www.czasopisma.pan.pl www.journals.pan.pl

Central European Journal of Economic Modelling and Econometrics

One-Period Joint Forecasts of Polish Inflation, Unemployment and Interest Rate Using Bayesian VEC-MSF Models

Justyna Wróblewska* and Anna Pajor†

Submitted: 28.10.2018, Accepted: 9.03.2019

Abstract

The paper aims at comparing forecast ability of VAR/VEC models with a non-changing covariance matrix and two classes of Bayesian Vector Error Correction – Stochastic Volatility (VEC-SV) models, which combine the VEC representation of a VAR structure with stochastic volatility, represented by the Multiplicative Stochastic Factor (MSF) process, the SBEKK form or the MSF-SBEKK specification.
Based on macro-data coming from the Polish economy (time series of unemployment, inflation and interest rates) we evaluate predictive density functions employing of such measures as log predictive density score, continuous rank probability score, energy score, probability integral transform. Each of them takes account of different feature of the obtained predictive density functions.

Keywords: cointegration, stochastic volatility, Bayesian analysis, forecast verification

JEL Classification: C11, C32

*Cracow University of Economics; e-mail: wroblewj@uek.krakow.pl; ORCID: 0000-0002-9789-2601
†Cracow University of Economics; Jagiellonian University in Kraków; ORCID: 0000-0001-5643-0649

1 Introduction

The vector autoregression (VAR) models with error correction mechanism and the changing conditional covariance matrix became popular in the analysis of macroeconomic time series, as such models enable to take into account both the long-run relationships (expressed by cointegrating relations) and the possibility of the changing variability (measured by the conditional covariance matrix). There is a growing literature discussing the forecast accuracy of Bayesian VAR models, possibly with time-varying parameters or with constant conditional mean but with the time-varying covariance structure or with other features and restrictions (see, e.g. Rossi, Sekhposyan 2014, Clark, Ravazzolo 2015, Berg 2017, Abbate, Marcellino 2017), but there is still not many papers comparing the forecast ability of vector error correction (VEC) models. Therefore the paper aims at comparing the forecast ability obtained within the set of Bayesian Vector Error Correction - Stochastic Volatility (VEC-SV) models, which combine the VEC representation of a VAR structure with stochastic volatility, represented by either the Multiplicative Stochastic Factor (MSF) process or the MSF-SBEKK specification. Additionally, similar to Pajor and Wróblewska (2017), we extend this set by Bayesian VEC models with non-changing covariance matrix. We also take into account VEC-SBEKK specifications, so the comparison is conducted within a slightly larger set then the one considered by Pajor and Wróblewska (2017). The predictive Bayes factors, the energy score and the continuous rank probability score are employed to compare the predictive ability of the studied models. To assess the calibration of the obtained forecast densities we apply the probability integral transform.
The paper is organized as follows. In Section 2 the basic framework is set and the Bayesian models are formulated. In Section 3 we outline the Bayesian forecast principles. Section 4 depicts the density forecast accuracy measures which are employed in Section 5, devoted to an empirical study, in which we use a standard set of the Polish macroeconomics variables (including the unemployment, inflation and interest rates). Section 6 concludes.

2 Bayesian VEC-SV models

Similar to Pajor and Wróblewska (2017) we consider the set of a linear n-variate and k-order vector autoregressive (VAR(k)) process with deterministic terms and a stochastic volatility (SV) structure. Three alternative structures for matrix Σ_t are analysed - Multiplicative Stochastic Factor (MSF), SBEKK and hybrid MSF-SBEKK (type I; see Osiewalski and Pajor 2009).
Equation (1) displays the VAR-SV process in the vector error correction (VEC) form, i.e. Vector Error Correction with Stochastic Volatility (VEC-SV) process:

www.czasopisma.pan.pl PAN POLSKA AKADEMIA NAUK www.journals.pan.pl

$$\Delta x_t = \tilde{\Pi} x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \Phi D_t + \varepsilon_t, \qquad t = 1, 2, \ldots, T, \tag{1}$$

$$\varepsilon_t | \psi_{t-1}, q_t, \theta \ \sim \ N\left(0, \Sigma_t\right), \tag{2}$$

where x_t is an $n \times 1$ random vector, $\{\epsilon_t\}$ is a vector white noise with some covariance matrix Σ (i.e. $\{\epsilon_t\} \sim WN(0, \Sigma)$), $\tilde{\Pi}$ and Γ_i are $n \times n$ matrices of real coefficients ($i = 1, \ldots k-1$), matrix D_t is comprised of deterministic variables such as the constant and centred seasonal dummies, Φ is a parameter matrix, q_t is the vector of latent variables, θ is a vector of parameters, ψ_{t-1} denotes the past of the process $\{x_t\}$ up to time $t-1$, and $\Sigma_t = \Sigma(q_t, \psi_{t-1})$. Moreover, $\tilde{\Pi} = \alpha\tilde{\beta}'$, with α and $\tilde{\beta}'$ being some $n \times r$ matrices, where $r < n$ is the number of cointegration relationships (if they exist). The initial conditions x_{-k+1}, x_{-k+2}, ..., x_0 are assumed to be known.
The Multiplicative Stochastic Factor structure for matrix Σ_t is as follows:

$$\Sigma_t = q_t \Sigma, \tag{3}$$

with $\ln q_t = \phi \ln q_{t-1} + \sigma_q \eta_t$, $\{\eta_t\} \sim iiN(0,\ 1)$, and $\epsilon_t \perp \eta_s$, for $t, s \in \mathbb{Z}$.
Henceforth, this specification will be referred to as the VEC-MSF (VEC with Multiplicative Stochastic Factor) process. In the VEC-MSF process the same q_t factor drives the dynamics of each element of Σ_t, so although conditional covariances vary over time, the conditional correlations remain constant. Such assumption may be empirically to restrictive, that is why we consider the MSF-SBEKK structure as an alternative MSV specification (see Osiewalski 2009, Osiewalski and Pajor 2009). We assume the so-called type I hybrid MSF-SBEKK process for matrix Σ_t:

$$\Sigma_t = q_t \tilde{\Sigma}_t, \tag{4}$$

$$\tilde{\Sigma}_t = (1 - a - b)\Sigma + b\left(\varepsilon_{t-1}\varepsilon_{t-1}'\right) + a\tilde{\Sigma}_{t-1}, \tag{5}$$

with $\ln q_t = \phi \ln q_{t-1} + \sigma_q \eta_t$, $\{\eta_t\} \sim iiN(0,\ 1)$, and $\epsilon_t \perp \eta_s$, for $t, s \in \mathbb{Z}$, $a, b \in \mathbb{R}$.
Matrix $\tilde{\Sigma}_t$ is square, of order n and follows the scalar BEKK(1,1) structure. The specification is further referred to as VEC-MSF-SBEKK. The presence of the scalar BEKK(1,1) structure in the conditional covariance matrix allows us to model time-varying conditional correlations without introducing more latent processes. Note that for $b = 0$ and $a = 0$ we obtain the VEC-MSF structure. In the limiting case when $\sigma_q \to 0$ and $\phi = 0$, the VEC-MSF-SBEKK model becomes the VEC-SBEKK one. As regards the initial conditions for $\tilde{\Sigma}_t$, we assume $\varepsilon_0 = 0$, $\tilde{\Sigma}_0 = s_{0,\Sigma} I_n$, where $s_{0,\Sigma} > 0$ and I_n denotes the identity matrix of size n.
Equation (1) can be decomposed and written as:

$$\begin{aligned} \Delta x_t &= \alpha[\tilde{\beta}', \Phi_1'] \begin{bmatrix} x_{t-1} \\ D_t^{(1)} \end{bmatrix} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \Phi_2 D_t^{(2)} + \varepsilon_t = \\ &= \alpha\beta' z_{1,t} + \Gamma' z_{2,t} + \Gamma_s' z_{3,t} + \varepsilon_t, \end{aligned} \tag{6}$$

where $\beta' = [\tilde{\beta}', \Phi_1']$, $z_{1,t} = [x_{t-1}', D_t^{(1)'}]'$, $z_{2,t} = (\Delta x_{t-1}', \Delta x_{t-2}', \ldots, \Delta x_{t-k+1}')'$, $z_{3,t} = D_t^{(2)}$, $\Gamma = [\Gamma_1, \Gamma_2, \ldots, \Gamma_{k-1}]'$, and $\Gamma_s = \Phi_2'$.
To simplify the notation we write the basic model (6) in a matrix form:

$$Z_0 = Z_1\Pi' + Z_2\Gamma + Z_3\Gamma_s + E, \tag{7}$$

where
$\Pi = \alpha\beta'$, $Z_0 = [\Delta x_1, \Delta x_2, \ldots, \Delta x_T]' = [z_{0,1}, z_{0,2}, \ldots, z_{0,T}]'$, $Z_1 = [z_{1,1}, z_{1,2}, \ldots, z_{1,T}]'$, $Z_2 = [z_{2,1}, z_{2,2}, \ldots, z_{2,T}]'$, $Z_3 = [z_{3,1}, z_{3,2}, \ldots, z_{3,T}]'$, $E = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T]'$.
The conditional distribution of x_t (given the past of the process, ψ_{t-1}, the parameters and the latent variable vector q_t) is n-variate Normal with the mean $\mu_t = x_{t-1} + \tilde{\Pi}x_{t-1} + \sum_{i=1}^{k-1}\Gamma_i\Delta x_{t-i} + \Phi D_t$ and the covariance matrix Σ_t:

$$p\left(x_t|\psi_{t-1}, \alpha, \beta, \Gamma, \Gamma_s, q_t, \Sigma, \theta_\Sigma, q_{0,\Sigma}\right) = f_{N,n}\left(x_t|\mu_t, \Sigma_t\right), \tag{8}$$

where θ_Σ and $q_{0,\Sigma}$ are the vectors of the stochastic volatility parameters: in the VEC-MSF model $\theta_\Sigma = (\phi, \sigma_q^2)'$ and $q_{0,\Sigma} = \ln q_0$, in the VEC-MSF-SBEKK model we have $\theta_\Sigma = (\phi, \sigma_q^2, a, b)'$ and $q_{0,\Sigma} = (\ln q_0, s_{0,\Sigma})'$, whereas in the VEC-SBEKK specification $\theta_\Sigma = (a, b)'$ and $q_{0,\Sigma} = s_{0,\Sigma}$. The vector $q_{0,\Sigma}$ is treated as an additional vector of parameters and is estimated jointly with other parameters. The density of the data (given the parameters) is the mixture (over $q = (q_1, q_2, \ldots, q_T)'$) distribution:

$$p\left(x|\alpha, \beta, \Gamma, \Gamma_s, \Sigma, \theta_\Sigma, q_{0,\Sigma}\right) = \int p\left(x|\alpha, \beta, \Gamma, \Gamma_s, \Sigma, \theta_\Sigma, q_{0,\Sigma}, q\right) p\left(q|\theta_\Sigma, q_{0,\Sigma}\right)\, dq, \tag{9}$$

where $x = [x_1, x_2, \ldots, x_T]'$ denotes the full data set. The two densities on the right hand side of (9) are given as:

$$p\left(x|\alpha, \beta, \Gamma_1, \Gamma_s, \Sigma, \theta_\Sigma, q_{0,\Sigma}, q\right) = \prod_{t=1}^{T} f_{N,n}\left(x_t|\mu_t, \Sigma_t\right),$$

and

$$p\left(q|\theta_\Sigma, q_{0,\Sigma}\right) = \prod_{t=1}^{T} q_t^{-1} f_{N,1}\left(\ln q_t|\phi\ln q_{t-1}, \sigma_q^2\right).$$

The Bayesian model is defined by the joint density of the vector of observations, latent variables and parameters:

$$\begin{aligned} &p\left(x, q, \theta\right) = p\left(x|q, \theta\right) p\left(q|\theta\right) p\left(\theta\right) = \\ &= p\left(\theta\right)\left[\prod_{t=1}^{T} f_{N,n}\left(x_t|\mu_t, \Sigma_t\right)\right]\left[\prod_{t=1}^{T} q_t^{-1} f_{N,1}\left(\ln q_t|\phi\ln q_{t-1}, \sigma_q^2\right)\right], \end{aligned} \tag{10}$$

where $\theta = (vec\alpha', vec\beta', vec\Gamma', vec\Gamma_s', vech\Sigma', \theta_\Sigma', q_{0,\Sigma})'$ is the parameter vector, q denotes the latent variable vector and $p(\theta)$ the density of prior distribution. The density (10) is also conditioned by certain initial observations, which are omitted from the notation. To complete the Bayesian models, we specify the prior distribution of the parameter vector θ. We assume that certain blocks of the parameters are *a priori* independent, and that the stability condition is imposed on the parameters of the VEC process:

$$p(\theta) = p(\alpha|\beta)p(\beta)p(\Gamma)p(\Gamma_s)p(\Sigma)p(\theta_\Sigma)p(q_{0,\Sigma})I_{[0,1]}(|\lambda|_{\max}), \tag{11}$$

where $I_{[a,b]}(.)$ denotes the indicator function of the interval $[a, b]$ and λ is the vector of the eigenvalues of the companion matrix, i.e. the matrix which makes it possible to write the analysed process in the VAR(1) form. As proposed by Koop *et al.* (2010) two parameterisation for matrix Π are considered:

$$\alpha\beta' = (\alpha M_\Pi)\left(\alpha {M_\Pi}^{-1}\right)' \equiv AB',$$

where M_Π is an $r \times r$ symmetric positive-definite matrix, A and B are unrestricted matrices, $\alpha = A(B'B)^{1/2}$, and $\beta = B(B'B)^{-1/2}$, so β has orthonormal columns, and it is an element of the Stiefel manifold $V_{r,m}$ (represented by the matrix space of $m \times r$ matrices with orthonormal columns). The data inform only about the cointegration space, which is the element of the Grassmann manifold $G_{r,m-r}$, i.e. the space of r-dimensional hyperplanes in $\mathbb{R}^m$. Additionally, we normalise the columns of β to have positive first elements (using a diagonal matrix whose elements are equal to either 1 or -1).

The prior distributions are the same as in Pajor and Wróblewska (2017) and are presented in section 4 (for the discussion see also Koop *et al.* 2010, Osiewalski, Pajor 2009 and Pajor, Osiewalski 2012).

3 Bayesian forecasting

The Bayesian forecast is based on the predictive distribution which is the conditional on the data distribution of the future values of the analysed series.

The joint density function of the observed data, the possible future values $(x^f = (x_{T+1}, x_{T+2}, \dots, x_S)')$, the forecasted latent variables $(q^f = (q_{T+1}, q_{T+2}, \dots, q_S)')$, the parameters and the latent variables up to time T can be decomposed as follows:

$$p\left(x, q, x^f, q^f, D, \theta\right) = p\left(x_f, q_f|x, q, D, \theta\right) p\left(x, q|\theta\right) p\left(\theta\right), \tag{12}$$

where D is the matrix with deterministic variables in the forecast period $D = (D_{T+1}, D_{T+2}, \dots, D_S)'$, $x = (x_1, x_2, \dots, x_T)'$ – the matrix of observable n-dimensional variables and $q = (q_1, q_2, \dots, q_T)$ – the vector of latent variables. The joint density is also conditioned by initial values, but we omit them from our notation. The predictive distribution is obtained as the average of the sampling predictive density over the parameters and latent variables space, with the use of the posterior density as the weight function:

$$p\left(x^f, q^f|x^o, D\right) = \int p\left(x^f, q^f|x^o, q, D, \theta\right) p\left(q, \theta|x^o\right) dq d\theta, \tag{13}$$

x^o contains the *ex post* observed values of x. If one is interested in predicting the future values of x, the predictive distribution of x^f is obtained:

$$p\left(x^f|x^o, D\right) = \int p\left(x^f, q^f|x^o, q, D, \theta\right) p\left(q, \theta|x^o\right) dq d\theta dq^f, \tag{14}$$

where

$$p\left(x^f, q^f|x^o, q, D, \theta\right) = \left[\prod_{t=T+1}^{S} f_{N,n}\left(x_t|\mu_t, \Sigma_t\right)\right]\left[\prod_{t=T+1}^{S} q_t^{-1} f_{N,1}\left(\ln q_t \ |\phi \ln q_{t-1} \ , \sigma_q^2\right)\right].$$

This predictive distribution summarises the whole information about the future values of x^f and takes into account uncertainty of the latent process and the parameters, conditional on the data, sampling model and a prior distribution.

4 Forecast accuracy measures

As was previously outlined as a result of Bayesian prediction we obtain the whole predictive density function, so we can compare predictive ability of the considered models by evaluating these functions. We use three measures to assess the obtained predictive densities. Each of them evaluates another feature of the forecast density.
We start by analysing predictive Bayes factors (see e.g. Geweke, Amisano 2010). They can be used to compare models both in the modelled and forecasted periods. In this paper we focus only on the forecast. The cumulative predictive Bayes factor in favour of a model M_i over a model M_j in the period $T+1$ through S, where $S > T$ reads as follows:

$$B_{ij} = \frac{p\left(x_S|x_T, M_i\right)}{p\left(x_S|x_T, M_j\right)} = \prod_{t=T+1}^{S} \frac{PL_{M_i}(t)}{PL_{M_j}(t)}, \tag{15}$$

where $PL_{M_i}(t)$ denotes the one-step-ahead predictive likelihood evaluated at time t. The analysis of the predictive Bayes factors' changes together with the path of the series in the forecast period may help to capture the reasons why one model is superior to the other.
As the log predictive density score is sensitive to outliers we also compute continuous rank probability score (CRPS) which is robust to extreme values (see e.g. Berg 2017, Clark, Ravazzolo 2015, Gneting, Raftery 2007, Hersbach 2000, Stelmasiak,

Szafrański 2016). The CRPS is intended to measure the distance between the occurred cumulative distribution ($I\{x \geq x_s\}$) and the predicted one (P):

$$CRPS\,(P, x_s) = \int_{-\infty}^{+\infty} \left(P\,(y) - I\,\{x \geq x_s\}\right)^2 dx = E_P\,|X - x_s| - \frac{1}{2} E_P\,|X - X'|\,, \quad (16)$$

where x_s denotes the realized value, X and X' are independent copies of a random variable with cumulative distribution function P, ($I\{x \geq x_s\}$) equals zero for values lower than the observed x_s and 1 in other cases. Note that the lower the value of CRPS the better the predictive ability of the model. The minimal value CRPS is equal to 0 and is achieved in the case of the perfect deterministic forecast. The CRPS is expressed in the same units as the predicted variables. To evaluate the obtained predictive densities for each of the analysed variables we will use the mean of CRPS values obtained within the whole forecast horizon:

$$\overline{CRPS}_{T+1}^{S} = \frac{1}{S-T} \sum_{t=T+1}^{S} CRPS\,(P, x_t). \quad (17)$$

Note that in the case of the deterministic forecast CRPS reduces to the mean absolute error.

If we are willing to assess the multidimensional predictive density the energy score (ES) is recommended (see e.g. Gneiting, Raftery 2007):

$$ES\,(P, x_s) = E_P \|X - x_s\|^{\beta} - \frac{1}{2} E_P \|X - X'\|^{\beta}, \quad (18)$$

where $\beta \in (0, 2)$, X and X' are independent copies of a random vector with distribution P, $\|.\|$ denotes the Euclidean norm. In this paper ES is used with $\beta = 1$. Note that for one-dimensional X and $\beta = 1$ the energy score equals the continuous rank probability score. Similarly to CRPS, we will present the mean of ES values obtained within the whole forecast horizon:

$$\overline{ES}_{T+1}^{S} = \frac{1}{S-T} \sum_{t=T+1}^{S} ES\,(P, x_t), \quad (19)$$

We make use of the above outlined measures to compare different forecast densities, whereas to assess the calibration of the forecast densities we employ probability integral transform (PIT), which is based on the evaluated predictive distribution obtained within model M_i and the observations (see e.g. Gneiting *et al.*, 2007, Geweke, Amisano 2010, Mitchell, Wallis 2011, Rossi, Sekhposyan 2014, Berg, Henzel 2015, Berg 2017):

$$PIT_1(t; M_i) = P\,(x_t | x_{t-1}, M_i)\,, \quad (20)$$

where the subscript 1 corresponds to the horizon of prediction, which is settle to one, x_{t-1} is a vector containing observations up to time $t-1$, x_t denotes the realization

of the forecasted value and P is a cumulated distribution function.
If the evaluated predictive density is consistent with the true one, then $\{PIT_1(t;M_i)\}_{t=T+1}^{S}$ are independent (note that when the forecast horizon h is longer than 1, then, by construction, PIT_h are dependent at least up to the lag $h-1$) and uniformly distributed: $\{PIT_1(t;M_i)\}_{t=T+1}^{S} \sim iiU(0,1)$. The visual inspection of the obtained PIT histograms can give some suggestions about deficiencies of the predictive density. Hump-shaped histograms indicate for over-dispersed distributions, U-shaped for too narrow and the triangle-shape are connected with biased ones (Gneiting *et al.* 2007). Following Rossi and Sekhposyan (2014) we employ Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) tests to formally check whether there are any violations of uniformity, and the Ljung-Box test to examine autocorrelation of PITs. KS test measures the difference between the empirical cumulative distribution of PITs and cdf of uniform distribution U(0,1):

$$KS = \sqrt{S-T} \max_{j=1,\ldots,S-T} \max\left\{\left|z_j^* - j/(S-T)\right|, \left|z_j^* - (j-1)/(S-T)\right|\right\} \quad , \qquad (21)$$

so this statistic treats all the points equally, whereas AD test places more weight on the tails of the empirical distribution:

$$AD = -(S-T) - \frac{1}{S-T}\sum_{j=1}^{S-T}(2j-1)\ln\left(z_j^*\left(1-z_{S-T+1-j}^*\right)\right). \qquad (22)$$

In the above presented statistics z_j^* denotes the values of $PIT_1(t;M_i)$ in ascending order.
Both statistics (KS and AD) have non-standard distributions.
The statistics of the Ljung-Box test of autocorrelation of lag length $\tilde{L}$ reads as follows:

$$Q = (S-T)(S-T+2)\sum_{l=1}^{\tilde{L}}\left(\frac{\rho_l^2}{S-T-l}\right) \sim_{as} \chi^2\left(\tilde{L}\right), \qquad (23)$$

where ρ_l^2 is the serial coefficient of PITs at lag l.
Although the main aim of the paper is to compare density forecast accuracy of the analysed set of models, we also present the weighted trace mean squared forecast error (WTMSFE) which assesses the accuracy of the point forecast based on the predictive mean (Carriero *et al.* 2011). We will calculate the mean of WTMSFE values obtained within the whole forecast horizon as:

$$\overline{WTMSFE}_{T+1}^{S} = \frac{1}{S-T}\sum_{t=T+1}^{S} WTMSFE\left(\hat{x}_{t+1}, x_t\right), \qquad (24)$$

where $\hat{x}_{t+1}$ denotes the one-step-ahead point forecast of the vector x_t, $WTMSFE\left(\hat{x}_{t+1}, x_t\right) = tr\left[\left(\hat{x}_{t+1}-x_t\right)' W\left(\hat{x}_{t+1}-x_t\right)\right]$ with W set as a diagonal matrix of weights accounting for different volatilities of the predicted variables. As

weights, we take the inverse of variance in the forecasted period (see also Carriero *et al.* 2011).
Note that the predictive mean is the optimal point forecast for the researchers with the quadratic loss function and in such a case the root mean squared error is an appropriate measure for discriminating among models (see e.g. Weiss 1996, Berg, Henzel 2015).

5 Results of forecast comparison

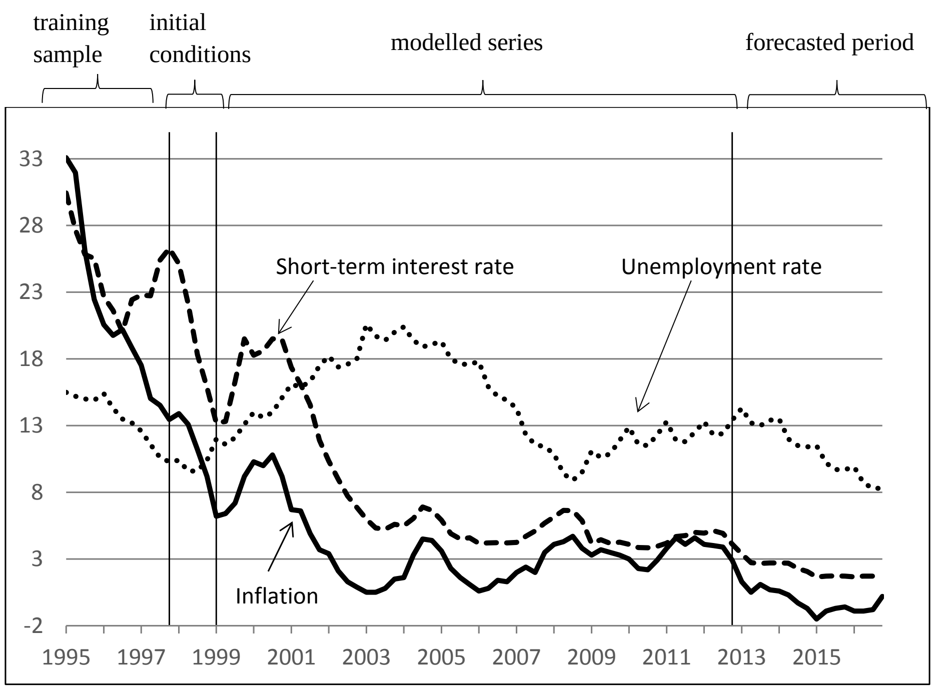
In this section the above presented methods are employed to assess the forecast accuracy of inflation of consumer prices (Δp_t) unemployment rate (U_t) and short-term interest rates (r_t) performed within the so called small model of monetary policy (see e.g. Primiceri 2005). This is the same model as was considered for the data from the Polish economy by Pajor and Wróblewska (2017) presenting within sample Bayesian comparison of different VEC-MSF specifications. The aim of this research is to assess forecast ability of these competing specifications and compare the obtained results to the ranking displayed by Pajor and Wróblewska (2017). For this reason we focus on the very same dataset. The modelled data cover the period from 1995Q1 to 2012Q4. The forecast evaluation is performed for 16 quarters ranging from 2013Q1 to 2016Q4. The analysed series are seasonally unadjusted and their seasonality is modelled and forecasted in a deterministic manner, i.e. using zero-mean seasonal dummies. The prior structure is the same as in Pajor and Wróblewska (2017), see also Koop *et al.* (2010), Osiewalski and Pajor (2009), Pajor and Osiewalski (2012) for the discussion. Similar to Pajor and Wróblewska (2017), the first three years of modelled data are used as a training sample to determine the hyperparameters of the prior distribution for the cointegration space (the Johansen procedure in the model with two lags and 2 cointegrating vectors β has been used, which further have been utilized to calculate $P_{0.5} = \beta\beta' + 0.5\beta_{\perp}\beta'_{\perp}$). The next 5 quarters (1998Q1 – 1999Q1) are sacrificed as initial conditions. Eventually, there are 55 modelled observations (see Figure 1).
The imposed priors are as follows:

1. the matrix normal distribution for B: $p\left(B|\, r\right) = f_{mN}\left(B|\, 0,\, I_r, P_{0.5}\right)$, which leads to the matrix angular central Gaussian (MACG) distribution for β: $p\left(\beta|\, r\right) = f_{MACG}\left(\beta|P_{0.5}\right)$ (see e.g. Chikuse 2002), $E\left(B\right) = 0$, $V\left(vec\ B\right) = I_r \otimes P_{0.5}$,

2. the matrix normal distribution for A: $p\left(A|\, \nu, r\right) = f_{mN}\left(A|0, \nu I_r, I_n\right)$ with inverse gamma distribution for ν: $p\left(\nu\right) = f_{IG}\left(\nu|\, 3, 2\right)$, so $E\left(\nu\right) = 1,\ V\left(\nu\right) = 1$ and $E\left(A\right) = 0,\ V\left(vec\ A|\nu\right) = \nu I_r \otimes I_n = \nu I_{rn}$,

3. the matrix normal distribution for Γ: $p\left(\Gamma|\, h\right) = f_{mN}\left(\Gamma|0, I_n, hI_l\right)$ with inverse gamma distribution for h: $p\left(h\right) = f_{IG}\left(h|\, 3, 2\right)$, $l = n\left(k-1\right)$, $E\left(h\right) = 1$, $V\left(h\right) = 1$, $E\left(\Gamma\right) = 0$, $V\left(vec\ \Gamma|h\right) = I_n \otimes hI_l = hI_{ln}$,

Figure 1: The analysed macro-data series (1995Q1 – 2016Q4)



4. the matrix normal distribution for Γ_s: $p(\Gamma_{\mathrm{s}}|\ h_s) = f_{mN}(\Gamma_{\mathrm{s}}|0, I_n, h_s I_{l_s})$ with inverse gamma distribution for h_s: $p(h_s) = f_{IG}(h_s|\ 3, 2)$, l_s denotes the number of deterministic terms in D_t, $E(h_s) = 1$, $V(h_s) = 1$, $E(\Gamma_{\mathrm{s}}) = 0$, $V(vec\ \Gamma_{\mathrm{s}}|h_s) = I_n \otimes h_s I_{l_s} = h_s I_{l_s n}$

5. the inverse Wishart distribution for Σ: $p(\Sigma|I_n, n+2)$, so $E(\Sigma) = I_n$,

6. the normal distribution for $\ln q_0$: $p(\ln q_0\ ) = f_{N,1}(\ln q_0\ |0, 1)$, so $E(\ln q_0\ ) = 0$, $V(\ln q_0\ ) = 1$,

7. the normal distribution for ϕ, truncated by the restriction $|\phi| < 1$:$p(\phi) \propto f_{N,1}(\phi|0.8,\ 0.2)\, I_{(-1,1)}(\phi)$,

8. the inverse gamma distribution for σ_q^2: $f_{IG}\left(\sigma_q^2|1.1,\ 0.04\right)$, so $E\left(\sigma_q^2\right) = 0.4$,

9. the uniform distribution over the unit simplex for a and b: $p(a,\ b) \propto I_{(0,1)}(a+b)$,

10. the exponential distribution for $s_{0,\Sigma}$: $p(s_{0,\Sigma}) = f_{Exp}(s_{0,\Sigma}|1)$, so $E(s_{0,\Sigma}) = 1$,

We analyse the set consisting of 96 non-nested specifications. Along with the number of latent processes driving covariances ($l = 0$ in the VEC and the VEC-SBEKK models, $l = 1$ in the VEC-MSF and VEC-MSF-SBEKK forms), the models can differ in the VAR order ($k \in \{2, 3, 4, 5\}$), the type of incorporated deterministic term ($d = 4$ denotes a constant restricted to the cointegration space, $d = 3$ – an unrestricted one) and the cointegration rank (r can be equal to 1 or 2). Additionally, we consider VAR models for the first differences of the analysed processes ($r = 0$) and VAR models for the levels of these processes ($r = 3$).
We start the analysis of models forecast accuracy by recalling the results of models comparison presented in Pajor and Wróblewska (2017).

Table 1: The most probable models

VAR order (k)	deterministic term (d)	cointegration rank (r)	model	$P(M_{(k,d,r,type)}\|x)$
2	3	0	VEC-MSF-SBEKK	0.98737
2	3	0	VEC-MSF	0.01116
2	4	1	VEC-MSF	0.00130
3	3	0	VEC-MSF-SBEKK	0.00005
2	4	1	VEC-MSF-SBEKK	0.00004
2	3	2	VEC-MSF	0.00003
5	3	0	VEC-MSF-SBEKK	0.00003
2	3	1	VEC-MSF	0.000014
2	4	2	VEC-MSF	0.000003
3	3	0	VEC-MSF	0.000002

Source: Pajor and Wróblewska (2017)

As can be noticed models with changing covariance structure gathered almost all the posterior probability. According to economic theory we could expect at least one cointegrating relation among the analysed series, but in this set of Bayesian models, these ones built for first differences of the original data are ranked first. The most probable model with one cointegrating relation took the third place.
We begin the evaluation of the predictive densities by discussing the obtained values of predictive likelihood. Table 2 displays the log predictive likelihood for the considered forecast period ($\log(P(x_f|x, M_{(k,d,r,type)}))$, and the log predictive Bayes factors in favour the model with the highest value of the predictive likelihood (i.e. the specification with two lags, two cointegrating relations, a constant restricted to the cointegration space and the MSF-SBEKK covariance structure) over some chosen models ($\log(B_{1i})$).
Among the models ranked first in Table 2 there are only those with changing covariance-matrix and with two cointegrating vectors. Pajor and Wróblewska (2017) found the changing covariance structure also very important in the modelled period, but contrary to the economic theory and prior expectations the model without

Table 2: The log predictive Bayes factor (2013Q1 – 2016Q4)

VAR order (k)	determin. term (d)	cointegration rank (r)	model	$\log\left(P\left(x_f \mid M_{(k,d,r,type)}\right)\right)$	$\log(B_{1i})$
2	4	2	VEC-MSF-SBEKK	−5.471	0
2	4	2	VEC-MSF	−5.861	0.393
2	3	2	VEC-MSF-SBEKK	−5.917	0.445
2	3	2	VEC-MSF	−6.016	0.544
2	3	1	VEC-MSF	-6.111	0.64
2	4	1	VEC-MSF-SBEKK	−6.126	0.655
2	4	1	VEC-MSF	−6.248	0.776
2	**3**	**0**	**VEC-MSF**	**−6.577**	**1.11**
3	3	0	VEC-MSF-SBEKK	−6.718	1.247
2	**3**	**0**	**VEC-MSF-SBEKK**	**−6.724**	**1.25**
2	3	3	VEC-MSF	−7.127	1.655
3	3	0	VEC-MSF	−7.233	1.762
2	4	2	VEC-SBEKK	−7.672	2.201
2	4	1	VEC-SBEKK	−7.878	2.407
2	3	2	VEC-SBEKK	−8.205	2.733
5	3	0	VEC-MSF-SBEKK	−8.772	3.301
2	3	1	VEC-SBEKK	−11.287	5.816
2	4	2	VEC	−12.72	7.245
2	4	1	VEC	−12.79	7.315
2	3	2	VEC	−13.05	7.574
3	3	0	VEC	−13.33	7.858

Source: own calculations based on 100 000 accepted draws

cointegrating relations gathered almost all the posterior probability. They also showed that when the priors for the cointegration space is settled with taking into account the cointegration rank of the analysed specification, models with one cointegrating vector are ranked first. The possible reason for this observed sensitivity for the prior assumptions is the shortness of the analysed time-series. It should be emphasised that models with long run relationships turned out to have the best predictive likelihood, so including the economic relationships into the model helps to achieve better prediction, in the sense of the predictive probability densities evaluated at the realized return. Basing on the results presented in Table 2, we can also asses which of the two considered model features turned out to be more important in this forecast exercise – changing covariance structure or cointegrating relations. The models with a constant covariance matrix are ranked last in Table 2, so the changing covariance structure turned out to be more important than economic relationships. Even the model with two cointegrating relations, but with constant covariance matrix, according to the rules proposed by Kass and Raftery (1995) is strongly rejected by the data, as its log Bayes factor equals 7.2. Contrary, log Bayes factors for the best specifications without cointegrating relations but with changing covariances are a little bit higher

www.czasopisma.pan.pl PAN POLSKA AKADEMIA NAUK www.journals.pan.pl

than one.
Table 3 displays models with the lowest values of the mean of energy score calculated for one-step-ahead forecasts in the considered evaluation period. Among them there are only models with time varying covariances. The first model with a constant covariance matrix ($M_{(2,4,1,VEC)}$) is ranked on the 40^{th} position with the mean energy score 17.3% higher than the ES of the best model ($M_{(2,4,2,VEC-MSF-SBEKK)}$). Summing up to the results presented in Tables 2 and 3, we can conclude that accounting for the changing covariance matrix, even with very simple structure as MSF, is more important than the number of long-run relationships. Additionally, the models with lower lag length are preferred.

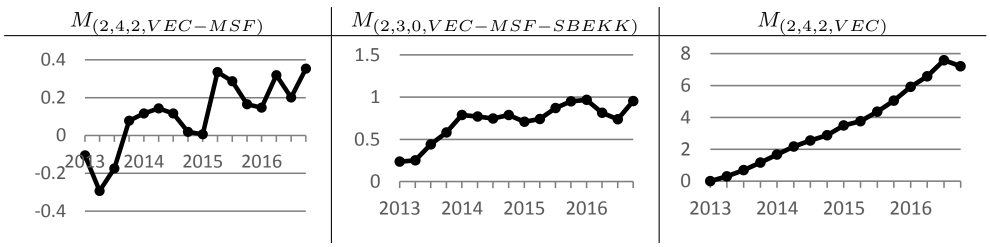
Table 3: The energy score (mean in the period 2013Q1 – 2016Q4) and the percentage loss to the best model,
$pl_{ES} = 100 \left(\overline{ES}_{T+1}^{T+16} \left(M_{(k,d,r,type)} \right) / \ \overline{ES}_{T+1}^{T+16} \left(M_{(2,4,2,VEC-MSF-SBEKK)} \right) - 1 \right)$

VAR order (k)	determin. term (d)	cointegration rank (r)	model	$\overline{ES}_{T+1}^{T+16} \left(M_{(k,d,r,type)} \right)$	pl_{ES}
2	4	2	VEC-MSF-SBEKK	0.4059	0.00
2	3	2	VEC-MSF	0.4070	0.25
2	4	2	VEC-MSF	0.4082	0.56
2	3	1	VEC-MSF	0.4111	1.28
2	4	1	VEC-MSF-SBEKK	0.4120	1.49
2	4	1	VEC-MSF	0.4127	1.68
2	3	0	VEC-MSF	0.4168	2.67
2	3	2	VEC-MSF-SBEKK	0.4191	3.25
2	3	1	VEC-MSF-SBEKK	0.4210	3.72
3	4	2	VEC-MSF	0.4217	3.88
4	3	0	VEC-MSF	0.4230	4.20
3	3	0	VEC-MSF	0.4231	4.23
4	3	0	VEC-MSF-SBEKK	0.4252	4.75
2	3	0	VEC-MSF-SBEKK	0.4266	5.10
4	3	2	VEC-MSF	0.4270	5.20
4	4	2	VEC-MSF-SBEKK	0.4271	5.22
4	3	0	VEC-SBEKK	0.4274	5.30
⋮	⋮	⋮	⋮	⋮	⋮
2	4	1	VEC	0.4762	17.30

Source: own calculations based on 100 000 accepted draws

Following Geweke and Amisano (2010) we present (Figure 2) the cumulative log predictive Bayes factors in favour of the $M_{(2,4,2,VEC-MSF-SBKK)}$ model over three models: $M_{(2,4,2,VEC-MSF)}$, which is ranked on the third place according to the energy score, $M_{(2,3,0,VEC-MSF-SBEKK)}$, which is the winner of the Bayesian model comparison within the modelled time and $M_{(2,4,2,VEC)}$, so the one with the highest log predictive likelihood among all models with the non-changing error covariance matrix.

Figure 2: Cumulative log predictive Bayes factors in favour of the $M_{(2,4,2,VEC-MSF-SBKK)}$ model



It can be noticed that the superiority of the $M_{(2,4,2,VEC-MSF-SBKK)}$ model over the model with constant error covariances raises quickly and almost steadily in the whole forecast period, whereas although at the endpoint the data weakly favour the $M_{(2,4,2,VEC-MSF-SBKK)}$ model over $M_{(2,4,2,VEC-MSF)}$, at some data-points during the evaluation period (especially at the beginning) the log predictive density of the letter model is higher. The predominance of $M_{(2,4,2,VEC-MSF-SBKK)}$ model over the model with the same covariance structure, but without long-run relationships also rises, but not-constantly and the speed of these changes is much lower than that one observed in the case of the $M_{(2,4,2,VEC)}$ model. The presented discussion of the observed time changes in the cumulative log predictive Bayes factors also confirms the importance of the changing covariance structure for the quality of the predictive distributions.

The results in Table 4 indicate that models with time-changing conditional covariances yield point forecasts that are more accurate than forecasts from models with constant conditional covariances. In general, WTMSFE of point forecast obtained within models with the MSF structure is the lowest, whereas the mean of WTMSFE of forecast from the VEC models without stochastic volatility is 19.5% higher than WTMSFE of forecast from the VEC-MSF models. For forecast from the VEC-MSF-SBEKK and the VEC-SBEKK models the means of WTMSFE are, accordingly, 3% and 4.5% higher. These results reinforce the conclusion of the importance of SV structures for the forecast ability of the analysed models.

Tables 2, 3 and 4 display almost the same models, but with a slightly different order. The main difference concerns the importance of the number of cointegrating relations. It turned out that for the point forecast accuracy the exact number of long-run relations is less important than for the accuracy of the density forecast. However, it is important to emphasize that according to the results presented in tables 2 through 4 the VAR models built for the levels of the analysed series performed the worst.

Table 4: The weighted trace mean squared forecast error (mean in the period 2013Q1 – 2016Q4) and the percentage loss to the best model, $pl_{WTMSFE} =$
$= 100\left(\overline{WTMSFE}_{T+1}^{T+16}\left(M_{(k,d,r,type)}\right)/WTMSFE_{T+1}^{T+16}\left(M_{(2,3,1,VEC-MSF)}\right)-1\right)$

VAR order (k)	determin. term (d)	cointegration rank (r)	model	$\overline{WTMSFE}_{T+1}^{T+16}(M_{\cdot})$	pl_{WTMSFE}
2	3	1	VEC-MSF	0.6198	0.000
2	3	0	VEC-MSF	0.6258	0.962
2	4	2	VEC-MSF	0.6344	2.356
2	4	1	VEC-SBEKK	0.6364	2.681
4	3	0	VEC-SBEKK	0.6385	3.016
2	4	1	VEC-MSF-SBEKK	0.6503	4.908
3	4	1	VEC-MSF-SBEKK	0.6516	5.119
2	3	2	VEC-MSF	0.6563	5.882
2	4	2	VEC-SBEKK	0.6608	6.615
2	3	0	VEC-MSF-SBEKK	0.6613	6.698
2	4	1	VEC-MSF	0.6632	4.334
3	3	1	VEC-MSF	0.6634	4.414
2	4	2	VEC-MSF-SBEKK	0.6643	4.467
4	3	0	VEC-MSF-SBEKK	0.6670	4.791
2	3	0	VEC-SBEKK	0.6673	4.973
3	3	1	VEC-SBEKK	0.6792	9.579
⋮	⋮	⋮	⋮	⋮	⋮
2	3	0	VEC	0.7389	19.208

Source: own calculations based on 100 000 accepted draws

Using the log predictive likelihood and the energy sore we have evaluated the three-dimensional forecast density, now we are going to examine the one-dimensional forecast densities – for each variable separately. We employ the continuous rank probability score and the probability integral transform.

Table 5 displays ten models with the lowest CRPS for the analysed variables and the percentage loss to the best model.

The inclusion of stochastic volatility and long-run relationships improves density forecast accuracy of unemployment rate and interest rate. In the case of unemployment rate MSF seems to be sufficient structure of a changing covariance matrix, whereas allowing for time-varying conditional correlations is essential for the forecasting performance of the compared models in the case of interest rate. The exact number of cointegrating relations is less important, but on average for density forecast of both the unemployment rate and the interest rate models with two cointegrating vectors perform better than those with only one vector. It is interesting that in this forecast exercise the models without the latent process and without long-run relationships generally obtained lower CRPS values.

Table 5: The continuous rank probability score (mean in the period 2013Q1 – 2016Q4) and the percentage loss to the best model $M^{B}_{(k,d,r,type)}$, $pl_{CRPS} = 100\left(\overline{CRPS}^{T+16}_{T+1}\left(M_{(k,d,r,type)}\right)/\overline{CRPS}^{T+16}_{T+1}\left(M^{B}_{(k,d,r,type)}\right) - 1\right)$

	VAR order (k)	determin. term (d)	cointegration rank (r)	model	$\overline{CRPS}^{T+16}_{T+1}\left(M_{(k,d,r,type)}\right)$	pl_{CRPS}
inflation	4	3	0	VEC-SBEKK	0.3226	0.00
	2	3	0	VEC	0.3272	1.45
	4	3	0	VEC-MSF-SBEKK	0.3314	2.73
	2	3	2	VEC-MSF	0.3319	2.90
	2	4	2	VEC-MSF-SBEKK	0.3321	2.96
	3	3	0	VEC-MSF	0.3322	3.00
	2	3	0	VEC-MSF	0.3325	3.08
	4	3	0	VEC	0.3332	3.30
	2	4	1	VEC	0.3333	3.34
	2	4	1	VEC-MSF-SBEKK	0.3335	3.39
unemployment rate	2	4	2	VEC-MSF	0.1161	0.00
	4	3	2	VEC-MSF	0.1163	0.20
	2	3	2	VEC-MSF	0.1177	1.42
	2	3	1	VEC-MSF	0.1178	1.44
	4	4	1	VEC-MSF	0.1189	2.38
	2	3	2	VEC-MSF-SBEKK	0.1211	4.30
	2	4	1	VEC-MSF	0.1212	4.39
	4	4	2	VEC-MSF-SBEKK	0.1213	4.46
	4	4	2	VEC-MSF	0.1221	5.19
	4	3	2	VEC-MSF-SBEKK	0.1231	6.05
interest rate	2	4	2	VEC-MSF-SBEKK	0.1183	0.00
	2	4	1	VEC-MSF-SBEKK	0.1187	0.37
	2	3	2	VEC-MSF-SBEKK	0.1251	5.77
	2	3	0	VEC-MSF-SBEKK	0.1256	6.20
	3	4	1	VEC-MSF-SBEKK	0.1267	7.12
	3	4	2	VEC-MSF-SBEKK	0.1269	7.29
	2	3	1	VEC-SBEKK	0.1273	7.67
	4	4	2	VEC-MSF-SBEKK	0.1280	8.25
	2	4	2	VEC-SBEKK	0.1282	8.41
	2	3	1	VEC-MSF-SBEKK	0.1283	8.50

Source: own calculations based on 100 000 accepted draws

Generally models with lower lag length perform better, whereas the type of the deterministic term is negligible.

Figures 3 through 5 display histograms of PIT values obtained in models which are ranked first according to Bayesian model comparison within the modelled sample, log predictive density score, ES and CRPS.

Figure 3: Probability integral transform – **inflation**, p-values are calculated with the use of R procedures (ks.test in the package “stats” and ad.test in the package “goftest”)

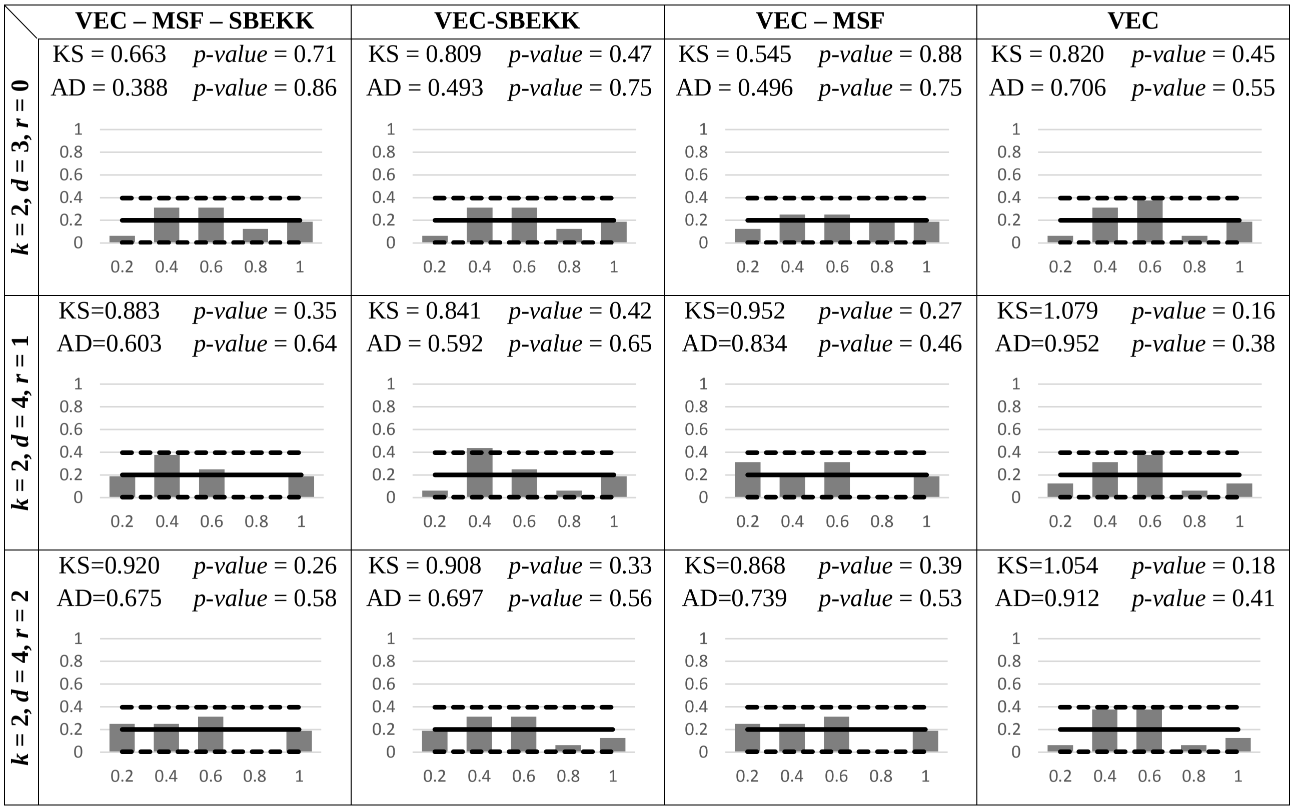


Figure 4: Probability integral transform – **unemployment rate**, *p*-values are calculated with the use of R procedures (ks.test in the package “stats” and ad.test in the package “goftest”)

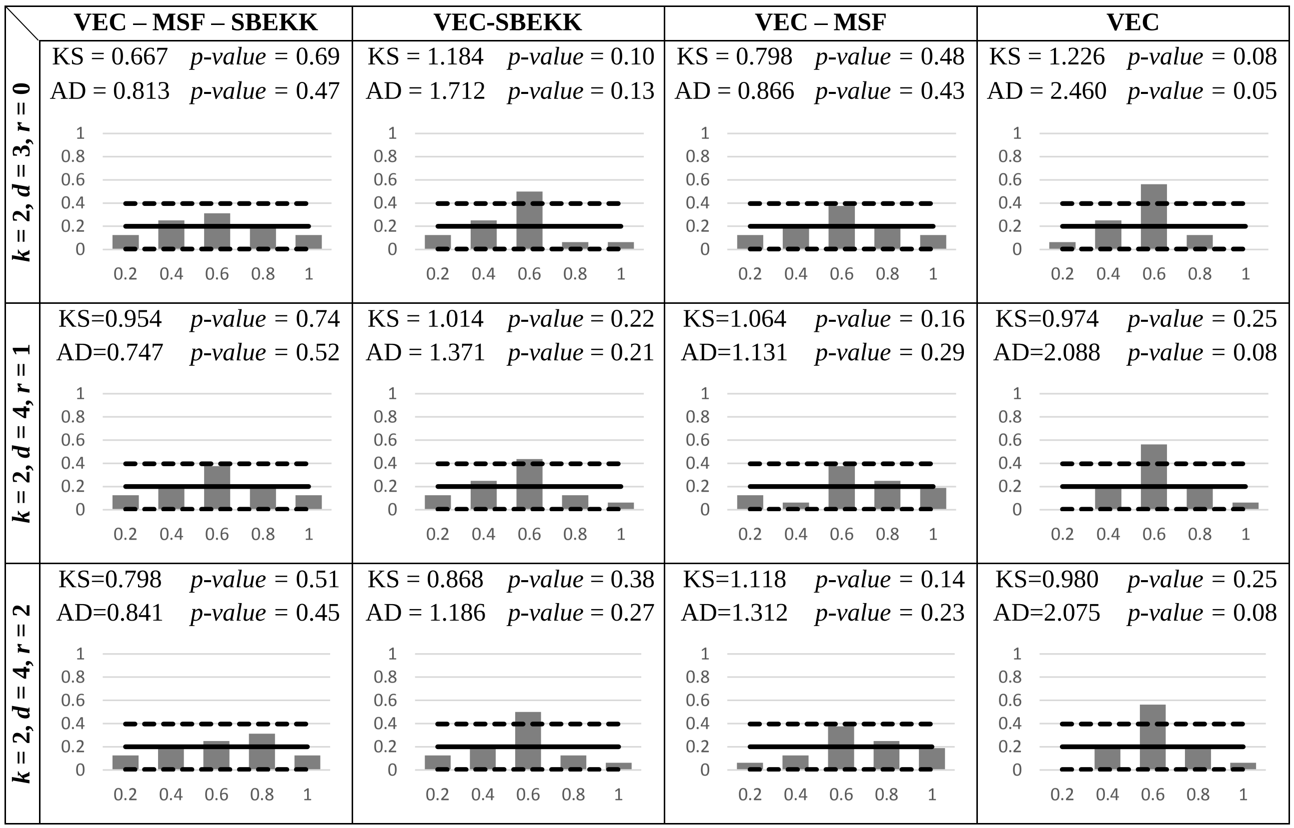
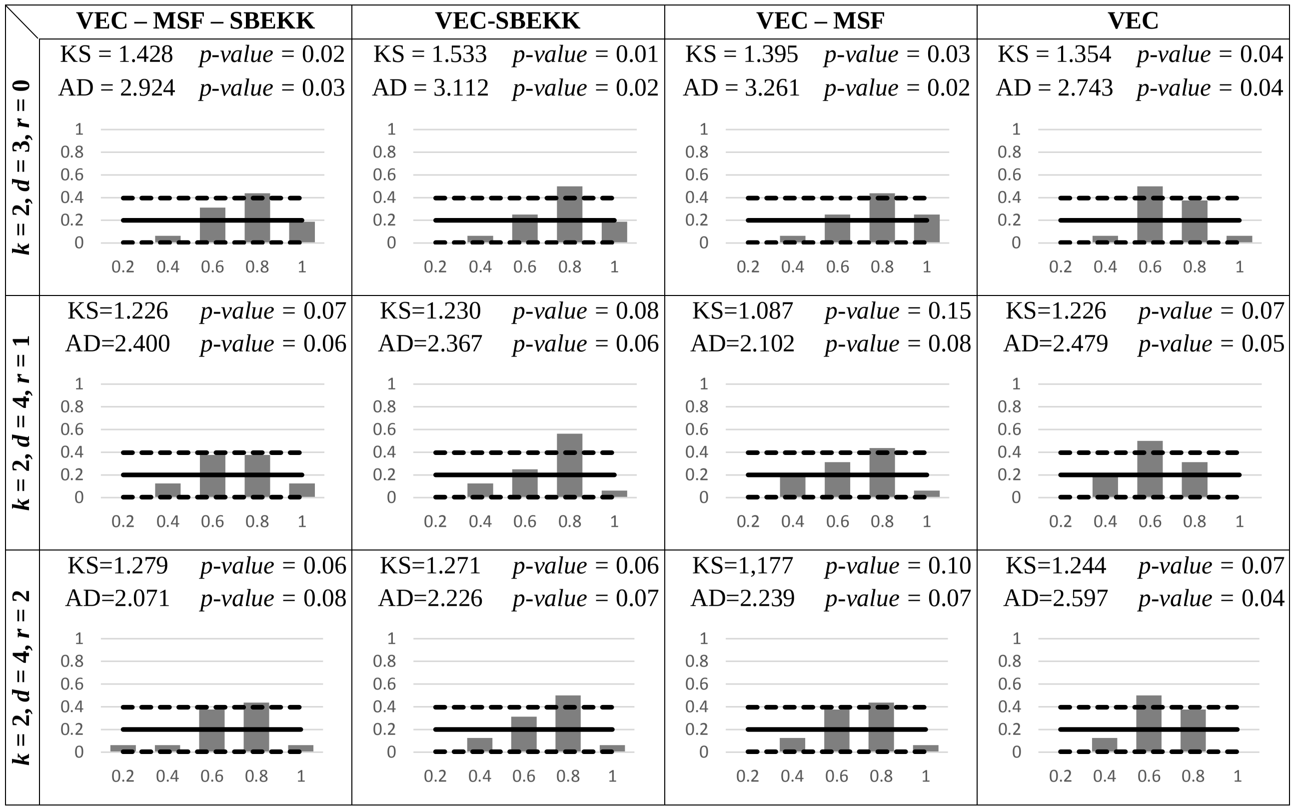


Figure 5: Probability integral transform – **interest rate**, p-values are calculated with the use of R procedures (ks.test in the package “stats” and ad.test in the package “goftest”)



Altogether we take a closer look on the results obtained within models with two lags which differ in the number of long-run relations ($r = 0$, $r = 1$, $r = 2$) with a constant restricted to the cointegrating relations and four considered types of covariance structures. PITs enable us to assess whether these winning models calibrate the predictive densities correctly. We follow Rossi and Sekhposyan (2014) and divide the unit interval into 5 equally sized bins and show the fraction of the PITs which follow into each bin. In the ideal case of PITs uniformity each bin should contain $\hat{p} = 0.2$ of the obtained values, which is depicted by the solid lines. The broken lines represent the 2.5^{th} and 97.5^{th} percentiles of the distribution of $\hat{p}$. They are constructed with the use of a normal approximation ($\hat{p}\pm 1.96\sqrt{\hat{p}\,(1-\hat{p})/16}$). Additionally, Figures 3 through 5 report results for the Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) tests. Table 6 displays p-value for autocorrelation test of PIT and PIT^2.

Table 6: p-value for autocorrelation test of PIT and PIT^2 for selected models (lags: 4)

model		inflation		unemployment		interest	
		PIT	PIT^2	PIT	PIT^2	PIT	PIT^2
$k = 2,\ d = 3,\ r = 0$	VEC-MSF-SBEKK	0.87	0.72	0.007	0.001	0.51	0.25
	VEC-SBEKK	0.77	0.62	0.004	0.001	0.54	0.40
	VEC-MSF	0.61	0.55	0.01	0.001	0.58	0.36
	VEC	0.72	0.59	0.02	0.01	0.01	0.01
$k = 2,\ d = 4,\ r = 1$	VEC-MSF-SBEKK	0.73	0.61	0.005	0.001	0.43	0.20
	VEC-SBEKK	0.72	0.59	0.002	0.0002	0.48	0.33
	VEC-MSF	0.57	0.50	0.01	0.002	0.53	0.31
	VEC	0.73	0.61	0.01	0.004	0.006	0.005
$k = 2,\ d = 4,\ r = 2$	VEC-MSF-SBEKK	0.77	0.68	0.004	0.001	0.55	0.25
	VEC-SBEKK	0.81	0.65	0.001	0.0002	0.46	0.26
	VEC-MSF	0.54	0.48	0.01	0.002	0.56	0.34
	VEC	0.71	0.60	0.01	0.004	0.006	0.005

Source: own calculations

In the case of inflation all the PIT histograms are essentially uniform, which is formally confirmed by the results of both tests – KS and AD. The hypotheses of non-autocorrelation of PIT and PIT^2 cannot be rejected for all 9 analysed models. The PIT histograms for forecasts obtained within models $M_{(2,3,0,VEC-MSF-SBKK)}$ and $M_{(2,3,0,VEC-MSF)}$ are almost equal to the histogram of uniform distribution.
For unemployment rate, all the considered models with a stable covariance matrix give significantly over-dispersed forecast densities. Moreover, there is a significant autocorrelation in PIT and PIT^2, so the PITs cannot be traded as realisations of independent distributions. This results indicate that none of the obtained forecast densities is correctly specified.
Turning to the interest rate, the outcomes of the Ljung-Box test suggest that the forecast densities from models with a constant covariance matrix are not correctly

www.czasopisma.pan.pl www.journals.pan.pl

specified. The PIT histograms and the KS and AD tests indicate that in models without long-run relationships forecasts are in general over-dispersed and placed too much weight to lower values of the interest rate. The distance between the PIT histogram for the model $M_{(2,4,1,VEC-MSF-SBEKK)}$ and the density of uniform distribution is the smallest. Note that is the model ranked second according to CRPS. The observed over-dispersion on forecast densities obtained within models with a non-changing covariance structure is not surprising as in such case models try to encompass changing variances by dispersing the distributions.

6 Concluding remarks

In this paper we have compared from the forecasting perspective the set of VAR possibly cointegrated models with changing conditional covariances, built for inflation, unemployment rate and interest rate from the Polish economy. With the use of log predictive density score, continuous rank probability score and energy score the predictive densities have been evaluated and the models have been ranked. The weighted trace mean squared forecast error has been employed to asses point forecast accuracy. According to these measures the most important feature for the predictive ability of the model is allowing for changing covariance matrices, the number of long-run relationships is less important. Besides, the models with lower lag order perform better. These outcomes are in accordance with the results of within-sample Bayesian model comparison performed by Pajor and Wróblewska (2017).

Additionally, the probability integral transform has been employed to assess the calibration of the forecast densities. According to the results, the forecast density for inflation is well calibrated contrary to not well calibration of the unemployment's forecast densities, whereas in the case of interest rate the forecast densities obtained within models with long-run relationships and changing in time conditional correlations are quite well calibrated. There is a visible cyclical pattern in the unemployment series, in which case the structure of VAR/VEC models is probably not sufficient to include this feature.

We conclude and emphasise that, from the macroeconomic forecasting perspective, taking into account the possibility of time-changing conditional covariances improves both point and density forecast ability of the vector autoregression models both with and without cointegrating relations. For the point forecast accuracy, the SV structures are much more important than the proper assessment of the exact number of long-term relationships. For the density forecast ability, the impact of the SV structures is also visible, but generally models with two cointegrating relations perform better than the others. In the small monetary model we expect one or two long-run relations, and their inclusion improves the quality of forecasts (even for one horizon), because through adjusting to them the series corrects their trajectories at every single point of time.

Acknowledgments

We would like to thank the anonymous referees for useful comments and suggestions that helped us to prepare the final version of the paper. All the remaining errors are ours.
We acknowledge support from research funds granted to the Faculty of Management (the first author) and the faculty of Finance and Law (the second author) at Cracow University of Economics, within the framework of the subsidy for the maintenance of research potential.

References

[1] Abbate A., Marcellino M., (2018), Point, interval and density forecast of exchange rates with time varying parameter models, *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 181, 155–179.

[2] Berg T.O., Henzel S.R., (2015), Point and density forecasts for the euro area using Bayesian VARs, *International Journal of Forecasting* 31, 1067–1095.

[3] Berg T. O., (2017), Forecast accuracy of a BVAR under alternative specifications of the zero lower bound, *Studies in Nonlinear Dynamics and Econometrics* 21, 1-29.

[4] Carriero A., Kapetanios G., Marcellino M., (2011), Forecasting large datasets with Bayesian reduced rank multivariate models, *Journal of Applied Econometrics* 26, 735–761.

[5] Chikuse Y., (2002), Statistics on special manifolds. *Lecture Notes in Statistics* 174, Springer-Verlag, New York.

[6] Clark T. E., Ravazzolo F., (2015), Macroeconomic forecasting performance under alternative specifications of time-varying volatility, *Journal of Applied Econometrics* 30, 551–575.

[7] Geweke J., Amisano G., (2010), Comparing and evaluating Bayesian predictive distributions of asset returns, *International Journal of Forecasting* 26, 216–230.

[8] Gneiting T., Balabdaoui F., Raftery A.E., (2007), Probabilistic forecasts, calibration and sharpness, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 69, 243–268.

[9] Gneiting T., Raftery A.E., (2007), Strictly proper scoring rules, prediction, and estimation, *Journal of the American Statistical Association* 102, 359–378.

[10] Hersbach H., (2000), Decomposition of the continuous ranked probability score for ensemble prediction systems, *Weather and Forecasting* 15, 559–570.

www.czasopisma.pan.pl PAN POLSKA AKADEMIA NAUK www.journals.pan.pl

[11] Kass R. E., Raftery A. E., (1995), Bayes Factors, *Journal of the American Statistical Association* 90, 773–795.

[12] 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.

[13] Mitchell J., Wallis K.F., (2011), Evaluating density forecasts: forecast combinations, model mixtures, calibration and sharpness, *Journal of Applied Econometrics* 26, 1023–1040.

[14] Osiewalski J., (2009), New Hybrid Models of Multivariate Volatility (a Bayesian Perspective), *Przegląd Statystyczny* (*Statistical Review*) 56, 15–22.

[15] Osiewalski J., Pajor A., (2009), Bayesian analysis for hybrid MSF-SBEKK models of multivariate volatility, *Central European Journal of Economic Modelling and Econometrics* 1, 179–202.

[16] 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, B-101-B-109.

[17] Pajor A., Wróblewska J., (2017), VEC-MSF models in Bayesian analysis of short- and long-run relationships, *Studies in Nonlinear Dynamics and Econometrics* 21, 1–22.

[18] Primiceri G. E., (2005), Time Varying Structural Vector Autoregressions and Monetary Policy, *Review of Economic Studies* 72, 821–852.

[19] Rossi B., Sekhposyan T., (2014), Evaluating predictive densities of US output growth and inflation in a large macroeconomic data set, *International Journal of Forecasting* 30, 662–682.

[20] Stelmasiak D., Szafrański G., (2016), Forecasting the Polish inflation using Bayesian VAR models with seasonality, *Central European Journal of Economic Modelling and Econometrics* 8, 21–42.

[21] Weiss A.A., (1996), Estimating time series models using the relevant cost function, *Journal of Applied Econometrics* 11, 539–560.