictive likelihood
FMA-aic	Forecasting combination based on AIC weights
FMA-bic	Forecasting combination based on BIC weights
FMA-hq	Forecasting combination based on Hannan-Quinn weights
	Benchmarks
RW	Random walk model (for MAE and MSE)
RW _{int}	Random walk model with intercept (for DA and DV)
BH	Buy-and-hold trading strategy (for TS ^S and TS ^F)
Forward rate	Rule based on the forward rate (for TS ^S and TS ^F)
MA(m, n)	Rule based on differences between moving averages over m and n periods (for TS ^S and TS ^F)
Filter	Filter rule based trading strategy (for TS ^S and TS ^F)
CT	Carry trade rule (for TS ^S and TS ^F)

Table 2: Forecasts of the monetary model for the EUR/USD exchange rate

Model	MAE	MSE	DA	DV	R ^S	R ^F	SR ^S	SR ^F
1-step								
r-VAR(12)	0.712	0.016	60.417	68.887	8.672	7.481	0.313	0.266
rs-VAR(2)	0.678	0.015	60.417	67.735	8.046	4.755	0.288	0.166
mean	0.712	0.016	54.167	62.856	5.751	3.801	0.202	0.132
tmean	0.710	0.016	54.167	60.016	4.490	3.801	0.156	0.132
median	0.690	0.015	58.333	63.627	6.066	4.755	0.213	0.166
OLS	0.769	0.019	52.083	59.360	4.267	3.177	0.148	0.110
PC	0.684	0.016	45.833	52.047	1.130	5.544	0.039	0.194
DMSFE	0.706	0.016	54.167	60.016	4.490	3.801	0.156	0.132
HR	0.712	0.016	54.167	62.856	5.751	3.801	0.202	0.132
EHR	0.732	0.018	50.000	52.306	0.892	6.643	0.031	0.235
EEDF	0.713	0.016	54.167	62.856	5.751	3.801	0.202	0.132
BMA	0.793	0.021	47.917	54.906	2.056	6.643	0.071	0.235
FMA-aic	0.755	0.019	45.833	47.117	-1.405	6.103	-0.048	0.215
FMA-bic	0.716	0.017	58.333	67.405	7.771	3.243	0.278	0.112
FMA-hq	0.733	0.017	52.083	61.683	5.195	3.801	0.182	0.132
RW	0.686	0.016	47.917	46.675				
MA(1,2)					6.893	6.893	0.244	0.244
6-steps								
r-VAR(12)	2.556	0.192	50.000	61.837	2.713	2.140	0.210	0.164
r-VEC(2,11)	2.617	0.206	52.083	60.650	2.360	-0.136	0.182	-0.010
s-DVAR(1)	1.955	0.122	45.833	42.485	-1.574	-1.156	-0.12	-0.088
rs-VAR(2)	1.990	0.117	52.083	62.135	2.666	0.955	0.206	0.073
mean	2.136	0.138	35.417	36.835	-2.796	-0.998	-0.217	-0.076
tmean	2.119	0.137	37.500	39.890	-2.168	-1.072	-0.166	-0.081
median	1.998	0.124	37.500	39.879	-2.191	-1.046	-0.168	-0.079
OLS	2.822	0.255	43.750	45.732	-0.829	-1.403	-0.063	-0.107
PC	1.896	0.109	72.917	80.388	6.601	<i>7.024</i>	<i>0.579</i>	<i>0.631</i>
DMSFE	2.103	0.134	33.333	36.727	-2.818	-1.095	-0.219	-0.083
HR	2.168	0.142	41.667	46.645	-0.696	-0.998	-0.053	-0.076
EHR	3.112	0.284	45.833	54.930	1.010	0.041	0.077	0.003
EEDF	2.181	0.144	37.500	39.815	-2.184	-0.998	-0.168	-0.076
BMA	2.743	0.224	41.667	45.052	-1.098	-1.481	-0.083	-0.113
FMA-aic	2.398	0.177	47.917	53.282	0.689	0.009	0.052	0.001
FMA-bic	2.033	0.131	45.833	46.471	-0.813	-0.628	-0.062	-0.048
FMA-hq	2.174	0.150	43.750	43.262	-1.464	-0.845	-0.112	-0.064
RW	1.916	0.113	50.000	41.065				
CT					4.116	4.116	0.328	0.328
12-steps								
r-VAR(12)	3.660	0.433	50.000	64.691	1.943	1.133	0.254	0.145
rs-VAR(2)	2.622	0.191	56.25	61.821	<u>1.583</u>	0.509	0.205	0.065
r-BDVAR(4)	2.487	0.188	39.583	32.598	-2.311	-3.075	-0.306	-0.423
mean	3.027	0.255	33.333	36.021	-1.820	-2.074	-0.237	-0.272
tmean	2.951	0.240	35.417	38.385	-1.525	-1.84	-0.197	-0.239
median	2.670	0.193	37.500	38.612	-1.522	-1.635	-0.196	-0.211
OLS	9.012	6.963	52.083	58.318	1.098	1.107	0.140	0.141
PC	2.414	0.172	72.917	83.300	4.379	<i>4.406</i>	<i>0.669</i>	0.674
DMSFE	3.061	0.259	33.333	36.021	-1.820	-2.498	-0.237	-0.333
HR	3.139	0.275	33.333	36.021	-1.820	-2.369	-0.237	-0.314
EHR	4.130	0.491	25.000	25.093	-3.250	-3.310	-0.452	-0.462
EEDF	3.179	0.284	33.333	36.021	-1.820	-2.378	-0.237	-0.316
BMA	3.737	0.443	37.500	41.591	-1.099	-1.629	-0.140	-0.211
FMA-aic	3.408	0.355	39.583	44.390	-0.745	-1.347	-0.095	-0.173
FMA-bic	2.805	0.224	41.667	38.670	-1.506	-2.340	-0.194	-0.310
FMA-hq	3.129	0.284	41.667	44.201	-0.796	-1.551	-0.101	-0.200
RW	2.330	0.163	43.750	41.265				
MA(12,14)					2.335	2.335	0.310	0.310

See Table 1 for the abbreviation of the models. Underlined **bold** figures indicate that the null hypothesis that the model does not outperform the benchmark model is rejected at the 5% significance level and underlined *italic* figures indicate that the null hypothesis is rejected at the 10% significance level using the stepM-SPA test. The stepM-SPA test is performed setting the dependence parameter q equal to 0.9 and the number of bootstrap simulations is equal to 5000. Both MAE and MSE loss measures are reported in exchange rate levels.

Table 3: Forecasts of the capital flows model for the EUR/USD exchange rate

Model	MAE	MSE	DA	DV	R ^S	R ^F	SR ^S	SR ^F
1-step								
r-DVAR(11)	0.729	0.017	54.167	62.952	5.817	10.163	0.204	0.374
rs-DVAR(1)	0.652	0.015	<i>66.667</i>	<i>70.078</i>	9.149	4.719	0.332	0.165
r-BDVAR(4)	0.662	0.015	<i>66.667</i>	<i>71.955</i>	10.026	4.755	0.368	0.166
mean	0.686	0.016	56.250	56.644	3.063	10.163	0.106	0.374
tmean	0.685	0.016	56.250	56.644	3.063	10.163	0.106	0.374
median	0.684	0.016	60.417	60.497	4.835	7.861	0.169	0.281
OLS	0.870	0.027	47.917	47.895	-1.055	6.737	-0.036	0.238
PC	0.657	0.015	54.167	57.24	3.435	8.076	0.119	0.289
DMSFE	0.677	0.015	56.250	56.644	3.063	10.163	0.106	0.374
HR	0.684	0.016	56.250	56.644	3.063	10.163	0.106	0.374
EHR	0.678	0.015	<i>66.667</i>	75.064	11.365	7.597	0.425	0.271
EEDF	0.685	0.016	56.250	56.644	3.063	10.163	0.106	0.374
BMA	0.790	0.021	<i>62.500</i>	59.263	4.221	8.487	0.147	0.305
FMA-aic	0.727	0.018	60.417	58.135	3.743	7.792	0.130	0.278
FMA-bic	0.679	0.016	58.333	56.500	3.092	4.755	0.107	0.166
FMA-hq	0.679	0.016	58.333	59.368	4.327	8.155	0.151	0.292
RW	0.686	0.016	47.917	46.675				
MA(1,2)					6.893	6.893	0.244	0.244
6-steps								
rs-VAR(2)	2.035	0.128	54.167	62.143	2.689	2.318	0.208	0.178
rs-DVAR(1)	1.949	0.122	54.167	52.331	0.428	-0.349	0.032	-0.026
r-BDVAR(4)	1.951	0.121	52.083	48.323	-0.428	1.570	-0.032	0.12
mean	2.062	0.137	47.917	46.624	-0.715	-0.358	-0.054	-0.027
tmean	2.083	0.138	45.833	45.969	-0.853	-0.832	-0.065	-0.063
median	2.110	0.136	45.833	45.969	-0.853	-1.090	-0.065	-0.083
OLS	2.618	0.252	47.917	48.809	-0.115	0.479	-0.009	0.036
PC	1.813	0.101	75.000	81.840	6.930	<i>7.137</i>	<i>0.619</i>	0.645
DMSFE	2.035	0.133	45.833	43.644	-1.327	-0.648	-0.101	-0.049
HR	2.071	0.137	47.917	46.624	-0.715	-0.648	-0.054	-0.049
EHR	2.302	0.150	37.500	46.734	-0.552	0.037	-0.042	0.003
EEDF	2.087	0.139	47.917	46.624	-0.715	-0.358	-0.054	-0.027
BMA	3.048	0.261	43.750	43.655	-1.344	-0.950	-0.102	-0.072
FMA-aic	2.464	0.183	43.750	45.865	-0.875	-0.868	-0.066	-0.066
FMA-bic	1.970	0.125	41.667	41.620	-1.788	0.649	-0.137	0.049
FMA-hq	2.069	0.138	39.583	40.221	-2.088	-0.363	-0.160	-0.028
RW	1.916	0.113	50.000	41.065				
CT					4.116	4.116	0.328	0.328
12-steps								
DVAR(12)	3.216	0.269	45.833	59.200	1.248	0.761	0.160	0.097
rs-VAR(2)	2.850	0.246	56.250	58.735	1.180	1.845	0.151	0.240
r-BDVAR(4)	2.464	0.186	39.583	32.598	-2.311	-1.971	-0.306	-0.258
mean	2.659	0.197	45.833	49.934	0.052	-0.618	0.007	-0.078
tmean	2.683	0.199	39.583	45.645	-0.516	-1.259	-0.065	-0.161
median	2.653	0.199	31.250	33.107	-2.178	-1.514	-0.287	-0.195
OLS	6.391	2.746	54.167	49.094	-0.181	-0.267	-0.023	-0.034
PC	2.474	0.170	77.083	82.969	4.324	4.370	<i>0.656</i>	<i>0.666</i>
DMSFE	2.616	0.193	45.833	49.934	0.052	-1.009	0.007	-0.129
HR	2.684	0.201	45.833	49.934	0.052	-0.618	0.007	-0.078
EHR	2.781	0.221	50.000	61.439	1.545	-0.978	0.199	-0.125
EEDF	2.689	0.202	45.833	49.934	0.052	-0.618	0.007	-0.078
BMA	3.983	0.431	35.417	41.152	-1.096	-2.702	-0.140	-0.364
FMA-aic	3.229	0.284	33.333	35.863	-1.795	-2.411	-0.233	-0.321
FMA-bic	2.522	0.191	37.500	32.177	-2.352	-1.184	-0.312	-0.152
FMA-hq	2.736	0.212	35.417	36.085	-1.804	-2.083	-0.235	-0.273
RW	2.330	0.163	43.750	41.265				
MA(12,24)					2.335	2.335	0.310	0.310

See Table 1 for the abbreviation of the models. Underlined **bold** figures indicate that the null hypothesis that the model does not outperform the benchmark model is rejected at the 5% significance level and underlined *italic* figures indicate that the null hypothesis is rejected at the 10% significance level using the stepM-SPA test. The stepM-SPA test is performed setting the dependence parameter q equal to 0.9 and the number of bootstrap simulations is equal to 5000. Both MAE and MSE loss measures are reported in exchange rate levels.