Bayesian Analysis (2024) **19**, Number 3, pp. 751–783

Bayesian ex Post Evaluation of Recursive Multi-Step-Ahead Density Prediction*

Anna Pajor†, Jacek Osiewalski‡, Justyna Wróblewska‡, and Łukasz Kwiatkowski‡

Abstract. This research is focused on a formal Bayesian method of recursive multi-step-ahead density prediction and its ex post evaluation. Our approach remains within the framework of the standard (classical or orthodox) Bayesian paradigm based on the Bayes factor and on the use of the likelihood-based update. We propose a new decomposition of the predictive Bayes factor into the product of partial Bayes factors, for both a finite number of consecutive k-step-ahead forecasts (where $k > 1$) and the recursive updates of the posterior odds ratios based on updated data sets. The first factor in the decomposition is related to the relative k-step-ahead forecasting ability of models, while the second one measures the updating effect.

To illustrate the usefulness of the proposed measures, we apply the new decomposed predictive Bayes factors to compare the forecasting ability of models when the true data generating process (DGP) is known, using simulated data sets. Taking into account the effect of updating, the posterior odds ratios leads to the conclusion that the best model coincides with the true DGP. However, the highest k-step-ahead forecasting ability (considered alone) can be achieved by some other, less adequate models. Next, we investigate the predictive ability of different Vector Error Correction (VEC) models with conditional heteroscedasticity, combining three macroeconomic variables: unemployment, inflation and interest rates, separately for the US and Polish economies. The results show that the inference about the models' predictive performance depends on the forecast horizon as well as on taking into account the updating effect.

MSC2020 subject classifications: 62F15, 60G25, 62M20.

Keywords: predictive Bayes factor, density forecast, cointegration, time-varying volatility.

1 Introduction

The predictive capacity of econometric models can be evaluated and compared in several ways. Many accuracy measures of prediction, both Bayesian and non-Bayesian, have

*Research supported by a grant from the National Science Center (NCN, Poland) under decision no. UMO-2018/31/B/HS4/00730. Jacek Osiewalski's contribution is financed from a subsidy granted to Krakow University of Economics – Project nr 080/EIE/2022/POT.

†Department of Mathematics, Krakow University of Economics, ul. Rakowicka 27, 31-510 Kraków, Poland; and Institute of Mathematics, Jagiellonian University, ul. Prof. Stanisława Łojasiewicza 6, 30-348 Kraków, Poland, pajora@uek.krakow.pl

‡Department of Econometrics and Operations Research, Krakow University of Economics, ul. Rakowicka 27, 31-510 Kraków, Poland, eeosiewa@cyf-kr.edu.pl; wroblewj@uek.krakow.pl; kwiatkol@uek.krakow.pl

 https://doi.org/10.1214/23-BA1367

been proposed and discussed in the literature. Reviews of Bayesian predictive model assessment and related concepts can be found in, e.g., Vehtari and Ojanen (2012), Vehtari and Ojanen (2014), Piironen and Vehtari (2017). A very popular approach to evaluation of forecasting ability hinges on the so-called scoring rules, proposed by Good (1952) and Bernardo and Smith (1994), and discussed in numerous papers, e.g., Winkler and Murphy (1968), Murphy and Winkler (1970), Winkler (1996), Gneiting and Raftery (2007), Carvalho (2016). Some of the most common scoring rules include: the quadratic score, the logarithmic score, the continuous ranked probability score (CRPS), the energy score (ES), as well as scoring rules depending on the first and second moments (see, e.g., Dawid and Sebastiani, 1999; Gneiting and Raftery, 2007; Yao et al., 2018).

In this paper, we focus on a formal Bayesian recursive multi-step-ahead density prediction and its ex post evaluation. Our approach remains within the framework of the standard (classical, orthodox) Bayesian paradigm based on the Bayes factor and the use of the likelihood-based update – an idea introduced in Jeffreys (1961). Note that also the concept of non-likelihood-based update has been entertained in the literature, see, e.g., Bissiri et al. (2016) and Loaiza-Maya et al. (2021), where the likelihood is replaced with the exponential of a problem-specific loss function.

The classical Bayesian approach to time series modelling makes it possible to formally and coherently compare the quality of predictive distributions for multiple horizons. The comparison is based on the so-called predictive likelihood function, i.e. the probability density function of the predictive distribution of future (unobserved) data conditional on the observed data, evaluated at the future values after they are observed (see Geweke, 2005; Geweke and Whiteman, 2006). There is a relative lack of fully, inherently Bayesian tools for comparing the predictive ability of models, based on a finite number of consecutive k-step-ahead forecasts with $k > 1$. In practice, to that end, the average difference between the log predictive scores (or, the log predictive likelihoods) of competing models is employed (see, e.g., Geweke and Amisano, 2010; Koop and Korobilis, 2012; Warne et al., 2013; Giannone et al., 2015; Cross et al., 2020). While the literature on density forecasting is abundant, it appears that, to the best of our knowledge, a fully classical Bayesian approach to evaluation and models' multi-step-ahead predictive performance comparison has not been developed yet. We aim at filling this gap and propose a new measure, based on the decomposition of the predictive Bayes factor, for comparing the forecasting performance of Bayesian models for both n consecutive k-step-ahead forecasts and updating posterior odds ratios. The first factor in the decomposition is related to the relative k-step-ahead forecasting ability of models, while the second one measures the updating effect. Jointly, the proposed predictive Bayes factor of order $(k,\, s)$ at time T shows how the posterior odds ratio depends (on average) on the updated observations.

The paper is organized as follows. In Section 2, scoring rules in the context of the Bayesian approach are briefly discussed. Section 3 outlines the Bayesian principles of model comparison and predictive power relative assessment, within the framework of which we propose a new decomposition of the predictive Bayes factor for comparing the forecasting performance of models. Section 4 presents simulation studies. In Section 5, to illustrate the proposed measure, we use data for the US and, separately, Polish three macroeconomic variables: the unemployment, inflation and interest rate, modelled jointly within the conditionally heteroscedastic vector error correction (VEC) models. Section 6 concludes.

2 A short note on scoring rules in the Bayesian approach

In the Bayesian approach, most inferences and predictions are formulated in probabilistic terms. To be able to measure how adequate probabilistic forecasts are, some manners of their evaluation are needed. The measures frequently used for such evaluations include strictly proper scoring rules. As noted by Winkler (1996): "In an ex ante sense, strictly proper scoring rules provide an incentive for careful and honest forecasting by the forecaster or forecast system. In an ex post sense, they reward accurate forecasts and penalize inferior forecasts." A broad discussion of the scoring rules can be found, e.g., in (Bernardo and Smith, 1994, Section 2.7). The best known proper scoring rules are the quadratic, logarithmic, and spherical rules (see, e.g., Brier, 1950; Good, 1952; Winkler, 1996; Savage, 1971). Bernardo and Smith (1994) have shown that only the logarithmic score function is smooth (continuously differentiable), proper (the maximum value is attained for the true predictive distribution), and local (the value of the function at a particular event depends only on the probability assigned to that event, and it is independent of probabilities of the remaining events). Thus, if one is interested in choosing such a scoring rule for assessing probabilistic forecasts that is smooth, proper and local, then the choice can be restricted to the logarithmic rule.

Scoring rules provide measures for the evaluation of probabilistic forecasts by assigning a numerical score based on the forecast and on the realised observation (see Gneiting and Raftery, 2007). From the Bayesian perspective, scores can be referred to as utilities – in fact, the expected utility of a predictive distribution can be maximised over the space of competing models (see Bernardo and Smith, 1994). The standard (and classical) Bayesian approach to comparing two models uses the Bayes factor, the logarithm of which can be interpreted as the difference between the two models' log scores. A survey of Bayesian predictive methods for model assessment, selection and comparison can be found in Vehtari and Ojanen (2012, 2014).

3 Bayesian predictive model assessment

Let $y_1^{T+k+n} = [y_1 \ \ldots \ y_{T+k+n}]$ be the matrix of ordered observables, θ_i be the vector of unknown parameters, and $h_{i,1}^{T+k+n} = [h_{i,1} \ \ldots \ h_{i,T+k+n}]$ be the matrix of latent variables in model M_i, $i = 1, \ldots, m$. Moreover, let $M_1, \ldots, M_m$ be a set of mutually exclusive (non-nested) and jointly exhaustive models, with a prior probability for each model, $p(M_i)$.[1] The Bayesian model M_i is defined by the joint distribution of the observables,

[1] The assumptions of mutually exclusive and jointly exhaustive models ensure that the models $M_1, \ldots, M_m$ form a complete set of non-nested alternative specifications, which is required to introduce a probability measure on the model space. Then, $p(\{M_1, \ldots, M_m\}) = p(M_1) + \cdots + p(M_m) = 1$.

parameters (θ_i) and latent variables ($h_{i,1}^{T+k+n}$):

$$p\left(y_1^{T+k+n}, h_{i,1}^{T+k+n}, \theta_i | M_i\right) = p\left(y_1^{T+k+n} | h_{i,1}^{T+k+n},\, \theta_i,\, M_i\right) p\left(h_{i,1}^{T+k+n} | \theta_i, M_i\right) p\left(\theta_i | M_i\right),$$

where $p(y_1^{T+k+n}|h_{i,1}^{T+k+n},\, \theta_i, M_i)$ is the conditional sampling density for the observables y_1^{T+k+n}, $p(h_{i,1}^{T+k+n}, \theta_i|M_i) = p(h_{i,1}^{T+k+n}|\theta_i, M_i)\, p(\theta_i|M_i)$ is the prior distribution of the parameters and latent variables, $p(h_{i,1}^{T+k+n}|\theta_i,\, M_i)$ is the density of the prior distribution of the latent variables, given the vector of parameters, and finally, $p(\theta_i|M_i)$ is the prior density for θ_i. Initial conditions are omitted in our notations.[2] Notice that the observables are the same in the models, but the unobservables need not to be the same.

In order to consider the predictive performance of the models, we split the matrices $h_{i,1}^{T+k+n}$ and y_1^{T+k+n} into two parts with T and $n+k$ elements, respectively: $h_{i,1}^{T+k+n} = [h_{i,1}^{T}\ h_{i,T+1}^{T+k+n}]$ and $y_1^{T+k+n} = [y_1^T\ y_{T+1}^{T+k+n}]$, where y_1^T is observed and y_{T+1}^{T+k+n} is unobserved (to be forecasted). Our purpose is to compare the predictive ability of the models in the period $T+1, \ldots, T+k+n$. Thus, the initial observations (the data) from $t = 1$ to $t = T$, collected in the matrix $y_1^{T,o}$, can be treated as a "training sample".

The posterior probability of the model M_i after observing the data, $y_1^{T,o}$, follows from the Bayes rule:

$$p\left(M_i | y_1^{T,o}\right) = \frac{p\left(y_1^{T,o} | M_i\right) p\left(M_i\right)}{\sum_{j=1}^{m} p\left(y_1^{T,o} | M_j\right) p\left(M_j\right)}, \tag{3.1}$$

where $p(M_i)$ is the prior probability of the model M_i, whereas

$$p\left(y_1^{T,o} | M_i\right) = \int p\left(y_1^{T,o} | h_{i,1}^{T}, \theta_i, M_i\right) p\left(h_{i,1}^{T}, \theta_i | M_i\right)\, dh_{i,1}^{T}\, d\theta_i$$

is the marginal likelihood (the marginal data density value at $y_1^{T,o}$) for the model M_i. The marginal likelihood of the model M_i is the measure of how well the model predicted the data $y_1^{T,o}$. The posterior probabilities of the models under consideration provide a formal basis for choosing the best specification from among $M_1, \ldots, M_m$, as well as for weighted averaging of the competing models, commonly referred to as Bayesian model averaging or Bayesian pooling approach.

The main criterion of comparison between the two models M_i and M_j is their posterior odds ratio:

$$\frac{p\left(M_i | y_1^{T,o}\right)}{p\left(M_j | y_1^{T,o}\right)} = \frac{p\left(y_1^{T,o} | M_i\right)}{p\left(y_1^{T,o} | M_j\right)} \frac{p\left(M_i\right)}{p\left(M_j\right)}. \tag{3.2}$$

Ultimately, it enables the calculation of the posterior probabilities of the models, $p(M_i|y)$, according to the formula given in (3.1).

[2] Initial conditions can be included into the parameter vector, and be subject to estimation. In the empirical part of the paper, some of them are assumed to be fixed, and thus not estimated, while the others constitute additional model parameters.

The ratio $B_{i,j}(y_1^{T,o}) = \frac{p(y_1^{T,o}|M_i)}{p(y_1^{T,o}|M_j)}$ is referred to as the Bayes factor in favour of the model M_i versus the model M_j, while $\frac{p(M_i)}{p(M_j)}$ is the prior odds ratio (which may depend on initial conditions). The prior odds ratio is the degree to which one's beliefs favour the model M_i over the model M_j (see, e.g., Morey et al., 2016). In turn, the posterior odds ratio informs one about the degree to which the model M_i is favoured over the model M_j after having observed the data. In other words, the posterior odds ratio indicates how many times the model M_i is more probable than the model M_j after having observed the data and under the given prior beliefs. If the prior odds ratio is equal to one (i.e. both models are equally probable *a priori*, $p(M_i) = p(M_j)$), then the posterior odds ratio between the two models coincides with the Bayes factor, $B_{i,j}(y_1^{T,o})$, which measures the relative within-sample predictive power of M_i and M_j (see, e.g., Osiewalski and Steel, 1993). The Bayes factor shows how the observations $y_1^{T,o}$ contribute to the evidence in favour of the model M_i over the model M_j. Let us point to the fact that $B_{i,j}(y_1^{T,o})$ can be expressed as the ratio of the posterior and prior odds ratios. Thus, it indicates the change in belief for the ratio of the model probabilities due to observing the data $y_1^{T,o}$.

Now, let us start with the distribution of y_{T+1}^{T+k+n} and latent variables $h_{i,T+1}^{T+k+n}$ conditional on y_1^T, $h_{i,1}^T, \theta_i$ in the model M_i:

$$\begin{aligned} p\left(y_{T+1}^{T+k+n},\, h_{i,T+1}^{T+k+n}|y_1^T, h_{i,1}^T, \theta_i, M_i\right) = p\left(y_{T+1}^{T+k+n}|h_{i,T+1}^{T+k+n}, y_1^T, h_{i,1}^T\theta_i, M_i\right) \times \\ \times\, p\left(h_{i,T+1}^{T+k+n}|y_1^T, h_{i,1}^T, \theta_i, M_i\right). \end{aligned} \tag{3.3}$$

The predictive density of y_{T+1}^{T+k+n} conditional on $y_1^{T,o}$ in the model M_i amounts to the following integral:

$$\begin{aligned} p\left(y_{T+1}^{T+k+n}|y_1^{T,o}, M_i\right) = \int p\left(y_{T+1}^{T+k+n}, h_{i,T+1}^{T+k+n}|y_1^{T,o}, h_{i,1}^T, \theta_i, M_i\right) \times \\ \times\, p\left(h_{i,1}^T, \theta_i|y_1^{T,o}, M_i\right)\, dh_{i,T+1}^{T+k+n}\, dh_{i,1}^T\, d\theta_i. \end{aligned} \tag{3.4}$$

Thus, the predictive density is obtained by integrating out the parameters and latent variables from the joint density function. This predictive distribution describes our beliefs about the future observation given the observed data $y_1^{T,o}$. Once y_{T+1}^{T+k+n} is known, one can evaluate $p(y_{T+1}^{T+k+n}|y_1^{T,o}, M_i)$ at the observed values, $y_{T+1}^{T+k+n,o}$, obtaining some real number: $p(y_{T+1}^{T+k+n,o}|y_1^{T,o}, M_i)$, which is called the predictive likelihood, see Geweke (2005). In other words, the predictive likelihood is the predictive density for y_{T+1}^{T+k+n} (given data up to time T) evaluated at the observed data. In order to assess how well alternative models predict the same set of observations, simply their predictive likelihoods are compared.

Notice that the posterior odds ratio given $y_1^{T+k+n,o}$ can be expressed as the product of the ratio of the predictive likelihoods (the predictive Bayes factor) and the posterior odds ratio given the initially observed data $y_1^{T,o}$:

$$\frac{p\left(M_i|y_{T+1}^{T+k+n,o}, y_1^{T,o}\right)}{p\left(M_j|y_{T+1}^{T+k+n,o}, y_1^{T,o}\right)} = \frac{p\left(y_{T+1}^{T+k+n,o}|y_1^{T,o},\, M_i\right)}{p\left(y_{T+1}^{T+k+n,o}|y_1^{T,o}, M_j\right)} \frac{p\left(M_i|y_1^{T,o}\right)}{p\left(M_j|y_1^{T,o}\right)}. \tag{3.5}$$

Thus, the predictive Bayes factor for the period from $T+1$ to $T+k+n$,

$$B_{i,j}\left(y_{T+1}^{T+k+n,o}|y_1^{T,o}\right) = \frac{p\left(y_{T+1}^{T+k+n,o}|y_1^{T,o},\, M_i\right)}{p\left(y_{T+1}^{T+k+n,o}|y_1^{T,o}, M_j\right)}, \tag{3.6}$$

shows how the observations collected in $y_{T+1}^{T+k+n,o}$ contribute to the evidence in favour of the model M_i over the model M_j. The predictive Bayes factor over the period $T+1$ to $T+k+n$ updates the ratio of the posterior probabilities based on the first T observations after having observed the predicted data y_{T+1}^{T+k+n}. In other words, it determines how the posterior beliefs change after having observed the data $y_{T+1}^{T+k+n,o}$. Therefore, $B_{i,j}(y_{T+1}^{T+k+n,o}|y_1^{T,o})$ indicates how much the additional data, $y_{T+1}^{T+k+n,o}$, changes the odds of the two models. Note that this predictive Bayes factor can be expressed as a part of the "full" Bayes factor (see, e.g., O'Hagan, 1995):

$$\begin{aligned} B_{i,j}\left(y_1^{T+k+n,o}\right) &= \frac{p\left(y_1^{T+k+n,o}|M_i\right)}{p\left(y_1^{T+k+n,o}|M_j\right)} = \\ &= \frac{p\left(y_{T+1}^{T+k+n,o}|y_1^{T,o}, M_i\right)}{p\left(y_{T+1}^{T+k+n,o}|y_1^{T,o}, M_j\right)} \frac{p\left(y_1^{T,o}|M_i\right)}{p\left(y_1^{T,o}|M_j\right)} = B_{i,j}\left(y_{T+1}^{T+k+n,o}|y_1^{T,o}\right) B_{i,j}\left(y_1^{T,o}\right). \end{aligned} \tag{3.7}$$

In turn, it is obvious that

$$\frac{p\left(y_{T+1}^{T+k+n,o}|y_1^{T}, M_i\right)}{p\left(y_{T+1}^{T+k+n,o}|y_1^{T}, M_j\right)} = \frac{\prod_{s=1}^{k+n} p\left(y_{T+s}^{o}|y_1^{T+s-1,o}, M_i\right)}{\prod_{s=1}^{k+n} p\left(y_{T+s}^{o}|y_1^{T+s-1,o}, M_j\right)} = \prod_{s=1}^{k+n} \frac{p\left(y_{T+s}^{o}|y_1^{T+s-1,o}, M_i\right)}{p\left(y_{T+s}^{o}|y_1^{T+s-1,o}, M_j\right)}; \tag{3.8}$$

see Geweke and Amisano (2010). Hence,

$$\log B_{i,j}\left(y_{T+1}^{T+k+n,o}|y_1^{T,o}\right) = \sum_{s=1}^{k+n} \log B_{i,j}\left(y_{T+s}^{o}|y_1^{T+s-1,o}\right), \tag{3.9}$$

and, in consequence, the predictive Bayes factor can be calculated based on the repeated one-step-ahead forecasts. It is worth mentioning that the predictive Bayes factor $B_{i,j}(y_{T+1}^{T+k+n,o}|y_1^{T,o})$ measures the relative predictive power of M_i and M_j within the whole period $T+1,\ldots,T+k+n$, given $y_1^{T,o}$, and it can be viewed as the difference of the two models' logarithmic scores (see Gneiting and Raftery, 2007). In fact, the same rank of models is obtained by the use of the average logarithmic score:

$$\begin{aligned} ALogS\left(1, k+n, T, M_i\right) &= \frac{1}{n+k} \log p\left(y_{T+1}^{T+k+n,o}|y_1^{T}, M_i\right) = \\ &= \frac{1}{n+k} \sum_{s=1}^{n+k} \log p\left(y_{T+s}^{o}|y_1^{T+s-1,o}, M_i\right). \end{aligned} \tag{3.10}$$

Some issue with the interpretation of the average of (or sum of) logarithmic scores arises when one is to compare the k-step-ahead forecasting ability of models M_i and M_j, for $k > 1$. The formulation of this problem and a proposition of its solution are presented in the following section.

4 Comparing models in respect to their k-step-ahead forecasting performance. A new decomposition of the predictive Bayes factor

In order to compare the k-step-ahead forecasting ability of two competing Bayesian models, M_i and M_j, let us start with the posterior odds ratio given the matrix of observations $y_1^{T,o}$ and the realised value of y_{T+k}, that is y_{T+k}^o:

$$\frac{p\left(M_i|y_{T+k}^o, y_1^{T,o}\right)}{p\left(M_j|y_{T+k}^o, y_1^{T,o}\right)} = \frac{p\left(y_{T+k}^o|y_1^{T,o},\, M_i\right)}{p\left(y_{T+k}^o|y_1^{T,o}, M_j\right)} \frac{p\left(M_i|y_1^{T,o}\right)}{p\left(M_j|y_1^{T,o}\right)}. \tag{4.1}$$

The predictive Bayes factor of the form: $B_{i,j}(y_{T+k}^o|y_1^{T,o}) = \frac{p(y_{T+k}^o|y_1^{T,o},M_i)}{p(y_{T+k}^o|y_1^{T,o},M_j)}$, for the k-step ahead forecast, indicates how much the additional data (the k-th observation as from T), y_{T+k}^o, modifies the odds of the two models based on the data up to time T. Henceforth, we refer to this quantity as the predictive Bayes factor of order k at time T. In addition, following Geweke (2005), it is easy to show that the predictive likelihood in the model M_i, $p(y_{T+k}^o|y_1^{T,o}, M_i)$, is the ratio of the corresponding marginal likelihoods:

$$p\left(y_{T+k}^o|y_1^{T,o}, M_i\right) = \frac{p\left(y_{T+k}^o, y_1^{T,o}|M_i\right)}{p\left(y_1^{T,o}|M_i\right)}. \tag{4.2}$$

Thus, it is the multiplicative updating factor applied to the marginal likelihood $p(y_1^{T,o}|M_i)$ after observing y_{T+k}, that produces the new marginal likelihood $p(y_{T+k}^o, y_1^{T,o}|M_i)$.

To approximate the predictive likelihood, $p(y_{T+k}^o|y_1^{T,o}, M_i)$, we first draw $h_{i,1}^{T+k,(q)}$ and $\theta_i^{(q)}$, for $q = 1, \ldots, N$, from the posterior distribution (conditional on $y_1^{T,o}$). Then, if $k > 1$, for each q, a vector $y_{T+1}^{T+k-1,(q)}$ is simulated from the conditional sampling distribution of observations given $y_1^{T,o}$, $h_{i,1}^{T+k,(q)}$ and $\theta_i^{(q)}$. Finally, the arithmetic mean is calculated:

$$\hat{p}\left(y_{T+k}^o|y_1^{T,o}, M_i\right) = \frac{1}{N}\sum_{q=1}^{N} p\left(y_{T+k}^o|y_1^{T,o}, y_{T+1}^{T+k-1,(q)}, h_{i,1}^{T+k,(q)}, \theta_i^{(q)}, M_i\right).$$

Now let us consider the following setup of recursive k-step-ahead forecasts. The set of actual observations is extended by a progressive inclusion of consecutive observations ($s = 1, \ldots, n$), and each time the forecasting is carried out using the k-step-ahead predictive density:

$$\begin{aligned} p\left(y_{T+k+s}|y_1^{T+s,o},M_i\right) = \int p\left(y_{T+k+s},h_{i,T+k+s}|y_1^{T+s,o},h_{i,1}^{T+s},\theta_i,M_i\right)\times \\ \times\, p\left(h_{i,1}^{T+s},\theta_i|y_1^{T+s,o},M_i\right)\,dh_{i,T+k+s}\,dh_{i,1}^{T+s}\,d\theta_i. \end{aligned} \tag{4.3}$$

The following key question arises now: how to compare, from a Bayesian perspective, the k-step-ahead forecasting ability of the models under consideration in the period from $T+1$ to $T+k+n$?

As mentioned above, for $k=1$, the answer is very simple: we can use the predictive Bayes factor $B_{i,j}(y_{T+1}^{T+k+n,o}|y_1^{T,o})$ or, equivalently, we can evaluate the predictive performance of the model M_i by means of the average log predictive likelihoods for the one-period-ahead forecasts. Thus, the Bayesian comparison of the accuracy of density forecasts, based on the predictive Bayes factor, can be regarded as using the log predictive score, with the logarithm of the predictive density evaluated at the realised value of the time series.

Let us now proceed to the case of in which k is fixed and $k>1$. In the literature (see, e.g., Gneiting and Raftery, 2007; Koop and Korobilis, 2012; Cross et al., 2020), a popular measure of the predictive ability of the model M_i for the k-step-ahead forecasts is the average (or sum) of the log predictive likelihoods:

$$AlogPL\,(k,n,T,M_i) = \frac{1}{n+1}\sum_{s=0}^{n}\log p\left(y_{T+k+s}^{o}|y_1^{T+s,o},\,M_i\right). \tag{4.4}$$

However, this measure has no direct Bayesian interpretation. Obviously, it is straightforward to show from (4.1) that the average of the log posterior odds ratios based on all observations is equal to the sum of two averages: the average of the log predictive Bayes factors of order k, and the one of the log posterior odds ratios obtained with the k-th observations excluded:

$$\begin{aligned} &\frac{1}{n+1}\sum_{s=0}^{n}\log\frac{p\left(M_i|y_{T+k+s}^{o},y_1^{T+s,o}\right)}{p\left(M_j|y_{T+k+s}^{o},y_1^{T+s,o}\right)} = \\ &= \frac{1}{n+1}\sum_{s=0}^{n}\log B_{i,j}\left(y_{T+k+s}^{o}|y_1^{T+s,o}\right) + \frac{1}{n+1}\sum_{s=0}^{n}\log\frac{p\left(M_i|y_1^{T+s,o}\right)}{p\left(M_j|y_1^{T+s,o}\right)}. \end{aligned} \tag{4.5}$$

Thus, the average of the log predictive Bayes factors of order k informs one on how the average of the log posterior odds ratios changes after the k-th observations have realized. However, when we want the average of the log posterior odds ratios based on all observations, $\frac{1}{n+1}\sum_{s=0}^{n}\log\frac{p(M_i|y_{T+k+s}^{o},y_1^{T+s,o})}{p(M_j|y_{T+k+s}^{o},y_1^{T+s,o})}$, to be compared to the log posterior odds ratio based on the first T observations, that is $\log\frac{p(M_i|y_1^{T,o})}{p(M_j|y_1^{T,o})}$, we should consider another term different from the average of log predictive Bayes factors of order k. Therefore, here we define the predictive Bayes factor of order $(k,\ s)$ for recursive k-step-ahead forecasts, and we provide a new decomposition of this Bayes factor, delivering a simple interpretation of its factors.

Let us start with the posterior odds ratio given the matrix of observations $y_1^{T+s,o}$ and the realised value of y_{T+k+s} (that is y^o_{T+k+s}) for $s=1,\dots,n$:

$$\frac{p\left(M_i|y^o_{T+k+s},y_1^{T+s,o}\right)}{p\left(M_j|y^o_{T+k+s},y_1^{T+s,o}\right)}=\frac{p\left(y^o_{T+k+s}|y_1^{T+s,o},M_i\right)}{p\left(y^o_{T+k+s}|y_1^{T+s,o},M_j\right)}\frac{p\left(M_i|y_1^{T+s,o}\right)}{p\left(M_j|y_1^{T+s,o}\right)}. \tag{4.6}$$

Next, we have:

$$\frac{p\left(M_i|y^o_{T+k+s},y_1^{T+s,o}\right)}{p\left(M_j|y^o_{T+k+s},y_1^{T+s,o}\right)}=\frac{p\left(y^o_{T+k+s},y_{T+1}^{T+s,o}|y_1^{T,o},M_i\right)}{p\left(y^o_{T+k+s},y_{T+1}^{T+s,o}|y_1^{T,o},M_j\right)}\frac{p\left(M_i|y_1^{T,o}\right)}{p\left(M_j|y_1^{T,o}\right)}=$$

$$=\frac{p\left(y^o_{T+k+s}|y_1^{T+s,o},M_i\right)}{p\left(y^o_{T+k+s}|y_1^{T+s,o},M_j\right)}\frac{p\left(y_{T+1}^{T+s,o}|y_1^{T,o},M_i\right)}{p\left(y_{T+1}^{T+s,o}|y_1^{T,o},M_j\right)}\frac{p\left(M_i|y_1^{T,o}\right)}{p\left(M_j|y_1^{T,o}\right)}=$$

$$=\frac{p\left(y^o_{T+k+s}|y_1^{T+s,o},M_i\right)}{p\left(y^o_{T+k+s}|y_1^{T+s,o},M_j\right)}\frac{\prod_{l=1}^{s}p\left(y^o_{T+l}|y_1^{T+l-1,o},M_i\right)}{\prod_{l=1}^{s}p\left(y^o_{T+l}|y_1^{T+l-1,o},M_j\right)}\frac{p\left(M_i|y_1^{T,o}\right)}{p\left(M_j|y_1^{T,o}\right)}.$$

Thus, the posterior odds ratio given the matrix of observations $y_1^{T+s,o}$ and the realised value of y_{T+k+s} is the product of the three: first, the predictive Bayes factor related to the k-step-ahead forecast at time $T+s$; second, the predictive Bayes factor based on the forecasts for the period (of a length of s) ranging from $t=T+1$ to $t=T+s$; and finally, the posterior odds ratio given the matrix of observations $y_1^{T,o}$:

$$\frac{p\left(M_i|y^o_{T+k+s},y_1^{T+s,o}\right)}{p\left(M_j|y^o_{T+k+s},y_1^{T+s,o}\right)}=B_{i,j}\left(y^o_{T+k+s}|y_1^{T+s,o}\right)B_{i,j}\left(y_{T+1}^{T+s,o}|y_1^{T,o}\right)\frac{p\left(M_i|y_1^{T,o}\right)}{p\left(M_j|y_1^{T,o}\right)}. \tag{4.7}$$

We define this product of the two predictive Bayes factors in the equation above as the predictive Bayes factor of order (k,s) at time T, and denote it as $B_{i,j}(k,s,T)$:

$$B_{i,j}(k,s,T)=B_{i,j}\left(y^o_{T+k+s}|y_1^{T+s,o}\right)B_{i,j}\left(y_{T+1}^{T+s,o}|y_1^{T,o}\right)\ \text{ for } s=0,1,\dots,n, \tag{4.8}$$

where according to our convention:

$$B_{i,j}\left(y_{T+1}^{T+s,o}|y_1^{T,o}\right)=1 \text{ for } s=0. \tag{4.9}$$

Of course, identity (4.8) can be expressed as:

$$B_{i,j}(k,s,T)=B_{i,j}\left(y^o_{T+k+s}|y_1^{T+s,o}\right)\prod_{l=1}^{s}B_{i,j}\left(y^o_{T+l}|y_1^{T+l-1,o}\right). \tag{4.10}$$

The predictive Bayes factor of order (k, s) at time T provides information about how much the k-th observation as from $T+s$, y^o_{T+k+s}, as well as all the observations from $T+1$ to $T+s$ change the posterior odds ratio of the two competing models based on the data up to time T. In other words, $B_{i,j}(k,s,T)$ informs about the predictive ability of two models not only related to y^o_{T+k+s}, but to the entire predicted (updated) path y^{T+s}_{T+1}. Clearly, for $k=1$, the predictive Bayes factor of order (k,s) at time T is equal to the predictive Bayes factor over the period $T+1,\ldots,T+s+1$: $B_{i,j}(1,s,T) = B_{i,j}(y^{T+1+s,o}_{T+1}|y^{T,o}_1)$.

It is worth noting that the first factor in equality (4.8) can be used to compare the predictive ability of two models for the k-step-ahead forecast at time $T+s$, while the second factor is the effect of updating the posterior odds ratios. Thus, the predictive Bayes factor of order (k,s) at time T captures both the k-step-ahead forecast ability and the effect of updating the posterior model probability. Note, that for $s=1,\ldots,n$:

$$\begin{aligned}\log B_{i,j}(k,s,T) &= \log p\left(y^o_{T+k+s}, y^{T+s,o}_{T+1}|y^{T,o}_1, M_i\right) - \log p\left(y^o_{T+k+s}, y^{T+s,o}_{T+1}|y^{T,o}_1, M_j\right) = \\ &= \log p\left(y^o_{T+k+s}|y^{T+s,o}_1, M_i\right) + \sum_{l=1}^{s}\log p\left(y^o_{T+l}|y^{T+l-1,o}_1, M_i\right) + \\ &\quad - \log p\left(y^o_{T+k+s}|y^{T+s,o}_1, M_j\right) - \sum_{l=1}^{s}\log p\left(y^o_{T+l}|y^{T+l-1,o}_1, M_j\right).\end{aligned} \tag{4.11}$$

Therefore, the log predictive Bayes factor of order (k,s) at time T can be treated as the difference of the corresponding log predictive score functions (cf. Geweke, 2005; Geweke and Amisano, 2011).

Let us now consider the average of the log posterior odds ratios. From identity (4.7), we obtain:

$$\begin{aligned}&\frac{1}{n+1}\sum_{s=0}^{n}\log\frac{p\left(M_i|y^o_{T+k+s}, y^{T+s,o}_1\right)}{p\left(M_j|y^o_{T+k+s}, y^{T+s,o}_1\right)} = \\ &= \frac{1}{n+1}\left[\sum_{s=0}^{n}\log B_{i,j}\left(y^o_{T+k+s}|y^{T+s,o}_1\right) + \sum_{s=1}^{n}\log B_{i,j}\left(y^{T+s,o}_{T+1}|y^{T,o}_1\right)\right] + \\ &+ \log\frac{p\left(M_i|y^{T,o}_1\right)}{p\left(M_j|y^{T,o}_1\right)}.\end{aligned} \tag{4.12}$$

As we can see from identity (4.12), the average (with respect to s) of the Bayes factors of order (k,s) at time T informs us how the posterior odds ratio, on average, changes by successively updating the observations: y^o_{T+k+s} and y^o_{T+s} for $s=1,\ldots,n$.

Note that the average of the logarithmic scores for the model M_i (in other words, the average of the cumulative predictive likelihoods for the model M_i) can be written as follows:

$$\begin{aligned} &AClogPL\left(k,T,M_i\right)= \\ &=\frac{1}{n+1}\left[\log p\left(y^o_{T+k}|y_1^{T,o},\ M_i\right)+\sum_{s=1}^{n}\log p\left(y^o_{T+k+s},y_{T+1}^{T+s,o}|y_1^{T,o},M_i\right)\right]= \\ &=\frac{1}{n+1}\left[\sum_{s=0}^{n}\log p\left(y^o_{T+k+s}|y_1^{T+s,o},M_i\right)+\sum_{s=1}^{n}\sum_{l=1}^{s}\log p\left(y^o_{T+l}|y_1^{T+l-1,o},M_i\right)\right]. \end{aligned} \tag{4.13}$$

Hence, the difference between $AClogPL(k,T,M_i)$ and $AClogPL(k,T,M_j)$ can be expressed with the use of the log predictive Bayes factors of order (k,s) at time T:

$$\begin{aligned} &AClogPL\left(k,T,M_i\right)-AClogPL\left(k,T,M_j\right)= \\ &=\frac{1}{n+1}\left[\sum_{s=0}^{n}\log B_{i,j}\left(y^o_{T+k+s}|y_1^{T+s,o}\right)+\sum_{s=1}^{n}\log B_{i,j}\left(y_{T+1}^{T+s,o}|y_1^{T,o}\right)\right]= \\ &=\frac{1}{n+1}\sum_{s=0}^{n}\log B_{i,j}\left(y^o_{T+k+s}|y_1^{T+s,o}\right)+\frac{1}{n+1}\sum_{s=1}^{n}\log B_{i,j}\left(y_{T+1}^{T+s,o}|y_1^{T,o}\right). \end{aligned} \tag{4.14}$$

The first component on the right-hand side of (4.14) informs one about the relative predictive ability of the models for the k-step-ahead forecasting in the period $T+k$, $\dots, T+n+k$. In turn, the second component is connected with the recursive updates of the posterior odds ratios based on the updated data sets. Formula (4.14) shows that the average (with respect to s) of the log predictive Bayes factors of order (k,s) at time T is equal to the difference between the averages of the corresponding log predictive scores (the average of the corresponding log predictive likelihoods), which can also be decomposed in a natural manner into two components: the k-step-ahead forecast ability and the updating effect, respectively. By re-estimating both models n times (for $s=1,\dots,n$, each time extending the matrix of observations by consecutive data points) and calculating the corresponding Bayes factors of order (k,s) for a fixed k, we can both: (i) compare the predictive ability of the models for the k-step-ahead forecasting at each time point $T+s$, and (ii) include the effect of updating the data set.

We end this section by providing some intuitions pertaining to the methodology developed above. Let us consider two very simple hypothetical examples. Suppose that $k>1$ is fixed, and assume:

$$\frac{p\left(M_1|y_1^{T,o}\right)}{p\left(M_2|y_1^{T,o}\right)}=2,$$

$$B_{1,2}\left(y^o_{T+k}|y_1^{T,o}\right)=0.1,\quad B_{1,2}\left(y^o_{T+k+1}|y_1^{T+1,o}\right)=0.01,$$

$$B_{1,2}\left(y^o_{T+k+2}|y_1^{T+2,o}\right)=1000,\quad B_{1,2}\left(y^o_{T+k+3}|y_1^{T+3,o}\right)=10.$$

Moreover:

$$B_{1,2}\left(y^o_{T+3}|y_1^{T+2,o}\right)=B_{1,2}\left(y^o_{T+2}|y_1^{T+1,o}\right)=B_{1,2}\left(y^o_{T+1}|y_1^{T,o}\right)=0.1.$$

s	$\frac{p(M_1 \vert y^o_{T+k+s}, y_1^{T+s,o})}{p(M_2 \vert y^o_{T+k+s}, y_1^{T+s,o})}$	$B_{1,2}(y^o_{T+k+s} \vert y_1^{T+s,o})$	$B_{1,2}(y_{T+1}^{T+s,o} \vert y_1^{T,o})= =0.1^s$	$\frac{p(M_1 \vert y_1^{T,o})}{p(M_2 \vert y_1^{T,o})}$
0	0.2	0.1	1	2
1	0.002	0.01	0.1^1	2
2	20	1000	0.1^2	2
3	0.02	10	0.1^3	2
average of logs	≈ -0.95	0.25	-1.5	≈ 0.30

Table 1: The first hypothetical example.

From identity (4.5) we obtain:

$$\frac{p\left(M_1|y^o_{T+k}, y_1^{T,o}\right)}{p\left(M_2|y^o_{T+k}, y_1^{T,o}\right)} = 0.2, \quad \frac{p\left(M_1|y^o_{T+k+1}, y_1^{T+1,o}\right)}{p\left(M_2|y^o_{T+k+1}, y_1^{T+1,o}\right)} = 0.002,$$

$$\frac{p\left(M_1|y^o_{T+k+2}, y_1^{T+2,o}\right)}{p\left(M_2|y^o_{T+k+2}, y_1^{T+2,o}\right)} = 20, \quad \frac{p\left(M_1|y^o_{T+k+3}, y_1^{T+3,o}\right)}{p\left(M_2|y^o_{T+k+3}, y_1^{T+3,o}\right)} = 0.02.$$

All calculations are summarized in Table 1.

The average of the logarithms of the predictive Bayes factors of order k, $B_{1,2}(y^o_{T+k+s}|y_1^{T+s,o})$, for $s = 0, 1, 2, 3$, is equal to 0.25, which indicates that the predictive capacity of the model M_2 for sheer k-step-ahead forecasting is better than that of the model M_1. However, when we take into account the updating effect, represented by the average log predictive Bayes factor over the period $T+1, \ldots, T+3$, (the average of $\log B_{1,2}(y_{T+1}^{T+s,o}|y_1^{T,o})$ with respect to s), which is equal to -1.5, we see that the posterior odds for the data up to time T, on average, decreases in the period from $T+1$ to $T+3$. In consequence, the average of the log posterior odds ratios for both: (i) the four consecutive k-step-ahead forecasts considered, and (ii) the updating of the data sets is equal to -0.95, whereas $\log \frac{p(M_1|y_1^{T,o})}{p(M_2|y_1^{T,o})} \approx 0.3$.

In the second example, it is assumed that for the one-step-ahead forecasts both of the models have the same predictive ability measured by the predictive Bayes factors, $B_{1,2}(y^o_{T+l}|y_1^{T+l-1,o}) = 1$ for $l = 1, \ldots, s$, and $s > 0$. The values of $B_{1,2}(y^o_{T+k+s}|y_1^{T+s,o})$ for $s = 0, 1, 2, 3$ remain the same as in the first example. In consequence, the updating affect equals zero, while the posterior odds ratio increases, on average (see the second column in Table 2).

The examples above show that the average of the log predictive Bayes factors of order k, $B_{1,2}(y^o_{T+k+s}|y_1^{T+s,o})$, measures only the k-step-ahead forecasting ability, without taking into account the effect of the data updates (which are always present in ex post analyses), whereas the average of the log predictive Bayes factors of order (k, s) at time T does measure the change in the average odds.

s	$\frac{p(M_1\|y^o_{T+k+s},y_1^{T+s,o})}{p(M_2\|y^o_{T+k+s},y_1^{T+s,o})}$	$B_{1,2}(y^o_{T+k+s}\|y_1^{T+s,o})$	$B_{1,2}(y_{T+1}^{T+s,o}\|y_1^{T,o})$	$\frac{p(M_1\|y_1^{T,o})}{p(M_2\|y_1^{T,o})}$
0	0.2	0.1	1	2
1	0.02	0.01	1	2
2	2000	1000	1	2
3	20	10	1	2
average of logs	≈ 0.55	0.25	0	≈ 0.30

Table 2: The second hypothetical example.

5 Simulation examples

In this section, we illustrate the usefulness of our new decomposition of the predictive Bayes factor of order (k, s), using simulated data. We apply the new measures of models' forecasting ability to compare forecasts from different models when the assumed data generating process (DGP) is actually known. We consider two 3-dimentional vector autoregressive processes of order two, each with two cointegrating relations and a constant restricted to the cointegration space:

1. VEC with constant conditional covariance matrix,
2. VEC with log-normal Multiplicative Stochastic Factor structure, VEC-LN-MSF.

Further details about the DGPs will be provided in the next section, where the models for empirical data are presented. We simulate time series of length $T + n = 275$, whereof the first 175 observations are used as a training sample, and the last $n = 100$ data points for evaluating the performance of our new measure. The DGPs under consideration are related to the empirical part of the paper Pajor et al. (2022), where VAR models with time-varying conditional covariances are analysed. Using the average log predictive score and the new measure, we compare one- to eight-step-ahead forecasts generated by the true model to various other specifications, which differ from the previous one in terms of the number of cointegrating relations and the structure of conditional covariances. The results are collected in Tables 3–8 and Figures 1–2.

Tables 3–8 reveal that, in terms of the sheer k-step-ahead forecasting ability (measured by the average log predictive Bayes factor of order k), the true model does not necessarily turn out as the best one if $k > 1$. However, for the purpose of the overall model comparison, the average of the log predictive Bayes factors of order (k, s) is by far more appropriate. For all forecast horizons this new, joint measure pinpoints the true DGP model, thus appears to be working well here.

6 Empirical results

In this section, the methods presented above are employed to assess the forecast accuracy of inflation of consumer prices (Δp_t), unemployment rate (U_t) and short-term interest

No.	Model	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$
1	VEC(0)	0.134	0.204	0.294	0.36	0.412	0.461	0.506	0.545
2	VEC(1)	0.055	0.067	0.108	0.135	0.152	0.176	0.196	0.213
3	**VEC(2), true DGP**	0	0	0	0	0	0	0	0
4	VEC-LN-MSF(2)	0.004	0.004	0.002	0.001	-0.001	-0.001	-0.002	-0.002
5	VEC-t-DBEKK(2)	0.013	0.016	0.017	0.012	0.013	0.007	0.006	0.005
6	VEC-LN-MSF-DBEKK(2)	0.005	0.005	0.006	0.004	0.005	0.003	0.003	0.001
	The best model	3	3	3	3	4	4	4	4

Table 3: Average log predictive Bayes factor of order k, $\frac{1}{n+1}\sum_{s=0}^{n}\log B_{\text{VEC}(2),j}(y^o_{T+k+s}|y_1^{T+s,o})$, for VEC(2) as the DGP. The number of cointegrating relations is given in parentheses.

No.	Model	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$
1	VEC(0)	7.156	7.225	7.315	7.381	7.433	7.482	7.527	7.566
2	VEC(1)	2.857	2.869	2.91	2.938	2.954	2.979	2.998	3.015
3	**VEC(2), true DGP**	0	0	0	0	0	0	0	0
4	VEC-LN-MSF(2)	0.224	0.224	0.222	0.221	0.219	0.219	0.218	0.218
5	VEC-t-DBEKK(2)	1.001	1.004	1.005	1	1.001	0.995	0.994	0.993
6	VEC-LN-MSF-DBEKK(2)	0.33	0.33	0.331	0.329	0.331	0.328	0.328	0.327
	The best model	3	3	3	3	3	3	3	3

Table 4: Average log predictive Bayes factor of order (k,s), $\frac{1}{n+1}\left[\sum_{s=0}^{n}\log B_{\text{VEC}(2),j}(y^o_{T+k+s}|y_1^{T+s,o}) + \sum_{s=1}^{n}\log B_{\text{VEC}(2),j}(y_{T+1}^{T+s,o}|y_1^{T,o})\right]$, for VEC(2) as the DGP. The number of cointegrating relations is given in parentheses.

No.	Model (no. of coint. relations)	Average of $\log B_{\text{VEC}(2),j}(y_{T+1}^{T+s,o}|y_1^{T,o})$
1	VEC(0)	7.022
2	VEC(1)	2.802
3	**VEC(2), true DGP**	**0**
4	VEC-LN-MSF(2)	0.220
5	VEC-t-DBEKK(2)	0.988
6	VEC-LN-MSF-DBEKK(2)	0.325
	The best model	3

Table 5: Average of $\log B_{\text{VEC}(2),j}(y_{T+1}^{T+s,o}|y_1^{T,o})$ for VEC(2) as the DGP.

rate (r_t) within the framework of the so-called small model of monetary policy (see, e.g., Primiceri, 2005). This model has been applied for the data for the Polish economy by Pajor and Wróblewska (2017), and Wróblewska and Pajor (2019), where different VEC-MSF specifications have been compared. In this paper, similarly as in Pajor et al. (2022), we focus on a larger class of specifications and on two (rather than one) quarterly data sets, representing not only the Polish but also the US economy (see Figure 3). The data for Poland ranges from 1998Q1 to 2020Q2, with the final 30 quarters (i.e. 2013Q1 to 2020Q2) designated for forecast evaluation. The series are seasonally unadjusted and their seasonality is modelled and forecasted in a deterministic manner, i.e. using

No.	Model	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$
1	VEC-LN-MSF(0)	0.159	0.207	0.257	0.28	0.309	0.348	0.386	0.399
2	VEC-LN-MSF(1)	0.067	0.06	0.057	0.067	0.076	0.079	0.081	0.088
3	**VEC-LN-MSF(2), true DGP**	0	0	0	0	0	0	0	0
4	VEC(2)	0.136	0.105	0.082	0.048	0.049	0.046	0.06	0.063
5	VEC-t-DBEKK(2)	0.064	0.039	0.013	−0.004	−0.007	−0.01	0.003	0.004
6	VEC-LN-MSF-DBEKK(2)	0.023	0.013	0.018	0.021	0.012	0.021	0.022	0.019
	The best model	3	3	3	5	5	5	3	3

Table 6: Average log predictive Bayes factor of order k $\frac{1}{n+1}\sum_{s=0}^{n}\log B_{\text{LN-MSF}(2),j}(y^o_{T+k+s}|y_1^{T+s,o})$, for LN-MSF(2) as the DGP. The number of cointegrating relations is given in parentheses.

No.	Model	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$
1	VEC-LN-MSF(0)	7.572	7.619	7.67	7.693	7.722	7.76	7.798	7.812
2	VEC-LN-MSF(1)	2.9	2.891	2.89	2.9	2.908	2.911	2.913	2.921
3	**VEC-LN-MSF(2), true DGP**	0	0	0	0	0	0	0	0
4	VEC(2)	3.389	3.357	3.335	3.301	3.301	3.298	3.312	3.315
5	VEC-t-DBEKK(2)	1.462	1.436	1.411	1.394	1.39	1.387	1.4	1.402
6	VEC-LN-MSF-DBEKK(2)	0.974	0.963	0.968	0.972	0.962	0.97	0.972	0.969
	The best model	3	3	3	3	3	3	3	3

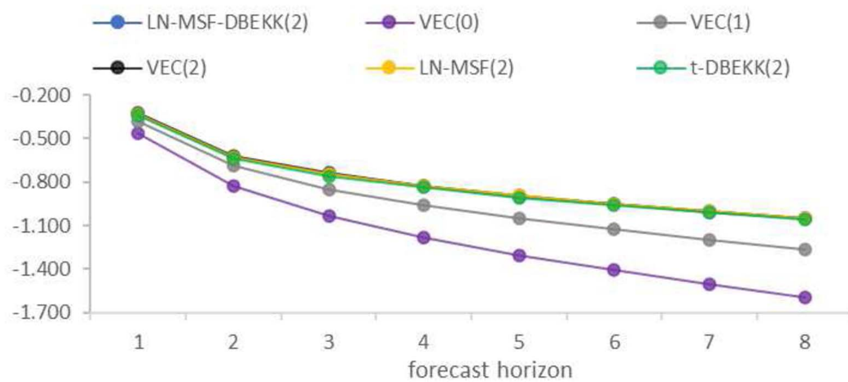
Table 7: Average log predictive Bayes factor of order (k,s), $\frac{1}{n+1}\left[\sum_{s=0}^{n}\log B_{\text{LN-MSF}(2),j}(y^o_{T+k+s}|y_1^{T+s,o})+\sum_{s=1}^{n}\log B_{\text{LN-MSF}(2),j}(y_{T+1}^{T+s,o}|y_1^{T,o})\right]$, for LN-MSF(2) as the DGP. The number of cointegrating relations is given in parentheses.

No.	Model (no. of coint. relations)	Average of $\log B_{\text{LN-MSF}(2),j}(y_{T+1}^{T+s,o}|y_1^{T,o})$
1	VEC-LN-MSF(0)	7.413
2	VEC-LN-MSF(1)	2.833
3	**VEC-LN-MSF(2), true DGP**	0
4	VEC(2)	3.253
5	VEC-t-DBEKK(2)	1.398
6	VEC-LN-MSF-DBEKK(2)	0.951
	The best model	3

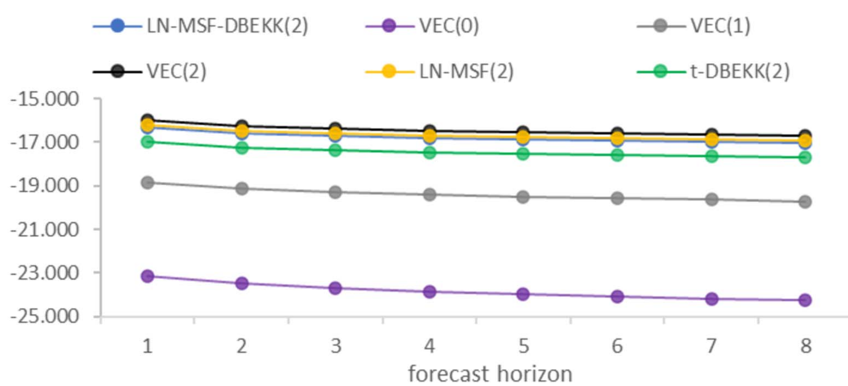
Table 8: Average of $\log B_{\text{LN-MSF}(2),j}(y_{T+1}^{T+s,o}|y_1^{T,o})$ for LN-MSF(2) as the DGP.

zero-mean seasonal dummies. The prior structure is the same as in Pajor et al. (2022). The US data covers the period from 1960Q1 to 2015Q4, and the forecast evaluation is performed for the final 56 quarters (i.e. 2002Q1 to 2015Q4). The analysed series are seasonally adjusted, and the Wu-Xia shadow federal funds rate (the last business day of a quarter) is used instead of the nominal rate.[3]

[3] https://www.frbatlanta.org/cqer/research/wu-xia-shadow-federal-funds-rate.aspx (see also Wu and Xia, 2016).



(a) Average of $\log p(y^o_{T+k+s}|y_1^{T+s,o}, M_i)$



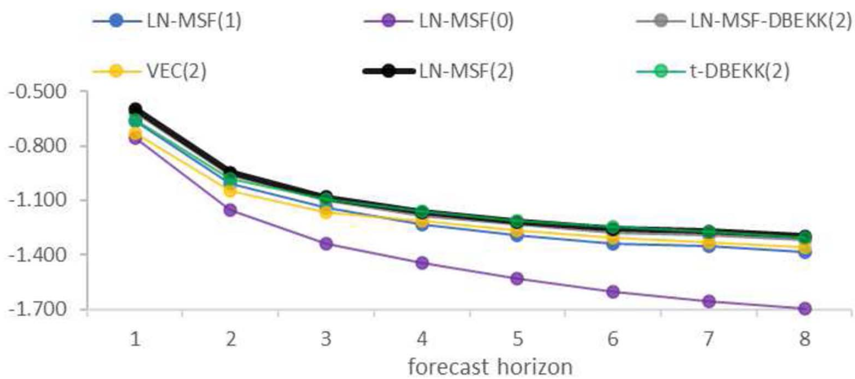
(b) Average of $\log p(y^o_{T+k+s}, y_{T+1}^{T+s,o}|y_1^{T,o}, M_i)$

Figure 1: Average log predictive density values for VEC(2) as the DGP: (a) without the updating effect, (b) with the updating effect. The number of cointegrating relations is given in parentheses.

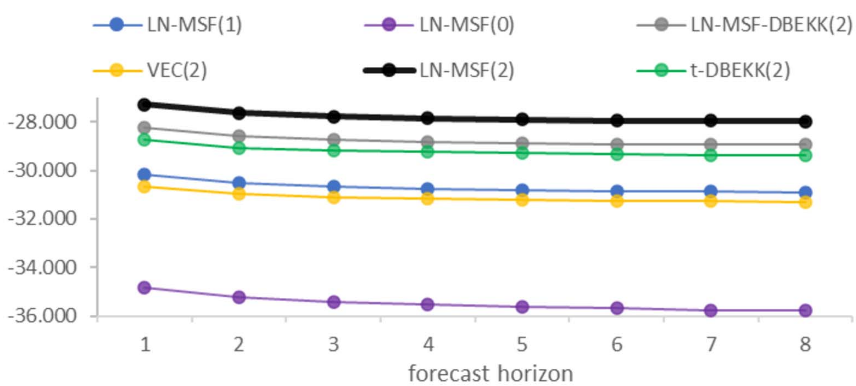
6.1 Bayesian models

The models considered here are motivated by Pajor et al. (2022), where numerous VAR structures with time-varying conditional covariances are analysed. We consider 3-variate vector autoregressive processes of order 2, in the vector error correction (VEC) form, with deterministic terms, two cointegration relationships, and time-varying conditional covariance structures:

$$\Delta y_t = \alpha\beta' \begin{bmatrix} y_{t-1} \\ D_t^{co} \end{bmatrix} + \Gamma_1 \Delta y_{t-1} + \Phi D_t + \varepsilon_t, \quad \varepsilon_t = \Sigma_t^{1/2}\zeta_t, \quad \zeta_t \sim iiN(0, I_3), \tag{6.1}$$



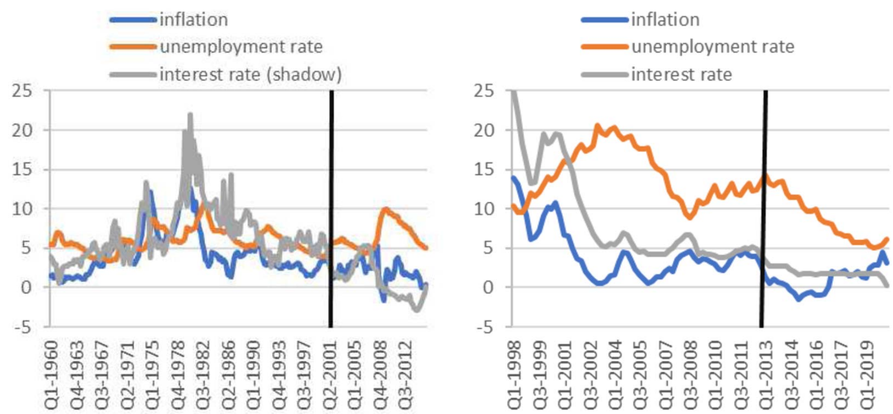
(a) Average of $\log p(y^o_{T+k+s}|y_1^{T+s,o}, M_i)$



(b) Average of $\log p(y^o_{T+k+s}, y_{T+1}^{T+s,o}|y_1^{T,o}, M_i)$

Figure 2: Average log predictive density values for LN-MSF(2) as the DGP: (a) without the updating effect, (b) with the updating effect. The number of cointegrating relations is given in parentheses.

where y_t is the observed and forecasted 3-variate random variable, $t = 1, \ldots, T+k+n$, α is a (3×2) matrix of adjustment coefficients, β' is a (2×4) cointegration matrix of full rank $r = 2$ (with only non-negative elements in the first column), D_t^{co} is a deterministic term (constant) included in the cointegration relations, D_t is the matrix of remaining deterministic variables (e.g. seasonal dummy variables), and Σ_t is a (3×3) symmetric and positive-definite matrix, which may depend on ψ_{t-1} (the past of y_t up to time $t-1$) and q_t (a latent variable from a scalar or vector random process $\{q_t\}$). Let θ be the vector of all parameters involved. As $\varepsilon_t|\psi_{t-1}, q_t, \theta \sim N(0, \Sigma_t)$, Σ_t is the conditional covariance matrix: $\Sigma_t = V(\varepsilon_t|\psi_{t-1}, q_t, \theta)$. We assume that $\Sigma_t = q_t\tilde{\Sigma}_t$, where $q_t > 0$ is a multiplicative stochastic factor (MSF) part, while $\tilde{\Sigma}_t$ may depend on ψ_{t-1}, as in the GARCH-type processes. We start with the Diagonal BEKK (DBEKK) form of the



(a) The US data (b) The Polish data

Figure 3: Analysed data. The vertical (black) line represents the beginning of the forecasting period.

matrix $\tilde{\Sigma}_t$:

$$\tilde{\Sigma}_t = \Sigma + A(\varepsilon_{t-1}\varepsilon_{t-1}')A + G\tilde{\Sigma}_{t-1}G, \tag{6.2}$$

where Σ is a $(3{\times}3)$ symmetric and positive-definite matrix, $A = diag(a_{11}, a_{22}, a_{33})$, $G = diag(g_{11}, g_{22}, g_{33})$, $a_{11} > 0$, $g_{11} > 0$ and $a_{ii}^2 + g_{ii}^2 < 1$ for $i = 1, 2, 3$. If $a_{ii} = a$ and $g_{ii} = g$ for $i = 1, 2, 3$, then the DBEKK form of $\tilde{\Sigma}_t$ reduces to a simpler case of Scalar BEKK (SBEKK).

The MSF component, representing the class of multivariate stochastic volatility models, is considered here in two forms: log normal (LN) and inverse gamma (IG). The logarithm of variable q_t is defined accordingly:

$$\ln q_t = \begin{cases} \phi \ln q_{t-1} + \eta_t, \ \ |\phi| < 1, \ \ \eta_t \perp \zeta_s, \ \ \{\eta_t\} \sim iiN\left(0,\, \sigma_q^2\right): \text{ LN-MSF}, \\ \phi \ln q_{t-1} + \ln \gamma_t, \ \ |\phi| < 1, \ \ \ \gamma_t \perp \zeta_s, \ \ \{\gamma_t\} \sim iiIG\left(\frac{\nu}{2},\, \frac{\nu}{2}\right): \text{ IG-MSF}. \end{cases} \tag{6.3}$$

The LN-MSF specification means that the univariate latent process $\{q_t\}$ is in fact the same as in the basic stochastic volatility (SV) process, where the log of unobserved volatility follows a stationary and causal Gaussian AR(1) process. In the IG-MSF specification, the log volatility follows a stationary, causal but non-Gaussian AR(1) process. Such $\{q_t\}$, multiplied by $\tilde{\Sigma}_t$ of either the SBEKK or DBEKK form, leads to simple hybrid MSV-MGARCH structures, proposed and developed in Osiewalski (2009), Osiewalski and Pajor (2009, 2018, 2019) and Pajor et al. (2022). All other model specifications with the conditional covariance matrix of the form $\Sigma_t = q_t\tilde{\Sigma}_t$, considered in this paper, are special or limiting cases of these basic two specifications; see Table 9 for the details.

Apart from the continuously-valued volatility processes presented above, we also consider a Markov-switching approach to capturing heteroscedasticity, in which the volatility dynamics is limited to discrete breaks over a finite number of states. The latter was also (though the only) of interest in Kwiatkowski (2020a,b), where the predictive

Process	Description of process $\{\varepsilon_t\}$: $\varepsilon_t = \zeta_t \tilde{\Sigma}_t^{1/2} q_t^{1/2}, \quad \{\zeta_t\} \sim iiN(0,\, I)$
LN-MSF-DBEKK	$\tilde{\Sigma}_t = \Sigma + A\left(\varepsilon_{t-1}\varepsilon'_{t-1}\right)A + G\tilde{\Sigma}_{t-1}G$, $A = diag(a_{11}, a_{22}, a_{33}), \quad G = diag(g_{11}, g_{22}, g_{33})$, $\ln q_t = \varphi \ln q_{t-1} + \eta_t, \quad \{\eta_t\} \sim iiN(0,\, \sigma_q^2)$
IG-MSF-DBEKK	$\tilde{\Sigma}_t = \Sigma + A\left(\varepsilon_{t-1}\varepsilon'_{t-1}\right)A + G\tilde{\Sigma}_{t-1}G$, $A = diag(a_{11}, a_{22}, a_{33}), \quad G = diag(g_{11}, g_{22}, g_{33})$, $\ln q_t = \varphi \ln q_{t-1} + \ln \gamma_t, \quad \{\gamma_t\} \sim iiIG(v/2,\, v/2)$
LN-MSF-SBEKK	$\tilde{\Sigma}_t = \Sigma + A\left(\varepsilon_{t-1}\varepsilon'_{t-1}\right)A + G\tilde{\Sigma}_{t-1}G$, $A = diag(a_{11}, a_{11}, a_{11}), \quad G = diag(g_{11}, g_{11}, g_{11})$, $\ln q_t = \varphi \ln q_{t-1} + \eta_t, \quad \{\eta_t\} \sim iiN(0,\, \sigma_q^2)$
IG-MSF-SBEKK	$\tilde{\Sigma}_t = \Sigma + A\left(\varepsilon_{t-1}\varepsilon'_{t-1}\right)A + G\tilde{\Sigma}_{t-1}G$, $A = diag(a_{11}, a_{11}, a_{11}), \quad G = diag(g_{11}, g_{11}, g_{11})$, $\ln q_t = \varphi \ln q_{t-1} + \ln \gamma_t, \quad \{\gamma_t\} \sim iiIG(v/2,\, v/2)$
IG-DBEKK (t-DBEKK)	$\tilde{\Sigma}_t = \Sigma + A\left(\varepsilon_{t-1}\varepsilon'_{t-1}\right)A + G\tilde{\Sigma}_{t-1}G$, $A = diag(a_{11}, a_{22}, a_{33}), \quad G = diag(g_{11}, g_{22}, g_{33})$, $\{q_t\} \sim iiIG(v/2,\, v/2)$
IG-SBEKK (t-SBEKK)	$\tilde{\Sigma}_t = \Sigma + A\left(\varepsilon_{t-1}\varepsilon'_{t-1}\right)A + G\tilde{\Sigma}_{t-1}G$, $A = diag(a_{11}, a_{11}, a_{11}), \quad G = diag(g_{11}, g_{11}, g_{11})$, $\{q_t\} \sim iiIG(v/2,\, v/2)$
LN-MSF	$\tilde{\Sigma}_t = \Sigma$, $\ln q_t = \varphi \ln q_{t-1} + \eta_t, \quad \{\eta_t\} \sim iiN(0,\, \sigma_q^2)$
IG-MSF	$\tilde{\Sigma}_t = \Sigma$, $\ln q_t = \varphi \ln q_{t-1} + \ln \gamma_t, \quad \{\gamma_t\} \sim iiIG(v/2,\, v/2)$
DBEKK	$\tilde{\Sigma}_t = \Sigma + A\left(\varepsilon_{t-1}\varepsilon'_{t-1}\right)A + G\tilde{\Sigma}_{t-1}G$, $A = diag(a_{11}, a_{22}, a_{33}), \quad G = diag(g_{11}, g_{22}, g_{33})$, $q_t \equiv 1$
SBEKK	$\tilde{\Sigma}_t = \Sigma + A\left(\varepsilon_{t-1}\varepsilon'_{t-1}\right)A + G\tilde{\Sigma}_{t-1}G$, $A = diag(a_{11}, a_{11}, a_{11}), G = diag(g_{11}, g_{11}, g_{11})$, $q_t \equiv 1$
Student's t white noise	$\tilde{\Sigma}_t = \Sigma, \quad \{q_t\} \sim iiIG(v/2,\, v/2)$
Gaussian white noise	$\tilde{\Sigma}_t = \Sigma, \quad q_t \equiv 1$

Table 9: The model assumptions.

performance of two- and three-state VEC models with Markov-switching heteroscedasticity (VEC-MSH) is evaluated, much in the spirit of the current study, although using shorter time series and evaluating only the one-step-ahead predictions. The two-state VEC-MSH specifications were also examined in Pajor et al. (2022), yet again, in terms of the one-step-ahead forecasts. To the best of our knowledge, except for the cited works and also the paper by Hauzenberger et al. (2021), hardly can be found in the literature studies examining the forecasting performance of the (Bayesian) VEC-MSH models, and even more so in the context of multiple-step-ahead predictions. Therefore, our paper also extends the cited works on VECs with Markovian regime changes in conditional volatility, maintaining the logic followed therein and based on a simple premise that for some data sets (such as relatively short macroeconomic time series) allowing for only discrete rather than continuously-valued volatility changes may just prove empirically sufficient.

In a K-state VEC-MSH model, where $K \in \{2, 3, \dots\}$ is the number of allowed regimes, the conditional covariance matrix of ε_t is indexed with a random variable $S_t \in \{1, 2, \dots, K\}$, denoting the state in which the modelled system resides at time t: $\Sigma_t = \Sigma_{S_t}$, so that $\varepsilon_t|S_t, \theta \sim N(0, \Sigma_{S_t})$. The sequence $\{S_t\}$ forms a latent, K-state, homogeneous and ergodic Markov chain with transition probabilities $p_{ij} = p(S_t = j|S_{t-1} = i)$, for $i, j = 1, \dots, K$.

Although various numbers of regimes could be allowed in the VEC-MSH structures, we limit our attention to two-state specifications only, i.e. $K = 2$, thereby enabling the system to switch between the high- and low-volatility regimes, with the conditional probabilities of remaining in a given state equal to p_{11} and p_{22}, respectively. As evidenced by Kwiatkowski (2020b), although allowing for yet another state for the U.S. data analysed therein could emerge empirically valid in terms of the in-sample inference, the predictive performance of the three-state models turns out to be just on a par with that of the simpler, two-state specifications. As for the Polish data, our preliminary estimation results (not reported in this study) for three-state VEC–MSH models do not provide empirical support for a third state whatsoever.

Similarly to Kwiatkowski (2020a,b), in order to address the problem of label switching, inherent to (Markov) mixture models, in all the VEC–MSH specifications (for the US and Poland) we impose an identification restriction of the form: $Var(r_t|S_t = 1, \psi_{t-1}, \theta) > Var(r_t|S_t = 2, \psi_{t-1}, \theta)$, so that the first regime displays a higher volatility of interest rates. However, we note that the estimation and forecasting results presented below prove insensitive to the choice of the state-identifying variable.

For the sake of comparison, we also include in this study two univariate, conditionally Gaussian and homoscedastic models, considered separately for each variable: stationary AR(1) for the first differences, and random walk (RW) for the levels. Note that, to remain consistent with the assumption of a constant present only in cointegration relations (see 6.1), neither of the two models features an intercept. However, for the Polish data, both structures are equipped with seasonal dummies. Admittedly, it may be expected that such models would perform rather poorly as they ignore the dependencies between the variables. Nevertheless, they are widely recognized as benchmark specifications in forecasting exercises.

The Bayesian statistical model amounts to specifying the joint distribution of all observations, latent variables and parameters. The assumptions presented so far determine the conditional distribution of the matrix of observations and the vector of latent variables given the parameters. Thus, what remains to be done is to formulate the marginal distribution of the parameters (the prior distribution). We start with the prior for α and β. For matrix $\alpha\beta'$ we use the parametrisation proposed by Koop et al. (2009):

$$\alpha\beta' = (\alpha M_\Pi)\left(\beta M_\Pi^{-1}\right)' \equiv QB', \tag{6.4}$$

where M_Π is a $(2{\times}2)$ symmetric and positive-definite matrix, and Q and B are unrestricted matrices. Identity (6.4) holds, in particular, for $\alpha = Q(B'B)^{1/2}$ and $\beta =$

$B(B'B)^{-1/2}$. The prior distributions for matrices Q and B imply the prior distributions for α and β. Moreover, we assume independence among groups of parameters. To specify the prior structures, we follow Osiewalski and Pajor (2019), Pajor and Wróblewska (2017) and Pajor et al. (2022). In particular, for the conditional mean parameters (common to all the models under consideration), we set:[4]

- the matrix normal distribution for B: $p\,(B|r=2) = f_{mN}\,(B|0, I_2, 0.01I_m)$, which leads to the matrix angular central Gaussian (MACG) distribution for β: $p\,(\beta|r=2) = f_{MACG}\,(\beta|0.01I_m) = f_{MACG}\,(\beta|I_m)$ (see, e.g., Chikuse, 2002), $E(B)=0$, $V(vecB) = 0.01I_r \otimes I_m = 0.01I_{rm}$,
- the matrix normal distribution for Q: $p\,(Q|\nu, r=2) = f_{mN}\,(Q|0, \nu I_2, 0.01I_3)$ with inverse gamma distribution for ν: $p\,(\nu) = f_{IG}\,(\nu|3,\,2)$, so $E(\nu)=1$, $Var(\nu)=1$ and $E(Q|r=2)=0$, $V(vecQ|\nu, r=2) = \nu I_2 \otimes 0.01I_3 = 0.01\nu I_6$,
- the matrix normal distribution for Γ_1: $p\,(\Gamma_1|h) = f_{mN}\,(\Gamma_1|0, I_3, hI_3)$ with inverse gamma distribution for h: $p\,(h) = f_{IG}\,(h|3,\,2)$, $E\,(h)=1$, $Var\,(h)=1$, $E(\Gamma_1) = 0$, $V(vec\,\Gamma_1|h) = I_3 \otimes hI_3 = hI_9$,
- the matrix normal distribution for Φ: $p\,(\Phi|h_s) = f_{mN}\,(\Phi|0, I_3, h_sI_{l_s})$ 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, $E(h_s)=1$, $Var(h_s)=1$, $E(\Phi)=0$, $V(vec\,\Phi|h_s) = I_3 \otimes h_sI_{l_s} = h_sI_{3l_s}$.

For the parameters related to the LN-MSF-DBEKK and IG-MSF-DBEKK volatility structures, we assume the following priors:

- the inverse Wishart distribution for Σ: $p\,(\Sigma) = f_{IW}\,(\Sigma|I_3, 5)$, so $E(\Sigma) = I_3$,
- the gamma distribution for v: $f_G\,(v|3,\,0.1)$, so $E(v) \approx 30$, $Mode(v) \approx 20$,
- the normal distribution for ϕ, truncated by the restriction $|\phi| < 1$: $p\,(\phi) \propto f_N\,(\phi|0,\,100)\,I_{(-1,\,1)}(\phi)$,
- the inverse gamma distribution for σ_q^2: $p\left(\sigma_q^2\right) = f_{IG}\left(\sigma_q^2|2.5,\,0.16\right)$, so $E(\sigma_q^2) \approx 0.107$ (see Osiewalski and Pajor, 2019),

[4]The following symbols are used:
$f_{mN}(X|M,U,V)$ – the probability density function of the matrix normal distribution with mean M, and positive definite matrices U and V;
$f_{IG}(x|a,b)$ – the probability density function of the inverse gamma distribution with mean $b/(a-1)$ for $a>1$ and variance $b^2/[(a-2)(a-1)^2]$ for $a>2$;
$f_{IW}(X|V,q)$ – the probability density function of the inverse Wishart distribution with a $(p \times p)$ scale matrix V and q degrees of freedom;
$f_N(x|a,b)$ – the probability density function of the normal distribution with mean a and variance b;
$f_G(x|a,b)$ – the probability density function of the gamma distribution with mean a/b and variance a/b^2;
$f_{Exp}(x|\lambda)$ – the probability density function of the exponential distribution with mean $1/\lambda$;
$f_{Be}(x|a,b)$ – the probability density function of the beta distribution with $a,\,b>0$;
$I_{(a,\,b)}(x)$ – the indicator function of the interval $(a,\,b)$.

- the uniform distribution over the unit square $[0,\,1]^2$ for a_{11} and g_{11}, truncated by the restriction $a_{11}^2 + g_{11}^2 < 1$: $p\,(a_{11},\, g_{11}) \propto I_{[0,\,1)}\left(a_{11}^2 + g_{11}^2\right)$,

- the uniform distribution over the square $[-1,\,1]^2$ for a_{ii} and g_{ii}, truncated by the restriction $a_{ii}^2 + g_{ii}^2 < 1$: $p\,(a_{ii},\, g_{ii}) \propto I_{[0,1)}\left(a_{ii}^2 + g_{ii}^2\right)$, $i = 2, 3$,

- the exponential distribution for σ_0^2: $p\left(\sigma_0^2\right) = f_{Exp}(\sigma_0^2|1)$, so $E(\sigma_0^2) = 1$.

Finally, for the volatility parameters in the VEC-MSH models, we specify the following priors (see Kwiatkowski, 2020a,b):

- the inverse Wishart distribution for Σ_i: $p\,(\Sigma_i) = f_{IW}\,(\Sigma_i|I_3, 5)$, so $E(\Sigma_i) = I_3$, $i = 1, 2$,

- the beta distribution for p_{11} and p_{22}: $p\,(p_{ii}) = f_{Be}\,(p_{ii}|1,\,1)$, $i = 1, 2$.

As regards the initial conditions for $\tilde{\Sigma}_t$, we take $\tilde{\Sigma}_0 = \sigma_0^2 I_3$ and treat $\sigma_0^2 > 0$ as an additional parameter, exponentially distributed a priori with mean one, whereas the initial value of q_t, i.e. q_0, is assumed to be equal to one. As the initial conditions for y_t, the first two vectors of observations are used: y_{-1} and y_0. Finally, the initial state of the latent Markov chain in the VEC-MSH models, S_0, is modelled as a binary random variable with probability $p_0 = p(S_0 = 1)$ treated as an additional parameter with a uniform prior over $[0,\,1]$.

For the univariate AR(1) models, the priors for Γ_1 (now a diagonal matrix), Σ (also diagonal) and, as in the case of the Polish data, Φ need to be specified. To that end, and to ensure the prior coherence (in the sense presented by Poirier, 1985) with the other models considered in this paper, the distributions: $p\,(\Gamma_1|h)$ and $p\,(\Sigma)$ are induced from the ones introduced above under zero restrictions for the off-diagonal elements. Specifically:

- the normal distribution for each diagonal element of Γ_1, i.e. $\Gamma_{1,ii}$, $i = 1, 2, 3$: $p\,(\Gamma_{1,ii}|h) = f_N\,(\Gamma_{1,ii}|0, h)$, with the hyperparameter h being the prior variance (common to all the coordinates of $diag(\Gamma_1)$),

- the inverse gamma distribution for each diagonal element of Σ, i.e. Σ_{ii}, $i = 1, 2, 3$: $p\,(\Sigma_{ii}) = f_{IG}\,(\Sigma_{ii}|3, 0.5)$, so $E(\Sigma_{ii}) = 0.25$ and $Var(\Sigma_{ii}) = 0.0625$.

Finally, $p\,(\Phi)$ remains the same as specified previously, i.e. $p\,(\Phi|h_s) = f_{mN}(\Phi|0, I_3, h_s I_{l_s})$.

In the univariate random walks, Γ_1 is further assumed to be the zero matrix, while $p\,(\Sigma_{ii})$, $i = 1, 2, 3$, and $p\,(\Phi)$ remain the same as for the AR(1) models.

Note that in the AR(1) models, the prior distributions of $\Gamma_{1,ii}$, $i = 1, 2, 3$, share the same hyperparameter h. Therefore, the models are only conditionally independent. Otherwise, considering unconditionally independent models would break the prior coherence. A similar reasoning pertains also to the seasonal parameters, Φ, sharing a common hyperparameter, h_s, across the individual AR(1) (or RW) equations. Incidentally, due to the lack of the seasonal component, the RW models for the US data do enjoy the unconditional independence.

In all models under study, the posterior distributions of the parameters are non-standard and intractable for analytical computations, thereby necessitating a use of Markov Chain Monte Carlo (MCMC) simulations, combining the Gibbs sampler and the Metropolis-Hastings algorithm. The conditional mean parameters in the VEC–LN-MSF-DBEKK and VEC–IG-MSF-DBEKK models are drawn in the same manner as in Pajor and Wróblewska (2017, 2022) and Wróblewska and Pajor (2019), while the volatility parameters are sampled following Osiewalski and Pajor (2018, 2019). Finally, to sample the VEC–MSH models' parameters we employ Jochmann and Koop (2015) approach, hinged on the Forward-Filtering-Backward-Sampling (FFBS) scheme designed by Carter and Kohn (1994) and Chib (1996).

6.2 Results of forecast comparison

All models considered here were re-estimated at a quarterly frequency. In each model, the results are based on 200000 posterior samples, preceded by 20000 burnt-in cycles. Table 10 displays the average log predictive Bayes factor of order k for each model for the US data and for selected forecast horizons: $k = 1, 4, 8$ (for brevity). Also, except for models with two cointegration relations, for both economies we additionally report results on two model specifications with zero long-run relationships: the one with the volatility structure that for $k = 1$ emerged the best for a given economy (i.e. VEC-t-DBEKK(0) for the US, and VEC-LN-MSF-DBEKK(0) for Poland), and the basic homoscedastic, conditionally normal VEC, denoted as VEC(0). Note that the latter is equivalent to a simple VAR(1) process for the first differences of the modelled variables.

For the US data, it is the VEC-t-DBEKK(2) specification that turned out the best for $k = 1$ and $k = 2$. Surprisingly, this model ranks only 8^{th} for the 4- to 8-step-ahead forecasts. For $k = 4, 6, 7$, the first position in the ranking is occupied by the VEC-LN-MSF-SBEKK(2) model. In turn, the VEC-LN-MSF(2) specification is the best for $k = 5, 8$. For $k = 3$, the best model is VEC-LN-MSF-DBEKK(2). For $k = 1$, the second position is occupied by VEC-t-DBEKK(0), but for $k = 2$ it ranks only 7^{th}. In the case of higher forecast horizons ($k = 3, \dots, 8$), the predictive performance of the VEC-t-DBEKK(0) model deteriorates dramatically, by taking $10 - 13$ positions. Further, note only a poor performance of the Markov-switching specification, ranked 13 (for $k = 1, 4$) and 10 (for $k = 8$), indicating that for the sake of prediction, allowing for only discrete volatility breaks is way insufficient. Finally, for each horizon, the ranking closes with the bundles of (conditionally independent) AR(1) and RW models, which remains in line with our expectations.

As seen in Table 10, the rank correlation coefficient drops markedly with the increase of the forecast horizon. Therefore, it may be concluded that when the effect of updating the posterior odds ratios is disabled, the ranking of the models depends on the forecast horizon. However, the rankings are more stable across the forecast horizons while using the average log predictive Bayes factors of order (k, s) than while using the average log predictive Bayes factors of order k, see Table 11. For each forecast horizon,

	Forecast horizon and rank					
Model	$k=1$	rank	$k=4$	rank	$k=8$	rank
VEC-LN-MSF-DBEKK(2)	0.071	5	0.004	2	0.058	6
VEC-IG-MSF-DBEKK(2)	0.051	4	0.020	5	0.064	7
VEC-LN-MSF-SBEKK(2)	0.107	9	**0**	**1**	0.004	2
VEC-IG-MSF-SBEKK(2)	0.084	7	0.025	7	0.016	3
VEC-LN-MSF(2)	0.125	11	0.008	3	**0**	**1**
VEC-IG-MSF(2)	0.121	10	0.011	4	0.02	4
VEC-t-DBEKK(2)	**0**	**1**	0.038	8	0.084	8
VEC-t-SBEKK(2)	0.036	3	0.024	6	0.025	5
VEC-DBEKK(2)	0.082	6	0.079	10	0.123	12
VEC-SBEKK(2)	0.203	12	0.076	9	0.098	9
VEC-MSH(2)	0.298	13	0.178	13	0.100	10
t-VEC(2)	0.097	8	0.081	11	0.110	11
VEC(2)	0.426	14	0.300	15	0.184	15
VEC-t-DBEKK(0)	0.035	2	0.102	12	0.123	13
VEC(0)	0.428	15	0.260	14	0.164	14
cond. indep. AR(1)	0.479	16	0.432	16	0.242	16
cond. indep. RW	0.559	17	0.549	17	0.640	17
corrRank ($k=1$; $k=j$)		1.000		0.544		0.436

Table 10: Average log predictive Bayes factor of order k, $\frac{1}{n+1}\sum_{s=0}^{n}\log B_{\text{the best m.},j}(y^o_{T+k+s}|y_1^{T+s,o})$, and ranks for selected forecast horizons. The US data, $n=48$. The number of cointegrating relations is given in parentheses.

the VEC-t-DBEKK(2) model ranks first, followed by VEC-t-DBEKK(0). In turn, the VEC-LN-MSF(2) specification is only 11^{th} in the rankings. Now the rank correlation coefficients are very close to one (see the last row in Table 11).

Since the average log predictive Bayes factor of order one ($k=1$) for two the best models (i.e. VEC-t-DBEKK(2) and VEC-t-DBEKK(0)) is 0.572 (and $10^{0.572}\approx 3.7$), the first observation from $T+s$, i.e. y^o_{T+1+s}, and all observations from T to $T+s$, for $s=0,\dots,48$, increase the posterior odds (based on the data up to time T) of the two competing models about 3.7 times, on average. In the forecasting period, the average increase of the posterior odds (obtained on the basis of the data up to time T) of the VEC-t-DBEKK(2) model versus VEC-t-DBEKK(0) for $k=4$ and $k=8$ is equal to about 0.601 and 0.577 orders of magnitude, respectively. In other words, the average posterior odds ratio is about 4 and 3.8 times higher, respectively, than the initial odds ratio. For the 4-step-ahead forecasts, the average of the posterior odds ratio in favour of VEC-t-DBEKK(2) and against VEC-LN-MSF(2) for the data sets up to time $T+s$ and including also the $(T+s+4)$-th observation, $(y^o_1,\dots,y^o_T,\dots,y^o_{T+s},y^o_{T+4+s})$, where $s=0,\dots,48$, is about 548.3 times higher than the posterior odds ratio for the data set up to time T. Similarly, for the 8-step-ahead forecasts, the average posterior odds ratio in favour of VEC-t-DBEKK(2) and against VEC-LN-MSF(2) for the data sets up to time $T+s$ and including also y^o_{T+s+8}, where $s=0,\dots,48$, is about 486.4 times higher than the posterior odds ratio for the data set up to time T.

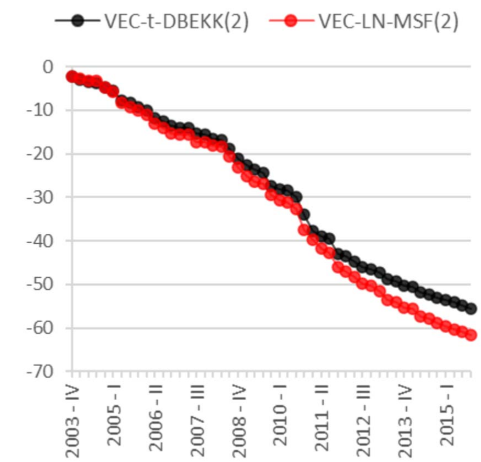
	Forecast horizon and rank					
Model	$k=1$	rank	$k=4$	rank	$k=8$	rank
VEC-LN-MSF-DBEKK(2)	1.561	5	1.456	5	1.466	5
VEC-IG-MSF-DBEKK(2)	1.129	4	1.059	4	1.059	4
VEC-LN-MSF-SBEKK(2)	2.344	9	2.199	8	2.158	8
VEC-IG-MSF-SBEKK(2)	1.785	6	1.688	6	1.634	6
VEC-LN-MSF(2)	2.895	11	2.739	11	2.687	11
VEC-IG-MSF(2)	2.856	10	2.708	10	2.672	10
VEC-t-DBEKK(2)	**0**	**1**	**0**	**1**	**0**	**1**
VEC-t-SBEKK(2)	0.757	3	0.707	3	0.663	3
VEC-DBEKK(2)	1.848	7	1.807	7	1.805	7
VEC-SBEKK(2)	4.484	12	4.319	12	4.296	12
VEC-MSH(2)	7.254	13	7.096	13	6.972	13
t-VEC(2)	2.275	8	2.222	9	2.206	9
VEC(2)	9.717	14	9.553	14	9.392	14
VEC-t-DBEKK(0)	0.572	2	0.601	2	0.577	2
VEC(0)	9.770	15	9.564	15	9.422	15
cond. indep. AR(1)	11.830	16	11.746	16	11.510	16
cond. indep. RW	13.522	17	13.473	17	13.520	17
corrRank ($k=1$; $k=j$)		1.000		0.998		0.995

Table 11: Average log predictive Bayes factor of order (k,s), $\frac{1}{n+1}\left[\sum_{s=0}^{n}\log B_{\text{the best m.},j}(y^o_{T+k+s}|y_1^{T+s,o}) + \sum_{s=1}^{n}\log B_{\text{the best m.},j}(y_{T+1}^{T+s,o}|y_1^{T,o})\right]$, and ranks for selected forecast horizons. The US data, $n=48$. The number of cointegrating relations is given in parentheses.

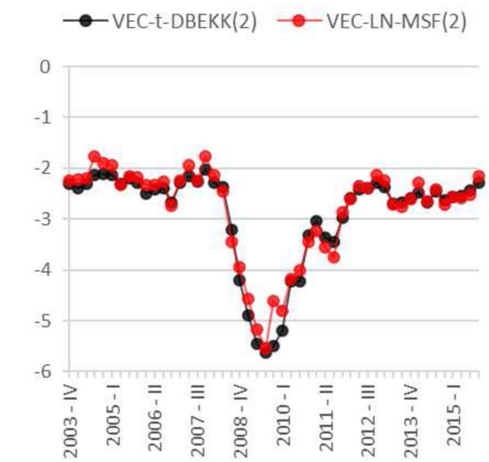
In Figure 4, for the two models discussed above we present the cumulative log predictive likelihoods for the 8-step-ahead forecasts. As can be seen, for each $s\in\{4,\dots,48\}$:

$$p(y^o_{T+s+8}, y_{T+1}^{T+s,o}|y_1^{T,o}, \text{VEC-t-DBEKK(2)}) > p(y^o_{T+s+8}, y_{T+1}^{T+s,o}|y_1^{T,o}, \text{VEC-LN-MSF(2)}).$$

Consequently, the predictive power of the VEC-t-DBEKK(2) model within the whole period $T{+}5,\dots,T{+}49$ and at $T{+}49{+}8$ dominates the VEC-LN-MSF(2) model. In turn, the log predictive likelihoods $p(y^o_{T+s+8}|y_1^{T+s,o}, \text{VEC-t-DBEKK(2)})$ for $s=0,\dots,48$ are higher than those for the VEC-LN-MSF(2) model only in 16 out of 49 cases, and, in consequence, the average of the log predictive likelihoods for the VEC-t-DBEKK(2) model is lower than that for VEC-LN-MSF(2). Thus, as we can see from Table 11, the VEC-t-DBEKK(2) model fits the predicted data better than the remaining models (in terms of the average log predictive Bayes factor of order (k,s)). However, the conclusion is different when we use the average predictive Bayes factor of order k. For example, for $k=8$ the predictive power of the VEC-LN-MSF(2) model dominates all remaining models under consideration, see Table 10. The updating component in the log predictive Bayes factors of order (k,s) has a major impact on the rankings. In fact, the predictive Bayes factors of order (k,s) and the corresponding predictive Bayes factors for the period of length s produce similar results, as can be inferred from Tables 11–12.



(a) Cumulative log predictive likelihood: $\log p(y^o_{T+k+s}, y^{T+s,o}_{T+1} | y^{T,o}_1, M_i)$



(b) Log predictive likelihood: $\log p(y^o_{T+k+s} | y^{T+s,o}_1, M_i)$

Figure 4: Cumulative log predictive likelihood and log predictive likelihood for $k = 8$. The US data, $n = 48$.

Model	$k = 1 \ldots, 8$	Rank
VEC-LN-MSF-DBEKK(2)	1.490	5
VEC-IG-MSF-DBEKK(2)	1.078	4
VEC-LN-MSF-SBEKK(2)	2.237	9
VEC-IG-MSF-SBEKK(2)	1.701	6
VEC-LN-MSF(2)	2.770	11
VEC-IG-MSF(2)	2.735	10
VEC-t-DBEKK(2)	**0**	**1**
VEC-t-SBEKK(2)	0.721	3
VEC-DBEKK(2)	1.766	7
VEC-SBEKK(2)	4.281	12
VEC-MSH(2)	6.956	13
t-VEC(2)	2.178	8
VEC(2)	9.292	14
VEC-t-DBEKK(0)	0.537	2
VEC(0)	9.342	15
cond. indep. AR(1)	11.351	16
cond. indep. RW	12.963	17

Table 12: Updating effect measured by $\frac{1}{n+1}\sum_{s=1}^{n} \log B_{\text{the best m.,j}}(y^{T+s,o}_{T+1} | y^{T,o}_1)$. The US data, $n = 48$. The number of cointegrating relations is given in parentheses.

	Forecast horizon and rank					
Model	$k=1$	rank	$k=4$	rank	$k=8$	rank
VEC-LN-MSF-DBEKK(2)	**0**	**1**	0.025	2	0.056	4
VEC-IG-MSF-DBEKK(2)	0.106	5	0.063	6	0.069	6
VEC-LN-MSF-SBEKK(2)	0.028	2	0.052	4	0.086	8
VEC-IG-MSF-SBEKK(2)	0.169	10	0.205	11	0.229	11
VEC-LN-MSF(2)	0.108	6	**0**	**1**	**0**	**1**
VEC-IG-MSF(2)	0.142	9	0.059	5	0.04	3
VEC-t-DBEKK(2)	0.079	4	0.036	3	0.027	2
VEC-t-SBEKK(2)	0.131	7	0.192	10	0.221	10
VEC-DBEKK(2)	0.140	8	0.079	7	0.066	5
VEC-SBEKK(2)	0.209	11	0.216	12	0.241	14
VEC-MSH(2)	0.733	15	0.541	15	0.236	13
t-VEC(2)	0.211	12	0.121	8	0.081	7
VEC(2)	0.441	13	0.266	13	0.177	9
VEC-LN-MSF-DBEKK(0)	0.037	3	0.155	9	0.234	12
VEC(0)	0.460	14	0.346	14	0.326	15
cond. indep. AR(1)	0.790	16	1.047	17	0.692	17
cond. indep. RW	1.074	17	0.573	16	0.411	16
corrRank ($k=1$; $k=j$)		1.000		0.860		0.672

Table 13: Average log predictive Bayes factor of order k, $\frac{1}{n+1}\sum_{s=0}^{n} \log B_{\text{the best m.,j}}(y^o_{T+k+s}|y_1^{T+s,o})$, and ranks for selected forecast horizons. The Polish data, $n=22$. The number of cointegrating relations is given in parentheses.

As regards the predictive ability of the models considered for Poland, the best predictive model, in terms of the average log predictive Bayes factor of order k, turns out to be VEC-LN-MSF-DBEKK(2), although only for $k=1,2,3$. On the other hand, the very same specification proves the best also according to the average log predictive Bayes factor of order (k,s), yet now for all of the considered forecast horizons (see Tables 13–15 and Figure 5). The ranking with respect to the average log predictive Bayes factor of order k for $k=4,5,6,7,8$ is topped by the VEC-LN-MSF(2) model. It follows from Table 14 that in terms of the average log predictive Bayes factors of order (k,s), the VEC-LN-MSF(2) specification ranks 4^{th} for all $k=1,\dots,8$. Generally, the rankings of the models vary depending on which of the components of the log predictive Bayes factor considered here is used. Taking into account both the k-step-ahead predictive ability and the updating effect leads to different results from those based only on one of the components of the log predictive Bayes factor of order (k,s).

Surprisingly, the Markov-switching model turns out to be outperformed by most of the specifications under consideration, regardless of the forecasting performance measure. Moreover, it almost always falls (way) behind even the homoscedastic VECs (except for VEC(0) for $k=8$ in terms of the average log predictive Bayes factor of order k, see Table 13). Nevertheless, the rankings according to both of the measures are still closed by the bundles of the AR(1) and RW models throughout.

	Forecast horizon and rank					
Model	$k=1$	rank	$k=4$	rank	$k=8$	rank
VEC-LN-MSF-DBEKK(2)	0	1	0.000	1	0	1
VEC-IG-MSF-DBEKK(2)	1.458	7	1.390	7	1.365	7
VEC-LN-MSF-SBEKK(2)	0.504	3	0.503	3	0.506	3
VEC-IG-MSF-SBEKK(2)	2.132	10	2.144	10	2.136	10
VEC-LN-MSF(2)	0.983	4	0.850	4	0.819	4
VEC-IG-MSF(2)	1.398	6	1.291	6	1.24	6
VEC-t-DBEKK(2)	1.222	5	1.154	5	1.114	5
VEC-t-SBEKK(2)	2.046	9	2.082	9	2.080	9
VEC-DBEKK(2)	1.962	8	1.876	8	1.832	8
VEC-SBEKK(2)	2.973	12	2.955	12	2.949	12
VEC-MSH(2)	9.112	15	8.895	15	8.559	15
t-VEC(2)	2.442	11	2.327	11	2.256	11
VEC(2)	5.251	13	5.052	13	4.931	13
VEC-LN-MSF-DBEKK(0)	0.318	2	0.411	2	0.459	2
VEC(0)	5.488	14	5.349	14	5.297	14
cond. indep. AR(1)	9.514	16	9.746	16	9.360	16
cond. indep. RW	13.175	17	12.648	17	12.456	17
corrRank ($k=1$; $k=j$)		1.000		1.000		1.000

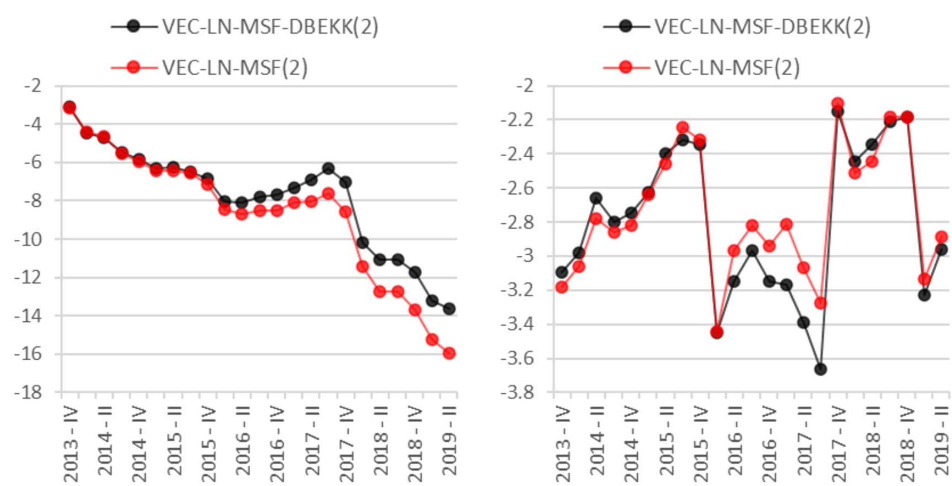
Table 14: Average log predictive Bayes factor of order (k,s), $\frac{1}{n+1}\left[\sum_{s=0}^{n}\log B_{\text{the best m.,j}}(y^{o}_{T+k+s}|y_1^{T+s,o})+\sum_{s=1}^{n}\log B_{\text{the best m.,j}}(y^{T+s,o}_{T+1}|y_1^{T,o})\right]$, and ranks for selected forecast horizons. The Polish data, $n=22$. The number of cointegrating relations is given in parentheses.

7 Conclusions

In the paper, a new measure is proposed for comparing the forecasting performance of Bayesian models for n consecutive k-step-ahead forecasts. Also, we introduced a new decomposition of the predictive Bayes factor into the product of partial Bayes factors accounting for both a finite number of consecutive k-step-ahead forecasts and recursive updates of the posterior odds ratios based on updated data sets. The simulation results suggest that the predictive Bayes factor of order (k,s) introduced in the paper allows to identify the model based on the true data generating process. The empirical results show that the rankings of the models vary depending on which of the measures (i.e. the components of the log predictive Bayes factor discussed in the paper) is used. Taking into account both the predictive ability for the k-step-ahead forecasts and the effect of updating the data set leads to different results from those based only on one of the components of the log predictive Bayes factor of order (k,s). Very similar results and highest rank correlations are found for the average log predictive Bayes factors of order (k,s), which accounts for both the predictive ability for the k-step-ahead forecasts and the updating effect in a forecasting period of length s, and the log predictive Bayes factor for the period from $T+1$ to $T+s$, which accounts only for the updating effect. Thus, the latter has a major impact on such Bayesian model comparison that takes

Model	$k = 1, \ldots, 8$	Rank
VEC-LN-MSF-DBEKK(2)	0.000	1
VEC-IG-MSF-DBEKK(2)	1.116	5
VEC-LN-MSF-SBEKK(2)	0.476	3
VEC-IG-MSF-SBEKK(2)	1.964	9
VEC-LN-MSF(2)	1.316	7
VEC-IG-MSF(2)	2.048	10
VEC-t-DBEKK(2)	0.840	4
VEC-t-SBEKK(2)	1.223	6
VEC-DBEKK(2)	1.822	8
VEC-SBEKK(2)	2.137	11
VEC-MSH(2)	8.379	15
t-VEC(2)	2.859	12
VEC(2)	4.811	13
VEC-LN-MSF-DBEKK(0)	0.281	2
VEC(0)	5.028	14
cond. indep. AR(1)	8.724	16
cond. indep. RW	12.101	17

Table 15: Updating effect measured by $\frac{1}{n+1}\sum_{s=1}^{n} \log B_{\text{the best m.,j}}(y_{T+1}^{T+s,o}|y_1^{T,o})$. The Polish data, $n = 22$. The number of cointegrating relations is given in parentheses.



(a) Cumulative log predictive likelihood: $\log p(y^o_{T+k+s}, y_{T+1}^{T+s,o}|y_1^{T,o}, M_i)$

(b) Log predictive likelihood: $\log p(y^o_{T+k+s}|y_1^{T+s,o}, M_i)$

Figure 5: Cumulative log predictive likelihood and log predictive likelihood for $k = 8$. The Polish data, $n = 22$.

into account both the k-step-ahead forecasting ability of models and the necessity of updating the observations (model re-estimation) in empirical assessment of predictive performance.

Both the simulation as well as empirical results presented in this paper may prompt further, but now theoretical, research on Bayesian model choice consistency (in the sense defined, e.g, by Berger and Pericchi (2001)) through the predictive Bayes factor of order (k, s), proposed here. Conceivably, providing rigorous proofs in that matter, even for some simple and popular econometric models, may be far from straightforward, yet most desirable.

References

Berger, J. O. and Pericchi, L. R. (2001). "Objective Bayesian methods for model selection: Introduction and comparison." In Lahiri, P. (ed.), *Model Selection. Institute of Mathematical Statistics Lecture Notes—Monograph Series*, volume 38, 135–207. Institute of Mathematical Statistics. MR2000753. doi: https://doi.org/10.1214/lnms/1215540968. 780

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*. New York: Wiley. MR1274699. doi: https://doi.org/10.1002/9780470316870. 752, 753

Bissiri, P. G., Holmes, C. C., and Walker, S. G. (2016). "A general framework for updating belief distributions." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(5): 1103–1130. MR3557191. doi: https://doi.org/10.1111/rssb.12158. 752

Brier, G. W. (1950). "Verification of forecasts expressed in terms of probability." *Monthly weather review*, 78(1): 1–3. 753

Carter, C. K. and Kohn, R. (1994). "On Gibbs sampling for state space models." *Biometrika*, 81(3): 541–553. MR1311096. doi: https://doi.org/10.1093/biomet/81.3.541. 773

Carvalho, A. (2016). "An overview of applications of proper scoring rules." *Decision Analysis*, 13(4): 223–242. MR3592366. doi: https://doi.org/10.1287/deca.2016.0337. 752

Chib, S. (1996). "Calculating posterior distributions and modal estimates in Markov mixture models." *Journal of Econometrics*, 75(1): 79–97. MR1414504. doi: https://doi.org/10.1016/0304-4076(95)01770-4. 773

Chikuse, Y. (2002). *Statistics on special manifolds*, volume 174. Springer Science & Business Media. MR1960435. doi: https://doi.org/10.1007/978-0-387-21540-2. 771

Cross, J. L., Hou, C., and Poon, A. (2020). "Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity." *International Journal of Forecasting*, 36(3): 899–915. 752, 758

Dawid, A. P. and Sebastiani, P. (1999). "Coherent dispersion criteria for optimal experimental design." *Annals of Statistics*, 65–81. MR1701101. doi: https://doi.org/10.1214/aos/1018031101. 752

Geweke, J. (2005). *Contemporary Bayesian econometrics and statistics*, volume 537. John Wiley & Sons. MR2166278. doi: https://doi.org/10.1002/0471744735. 752, 755, 757, 760

Geweke, J. and Amisano, G. (2010). "Comparing and evaluating Bayesian predictive distributions of asset returns." *International Journal of Forecasting*, 26(2): 216–230. 752, 756

Geweke, J. and Amisano, G. (2011). "Optimal prediction pools." *Journal of Econometrics*, 164(1): 130–141. MR2821798. doi: https://doi.org/10.1016/j.jeconom.2011.02.017. 760

Geweke, J. and Whiteman, C. (2006). "Bayesian forecasting." In Elliott, G., Granger, C. W. J., and Timmermann, A. G. (eds.), *Handbook of economic forecasting*, 3–8. Amsterdam: North Holland: Elsevier. 752

Giannone, D., Lenza, M., and Primiceri, G. E. (2015). "Prior selection for vector autoregressions." *Review of Economics and Statistics*, 97(2): 436–451. 752

Gneiting, T. and Raftery, A. E. (2007). "Strictly proper scoring rules, prediction, and estimation." *Journal of the American statistical Association*, 102(477): 359–378. MR2345548. doi: https://doi.org/10.1198/016214506000001437. 752, 753, 756, 758

Good, I. J. (1952). "Rational Decisions." *Journal of the Royal Statistical Society Series B*, 14: 107–114. MR0077033. 752, 753

Hauzenberger, N., Huber, F., Pfarrhofer, M., and Zörner, T. O. (2021). "Stochastic model specification in Markov switching vector error correction models." *Studies in Nonlinear Dynamics & Econometrics*, 25(2): 1–17. MR4255821. doi: https://doi.org/10.1515/snde-2018-0069. 769

Jeffreys, H. (1961). *The theory of probability*. Clarendon Press, Oxford, 3 edition. MR0187257. 752

Jochmann, M. and Koop, G. (2015). "Regime-switching cointegration." *Studies in Nonlinear Dynamics & Econometrics*, 19(1): 35–48. MR3315033. doi: https://doi.org/10.1515/snde-2012-0064. 773

Koop, G. and Korobilis, D. (2012). "Forecasting inflation using dynamic model averaging." *International Economic Review*, 53(3): 867–886. MR2959266. doi: https://doi.org/10.1111/j.1468-2354.2012.00704.x. 752, 758

Koop, G., León-González, R., and Strachan, R. W. (2009). "Efficient posterior simulation for cointegrated models with priors on the cointegration space." *Econometric Reviews*, 29(2): 224–242. MR2747499. doi: https://doi.org/10.1080/07474930903382208. 770

Kwiatkowski, Ł. (2020a). “Bayesian VEC models with Markov-switching heteroscedasticity in forecasting macroeconomic time series.” In Papież, M. and Śmiech, S. (eds.), *The 14th Professor Aleksander Zeliaś International Conference on Modelling and Forecasting of Socio-Economic Phenomena*, 78–87. 768, 770, 772

Kwiatkowski, Ł. (2020b). “Predictive Power Comparison of Bayesian Homoscedastic vs. Markov-switching Heteroscedastic VEC Models.” In Ciałowicz, B. (ed.), *Quantitative methods in the contemporary issues of economics*, 67–76. 768, 770, 772

Loaiza-Maya, R., Martin, G. M., and Frazier, D. T. (2021). “Focused Bayesian prediction.” *Journal of Applied Econometrics*, 36(5): 517–543. MR4309597. doi: https://doi.org/10.1002/jae.2810. 752

Morey, R. D., Romeijn, J.-W., and Rouder, J. N. (2016). “The philosophy of Bayes factors and the quantification of statistical evidence.” *Journal of Mathematical Psychology*, 72: 6–18. MR3506021. doi: https://doi.org/10.1016/j.jmp.2015.11.001. 755

Murphy, A. H. and Winkler, R. L. (1970). “Scoring rules in probability assessment and evaluation.” *Acta psychologica*, 34: 273–286. 752

O’Hagan, A. (1995). “Fractional Bayes factors for model comparison.” *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1): 99–118. MR1325379. 756

Osiewalski, J. (2009). “New hybrid models of multivariate volatility (a Bayesian perspective).” *Przeglad Statystyczny*, 56(1): 15–22. 768

Osiewalski, J. and 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. 768

Osiewalski, J. and Pajor, A. (2018). “A hybrid MSV-MGARCH generalisation of the t-MGARCH model.” In Papież, M. and Śmiech, S. (eds.), *The 12th Professor Aleksander Zeliaś International Conference on Modelling and Forecasting of Socio-Economic Phenomena: Conference Proceedings*, 1, 345–354. 768, 773

Osiewalski, J. and Pajor, A. (2019). “On sensitivity of Inference in Bayesian MSF-MGARCH models.” *Central European Journal of Economic Modelling and Econometrics*, 3. 768, 771, 773

Osiewalski, J. and Steel, M. F. (1993). “Una perspectiva bayesiana en selección de modelos (A Bayesian perspective on model selection).” *Cuadernos económicos de ICE*, 55: 327–351.
URL https://home.cyf-kr.edu.pl//~eeosiewa. 755

Pajor, A. and Wróblewska, J. (2017). “VEC-MSF models in Bayesian analysis of short- and long-run relationships.” *Studies in Nonlinear Dynamics & Econometrics*, 21(3). MR3665986. doi: https://doi.org/10.1515/snde-2016-0004. 764, 771, 773

Pajor, A. and Wróblewska, J. (2022). “Forecasting performance of Bayesian VEC-MSF models for financial data in the presence of long-run relationships.” *Eurasian Economic Review*, 1–22. 773

Pajor, A., Wróblewska, J., Kwiatkowski, Ł., and Osiewalski, J. (2022). "Hybrid MSV-MGARCH, t-GARCH and Markov Switching covariance structures in VEC models – which is better from a predictive perspective?" *manuscript, under review in International Statistical Review*. 763, 764, 765, 766, 768, 769, 771

Piironen, J. and Vehtari, A. (2017). "Comparison of Bayesian predictive methods for model selection." *Statistics and Computing*, 27(3): 711–735. MR3613594. doi: https://doi.org/10.1007/s11222-016-9649-y. 752

Poirier, D. (1985). "Bayesian hypothesis testing in linear models with continuously induced conjugate priors across hypotheses." *Bayesian Statistics*, 2: 711–722. MR0862514. 772

Primiceri, G. E. (2005). "Time varying structural vector autoregressions and monetary policy." *The Review of Economic Studies*, 72(3): 821–852. MR2148143. doi: https://doi.org/10.1111/j.1467-937X.2005.00353.x. 764

Savage, L. J. (1971). "Elicitation of personal probabilities and expectations." *Journal of the American Statistical Association*, 66(336): 783–801. MR0331571. 753

Vehtari, A. and Ojanen, J. (2012). "A survey of Bayesian predictive methods for model assessment, selection and comparison." *Statistics Surveys*, 6: 142–228. MR3011074. doi: https://doi.org/10.1214/12-SS102. 752, 753

Vehtari, A. and Ojanen, J. (2014). "Errata: A survey of Bayesian predictive methods for model assessment, selection and comparison." *Statistics Surveys*, 8: 1–1. MR3178542. doi: https://doi.org/10.1214/14-SS105. 752, 753

Warne, A., Coenen, G., and Christoffel, K. P. (2013). "Predictive likelihood comparisons with DSGE and DSGE-VAR models." *ECB Working Paper*. 752

Winkler, R. L. (1996). "Scoring rules and the evaluation of probabilities." *Test*, 5(1): 1–60. MR1410455. doi: https://doi.org/10.1007/BF02562681. 752, 753

Winkler, R. L. and Murphy, A. H. (1968). "“Good” probability assessors." *Journal of Applied Meteorology and Climatology*, 7(5): 751–758. 752

Wróblewska, J. and 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. 764, 773

Wu, J. C. and Xia, F. D. (2016). "Measuring the macroeconomic impact of monetary policy at the zero lower bound." *Journal of Money, Credit and Banking*, 48(2–3): 253–291. 765

Yao, Y., Vehtari, A., Simpson, D., and Gelman, A. (2018). "Using stacking to average Bayesian predictive distributions (with discussion)." *Bayesian Analysis*, 13(3): 917–1007. MR3853125. doi: https://doi.org/10.1214/17-BA1091. 752

Acknowledgments

We are very grateful to the Associate Editor and two anonymous Reviewers for their valuable comments and suggestions which helped us to improve our paper.