



Forecasting performance of Bayesian VEC-MSF models for financial data in the presence of long-run relationships

Anna Pajor^{1,2}  · Justyna Wróblewska³

Received: 28 July 2021 / Revised: 29 November 2021 / Accepted: 6 January 2022 /
Published online: 9 February 2022
© The Author(s) 2022

Abstract

The paper is focused on comparing the forecasting performance of two relatively new types of Vector Error Correction - Multiplicative Stochastic Factor (VEC-MSF) specifications: VEC-MSF with constant conditional correlations, and VEC-MSF-SBEKK with time-varying conditional correlations. For the sake of comparison, random walks, vector autoregressions (VAR) with constant conditional covariance matrix, and VAR-SBEKK models are also considered. Based on daily quotations on three exchange rates: PLN/EUR, PLN/USD, and EUR/USD, where the cointegrating vector may be assumed to be known *a priori*, we show that in econometric models it can be more important to allow for cointegration relationships than for time-varying conditional covariance matrix.

Keywords Multivariate time series · Cointegration · Stochastic volatility · Predictive Bayes factor · Exchange rate

JEL Classification C11 · C53

1 Introduction

Widely-used models for financial time series are based on conditional heteroscedasticity processes. The most popular ones are the Generalized Autoregressive Conditional Heteroskedastic (GARCH) and the Stochastic Volatility (SV)

✉ Anna Pajor
pajora@uek.krakow.pl

¹ Department of Mathematics, Cracow University of Economics, Rakowicka 27, 31-510 Kraków, Poland

² Department of Financial Mathematics, Jagiellonian University in Kraków, prof. Stanisława Łojasiewicza 6, 30-348 Kraków, Poland

³ Department of Econometrics and Operations Research, Cracow University of Economics, Rakowicka 27, 31-510 Kraków, Poland

processes. The latter are less often used, though (for a review see, e.g., Silvennoinen & Teräsvirta, 2009; Asai et al., 2006; Shephard & Andersen, 2009; Chib et al., 2009; Eratalay, 2016). However, some research has shown that apart from the fact that the variability of financial time series (measured, for example, by the conditional covariance matrix) varies over time, on financial markets there can also exist long-term relationships. Therefore, it appears essential to construct such models in which the possible presence of long-run relationships and time-variable volatility are simultaneously taken into account. As the vector autoregression (VAR) models underlie the cointegration analysis, the efforts have been focused on suitably tailoring the vector error correction (VEC) representation of VAR so as to accommodate for time-varying conditional covariances, see, e.g., Seo (2007), Herwartz and Lütkepohl (2011), Koop et al. (2011), Osiewalski and Osiewalski (2013), Osiewalski and Osiewalski (2016), Cavaliere et al. (2015), Pajor and Wróblewska (2017), Cavaliere et al. (2018).

Although there exists a growing number of publications discussing the forecast accuracy of Bayesian VAR models with time-varying parameters or with the time-varying covariance structure (see, e.g., Clark, 2011; D'Agostino et al., 2013; Rossi & Skhposyan, 2014; Clark & Ravazzolo, 2015; Berg, 2017; Abbate & Marcellino, 2018; Chan & Eisenstat, 2018; Vardar et al., 2018; Huber et al., 2020; Kastner & Huber, 2021), one can find rather few publications in which different vector error correction models are compared in respect to their forecasting potential (see, e.g., Anderson et al., 2002; Swanson, 2002; Kuo, 2016; Huber & Zörner, 2019). The main conclusion of the papers mentioned above is that VAR models with time-varying parameters and with time-varying covariance structure are often found to have better forecasting performance than their constant coefficient variants. On the other hand, Kuo (2016) showed that VEC models outperform VAR models in forecasting stock prices on Taiwan markets.

Our research is focused on Bayesian Vector Error Correction - Multiplicative Stochastic Factor (VEC-MSF) models, proposed by Pajor and Wróblewska (2017). These models integrate the VEC representation of a VAR structure with stochastic volatility. In consequence, the VEC-MSF models enable us to capture long-run relationships among processes. Also, they make it possible to formally examine the presence of time-variation in the conditional covariances. In the VEC-MSF models proposed by Pajor and Wróblewska (2017), conditional heteroskedasticity is taken into account by the Multiplicative Stochastic Factor (MSF) process, or, alternatively, by one of its generalizations, namely, by the hybrid Multiplicative Stochastic Factor - Scalar BEKK (MSF-SBEKK) specification. It is worth mentioning that, in the VEC-GMSF-SBEKK models proposed by Osiewalski and Osiewalski (2016), the authors consider only two cases of reduced rank long-run multiplier matrix, Π . In their paper only three partial and one global relationships are analyzed. In our paper, by contrast, all possible cases of the long-run multiplier matrix Π are considered.

The main purpose of this paper is to discuss properties of Bayesian VEC-MSF models in the context of modeling financial time series. Another purpose is to compare the forecasting performance of three types of model specifications: the VEC-MSF model with constant conditional correlations, the VEC-MSF-SBEKK and

VEC-SBEKK models with varying conditional correlations, and the VEC model with constant conditional covariance matrix.

Based on daily quotations on three exchange rates: PLN/EUR, PLN/USD and EUR/USD, the predictive capacity of the models under consideration is compared. By modeling the exchange rates (where, as arbitrage opportunities are lacking, the cointegrating vector may be assumed to be known *a priori*, see Osiewalski and Pipień, 2016), it will be possible to show how important for prediction it is to incorporate a long-run relationship in financial data. It is important to stress that our attention is focused on pure time-series specifications for daily data, without applying any extra variables suggested by economic or financial theory. The fundamentals of exchange rate modeling are explained in Kębłowski et al. (2020), where the classical VEC model with constant conditional covariance matrix is used for monthly data spanning a period from January 2000 to June 2017.

The main criterion used in this study for drawing this comparison is the predictive Bayes factor. The Probability Integral Transform (PIT) of Rosenblatt (1952) is also applied. As it is pointed out by Geweke and Amisano (2010, p. 229) “the two approaches can be complementary, each identifying strengths and weaknesses in the models”. A popular measure of forecasting performance is the so-called continuous ranked probability score (CRPS). The definition of CRPS and its generalization to energy score can be found in Gneiting and Raftery (2007). As the analysis conducted here is mainly Bayesian, the predictive Bayes factor will be used as the main criterion applied for evaluating the forecast performance of models under consideration.

The paper is organized as follows. In Sect. 2 Bayesian VEC-MSF models are presented. Sections 3 and 4 are devoted to the predictive Bayes factor and to the PIT in the context of predictive performance of models. Section 5 contains empirical results, and Sect. 6 concludes the paper.

2 Bayesian VEC models with stochastic volatility

Let us start with a linear n -variate and k th-order vector autoregressive (VAR) process:

$$\mathbf{x}_t = \mathbf{A}_1 \mathbf{x}_{t-1} + \mathbf{A}_2 \mathbf{x}_{t-2} + \dots + \mathbf{A}_k \mathbf{x}_{t-k} + \mathbf{\Phi} \mathbf{D}_t + \varepsilon_t, \quad t = 1, 2, \dots, T + h, \quad (1)$$

where \mathbf{x}_t is an $n \times 1$ random vector, $\{\varepsilon_t\}$ is a vector white noise process with the covariance matrix $\mathbf{\Sigma}$ (i.e. $\{\varepsilon_t\} \sim WN(0, \mathbf{\Sigma})$), \mathbf{A}_i is an $n \times n$ matrix of real coefficients ($i = 1, 2, \dots, k$), matrix \mathbf{D}_t is comprised of deterministic variables, $\mathbf{\Phi}$ is a parameter matrix, T is the number of observations, h is the forecasting horizon, and $\mathbf{x}_{1-k}, \mathbf{x}_{2-k}, \dots, \mathbf{x}_0$ are the initial conditions. The $\{\mathbf{x}_t\}$ process takes on the following VEC representation:

$$\Delta \mathbf{x}_t = \tilde{\mathbf{\Pi}} \mathbf{x}_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta \mathbf{x}_{t-i} + \mathbf{\Phi} \mathbf{D}_t + \varepsilon_t, \quad (2)$$

where $\Delta \mathbf{x}_t = \mathbf{x}_t - \mathbf{x}_{t-1}$, $\tilde{\Pi} = \sum_{i=1}^k \mathbf{A}_i - \mathbf{I}$, $\Gamma_i = -\sum_{j=i+1}^k \mathbf{A}_j$. Moreover, $\tilde{\Pi} = \alpha \tilde{\beta}'$, with α and $\tilde{\beta}'$ being some $(n \times r)$ -sized matrices, and $r < n$ is the number of cointegration relationships (if they exist).

The definition of $\{\mathbf{x}_t\}$ provided in (2), has been generalized in Pajor and Wróblewska (2017) by introducing random-variability into elements of Σ :

$$\Delta \mathbf{x}_t = \alpha \tilde{\beta}' \mathbf{x}_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta \mathbf{x}_{t-i} + \Phi \mathbf{D}_t + \varepsilon_t, \quad t = 1, 2, \dots, T + h, \tag{3}$$

$$\varepsilon_t | \psi_{t-1}, \mathbf{D}_t, q_t, \theta \sim N(\mathbf{0}, \Sigma_t), \tag{4}$$

where ψ_{t-1} denotes the past of the process $\{\mathbf{x}_t\}$ up to time $t - 1$, q_t is a latent variable, θ is a vector of parameters, and $\Sigma_t = \Sigma(q_t, \psi_{t-1}, \mathbf{D}_t, \theta)$.

In order to investigate the influence of assumptions (pertaining to the conditional covariance matrix) on the predictive abilities of the models we consider four structures for matrix Σ_t : constant, Multiplicative Stochastic Factor (MSF), SBEEKK, and hybrid MSF-SBEEKK [type I; see, e.g., Osiewalski (2009); Osiewalski and Pajor (2009); Pajor and Wróblewska (2017)].

The Multiplicative Stochastic Factor structure for matrix Σ_t is as follows:

$$\Sigma_t = q_t \Sigma, \tag{5}$$

where $\ln q_t = \varphi \ln q_{t-1} + \sigma_q \eta_t$, $|\varphi| < 1$, $\{\eta_t\} \sim iiN(0, 1)$, and $\varepsilon_t \perp \eta_s$, for $t, s \in \{1, 2, \dots, T + h\}$.

Although the VEC-MSF process features non-zero time-variable conditional covariances, conditional correlations remain constant in time. Such a result is attributable to the fact that the very same q_t factor drives the dynamics of each element of Σ_t . The process has been employed in, e.g., Osiewalski and Pajor (2009) and Pajor and Osiewalski (2012).

The type I hybrid MSF-SBEEKK structure for matrix Σ_t proposed by Osiewalski (2009) and Osiewalski and Pajor (2009) is the following:

$$\Sigma_t = q_t \tilde{\Sigma}_t, \tag{6}$$

$$\tilde{\Sigma}_t = (1 - a - b)\Sigma + b(\varepsilon_{t-1} \varepsilon'_{t-1}) + a\tilde{\Sigma}_{t-1}, \tag{7}$$

with $\ln q_t = \varphi \ln q_{t-1} + \sigma_q \eta_t$, $|\varphi| < 1$, $\{\eta_t\} \sim iiN(0, 1)$, and $\varepsilon_t \perp \eta_s$, for $t, s \in \{1, 2, \dots, T + h\}$, $a, b \in [0, 1]$, $a + b < 1$.

As regards the initial conditions for $\tilde{\Sigma}_t$, we assume $\varepsilon_0 = \mathbf{0}$, $\tilde{\Sigma}_0 = s_{0,\Sigma} \mathbf{I}_n$, where $s_{0,\Sigma} > 0$, and \mathbf{I}_n denotes identity matrix of size n . The presence of the scalar BEKK(1,1) structure in the conditional covariance matrix allows to model time-varying conditional correlations without introducing more latent processes. The hybrid model defined by (3, 4) and (6, 7) nests two simple basic structures. For $b = 0$ and $a = 0$ we obtain the VEC-MSF model. In the limiting case when $\sigma_q \rightarrow 0$ and $\varphi = 0$, the VEC-MSF-SBEEKK model becomes the VEC-SBEEKK one.

Equation (3) can be decomposed and expressed as follows:

$$\begin{aligned} \Delta \mathbf{x}_t &= \alpha[\tilde{\beta}', \Phi_1'] \left[\begin{matrix} \mathbf{x}_{t-1} \\ \mathbf{D}_t^{(1)} \end{matrix} \right] + \sum_{i=1}^{k-1} \Gamma_i \Delta \mathbf{x}_{t-i} + \Phi_2 \mathbf{D}_t^{(2)} + \varepsilon_t \\ &= \alpha \beta' \mathbf{z}_{1,t} + \Gamma' \mathbf{z}_{2,t} + \Gamma'_s \mathbf{z}_{3,t} + \varepsilon_t, \end{aligned} \tag{8}$$

where $\beta' = [\tilde{\beta}', \Phi_1']$, $\mathbf{z}_{1,t} = [\mathbf{x}'_{t-1}, \mathbf{D}'_t^{(1)}]'$, $\mathbf{z}_{2,t} = (\Delta \mathbf{x}'_{t-1}, \Delta \mathbf{x}'_{t-2}, \dots, \Delta \mathbf{x}'_{t-k+1})'$, $\mathbf{z}_{3,t} = \mathbf{D}_t^{(2)}$, $\Gamma = [\Gamma_1, \Gamma_2, \dots, \Gamma_{k-1}]'$, $\Gamma_s = \Phi_2'$, and $\alpha \Phi_1' \mathbf{D}_t^{(1)} + \Phi_2 \mathbf{D}_t^{(2)} = \Phi \mathbf{D}_t$.

The conditional distribution of \mathbf{x}_t (given the past of the process, ψ_{t-1} , the deterministic variables, \mathbf{D}_t , the parameters and the latent variable, q_t) is n -variate Normal with the mean $\mu_t = \mathbf{x}_{t-1} + \Pi \mathbf{z}_{1,t} + \Gamma' \mathbf{z}_{2,t} + \Gamma'_s \mathbf{z}_{3,t}$ and the covariance matrix Σ_t :

$$p(\mathbf{x}_t | \psi_{t-1}, \mathbf{D}_t, \alpha, \beta, \Gamma, \Gamma_s, q_t, \Sigma, \theta_\Sigma, q_{0,\Sigma}) = f_{N,n}(\mathbf{x}_t | \mu_t, \Sigma_t), \tag{9}$$

where $\Pi = \alpha \beta'$, θ_Σ and $q_{0,\Sigma}$ are the vectors of the stochastic volatility parameters: in the VEC-MSF model $\theta_\Sigma = (\varphi, \sigma_q^2)'$ and $q_{0,\Sigma} = \ln q_0$, in the VEC-MSF-SBEKK model we have $\theta_\Sigma = (\varphi, \sigma_q^2, a, b)'$ and $\mathbf{q}_{0,\Sigma} = (\ln q_0, s_{0,\Sigma})'$, whereas in the VEC-SBEKK specification $\theta_\Sigma = (a, b)'$ and $\mathbf{q}_{0,\Sigma} = s_{0,\Sigma}$.

2.1 Predictive distribution

Let us adopt the following symbols:

- $\theta = (\text{vec } \alpha', \text{vec } \beta', \text{vec } \Gamma', \text{vec } \Gamma'_s, \text{vech } \Sigma', \theta'_\Sigma, q_{0,\Sigma})'$ – the parameter vector,
- \mathbf{x}_t , $t = 1, 2, \dots, T$ – the observable random vectors,
- $\mathbf{x}_T = [\mathbf{x}_1 \mathbf{x}_2 \dots \mathbf{x}_T]$ – the matrix of observables,
- $\mathbf{x}_f = [\mathbf{x}_{T+1} \mathbf{x}_{T+2} \dots \mathbf{x}_{T+h}]$ – the matrix of future observables,
- $\mathbf{q}_T = (q_1, q_2, \dots, q_T)'$ – the vector of latent variables for the first T observations,
- $\mathbf{q}_f = (q_{T+1}, q_{T+2}, \dots, q_{T+h})'$ – the vector of future latent variables,
- $\mathbf{D} = [\mathbf{D}_1 \mathbf{D}_2 \dots \mathbf{D}_{T+h}]$ – the matrix of deterministic variables.

Inference on all unknown and unobserved quantities (parameters, latent variables and future observables) can be based on the joint posterior – predictive density function. The joint density function of the observed data, h forecasted values of the data and forecasted latent variables (given the parameters and latent variables up to time T) is as follows:

$$\begin{aligned} p(\mathbf{x}_T, \mathbf{x}_f, \mathbf{q}_f | \mathbf{D}, \theta, \mathbf{q}_T) &= \prod_{t=1}^{T+h} f_{N,n}(\mathbf{x}_t | \mu_t, \Sigma_t) \prod_{t=1}^{T+h} q_t^{-1} f_{N,1}(\ln q_t | \varphi \ln q_{t-1}, \sigma_q^2). \end{aligned} \tag{10}$$

This density depends on some initial values, which are not shown in our notation. The out-of-sample predictive density of h future values of observations and of latent variables is obtained through averaging the sampling predictive densities over the parameters and latent variables space, with the use of the posterior density as a weight function:

$$p(\mathbf{x}_f, \mathbf{q}_f | \mathbf{x}_T^o, \mathbf{D}) = \int p(\mathbf{x}_f, \mathbf{q}_f | \mathbf{x}_T^o, \mathbf{D}, \mathbf{q}_T, \theta) p(\mathbf{q}_T, \theta | \mathbf{x}_T^o, \mathbf{D}) d\mathbf{q}_T d\theta, \tag{11}$$

where

$$\begin{aligned} p(\mathbf{x}_f, \mathbf{q}_f | \mathbf{x}_T^o, \mathbf{D}, \mathbf{q}, \boldsymbol{\theta}) \\ = \prod_{t=T+1}^{T+h} f_{N,n}(\mathbf{x}_t | \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t) \prod_{t=T+1}^{T+h} q_t^{-1} f_{N,1}(\ln q_t | \varphi \ln q_{t-1}, \sigma_q^2), \end{aligned} \quad (12)$$

$\mathbf{x}_T^o = [\mathbf{x}_1^o \dots \mathbf{x}_T^o]$ denotes the *ex post*, observed, value of \mathbf{x}_T .

The main advantage of the Bayesian approach is that it provides a probability distribution for future observables (or any function of them). If we are interested only in prediction of future observables, we can use the following Bayesian predictive distribution:

$$p(\mathbf{x}_f | \mathbf{x}_T^o, \mathbf{D}) = \int p(\mathbf{x}_f, \mathbf{q}_f | \mathbf{x}_T^o, \mathbf{D}, \mathbf{q}_T, \boldsymbol{\theta}) p(\mathbf{q}_T, \boldsymbol{\theta} | \mathbf{x}_T^o, \mathbf{D}) d\mathbf{q}_T d\boldsymbol{\theta} d\mathbf{q}_f, \quad (13)$$

which takes into account uncertainty regarding parameters and latent variables, given the data, sampling model and a prior distribution.

2.2 Prior distribution and sampling scheme

It remains to formulate the prior distribution of the parameter vector $\boldsymbol{\theta}$. We assume independence among certain blocks of parameters, and we use the same prior distributions as in Pajor and Wróblewska (2017):

$$p(\boldsymbol{\theta}) = p(\boldsymbol{\alpha} | \boldsymbol{\beta}) p(\boldsymbol{\beta}) p(\boldsymbol{\Gamma}) p(\boldsymbol{\Gamma}_s) p(\boldsymbol{\Sigma}) p(\boldsymbol{\theta}_\Sigma) p(\mathbf{q}_{0,\Sigma}) I_{[0,1]}(|\boldsymbol{\lambda}|_{\max}), \quad (14)$$

where $I_{[0,1]}(\cdot)$ denotes the indicator function of the interval $[0, 1]$, and $\boldsymbol{\lambda}$ is the vector of eigenvalues of the companion matrix, connected with the VAR form.

For matrix $\boldsymbol{\Pi}$ we use the parameterization which has been proposed by Koop et al. (2010):

$$\boldsymbol{\alpha} \boldsymbol{\beta}' = (\boldsymbol{\alpha} \mathbf{M}_\Pi) (\boldsymbol{\beta} \mathbf{M}_\Pi^{-1})' \equiv \mathbf{A} \mathbf{B}',$$

where \mathbf{M}_Π is an $r \times r$ symmetric positive-definite matrix, and \mathbf{A} and \mathbf{B} are unrestricted matrices. It can be shown that $\boldsymbol{\alpha} = \mathbf{A}(\mathbf{B}'\mathbf{B})^{-\frac{1}{2}}$, $\boldsymbol{\beta} = \mathbf{B}(\mathbf{B}'\mathbf{B})^{-\frac{1}{2}}$. The prior distributions for matrices \mathbf{A} and \mathbf{B} imply the prior distributions for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. They are assumed to be as follows:

- for matrix \mathbf{B} the matrix normal distribution is used: $p(\mathbf{B}|r) = f_{mN}(\mathbf{B} | \mathbf{0}, \mathbf{I}_r, \mathbf{I}_n)$, which leads to the matrix angular central Gaussian (MACG) distribution for $\boldsymbol{\beta}$: $p(\boldsymbol{\beta}) = f_{MACG}(\boldsymbol{\beta} | \mathbf{I}_n)$. By imposing MACG distribution with identity matrix we set uniform prior for $\boldsymbol{\beta}$, (see, e.g., Chikuse, 2002).
- on matrix \mathbf{A} we also impose the matrix normal distribution: $p(\mathbf{A} | \nu, r) = f_{mN}(\mathbf{A} | \mathbf{0}, \nu \mathbf{I}_r, \mathbf{I}_n)$, where parameter ν has the inverse-gamma distribution: $p(\nu) = f_{IG}(\nu | 2, 3)$.

The priors for the remaining parameters are as follows:

- $p(\mathbf{\Gamma}|h) = f_{mN}(\mathbf{\Gamma}|\mathbf{0}, \mathbf{I}_n, h\mathbf{I}_l)$, where $l = n(k - 1)$,
- $p(\mathbf{\Gamma}_s|h_s) = f_{mN}(\mathbf{\Gamma}_s|\mathbf{0}, \mathbf{I}_n, h_s\mathbf{I}_{l_s})$, where l_s is the number of deterministic terms in \mathbf{D}_t , moreover, $p(h) = f_{IG}(h|2, 3)$ and $p(h_s) = f_{IG}(h_s|2, 3)$;
- $p(\varphi) \propto f_{N,1}(\varphi | 0, 100)I_{(-1,1)}(\varphi)$, – the normal density with mean zero and variance 100, truncated by the restriction $|\varphi| < 1$;
- $p(\sigma_q^2) = f_{IG}(\sigma_q^2 | 1, 0.005)$ – the density of the inverted gamma distribution, it is assumed that $E(\sigma_q^{-2}) = 200$, $Var(\sigma_q^{-2}) = 200^2$ (see Jacquier et al., 2004);
- $p(\mathbf{\Sigma}) = f_{IW}(\mathbf{\Sigma} | \mathbf{I}_n, n + 2, n)$ – the inverted Wishart distribution with mean I_n ;
- $p(a, b) \propto I_{(0,1)}(a + b)$ – the uniform distribution over the unit simplex;
- $p(\ln q_0) = f_{N,1}(\ln q_0 | 0, 100)$;
- $p(s_{0,\Sigma}) = f_{Exp}(s_{0,\Sigma} | 1)$ – the exponential distribution with mean 1.

In order to estimate our VAR-MSF models, which contain as many latent variables as the number of observations, we resort to MCMC methods which allow for generating a pseudorandom sample from the posterior distribution. The MCMC methods adapted to our models are presented in Pajor and Wróblewska (2017). Simulation from the predictive distribution is performed by generating pseudorandom values from posterior distribution of parameters and latent variables, and by inserting them into the sampling distributions of future observables and latent variables. In other words, at each step of Gibbs algorithm (for each simulated vector of parameters and latent variables) \mathbf{x}_f and \mathbf{q}_f are drawn from the sampling predictive distributions. The resulting sequence of future observables and latent variables is a simulation from the predictive distribution (expressed in 11). The sampling scheme is the following:

- Step 1. Set $i = 0$ and get starting values for the vector of parameters $\theta^{(i)}$ and for the vector of latent variables $\mathbf{q}_T^{(i)}$;
- Step 2. Generate $\theta_1^{(i+1)} = (a, b, s_{0,\Sigma}, \text{vec}(\mathbf{\Sigma}))^{(i+1)}$ from the full conditional posterior distribution of the parameter, using the sequential Metropolis-Hastings algorithm. The proposal distribution used to simulate $\theta_1^{(i+1)}$ is a truncated Student t distribution with 3 degrees of freedom, centered at previous values of the chain. The covariance matrix of the Student t distribution is determined by initial draws of the algorithm;
- Step 3. Draw $\varphi^{(i+1)}$ from a truncated normal distribution;
- Step 4. Draw $(\sigma_q^2)^{(i+1)}$ from an inverted gamma distribution;
- Step 5. Draw $\ln q_0^{(i+1)}$ from a normal distribution;
- Step 6. Generate $\theta_2^{(i+1)} = (\text{vec}(\mathbf{A})', \text{vec}(\mathbf{B})', \text{vec}(\mathbf{\Gamma})', \text{vec}(\mathbf{\Gamma}_s)', h, h_s, v)^{(i+1)}$ from the full conditional posterior distribution of the vector of parameters, using the sequential Metropolis-Hastings algorithm. The proposal distribution used to simulate $\theta_2^{(i+1)}$ is a truncated normal distribution;
- Step 7. For $t = 1, \dots, T$ generate $q_t^{(i+1)}$ from the full conditional posterior distribution of this latent variable, using the independence Metropolis - Hastings algorithm. The proposal distribution used to simulate q_t is an inverted gamma distribution;
- Step 8. For $j = 1, \dots, h$ draw $q_{T+j}^{(i+1)}$ from the sampling predictive distribution whose density is a product of log-normal densities;

Step 9. For $j = 1, \dots, h$ draw $\mathbf{x}_{T+j}^{(i+1)}$ from the conditional (given $q_{T+j}^{(i+1)}$) sampling predictive distribution whose density is a product of multivariate normal densities;
 Step 10. Set $i = i + 1$ and return to Step 2 until convergence is achieved.

Details on the full conditionals of parameters and the latent variables, as well as proposals distributions can be found in Pajor and Wróblewska (2017).

3 Predictive Bayes factor

Comparing and evaluating the predictive performance of models requires the use of inference methods for predictive accuracy. The Bayesian approach delivers the concept related to the marginal likelihood and predictive likelihood, which are relevant for the context of the comparison of models. Because “there is no distinction between a model’s adequacy and its out of sample prediction record.” (see Geweke, 2005, p. 23), we can start with the model comparison. Let us assume that M_1, \dots, M_m are mutually exclusive (non-nested) and jointly exhaustive models. The posterior probability of model M_i is

$$p(M_i|\mathbf{x}_T^o) = \frac{p(\mathbf{x}_T^o|M_i)p(M_i)}{\sum_{j=1}^m p(\mathbf{x}_T^o|M_j)p(M_j)}, \quad i = 1, 2, \dots, m, \quad (15)$$

where $p(M_i)$ and $p(\mathbf{x}_T^o|M_i)$ denote a prior probability of model M_i and the marginal likelihood, respectively. The marginal likelihood is defined as a value of the marginal probability density function for \mathbf{x}_T in observed data, \mathbf{x}_T^o , and thus it is a real number (see Geweke, 2005). The marginal probability density function for \mathbf{x}_T is

$$p(\mathbf{x}_T|M_i) = \int p(\mathbf{x}_T|\boldsymbol{\theta}_i, \mathbf{q}_{T_i}, M_i)p(\boldsymbol{\theta}_i, \mathbf{q}_{T_i}|M_i) d\boldsymbol{\theta}_i d\mathbf{q}_{T_i}, \quad (16)$$

where $\boldsymbol{\theta}_i$ and \mathbf{q}_{T_i} denote the vector of parameters and the vector of latent variables, respectively, in the model M_i . As it is pointed out by Geweke (2005, p. 23), the marginal density “is a prediction of what the data will be, before they are observed”.

To compare two econometric models, their ratio of posterior probabilities (called the posterior odds ratio) can be used. For any pair of competing models, M_i and M_j , the posterior odds ratio in favor of the model M_i is given by the following:

$$\frac{p(M_i|\mathbf{x}_T^o)}{p(M_j|\mathbf{x}_T^o)} = \frac{p(\mathbf{x}_T^o|M_i)}{p(\mathbf{x}_T^o|M_j)} \times \frac{p(M_i)}{p(M_j)}, \quad (17)$$

where $p(M_i)/p(M_j)$ is the prior odds ratio, and $p(\mathbf{x}_T^o|M_i)/p(\mathbf{x}_T^o|M_j)$ is the ratio of the marginal data density, called the Bayes factor. Thus the posterior odds ratio relating to models M_i and M_j is the product of the Bayes factor pertaining to models M_i and M_j and of the prior odds ratio. In turn, the marginal probability density function for \mathbf{x}_T in model M_i can be expressed as the product of conditional distributions:

$$p(\mathbf{x}_T|M_i) = \prod_{t=1}^T p(\mathbf{x}_t|\mathbf{x}_{t-1}, M_i). \tag{18}$$

Thus the Bayes factor in favor of the model M_i versus model M_j can be expressed as follows (Geweke and Amisano, 2010, p. 229):

$$\frac{p(\mathbf{x}_T^o|M_i)}{p(\mathbf{x}_T^o|M_j)} = \frac{\prod_{t=1}^T p(\mathbf{x}_t^o|\mathbf{x}_{t-1}^o, M_i)}{\prod_{t=1}^T p(\mathbf{x}_t^o|\mathbf{x}_{t-1}^o, M_j)} = \prod_{t=1}^T \frac{PL_{M_i}(t)}{PL_{M_j}(t)}, \tag{19}$$

where $PL_{M_i}(t) = p(\mathbf{x}_t^o|\mathbf{x}_{t-1}^o, M_i)$ denotes the one-step-ahead predictive likelihood in model M_i , $\mathbf{x}_{t-1}^o = [\mathbf{x}_1^o, \dots, \mathbf{x}_{t-1}^o]$. The one-step-ahead predictive likelihood, which is a real number, can be calculated as follows:

$$PL_{M_i}(t) = \int p(\mathbf{x}_t^o|\mathbf{x}_{t-1}^o, \boldsymbol{\theta}_i, \mathbf{q}_{.ii}, M_i) p(\boldsymbol{\theta}_i, \mathbf{q}_{.ii}|\mathbf{x}_{t-1}^o, M_i) d\boldsymbol{\theta}_i d\mathbf{q}_{.ii}, \tag{20}$$

where $\mathbf{q}_{.ii} = (q_{1i}, \dots, q_{ii})'$. The approximation of $PL_{M_i}(t)$ is:

$$P\hat{L}_{M_i}(t) = \frac{1}{M} \sum_{d=1}^M p(\mathbf{x}_t^o|\mathbf{x}_{t-1}^o, \boldsymbol{\theta}_i^{(d)}, \mathbf{q}_{.ii}^{(d)}, M_i), \tag{21}$$

where $\{\boldsymbol{\theta}_i^{(d)}, \mathbf{q}_{.ii}^{(d)}\}_{d=1}^M$ are drawn from $p(\boldsymbol{\theta}_i, \mathbf{q}_{.ii}|\mathbf{x}_{t-1}^o, M_i)$, see also Pajor (2021). $PL_{M_i}(t)$ constitutes a measure of how well model M_i predicts the observation \mathbf{x}_t^o conditional on \mathbf{x}_{t-1}^o . To compare how well alternative models predict the same set of observations, the predictive Bayes factor can be applied. The predictive Bayes factor in favor of M_i over M_j for observation t is defined as a ratio of the predictive likelihoods: $PL_{M_i}(t)/PL_{M_j}(t)$. From equation (19) it can be seen that the Bayes factor can be expressed in terms of predictive Bayes factors.

Let us assume that $T < S$. Then, the predictive likelihood, $p(\mathbf{x}_S^{o,T+1}|\mathbf{x}_T^o, M_i) = \prod_{t=T+1}^S PL_{M_i}(t)$, is a measure of how well the model M_i predicts the data $\mathbf{x}_S^{o,T+1} = [\mathbf{x}_{T+1}^o, \dots, \mathbf{x}_S^o]$ conditional on \mathbf{x}_T^o . For two models we obtain the following:

$$\frac{p(\mathbf{x}_S^{o,T+1}|\mathbf{x}_T^o, M_i)}{p(\mathbf{x}_S^{o,T+1}|\mathbf{x}_T^o, M_j)} = \prod_{t=T+1}^S \frac{PL_{M_i}(t)}{PL_{M_j}(t)}, \tag{22}$$

and consequently the cumulative log predictive Bayes factor for period $T + 1$ through S can be calculated as follows:

$$\log \left[\frac{p(\mathbf{x}_S^{o,T+1}|\mathbf{x}_T^o, M_i)}{p(\mathbf{x}_S^{o,T+1}|\mathbf{x}_T^o, M_j)} \right] = \sum_{t=T+1}^S \log \left[\frac{PL_{M_i}(t)}{PL_{M_j}(t)} \right]. \tag{23}$$

The formula (22) shows that the predictive Bayes factor is a multiplicative updating factor applied to the predictive likelihood, after the individual observation that produces the new predictive likelihood. Each individual observation contributes to the evidence of superiority of one model over another. In equations (22) and (23) the accumulation of this evidence starts at $t = T + 1$, and thus, if we were to regard $p(\theta_i, \mathbf{q}_{.Si} | \mathbf{x}_T^o, M_i)$ as the prior distribution for $\theta_i, \mathbf{q}_{.Si}$ (that is, if \mathbf{x}_T^o were to be treated as a training sample), then the ratio (22) would have the same interpretation as the Bayes factor. In other words, the first T observations are used to update the prior distribution, whereas the remaining ones are used to compare the models. Further details can be found in Geweke (2005) and Geweke and Amisano (2010).

It is worth mentioning that the use of the logarithm of the Bayes factor is equivalent to the use of the so-called ignorance or logarithmic score (see Gneiting & Raftery, 2007). The logarithm of the Bayes factor is the difference of the logarithmic scores for corresponding two models. Thus, the predictive Bayes factor is equivalent to the difference in the logarithmic score.

4 Forecast evaluation with probability integral transformations

The predictive Bayes factor measures the relative predictive performances of models. It depends only on the predictive probability densities evaluated at the observed values of the future data. An alternative, but non-Bayesian, assessment of the predictive performance of a model is provided by the probability integral transformation (PIT) of Rosenblatt (1952). The idea based on the PIT derives from the fact that, if a random variable z has the cumulative distribution function F , then $F(z)$ has the uniform distribution over the interval $[0, 1]$. Let $\left\{x_{j;t-1+h}^o\right\}_{t=T+1}^{T+N}$ be the realized values of $x_{j;t-1+h}$ at time $t = T + 1, \dots, T + N$, and let $\hat{p}_j(u | \mathbf{x}_{.t-1}^o, M_i)$ be the assessment of the conditional predictive density for the j -th component of vector \mathbf{x}_{t-1+h} in the model M_i . The PIT corresponding to model M_i as well as to the sequence $\left\{x_{j;t-1+h}^o\right\}_{t=T+1}^{T+N}$ is a sequence of values of the cumulative density function:

$$\begin{aligned} z_{j;t-1+h} &= \int_{-\infty}^{x_{j;t-1+h}^o} \hat{p}_j(u | \mathbf{x}_{.t-1}^o, M_i) du \\ &\equiv \hat{F}_j(x_{j;t-1+h}^o | \mathbf{x}_{.t-1}^o, M_i), j = 1, \dots, n. \end{aligned} \quad (24)$$

Under some regularity conditions (see Diebold et al., 1998) and if the realized values $\left\{x_{j;t-1+h}^o\right\}_{t=T+1}^{T+N}$ are generated by the predictive distribution $p_j(u | \mathbf{x}_{.t-1}^o, M_i)$, then $\hat{F}_j(x_{j;t-1+h}^o | \mathbf{x}_{.t-1}^o, M_i)$ has a uniform distribution over the interval $[0, 1]$. Moreover, if $h = 1$, then the sequence $\left\{z_{j;t-1+h}\right\}_{t=T+1}^{T+N}$ is distributed i.i.d. uniform $[0, 1]$ *ex ante* (see Geweke & Amisano, 2010). The uniformity of the PIT can be tested, e.g., using the Kolmogorov-Smirnov or Anderson–Darling tests (see Rossi & Skhposyan, 2014). The test statistics are as follows:

- Kolmogorov-Smirnov test (KS, Kolmogorov, 1933; Smirnov, 1948):

$$KS_j = \sqrt{N} \max_{t=1,2,\dots,N} \max \left\{ \left| z_{j,t}^* - t/N \right|, \left| z_{j,t}^* - (t-1)/N \right| \right\}$$

- Anderson-Darling test (AD, Anderson & Darling, 1952, 1954):

$$AD_j = -N - \frac{1}{N} \sum_{t=1}^N (2t-1) \ln \left(z_{j,t}^* \left(1 - z_{j,N+1-t}^* \right) \right)$$

where $\left\{ z_{j,t-1+h}^* \right\}_{t=T+1}^{T+N}$ denotes the values of $\left\{ z_{j,t-1+h} \right\}_{t=T+1}^{T+N}$ in ascending order.

5 Empirical results

In this part of the paper we analyze financial data of daily quotations on two major Polish exchange rates, PLN/EUR and PLN/USD, and on EUR/USD (see Fig. 1). We consider the exchange rates, where the cointegrating vector may be assumed to be known *a priori*, in order to compare the forecast performance of our models in the presence of the long-run relationship and time-varying volatility.

Following Pajor and Wróblewska (2017), we consider two average daily Polish official exchange rates downloaded from the website of the National Bank of Poland: the zloty (PLN) values of the euro (EUR/PLN, $y_{2,t}$), and the zloty values of the US dollar (USD/PLN, $y_{3,t}$), over the period from January 2, 2013 to February 3, 2017. Due to the fact that the Polish official exchange rates are linked to the exchange rates quoted on Forex, we introduce the euro value of the US dollar, downloaded from <http://stoq.com>: EUR/USD ($y_{1,t}$). The dataset of hundreds of logarithms of

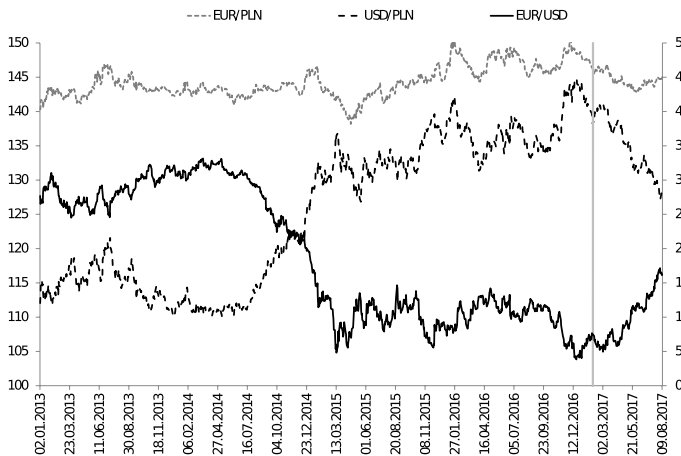


Fig. 1 The Polish official exchange rates (left axis) and the euro value of the US dollar (right axis), expressed in logarithms (in hundreds). Source of the data: https://www.nbp.pl/home.aspx?f=/kursy/arch_a.html in the case of the Polish official exchange rates; <https://stoq.com/> in the case of the EUR/USD exchange rate

exchange rates consists of 1032 observations (for each series). The first five observations are treated as an initial condition, thus $T = 1027$. We consider one-step-ahead predictions during the period beginning on February 3, 2017 and ending on August 8, 2017, thus $S = T + N = 1027 + 130 = 1157$. We calculate predictive distributions of \mathbf{x}_t based on the whole dataset available at time $T + i$ for each $i = 0, 1, \dots, N$ (up to $T + N = 1162$). Thus, we obtained 130 predictive distributions for a 1-day forecast horizon. The computations are based on the 30000 Markov chain Monte Carlo posterior samples (after having burnt 30000 cycles) in each model. Due to the typical properties of daily growth rates of exchange rates and the relationship: $(\text{EUR/PLN})/(\text{USD/PLN}) \approx \text{EUR/USD}$, we consider VEC-MSF, VEC-SBEKK, and VEC-MSF-SBEKK models with one and two cointegrating vectors ($r = 1, 2$), containing an unrestricted vector of constants ($d = 3$), and the number of lags equal to 2 ($k = 2$). The above approximate relation among exchange rates (in logarithmic terms) yields the cointegrating vector $(1, -1, 1)$ and the long-run relationship: $100 \ln y_{1,t} = 100 \ln y_{2,t} - 100 \ln y_{3,t}$, i.e., $x_{1,t} = x_{2,t} - x_{3,t}$. For the sake of comparison, we also consider random walk with drift processes ($k = 1, r = 0$), VEC models with constant conditional covariance matrix, VAR-SBEKK, and the VAR-MSF-SBEKK model for the levels of these processes ($r = 3$). The latter model is inadequate for the type of the data considered, but we use it to check the relative forecasting performance of the "wrong" model.

5.1 A comparison of models with the predictive Bayes factor

The overall ranking of the compared models M_i as well as $\log(p(\mathbf{x}_f | \mathbf{x}_T, M))$, and also the decimal logarithms of the predictive Bayes factors in favor of the VEC-MSF-SBEKK specification ($\log(B_{1,i})$) are shown in Table 1.

Table 1 provides the log predictive likelihoods and the log predictive Bayes factors in favor of the VEC-MSF-SBEKK model over thirteen selected specifications of the VEC model. As expected, the highest log predictive likelihood is in VEC-MSF-SBEKK with two cointegrating vectors ($r = 2$). The random walk model with drift and with constant conditional covariance matrix has the lowest predictive likelihood, rounding out the ranking. Models with time-varying condition covariance matrix and with one or two cointegrating vectors occupy four of the top positions. Using the VEC process together with the MSF-SBEKK covariance structure seems to be a powerful modeling strategy. Models with one or two cointegrating vectors have more explanatory power than models with $r = 0$. Even the VEC specification with constant conditional covariance matrix, but with one cointegrating vector, prevails over the VEC-SBEKK, the VEC-MSF and over the VEC-MSF-SBEKK structures without any cointegration relationship in terms of the predictive data density value (a natural Bayesian measure of fit). Capturing the long-run relationship (if one exists) turned out to be more important from a predictive point of view than allowing the conditional covariance matrix to be time-variable (stochastic). In other words, if the long-run relationships exist, allowing for the cointegration relationships in econometric models is more important than allowing conditional covariances to

Table 1 Decimal logs of predictive Bayes factors in favor of VEC-MSF-SBEKK ($d = 3, N = 130$)

VAR order (k)	Cointegration rank (r)	Model	$\log(p_{x_T})$	$\log(B_{1i})$
2	2	VEC-MSF-SBEKK	– 35.191	0.000
2	1	VEC-MSF-SBEKK	– 35.712	0.521
2	2	VEC-MSF	– 36.264	1.073
2	1	VEC-MSF	– 36.820	1.629
2	1 known <i>a priori</i>	VEC-MSF-SBEKK	– 38.498	3.307
2	1	VEC-SBEKK	– 40.525	5.334
2	2	VEC-SBEKK	– 42.410	7.216
2	1	VEC	– 43.958	8.767
2	0	VEC-MSF	– 54.211	19.020
2	0	VEC-MSF-SBEKK	– 54.726	19.535
2	0	VEC-SBEKK	– 58.642	23.451
2	0	VEC	– 59.916	24.725
1	0	VEC-MSF-SBEKK	– 75.114	39.922
1	0	VEC-MSF	– 75.893	40.702
1	0	VEC	– 80.977	45.785
2	3	VAR-MSF-SBEKK	– 119.363	84.172

$d = 3$ denotes an unrestricted constant, as in Pajor and Wróblewska (2017); $\log(p_{x_T}) = \log(p(x_T | x_T, M))$; scale for the strength of evidence against M_i (see Kass & Raftery, 1995): $0 < \log(B_{1i}) \leq 1/2$ – negligible, $1/2 < \log(B_{1i}) \leq 1$ – mild, $1 < \log(B_{1i}) \leq 2$ – strong, $2 < \log(B_{1i})$ – very strong

vary over time. It is worth mentioning that in the case of the Polish macroeconomic data the most important feature for the predictive ability of the vector autoregression models is allowing for changing covariance matrices, while the number of long-run relationships is of lesser importance, see Wróblewska and Pajor (2019).

It is no surprise that the VEC-SBEKK models fit the data much worse (in terms of the predictive Bayes factors) than their alternative variants with multiplicative stochastic factors, but still better than the VEC specifications with constant conditional covariance matrix. The VEC-MSF specification uses one latent process common to all conditional variances and covariances. This assumption of common dynamics leads to constant conditional correlations. Nevertheless, the VEC-SBEKK models, which do not use any latent processes and therefore are less flexible in dealing with outliers, appear still worse in term of the predictive Bayes factor.

The plots of cumulative log predictive Bayes factors for $t = T + 1, \dots, T + S$ are presented in Fig. 2. The sequence of cumulative log predictive Bayes factors in favor of the VEC-MSF-SBEKK structure with $k = 2, d = 3, r = 1$ (called MSF-SBEKK231) over the one with $k = 2, d = 3, r = 0$ (called MSF-SBEKK230) tends to increase with the rise of the number of predictions. The log predictive Bayes factors in favor of the MSF-SBEKK231 specification over the MSF-SBEKK232 are less than 1 each. According to the scale presented by Kass and Raftery (1995), it is a negligible strength of evidence against the MSF-SBEKK232 model. Furthermore, the cumulative log predictive Bayes factors in favor of the MSF-SBEKK231 model over the VAR-MSF with $k = 2, d = 3, r = 1$ (MSF231) indicate strong but not

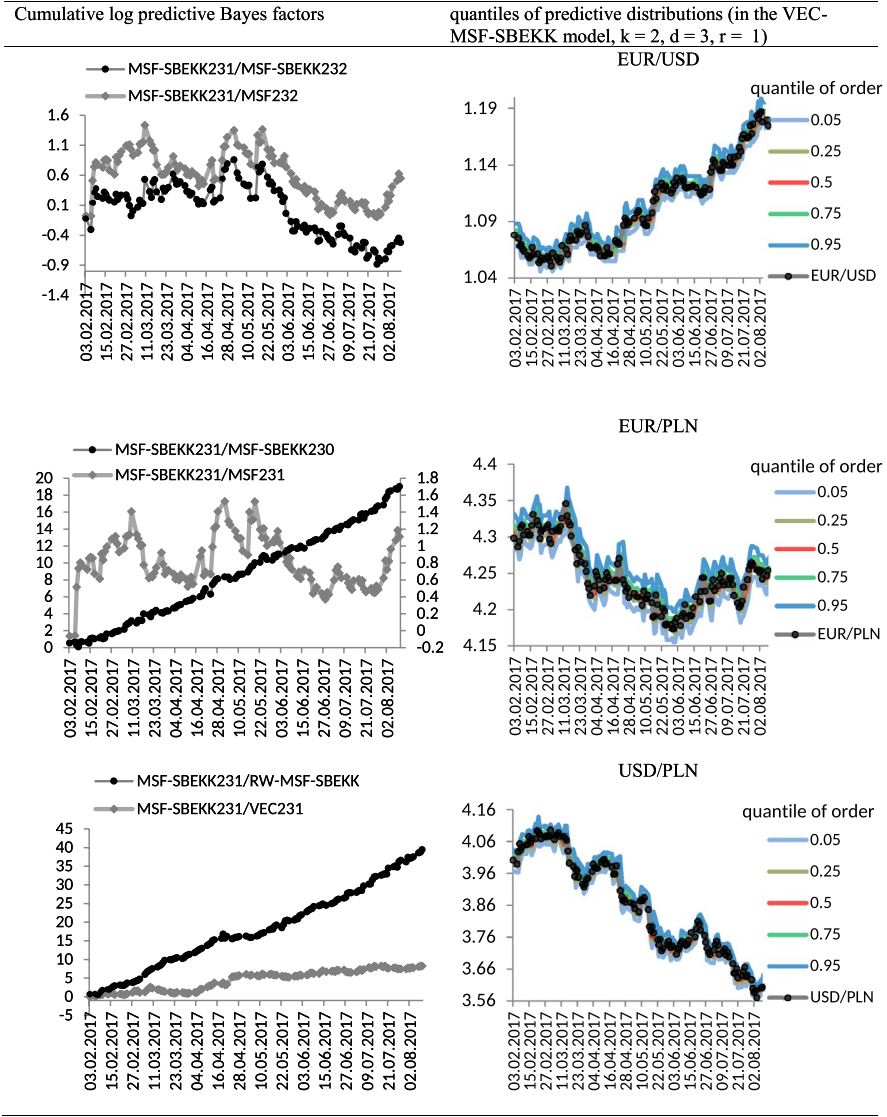


Fig. 2 Cumulative log predictive Bayes factors in favor of the VEC-MSF-SBEKK model over selected ones, and quantiles of predictive distributions

very strong (decisive) evidence against the latter. The right panels of Fig. 2 depict quantiles of predictive distributions obtained in the VEC-MSF-SBEKK model with $r = 1$ (very similar results were obtained for $r = 2$). Black points represent the real returns. As can be seen, the realized returns (out-of-sample data) lie in areas of high values of predictive densities. As expected, a very small value of each predictive Bayes factor in favor of VAR-MSF-SBEKK for the levels of the exchange rates

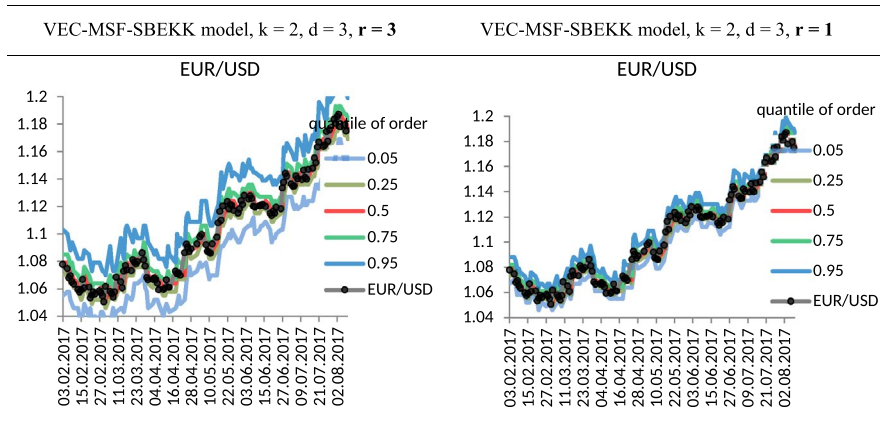


Fig. 3 Quantiles of predictive distributions and “future” observations

(with $r = 3$) indicates poor predictive potential of the model. As can be seen from the Fig. 3, the predictive distributions obtained in VEC-MSF-SBEKK model with $r = 3$ are more diffused than those obtained in the same specification of the model, but with $r = 1$. This unsuitable specification leads to more diffuse predictive distributions and large predictive Bayes factors in favor of the VEC-MSF models. It is worth noting that, for the median of the 1-day ahead predictive distribution as a point predictor the root mean square forecast errors (RMSFE) obtained in both models are almost the same. As a by-product, our results show that the RMSFE is not a good indicator of predictive performance of models.

5.2 Forecast evaluation with PITs

An alternative, but non-Bayesian, assessment of the predictive performance of a model is provided by the probability integral transformation (PIT). The PIT is the inverse of the sequence of *ex ante* predictive cumulative distribution functions evaluated on the sequence of “realized” observation (*ex post*). Figures 4, 5, 6 provide a further comparison of the predictive performances of the VEC-MSF-SBEKK, VEC-MSF, and VEC models with the use of PIT values. Histograms of PIT values allow for an informal assessment of their uniformity.

In all models (excluding the VAR-MSF-SBEKK one with $r = 3$), the histograms of the PITs for the EUR/USD exchange rates show that more realizations fall in the right tail of the predictive distribution than would be expected if the PITs were uniformly distributed. Overall, there is a lack of uniformity for density forecasts of the EUR/USD exchange rates. The correct specification of the predictive distributions of the EUR/USD exchange rates is rejected in all cases. The figures also provide a visual presentation of the misspecification in the PITs: future realizations of the EUR/USD exchange rates are under-predicts. The evident inadequacy of the predictive densities for EUR/USD in the right tail can be explained by the unexpected depreciation of the US dollar to the euro in the forecasting period. In Kębłowski

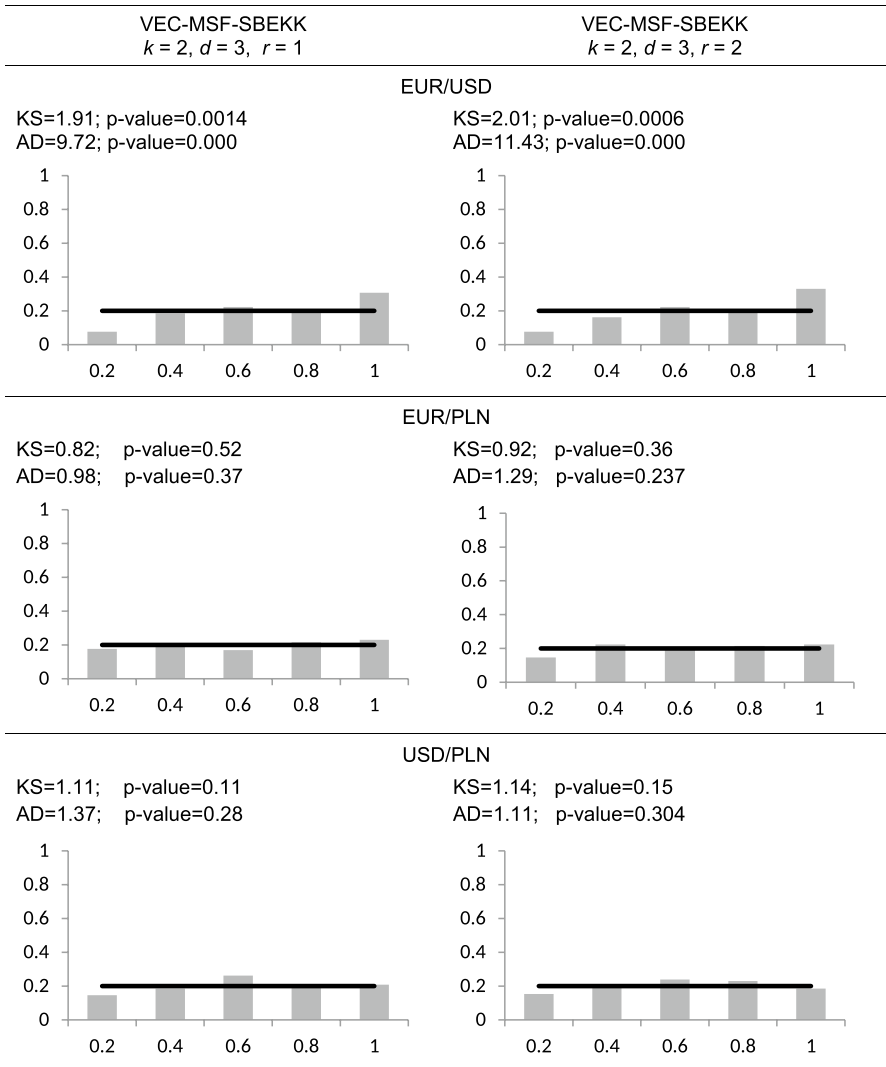


Fig. 4 PIT for selected VEC-MSF-SBEKK models with $r = 1$ and $r = 2$. The histograms depict the empirical distributions of the PITs. Solid lines represent the numbers of draws that are expected to fall into each bin under an $U(0, 1)$ distribution

et al. (2020) it has been shown that in the long-run the exchange rates of the EUR/USD and EUR/PLN are driven by crude oil prices, by the European Central Bank’s monetary policy and by the monetary policy decisions of the Federal Reserve System (see also Ghalayini, 2017).

The histograms of the PITs in the VAR-MSF-SBEKK (with $r = 3$) show that more “future” observations/realizations fall in the middle of the distribution than would be expected if the PITs had the uniform distributions.

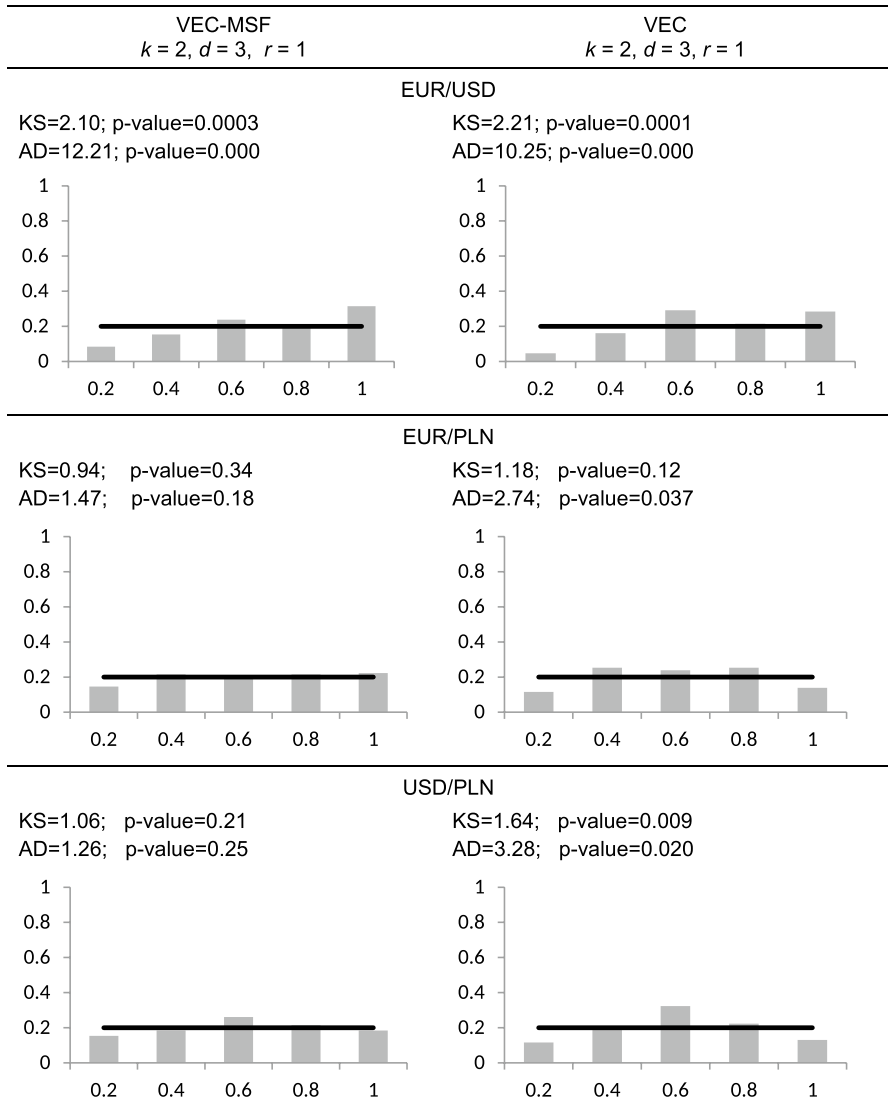


Fig. 5 PIT for selected VEC-MSF and VEC models. Meaning of symbols used the same as in Fig. 4

To test the correctness of the specification of the predictive densities (more specifically: whether the PIT is uniform) the Kolmogorov-Smirnov and Anderson-Darling tests are used. Figures 4, 5, 6 provide results for the Kolmogorov-Smirnov (labelled "KS") and Anderson-Darling (labelled "AD") tests of uniformity of the distribution of the PITs. In all models, the uniformity of PITs for the EUR/

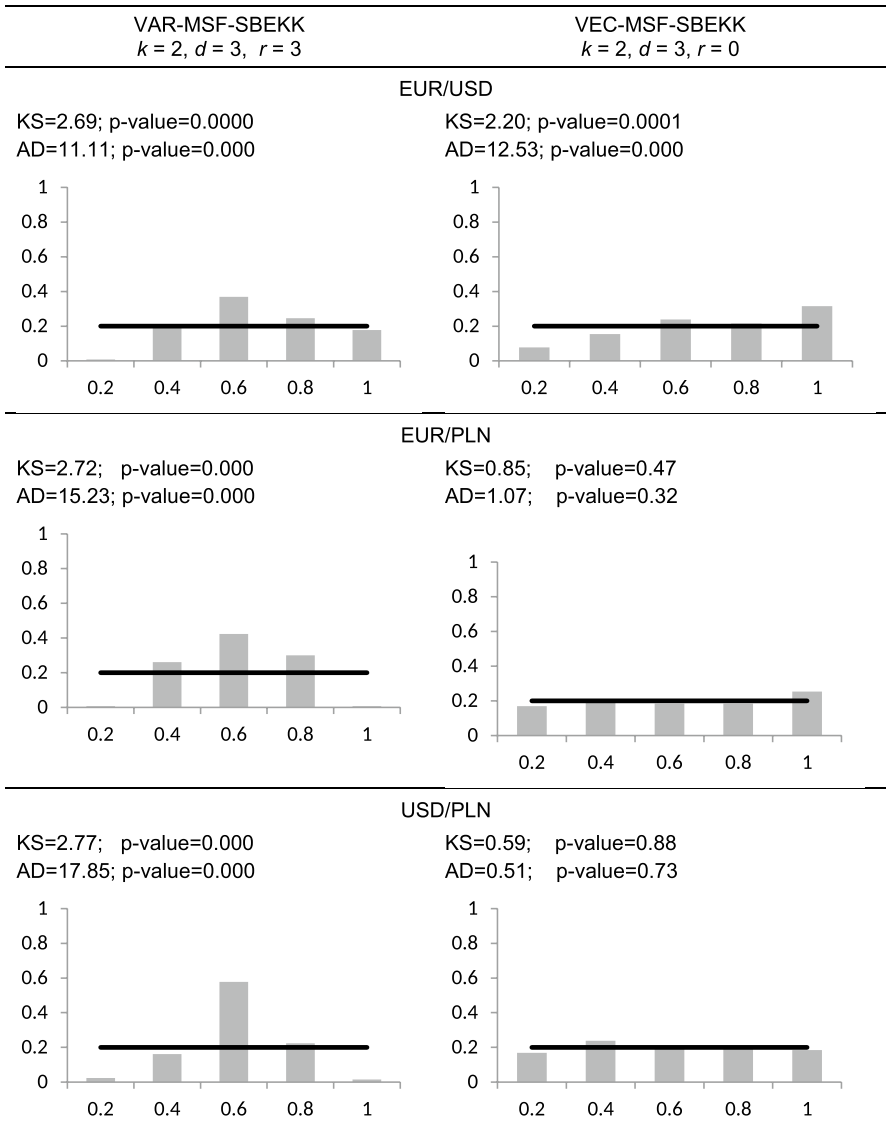


Fig. 6 PIT for selected VEC-MSF-SBEKK models with $r = 0$ and $r = 3$. Meaning of symbols used the same as in Fig. 4

USD exchange rates is rejected. Both the Kolmogorov-Smirnov and Anderson-Darling tests manifest empirical evidence against the correct specification of the VAR-MSF-SBEKK (with $r = 3$).

Table 2 p -value for autocorrelation test of PIT and PIT^2 for selected models (lags: 10, $k = 2$, $d = 3$)

Model	EUR/USD		EUR/PLN		USD/PLN	
	PIT	PIT^2	PIT	PIT^2	PIT	PIT^2
$r = 1$						
VEC-MSF-SBEKK	0.97	0.94	0.26	0.29	0.84	0.85
VEC-MSF	0.98	0.96	0.18	0.28	0.73	0.73
VEC	0.96	0.93	0.34	0.39	0.73	0.70
$r = 2$						
VEC-MSF-SBEKK	0.97	0.93	0.22	0.28	0.69	0.64
VEC-MSF	0.98	0.93	0.46	0.48	0.66	0.76
$r = 0$						
VEC-MSF-SBEKK	0.95	0.92	0.45	0.55	0.79	0.73

Regarding the independence of the PITs, the Ljung–Box test of no autocorrelation in the PITs and in the squares of the PITs has been used. Table 2 reports the p -values of the LB tests for serial correlation of PITs as well as of their squares. No such autocorrelation has been detected.

6 Conclusion

This paper evaluates the predictive performance of VEC-SV models in the context of modeling financial time series and in the presence of long-run relationships. Financial data of daily quotations on three exchange rates: PLN/EUR, PLN/USD, and EUR/USD was analyzed. As there exists at least one long-run relationship among the exchange rates (which is known *a priori*), we could check whether these models which account for long-run relationships perform better than those without the relationships. The empirical findings indicate that, overall, the VAR models with one or two cointegration relationships perform better than those without cointegration relationship. The VAR models with constant conditional covariance matrix are dominated by those with time-varying conditional covariances. However, if the long-run relationships exist, allowing for the cointegration relationships in econometric models is more important than allowing for conditional covariance matrix to vary over time. The analysis of PITs also shows that the predictive distributions are not perfectly fitted. In all models, the uniformity of PITs for the EUR/USD exchange rates is rejected. More "realized" observations fall in the right tail of the predictive distribution than would be expected if the PITs were uniformly distributed. The lack of uniformity of the distribution of the PIT suggests that the VEC-SV models considered should be modified by allowing for asymmetries of conditional distributions. This can be achieved by using, e.g., copulas or skew normal distributions.

Funding Research supported by a grant from the National Science Center in Poland under decision no. UMO-2018/31/B/HS4/00730.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abbate, A., & Marcellino, M. (2018). Point, interval and density forecast of exchange rates with time varying parameter models. *Journal of the Royal Statistical Society*, *181*(1), 155–179.
- Anderson, T. W., & Darling, D. A. (1952). Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes. *Annals of Mathematical Statistics*, *23*(2), 193–212.
- Anderson, T. W., & Darling, D. A. (1954). A test of goodness-of-fit. *Journal of the American Statistical Association*, *49*(268), 765–769.
- Anderson, R. G., Hoffman, D. L., & Rasche, R. H. (2002). A vector error-correction forecasting model of the U.S. Economy. *Journal of Macroeconomics*, *24*(4), 569–598.
- Asai, M., McAleer, M., & Yu, J. (2006). Multivariate stochastic volatility: A review. *Econometric Reviews*, *25*, 145–175.
- Berg, T. O. (2017). Forecast accuracy of a BVAR under alternative specifications of the zero lower bound. *Studies in Nonlinear Dynamics and Econometrics*, *21*(2), 1–29.
- Cavaliere, G., Angelis, L. D., Rahbek, A., & Taylor, A. M. R. (2015). A comparison of sequential and information-based methods for determining the co-integration rank in heteroskedastic VAR models. *Oxford Bulletin of Economics and Statistics*, *77*, 106–128.
- Cavaliere, G., Angelis, L. D., Rahbek, A., & Taylor, A. M. R. (2018). Determining the cointegration rank in heteroskedastic VAR models of unknown order. *Econometric Theory*, *34*(2), 349–82.
- Chan, J. C. C., & Eisenstat, E. (2018). Bayesian model comparison for time-varying parameter VARs with stochastic volatility. *Journal of Applied Econometrics*, *33*(4), 509–532.
- Chib, S., Omori, Y., & Asai, M. (2009). Multivariate stochastic volatility. In T. G. Andersen, R. A. Davis, J.-P. Kreiss, & T. Mikosch (Eds.), *Handbook of financial time series* (pp. 365–400). Springer-Verlag.
- Chikuse, Y. (2002). Statistics on special manifolds. *Lecture notes in statistics* (Vol. 174). Springer-Verlag.
- Clark, T. E. (2011). Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business and Economic Statistics*, *29*(3), 327–341.
- Clark, T. E., & Ravazzolo, F. (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, *30*(4), 551–575.
- D'Agostino, A., Gambetti, L., & Giannone, D. (2013). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics*, *28*, 82–101.
- Diebold, F. X., Gunther, T. A., & Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, *39*(4), 863–883.
- Eratalay, M. (2016). Estimation of multivariate stochastic volatility models: A comparative monte Carlo study. *International Econometric Review*, *8*(2), 19–52.
- Geweke, J. (2005). Contemporary Bayesian econometrics and statistics. *Wiley series in probability and statistics*. Wiley-Interscience [John Wiley and Sons].
- Geweke, J., & Amisano, G. (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, *26*(2), 216–230.
- Ghalayini, L. (2017). Modeling and forecasting spot oil price. *Eurasian Business Review*, *7*, 355–373.

- Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction and estimation. *Journal American Statistical Association*, 102, 359–378.
- Herwartz, H., & Lütkepohl, H. (2011). Generalized least squares estimation for cointegration parameters under conditional heteroskedasticity. *Journal of Time Series Analysis*, 32(3), 281–291.
- Huber, F., Koop, G., & Pfarrhofer, M. (2020). Bayesian inference in high-dimensional time-varying parameter models using integrated rotated gaussian approximations. Available at: <https://arxiv.org/abs/2002.10274> (Accessed on: 11 November, 2021)
- Huber, F., & Zörner, T. O. (2019). Threshold cointegration in international exchange rates: A Bayesian approach. *International Journal of Forecasting*, 35(2), 458–473.
- Jacquier, E., Polson, N., & Rossi, P. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics*, 122, 185–212.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795.
- Kastner, G., & Huber, F. (2021). Sparse Bayesian vector autoregressions in huge dimensions. *Journal of Forecasting*, 39(7), 1142–1165.
- Kębłowski, P., Leszkiewicz-Kędzior, K., & Welfe, A. (2020). Real exchange rates. *Oil Price Spillover Effects, and Tripolarity, Eastern European Economics*, 58(5), 415–435.
- Kolmogorov, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano Degli Attuari*, 4, 83–91.
- Koop, G., León-González, R., & Strachan, R. W. (2010). Efficient posterior simulation for cointegrated models with priors on the cointegration space. *Econometric Reviews*, 29, 224–242.
- Koop, G., León-González, R., & Strachan, R. W. (2011). Bayesian inference in the time varying cointegration model. *Journal of Econometrics*, 165, 210–220.
- Kuo, C. Y. (2016). Does the vector error correction model perform better than others in forecasting stock price? An application of residual income valuation theory. *Economic Modelling*, 52(B), 772–789.
- Osiewalski, J. (2009). New hybrid models of multivariate volatility (a Bayesian perspective). *Przegląd Statystyczny (Statistical Review)*, 56(1), 15–22.
- Osiewalski, K., & Osiewalski, J. (2013). A long-run relationship between daily prices on two markets: The Bayesian VAR(2)-MSF-SBEKK model. *Central European Journal of Economic Modelling and Econometrics*, 5(1), 65–83.
- Osiewalski, J., & Osiewalski, K. (2016). Hybrid MSV-MGARCH models general remarks and the GMSF-SBEKK specification. *Central European Journal of Economic Modelling and Econometrics*, 8(4), 241–271.
- Osiewalski, J., & Pajor, A. (2009). Bayesian analysis for hybrid MSF-SBEKK models of multivariate volatility. *Central European Journal of Economic Modelling and Econometrics*, 1(2), 179–202.
- Osiewalski, J., & Pipień, M. (2016). Bayesian comparison of bivariate GARCH processes. The role of the conditional mean specification. In A. Welfe (Ed.), *New directions in macromodelling* (pp. 173–196). Elsevier.
- Pajor, A. (2021). New estimators of the Bayes factor for models with high-dimensional parameter and/or latent variable spaces. *Entropy*, 23(4), 399.
- Pajor, A., & Osiewalski, J. (2012). Bayesian value-at-risk and expected shortfall for a large portfolio (multi- and univariate approaches). *Acta Physica Polonica A*, 121(2–B), B-101–B-109.
- Pajor, A., & Wróblewska, J. (2017). VEC-MSF models in Bayesian analysis of short- and long-run relationships. *Studies in Nonlinear Dynamics and Econometrics*, 21(3), 1–22.
- Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, 23(3), 470–472.
- Rossi, B., & Skhposyan, T. (2014). Evaluating predictive densities of us output growth and inflation in a large macroeconomic data set. *International Journal of Forecasting*, 30, 662–682.
- Seo, B. (2007). Asymptotic distribution of the cointegrating vector estimator in error correction models with conditional heteroskedasticity. *Journal of Econometrics*, 137, 68–111.
- Shephard, N., & Andersen, T. G. (2009). Stochastic volatility: Origins and overview. In T. G. Andersen, R. A. Davis, J.-P. Kreiss, & T. Mikosch (Eds.), *Handbook of financial time series* (pp. 233–254). Springer.
- Silvennoinen, A., & Teräsvirta, T. (2009). Multivariate GARCH models. In T. G. Andersen, R. A. Davis, J.-P. Kreiss, & T. Mikosch (Eds.), *Handbook of financial time series* (pp. 201–229). Springer.
- Smirnov, N. V. (1948). Tables for estimating the goodness of fit of empirical distributions. *The Annals of Mathematical Statistics*, 19(2), 279–281.

- Swanson, N. R. (2002). Comments on 'A vector error-correction forecasting model of the US economy'. *Journal of Macroeconomics*, 24(4), 599–606.
- Vardar, G., & Coskun, Y. (2018). Shock transmission and volatility spillover in stock and commodity markets: Evidence from advanced and emerging markets. *Eurasian Economic Review*, 8, 231–288.
- Wróblewska, J., & Pajor, A. (2019). One-period joint forecasts of polish inflation, unemployment and interest rate using Bayesian VEC-MSF models. *Central European Journal of Economic Modelling and Econometrics*, 11(1), 23–45.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.