Model averaging and variable selection in VAR models

Shutong Ding[*] Örebro University

Sune Karlsson[†] Örebro University

August 4, 2012

Abstract

Bayesian model averaging and model selection is based on the marginal likelihoods of the competing models. This can, however, not be used directly in VAR models when one of the issues is which - and how many - variables to include in the model since the likelihoods will be for different groups of variables and not directly comparable. One possible solution is to consider the marginal likelihood for a core subset of variables that are always included in the model. This is similar in spirit to a recent proposal for forecast combination based on the predictive likelihood. The two approaches are contrasted and their performance is evaluated in a simulation study and a forecasting exercise.

Keywords: Bayesian model averaging, marginalized likelihood, predictive likelihood

[*]Shutong.Ding@oru.se

[†]Sune.Karlsson@oru.se

1 Introduction

VAR models, in particular in their Bayesian flavor, have become a work horse model for macroeconomic forecasting and are also extensively used for policy analysis. In both cases the results can be highly dependent on the model specification. In a typical forecasting application the focus is on forecasting a few variables, for example GDP growth and inflation, and any additional variables are included because they are deemed to be important predictors for the variables of interest.

These are not clear cut choices and there is a multitude of models to choose from. Bayesian model averaging (BMA) or forecast combination where forecasts from individual models are averaged using the posterior model probabilities as weights is a natural solution to this dilemma. BMA will however run into fundamental difficulties as soon as one of the model specification issues is which variables should be modeled. The problem is that the (multivariate) likelihoods are no longer comparable when a variable is added, removed or variables are swapped in and out of the model. This in turn directly affects the marginal likelihoods which are the basis for calculating posterior model probabilities. There can thus be substantial differences in the marginal likelihoods and weights assigned to models with similar forecasting performance or similar marginal likelihoods and weights for models with quite different forecasting performance.

On possible solution, suggested by Andersson and Karlsson (2009), is to replace the marginal likelihood with the predictive likelihood for the variables of interest, that is after marginalizing out the other variables, in the calculation of posterior "probabilities" or model weights. This creates a focused measure and is attractive in a forecasting context since it directly addresses the forecasting performance of the different models. In this paper we propose an alternative approach – to base the model averaging on the marginalized marginal likelihood. The marginalization leads to a meaningful measure of the fit for the variables of interest that can be used to compare models while, by virtue of being based on the marginal likelihood and the full sample, it offers the promise of being able to distinguish more sharply between models than the marginalized predictive likelihood.

Our focus on the choice of variables to model differ from most of the (Bayesian) VAR literature which takes the set of variables as given when considering the model specification. The focus in the literature has largely been on devices for reducing the effective model size, e.g. by applying shrinkage as in the Minnesota prior of Litterman (1979, 1986), selecting lags of variables to include with the SSVS approach of George, Sun and Ni (2001) or summarizing a large number of variables by extracting the common factors as in the FAVAR of Bernanke, Boivin and Eliasz (2005). Notable exceptions are Andersson and Karlsson (2009), which we extend, and Jarociński and Maćkowiak (2011) who also base their approach on the marginal likelihood but in the context of a "super-model" containing all variables that are potentially useful in the forecasting or modeling exercise. Noting that block exogeneity restrictions on the super-model implies that a subset of the variables can be modeled separately Jarociński and Maćkowiak (2011) suggests basing the variable choice on the marginal likelihood for models imposing different block exogeneity restrictions. While interesting the procedure of Jarociński and Maćkowiak (2011) suffers from the potential drawback that it measures the overall fit of the super-model rather than focusing on the variables of interest.

The plan of the paper is as follows. Section two introduces the marginalized marginal and predictive likelihood and shows how they can be used in model averaging and model selection. Section three specializes the discussion to VAR models and gives details, including MCMC algorithms, on the calculation of the marginalized marginal likelihood for common prior distributions. In section four we conduct a small simulation exercise to evaluate and contrast the performance of the marginalized marginal and predictive likelihoods as basis for forecast combination and model selection. Section five contains an application to forecasting US GDP growth and inflation and section 6 concludes.

2 Bayesian model averaging and marginalized likelihoods

The standard Bayesian approach to model averaging obtains the posterior model probabilities for model $\mathcal{M}_i$, $i = 1, \ldots, M$, by straightforward application of Bayes rule,

$$p\left(\mathcal{M}_i | \mathbf{Y}\right) = \frac{m\left(\mathbf{Y} | \mathcal{M}_i\right) p\left(\mathcal{M}_i\right)}{\sum_{j=1}^{M} m\left(\mathbf{Y} | \mathcal{M}_j\right) p\left(\mathcal{M}_j\right)}$$

where $m\left(\mathbf{Y} | \mathcal{M}_i\right)$ is the *marginal* likelihood for model i,

$$m\left(\mathbf{Y} | \mathcal{M}_i\right) = \int L\left(\mathbf{Y} | \theta_i, \mathcal{M}_i\right) p\left(\theta_i | \mathcal{M}_i\right) d\theta_i.$$

$p\left(\mathcal{M}_i\right)$ the prior model probability, $p\left(\theta_i | \mathcal{M}_i\right)$ the prior on the parameters in model i and $L\left(\mathbf{Y} | \theta_i, \mathcal{M}_i\right)$ the model likelihood. This is straightforward and unproblematic when the same set of dependent variables $\mathbf{Y}$ enters into all the model likelihoods. If, as in our case, one of the fundamental model specification issues is which variables to include in the VAR models the use of the marginal likelihood becomes problematic as it no longer is comparable between models.

In forecasting and many other applications there is usually a core set of variables of primary interest that are retained in all the considered models. We can thus partition the matrix of dependent variables in model i, $\mathbf{Y}_i$, into the variables of primary interest $\mathbf{Y}_1$ and the "other" variables $\mathbf{Y}_{2i}$, $\mathbf{Y}_i = \left(\mathbf{Y}_1, \mathbf{Y}_{2i}\right)$. Our basic proposal is to focus on the variables of interest by marginalizing out $\mathbf{Y}_{2i}$. Building on the work of Eklund and Karlsson (2007), Andersson and Karlsson (2009) proposed marginalizing the predictive likelihood and base forecast combination and model selection on the predictive likelihood. This has the advantage for forecast combination that the measure is directly related to the out of sample predictive performance of the model but leads to several complications when applied to dynamic models where the marginalized predictive likelihood typically is not available in closed form. In this paper we propose to work with the marginalized marginal likelihood instead. This is basically an in-sample measure of fit and can make more efficient use of the data as the whole sample is used for model evaluation. The marginalized marginal likelihood is also more computationally convenient as it typically is available in closed form when conjugate priors are used.

2.1 Marginalized Predictive Likelihood

The predictive likelihood approach of Eklund and Karlsson (2007) is based on a split of the data, $\mathbf{Y} = (\mathbf{y}_1', \mathbf{y}_2', \ldots, \mathbf{y}_T')'$ where we drop the model specific subscript for notational simplicity, into two parts, the training sample, $\mathbf{Y}_n^* = (\mathbf{y}_1', \mathbf{y}_2', \ldots, \mathbf{y}_n')'$ of size n, and an evaluation or hold out sample, $\widetilde{\mathbf{Y}}_n = \left(\mathbf{y}_{n+1}', \mathbf{y}_{n+2}', \ldots, \mathbf{y}_T'\right)'$ of size $T - n$, where $\mathbf{y}_t$ is the vector of modeled variables. The training sample is used to convert the prior into a posterior and the predictive likelihood for the hold out sample is obtained by marginalizing out the parameters from the joint distribution of data and parameters,

$$p\left(\widetilde{\mathbf{Y}}_n \middle| \mathbf{Y}_n^*, \mathcal{M}_i\right) = \int L\left(\widetilde{\mathbf{Y}}_n \middle| \theta_i, \mathbf{Y}_n^*, \mathcal{M}_i\right) p\left(\theta_i | \mathbf{Y}_n^*, \mathcal{M}_i\right) d\theta_i.$$

Partitioning the hold out sample data into the variables of interest and the remaining variables, $\widetilde{\mathbf{Y}}_n = \left(\widetilde{\mathbf{Y}}_{1,n}, \widetilde{\mathbf{Y}}_{2,n}\right)$, the marginalized predictive likelihood is obtained by marginalizing out $\widetilde{\mathbf{Y}}_{2,n}$,

$$MPL\left(\widetilde{\mathbf{Y}}_{1,n} \middle| \mathbf{Y}_n^*, \mathcal{M}_i\right) = \int p\left(\widetilde{\mathbf{Y}}_n \middle| \mathbf{Y}_n^*, \mathcal{M}_i\right) d\widetilde{\mathbf{Y}}_{2,n}.$$

Predictive weights that can be used for model averaging or model selection are then calculated as

$$w\left(\mathcal{M}_i | \widetilde{\mathbf{Y}}_{1,n}, \mathbf{Y}_n^*\right) = \frac{MPL\left(\widetilde{\mathbf{Y}}_{1,n} \middle| \mathbf{Y}_n^*, \mathcal{M}_i\right) p\left(\mathcal{M}_i\right)}{\sum_{j=1}^{M} MPL\left(\widetilde{\mathbf{Y}}_{1,n} \middle| \mathbf{Y}_n^*, \mathcal{M}_j\right) p\left(\mathcal{M}_j\right)} \tag{1}$$

where $MPL\left(\widetilde{\mathbf{Y}}_{1,n} \middle| \mathbf{Y}_n^*, \mathcal{M}_i\right)$ is evaluated at the observed values of the variables of interest in the hold out sample.

While the predictive weights (1) strictly speaking can not be interpreted as posterior probabilities they have the advantage that proper prior distributions are not required for the parameters. The predictive likelihood is, in contrast to the marginal likelihood, well defined as long as the posterior distribution of the parameters conditioned on the training sample is proper.

The use of the predictive likelihood is complicated by the dynamic nature of VAR models. As noted by Andersson and Karlsson (2009) the predictive likelihood is the joint predictive distribution over lead times $h = 1$ to $T - n$. This will become increasingly uninformative for larger lead times and unrepresentative of lead times such as $h = 4$ or 8 usually considered in macroeconomic forecasting. At the same time the hold out sample needs to be relatively large in order to provide a sound basis for assessing the forecast performance of the models. To overcome this Andersson and Karlsson suggested focusing the measure to specific lead times $h_1, \ldots, h_k$ and using a series of predictive likelihoods,

$$g\left(\widetilde{\mathbf{Y}}_{1,n} | \mathcal{M}_i\right) = \prod_{t=n}^{T-h_k} MPL\left(y_{1,t+h_1}, \ldots, y_{1,t+h_k} | \mathbf{Y}_t^*, \mathcal{M}_i\right), \tag{2}$$

in the calculation of the predictive weights.

A final complication is that the predictive likelihood is not available in closed form for lead times $h > 1$ and must be estimated using simulation methods. With a normal likelihood the predictive likelihood for a VAR model and many other multivariate models will be normal *conditional* on the parameters and easy to evaluate. Andersson and Karlsson suggested estimating the multiple horizon marginalized predictive likelihood using a Rao-Blackwellization technique as

$$\widehat{MPL}\left(y_{1,t+h_1},\ldots,y_{1,t+h_k}\middle|\,\mathbf{Y}_t^*,\mathcal{M}_i\right)=\frac{1}{R}\sum_{i=1}^{R}p\left(y_{1,t+h_1},\ldots,y_{1,t+h_k}\middle|\,\mathbf{Y}_t^*,\mathcal{M}_i,\theta^{(i)}\right)$$

by averaging the conditional predictive likelihood $p\left(y_{1,t+h_1},\ldots,y_{1,t+h_k}\middle|\,\mathbf{Y}_t^*,\mathcal{M}_i,\theta^{(i)}\right)$ over draws of the parameters from the posterior distribution based on $\mathbf{Y}_t^*$. This leads to estimated predictive weights

$$\widehat{w}\left(\mathcal{M}_i\middle|\,\widetilde{\mathbf{Y}}_{1,n},\mathbf{Y}_n^*\right)=\frac{\widehat{g}\left(\widetilde{\mathbf{Y}}_{1,n}|\mathcal{M}_i\right)p\left(\mathcal{M}_i\right)}{\sum_{j=1}^{M}\widehat{g}\left(\widetilde{\mathbf{Y}}_{1,n}|\mathcal{M}_j\right)p\left(\mathcal{M}_j\right)}$$

with

$$\widehat{g}\left(\widetilde{\mathbf{Y}}_{1,n}|\mathcal{M}_i\right)=\prod_{t=n}^{T-h_k}\widehat{MPL}\left(y_{1,t+h_1},\ldots,y_{1,t+h_k}\middle|\,\mathbf{Y}_t^*,\mathcal{M}_i\right).$$

2.2 Marginalized Marginal Likelihood

With the marginalized marginal likelihood we start with the standard full sample marginal likelihood,

$$m\left(\mathbf{Y}|\,\mathcal{M}_i\right)=\int L\left(\mathbf{Y}|\,\theta_i,\mathcal{M}_i\right)p\left(\theta_i|\,\mathcal{M}_i\right)d\theta_i,$$

and marginalize out $\mathbf{Y}_2$ to obtain a measure focused on the variables of interest. With a slight abuse of notation[1] we write

$$MML\left(\mathbf{Y}_1|\mathcal{M}_i\right)=\int m\left(\mathbf{Y}|\mathcal{M}_i\right)d\mathbf{Y}_2$$

for the marginalized marginal likelihood. Posterior weights are then simply calculated as

$$w\left(\mathcal{M}_i|\,\mathbf{Y}\right)=\frac{MML\left(\mathbf{Y_1}|\,\mathcal{M}_i\right)p\left(\mathcal{M}_i\right)}{\sum_{j=1}^{M}MML\left(\mathbf{Y_1}|\,\mathcal{M}_j\right)p\left(\mathcal{M}_j\right)}.$$

With a conjugate prior the marginalized marginal likelihood will frequently be available in closed form whereas it must be estimated using simulation methods with non-conjugate priors. We give details of these calculations in the following section.

[1]The notation $MML\left(\mathbf{Y}_1|\mathcal{M}_i\right)=\int m\left(\mathbf{Y}|\mathcal{M}_i\right)d\mathbf{Y}_2$ is not entirely accurate in dynamic models where the marginal likelihood can be decomposed into a series of conditional distributions. To exemplify, with two time periods and $\mathbf{y}_t$ depending on $\mathbf{y}_{t-1}$ only we can write $m\left(\mathbf{Y}|\mathcal{M}_i\right)=p\left(\mathbf{y}_2|\mathbf{y}_1\right)p\left(\mathbf{y}_1\right)=p\left(\mathbf{y}_{1,2},\mathbf{y}_{2,2}|\mathbf{y}_{1,1},\mathbf{y}_{2,1}\right)p\left(\mathbf{y}_{1,1},\mathbf{y}_{2,1}\right)$. Our notation $\int m\left(\mathbf{Y}|\mathcal{M}_i\right)d\mathbf{Y}_2$ is shorthand for the operation $p\left(\mathbf{y}_{1,2}|\mathbf{y}_{1,1},\mathbf{y}_{2,1}\right)p\left(\mathbf{y}_{1,1}\right)=\int p\left(\mathbf{y}_{1,2},\mathbf{y}_{2,2}|\mathbf{y}_{1,1},\mathbf{y}_{2,1}\right)d\mathbf{y}_{2,2}\int p\left(\mathbf{y}_{1,1},\mathbf{y}_{2,1}\right)d\mathbf{y}_{2,1}$ and *not* $p\left(\mathbf{y}_{1,2}|\mathbf{y}_{1,1}\right)p\left(\mathbf{y}_{1,1}\right)=\int\int p\left(\mathbf{y}_{1,2},\mathbf{y}_{2,2}|\mathbf{y}_{1,1},\mathbf{y}_{2,1}\right)p\left(\mathbf{y}_{1,1},\mathbf{y}_{2,1}\right)d\mathbf{y}_{2,2}d\mathbf{y}_{2,1}$. Our marginalized marginal likelihood is in fact the product of a series of one step-ahead marginalized predictive likelihoods and a special case of (2) with the size of the training sample set to zero and a single horizon, $h=1$.

3 Marginalized marginal likelihoods for VAR models

We consider the m variable VAR model

$$\begin{aligned}\mathbf{y}_t' &= \sum_{i=1}^{p} \mathbf{y}_{t-i}'\mathbf{A}_i + \mathbf{x}_t'\mathbf{C} + \mathbf{u}_t' \\ &= \mathbf{z}_t'\boldsymbol{\Gamma} + \mathbf{u}_t'\end{aligned} \tag{3}$$

for $\mathbf{z}_t = \left(\mathbf{y}_{t-1}', \ldots, \mathbf{y}_{t-p}', \mathbf{x}_t'\right)'$ a $k \times 1$ vector or in matrix form

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\Gamma} + \mathbf{U}$$

with a normal likelihood, $\mathbf{u}_t \sim N(0, \boldsymbol{\Psi})$, and two common priors, the conjugate normal-Wishart prior and the non-conjugate independent normal Wishart prior.

3.1 Normal-Wishart prior

With the normal-Wishart prior,

$$\begin{aligned}\boldsymbol{\Gamma}|\boldsymbol{\Psi} &\sim MN_{km}\left(\underline{\boldsymbol{\Gamma}}, \boldsymbol{\Psi}, \underline{\boldsymbol{\Omega}}\right) \\ \boldsymbol{\Psi} &\sim iW\left(\underline{\mathbf{S}}, \underline{v}\right),\end{aligned}$$

the full sample marginal likelihood is readily available as a matricvariate t-distribution by noting that $\mathbf{Y}$ conditional on $\boldsymbol{\Psi}$ is matricvariate normal,

$$\mathbf{Y}|\boldsymbol{\Psi} = (\mathbf{Z}\boldsymbol{\Gamma} + \mathbf{U})|\boldsymbol{\Psi} \sim MN_{Tm}\left(\mathbf{Z}\underline{\boldsymbol{\Gamma}}, \boldsymbol{\Psi}, \mathbf{I}_T + \mathbf{Z}\underline{\boldsymbol{\Omega}}\mathbf{Z}'\right)$$

and then integrating out $\boldsymbol{\Psi}$ from the joint distribution of $\mathbf{Y}$ and $\boldsymbol{\Psi}$ to obtain the marginal likelihood as[2]

$$\mathbf{Y} \sim Mt_{Tm}\left(\mathbf{Z}\underline{\boldsymbol{\Gamma}}, \left(\mathbf{I}_T + \mathbf{Z}\underline{\boldsymbol{\Omega}}\mathbf{Z}'\right)^{-1}, \underline{\mathbf{S}}, \underline{v}\right).$$

To derive the marginalized marginal likelihood for the q variables in $\mathbf{Y}_1$, let $\mathbf{P}$ be the $m \times q$ selection matrix that yields $\mathbf{Y}_1 = \mathbf{Y}\mathbf{P}$, e.g. $\mathbf{P} = \left(\mathbf{I}_q, \mathbf{0}_{q\times(m-q)}\right)'$. We then have

$$\mathbf{Y}_1|\boldsymbol{\Psi} \sim MN_{Tq}\left(\mathbf{Z}\underline{\boldsymbol{\Gamma}}_1, \boldsymbol{\Psi}_1, \mathbf{I}_T + \mathbf{Z}\underline{\boldsymbol{\Omega}}\mathbf{Z}'\right)$$

for $\underline{\boldsymbol{\Gamma}}_1 = \underline{\boldsymbol{\Gamma}}\mathbf{P}$ and $\boldsymbol{\Psi}_1 = \mathbf{P}'\boldsymbol{\Psi}\mathbf{P}$. Theorem A.17 of Bauwens et al. (1999) implies that $\boldsymbol{\Psi}_1 \sim iW\left(\underline{\mathbf{S}}_1, \underline{v} - m + q\right)$ for $\underline{\mathbf{S}}_1 = \mathbf{P}'\underline{\mathbf{S}}\mathbf{P}$. We can thus marginalize out $\boldsymbol{\Psi}_1$ from the joint distribution of $\mathbf{Y}_1$ and $\boldsymbol{\Psi}_1$ to obtain the marginalized marginal likelihood as a matricvariate t-distribution,

$$\mathbf{Y}_1 \sim Mt_{Tn}\left(\mathbf{Z}\underline{\boldsymbol{\Gamma}}_1, \left(\mathbf{I}_T + \mathbf{Z}\boldsymbol{\Omega}\mathbf{Z}'\right)^{-1}, \underline{\mathbf{S}}_1, \underline{v} - m + q\right).$$

[2] We follow the notation of Bauwens, Lubrano and Richard (1999) for the matricvariate t and other distributions.

3.2 Independent normal Wishart prior

With the independent normal Wishart prior

$$\begin{gathered}\operatorname{vec}(\boldsymbol{\Gamma}) = \boldsymbol{\gamma} \sim N\left(\underline{\boldsymbol{\gamma}}, \underline{\boldsymbol{\Sigma}}_{\gamma}\right) \\ \boldsymbol{\Psi} \sim iW\left(\underline{\mathbf{S}}, \underline{v}\right)\end{gathered}$$

the marginal likelihood,

$$m(\mathbf{Y}) = \int\int L(\mathbf{Y}|\boldsymbol{\gamma}, \boldsymbol{\Psi})\, \pi(\boldsymbol{\gamma})\, \pi(\boldsymbol{\Psi})\, d\boldsymbol{\gamma} d\boldsymbol{\Psi},$$

and the marginalized marginal likelihood

$$MML(\mathbf{Y}_1) = \int\int\int L(\mathbf{Y}|\boldsymbol{\gamma}, \boldsymbol{\Psi})\, \pi(\boldsymbol{\gamma})\, \pi(\boldsymbol{\Psi})\, d\boldsymbol{\gamma} d\boldsymbol{\Psi} d\mathbf{Y}_2 \tag{4}$$

are not available in closed form.

Matters are simplified if the order of integration is changed. First integrate out $\mathbf{Y}_2$ to obtain the (conditional) marginalized likelihood for $\mathbf{Y}_1$ as the normal distribution

$$\operatorname{vec} \mathbf{Y}_1 | \boldsymbol{\gamma}, \boldsymbol{\Psi} \sim N\left[\left(\mathbf{P}' \otimes \mathbf{Z}\right) \boldsymbol{\gamma}, \mathbf{P}'\boldsymbol{\Psi}\mathbf{P} \otimes \mathbf{I}_T\right].$$

The marginalized marginal likelihood can then be obtained by integrating over the prior distributions for $\boldsymbol{\gamma}$ and $\boldsymbol{\Psi}$,

$$MML(\mathbf{Y}_1) = \int\int L(\mathbf{Y}_1|\boldsymbol{\gamma}, \boldsymbol{\Psi})\, \pi(\boldsymbol{\gamma})\, \pi(\boldsymbol{\Psi})\, d\boldsymbol{\gamma} d\boldsymbol{\Psi}. \tag{5}$$

Additional simplifications can be achieved by integrating $\boldsymbol{\gamma}$ analytically to obtain

$$\operatorname{vec} \mathbf{Y}_1 | \boldsymbol{\Psi} \sim N\left[\left(\mathbf{P}' \otimes \mathbf{Z}\right) \underline{\boldsymbol{\gamma}}, \mathbf{P}'\boldsymbol{\Psi}\mathbf{P} \otimes \mathbf{I}_T + \left(\mathbf{P}' \otimes \mathbf{Z}\right) \underline{\boldsymbol{\Sigma}}_{\gamma} \left(\mathbf{P} \otimes \mathbf{Z}'\right)\right] \tag{6}$$

and the marginalized marginal likelihood requires only integration over the prior distribution for $\boldsymbol{\Psi}$,

$$MML(\mathbf{Y}_1) = \int p(\mathbf{Y}_1|\boldsymbol{\Psi})\, \pi(\boldsymbol{\Psi})\, d\boldsymbol{\Psi}. \tag{7}$$

3.2.1 Numerical evaluation of $MML(\mathbf{Y}_1)$

Common methods for estimating the marginal likelihood such as the modified harmonic mean of Gelfand and Dey (1994) and Geweke (1999) or the methods of Chib (2008) and Chib and Jeliazkov (2001) relies on the identity

$$p(\boldsymbol{\theta}|\mathbf{Y}) = \frac{L(\mathbf{Y}|\boldsymbol{\theta})\, \pi(\boldsymbol{\theta})}{m(\mathbf{Y})}.$$

In terms of the marginalized marginal likelihood this corresponds to

$$p(\boldsymbol{\theta}|\mathbf{Y}_1) = \frac{L(\mathbf{Y}_1|\boldsymbol{\theta})\, \pi(\boldsymbol{\theta})}{\int L(\mathbf{Y}_1|\boldsymbol{\theta})\, \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta}} = \frac{L(\mathbf{Y}_1|\boldsymbol{\theta})\, \pi(\boldsymbol{\theta})}{MML(\mathbf{Y}_1)}.$$

Algorithm 1 Sampling from $p(\boldsymbol{\gamma}, \boldsymbol{\Psi}|\mathbf{Y}_1)$ using Gibbs within Metropolis-Hastings

For $j = 1, \ldots, B + R$

1. Generate $\boldsymbol{\gamma}^{(j)}$ from the normal distribution $p\left(\boldsymbol{\gamma}|\mathbf{Y}_1, \boldsymbol{\Psi}^{(j-1)}\right)$ in (8)
2. Generate a proposal $\boldsymbol{\Psi}'$ for $\boldsymbol{\Psi}$ from the full data posterior $p\left(\boldsymbol{\Psi}|\mathbf{Y}, \boldsymbol{\gamma}^{(j)}\right)$ or the prior $\pi\left(\boldsymbol{\Psi}\right)$ which both are inverse Wishart distributions and accept the proposal with probability

$$\alpha\left(\boldsymbol{\Psi}^{(j-1)}, \boldsymbol{\Psi}'\right) = \min\left(\frac{L\left(\mathbf{Y}_1|\boldsymbol{\gamma}^{(j)}, \boldsymbol{\Psi}'\right)\pi\left(\boldsymbol{\Psi}'\right)}{L\left(\mathbf{Y}_1|\boldsymbol{\gamma}^{(j)}, \boldsymbol{\Psi}^{(j-1)}\right)\pi\left(\boldsymbol{\Psi}^{(j-1)}\right)}\frac{p\left(\boldsymbol{\Psi}^{(j-1)}|\mathbf{Y}, \boldsymbol{\gamma}^{(j)}\right)}{p\left(\boldsymbol{\Psi}'|\mathbf{Y}, \boldsymbol{\gamma}^{(j)}\right)}, 1\right)$$

or

$$\alpha\left(\boldsymbol{\Psi}^{(j-1)}, \boldsymbol{\Psi}'\right) = \min\left(\frac{L\left(\mathbf{Y}_1|\boldsymbol{\gamma}^{(j)}, \boldsymbol{\Psi}'\right)}{L\left(\mathbf{Y}_1|\boldsymbol{\gamma}^{(j)}, \boldsymbol{\Psi}^{(j-1)}\right)}, 1\right)$$

Discard the first B draws as burn-in.

Analogous to the modified harmonic mean method for the marginal likelihood we then have, for a suitable function f that integrates to 1 over a subset of the domain of $\boldsymbol{\theta}$ and is zero outside of that subset, that

$$\begin{aligned}\frac{1}{MML\left(\mathbf{Y}_1\right)} &= \int_\Theta \frac{f\left(\boldsymbol{\theta}\right)}{MML\left(\mathbf{Y}_1\right)} d\boldsymbol{\theta} = \int_\Theta \frac{f\left(\boldsymbol{\theta}\right)}{L\left(\mathbf{Y}_1|\boldsymbol{\theta}\right)\pi\left(\boldsymbol{\theta}\right)} p\left(\boldsymbol{\theta}|\mathbf{Y}_1\right) d\boldsymbol{\theta} \\ &= E_{\boldsymbol{\theta}|\mathbf{Y}_1}\left(\frac{f\left(\boldsymbol{\theta}\right)}{L\left(\mathbf{Y}_1|\boldsymbol{\theta}\right)\pi\left(\boldsymbol{\theta}\right)}\right).\end{aligned}$$

We can thus use the harmonic mean method to estimate the marginalized marginal likelihood by sampling from the "posterior" distribution of $\boldsymbol{\gamma}$ and $\boldsymbol{\Psi}$ conditional on $\mathbf{Y}_1$ only. A Gibbs within Metropolis-Hastings sampler can be developed by noting that the partial data likelihood $L\left(\mathbf{Y}_1|\boldsymbol{\theta}\right)$ only identifies the parameters $\boldsymbol{\Gamma}_1$ relating to $\mathbf{Y}_1$. For the simple case where $\underline{\boldsymbol{\Sigma}}_\gamma$ is block diagonal this yields

$$\begin{aligned}\gamma_1\,|\mathbf{Y}_1, \boldsymbol{\Psi} &\sim N(\overline{\gamma}_1, \overline{\boldsymbol{\Sigma}}_1) \\ \gamma_2\,|\mathbf{Y}_1, \boldsymbol{\Psi} &\sim N(\underline{\gamma}_2, \underline{\boldsymbol{\Sigma}}_2)\end{aligned} \tag{8}$$

with

$$\begin{aligned}\overline{\boldsymbol{\Sigma}}_1 &= (\underline{\boldsymbol{\Sigma}}_1^{-1} + \boldsymbol{\Psi}_1^{-1} \otimes \mathbf{Z}'\mathbf{Z})^{-1} \\ \overline{\gamma}_1 &= \overline{\boldsymbol{\Sigma}}_1\left(\underline{\boldsymbol{\Sigma}}_1^{-1}\underline{\gamma}_1 + \mathrm{vec}(\mathbf{Z}'\mathbf{Y}_1\boldsymbol{\Psi}_1^{-1})\right)\end{aligned}$$

where $\boldsymbol{\Psi}_1 = \mathbf{P}'\boldsymbol{\Psi}\mathbf{P}$, $\underline{\boldsymbol{\Sigma}}_1 = (\mathbf{P} \otimes \mathbf{I}_k)'\underline{\boldsymbol{\Sigma}}_\gamma(\mathbf{P} \otimes \mathbf{I}_k)$, $\underline{\gamma}_1 = \mathrm{vec}(\underline{\boldsymbol{\Gamma}}_1)$, $\underline{\gamma}_2 = \mathrm{vec}(\underline{\boldsymbol{\Gamma}}\mathbf{P}^c)$, and $\underline{\boldsymbol{\Sigma}}_2 = (\mathbf{P}^c \otimes \mathbf{I}_k)'\underline{\boldsymbol{\Sigma}}_\gamma(\mathbf{P}^c \otimes \mathbf{I}_k)$ and $\mathbf{P}^c$ is the selection matrix for $\mathbf{Y}_2$. The MCMC algorithm is summarized in Algorithm 1.

An algorithm similar to the Chib and Jeliazkov (2001) method can be implemented by noting that

$$MML\left(\mathbf{Y}_1\right) = \frac{L\left(\mathbf{Y}_1|\boldsymbol{\theta}^*\right)\pi\left(\boldsymbol{\theta}^*\right)}{p\left(\boldsymbol{\theta}^*|\mathbf{Y}_1\right)}$$

for an arbitrary value of $\boldsymbol{\theta}^*$. By taking logarithms one obtains the expression

$$\log MML(Y_1) = \log L(\mathbf{Y}_1 \,|\boldsymbol{\gamma}^*, \boldsymbol{\Psi}^*) + \log \pi(\boldsymbol{\gamma}^*) + \log \pi(\boldsymbol{\Psi}^*) - \log p(\boldsymbol{\gamma}^*, \boldsymbol{\Psi}^* \,|\mathbf{Y}_1)$$

from which the marginalized marginal likelihood can be estimated by finding an estimate of the posterior ordinate $p\left(\boldsymbol{\gamma}^*, \boldsymbol{\Psi}^*|\mathbf{Y}_1\right) = p\left(\boldsymbol{\gamma}^*|\mathbf{Y}_1, \boldsymbol{\Psi}^*\right) p\left(\boldsymbol{\Psi}^*|\mathbf{Y}_1\right)$, this requires the estimation of $p\left(\boldsymbol{\Psi}^*|\mathbf{Y}_1\right)$. Following Chib and Jeliazkov (2001) the estimate can be based on the identity

$$p(\boldsymbol{\Psi}^* \,|\mathbf{Y}_1) = \frac{E_1[\alpha(\boldsymbol{\Psi}, \boldsymbol{\Psi}^* \,|\gamma, \mathbf{Y})p(\boldsymbol{\Psi}^* \,|\gamma, \mathbf{Y})]}{E_2[\alpha(\boldsymbol{\Psi}^*, \boldsymbol{\Psi} \,|\gamma, \mathbf{Y})]}$$

where the numerator expectation E_1 is with respect to the distribution $p(\boldsymbol{\gamma}, \boldsymbol{\Psi} \,|\mathbf{Y}_1)$ and the denominator expectation E_2 is with respect to the distribution $p(\boldsymbol{\gamma} \,|\boldsymbol{\Psi}^*, \mathbf{Y}_1) \times q(\boldsymbol{\Psi} \,|\gamma, \mathbf{Y})$ with $q(\cdot)$ and $\alpha(\cdot)$ the proposal density and acceptance probability from Algorithm 1. The expectation E_1 and E_2 requires running two Markov chains, one for $p\left(\boldsymbol{\gamma}, \boldsymbol{\Psi}|\mathbf{Y}_1\right)$ using Algorithm 1 and one for the distribution $p(\boldsymbol{\gamma} \,|\boldsymbol{\Psi}^*, \mathbf{Y}_1) \times q(\boldsymbol{\Psi} \,|\gamma, \mathbf{Y})$ where $p(\boldsymbol{\gamma} \,|\boldsymbol{\Psi}^*, \mathbf{Y}_1)$ is given by (8) and $q(\boldsymbol{\Psi} \,|\gamma, \mathbf{Y})$ is standard.

The simplest method for estimating the marginalized marginal likelihood follows from (7) where we can sample from the prior for $\boldsymbol{\Psi}$. If m is small this is a relatively low dimensional integral and direct Monte Carlo integration is straightforward but possibly inefficient if the prior is uninformative. As an alternative that can be more efficient we also consider importance sampling,

$$MML(\mathbf{Y}_1) = \int p(\mathbf{Y}_1 \,|\boldsymbol{\Psi}) \frac{\pi(\boldsymbol{\Psi})}{i(\boldsymbol{\Psi})} i(\boldsymbol{\Psi}) d\boldsymbol{\Psi}$$

with a mixture between the prior and the full data conditional posterior, $i(\boldsymbol{\Psi}) = \tau\pi(\boldsymbol{\Psi}) + (1-\tau)p(\boldsymbol{\Psi} \,|\mathbf{Y}, \boldsymbol{\gamma}^*)$, $0 \leq \tau \leq 1$, as importance function.

A similar simplification is also possible with the modified harmonic mean approach by writing

$$\frac{1}{MML\left(\mathbf{Y}_1\right)} = \frac{p\left(\boldsymbol{\Psi}|\mathbf{Y}_1\right)}{p\left(\mathbf{Y}_1|\boldsymbol{\Psi}\right)\pi\left(\boldsymbol{\Psi}\right)} = \int \frac{f\left(\boldsymbol{\Psi}\right)}{p\left(\mathbf{Y}_1|\boldsymbol{\Psi}\right)\pi\left(\boldsymbol{\Psi}\right)} p\left(\boldsymbol{\Psi}|\mathbf{Y}_1\right) d\boldsymbol{\Psi} \tag{9}$$

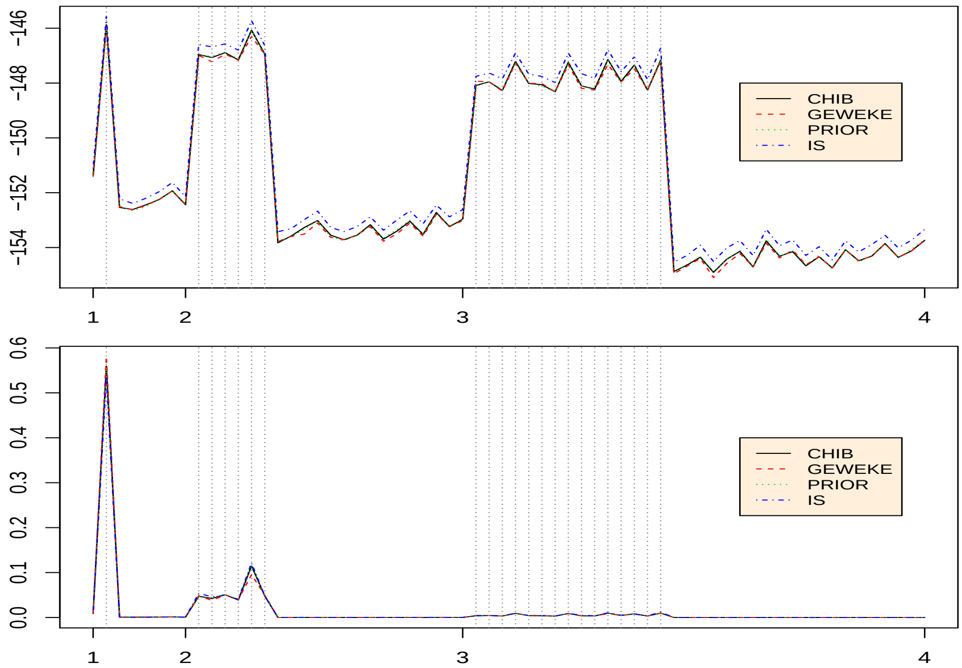
$$= E_{\boldsymbol{\Psi}|\mathbf{Y}_1}\left(\frac{f\left(\boldsymbol{\Psi}\right)}{p\left(\mathbf{Y}_1|\boldsymbol{\Psi}\right)\pi\left(\boldsymbol{\Psi}\right)}\right) \tag{10}$$

where $p\left(\mathbf{Y}_1|\boldsymbol{\Psi}\right)$ is given by (6). By performing part of the integration analytically the simulation noise is reduced and it should be easier to construct a good f function.

3.2.2 Performance of estimates of $MML\left(\mathbf{Y}_1\right)$

We conduct a small experiment to evaluate the performance of the different methods for estimating $MML\left(\mathbf{Y}_1\right)$ outlined above. Using two data sets of length 100 and 300 generated from DGP1 in the simulation study (see section 4 for more details) we study the stability of the estimated $\ln MML\left(\mathbf{Y}_1\right)$ and the required CPU time. The true DGP is a bivariate VAR consisting of y_1, the variable of interest, and y_2. In addition there are six extraneous variables that are considered for inclusion in the VAR in addition to y_2. We estimate $\ln MML\left(\mathbf{Y}_1\right)$ for the 64 possible VAR-models of dimension 1-4 with y_1 always included in the model.

Figure 1 Log-$MML(\mathbf{Y}_1)$ and the corresponding posterior weights using different methods, DGP1, $T = 100$, $\delta_k = 0.2$, Dimension: 1-4



The prior specification is the same as in the simulation study (see section 4.1). The estimates are based on 5000 MCMC draws after discarding 1000 draws as burn-in and Algorithm 1 is implemented with the prior for $\mathbf{\Psi}$ as the proposal density. In the Chib-Jeliazkov method (*CHIB*), $\boldsymbol{\gamma}^*$ and $\mathbf{\Psi}^*$ are set to the mean of the draws from Algorithm 1.

The modified harmonic mean (*GEWEKE*) estimate is based on (10) working with the Cholesky factor $\mathbf{U}$ of $\mathbf{\Psi}$ and the logarithm of the diagonal elements of $\mathbf{U}$ and $f(\cdot)$ a truncated multivariate normal distribution. Note that this requires an adjustment for the variable transformation with the Jacobian term,

$$\mathcal{J}(\boldsymbol{\psi} \to \mathbf{\Psi}) = \mathcal{J}(\boldsymbol{\psi} \to \mathbf{U}) \times \mathcal{J}(\mathbf{U} \to \mathbf{\Psi})$$
$$= \left(\prod_{i=1}^{m} u_{ii}\right)^{-1} \times 2^m \prod_{i=1}^{m} u_{ii}^{m-i+1}$$

where $\boldsymbol{\psi}$ contains the non-zero elements of $\mathbf{U}$ after taking logarithms and u_{ii} are the diagonal elements before taking the logarithm. After trying several values of the truncation factor ρ as suggested in Geweke (1999), we choose $\rho = 0.5$.

In the method of importance sampling (*IS*), the probability of drawing from the prior is set to $\tau = 0.2$. The method of direct integration using samples from prior is denoted as *PRIOR* in our study.

Figure 1 presents the log-$MML(\mathbf{Y}_1)$ and the corresponding posterior model weights of the 64 unique models of dimension from one to four. The true model is the second model consisting of $\{y_1, y_2\}$ and is marked by the first gray dashed line. The other gray dashed lines indicate larger models containing y_1 and y_2. Not surprisingly, the MMLs for a given model dimension are largest when the variable y_2 is in the model. Furthermore, the global maximum is obtained for the correct

Table 1 CPU time for estimating $\ln MML(Y_1)$, DGP1

Method	$T = 100$	$T = 300$
CHIB	16.675	20.439
GEWEKE	18.600	21.038
PRIOR	72.731	1569.483
IS	95.658	1610.684

model and the posterior weights clearly indicate the correct model. All the methods give similar estimates of the MMLs with the *IS* method deviating slightly.

Based on the results in Figure 1 there is little reason to prefer one method over another and we turn to the CPU times reported in Table 1. The *CHIB* and *GEWEKE* methods are the fastest and thus preferred while the *PRIOR* and *IS* methods show a dramatic increase in CPU time as the number of observations increase. The increase in CPU time for the latter methods can be explained by the need to invert the $qT \times qT$ covariance matrix in (6).

4 Simulation study

To evaluate the performance of the marginalized marginal likelihood and predictive likelihood methods we conduct a small simulation study with three different DGPs

DGP1:

$$\mathbf{y}_t = \begin{pmatrix} 0.5 & -0.2 \\ 0.3 & 0.7 \end{pmatrix} \mathbf{y}_{t-1} + \mathbf{u}_t$$

DGP2:

$$\mathbf{y}_t = \begin{pmatrix} 0.5 & -0.2 \\ 0.3 & 0.7 \end{pmatrix} \mathbf{y}_{t-1} + \begin{pmatrix} 0.2 & 0.2 \\ 0.1 & -0.3 \end{pmatrix} \mathbf{y}_{t-2} + \mathbf{u}_t$$

DGP3:

$$\mathbf{y}_t = \begin{pmatrix} 0.5 & -0.2 & 0.2 \\ 0.3 & 0.7 & 0.1 \\ 0.4 & 0.3 & 0.2 \end{pmatrix} \mathbf{y}_{t-1} + \mathbf{u}_t$$

In each case we have one variable of interest, y_1, which is always retained in the model and we consider models with up to four variables selected from the remaining variables in $\mathbf{y}$ and 6 extraneous variables

$$\begin{aligned}
x_{1,t} &= 0.3y_{1,t-1} + 0.5x_{1,t-1} + e_{1,t} \\
x_{2,t} &= 0.5y_{2,t-1} + 0.5x_{2,t-1} + e_{2,t} \\
x_{3,t} &= 0.2x_{3,t-1} + e_{3,t} \\
x_{4,t} &= 0.7x_{4,t-1} + e_{4,t} \\
x_{5,t} &= e_{5,t} \\
x_{6,t} &= e_{6,t}.
\end{aligned}$$

4.1 Preliminary setup

The prior specification is based on a Litterman type prior with prior mean zero for $\boldsymbol{\gamma}$ except for the own first lag where the prior mean is set to 0.9. The prior standard deviations are given by

$$\begin{aligned} &\frac{\pi_1}{k^{\pi_3}}, \text{ own lags, } k = 1, \ldots, p \\ &\frac{s_i \pi_1 \pi_2}{s_j k^{\pi_3}}, \text{ lags of variable } j \text{ in equation } i, \ k = 1, \ldots, p \\ &\pi_4, \text{ deterministic variables} \end{aligned}$$

where s_i is the residual standard deviation for equation i from the OLS fit of the VAR-model. For the independent normal Wishart we use $\pi_1 = 0.5$, $\pi_2 = 0.5$, $\pi_3 = 0.5$, $\pi_4 = 1$, $\underline{v} = 9$ and for the normal-Wishart $\pi_1 = 0.5$, $\pi_2 = 1$, $\pi_3 = 0.5$, $\pi_4 = 1$, $\underline{v} = 9$ with $\underline{\mathbf{\Omega}}$ and $\underline{\mathbf{S}}$ set as in Kadiyala and Karlsson (1997).

The model prior is given by

$$\begin{aligned} \pi\left(\mathcal{M}_j\right) &\propto \prod_{k=1}^{K} \delta_k^{d_k} \left(1 - \delta_k\right)^{d_k} \\ d_k &= 1 \text{ if variable } k \text{ is included} \\ \delta_k &= \text{prior variable inclusion probability} \end{aligned}$$

and we consider two settings for $\delta_k = 0.2$ or 0.5.

For each DGP we generate 100 data sets of length 112 and 312 where the last 12 observations set aside for forecast evaluation with forecast horizons $h = 1$ to 12. We use two settings for the lag length of the VAR-models, the true lag length and the true lag length + 1. A constant term is always included as a deterministic variable in all model specifications.

For the marginalized predictive likelihood we use a hold out sample of 70 observations with the small data sets and 90 observations for the large data sets and estimate the predictive weights for the single horizon $h = 1$ with 5000 draws from the training sample posterior. The marginalized marginal likelihood is estimated using 5000 draws from the prior with the independent normal Wishart prior. The final forecasts arise as the mean forecast from 5000 sample draws.

4.2 Results

4.2.1 Variable and model selection

Following Andersson and Karlsson (2009) we calculate the average variable inclusion probabilities to check the variable selection performance and the proportion of data sets where the true model is selected to investigate the model selection performance of the marginalized predictive likelihood and the marginalized marginal likelihood. The variable inclusion probability for variable k is given by

$$p(x_k \,|\mathbf{y}) = \sum_{j=1}^{M} 1(x_k \in \mathcal{M}_j) p(\mathcal{M}_j \,|\mathbf{y})$$

Table 2 Posterior variable inclusion probabilities, DGP 1, $T = 100$

Normal-Wishart prior

		Predictive likelihood			Marginal likelihood		
p	δ_k	$p(y_2)$	$max[p(x_i)]$	$\frac{p(y_2)}{\max[p(x_i)]}$	$p(y_2)$	$max[p(x_i)]$	$\frac{p(y_2)}{\max[p(x_i)]}$
1	0.2	0.69	0.15	4.58	0.86	0.07	11.51
	0.5	0.77	0.28	2.78	0.92	0.18	5.21
2	0.2	0.55	0.15	3.72	0.75	0.04	17.03
	0.5	0.67	0.27	2.50	0.86	0.12	7.22

Independent normal Wishart prior

		Predictive likelihood			Marginal likelihood		
p	δ_k	$p(y_2)$	$max[p(x_i)]$	$\frac{p(y_2)}{\max[p(x_i)]}$	$p(y_2)$	$max[p(x_i)]$	$\frac{p(y_2)}{\max[p(x_i)]}$
1	0.2	0.72	0.16	4.41	0.78	0.02	39.84
	0.5	0.82	0.30	2.74	0.86	0.04	21.17
2	0.2	0.63	0.16	4.01	0.70	0.04	16.83
	0.5	0.73	0.28	2.60	0.81	0.10	8.16

Table 3 Posterior variable inclusion probabilities, DGP 1, $T = 300$

Normal-Wishart prior

		Predictive likelihood			Marginal likelihood		
p	δ_k	$p(y_2)$	$max[p(x_i)]$	$\frac{p(y_2)}{\max[p(x_i)]}$	$p(y_2)$	$max[p(x_i)]$	$\frac{p(y_2)}{\max[p(x_i)]}$
1	0.2	0.90	0.19	4.69	1.00	0.05	18.63
	0.5	0.93	0.31	2.97	1.00	0.14	7.35
2	0.2	0.87	0.17	5.04	1.00	0.03	39.73
	0.5	0.91	0.29	3.14	1.00	0.07	14.92

Independent normal Wishart prior

		Predictive likelihood			Marginal likelihood		
p	δ_k	$p(y_2)$	$max[p(x_i)]$	$\frac{p(y_2)}{\max[p(x_i)]}$	$p(y_2)$	$max[p(x_i)]$	$\frac{p(y_2)}{\max[p(x_i)]}$
1	0.2	0.90	0.19	4.79	1.00	0.09	10.84
	0.5	0.93	0.31	3.03	1.00	0.21	4.69
2	0.2	0.88	0.17	5.15	1.00	0.07	15.04
	0.5	0.92	0.29	3.19	1.00	0.16	6.26

where $1(x_k \in M_j)$ is 1 if variable k is included in model j.

Table 2 - 3 reports on the variable inclusion probabilities for the simulated small and large data set under DGP1 respectively. Instead of reporting the variable inclusion probabilities of all 6 extraneous variables we only report the largest variable inclusion probability. All the results show that the procedure using either the marginalized marginal likelihood or predictive likelihood is able to make a clear choice between the "true" variable y_2 and the remaining variables. The performance is slightly better when the true lag length ($p = 1$) and smaller prior variable inclusion probability ($\delta_k = 0.2$) is used. This is consistent with the findings of Andersson and Karlsson (2009) that too large a lag length or a model prior favoring too large models has a negative impact on the discriminatory power. We also find that the marginalized marginal likelihood provides a sharper discrimination between models and variables as it is based on the full sample and makes more efficient use of the data.

Table 4 provides the model selection results for the small data set under DGP1 with $T = 100$. Overall, the marginalized marginal likelihood selects the correct model well (between 69% and 94% of the 100 replicates), and the posterior weights for the true model are quite large (mostly over 0.5). In contrast, the marginalized predictive likelihood in general selects the true model poorly and especially when the model prior favors larger models ($\delta_k = 0.5$) with the correct model select in less than 18% of the cases. Turning to the results for $T = 300$ in Table 5 we find that the results are similar in terms of the posterior weights but that the proportion of correct model selection increases substantially for both the marginalized marginal and predictive likelihoods. The results for DGP2 and DGP3 are qualitatively similar and are given in Appendix B.

4.2.2 Forecasting performance

For the forecasting performance, we will focus on the simulation results for the normal-Wishart prior presented in Figure 2 - 7. The qualitative results are quite similar for the independent normal Wishart prior, which can be found in Figure 15 - 20 in Appendix B. For all DGPs, we compute the mean square error (MSE) for the Bayesian forecast combination (BMA) and compare it to the model with the highest model weights (TOP). Furthermore, the performance is reported in terms of MSE relative to the MSE for a univariate AR(3) with a ratio less than 1 indicating superior performance.

The results for DGP1 and DGP3 follow the same pattern. For the shorter lead times, the forecast combination and the forecast using best model outperforms the forecasts from the AR(3) by a substantial margin and is quite similar for the longer lead times. The difference in performance is smaller for the larger sample size with a clear edge for the forecast combinations and the "top model" only for very short lead times. The results for DGP2 are qualitatively similar but with less pronounced gains for the small sample size.

The differences between the marginalized marginal and predictive likelihoods or forecasts based on model averaging or selecting a best model are very small. This can be explained by the model selection results in Tables 4 and 5. Both the marginalized marginal and predictive likelihoods almost always select the correct model and the

Table 4 Model Selection, DGP1, Average posterior weight and proportion selected for true model, $T = 100$.

Normal-Wishart prior					
		Predictive likelihood		Marginal likelihood	
p	δ_k	Weight	Selected	Weight	Selected
1	0.2	0.27	0.57	0.59	0.89
	0.5	0.08	0.17	0.29	0.83
2	0.2	0.23	0.33	0.60	0.78
	0.5	0.10	0.18	0.44	0.79
Independent normal Wishart prior					
		Predictive likelihood		Marginal likelihood	
p	δ_k	Weight	Selected	Weight	Selected
1	0.2	0.26	0.65	0.76	0.85
	0.5	0.07	0.11	0.76	0.94
2	0.2	0.23	0.45	0.64	0.72
	0.5	0.08	0.12	0.61	0.69

Table 5 Model Selection, DGP1, Average posterior probability and proportion selected for true model, $T = 300$.

Normal-Wishart prior					
		Predictive likelihood		Marginal likelihood	
p	δ_k	Weight	Selected	Weight	Selected
1	0.2	0.28	0.90	0.77	0.96
	0.5	0.05	0.04	0.44	0.86
2	0.2	0.30	0.81	0.90	0.99
	0.5	0.07	0.13	0.72	0.90
Indedependent normal Wishart prior					
		Predictive likelihood		Marginal likelihood	
p	δ_k	Weight	Selected	Weight	Selected
1	0.2	0.29	0.91	0.60	0.96
	0.5	0.05	0.07	0.22	0.68
2	0.2	0.30	0.86	0.72	0.93
	0.5	0.06	0.11	0.39	0.78

Figure 2 Forecast performance, DGP1, MSE relative to univariate AR(3), normal-Wishart prior, $T = 100$

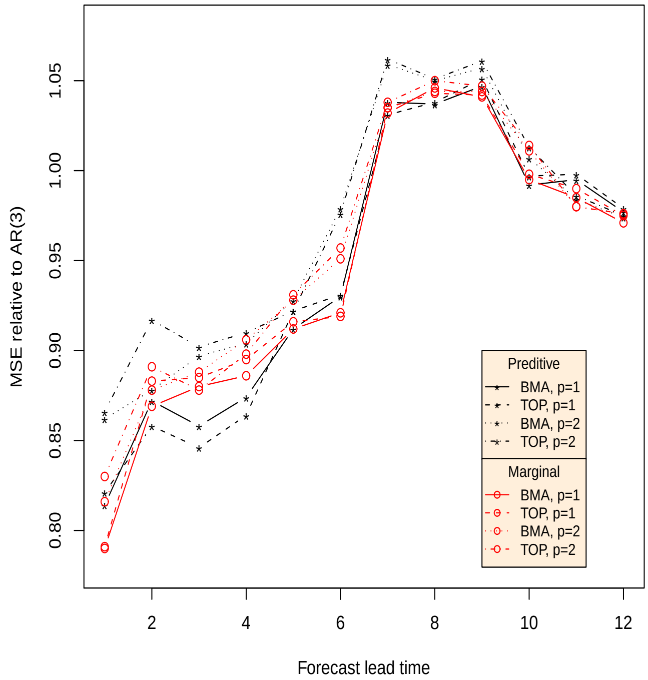


Figure 3 Forecast performance, DGP1, MSE relative to univariate AR(3), normal-Wishart prior, $T = 300$

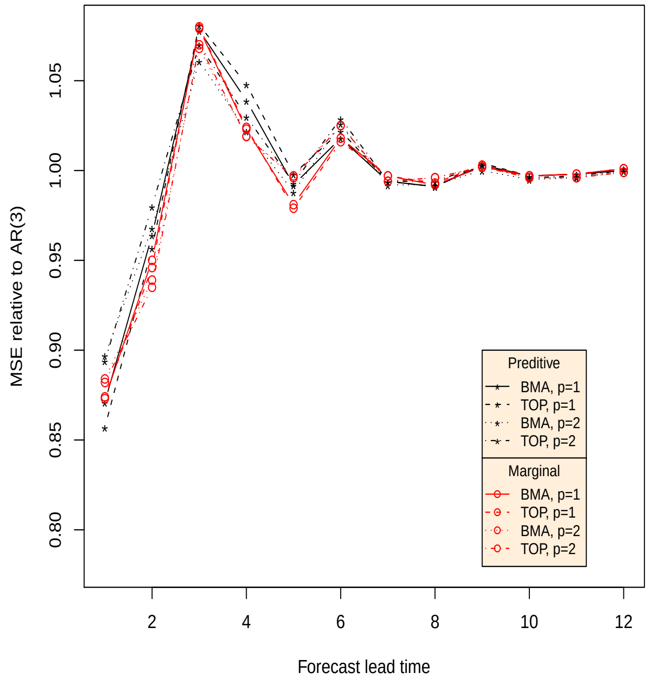


Figure 4 Forecast performance, DGP2, MSE relative to univariate AR(3), normal-Wishart prior, $T = 100$

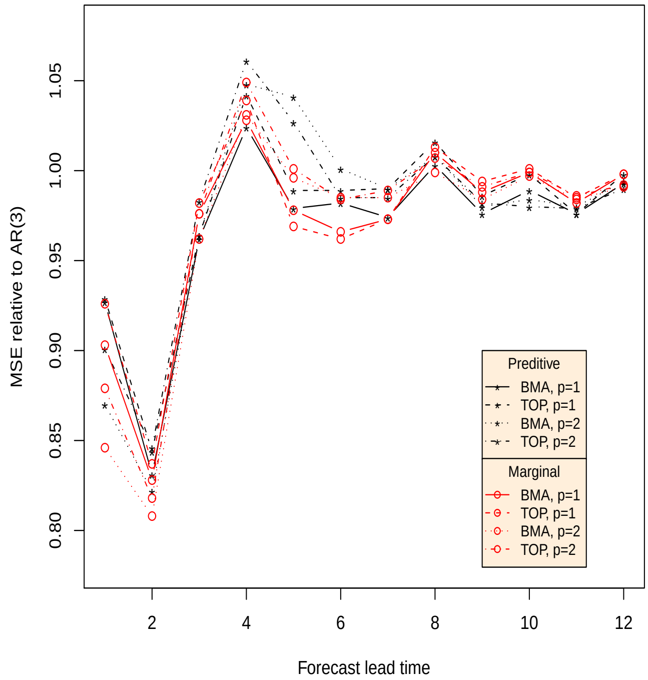


Figure 5 Forecast performance, DGP2, MSE relative to univariate AR(3), normal-Wishart prior, $T = 300$

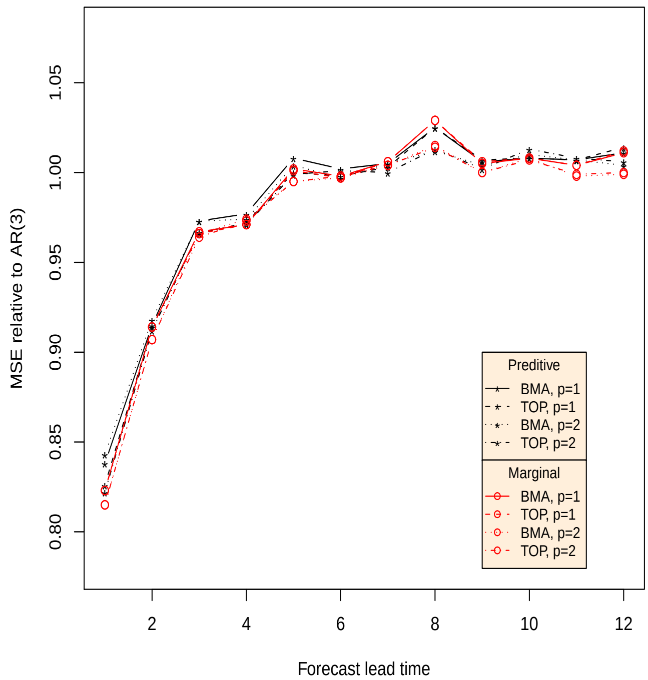


Figure 6 Forecast performance, DGP3, MSE relative to univariate AR(3), normal-Wishart prior, $T = 100$

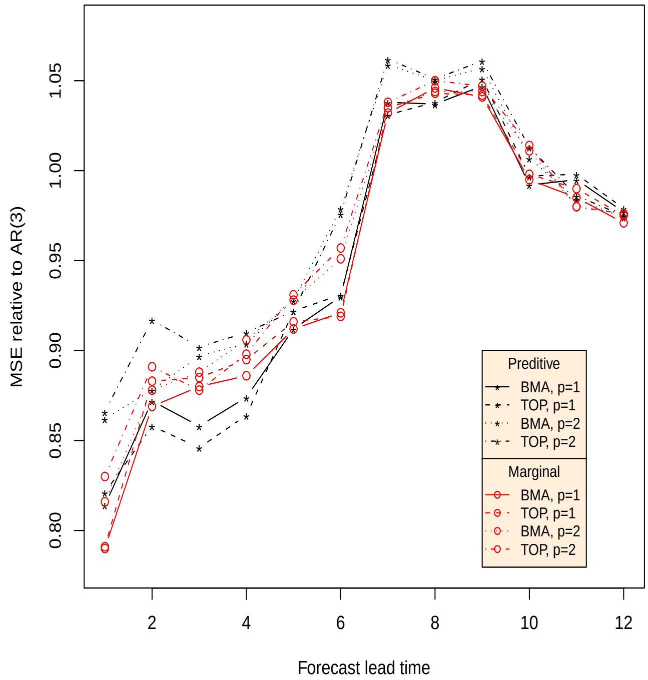


Figure 7 Forecast performance, DGP3, MSE relative to univariate AR(3), normal-Wishart prior, $T = 300$

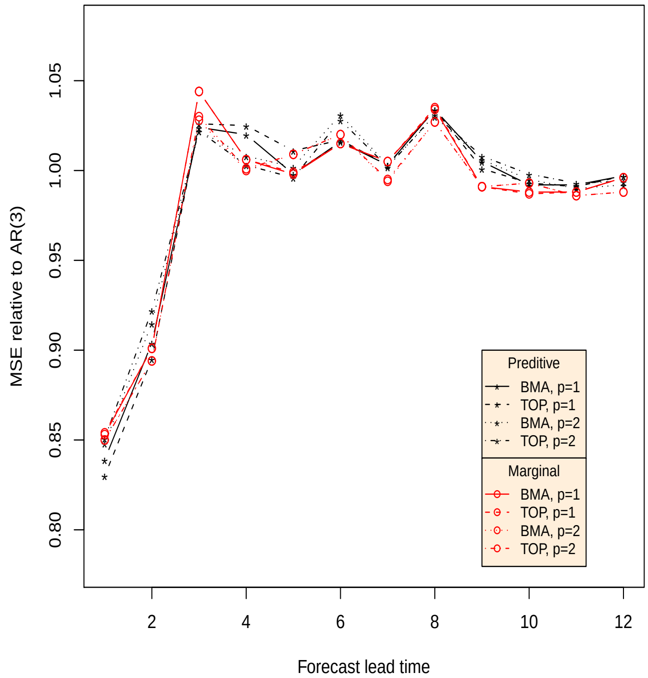
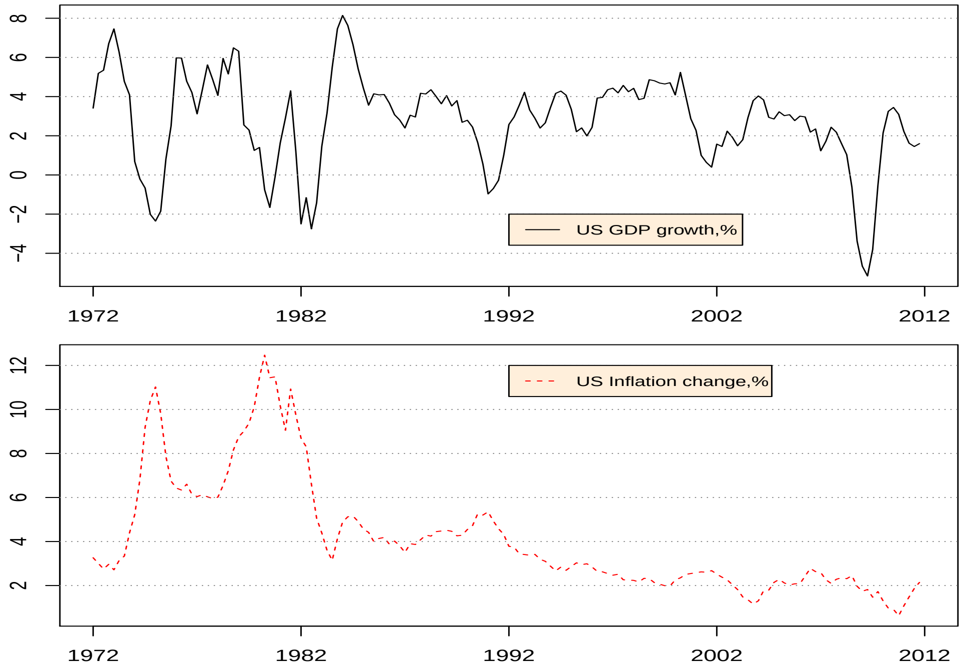


Figure 8 US GDP growth and inflation, in percent, 1972Q1-2011Q4



high average weight means that the correct model will dominate the combination forecasts, making them all very similar in the simulation exercise.

5 Forecasting US GDP growth and inflation

In this section we evaluate the forecast performance empirically by forecasting the U.S. real GDP growth and CPI inflation. The two variables of primary interest are displayed in Figure 8. In addition to GDP growth and inflation we consider an additional 15 variables for inclusion in the forecasting model. The set of variables is similar to Andersson and Karlsson (2009) and consists of aggregate real and nominal quarterly data from 1972Q1 to 2011Q4. The full list of variables can be found in Appendix A.

In addition to the two variables of interest we consider VAR models with up to four additional variables selected from the list in Appendix A for a total of 1941 distinct models. The forecasts are then produced by selecting a single model according to the marginalized predictive and marginal likelihoods and by forecast combinations based on the marginalized likelihoods.

The simulation study showed little or no gains from using the more complicated independent normal Wishart prior and we will only consider forecasts based on the normal-Wishart prior. The prior hyperparameters are set as in the simulation study, i.e. the overall tightness is set to $\pi_1 = 0.5$, the cross-equation tightness to $\pi_2 = 1$, the lag decay to $\pi_3 = 0.5$, the deterministic tightness to $\pi_4 = 5$ and the prior degrees of freedom to $\underline{v} = 9$. For the marginalized predictive likelihood a hold out sample of 50 observations is used and the predictive weights are estimated for the single horizon $h = 1$ using 5000 draws from the training sample posterior. For each model the forecast is the expected value of the predictive distribution which is simulated

Table 6 Forecast accuracy for US GDP growth, relative RMSE to AR(4), absolute RMSE for AR(4), 30 origins, 1997Q1-2004Q2

	Predictive likelihood		Marginal likelihood		
h	BMA	TOP	BMA	TOP	AR(4)
1	0.940	0.942	0.950	0.942	0.871
2	0.918	0.919	0.935	0.934	0.900
3	0.902	0.912	0.891	0.906	1.300
4	0.951	0.953	0.968	0.977	1.275
5	1.132	1.118	1.071	1.082	1.347
6	1.193	1.175	1.119	1.149	1.305
7	1.128	1.098	1.089	1.124	1.452
8	1.213	1.186	1.106	1.135	1.406

Stdev(GDP) = 1.441.

Table 7 Forecast accuracy for US GDP growth, relative RMSE to AR(4), absolute RMSE for AR(4), 30 origins, 2004Q3-2011Q4

	Predictive likelihood		Marginal likelihood		
h	BMA	TOP	BMA	TOP	AR(4)
1	0.991	1.021	0.982	0.984	1.740
2	1.065	1.074	1.097	1.104	1.824
3	0.994	0.994	1.006	1.015	3.060
4	1.084	1.103	1.106	1.114	3.082
5	1.115	1.129	1.133	1.142	3.414
6	1.093	1.098	1.089	1.103	3.402
7	1.123	1.126	1.115	1.127	3.295
8	1.077	1.070	1.034	1.057	3.320

Stdev(GDP) = 2.440.

using 5000 draws from the full sample posterior.

5.1 Forecast performance

The period 1997Q1 to 2011Q4 is set aside for evaluating the forecast performance and forecasts for 1 to 8 quarters ahead are generated using a recursive updating scheme. That is, the first set of forecasts covering 1997Q1 to 1998Q4 are based on data up to 1996Q4, the second set of forecasts for 1997Q2 to 1999Q1 are based on data up to 1997Q1 and so on.

Tables 6 and 7 report on the performance of the forecasts of GDP growth and Figures 9 and 10 shows the one quarter ahead forecast combination forecasts. We report separately on the two periods 1997Q1 to 2004Q2 and 2004Q3 to 2011Q4 in order to highlight the effect of the recent financial on the forecast performance. The tables report the root mean square error (RMSE) for a baseline univariate AR(4) forecasting model and the RMSE relative to the AR(4) for the forecast combinations

Figure 9 BMA forecast performance, US GDP growth, one quarter ahead forecast and forecast errors, 1997Q1-2004Q2

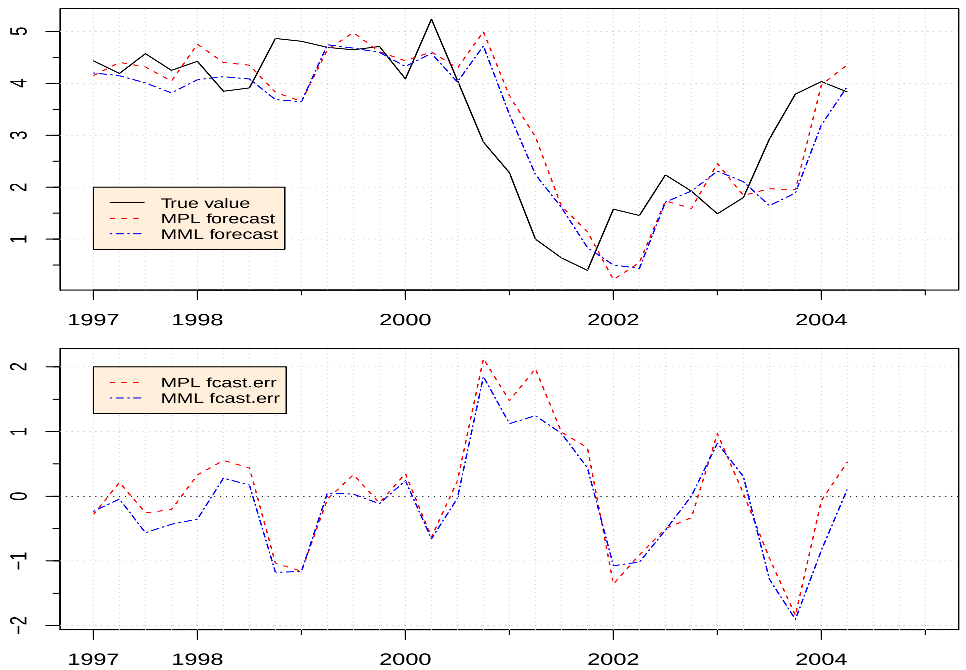


Figure 10 BMA forecast performance, US GDP growth, one quarter ahead forecast and forecast errors, 2004Q3-2011Q4

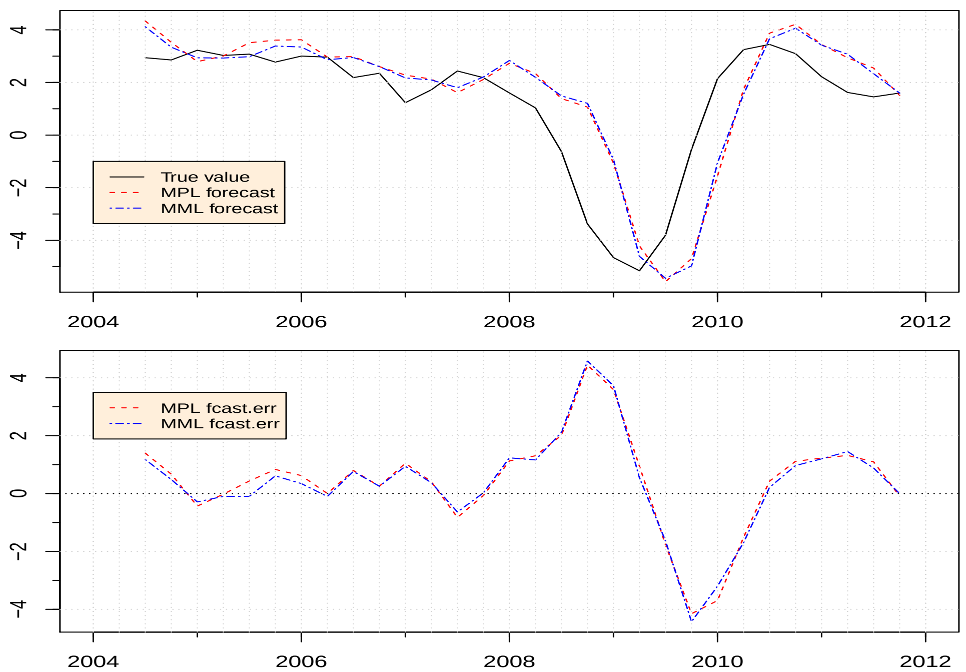


Table 8 Forecast accuracy for US Inflation, relative RMSE to AR(4), absolute RMSE for AR(4), 30 origins, 1997Q1-2004Q2

	Predictive likelihood		Marginal likelihood		
h	BMA	TOP	BMA	TOP	AR(4)
1	0.795	0.931	0.816	0.837	0.295
2	0.879	1.129	0.912	0.939	0.301
3	0.693	0.920	0.683	0.702	0.614
4	0.871	1.169	0.788	0.806	0.564
5	0.725	0.927	0.647	0.683	0.860
6	1.030	1.314	0.722	0.737	0.736
7	0.805	0.940	0.644	0.681	1.063
8	1.160	1.378	0.833	0.865	0.868

Stdev(CPI) = 0.409.

Table 9 Forecast accuracy for US Inflation, relative RMSE to AR(4), absolute RMSE for AR(4), 30 origins, 2004Q3-2009Q2

	Predictive likelihood		Marginal likelihood		
h	BMA	TOP	BMA	TOP	AR(4)
1	0.998	0.957	1.016	1.042	0.397
2	1.094	1.014	1.070	1.100	0.404
3	0.882	0.811	0.921	0.935	0.721
4	0.944	0.901	0.996	1.049	0.709
5	0.773	0.786	0.777	0.898	0.805
6	0.761	0.790	0.785	0.848	0.801
7	0.765	0.747	0.743	0.767	1.015
8	0.973	0.949	0.946	0.938	0.948

Stdev(CPI) = 0.537.

(BMA in the tables) and the forecasts from the model with the highest marginalized likelihood at each point in time (TOP in the tables).

For the period 1997Q1 to 2004Q2 (Table 6) the predictive likelihood performs slightly better than the marginal likelihood, and both procedures show some improvement on the common benchmark of AR(4) for short lead times. The forecast combination (BMA) performs slightly better than selecting a single model (TOP) with the predictive likelihood, while we observe the opposite result with the marginal likelihood. In Table 7, covering the financial crisis period, the forecast performance for all procedures are worse than the results for the "normal" period. The range of RMSE for the AR(4) is $(1.740, 3.414)$ in the crisis period, much higher than $(0.871, 1.452)$ for the normal period. The forecast combinations do better than selecting a single model but both procedures fail to improve on the AR(4) benchmark for the 2004Q3 to 2011Q4 period.

As can be expected, the forecast performance for the inflation rate reported in Tables 8 and 9 and Figures 11 and 12 is much better than for GDP growth. The

Figure 11 BMA forecast performance, US Inflation change, one quarter ahead forecast and forecast errors, 1997Q1-2004Q2

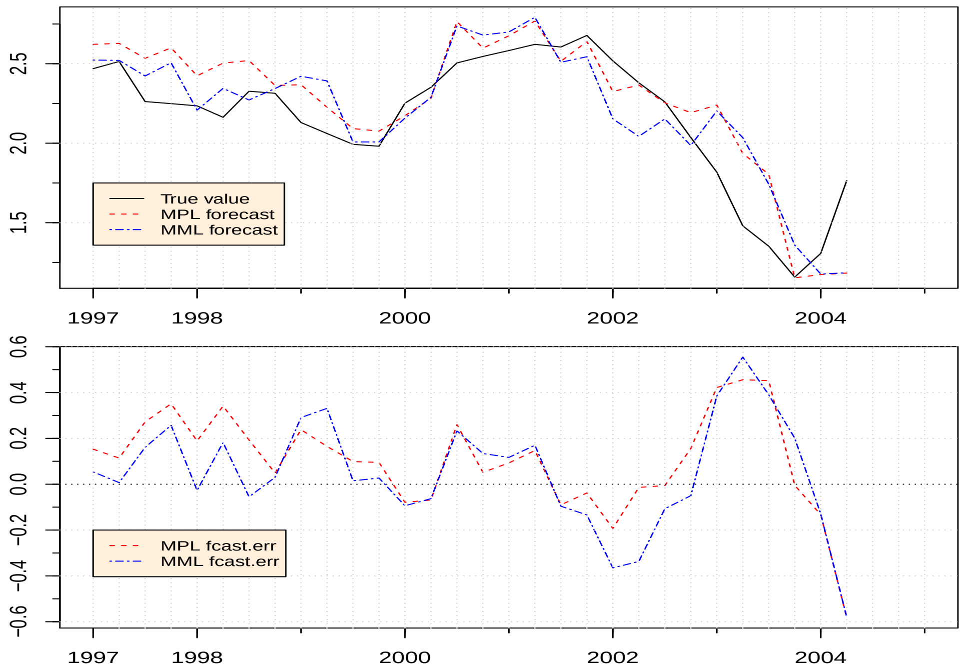


Figure 12 BMA forecast performance, US Inflation change, one quarter ahead forecast and forecast errors, 2004Q3-2011Q4

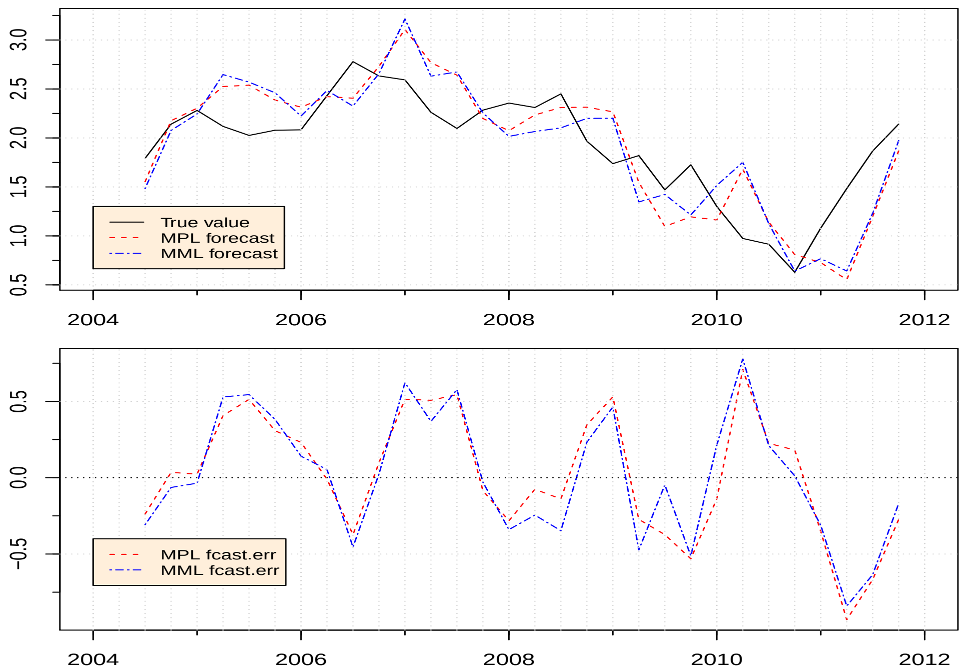
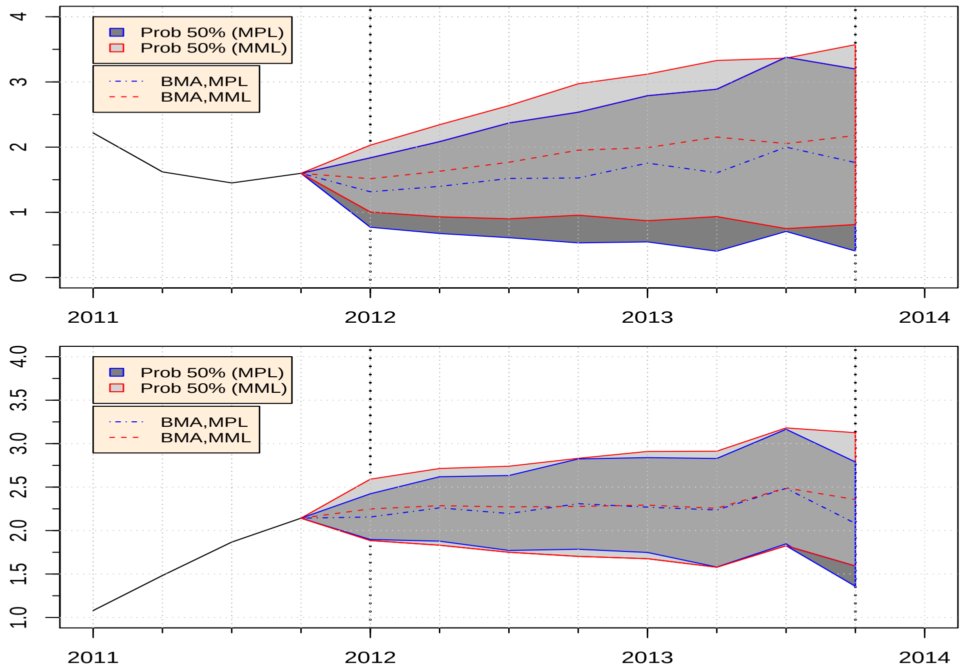


Figure 13 Forecasts of US GDP growth (first) and inflation (second) as of 2011Q4, MPL weights and MML weights



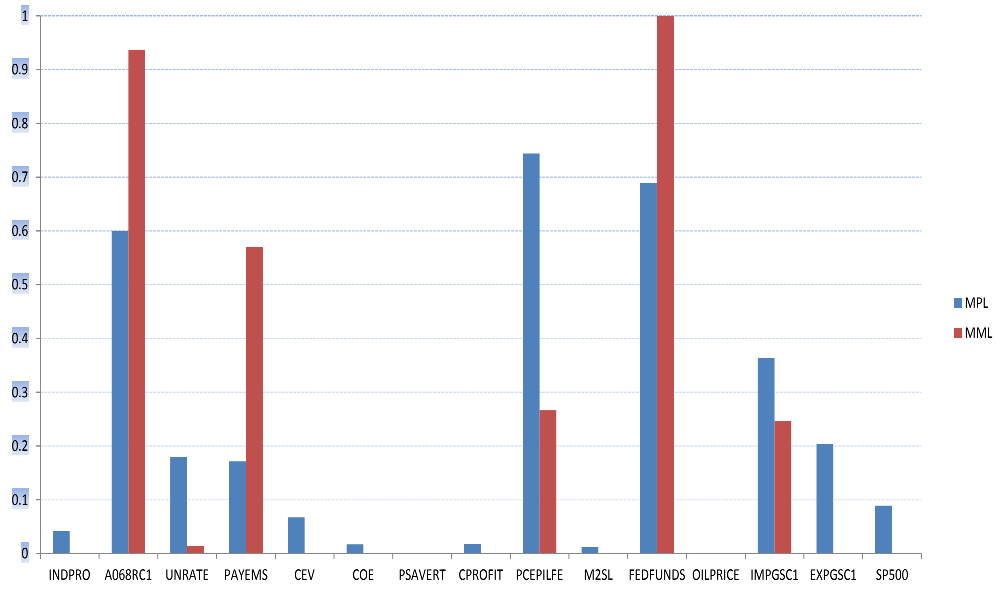
effect of the financial crisis on the forecastability is much smaller for inflation with small differences in absolute RMSEs between the two periods. Overall the forecast combination based on the marginalized marginal likelihood performs best, sometimes improving on the AR(4) benchmark by as much as 35% while only producing worse forecasts in two cases.

5.2 Out of sample forecasts

To provide further insight into the properties of the model averaging procedures we provide out of sample forecasts 1 to 8 quarters ahead based on both the predictive and marginal likelihoods in Figure 13. The 50% prediction intervals are based on the model averaged predictive distribution and accounts for both parameter and model uncertainty. While the forecasts are quite similar for the inflation rate the predictive likelihood forecast is more pessimistic about GDP growth.

To see why this is the case we turn to Figure 14 showing the variable inclusion "probabilities" and Tables 10 and 11 which show the top 10 models using the marginalized predictive and marginal likelihoods. The two measures clearly favors different sets of variables. The marginal likelihood focuses on fewer variables with the interest rate (FEDFUNDS), personal outlays (A068RC1) and employment (PAYEMS) being the most important followed by personal consumption expenditure (PCEPILFE) and imports (IMPGSC1). The predictive likelihood gives lower weights to the three variables most favored by the marginal likelihood and instead gives more weight to consumption expenditure (which is largely exchangeable with personal outlays) and gives substantial weights to exports (EXPGSC1) and unemployment (UNRATE). The difference in the forecasts can thus be attributed to the larger role for exports and unemployment in the models favored by the predictive

Figure 14 Variable selection, US GDP growth and inflation, MPL weights and MML weights



likelihood.

6 Summary

Standard Bayesian model averaging and model selection does not work with multivariate models subject to dimension changes. This paper proposes the use of the marginalized marginal likelihood as a solution to this problem. Similar to the marginalized predictive likelihood proposed by Andersson and Karlsson (2009) it has the additional advantage of being a focused indicator that measures the model fit for the variables of interest.

We show how to use the marginalized marginal likelihood in VAR models and two commonly used families of prior distributions, the normal-Wishart and independent normal Wishart priors. With the normal-Wishart prior the marginalized marginal likelihood is available in closed form while numerical methods are required with the independent normal Wishart prior. For the latter we propose and evaluate several different simulation based methods for estimating the marginalized marginal likelihood.

In a small simulation study and an application to forecasting US GDP growth and inflation we illustrate the variable and model selection properties and demonstrate that forecasting combinations based on the marginalized marginal likelihood can improve on forecasts based on single models.

Table 10 Top 10 models, US GDP growth and inflation, predictive weights

	Top model									
Variable	1	2	3	4	5	6	7	8	9	10
INDPRO	○	○	○	○	○	○	○	○	●	○
A068RC1	●	○	○	○	○	●	○	●	○	●
UNRATE	○	●	●	○	○	○	○	○	○	○
PAYEMS	○	○	○	●	●	○	●	○	○	○
CEV	○	○	○	○	●	○	○	○	○	○
COE	○	○	○	○	○	○	○	○	○	○
PSAVERT	○	○	○	○	○	○	○	○	○	○
CPROFIT	○	○	○	○	○	○	○	○	○	○
PCEPILFE	●	●	●	○	○	○	○	●	●	○
M2SL	○	○	○	○	○	○	○	○	○	○
FEDFUNDS	●	●	○	○	○	●	○	●	○	●
OILPRICE	○	○	○	○	○	○	○	○	○	○
IMPGSC1	○	●	●	●	●	○	●	●	○	●
EXPGSC1	○	○	○	●	●	○	●	○	●	○
SP500	○	○	○	○	○	○	●	○	○	○
Weight	0.511	0.069	0.068	0.067	0.044	0.028	0.026	0.019	0.012	0.011

Table 11 Top 10 models, US GDP growth and inflation, marginal weights

	Top model									
Variable	1	2	3	4	5	6	7	8	9	10
INDPRO	○	○	○	○	○	○	○	○	○	○
A068RC1	●	●	●	●	●	●	○	○	●	●
UNRATE	○	○	○	○	○	○	○	○	○	●
PAYEMS	●	○	○	○	●	●	○	●	○	○
CEV	○	○	○	○	○	○	○	○	○	○
COE	○	○	○	○	○	○	○	○	○	○
PSAVERT	○	○	○	○	○	○	○	○	○	○
CPROFIT	○	○	○	○	○	○	○	○	○	○
PCEPILFE	○	●	○	○	○	●	●	●	●	○
M2SL	○	○	○	○	○	○	○	○	○	○
FEDFUNDS	●	●	●	●	●	●	●	●	●	●
OILPRICE	○	○	○	○	○	○	○	○	○	○
IMPGSC1	○	○	●	○	●	○	●	○	●	●
EXPGSC1	○	○	○	○	○	○	○	○	○	○
SP500	○	○	○	○	○	○	○	○	○	○
Weight	0.426	0.169	0.112	0.097	0.074	0.034	0.026	0.018	0.011	0.009

A Data

The time series used for the US GDP forecasts are from Federal Reserve Economic Data (FRED), and the variables are chosen similarly as in Andersson and Karlsson (2009). The transformation codes for the time series are

T.code	Transformation
1	level (y_t)
2	log difference ($\ln y_t - \ln y_{t-1}$)
3	4 quarter difference ($y_t - y_{t-4}$)
4	4 quarter log difference ($\ln y_t - \ln y_{t-4}$)

Table 12 Variables used for forecasting US GDP

No	Variable	Description	T.code
1	GDPC1*	Real Gross Domestic Product, 1 Decimal	4
2	CPILFESL*	Consumer Price Index for All Urban Consumers: All Items Less Food & Energy	4
3	INDPRO	Industrial Production Index	4
4	A068RC1	Personal outlays	2
5	UNRATE	Civilian Unemployment Rate	3
6	PAYEMS	All Employees: Total nonfarm	2
7	CE16OV	Civilian Employment	2
8	COE	National Income: Compensation of Employees, Paid	2
9	PSAVERT	Personal Saving Rate	1
10	CPROFIT	Corporate Profits with Inventory Valuation Adjustment (IVA) and Capital Consumption Adjustment (CCAdj)	2
11	PCEPILFE	Personal Consumption Expenditures Excluding Food and Energy (Chain-Type Price Index)	2
12	M2SL	M2 Money Stock	2
13	FEDFUNDS	Effective Federal Funds Rate	3
14	OILPRICE	Spot Oil Price: West Texas Intermediate	4
15	IMPGSC1	Real Imports of Goods & Services, 1 Decimal	4
16	EXPGSC1	Real Exports of Goods & Services, 1 Decimal	4
17	SP500	S&P 500 Index	4

* - The variable of interest.

B Simulation Results

Table 13 Posterior variable inclusion probabilities, DGP 2, $T = 100$

		Normal-Wishart prior					
		Predictive likelihood			Marginal likelihood		
p	δ_k	$p(y_2)$	$max[p(x_i)]$	$\frac{p(y_2)}{\max[p(x_i)]}$	$p(y_2)$	$max[p(x_i)]$	$\frac{p(y_2)}{\max[p(x_i)]}$
2	0.2	0.73	0.16	4.72	0.89	0.05	18.61
	0.5	0.79	0.24	3.27	0.95	0.12	7.77
3	0.2	0.64	0.15	4.42	0.82	0.04	21.98
	0.5	0.73	0.24	3.02	0.90	0.09	10.01
		Independent normal Wishart prior					
		Predictive likelihood			Marginal likelihood		
p	δ_k	$p(y_2)$	$max[p(x_i)]$	$\frac{p(y_2)}{\max[p(x_i)]}$	$p(y_2)$	$max[p(x_i)]$	$\frac{p(y_2)}{\max[p(x_i)]}$
2	0.2	0.78	0.18	4.23	0.86	0.06	15.28
	0.5	0.83	0.29	2.88	0.91	0.11	8.14
3	0.2	0.74	0.19	3.92	0.80	0.12	6.43
	0.5	0.81	0.30	2.65	0.85	0.22	3.88

Table 14 Posterior variable inclusion probabilities, DGP 2, $T = 300$

		Normal-Wishart prior					
		Predictive likelihood			Marginal likelihood		
p	δ_k	$p(y_2)$	$max[p(x_i)]$	$\frac{p(y_2)}{\max[p(x_i)]}$	$p(y_2)$	$max[p(x_i)]$	$\frac{p(y_2)}{\max[p(x_i)]}$
2	0.2	0.93	0.18	5.11	1.00	0.02	45.31
	0.5	0.95	0.31	3.06	1.00	0.06	16.95
3	0.2	0.91	0.18	5.14	1.00	0.01	72.75
	0.5	0.94	0.30	3.15	1.00	0.04	27.55
		Independent normal Wishart prior					
		Predictive likelihood			Marginal likelihood		
p	δ_k	$p(y_2)$	$max[p(x_i)]$	$\frac{p(y_2)}{\max[p(x_i)]}$	$p(y_2)$	$max[p(x_i)]$	$\frac{p(y_2)}{\max[p(x_i)]}$
2	0.2	0.93	0.18	5.03	1.00	0.06	17.84
	0.5	0.94	0.29	3.26	1.00	0.11	8.96
3	0.2	0.91	0.19	4.79	1.00	0.05	20.83
	0.5	0.91	0.30	2.98	1.00	0.10	9.79

Table 15 Model Selection, DGP2, Average posterior weight and proportion selected for true model, $T = 100$.

Normal-Wishart prior					
		Predictive likelihood		Marginal likelihood	
p	δ_k	Weight	Selected	Weight	Selected
2	0.2	0.32	0.52	0.71	0.90
	0.5	0.13	0.19	0.47	0.82
3	0.2	0.30	0.39	0.68	0.82
	0.5	0.16	0.19	0.54	0.84
Independent normal Wishart prior					
		Predictive likelihood		Marginal likelihood	
p	δ_k	Weight	Selected	Weight	Selected
2	0.2	0.27	0.56	0.77	0.85
	0.5	0.08	0.17	0.68	0.80
3	0.2	0.24	0.40	0.52	0.62
	0.5	0.07	0.12	0.39	0.46

Table 16 Model Selection, DGP2, Average posterior weight and proportion selected for true model, $T = 300$.

Normal-Wishart prior					
		Predictive likelihood		Marginal likelihood	
p	δ_k	Weight	Selected	Weight	Selected
2	0.2	0.32	0.88	0.90	0.99
	0.5	0.07	0.12	0.72	0.93
3	0.2	0.33	0.80	0.96	0.99
	0.5	0.08	0.09	0.86	0.99
Independent normal Wishart prior					
		Predictive likelihood		Marginal likelihood	
p	δ_k	Weight	Selected	Weight	Selected
2	0.2	0.33	0.92	0.77	0.96
	0.5	0.08	0.17	0.62	0.84
3	0.2	0.32	0.81	0.78	0.97
	0.5	0.07	0.12	0.69	0.88

Table 17 Posterior variable inclusion probabilities, DGP 3, $T = 100$, Normal-Wishart prior

p	δ_k	$p(y_2)$	$p(y_3)$	$max[p(x_i)]$	$\frac{p(y_2)}{\max[p(x_i)]}$	$\frac{p(y_3)}{\max[p(x_i)]}$
			Predictive likelihood			
1	0.2	0.67	0.83	0.12	5.49	6.77
	0.5	0.75	0.88	0.21	3.52	4.18
2	0.2	0.54	0.76	0.12	4.58	6.43
	0.5	0.61	0.82	0.21	2.93	3.93
			Marginal likelihood			
1	0.2	0.85	0.90	0.06	15.34	16.15
	0.5	0.92	0.94	0.12	7.44	7.64
2	0.2	0.76	0.82	0.04	20.44	22.02
	0.5	0.86	0.89	0.09	9.29	9.69

Table 18 Posterior variable inclusion probabilities, DGP 3, $T = 100$, Independent normal Wishart prior

p	δ_k	$p(y_2)$	$p(y_3)$	$max[p(x_i)]$	$\frac{p(y_2)}{\max[p(x_i)]}$	$\frac{p(y_3)}{\max[p(x_i)]}$
			Predictive likelihood			
1	0.2	0.70	0.84	0.13	5.21	6.28
	0.5	0.78	0.90	0.22	3.59	4.13
2	0.2	0.60	0.78	0.14	4.31	5.58
	0.5	0.69	0.84	0.24	2.81	3.45
			Marginal likelihood			
1	0.2	0.61	0.68	0.04	16.27	18.05
	0.5	0.72	0.76	0.08	9.45	10.07
2	0.2	0.84	0.84	0.14	6.02	6.13
	0.5	0.70	0.79	0.20	3.51	3.95

Table 19 Posterior variable inclusion probabilities, DGP 3, $T = 300$, Normal-Wishart prior

p	δ_k	$p(y_2)$	$p(y_3)$	$max[p(x_i)]$	$\frac{p(y_2)}{\max[p(x_i)]}$	$\frac{p(y_3)}{\max[p(x_i)]}$
			Predictive likelihood			
1	0.2	0.90	0.91	0.13	6.69	6.82
	0.5	0.92	0.94	0.20	4.62	4.70
2	0.2	0.85	0.91	0.13	6.61	7.02
	0.5	0.88	0.93	0.20	4.47	4.73
			Marginal likelihood			
1	0.2	1.00	1.00	0.04	23.24	23.24
	0.5	1.00	1.00	0.10	10.48	10.48
2	0.2	1.00	1.00	0.02	43.10	43.11
	0.5	1.00	1.00	0.06	17.47	17.47

Table 20 Posterior variable inclusion probabilities, DGP 3, $T = 300$, Independent normal Wishart prior

p	δ_k	$p(y_2)$	$p(y_3)$	$max[p(x_i)]$	$\frac{p(y_2)}{\max[p(x_i)]}$	$\frac{p(y_3)}{\max[p(x_i)]}$
Predictive likelihood						
1	0.2	0.91	0.92	0.13	6.83	6.96
	0.5	0.93	0.95	0.20	4.72	4.81
2	0.2	0.87	0.92	0.13	6.83	7.22
	0.5	0.90	0.94	0.19	4.60	4.85
Marginal likelihood						
1	0.2	1.00	1.00	0.08	13.27	13.27
	0.5	1.00	1.00	0.14	7.14	7.14
2	0.2	1.00	1.00	0.06	16.53	16.53
	0.5	1.00	1.00	0.13	7.98	7.98

Table 21 Model Selection, DGP3, Average posterior weights and proportion selected for true model, $T = 100$.

Normal-Wishart prior					
		Predictive likelihood		Marginal likelihood	
p	δ_k	Weight	Selected	Weight	Selected
1	0.2	0.28	0.53	0.58	0.78
	0.5	0.15	0.20	0.40	0.88
2	0.2	0.22	0.37	0.53	0.67
	0.5	0.14	0.14	0.48	0.74
Independent normal Wishart prior					
		Predictive likelihood		Marginal likelihood	
p	δ_k	Weight	Selected	Weight	Selected
1	0.2	0.29	0.64	0.35	0.68
	0.5	0.14	0.15	0.44	0.85
2	0.2	0.24	0.40	0.31	0.59
	0.5	0.14	0.20	0.36	0.64

Table 22 Model Selection, DGP3, Average posterior weights and proportion selected for true model, $T = 300$.

Normal-Wishart prior

		Predictive likelihood		Marginal likelihood	
p	δ_k	Weight	Selected	Weight	Selected
1	0.2	0.38	0.84	0.82	0.97
	0.5	0.15	0.10	0.56	0.88
2	0.2	0.36	0.74	0.92	0.97
	0.5	0.16	0.11	0.78	0.93

Independent normal Wishart prior

		Predictive likelihood		Marginal likelihood	
p	δ_k	Weight	Selected	Weight	Selected
1	0.2	0.38	0.86	0.66	0.92
	0.5	0.16	0.09	0.35	0.71
2	0.2	0.38	0.76	0.76	0.93
	0.5	0.16	0.12	0.50	0.81

Figure 15 Forecast performance, DGP1, MSE relative to univariate AR(3), Independent normal-Wishart prior, $T = 100$

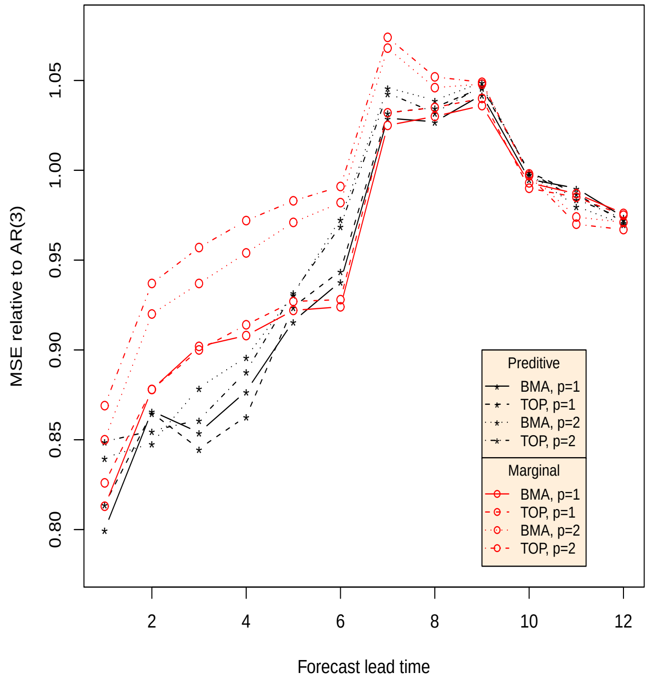


Figure 16 Forecast performance, DGP1, MSE relative to univariate AR(3), Independent normal Wishart prior, $T = 300$

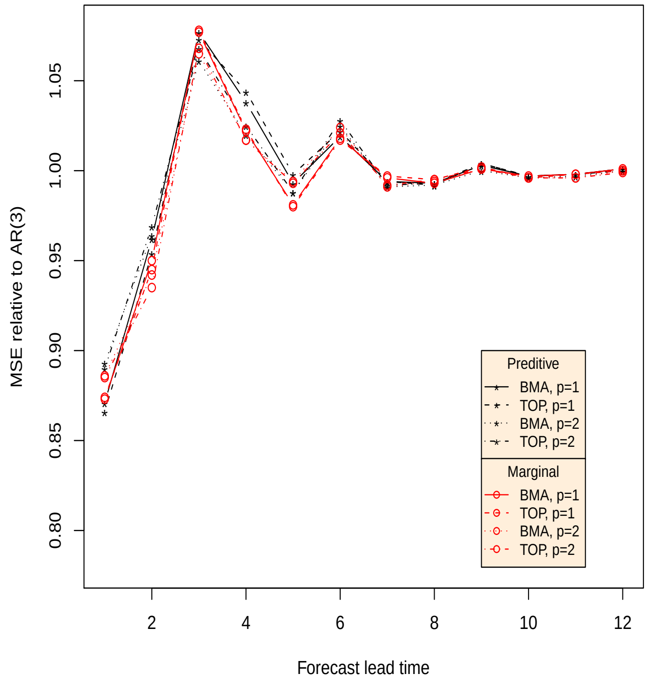


Figure 17 Forecast performance, DGP2, MSE relative to univariate AR(3), Independent normal Wishart prior, $T = 100$

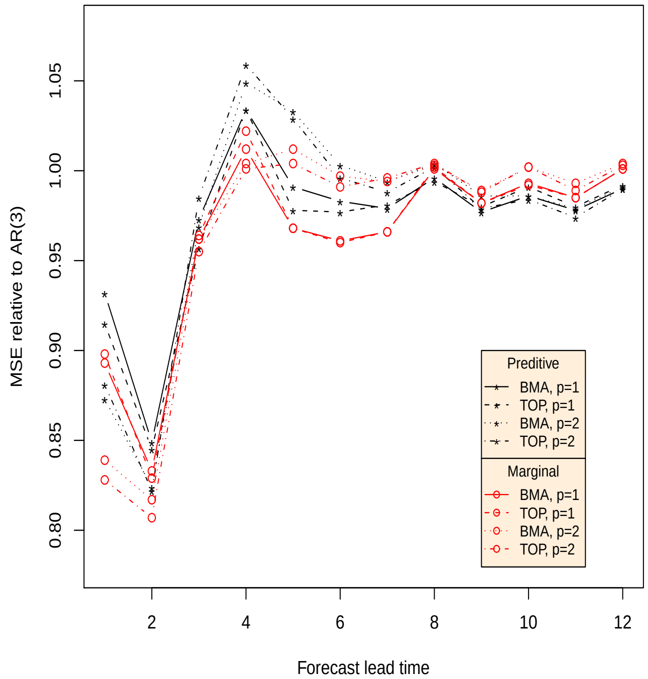


Figure 18 Forecast performance, DGP2, MSE relative to univariate AR(3), Independent normal Wishart prior, $T = 300$

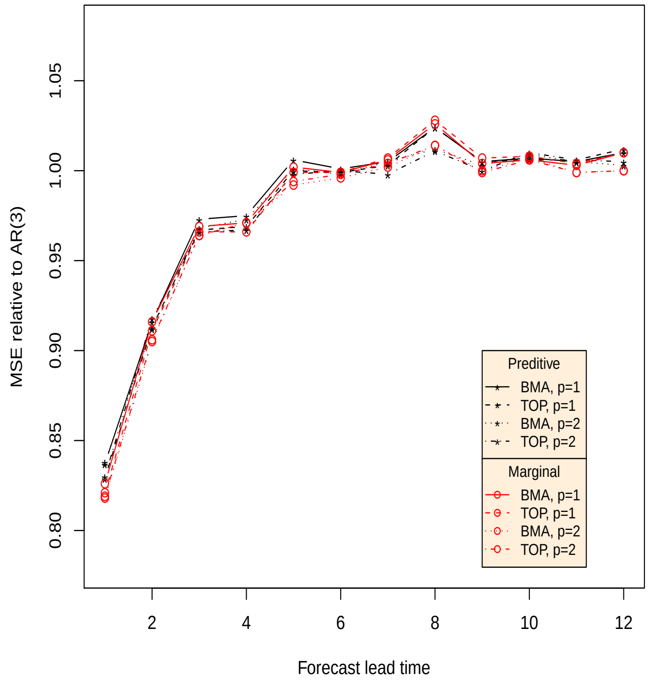


Figure 19 Forecast performance, DGP3, MSE relative to univariate AR(3), Independent normal Wishart prior, $T = 100$

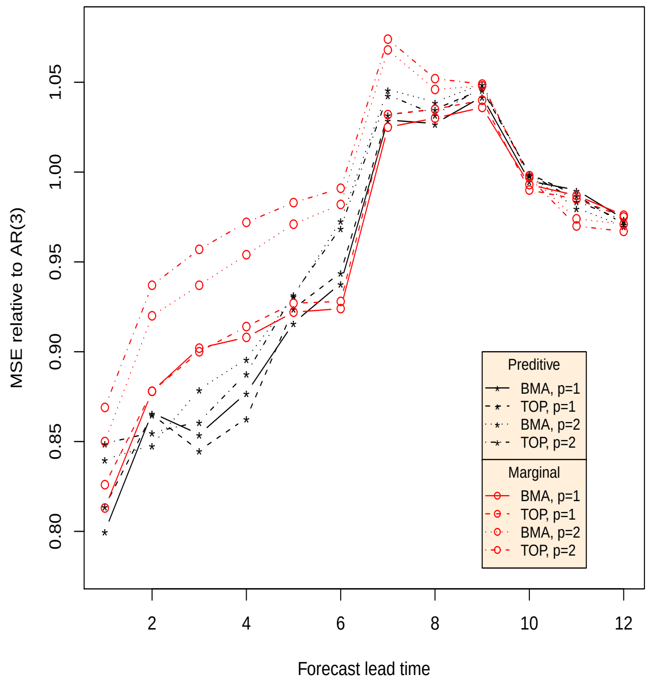
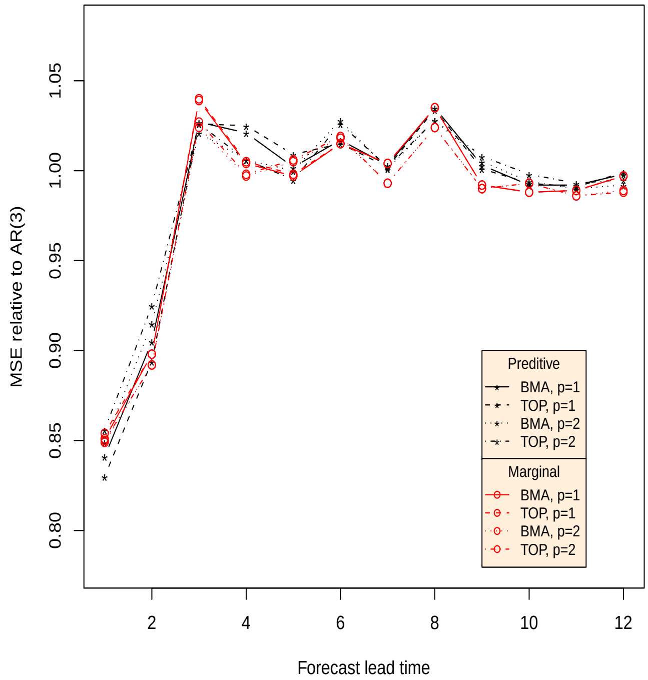


Figure 20 Forecast performance, DGP3, MSE relative to univariate AR(3), Independent normal Wishart prior, $T = 300$



References

Andersson, M. and Karlsson, S. (2009), Bayesian forecast combination for VAR models, *in* S. Chib, G. Koop and B. Griffiths, eds, 'Bayesian Econometrics', Vol. 23 of *Advances in Econometrics*, Emerald.

Bauwens, L., Lubrano, M. and Richard, J.-F. (1999), *Bayesian Inference in Dynamic Econometric Models*, Oxford University Press.

Bernanke, B. S., Boivin, J. and Eliasz, P. (2005), 'Measuring the effects of monetary policy: A factor-augmented vector autoregressive (favar) approach', *The Quarterly Journal of Economics* **120**, 387–422.

Chib, S. (2008), 'Marginal likelihood from the Gibbs output', *Journal of Econometrics* **142**, 553–580.

Chib, S. and Jeliazkov, I. (2001), 'Marginal likelihood from the Metropolis-Hastings output', *Journal of the American Statistical Association* **96**, 270–281.

Eklund, J. and Karlsson, S. (2007), 'Forecast combination and model averaging using predictive measures', *Econometric Reviews* **26**, 329–363.

Gelfand, A. E. and Dey, D. K. (1994), 'Bayesian model choice: Asymptotics and exact calculations', *Journal of the Royal Statistical Society, Ser. B* **56**, 501–514.

George, E. I., Sun, D. and Ni, S. (2001), 'Bayesian stochastic search for var model restrictions', *Journal of the American Statistical Association* **96**, 270–281.

Geweke, J. (1999), 'Using simulation methods for bayesian econometric models: Inference, development and communication', *Econometric Reviews* **18**, 1–126. with discussion.

Jarociński, M. and Maćkowiak, B. (2011), Choice of variables in vector autoregressions. Manuscript.

Kadiyala, K. R. and Karlsson, S. (1997), 'Numerical methods for estimation and inference in bayesian var-models', *Journal of Applied Econometrics* **12**, 99–132.

Litterman, R. B. (1979), Techniques of forecasting using vector autoregressions, Working Paper 115, Federal Reserve Bank of Minneapolis.

Litterman, R. B. (1986), 'Forecasting with bayesian vector autoregressions - five years of experience', *Journal of Business & Economic Statistics* **4**, 25–38.