Confidence must be earned

Amundi

ASSET MANAGEMENT

Working Paper | Cross asset Investment Strategy

WP-91-2019 | September 2019

Forecasting Combination: An Application For Exchange Rates

Research & Macro Strategy

Forecasting Combination:
An Application For Exchange Rates*

Elisa Baku
Paris School of Economics,
University Paris 1 Pantheon
Sorbonne
Elisa.Baku@etu.univ-paris1.fr

Edmond Lezmi
Quantitative Research
edmond.lezmi@amundi.com

Abstract

This paper tries to forecast exchange rates by comparing forecasting methods that take into account cointegration and methods that do not. The first finding is that taking into account cointegration provides better forecasting results.

Furthermore, the factor model with cointegration provides the smallest forecasting errors, but when compared with penalized maximum likelihood, the differences are not always significant. In addition, we show that a forecast combination of all the methods used provides better exchange rates forecast accuracy.

Keywords: Bayesian models, cointegration, factor model, forecasting exchange rates, penalized maximum likelihood, sparse estimation.

JEL classification: C01, C53, F31, F37

*Views expressed in this paper are those of the authors and do not necessarily reflect those of Amundi. The authors are very grateful to Jean-Bernard Chatelain, Catherine Doz, Thierry Roncalli and Ines Wilms for their helpful comments and suggestions.

About the authors

Elisa Baku

Elisa Baku joined Amundi in 2016 in the Quantitative Research Team as a PhD candidate. She is specialized in currency markets and develops macro-econometric models for analyzing and forecasting foreign exchange rates. Prior to that, she was a graduate student from Paris School of Economics and she has worked as an intern in the economics department of the American Embassy in Paris in 2014.

She holds a Bachelor Degree in Finance and Accounting at University of Tirana, a Master's in Economics from University Paris 1 Pantheon Sorbonne, and a Master's in Economic Theory and Empirics from Paris School of Economics and University of Paris 1 Pantheon Sorbonne. Currently she is a PhD candidate in Macroeconomics and Finance at Paris School of Economics. Since 2015, she is teaching assistant for graduate courses in Econometrics at University Paris 1 Pantheon Sorbonne.

Edmond Lezmi

Edmond Lezmi joined Amundi in 2002. He is currently Head of Multi-Asset Quantitative Research. Prior to that, he was Head of Quantitative Research at Amundi Alternative Investments (2008-2012), a derivatives and fund structurer at Amundi IS (2005-2008), and Head of Market Risk (2002-2005). Before joining Amundi, he was Head of Market Risk at Natixis, and an exotic FX derivatives quantitative developer at Société Générale. He started his working career with Thales in 1987 as a research engineer in signal processing. He holds an MSc in Stochastic processes from the University of Orsay.

1 Introduction

Forecasting macroeconomic variables is very important, not only for policy makers and monetary authorities, but also for governments. The majority of the forecasting literature does not account for cointegration. Wilms and Croux (2016), to our knowledge, are the pioneers of developing a cointegration method for high dimensional time series, called the penalized maximum likelihood. Their method consists in estimating the cointegration vectors in a sparse way[1]. They showed that the sparse cointegration method provides smaller forecasting errors for interest rate and consumption forecasting, compared to the traditional Maximum Likelihood estimator.

Historically, researchers have been using the vector auto regressions (VAR) model to forecast different macroeconomic variables. Only recently have they brought to attention that the VAR model has some limitations such as not imposing restrictions on the parameters used and including only a small number of variables. In order to overcome these limitations, a widely used method, which has proven successful in forecasting different time series, is the so-called factor model. Stock and Watson (2002) found that better forecasting results were obtained when extracting some factors from a large dataset and using them to augment a VAR model. Boivin and Giannoni (2006) linked DSGE models with factor analysis, while Giannoni *et al.*. (2008) merged different frequency data and used the factor model to form a forecast for real GDP, while Doz *et al.* (2011) developed the theory used by the former. Banerjee *et al.* (2014) applied the factor error correction model (FECM) introduced by Banerjee and Marcellino (2009) to forecast, among other applications, three bilateral exchange rates. They found that FECM provided lower MSE relative to the autoregressive model. Likewise, by using dynamic factor model, Ludvigson and Ng (2009) were able to show a link between macroeconomic fundamentals and bond returns. Moreover, Barigozzi *et al.* (2016), proposed a non-stationary dynamic factor model (DFM) for large datasets.

Alternative methods to improve forecasting accuracy, have been the prime focus of researchers. De Mol *et al.* (2008) used the ridge regression and concluded that for Bayesian regression to capture factors that explain most of the prediction's variation, the data should be highly collinear. Moreover, Litterman (1986) found that better forecasting results were obtained when the bayesian shrinkage was applied to a VAR model. Further, Banbura *et al.* (2010) showed that the bayesian VAR model was an appropriate tool for large panels of data and it provided better forecasting results than a normal VAR model. While the majority of the research has been focused on the small sample forecasts (Geweke, 1996; Reinsel, 1983; Camba-Mendez *et al.*, 2003), Carriero *et al.* (2011) focused on forecasting all the variables in a large dataset

[1]Sparse estimation has been shown to perform very well (Liao and Phillips, 2015; Fan *et al.*, 2011; Zhou *et al.*, 2014.). Some of the benefits of using sparse models are that they select only the variables that have non-zero coefficients, which improves model interpretation. Moreover, they reduce the variance that has a direct impact on the forecast performance and, lastly, they can be applied even when the number of time series exceeds the time series length (Wilms and Croux (2016).

using multivariate models. They considered three different models. The first model was a classical reduced rank regression (RR), as used by Velu *et al.* (1986). The second method applied bayesian shrinkage and rank reduction restrictions at the same time, called bayesian rank reduction (BRR), and they proposed a new method called reduced rank posterior (RRP), which applied rank reduction on the posterior estimates of a Bayesian VAR. They found that using shrinkage and rank reduction in combination improved the accuracy of forecasts. RRP and BRR performed very well in forecasting several US macroeconomic variables, such as industrial production growth, inflation and short-term interest rate.

The ability to forecast exchange rates remains debatable among economists. Meese and Rogoff (1983a, 1983b) demonstrated that the random walk forecasts exchange rates out-of-sample better than the monetary models. While Giacomini and Rossi (2010) and Rossi and Inoue (2012) found strong empirical evidence in favor of Taylor-rule fundamentals, Rogoff and Stavrakeva (2008) found that the empirical evidence in favor of Taylor-rule fundamentals was not robust. Among others, Chen and Rogoff (2003) focused on the commodity currencies and showed that the US dollar price of the commodity exports of Australia and New Zealand has an impact on their exchange rates movements. Moreover, Ferraro *et al.* (2015) showed the existence of a relationship between commodity prices and nominal exchange rates at daily frequency. Furthermore, Kilian and Zhou (2019) provided empirical support for the existence of the link between exchange rates, oil prices and interest rates.

Our goal in this paper is to try to forecast exchange rate movements by comparing different forecasting methods that have been proven to provide good forecasting accuracy when they have been used for different time series forecast. We acknowledge that forecasting exchange rates is difficult but as Henri Poincare states:

> "*It is better to foresee even without certainty than to not foresee at all*"
> (Henri Poincare in The Foundations of Science, 1913, page 129).

Following Wilms and Croux (2016), we apply their methodology to the exchange rates framework. The methods taken into consideration can be devised into two main groups of forecasting methods. In the first group are the methods that account for cointegration, such as penalized maximum likelihood estimation of the VECM and factor model for non-stationary time series. In the second group are the methods that do not account for cointegration such as Penalized maximum likelihood of the VAR, , factor model for stationary time series, bayesian estimation of the VAR and bayesian reduced rank regression.

Going further than Wilms and Croux (2016), we apply the so called "hedging" against model risk (Raviv *et al.*, 2016), which consists in aggregating different forecasts, inspired from the ensemble methods in machine learning[2]. Many researchers have used the forecast combination such as Stock and Watson (2004), Hansen (2008), Ravazzolo *et al.* (2007). On the exchange rates forecasts, Wright (2008) applied bayesian model averaging (BMA), which consists in taking forecasts from different

[2] Ensemble methods is a machine learning technique that combines several predictive models in order to produce the best predictive model. The "ForecastCombinations" package was used in R.

models and weighting them by the posteriors probabilities. By using this method, he showed positive out-of-sample forecasting results.

Furthermore, unlike other papers, we focus on forecasting exchange rates by their own movements. When doing so, we show that substantial gains can be obtained when accounting for cointegration and, consistent with Wilms and Croux (2016), we found that sparse estimation provides better forecasting results. Differently from them, the method that performs the best in terms of forecasting accurancy is the factor model with cointegration, even though the differences with the PML are small and not always significant. Likewise, we obtain remarkable results, when we do a forecast combination of all the methods used.

The remainder of the paper is as follows. Section two describes the forecasting models. Section three discusses the forecast accuracy. Section four presents the data based on the methodology used. The empirical results will be provided in Section five. Lastly, Section six will conclude.

2 Forecasting models

Our objective is to forecast exchange rates based on their own movements, by comparing different forecasting methods. Following Wilms and Croux (2016), the forecasting models used for the purpose of this paper are:

- Penalized maximum likelihood (PML) estimation of the VECM of Wilms and Croux (2016);
- Factor model for non-stationary time series of Barigozzi *et al.* (2017);
- Penalized maximum likelihood estimation of the VAR;
- Factor model for stationary time series of Stock and Watson (2002);
- Bayesian estimation of the VAR (BVAR) of Banbura *et al.* (2010);
- Bayesian reduced rank regression (BRR) of Carriero *et al.* (2011).

We compare the performance of the above six estimators, by performing a rolling window forecast, where the first two estimators account for cointegration and the other estimators do not account for cointegration[3].

2.1 Penalized maximum likelihood estimation of the VECM and VAR

Letting y_t be a q-dimensional multivariate time series and assuming it follows a VAR(p), which can be written as the following VECM representation:

$$\Delta y_t = \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \Pi y_{t-1} + \varepsilon_t, \qquad t = p+1, ..., T \tag{1}$$

where Γ_i and Π are both $q \times q$ matrices, [4] and ε_t is assumed to follow a $\mathcal{N}_q(0, \Sigma)$. Following Wilms and Croux (2016), being able to express $\Pi = \alpha\beta^\top$ where α and β are $q \times r$ matrices of full column rank r, then the linear combinations given by $\beta^\top y_t$ will be stationary and y_t will then be co-integrated with cointegration rank r, where the columns of β will be the co-integrating vectors and the elements of α will be the adjustment coefficients.

Rewritting Equation (1) in matrix form we will end up having the following:

$$\Delta Y = \Delta Y_L \Gamma + Y \Pi^\top + E \tag{2}$$

where $\Delta Y = (\Delta y_{p+1}, ..., \Delta y_T)^\top$; $\Delta Y_L = (\Delta X_{p+1}, ..., \Delta X_T)^\top$ with $\Delta X_t = (\Delta y_{t-1}^\top, ..., \Delta y_{t-p+1}^\top)$; $Y = (y_p, ..., y_{T-1})$; $\Gamma = (\Gamma_1, ..., \Gamma p - 1)$; and $E = (\varepsilon_{p+1}, ..., \varepsilon_T)$. The

[3] The methods are applied on log-transformed time series

[4] The former one captures the short-run effects and the latter one the rank r, $0 \le r \le q$

penalized negative log-likelihood for Equation (2) will be:

$$\mathcal{L}(\Gamma, \Pi, \Omega) = \frac{1}{T} tr \left(\left(\Delta Y - \Delta Y_L \Gamma - Y \Pi^\top \right) \Omega \left(\Delta Y - \Delta Y_L \Gamma - Y \Pi^\top \right)^\top \right) \\ - log|\Omega| + \lambda_1 P_1(\beta) + \lambda_2 P_2(\Gamma) + \lambda_3 P_3(\Omega) \tag{3}$$

where $\Omega = \Sigma^{-1}$, $tr(\cdot)$ denotes the trace and P_1, P_2 and P_3 are the penalties. Wilms and Croux (2016) used L_1 penalization on the cointegration vectors[5] β, on the short-run effects Γ and on the inverse error covariance matrix Ω. By doing so they are able to have a sparse solution of Equation (3), where some elements of β, Γ and Ω will be estimated as zero.

The goal is to select Γ, Π, Ω so as to minimize (3) subject to the constraint

$$\Pi = \alpha \beta^\top \tag{4}$$

where the matrices α and β are not uniquely defined and the normalization condition $\alpha^{'} \Omega \alpha = I_r$ is imposed for identifiability purposes. To find the minimum of the penalized negative log-likelihood in (3), one must iteratively solve for Π conditional on Γ, Ω; for Γ conditional on Π, Ω; and for Ω conditional on Γ, Π. Solving for Π conditional on Γ, Ω. The minimization problem for each one of the cases is shown[6] in Appendix B.1.

2.2 Factor model for stationary and non-stationary time series

Factor models have been recently widely used in the macroeconomic literature. For the purposed of this article two papers were used – Stock and Watson (2002) and Barigozzi *et al.* (2017) – the former as a representative of the use of factor model for stationary time series and the later as a representative of non-stationary time series. Following Stock and Watson (2002), we write the dynamic factor model[7] as:

$$y_{t+1} = \beta(L) f_t + \gamma(L) y_t + \varepsilon_{t+1} \tag{5}$$

$$X_{i,t} = \lambda_i(L) f_t + e_{i,t} \tag{6}$$

where y_{t+1} are the series to be forecast, X_t is a N-dimensional time series of predictor variables for $t = 1, ..., T$, f_t are the dynamic factor model for $i = 1, ...N$, $e_i = (e_{1,t}, ..., e_{N,t})$ is the $N \times 1$ idiosyncratic disturbance and $\lambda_i(L)$ and $\beta(L)$ are lag polynomials of L. A static representation of Equation (5) and (6) will be the following (Stock and Watson, 2002):

$$y_{t+1} = \beta^\top F_t + \gamma(L) y_t + \varepsilon_{t+1} \tag{7}$$

[5] $P_1(\beta) = \sum_{i=1}^{q} \sum_{j=1}^{r} |\beta_{ij}|$ is known as the Lasso (Tibshirani, 1996).

[6] For the unpenalized case ($\lambda_1 = 0$, $\lambda_2 = 0$ and $\lambda_3 = 0$); the objective function in Equation (3) boils down to the one introduced by Johansen (1988) or by the iterative algorithm described by Wilms and Croux (2016), which is described in Appendix B.1.

[7] The difference between the dynamic and static factor models is that the former considers that the lags of the factors will also affect the variables

$$X_t = \Lambda F_t + e_t \tag{8}$$

where $F_t = (f_t, ..., f_{t-q})$ is $r \times 1$, Λ is $(\lambda_{i0}, ..., \lambda_{iq})$, and $\beta = (\beta_0, ..., \beta_q)$. Proceeding in this way allows us to estimate the factors using the principal components method. Furthermore, their approach of forecasting it consists in h-step-ahead forecasts:

$$y_{t+h}^h = \alpha_h + \beta_h(L)F_t + \gamma_h(L)y_t + \varepsilon_{t+h}^h \tag{9}$$

where y_{t+h}^h is the h-step-ahead variable to be forecast and α_h is the constant term. In order to forecast, following them, first we extract the factors from the data, then we regress y_{t+1} into the estimated factors, a constant and y_t. By doing so, we get the $\hat{\alpha}_h$, $\hat{\beta}_h(L)$ and $\hat{\gamma}_h(L)$ and are able to build the forecast on y_{t+h}^h as $\hat{\alpha}_h + \hat{\beta}_h(L)F_t + \hat{\gamma}_h(L)y_t$.

Factor model without cointegration (Stock and Watson, 2002; Bai and Ng, 2002) estimated the factors and their loadings by the use of principal component, while the matrix of the lag operator is usually obtained by using the VAR model. For the variables that are not stationary, before extracting the factors, the first difference is taken. On the other hand, Barigozzi *et al.* (2017) considered the case of non-stationary variables. Under the assumptions that F_t are $I(1)$, singular and cointegrated[8]) and the idiosyncratic components are $I(1)$ or $I(0)$, they model F_t as a vector error correction model (VECM)[9].

2.3 Bayesian estimation of the VAR (BVAR) and BRR

When VAR is used in macroeconomic models, usually researchers have to make a choice on the number of variables, due to the fact that VAR has a limitation on the number of variables to be included. The Bayesian VAR model applies the bayesian shrinkage in order to handle large VAR.

Suppose we have the following VAR(p) model:

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p} + u_t \tag{10}$$

where $Y_t = (y_{1,t}, y_{2,t}, ..., y_{N,t})$ is a vector of variables, $X_t = (Y_{t-1}, Y_{t-2}, ..., Y_{t-p})$ for $t = 1, ..., T$ when means and trends have been removed. Defining $B = (A_1, A_2, ..., A_p)$, the above equation can be re-written as:

$$Y_t = B^\top X_t + u_t \tag{11}$$

Banbura *et al.* (2010) estimated the above model using a Bayesian VAR, by imposing prior beliefs on the parameters, known as Minnesota prior[10]. The intuition behind the Minnesota prior is that it shrinks the diagonal elements of A_1 towards

[8]Noting c as the cointegration rank. c is at least $r - q$, where q is the number of shocks, that is $c = r - q + d$ with $0 \leq d < q$, hence singular $I(1)$ vectors are always cointegrated. See, Barigozzi *et al.*(2017) for a full explanation on the model implementation.

[9]A VECM(p) with cointegration rank c can also be written as a VAR($p + 1$) with $r - c$ unit roots. So, a VECM is easily written as a VAR model.

[10]Like Banbura *et al.* (2010), we follow the version of the Minnesota prior proposed by Litterman with modifications made by Kadiyala and Karlsson (1997) and Sims and Zha (1998).

one and the remaining coefficients towards zero. Consolidating in this way the idea that more trustworthy information about one variable is given by its own recent lags (Banbura *et al.*, 2010). Under the Minnesota prior, the prior expectations and variances are:

$$\mathcal{E}\left[A_k^{(i,j)}\right] = \begin{cases} \delta_i & \text{for } j = i, k = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathcal{V}\left[A_k^{(i,j)}\right] = \begin{cases} \phi k^{-2} & \text{for } j = i, \forall k \\ \phi k^{-2}\theta\sigma_i^2\sigma_j^{-2} & \text{for } j \neq i, \forall k \end{cases}$$

where ϕ captures the overall tightness of the prior[11], the factor k^{-2} is the rate at which prior variance decreases with increasing lag length, while $\sigma_i^2\sigma_j^{-2}$ accounts for the different scale and variability of the data. Furthermore, the parameter $\theta \in [0,1]$ controls for the fact the other variables lags are not as important as its own lags.

Kadiyala and Karlsson (1997) modified the Minnesota prior proposed by Litterman, by setting the $\theta = 1$, in order to have more efficient computations and to avoid the assumptions of fixed and diagonal residual variance matrix. Following them we can write Equation (11) as a multivariate regression:

$$Y = XB + U \tag{12}$$

where $Y = (Y_1, ..., Y_T)'$ is a $T \times N$ matrix of dependent variables, and $X = (X_1, ..., X_T)'$ is a $T \times M$ matrix of explanatory variable, where $M = Np$. The matrix $U = (u_1, u_2, ..., u_T)'$ is the matrix of disturbances, which are assumed to be independent and identically distributed across observations; We define r as the rank of the $M \times N$ matrix of coefficients B, where $r \leq N$. The normal-inverted Wishart prior has the form[12]:

$$B|\Sigma \sim \mathcal{N}(B_0, \Sigma \otimes \Omega_0); \qquad \Sigma \sim IW(\vartheta_0, S_0) \tag{14}$$

where the parameters ϑ_0, S_0, B_0, Ω_0 are such that the expectation of the Σ matrix is equal to the fixed residual covariance matrix of the Minnesota prior, and the prior expectation and variance of B are that of the Minnesota prior with $\theta = 1$ (Carriero *et al.*, 2011).

As we stated earlier, the estimation of VAR model provides a lot of insignificant coefficients. Another way to handle the insignificant coefficients that result from the VAR estimation is the bayesian reduced rank (BRR), which is a combination of bayesian VAR and reduced rank. Geweke (1996) was the first to introduce such

[11] When $\phi = 0$, the prior is exactly imposed and the estimates are not influenced by the data, while as ϕ goes to ∞ the posterior estimates approach the OLS estimates (Carriero *et al.*, 2011). De Mol *et al.* (2008) showed that the more variables you have the more you should shrink the ϕ in order to avoid possible overfitting issues.

[12] The conditional posterior distributions are of the normal-inverted Wishart form as well:

$$B|\Sigma, Y \sim \mathcal{N}(\overline{B}, \Sigma \otimes \overline{\Omega}); \qquad \Sigma|Y \sim IW(\overline{\vartheta}, \overline{S}) \tag{13}$$

where the bar denotes that parameters of the posterior distribution. See Banbura *et al.* (2010) and Carriero *et al.* (2011) for more details.

a model, and then BRR was re-implemented by Carriero *et al.* (2011) for a larger dataset. The BRR not only uses shrinkage (as the BVAR does) but also rank reduction, such as imposing B to have a rank r, where $r < N$[13]. Such assumption corresponds to the following parametric specification:

$$Y = X\Psi\Phi + E \tag{15}$$

where Ψ and Φ are $M \times r$ and $r \times N$, respectively, matrices. Following Geweke (1996), the normalization used to identify Ψ and Φ is (See Appendix B.2 for more details)[14]:

$$\Phi = [I_r \Phi^*] \tag{16}$$

2.4 Forecast Combination

Going further, we analyze different forecast combination methods, known in the literature as "hedging against model risk" in order to examine whether a combination of all the forecasts provide better forecasting results. Following Raviv (2016), the following forecast methods are used:

- **Simple average** is the simple average of all the forecasting methods (Clemen, 1989; Genre *et al.*, 2013). The combined forecast will be then given by the following equation:

  $$F^c = \frac{\sum_{i=1}^{N} F_i}{N} \tag{17}$$

  where F^c is the combined forecast, F_i is the forecast obtained from each one of the six forecasting methods explained in Section 2 and N is the number of forecasts.

- **Ordinary least squares (OLS)** regression, the combined forecast is obtained as a linear function of the individual forecasts

  $$F^c = \hat{\alpha} + \sum_{i=1}^{N} \hat{\beta}_i F_i \tag{18}$$

  where $\hat{\alpha}$ and $\hat{\beta}_i$ are obtained by regressing the individual forecasts on the target itself (Granger and Ramanathan, 1984; Raviv, 2016).

- **Least absolute deviation (LAD)** regression, this method estimates the $\hat{\alpha}$ and $\hat{\beta}_i$'s not by minimizing the sum of squared errors like OLS, but by using the absolute sum of squares errors (Weiss *et al.*, 2018) .

- **Constrained Least Squares (CLS)** regression adds some constraints when minimizing the sum of squared errors i.e., constraining the $\hat{\beta}_i$ to allow only for positive solutions and to sum up to one.

[13] See Carriero *et al.* (2011) for an in depth explanation of the reduced rank regression (RR).

[14] For a discussion of the role of normalization in reduced rank models see, Kleibergen and van Dijk (1998) and Hamilton *et al.* (2007).

- **Variance-based**, the combined forecast is computed as follows:

$$F^c = \frac{\frac{1}{MSE_i}}{\sum_{i=1}^{N} \frac{1}{MSE_i}} F_i \tag{19}$$

where the MSE_i is the mean squared error, which is computed based on out-of-sample forecasts.

3 Forecast Accuracy

Diebold-Marino test was used to compare forecast performance among different methods (Diebold and Mariano, 1995). The cointegration rank is estimated using the rank selection criterion, which is explained in Appendix B.3 and the order of the VECM is chosen by the BIC criterion. The out-of-sample forecast accuracy is evaluated by performing rolling window forecasting, with window size S, which is the number of time points available for estimation. Different window sizes $S \in \{48, 96, 144\}$ were considered. Let h be the forecast horizon. We consider $h \in \{1, 3, 6, 12\}$. At each time point $t = S, ..., T - h$, the h-step-ahead forecasts of Equation (1) can be written as follows:

$$\Delta\hat{y}_{t+h} = \sum_{i=1}^{p-1} \hat{\Gamma}_i \Delta y_{t+1-i} + \Pi \hat{y}_t \tag{20}$$

for which the h-step-ahead multivariate forecast errors are computed: $\hat{e}_{t+h} = \Delta y_{t+h} - \Delta\hat{y}_{t+h}$.

In each simulation run, the overall multivariate forecast performance is then measured by the multivariate mean absolute forecast error (Carriero *et al.*, 2011):

$$MMAFE = \frac{1}{T-h-S+1} \sum_{t=S}^{T-h} \frac{1}{q} \sum_{i=1}^{q} \frac{|\Delta y_{t+h}^{(i)} - \Delta\hat{y}_{t+h}^{(i)}|}{\hat{\sigma}^{(i)}} \tag{21}$$

where $\hat{\sigma}^{(i)}$ is the standard deviation of the i^{th} time series in differences. The MMAFE depends on the forecast horizon h. Apart from the multivariate mean absolute forecast error, the mean absolute forecast error is computed for every individual time series:

$$MAFE = \frac{1}{T-h-S+1} \sum_{t=S}^{T-h} \frac{|\Delta y_{t+h}^{(i)} - \Delta\hat{y}_{t+h}^{(i)}|}{\hat{\sigma}^{(i)}} \tag{22}$$

Forecast errors are comparable among methods because the MMAFE and MAFE are computed for the time series in difference.

4 Data and summary statistics

The data used in the analysis consists of time series of the main frequently traded global exchange rates according to the 2013 Triennial Central Bank Survey from the

Bank for International Settlements (BIS)[15]. Data were collected at a monthly frequency from Bloomberg. The sample begins in December 2001 and ends in February 2016, and therefore contains a total of 171 months.

Table (4) in Appendix A.1 provides the list of all the currencies used. Exchange rates are the month-end values of the US Dollar versus the 30 most actively traded currencies. Some basic features of the data to help guide our empirical design are provided in Table (4). They provide reports, means and sample ranges of the variables. The mean varies between a segment of values of -0.51 and 9.19, where the extreme values belong to the Indonesian rupiah (IDR) and British pound (GBP), respectively.

The log of the nominal exchange rate levels was taken before implementing the methods discussed in Section 2. Figure (1) provides a graphical representation of all the currencies. Observing the movement of the exchange rates over time we can get the first intuition that the exchange rates might not be stationary.

Furthermore, Table (5) in Appendix A.2 shows the kurtosis for the exchange rates which are lower than three for all the countries. A value of kurtosis equal to three is considered to be the value for a normal distribution[16]. We proceed by performing two normality tests; Jarque-Bera and Shapiro-Wilk. From the results obtained, we reject the normality hypothesis of the test for all the currencies. So, we conclude that exchange rate changes are non-Gaussian for most countries and hence are not jointly and normally distributed.

5 Results

We firstly proceed by performing an ADF test to check whether the series are stationary or not. From Table (6) in Appendix A.2 we can conclude that the exchange rates in level are not stationary, but their first difference is stationary[17], which confirms that time series are integrated of order one.

The multivariate mean forecast errors (MMFE) were computed for three different window sizes; $60, 72$ and 144 [18] and for four forecast horizons $h = 1, 3, 6, 12$, for all the forecasting methods. The results reported in Tables (1), (2) and (3). The lowest forecasting errors are obtained by the dynamic factor model (DFM) that accounts for cointegration, as the values in bold indicate. The DFM that accounts for cointegration, for the first two window sizes, is superior compared to the other methods for all the forecast horizons. For the last window size, the results are relatively mixed, as shown in Table (3). For the first forecast horizon, the bayesian

[15] The same sample was used by Baku (2018) to create the global variable.

[16] In this case the distribution is called platykurtic and its tails are shorter and thinner and often its central peak is lower and broader, compared to a normal distribution.

[17] The p-value for the time series in levels is more than 5%, which means that we fail to reject the null hypothesis of non-stationarity, while the p-value for the time series in differences is smaller than 5%, so we reject the null hypothesis of non-stationarity.

[18] We also computed the multivariate mean forecast errors for the window size 96. The results were very similar to window size 72, so we did not provide the results for brevity reasons.

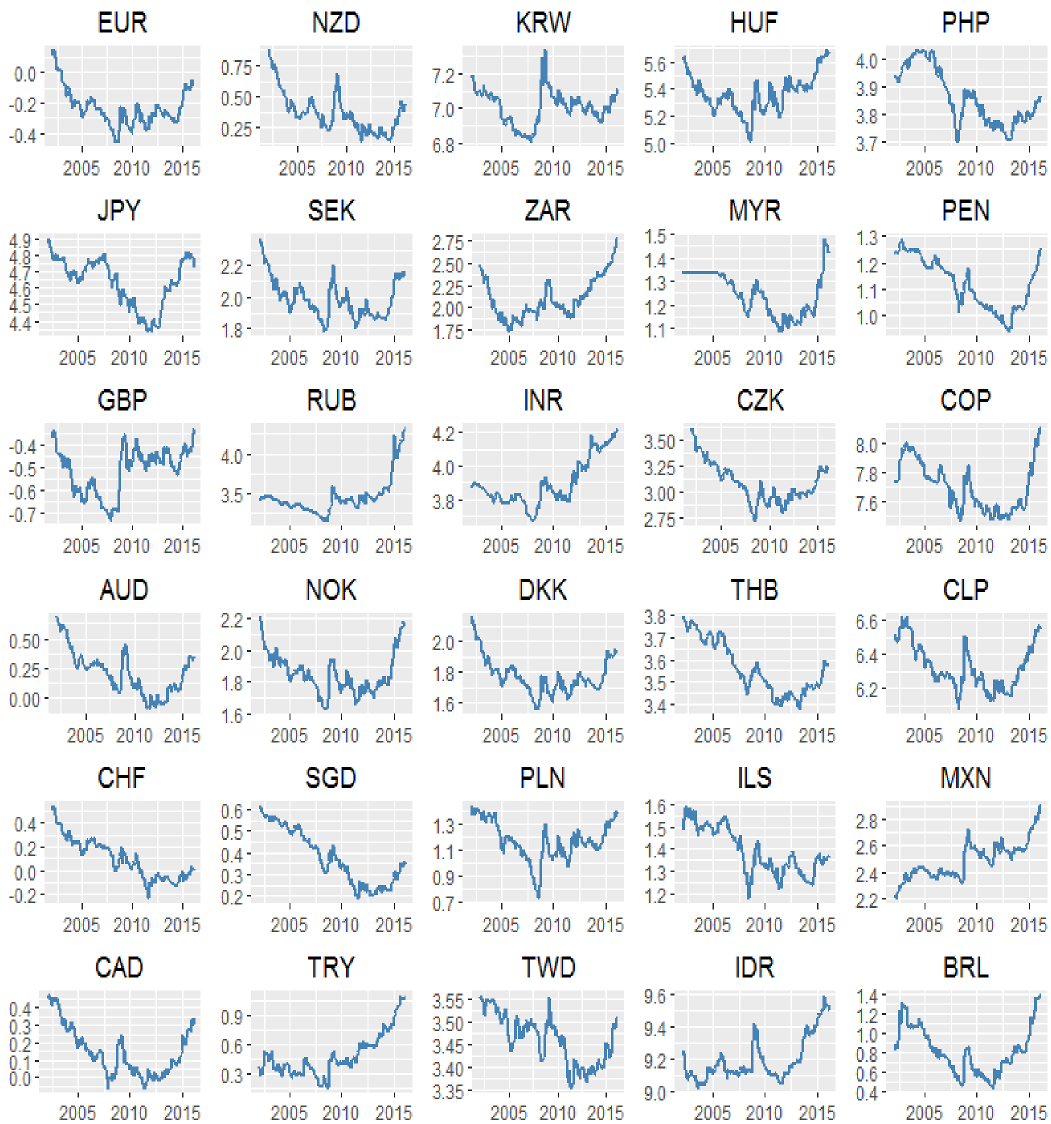


Figure 1: Exchange rate graphical representation

reduced rank provides the lowest forecasting errors followed by DFM that accounts for cointegration and VAR PML. When h is equal to two and six, the DFM and VAR penalized maximum likelihood provide the lowest forecasting errors, while when h is equal to 12, the factor model with and without cointegration provides the same forecasting results. Furthermore, among the methods that do not take into account cointegration, the Bayesian reduced rank and VAR penalized maximum likelihood perform better in long horizons.

Table 1: MMAFE for window size 60 and forecast horizon h

	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Coint PML	0.92	0.91	0.92	0.87
Coint DFM	**0.81***	**0.82***	**0.83***	**0.84***
VAR PML	0.91	0.86	0.86***	0.83***
BVAR	0.89	0.91	0.91	0.92
VAR BRR	0.89	0.91	0.91	0.91
VAR DFM	0.85	0.85	0.89	0.87

In addition, mean absolute forecast errors for each time series have been computed. From the results, shown in Tables (7), (8) and (9), we can conclude that the dynamic factor model that accounts for cointegration provides the smallest forecasting errors. This result is along the same lines as the one obtained by the multivariate mean forecast errors. Although it is important to state that the differences in terms of the forecasting errors between the DFM with cointegration and VAR PML are very small, for longer rolling windows they provide similar forecasting errors. Overall, we can conclude that taking into account cointegration will make the forecasting errors smaller and, of the methods that do take into account cointegration the best method is the dynamic factor model.

Table 2: MMAFE for window size 72 and forecast horizon h

	h =1	$h = 3$	$h = 6$	$h = 12$
Coint PML	0.97	0.90	0.91	0.82
Coint DFM	**0.83****	**0.82***	**0.83***	**0.77***
VAR PML	1.00	0.83**	0.85	**0.77***
BVAR	0.90	0.95	0.93	0.85
VAR BRR	0.89	0.95	0.92	0.85
VAR DFM	0.86	0.88	0.88	0.81

Moreover, we have implemented different forecast combination models, explained in Section 2.4. We did the forecast combination with and without the worst forecasting method, which was the maximum likelihood without cointegration. For reasons of brevity we have provided the results only for the window size of 60. Table

Table 3: MMAFE for window size 144 and forecast horizon h

	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Coint PML	0.71	0.72	0.80	0.81
Coint DFM	0.70	**0.71***	**0.76***	**0.77***
VAR PML	0.70	**0.71**	**0.76**	0.78
BVAR	0.71	0.88	0.88	0.90
VAR BRR	0.70	0.88	0.87	0.90
VAR DFM	**0.69**	0.74	0.78	**0.77**

(11) in Appendix C.1 shows the results for all the methods and for all the different forecasting horizons that were considered. As shown, the least absolute deviation (LAD) forecast method provides the smallest forecast errors, followed by ordinary least squares (OLS). Furthermore, graphical representations of the out-of-sample forecast are provided in Appendix A.6. Again for reasons of brevity we have only shown the graphical representations for three forecasting methods: LAD forecast combination, PML with cointegration and BVAR. Figures (2) and (3) show the out-of-sample forecasts for four commodities currencies, two for developed currencies – AUD, CAD – and two for emerging currencies – BRL, RUB. As shown the best forecasting method is the LAD. Furthermore, the out-of-sample forecasts for EUR, GBP, CHF and JPY are provided in Figures (4) and (5). The best forecast is provided by the LAD forecast combination method, while the forecast provided by BVAR comes with a lag for all these currencies. Lastly, two out-of-sample forecasts for the emerging currencies are shown in Figure (6). For the Turkish lira (TRY), the best forecasting method is PML with cointegration, while LAD is for Mexican peso (MXN).

Conclusively, we show that when forecasting highly collinear time series, accounting for cointegration provides substantial gains, and the sparse cointegration method achieves better forecasting results compared to more traditional methods. Furthermore, we show that even better forecasting results are obtained by implementing a forecast combination of all the forecasting methods used.

6 Conclusion

In this paper, we compare different exchange rate forecasting methods. The methods can be divided into two groups, that takes into account cointegration and that does not take into account cointegration. We conclude that taking cointegration into account provides better forecasting results and, the factor model followed by the penalized maximum likelihood method are the best performers. Moreover, this article confirms that sparsity provides better forecasting results. Furthermore, we combine the forecasts provided by all the methods implemented, in order to see whether a forecast combination of these methods can provide better forecasting results than individual model forecasts. We confirm that the LAD forecast combination provides better results, followed by OLS forecast combination.

Future research should be done in using other penalizations in the penalized likelihood method in order to see whether there could be additional gains in the forecast accuracy. Furthermore, it would be interesting to see whether the results will hold if a different sample is taken for forecasting exchange rates.

References

[1] Bai, J. and Ng, S. (2002), Determining the Number of Factors in Approximate Factor Models, *Econometrica*, 70(1), pp. 191-221.

[2] Baku, E. (2018), Exchange Rate Predictability in Emerging Markets, *International Economics*, 157(3), pp. 1-22.

[3] Banbura, M., Giannone, D. and Reichlin, L.(2010), Large Bayesian Vector Auto Regressions, *Journal of Applied Econometrics*, 25(1), pp 71-92.

[4] Banerjee, A. and Marcellino, M. (2009), Factor-Augmented Error Correction Models, *in: Castle and Shephard, The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, pp. 227-254.

[5] Banerjee, A., Marcellino, M. and Masten, I. (2014), Forecasting with Factor-Augmented Error Correction Models, *International Journal of Forecasting*, 30(3), pp. 589-612.

[6] Bank of International Settlements (2013), Trinnial Central Bank Survey: OTC Interest Rate Derivated Turnover in April 2013, *BIS Quarterly Review*, 30(3), pp. 589-612.

[7] Barigozzi, M., Lippi, M.and Luciani, M. (2017), Non-Stationary Dynamic Factor Models for Large Datasets, *Board of Governors of the Federal Reserve System*.

[8] Boivin, J. and Giannoni, M. (2006), DSGE Models in a Data-Rich Environment, *National Bureau of Economic Research*, 12772.

[9] Bunea, F., She, Y. and Wegkamp, M.H. (2011), Optimal Selection of Reduced Rank Estimators of High-Dimensional Matrices, *The Annals of Statistics*, 39(2), pp. 1282-1309.

[10] Camba-Mendez, G., Kapetanios, G., Smith, R.J. and Weale, M.R.(2003), Tests of Rank in Reduced Rank Regression Models, *Journal of Business & Economic Statistics*, 21(1), pp. 145-155.

[11] Carriero, A., Kapetanios, G. and Marcellino, M. (2011), Forecasting Large Datasets with Bayesian Reduced Rank Multivariate Models, *Journal of Applied Econometrics*, 26(5), pp. 735-761.

[12] Chen, Y.C. and Rogoff, K. (2003), Commodity Currencies, *Journal of International Economics*, 60(1), pp. 133-160.

[13] Chen, L. and Huang, J. Z. (2012), Sparse Reduced-Rank Regression for Simultaneous Dimension Reduction and Variable Selection, *Journal of the American Statistical Association*, 107(500), pp. 1533-1545.

[14] Clemen, R. T. (1989), Combining Forecasts: A Review and Annotated Bibliography, *International journal of forecasting*, 5(4), pp. 559-583.

[15] De Mol, C., Giannone, D. and Reichlin, L. (2008), Forecasting using a Large Number of Predictors: Is Bayesian Shrinkage a Valid Alternative to Principal Components?, *Journal of Econometrics*, 146(2), pp. 318-328

[16] Diebold, F. and Mariano, R. (1995), Comparing Predictive Accuracy, *Journal of Business & Economic Statistics*, 13(3), pp. 253-263.

[17] Doz, C., Giannone, D. and Reichlin, L. (2011), A Two-Step Estimator for Large Approximate Dynamic Factor Models based on Kalman Filtering, *Journal of Econometrics*, 164(1), pp. 188-205.

[18] Fan, J., Lv, J. and Qi, L. (2011), Sparse High-Dimensional Models in Economics, *Annual Review of Economics*, 3, pp. 291-317.

[19] Ferraro, D., Rogoff, K. and Rossi, B. (2015), Can Oil Prices Forecast Exchange Rates? An Empirical Analysis of the Relationship between Commodity Prices and Exchange Rates, *Journal of International Money and Finance*, 54, pp. 116-141.

[20] Genre, V., Kenny, G., Meyler, A., and Timmermann, A. (2013), Combining Expert forecasts: Can Anything Beat the Simple Average?, *International Journal of Forecasting*, 29(1), pp. 108-121.

[21] Geweke, J. (1996), Bayesian Reduced Rank Regression in Econometrics, *Journal of Econometrics*, 75(1), pp. 121-146.

[22] Giacomini, R. and Rossi, B. (2010), Forecast Comparisons in Unstable Environments, *Journal of Applied Econometrics*, 25(4), pp. 595-620.

[23] Giannone, D., Reichlin, L. and Small, D. (2008), Nowcasting: The Real-Time Informational Content of Macroeconomic Data, *Journal of Monetary Economics*, 55(4), pp. 665-676.

[24] Granger, C. W. and Ramanathan, R. (1984), Improved Methods of Combining Forecasts, *Journal of Forecasting*, 3(2), pp. 197-204.

[25] Hamilton, J.D., Waggoner, D.F. and Zha, T. (2007), Normalization in Econometrics, *Econometric Reviews*, 26(2-4), pp. 221-252.

[26] Hansen, B. E. (2008), Least-Squares Forecast Averaging, *Journal of Econometrics*, 146(2), pp. 342-350.

[27] Johansen, S. (1988), Statistical Analysis of Cointegration Vectors, *Journal of Economic Dynamics and Control*, 12(2-3), pp. 231-254.

[28] Kadiyala, K.R. and Karlsson, S. (1997), Numerical Methods for Estimation and Inference in Bayesian VAR-Models, *Journal of Applied Econometrics*, 12(2), pp. 99-132.

[29] Kilian, L. and Zhou, X. (2019), Oil Prices, Exchange Rates and Interest Rates, *CESifo Working paper*, 13478.

[30] Kleibergen, F. and Van Dijk, H.K. (1998), Bayesian Simultaneous Equations Analysis using Reduced Rank Structures, *Econometric Theory*, 14(6), pp. 701-743.

[31] Liao, Z. and Phillips, P. (2015), Automated Estimation of Vector Error Correction Models, *Econometric Theory*, 31, pp. 581-646.

[32] Lissitz, R.W., Schonemann, P.H., and Lingoes, J.C. (1976), A Solution to the Weighted Procrustes Problem in which the Transformation is in Agreement with the Loss Function, *Psychometrika*, 41(4), pp. 547-550.

[33] Litterman, R.B. (1986), Forecasting with Bayesian Vector Autoregressions–Five Years of Experience, *Journal of Business & Economic Statistics*, 4(1), pp. 25-38.

[34] Ludvigson, S.C. and Ng, S. (2009), Macro Factors in Bond Risk Premia, *Review of Financial Studies*, 22(12), pp. 5027-5067.

[35] Meese, R.A. and Rogoff, K.(1983), Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?, *Journal of International Economics*, 14(1-2), pp. 3-24.

[36] Meese, R.A. and Rogoff, K.(1983), The Out-of-Sample Failure of Empirical Exchange Rate Models: Sampling Error or Misspecification?, *In Exchange Rates and International Macroeconomics, University of Chicago Press*, pp. 67-112.

[37] Poincare, H. (1913), The Foundations of Science: Science and Hypothesis, The Value of Science, Science and Method.

[38] Ravazzolo, F., Verbeek, M. and Van Dijk, H.K. (2007), Predictive Gains from Forecast Combinations Using Time Varying Model Weights, *SSRN*, 1012574.

[39] Raviv, E. (2016), Forecast Combinations in R using the ForecastCombinations Package: A Manual.

[40] Reinsel, G.C. (1983), Some Results on Multivariate Autoregressive Index Models, *Biometrika*, 70(1), pp. 145-156.

[41] Rogoff, K.S. and Stavrakeva, V. (2008), The Continuing Puzzle of Short Horizon Exchange Rate Forecasting, *National Bureau of Economic Research*, 14071.

[42] Rossi, B. and Inoue, A. (2012), Out-of-Sample Forecast Tests Robust to the Choice of Window Size, *Journal of Business & Economic Statistics*, 30(3), pp. 432-453.

[43] Tibshirani, R. (1996), Regression Shrinkage and Selection via the Lasso, *Journal of the Royal Statistical Society Series B*, 58(1), pp. 267-288.

[44] Sims, C.A. and Zha, T. (1998), Bayesian Methods for Dynamic Multivariate Models, *International Economic Review*, 39(4), pp. 949-968.

[45] Stock, J.H. and Watson, M.W. (2002), Macroeconomic Forecasting using Diffusion Indexes, *Journal of Business and Economic Statistics*, 20(2), pp. 147-162.

[46] Stock, J.H. and Watson, M.W. (2004), Combination Forecasts of Output Growth in a Seven-Country Data Set, *Journal of Forecasting*, 23(6), pp. 405-430.

[47] Velu, R.P., Reinsel, G.C. and Wichern, D.W. (1986), Reduced Rank Models for Multiple Time Series, *Biometrika*, 73(1), pp. 105-118.

[48] Weiss, C.E., Raviv, E. and Roetzer, G. (2018), Forecast Combinations in R using the ForecastComb Package, *The R Journal*.

[49] Wilms, I., and Croux, C. (2016), Forecasting using Sparse Cointegration, *International Journal of Forecasting*, 32(4), pp. 1256-1267.

[50] Wright, J.H. (2008), Bayesian Model Averaging and Exchange Rate Forecasts, *Journal of Econometrics*, 146(2), pp. 329-341.

[51] Zhou, X., Nakajima, J. and West, M. (2014), Bayesian Forecasting and Portfolio Decisions using Dynamic Sparse Factor Models, *International Journal of Forecasting*, 30(4), pp. 963-980.

A Appendix

A.1 List of variables

Table 4: Summary statistics of the most traded exchange rates

Iso Code	Currency name	N	Mean	St. Dev.	Min	Max
BRL	Brazilian real	171	0.82	0.24	0.44	1.39
RUB	Russian ruble	171	3.46	0.24	3.15	4.32
ZAR	South African rand	171	2.11	0.23	1.73	2.77
TRY	Turkish lira	171	0.49	0.22	0.15	1.11
AUD	Australian dollar	171	0.21	0.19	−0.09	0.68
CZK	Czech koruna	171	3.09	0.19	2.72	3.61
COP	Colombian peso	171	7.70	0.17	7.47	8.10
NZD	New Zealand dollar	171	0.37	0.17	0.13	0.88
CHF	Swiss franc	171	0.21	0.19	−0.09	0.68
JPY	Japanese yen	171	4.64	0.15	4.33	4.90
PLN	Polish zloty	171	1.17	0.15	0.72	1.44
MXN	Mexican peso	171	0.82	0.24	0.44	1.39
CAD	Canadian dollar	171	0.15	0.14	−0.06	0.47
INR	Indian rupee	171	3.90	0.14	3.67	4.23
HUF	Hungarian forint	171	5.37	0.14	5.01	5.68
CLP	Chilean peso	171	6.33	0.13	6.08	6.62
SEK	Swedish krona	171	1.99	0.13	1.78	2.37
NOK	Norwegian krone	171	1.86	0.13	1.63	2.21
SGD	Singapore dollar	171	0.38	0.13	0.19	0.62
IDR	Indonesian rupiah	171	9.19	0.13	9.02	9.59
EUR	Euro	171	−0.23	0.12	−0.46	0.15
DKK	Danish krone	171	1.78	0.12	1.55	2.16
THB	Thai baht	171	3.56	0.12	3.38	3.79
PEN	Peruvian sol	171	1.12	0.10	0.94	1.29
GBP	British pound	171	−0.51	0.10	−0.73	−0.33
KRW	South Korean won	171	7.01	0.10	6.80	7.34
ILS	Israeli new shekel	171	1.39	0.10	1.17	1.59
PHP	Philippine peso	171	3.86	0.10	3.70	4.03
MYR	Malaysian ringgit	171	1.25	0.09	1.09	1.48
TWD	New Taiwan dollar	171	3.46	0.06	3.35	3.56

Source: BIS (2013).

A.2 Additional results

Table 5: Normality test

Exchange Rate	Kurtosis	Shapiro-Wilk	Jarque-Bera
SGD	−1.340	0.000	0.000
PEN	−1.320	0.000	0.003
PHP	−1.228	0.000	0.001
ILS	−1.201	0.000	0.005
THB	−1.199	0.000	0.001
TWD	−1.105	0.000	0.011
COP	−0.938	0.000	0.004
MYR	−0.910	0.000	0.030
CLP	−0.875	0.000	0.002
JPY	−0.834	0.000	0.001
GBP	−0.813	0.000	0.004
CHF	−0.574	0.000	0.032
BRL	−0.561	0.000	0.016
INR	−0.505	0.000	0.000
CAD	−0.486	0.000	0.000
AUD	−0.317	0.000	0.026
ZAR	−0.211	0.000	0.000
CZK	−0.193	0.000	0.008
HUF	−0.108	0.014	0.003
NOK	−0.013	0.000	0.001
KRW	0.027	0.000	0.000
MXN	0.066	0.000	0.008
PLN	0.112	0.000	0.003
SEK	0.215	0.000	0.000
IDR	0.434	0.000	0.000
TRY	0.513	0.000	0.000
NZD	0.838	0.000	0.000
DKK	0.994	0.000	0.000
EUR	0.998	0.000	0.000
RUB	3.985	0.000	0.000

For period: December 2001 to February 2016.

Table 6: Stationarity test

Exchange Rate	Level (p-values)	Difference (p-values)
GBP	0.216	0.01
ILS	0.222	0.01
SEK	0.270	0.01
CHF	0.319	0.01
HUF	0.351	0.01
NZD	0.401	0.01
KRW	0.407	0.01
MXN	0.411	0.01
EUR	0.436	0.01
DKK	0.439	0.01
PLN	0.573	0.01
IDR	0.615	0.01
INR	0.642	0.01
PHP	0.650	0.01
ZAR	0.696	0.01
AUD	0.749	0.01
TWD	0.753	0.01
CZK	0.818	0.01
NOK	0.863	0.01
CLP	0.897	0.01
JPY	0.938	0.01
CAD	0.945	0.01
TRY	0.950	0.01
BRL	0.968	0.01
SGD	0.974	0.01
COP	0.980	0.01
THB	0.981	0.01
PEN	0.990	0.01
RUB	0.990	0.01
MYR	0.990	0.01

B Appendix

B.1 Algorithm for the minimization problem of the PML

When Γ and Ω are fixed, the minimization problem in (3) with $\Pi = \alpha\beta^\top$ is equivalent to:

$$\left(\hat{\alpha}, \hat{\beta}\right)|\Gamma, \Omega = \arg\min_{\alpha,\beta} \frac{1}{T} \operatorname{tr}\left((\Delta Y - \Delta Y_L \Gamma - Y\beta\alpha^\top)\Omega(\Delta Y - \Delta Y_L \Gamma - Y\beta\alpha^\top)^\top\right) + \lambda_1 P_1(\beta) \tag{23}$$

which boils down to a penalized reduced rank regression (Chen and Huang, 2012). We first estimate α conditional on β, next we estimate β conditional on α.

For fixed β, the minimization problem in (5) reduces to:

$$\hat{\alpha}|\Gamma, \Omega, \beta = \arg\min_{\alpha} \frac{1}{T} \operatorname{tr}\left((\Delta Y - \Delta Y_L \Gamma - Y\beta\alpha^\top)\Omega(\Delta Y - \Delta Y_L \Gamma - Y\beta\alpha^\top)^\top\right) \tag{24}$$

s.t. $\alpha^\top\Omega\alpha = I_r$, which is a weighted procrustes problem (Lissitz *et al.*, 1976). This weighted procrustes problem for α can be seen as an unweighted procrustes problem for $\alpha^* = \Omega^{-1/2}\alpha$. The solution is:

$$\hat{\alpha} = \Omega^{-1/2} V U^\top \tag{25}$$

where U and V are obtained from the singular value decomposition of:

$$\hat{\beta} Y^\top (\Delta Y - \Delta Y_L \Gamma)\Omega^{1/2} = UDV^\top \tag{26}$$

where Chen and Huang (2012) ony consider the case $\Omega = I$, and use a procrustes problem to solve for α. A weighted procrustes problem takes the covariance structure into account.

For fixed α, the minimization problem in (5) reduces to:

$$\hat{\beta}|\Gamma, \Omega, \alpha = \arg\min_{\beta} \frac{1}{T} \operatorname{tr}\left((\Delta Y - \Delta Y_L \Gamma - Y\beta\alpha^\top)\Omega(\Delta Y - \Delta Y_L \Gamma - Y\beta\alpha^\top)^\top\right) + \lambda_1 P_1(\beta) \tag{27}$$

Since $\alpha^{*\top}\alpha^* = I_r$, there exists a matrix $\alpha^{*\perp}$ with orthonormal columns such that $(\alpha^*, \alpha^{*\perp})$ is an orthogonal matrix. Then, with $\tilde{Y} = \Delta Y - \Delta Y_L \Gamma$, we obtain:

$$\begin{aligned}
\operatorname{tr}\left((\tilde{Y} - Y\beta\alpha^\top)\Omega(\tilde{Y} - Y\beta\alpha^\top)^\top\right) &= \left\|\left(\tilde{Y} - Y\beta\alpha^\top\right)\Omega^{1/2}\right\|^2 \\
&= \left\|\left(\tilde{Y}\Omega^{1/2} - Y\beta\alpha^{*\top}\right)\right\|^2 \\
&= \left\|\left(\tilde{Y}\Omega^{1/2} - Y\beta\alpha^{*\top}\right)\left(\alpha^*, \alpha^{*\perp}\right)\right\|^2 \\
&= \left\|\tilde{Y}\Omega^{1/2}\alpha^* - Y\beta\right\|^2 + \left\|\tilde{Y}\Omega^{1/2}\alpha^{*\perp}\right\|^2
\end{aligned}$$

where $\|\cdot\|$ denotes the Frobenius norm for a matrix. Since the second term on the left-hand-side does not involve β, the minimization problem reduces to:

$$\hat{\beta}|\Gamma,\Omega,\alpha = \arg\min_{\beta} \frac{1}{T}\operatorname{tr}\left((\tilde{Y}\Omega^{1/2}\alpha^* - Y\beta)(\tilde{Y}\Omega^{1/2}\alpha^* - Y\beta)^\top\right) + \lambda_1 P_1(\beta) \tag{28}$$

which is a penalized multivariate least squares regression of $\tilde{Y}\Omega^{1/2}\alpha^*$ on Y.

When solving for Γ conditional on Π, Ω, the minimization problem in (3) is a penalized multivariate regression of $(\Delta Y - Y\Pi^\top)$ on ΔY_L (Rothman et al., 2010). When solving for Ω conditional on Γ, Π, the minimization problem in (3) is a penalized covariance estimation (Friedman *et al.*, 2008).

B.2 Bayesian reduced rank

A proper prior given the above normalization is:

$$|\Sigma|^{-(N+v_0+1)} \exp\left[-\frac{1}{2}\operatorname{tr}\left(S_0\Sigma^{-1}\right)\right] \exp\left[-\frac{\tau^2}{2}\left(\operatorname{tr}\left(\Phi^{*'}\Phi^*\right) + \operatorname{tr}\left(\Psi'\Psi\right)\right)\right] \tag{29}$$

namely a product of an independent Wishart distribution for Σ with v_0 degrees of freedom and matrix parameter S_0, and independent $\mathcal{N}(0,\tau^{-2})$ shrinkage priors for each element of the coefficient matrices Φ^* and Ψ. The conditional posterior distribution of Σ is

$$\Sigma\,|(\,\Phi^*,\Psi,X,Y) \sim IW[T+v, S_0 + (Y-XB)^\top(Y-XB)] \tag{30}$$

The conditional posterior distributions of the coefficients Φ^*, Ψ, are multivariate normals. In particular, the conditional posterior distribution of Φ^* is:

$$\operatorname{vec}(\Phi^*)|(\Psi,\Sigma,X,Y) \sim \mathcal{N}\left(\Pi_\Phi \cdot \operatorname{vec}(\hat{\Phi}^*), \Pi_\Phi\right) \tag{31}$$

where

$$\begin{aligned}
\hat{\Phi}^* &= (\Psi'X'X\Psi)^{-1}\Psi'X'Y_1\Sigma^{12}(\Sigma^{22})^{-1} - \Sigma^{12}(\Sigma^{22})^{-1} + \\
&\quad (\Psi'X'X\Psi)^{-1}\Psi'X' \\
\Pi_\Phi &= [(\Sigma^{22})^{-1} \otimes (\Psi'X'X\Psi)^{-1} + \tau^2 I_{r(N-r)}]^{-1}
\end{aligned}$$

where $Y = [Y_1 Y_2]$ is a partitioning of Y into its first r and last $N-r$ columns and where Σ^{ij} denotes the partitioning of Σ^{-1} into its first r and last $N-r$ rows and columns.

The conditional posterior distribution of Ψ is:

$$\operatorname{vec}(\Psi)|(\Phi,\Sigma,X,Y) \sim \mathcal{N}\left(\Pi_\Psi \cdot \operatorname{vec}(\hat{\Psi}), \Pi_\Psi\right) \tag{32}$$

where:

$$\begin{aligned}
\hat{\Psi} &= \hat{B}\left[\Phi^+ + \Phi^0\tilde{\Sigma}^{21}(\tilde{\Sigma}^{11})^{-1}\right] \\
\Pi_\Psi &= \left[\tilde{\Sigma}^{11} \otimes X'X + \tau^2 I_{Mr}\right]^{-1}
\end{aligned}$$

where $\hat{B}$ is the OLS estimator, Φ^+ is the generalized inverse of Φ, Φ^0 is column-wise orthogonal to Φ^+, and $\tilde{\Sigma}^{ij}$ denotes the partitioning of $\tilde{\Sigma}^{-1} = \left([\Phi^+\Phi^0]'\Sigma[\Phi^+\Phi^0]\right)^{-1}$ into its first r and last $N-r$ rows and columns [19].

B.3 Cointegration rank

To determine the cointegration rank we followed Bunea *et al.* (2011) and Wilms and Croux (2016). The cointegration rank r was choosen after an iterative procedure based on the rank selection criterion (RSC). The starting point is an initial value of the cointegration rank $r_{rank} = q$, for which, $\hat{\Gamma}$ is obtained. Bunea *et al.* (2011) argued that $\hat{r}$ is given by the number of eigenvalues of the matrix $\Delta\tilde{Y}^\top P\Delta\tilde{Y}$ that exceeds the following threshold μ:

$$\tilde{r} = max\{r : \lambda_r(\Delta\tilde{Y}^\top P\Delta\tilde{Y}) \geq \mu\} \tag{33}$$

where $\Delta\tilde{Y} = \Delta Y - \Delta Y_L\hat{\Gamma}$ and $P = Y(Y^\top Y)^- Y^\top$ the projection matrix onto the column space of Y. Taking $l = rank|Y|$ and assuming that $l < T$, then the threshold μ will be equal to $\mu = 2S^2(q+l)$ (Bunea *et al.*, 2011), where

$$S^2 = \frac{\left\|\Delta\tilde{Y} - P\Delta\tilde{Y}\right\|^2}{Tq - lq} \tag{34}$$

This procedure was prevented from repeating when the cointegration rank in two iterations was unchanged.

C Appendix

C.1 Estimated parameters

Table 7: TC-MAFE for window size 60 and forecast horizon h

	TC-MAFE1	TC-MAFE3	TC-MAFE6	TC-MAFE12
Coint PML	0.96	0.90	0.93	0.91
Coint DFM	0.81	0.83	0.80	0.83
VAR PML	0.90	0.79	0.82	0.82
BVAR	0.88	0.94	0.91	0.97
VAR BRR	0.88	0.94	0.89	0.96
VAR DFM	0.83	0.83	0.84	0.86

[19] See Geweke (1996) and Carriero *et al.* (2011) for more details on the computations.

Table 8: TC-MAFE for window size 72 and forecast hozizon h

	TC-MAFE1	TC-MAFE3	TC-MAFEC6	TC-MAFE12
Coint PML	1.00	1.02	0.97	0.80
Coint DFM	0.82	0.82	0.83	0.77
VAR PML	0.87	0.82	0.84	0.77
BVAR	0.87	1.01	0.94	0.89
VAR BRR	0.87	1.01	0.92	0.89
VAR DFM	0.88	0.95	0.92	0.83

Table 9: TC-MAFE for window size 144 and forecast hozizon h

	TC-MAFE1	TC-MAFE3	TC-MAFE6	TC-MAFE12
Coint PML	0.70	0.66	0.82	0.75
Coint DFM	0.68	0.68	0.74	0.75
VAR PML	0.68	0.68	0.74	0.75
BVAR	0.67	0.93	0.79	0.98
VAR BRR	0.69	0.91	0.83	0.98
VAR DFM	0.69	0.76	0.81	0.77

Table 10: Forecast Combination

Method	AUD	CAD	BRL	RUB	EUR
Simple	0.0813	0.0468	0.0671	0.0841	0.0317
OLS	0.0311	0.0213	0.0342	0.0354	0.0227
CLS	0.0309	0.0222	0.0339	0.0358	0.0227
LAD	0.0302	0.0205	0.0331	0.0341	0.0219
Variance based	0.0438	0.0313	0.0485	0.0530	0.0346
Method	GBP	CHF	JPY	TRY	MXN
Simple	0.0469	0.0531	0.0317	0.0638	0.0570
OLS	0.0202	0.0265	0.0227	0.0265	0.0242
CLS	0.0204	0.0264	0.0227	0.0271	0.0242
LAD	0.0198	0.0253	0.0219	0.0261	0.0236
Variance based	0.0281	0.0365	0.0292	0.0382	0.0363

Table 11: Mean Absolute Forecast Errors For LAD

Currency	Value
AUD	0.5825
CAD	0.6339
BRL	0.6078
RUB	0.5893
EUR	0.6254
GBP	0.6698
CHF	0.6304
JPY	0.6836
TRY	0.6085
MXN	0.5854

C.2 Graphs

C.3 Commodity currencies

Figure 2: Developed markets

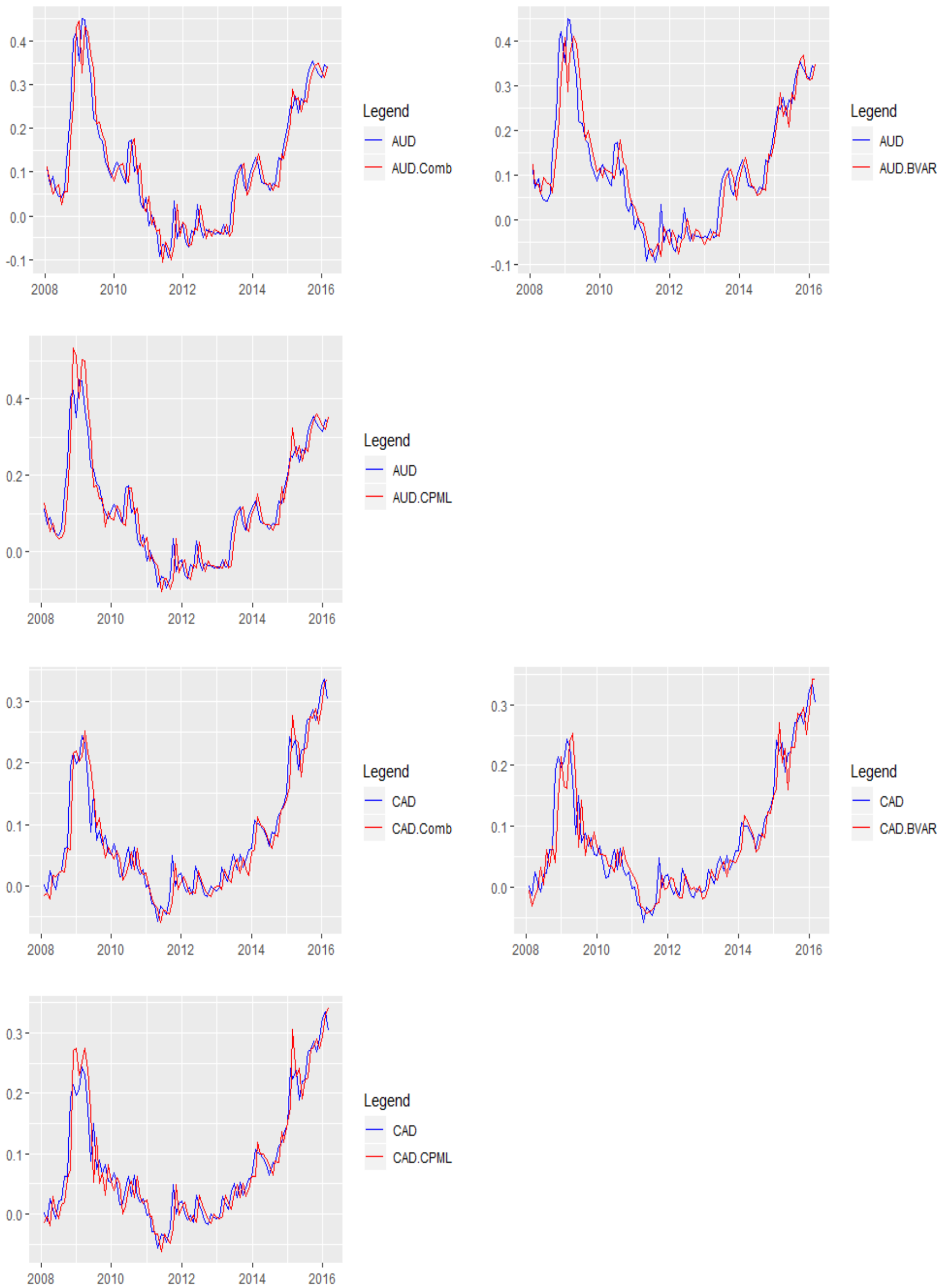
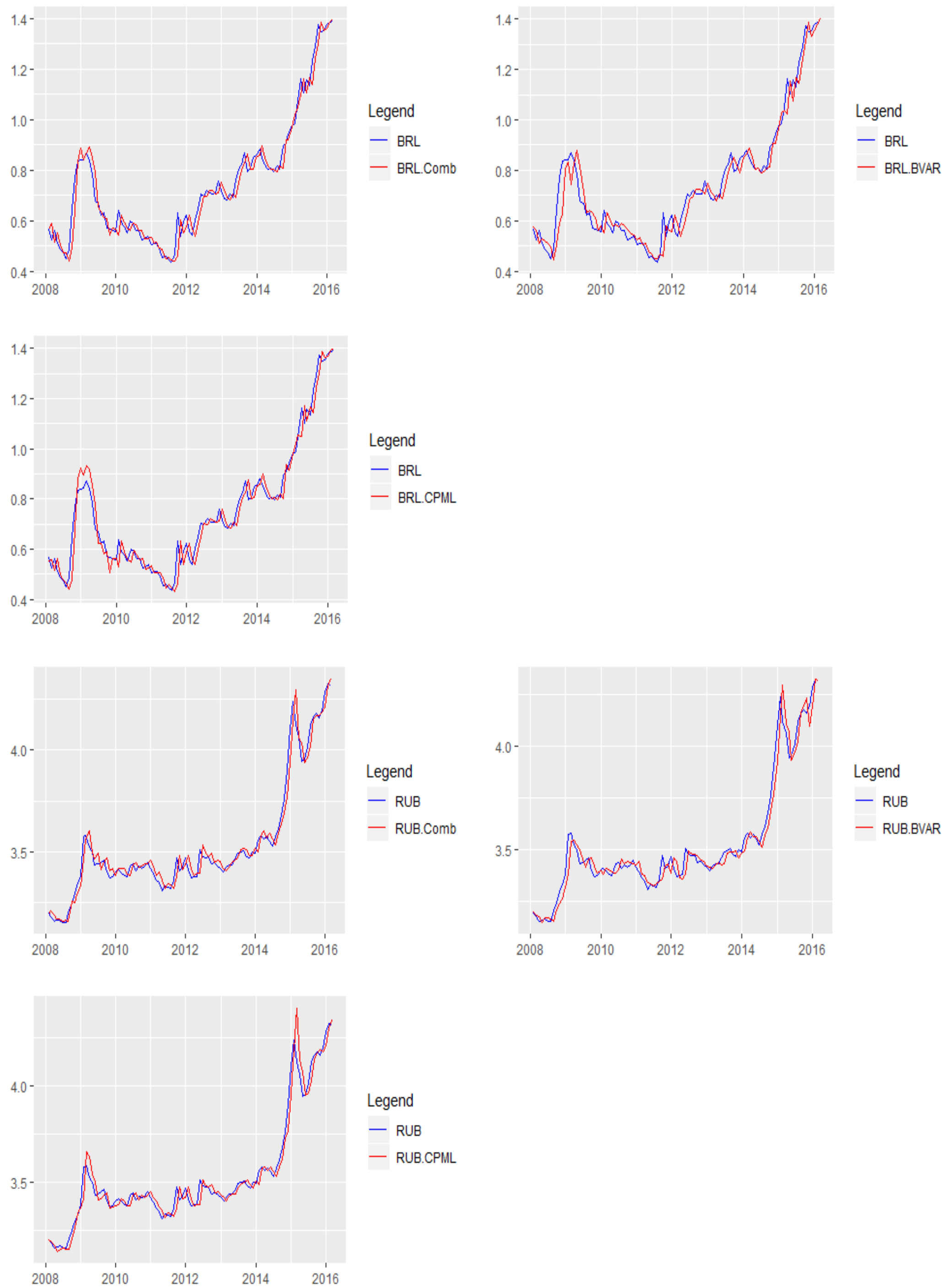


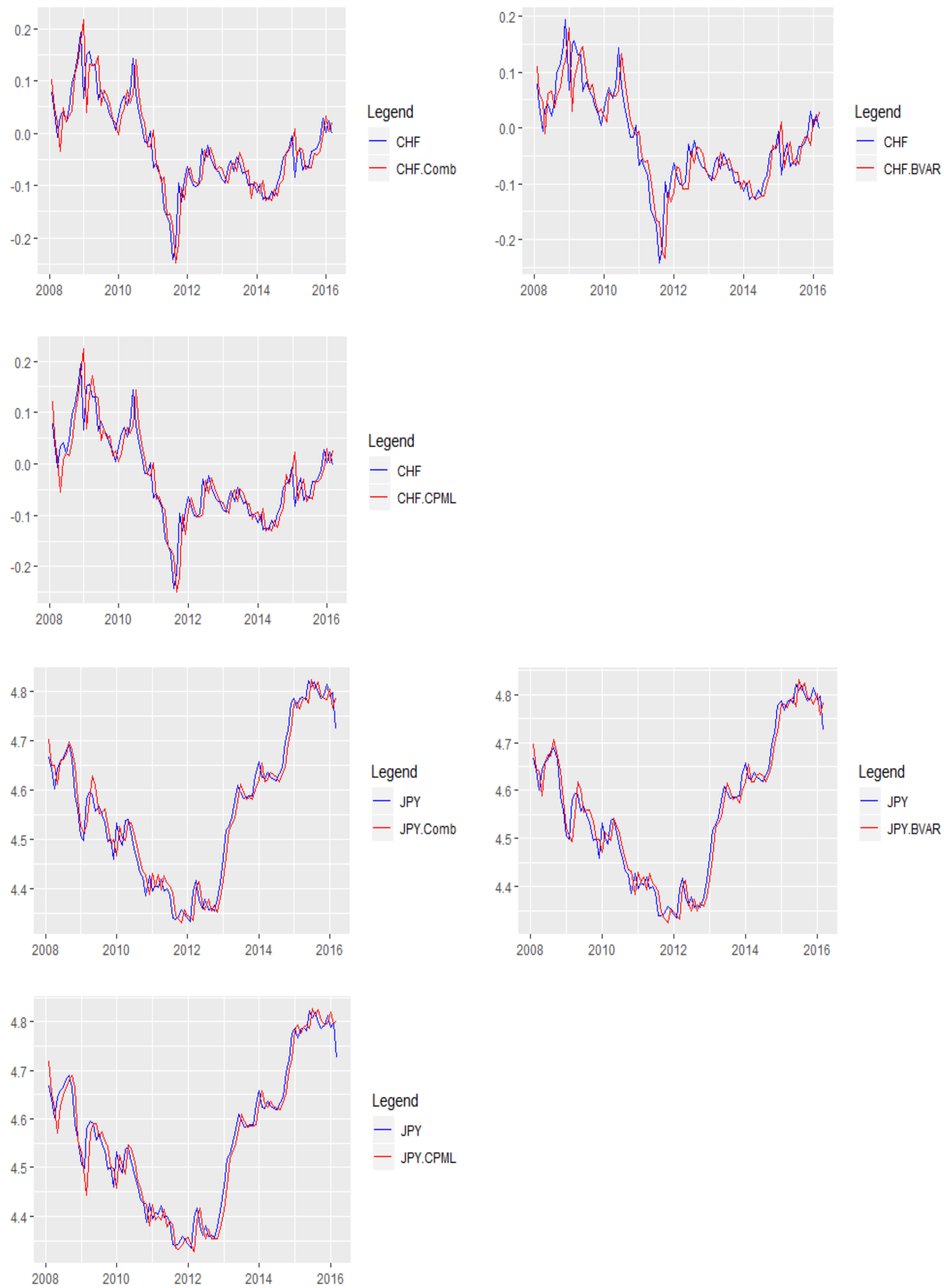
Figure 3: Emerging markets



C.4 Developed currencies

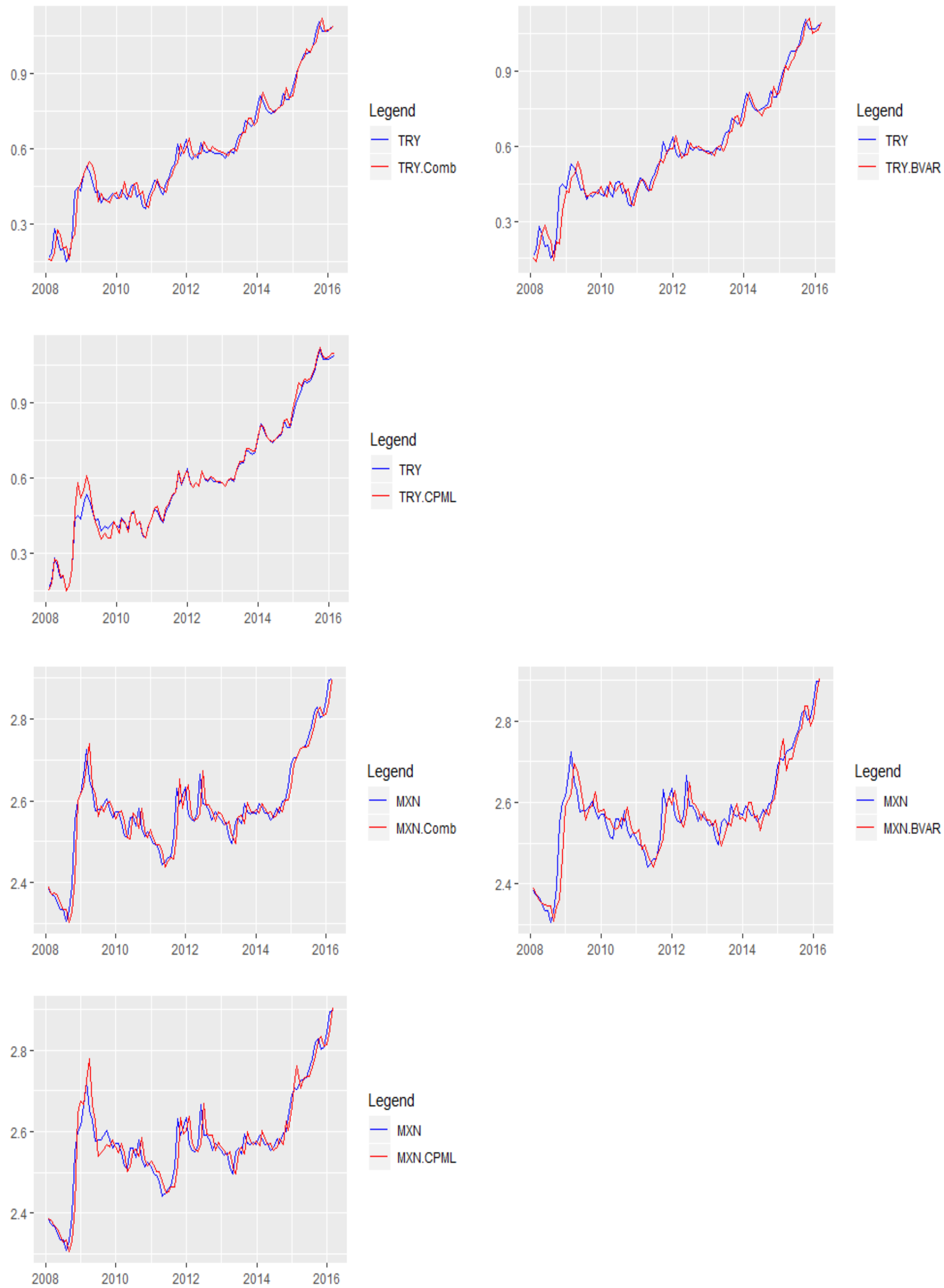
Figure 4: EUR and GBP

Figure 5: CHF and JPY



C.5 Emerging currencies

Figure 6: TRY and MXN



Chief Editors

Pascal BLANQUÉ
Chief Investment Officer

Philippe ITHURBIDE
Global Head of Research

Cross asset
INVESTMENT STRATEGY
September 2019 | Working paper

DISCLAIMER

Document issued by Amundi, "société par actions simplifiée"- SAS with a capital of €1,086,262,605 - Portfolio manager regulated by the AMF under number GP04000036 – Head office: 90 boulevard Pasteur – 75015 Paris – France – 437 574 452 RCS Paris - www.amundi.com

Photo credit: iStock by Getty Images - monsitj/Sam Edwards

Find out more about
Amundi research team

research-center.amundi.com

Amundi
ASSET MANAGEMENT