WORKING PAPER SERIES

WORKING PAPER NO 13, 2007

ESI

Bayesian Forecast Combination for VAR Models

By

Michael K Andersson
Sveriges Riksbank

Sune Karlsson
Örebro University

Michael.Andersson@riksbank.se

Sune.Karlsson@esi.oru.se

http://www.oru.se/esi/wps

SE-701 82 Örebro

SWEDEN

ISSN 1403-0586

Bayesian forecast combination for VAR models*

Michael K Andersson† Sune Karlsson‡
Sveriges Riksbank Örebro University

November 12, 2007

Abstract

We consider forecast combination and, indirectly, model selection for VAR models when there is uncertainty about which variables to include in the model in addition to the forecast variables. The key difference from traditional Bayesian variable selection is that we also allow for uncertainty regarding which endogenous variables to include in the model. That is, all models include the forecast variables, but may otherwise have differing sets of endogenous variables. This is a difficult problem to tackle with a traditional Bayesian approach. Our solution is to focus on the forecasting performance for the variables of interest and we construct model weights from the predictive likelihood of the forecast variables. The procedure is evaluated in a small simulation study and found to perform competitively in applications to real world data.

Keywords: Bayesian model averaging, Predictive likelihood, GDP forecasts
JEL-codes: C11, C15, C32, C52, C53

---

*The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Executive Board of Sveriges Riksbank. We have benefitted from discussions with Martin Sköld and Mattias Villani.

†Michael.Andersson@riksbank.se

‡Sune.Karlsson@esi.oru.se

1 Introduction

The increasing availability of data has spurred the interest in forecasting procedures that can extract information from a large number of variables in an efficient manner. Examples include the diffusion indexes of Stock and Watson (2002*b*) and procedures based on combining forecasts from many models as in Jacobson and Karlsson (2004), see Stock and Watson (2006) for a recent review and additional references. While this development has clear implications for policy makers such as central banks (see e.g. Bernanke and Boivin (2003)) procedures of this type are not particularly widespread in central banks. Notable practitioners are Sveriges Riksbank, the Bank of England and the Bank of Canada. These central banks employ a wide variety of model approaches, ranging from simple univariate time series models to highly sophisticated multivariate non-linear models. While a great many models are used, the procedures are easy to manage and highly automated (see, for example, Andersson and Löf (2007) and Kapetanios, Labhard and Price (2007)).

One possible reason for the apparent lack of interest in the possibilities offered by these procedures is that the literature has largely focused on univariate forecasting procedures. This paper attempts to bridge this gap by proposing a Bayesian procedure for combining forecasts from multivariate forecasting models, e.g. VAR models. Standard applications of Bayesian model averaging suffer from a basic difficulty in this context, when additional variables are included and modelled the connection between the overall measure of fit for the model, the marginal likelihood, and the expected forecasting performance for the variables of interest is lost. It is easy to see that the (multivariate) marginal likelihood can change when a model is modified by adding, removing or exchanging variables without this having the corresponding effect on the predictive ability for the variable of interest.

We circumvent this problem by focusing on the predictive performance for the variables of interest and base the forecast combination on the predictive likelihood as proposed by Eklund and Karlsson (2007) in the context of univariate forecasting models. While the basic predictive likelihood is also multivariate it is meaningful to marginalize the predictive distribution with respect to the auxiliary variables yielding a univariate predictive distribution and corresponding predictive likelihood. Forecasts from different models can then be combined using weights based on the univariate predictive likelihood.

Specifically we consider forecast combination and, indirectly, model selection for VAR models when there is uncertainty about which additional variables to include in the model. Given a set of auxiliary variables that are expected to be useful for modelling and forecasting the variable of interest we consider the set of models that arise when taking all possible combinations of the auxiliary variables. The forecasts from these models are then combined using weights based on the predictive likelihood at the relevant forecast horizon.

In most cases the predictive likelihood will not be available in closed form. Instead we use MCMC methods to simulate the predictive distribution and estimate the density function from the MCMC output. In addition the MCMC output is used to obtain forecast intervals both for forecasts based on a single model and the combined forecast.

The procedure is evaluated in a simulation study and found to perform competitively in an application to forecasting the growth rate of US GDP.

2 Bayesian Forecast Combination

Bayesian forecast combination is a straightforward application of Bayesian model averaging (see Hoeting, Madigan, Raftery and Volinsky (1999) for an introduction to Bayesian model averaging and Min and Zellner (1993), Jacobson and Karlsson (2004) and Koop and Potter (2004) for applications of Bayesian model averaging to forecasting and Timmermann (2006) for a review of forecast combination). Suppose that the forecaster has a set, $\mathfrak{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_M\}$, of M possible forecasting models available, each specified in terms of a likelihood function $L(\mathbf{y}|\theta_i, \mathcal{M}_i)$ and prior distribution for the parameters in the model, $p(\theta_i|\mathcal{M}_i)$. In addition the forecaster assigns prior probabilities, $p(\mathcal{M}_i)$, to each model, reflecting the forecasters prior confidence in the models. The posterior model probabilities can then be obtained by routine application of Bayes theorem

$$p(\mathcal{M}_i|\mathbf{y}) = \frac{m(\mathbf{y}|\mathcal{M}_i)\, p(\mathcal{M}_i)}{\sum_{j=1}^{M} m(\mathbf{y}|\mathcal{M}_j)\, p(\mathcal{M}_j)} \tag{1}$$

where

$$m(\mathbf{y}|\mathcal{M}_i) = \int L(\mathbf{y}|\theta_i, \mathcal{M}_i)\, p(\theta_i|\mathcal{M}_i)\, d\theta_i \tag{2}$$

is the marginal likelihood of model $\mathcal{M}_i$. The combined forecast is obtained as

$$E(y_{T+h}|\mathbf{y}) = \sum_{j=1}^{M} E(y_{T+h}|\mathbf{y}, \mathcal{M}_j)\, p(\mathcal{M}_j|\mathbf{y})$$

by weighting the forecasts from each model by the posterior model probabilities. It is easily seen that the Bayesian forecast combination is a special case of the general result that the marginal (over all models) posterior distribution for some function ϕ of the parameters is

$$p(\phi|\mathbf{y}) = \sum_{j=1}^{M} p(\phi|\mathbf{y}, \mathcal{M}_j)\, p(\mathcal{M}_j|\mathbf{y}). \tag{3}$$

The crucial feature of the marginal distribution (3) is that it takes account of both parameter and model uncertainty. It is thus relatively easy to produce prediction intervals that incorporates model uncertainty.

The marginal likelihood (2) is the basic Bayesian measure of fit of a model and is a joint assessment of how well the likelihood and parameter prior agrees with the data. It is the key quantity for determining the posterior model probabilities and hence the weights assigned to the forecasts from the different models.

2.1 Predictive Likelihood

The marginal likelihood is well suited for combination of univariate forecasting models but, unfortunately, problematic when it comes to the combination of forecasts from multivariate forecasting models. Multivariate forecasting models, e.g. VAR-models, are typically built with the express purpose of forecasting a single variable and the remaining dependent variables in the model are only included if they are deemed to improve the forecasting performance for the variable of interest. As the marginal likelihood measures the fit of the whole model it is easy to see that the forecast performance can remain unaffected by a change in the model that either increases or decreases the marginal likelihood. This can happen when a dependent is exchanged for another variable or the dimension of the model changes as variables are added or dropped from the model.

To overcome these problems with the marginal likelihood we propose to base the forecast combination on the predictive likelihood as suggested by Eklund and Karlsson (2007) in the context of univariate forecasting models. Our primary motivation for using the predictive likelihood is that it is meaningful to marginalize this over the non-forecasted variables to obtain a measure that is focused on the variable of interest. An added benefit of the predictive likelihood is that it is a true out of sample measure of fit whereas the marginal likelihood depends on the predictive content of the parameter prior. When combining the forecasts from a large set of models it is often to time consuming to provide well thought out parameter priors for all the models. Instead uninformative default priors such as the ones suggested by Fernández, Ley and Steel (2001) are used and with this type of prior the marginal likelihood essentially reduces to an in-sample measure of fit.

Our use of the predictive likelihood is based on a split of the data, $\mathbf{Y} = (\mathbf{y}_1', \mathbf{y}_2', \ldots, \mathbf{y}_T')'$, 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 $m = T - n$, where $\mathbf{y}_t = (y_{1t}, \ldots, y_{qt})$ is the vector of modelled variables. The training sample is used to convert the prior into a posterior and the predictive likelihood is obtained by marginalizing out the parameters from the joint distribution of data and parameters,

$$p\left(\widetilde{\mathbf{Y}}_n \middle| \mathbf{Y}_n^*, \mathcal{M}_i\right) = \int L\left(\widetilde{\mathbf{Y}}_n \middle| \theta_i, \mathbf{Y}_n^*, \mathcal{M}_i\right) p\left(\theta_i \middle| \mathbf{Y}_n^*, \mathcal{M}_i\right) d\theta_i. \tag{4}$$

Technically this is the predictive distribution of an unknown $\widetilde{\mathbf{Y}}_n$ conditional on the training sample, $\mathbf{Y}_n^*$. When evaluated at the observed $\widetilde{\mathbf{Y}}_n$ (4) provides a measure of the out of sample predictive performance and we refer to this as the predictive likelihood. Since our primary interest is to forecast a subset of the q modelled variables the multivariate predictive likelihood (4) suffers from the same drawback as the marginal likelihood in that it is not directly informative about the forecasting performance for the variable of interest. To overcome this we marginalize the predictive distribution of $\widetilde{\mathbf{Y}}_n$ with respect to the auxiliary variables, with y_1 the variable of interest we have

$$p\left(\widetilde{\mathbf{y}}_{1,n} \middle| \mathbf{Y}_n^*, \mathcal{M}_i\right) = \int p\left(\widetilde{\mathbf{Y}}_n \middle| \mathbf{Y}_n^*, \mathcal{M}_i\right) d\widetilde{\mathbf{y}}_{2,n} \ldots d\widetilde{\mathbf{y}}_{q,n} \tag{5}$$

the marginal predictive likelihood for the hold of sample of y_1 as a measure of the average predictive performance for the variable of interest.

Replacing the marginal likelihood with the marginal predictive likelihood in (1) yields the predictive weights

$$w\left(\mathcal{M}_i\right|\widetilde{\mathbf{y}}_{1,n},\mathbf{Y}_n^*) = \frac{p\left(\tilde{\mathbf{y}}_{1,n}\right|\mathbf{Y}_n^*,\mathcal{M}_i)\,p\left(\mathcal{M}_i\right)}{\sum_{j=1}^{M} p\left(\tilde{\mathbf{y}}_{1,n}\right|\mathbf{Y}_n^*,\mathcal{M}_j)\,p\left(\mathcal{M}_j\right)} \tag{6}$$

and the combined forecast

$$\hat{y}_{T+h} = \sum_{j=1}^{M} E\left(y_{T+h}\right|\mathbf{Y},\mathcal{M}_j)\,w\left(\mathcal{M}_i\right|\widetilde{\mathbf{y}}_{1,n},\mathbf{Y}_n^*).$$

Note that the forecasts from each model is conditional on all available data up to time T. That is, the model specific posterior is based on the full sample and the forecast is the expected value of y_{T+h} with respect to this posterior. The sample split is only used for the purpose of calculating the predictive weights.

Comparing (6) with the posterior model probabilities in (1) it is clear that there are two distinct differences between using predictive weights for forecast combination and standard Bayesian model averaging. The first difference is the use of prior model probabilities in (6) instead of posterior model probabilities based on the training sample as suggested by (1). This is the sample split idea; the training sample is used to learn about the parameters of each model and the hold out sample is used to assess the forecasting performance and update the model weights. The second difference is that we marginalize out the auxiliary variables from the predictive likelihood to produce a measure of forecast performance that focuses on the variable of interest.

While the predictive weights (6) strictly speaking can not be interpreted as posterior probabilities they have several appealing properties in addition to providing a basis for meaningful marginalization with respect to the auxiliary variables in the model.

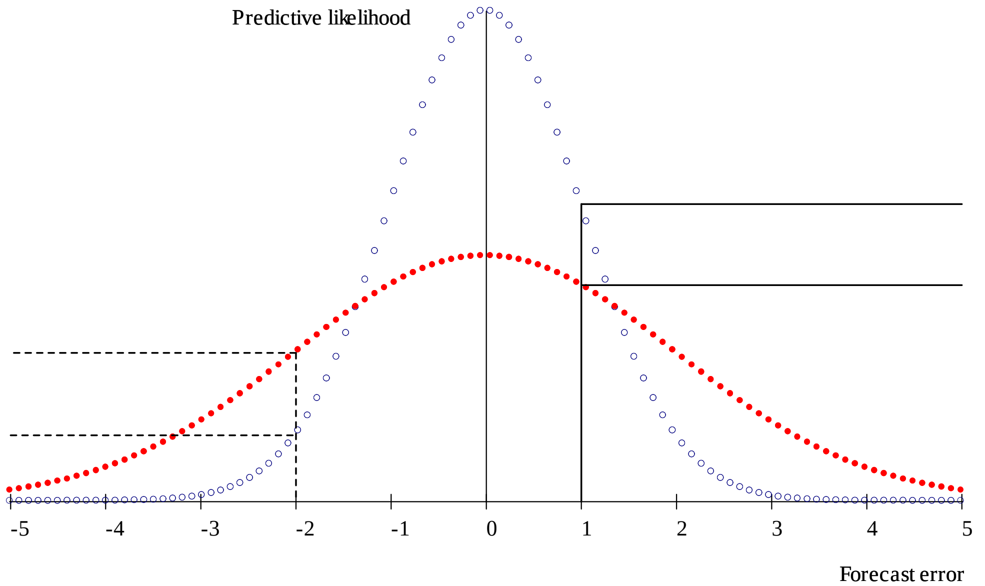
- Proper prior distributions are not required for the parameters. The predictive likelihood is, in contrast to the marginal likelihood, well defined as long as the posterior distribution of the parameters conditioned on the training sample is proper.

- The predictive likelihood is not an absolute measure of forecasting performance. Instead it is relative to the precision of forecasts implied by the model and models with a good in-sample fit are penalized when a "good" and "bad" model forecast both forecasts poorly. This is illustrated in Figure 1. If the forecast error is small (1, solid lines) as can be expected from a model with good in-sample fit, the predictive likelihood prefers the "good" model but the "bad" model is favoured if the forecast error (-2, dashed lines) is larger than what can be expected from the "good" model. The predictive weights will thus be small for models that overfit the data or models that suffer from structural breaks.

2.2 Dynamic Models

The predictive densities (4) and (5) are joint predictive distributions for lead times $h = 1$ through $h = m = T - n$. For dynamic models where the forecast precision

Figure 1 Predictive likelihood for a "good" and a "bad" model



Predictive likelihood for a "good" model (circles) and a "bad" model (dots) evaluated at small (solid lines) and large (dashed lines) forecast errors.

typically deteriorates as the lead time increases these will not be appropriate measures of forecast performance if the focus is on producing forecasts for a few select lead times. One solution is to set m to the largest lead time, H, considered but this will typically be small (say 8 quarters) and the Monte Carlo experiments in Eklund and Karlsson (2007) indicates that the hold out sample should be large, on the order of 70% of the data. To combine these two requirements we suggest using a series of short horizon predictive likelihoods,

$$g\left(\mathbf{Y},n|\mathcal{M}_i\right) = \prod_{t=n}^{T-h_k} p\left(y_{1,t+h_1},\ldots,y_{1,t+h_k}|\,\mathbf{Y}_t^*,\mathcal{M}_i\right) \tag{7}$$

where $h_1,\ldots,h_k$ represents the lead times at which we wish to evaluate the forecast performance.

The use of the predictive likelihood in dynamic models is complicated by the fact that the predictive likelihood is not available in closed form for lead times $h > 1$. Instead the predictive distribution must be simulated and the predictive likelihood estimated from the simulation output. Standard density estimation techniques can be used for this purpose and works quite well if the predictive likelihood is evaluated at a single lead time. Evaluating the predictive likelihood at multiple horizons leads to more complex multivariate density estimation.

To facilitate the use of multiple horizon predictive likelihoods we take advantage of the model structure and use the idea of Rao-Blackwellization to estimate the predictive likelihood. Consider the task of evaluating the unknown density f_u at $u = x$ when we have draws from the joint distribution of (u,v) or only the marginal distribution of v and the conditional density $f_{u|v}$ is known. We want $f_u(x) = \int f_{u,v}(x,v)\,dv = \int f_{u|v}(x,v)\,f_v(v)\,dv = E_v\left[f_{u|v}(x,v)\right]$ where we make the dependence of $f_{u|v}$ on v explicit by including it as an argument to the function. A simple Monte Carlo estimate is then given by $\widehat{f}_u(x) = \frac{1}{R}\sum_{i=1}^{R} f_{u|v}(x, v_i^*)$ where v_i^* are the draws from the marginal distribution of v. The Rao-Blackwellized estimate will in general be quite precise even for moderate sample sizes and preserves any smoothness properties of the underlying density.

In our case the conditioning variables are the parameters of the VAR-model, $\mathbf{\Gamma}$ and $\mathbf{\Psi}$ and we wish to estimate the predictive likelihood for the subset of variables and lead times, $y_{1,t+h_1}, \ldots, y_{1,t+h_k}$. For the VAR-model

$$\begin{aligned}\mathbf{y}_t &= \sum_{i=1}^{p} \mathbf{y}_{t-i}\mathbf{A}_i + \mathbf{x}_t\mathbf{C} + \mathbf{u}_t \\ &= \mathbf{z}_t\mathbf{\Gamma} + \mathbf{u}_t\end{aligned} \tag{8}$$

with $\mathbf{z}_t = (\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-p}, \mathbf{x}_t)$ and normally distributed errors, $\mathbf{u}_t \sim N(0, \mathbf{\Psi})$, the joint lead time 1 through $H = \max(h_1, \ldots, h_k)$ predictive distribution conditional on the parameters, $p(\mathbf{y}_{t+1}, \ldots, \mathbf{y}_{t+H} | \mathbf{Y}_t^*, \mathcal{M}_i, \mathbf{\Gamma}, \mathbf{\Psi})$, is multivariate normal (see Lütkepohl (1993) for details). Consequently, the conditional predictive distribution for the subset of interest, $p(y_{1,t+h_1}, \ldots, y_{1,t+h_k} | \mathbf{Y}_t^*, \mathcal{M}_i, \mathbf{\Gamma}, \mathbf{\Psi})$, is also multivariate normal. The Rao-Blackwellized estimate of $p(y_{1,t+h_1}, \ldots, y_{1,t+h_k} | \mathbf{Y}_t^*, \mathcal{M}_i)$ is then obtained as

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

by averaging over draws $\mathbf{\Gamma}^{(i)}$ and $\mathbf{\Psi}^{(i)}$ from the posterior distribution based on $\mathbf{Y}_t^*$. The draws from the posterior distribution of the parameters are, in our case, obtained from a standard Gibbs sampler.

The estimates of the predictive weights, finally, are formed as

$$\widehat{w}(\mathcal{M}_i | \widetilde{\mathbf{y}}_{1,n}, \mathbf{Y}_n^*) = \frac{\widehat{g}(\mathbf{Y}, n | \mathcal{M}_i)\, p(\mathcal{M}_i)}{\sum_{j=1}^{M} \widehat{g}(\mathbf{Y}, n | \mathcal{M}_j)\, p(\mathcal{M}_j)} \tag{9}$$

with

$$\widehat{g}(\mathbf{Y}, n | \mathcal{M}_i) = \prod_{t=n}^{T-h_k} \widehat{p}(y_{1,t+h_1}, \ldots, y_{1,t+h_k} | \mathbf{Y}_t^*, \mathcal{M}_i). \tag{10}$$

3 Prior Specification

We use a Normal-Diffuse prior on the parameters in the VAR-model (8), i.e. $\operatorname{vec}(\mathbf{\Gamma}) \sim N(\gamma_0, \mathbf{\Sigma}_0)$ and $\pi(\mathbf{\Psi}) \propto |\mathbf{\Psi}|^{-(q+1)/2}$, see Kadiyala and Karlsson (1997) for details and the Gibbs sampler for simulating from the posterior distribution of $\mathbf{\Gamma}$ and $\mathbf{\Psi}$. The prior for $\mathbf{\Gamma}$ is a Litterman type prior. That is, γ_0 is zero except for elements corresponding to the first own lag of variables. These are set to unity for variables believed to be non-stationary and to 0.9 for stationary variables. $\mathbf{\Sigma}_0$ is a diagonal matrix and 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.

The model prior is given by

$$\pi(\mathcal{M}_j) \propto \prod_{k=1}^{K} \delta_k^{d_k} (1 - \delta_k)^{1-d_k}$$

where $d_k = 1$ if variable k is included in the model and δ_k is the prior inclusion probability of variable k.

4 Monte Carlo Experiment

We use three small Monte Carlo experiments to evaluate the forecasting performance of forecast combinations based on the predictive weights (9). The data generating processes are a bivariate VAR(1),

$$\text{DGP 1: } \mathbf{y}_t = \mathbf{y}_{t-1} \begin{pmatrix} 0.5 & 0.2 \\ 0.5 & 0.5 \end{pmatrix} + \mathbf{u}_t, \tag{11}$$

a bivariate VAR(2),

$$\text{DGP 2: } \mathbf{y}_t = \mathbf{y}_{t-1} \begin{pmatrix} 0.5 & 0.2 \\ 0.5 & 0.5 \end{pmatrix} + \mathbf{y}_{t-2} \begin{pmatrix} 0.1 & 0.1 \\ 0.2 & -0.3 \end{pmatrix} + \mathbf{u}_t, \tag{12}$$

a trivariate VAR(1),

$$\text{DGP 3: } \mathbf{y}_t = \mathbf{y}_{t-1} \begin{pmatrix} 0.5 & 0.2 & 0.1 \\ 0.5 & 0.5 & 0.1 \\ 0.5 & 0.3 & 0.2 \end{pmatrix} + \mathbf{u}_t, \tag{13}$$

and, finally, a univariate AR(2)

$$\text{DGP 4: } y_t = 0.5y_{t-1} + 0.3y_{t-2} + u_t. \tag{14}$$

In addition we generate a set of 5 extraneous variables as

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

with $u_{i,t}$, and $e_{i,t}$ iid standard normal random variables. The last, white noise, extraneous variable is dropped with the trivariate VAR-model and a second white noise series is added for the univariate AR-model so that the generated data sets in each Monte Carlo experiment consists of seven variables. For each experiment we generate 100 data sets of length 112 with the last 12 observations set aside for forecast evaluation.

The variable to be forecasted is $y_{1,t}$. For the bivariate DGPs we consider the 42 models arising from modelling $y_{1,t}$ alone or together with combinations of $y_{2,t}$ and $z_{1,t}, \ldots, z_{5,t}$ with a maximum of four variables in the model, for the trivariate DGP we consider the 57 possible models when allowing a maximum of five variables in the model. For the univariate AR(2) z_2 simplifies to $z_{2,t} = 0.5 z_{2,t-1} + e_{2,t}$ and we consider models with a maximum of 4 variables. We use two settings for the lag length of the VAR-models, $p = 2$ and $p = 4$.

We are particularly concerned about the number of observations needed for the hold out sample, for this we consider three cases, $m = 30$, $m = 50$ and $m = 70$, ($m = 70$ is not used in combination with lag length 4 in the estimated models since this would reduce the number of available observations too much) and the effect of the lead time used for the calculation of the predictive weights, here we consider eight alternatives, the single lead times $h = 1$, 2, 3, 4 and 8 and the multiple lead times $h = (1, 2, 3, 4)$, $h = (1, 2, 3, 4, 5, 6, 7, 8)$ and $h = (1, 4, 8)$. We also experiment with two specifications of the model prior, setting $\delta_k = 0.2$ implying a prior expected model size of 2.15 when we allow for four variables in the model and 2.19 when we allow five variables. The other settings $\delta_k = 0.5$, with all models equally likely and prior expected model sizes 3.29 and 3.74. The prior for $\mathbf{\Gamma}$ is specified with $\pi_1 = 0.5$, $\pi_2 = 0.5$, $\pi_3 = 1$ and $\pi_4 = 5.0$.

When conducting the Monte Carlo exercise we simplify the estimation of the predictive likelihoods by not updating the posterior distribution of the parameters as t increases in the product (10), this allow us to perform all the calculations for the predictive weights within a single Gibbs sampler run instead of running one Gibbs sampler for each value of t.[1] The predictive likelihoods are estimated based on 5000 draws from Markov chain and the final forecast, $E\left(y_{T+h} \middle| \mathbf{Y}, \mathcal{M}_j\right)$, is estimated from 5000 draws from the Markov chain based on the full sample. To increase the precision of the estimate we use antithetic variates where an antithetic draw of $\mathbf{\Gamma}$, conditional on $\mathbf{\Psi}$, is obtained in each step of the Markov chain.

4.1 Results

We will focus on DGP 1, a bivariate VAR(1), when the models are estimated with lag length $p = 2$ when reporting the results. The qualitative results are similar for the other DGPs as well as models estimated with $p = 4$. A comprehensive set of results are available in Appendix B.

Table 1 reports on the posterior variable inclusion "probabilities", or more precisely the sum of the predictive weights for the set of models containing the variable. It is clear that the procedure is able to discriminate between the variable y_2 which is in the true model and the extraneous variables. The strongest discrimination is achieved when the predictive likelihood is evaluated at $h = 1$. This is not too surprising given that prediction intervals rapidly becomes very wide as the forecast horizon increases with a correspondingly diminishing discriminatory power. Longer lead times might, however, be important for seasonal or cyclical data. This is to

[1]We do a limited check on the effect of not updating the prior by rerunning a few experiments for the first DGP with the posterior updated as new observations are added. The results are slightly better when the posterior is updated, particularly for $m = 70$, but overall the differences are small. Se Tables B2, B5, B7 and B9 for details on the performance.

Table 1 Predictive weights for variables, DGP 1, models estimated with lag length $p = 2$

Model prior, $\delta_k = 0.2$

h	hold out sample, $m = 30$			hold out sample, $m = 70$		
	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$
1	0.79	0.17	4.71	0.92	0.15	6.11
4	0.42	0.19	2.26	0.49	0.20	2.47
8	0.31	0.19	1.57	0.28	0.20	1.40
$1-4$	0.76	0.17	4.38	0.79	0.19	4.10
$1-8$	0.70	0.18	3.81	0.66	0.18	3.76
$1,4,8$	0.76	0.17	4.49	0.76	0.16	4.68

Model prior, $\delta_k = 0.5$

h	hold out sample, $m = 30$			hold out sample, $m = 70$		
	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$
1	0.88	0.31	2.79	0.96	0.28	3.48
4	0.60	0.36	1.67	0.63	0.32	1.96
8	0.49	0.37	1.32	0.40	0.32	1.27
$1-4$	0.85	0.30	2.89	0.84	0.26	3.25
$1-8$	0.78	0.28	2.80	0.71	0.22	3.27
$1,4,8$	0.85	0.29	2.88	0.82	0.23	3.64

$p(\cdot)$ denotes the predictive weight for the variable.

some extent indicated by the results for DGP 2 which contains a cycle. Evaluating the predictive likelihood at multiple horizons discriminates almost as well as the single $h = 1$ and can be a useful alternative. Increasing the size of the hold out sample is beneficial for discriminating between the variables although the estimation sample can obviously not be made too small (in particular when the posterior is not updated with new observations and always based on the first $T - m$ observations). As can be expected we also achieve better discrimination with the $\delta_k = 0.2$ model prior which favours small models.

Table 2 summarizes the model selection properties of the predictive likelihood. The predictive weights for the true model are not particularly large but the performance is reasonable in terms of model selection. With the $\delta_k = 0.2$ model prior the correct model is selected in between 70% and 87% of the Monte Carlo replicates when the predictive likelihood is evaluated at $h = 1$. Performance is, on the other hand, quite poor with the uninformative model prior which favours large models.

Figure 2 summarizes the forecast performance for DGP 1 and models estimated with lag length $p = 2$. The figure compare the root mean square forecast error (RMSE) for the forecast combination to that of the forecasts from the model with only $y_{1,t}$, i.e. an AR(2). There is clearly a substantial gain for shorter forecast lead times. The larger hold out sample, $m = 70$, provides the best forecasts together with predictive criteria that puts weight on lead time 1. The difference between the $\delta_k = 0.2$ and $\delta_k = 0.5$ model priors is small for this DGP and models estimated with lag length $p = 4$ gives slightly worse forecasts.

Table 2 Model selection, DGP 1, models estimated with lag length $p = 2$. Average predictive weight and proportion selected for true model.

h	Model prior, $\delta_k = 0.2$				Model prior, $\delta_k = 0.5$			
	hold out, $m = 30$		hold out, $m = 70$		hold out, $m = 30$		hold out, $m = 70$	
	Weight	Selected	Weight	Selected	Weight	Selected	Weight	Selected
1	0.31	0.87	0.44	0.70	0.08	0.20	0.18	0.39
4	0.16	0.29	0.23	0.34	0.05	0.18	0.15	0.26
8	0.12	0.19	0.12	0.13	0.05	0.25	0.10	0.15
$1-4$	0.33	0.61	0.42	0.46	0.13	0.28	0.33	0.37
$1-8$	0.33	0.50	0.31	0.34	0.19	0.30	0.28	0.28
$1,4,8$	0.34	0.66	0.41	0.45	0.14	0.31	0.34	0.40

The results for DGP 2 shown in Figure 3 show a larger improvement from the forecast combination at lead time 1 than for DGP 1 but the results are slightly worse than an AR(2) at the longer lead times. Again, the forecasts combinations based on the predictive likelihood evaluated at $h = 4$ and 8 provides the least improvement on an AR(2). Performance is slightly better for the $\delta_k = 0.5$ model prior with smaller differences between combinations based on predictive likelihoods evaluated at different horizons.

With DGP 3 (Figure 4) the forecast combination improves on an AR(2) at all but the longest lead times. The difference between the different forecast combinations is small except for when the predictive likelihood is evaluated at $h = 8$ which performs worse than the other combinations. The difference between model priors is very small, the $\delta_k = 0.5$ prior does slightly better at longer lead times and the $\delta_k = 0.2$ prior does slightly better at short lead times.

DGP 4, finally, provides a check on the forecast performance when the true model is a univariate AR(2). The performance of the predictive likelihood forecast combination, depicted in Figure 5, is very close to the AR(2) and the RMSE never exceeds that of the AR(2) by more than 1.6% in the experiments we performed. The forecast performance is very stable across the prior settings, the horizons at which the predictive likelihood is evaluated and the size of the hold out sample. There are, on the other hand, substantial differences in terms of variable and model selection. With the uniform $\delta_k = 0.5$ prior, which favours large models, the correct univariate model is never selected in more than 38% of the replicates. This indicates the importance of the shrinkage prior on the parameters as a protection agains overfitting. The model selection performs better with the $\delta_k = 0.2$ prior where the correct model is selected in 39% to 89% of the replicates.

Overall it is clear that forecast combination based on the predictive likelihood can improve substantially on the common benchmark of a univariate AR-model. The improvement is larger for short lead times and is also larger for more complex DGPs. The performance is in general better when the predictive likelihood is evaluated at a single short horizon although the use of multiple horizons may be more robust. With a single horizon the use of standard density estimation techniques is uncomplicated and the procedure generalizes readily to situations where the model structure does not allow the use of the Rao-Blackwellization device.

Figure 2 RMSE for forecast combination relative to AR(2), DGP 1, $\delta_k = 0.2$

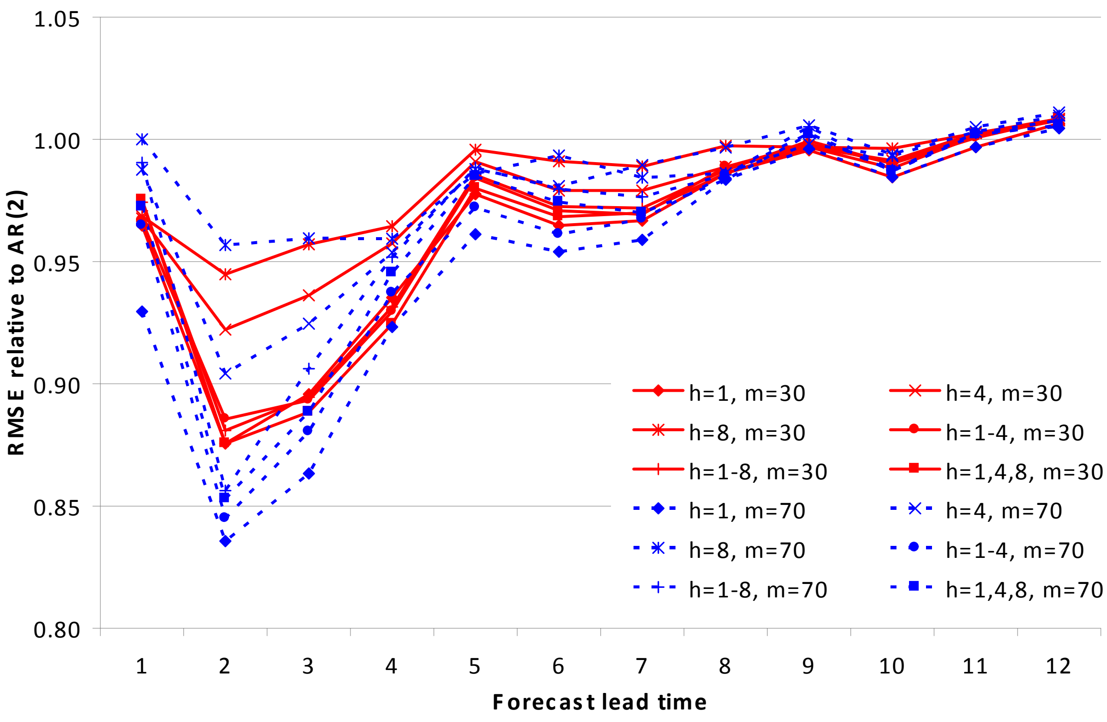


Figure 3 RMSE for forecast combination relative to AR(2), DGP 2, $\delta_k = 0.2$

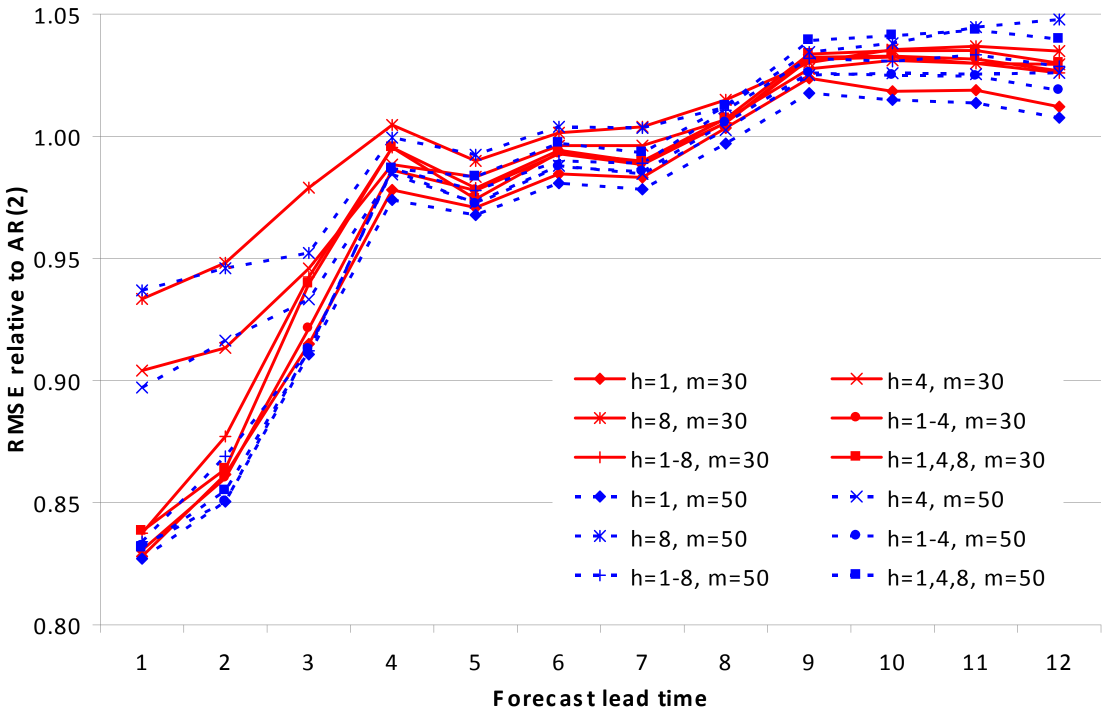


Figure 4 RMSE for forecast combination relative to AR(2), DGP 3, $\delta_k = 0.2$

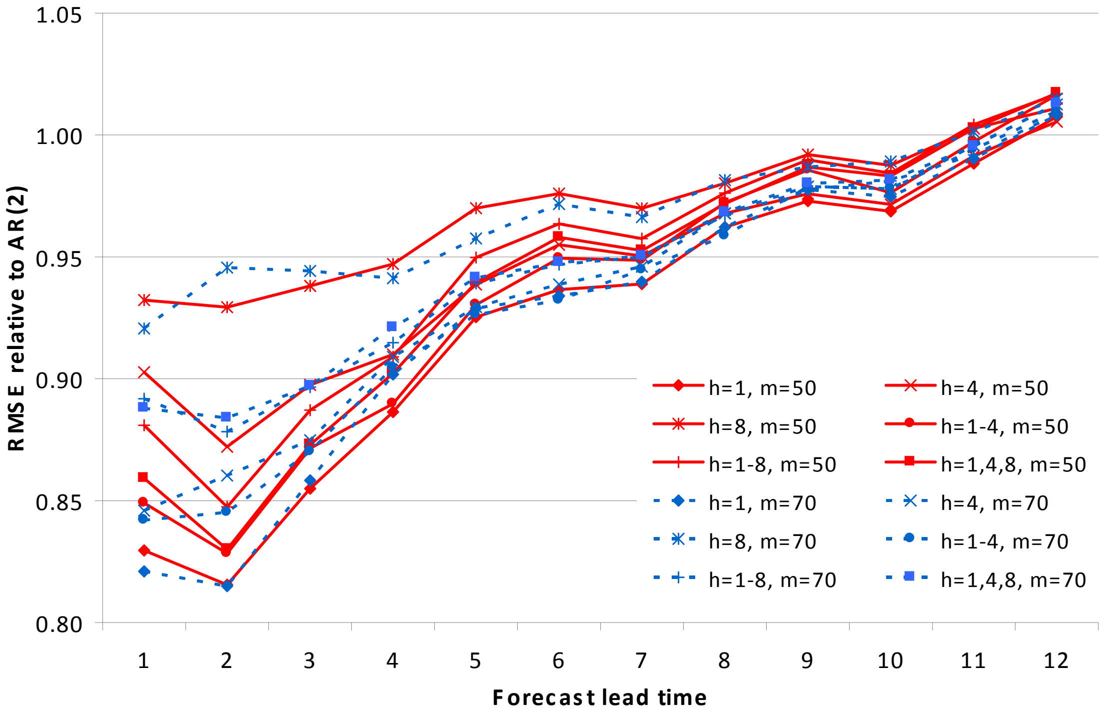


Figure 5 RMSE for forecast combination relative to AR(2), DGP 4, $\delta_k = 0.2$

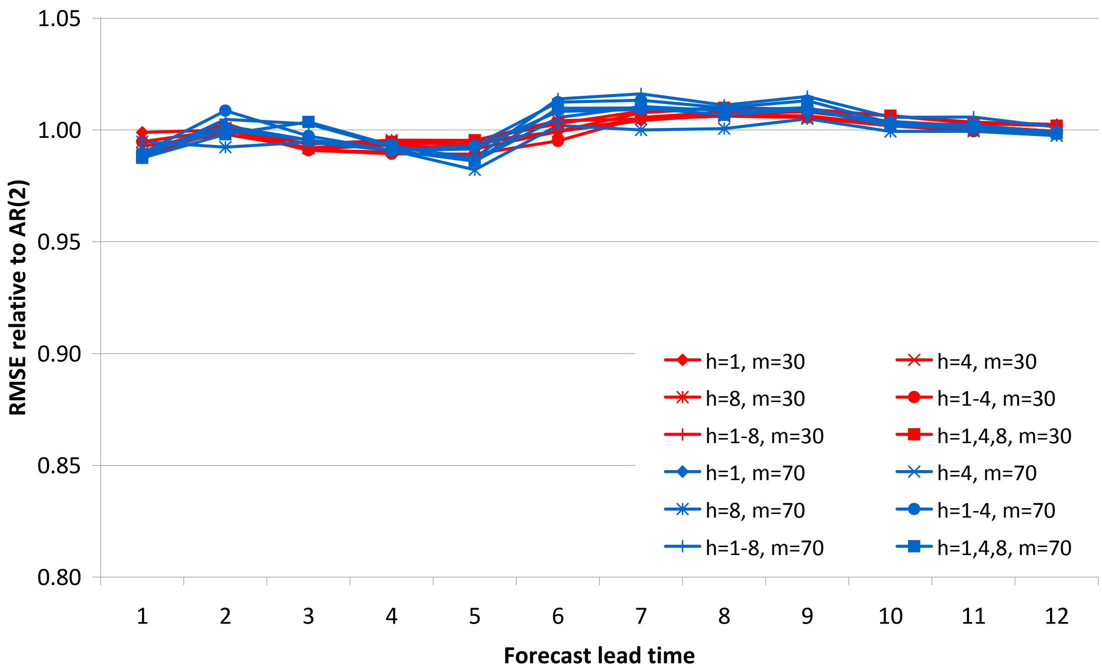
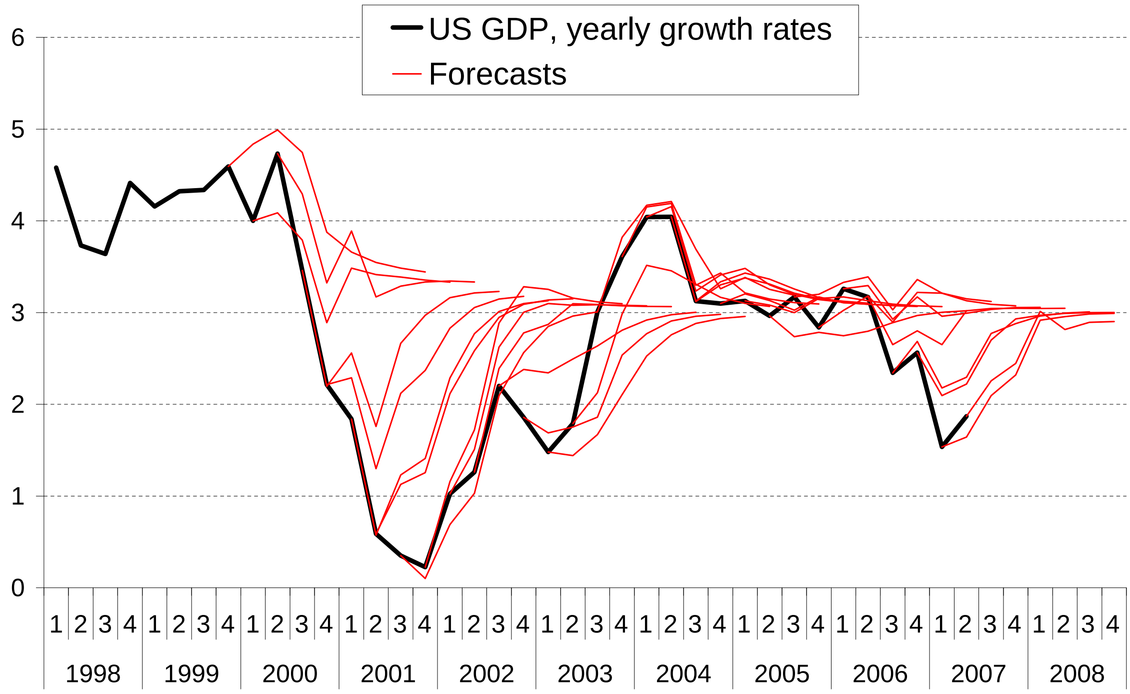


Figure 6 Sequential forecasts from 2000:1 to 2008:3.



Note: The figure presents the median of the predictive distribution.

5 Forecasting US GDP

This section illustrates the predictive likelihood forecast combination procedure at work. The forecast variable is U.S. gross domestic product (GDP). The VAR models are of dimensions one to four and we use a data set of 19 series (GDP included) ranging from second quarter 1971 to the second quarter 2007. The full list variables can be found in Appendix A. This implies estimation of 988 (unique) model combinations. The series are modelled in their first differences or in the levels, but in the presentation the forecasts, as well as the data, are in the fourth log-differences (as an approximation to yearly growth rates).

The prior variable probabilities, δ_k, are all set to 0.2, but we have also tried a value of 0.5 (which is equivalent to a uniform prior over the model space). The final results do not change much when the prior distribution is changed. However, the procedure puts a larger predictive weights on larger systems when the prior 0.5 is used. The predictive likelihood is evaluated at lead time 1 and computed through 5000 Gibbs samples and 50 evaluation points in time. The final forecasts arises as the mean forecast from 1000 Gibbs samples. The prior specification for the parameters is of the same Litterman type as in the Monte Carlo experiment; we set the first (own) lag mean to zero for difference stationary variables and the first lag mean to 0.9 for stationary series. The overall tightness (π_1) is 0.2, the cross-equation tightness (π_2) is 0.5 the lag decay (π_3) is 1 and the tightness on the constant term (π_4) is 5.

In order to compare the general forecasting performance of our procedure, we compute (pseudo out-of-sample) root of the mean squared errors (RMSE) for the combination estimator and compare it to the model with the highest model pre-

Table 3 Forecast accuracy, absolute and relative RMSE, US GDP growth

Lead	For. comb.	Top mod.	AR(2)	R. Walk	Rec. mean	No Fcsts
1	0.41	1.11	1.23	1.61	3.32	30
2	0.58	1.10	1.23	1.73	2.71	29
3	0.93	1.01	1.11	1.44	1.88	28
4	1.23	0.96	1.04	1.33	1.56	27
5	1.36	0.94	0.97	1.31	1.49	26
6	1.41	0.94	0.95	1.36	1.48	25
7	1.33	0.95	0.95	1.47	1.54	24
8	1.22	0.95	0.96	1.64	1.61	23

$Stdev\,(GDP)$ 1.15

RMSE for forecast combination. Ratio of RMSE to RMSE for forecast combination for other procedures.

dictive weight (which may be a different model for different forecast occasions). Furthermore, the performance is also compared to a Bayesian second-order autoregressive model, a random walk forecasts and a recent mean construct (based on the last eight quarters of data). The RMSE's, for horizons 1-8, are calculated for forecasts ranging from first quarter 2000 to second quarter 2007. The reported results concern average performance of the procedure (in terms of RMSE:s) but we also present a current situation analysis (in terms of forecasts and predictive weights).

5.1 Average Forecasting Performance

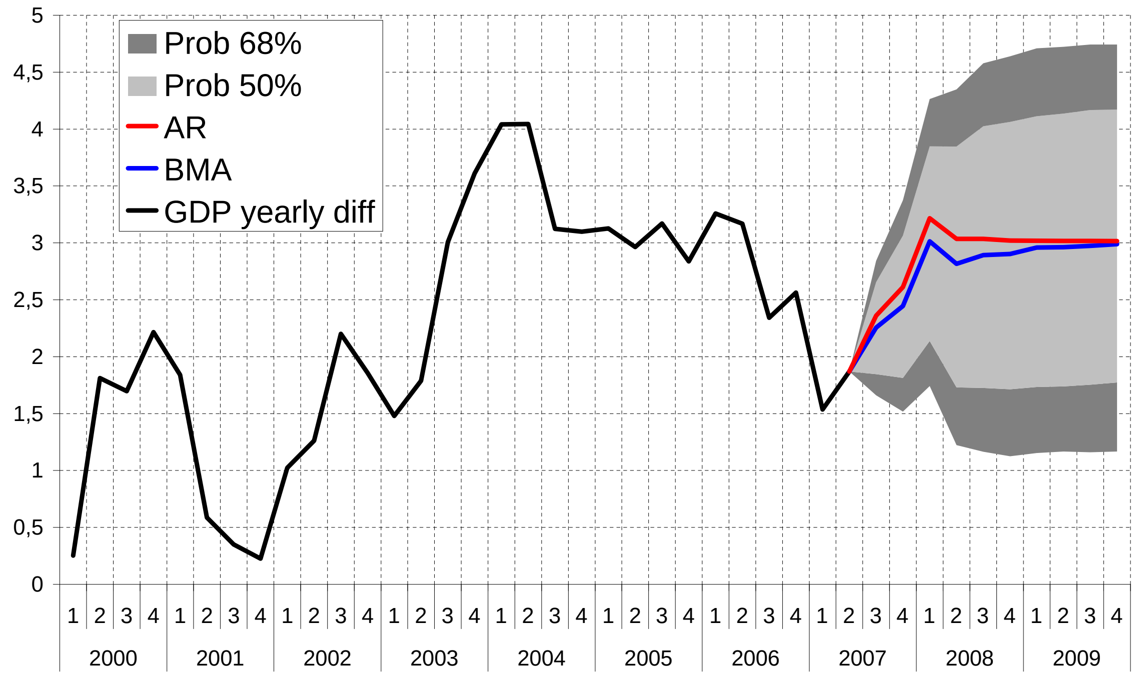
Figure 6 presents a cascade plot of forecasts (one to eight steps ahead) from different points in time. This picture reveals how well the forecasts track the development of GDP growth (if we neglect data revisions which may be sizeable). For example, the first BMA forecast is constructed with data up to the last quarter 1999. From the forecast cascade it is demonstrated that the BMA procedure underestimated the weakness of the economy during 2001, but predicts the period 2002 to 2005 reasonably well. The forecasts did not quite catch the down turn in the recent past and GDP growth is somewhat overpredicted, but not to the same degree as in 2001.

Turning to a more formal evaluation of the forecasts, Table 3 shows that the forecast combination improves on the top model and especially the AR(2) for shorter lead times but does slightly worse than the top model for lead times 4 and higher and worse than the AR(2) for lead times 5 and higher.[2] Due to the small evaluation sample no formal testing is performed. This improvement is somewhat more articulated when we use the uniform prior for the models, $\delta_k = 0.5$. The two simplest alternative forecasts, namely the random walk and the recent mean forecasts, perform notably worse than the other forecasts.

The size of the RMSE of the forecast combination for lead times $h = 4$ and higher is approximately the same as the standard deviation of the GDP series. Our procedure can thus be regarded as a complement to traditional forecasts for short

[2]The ability of autoregressions to compare well with more sophisticated approaches is a familiar phenomenon. See, for example, Stock and Watson (2002*a*) and Stock and Watson (2004).

Figure 7 Forecast from 2007:2. Posterior mean and probability intervals for forecast combination and mean forecast from a Bayesian AR(2).



horizons. This is in line with previous studies, see for instance Galbraith and Tkacz (2006).

5.2 Contemporaneous Forecasts from the Procedure

Figure 7 presents the posterior **median** of the combination forecasts given data up to second quarter 2007. The forecast cover the period 2007:3 to 2009:4. The figure also presents the associated 50 and 68 per cent probability intervals for the forecast combination and the forecasts from a Bayesian autoregression. The intervals demonstrate that there is considerable forecast uncertainty. The combination forecast suggests that the US economy will slowly approach the potential growth rate. The autoregressive forecast only considers the dynamics contained in GDP itself, whereas, the combination procedure also takes the other eighteen variables into account. Figure 7 demonstrate that the information contained in the indicator variables leads to a lower forecast for the whole forecast period compared to not using the indicator information. Thus, the indicators contain a signal of a weaker growth than the GDP series by itself.

Figure 8 presents the predictive weights (i.e. the sum of the predictive weights for the models containing the variable) for each variable, based on the present full data-set. This information may be useful by itself, e.g., this information may be incorporated in judgementally based forecasting schemes. The highest variable inclusion probability is found for the jobless claims (JOBLESS). The other real variables exhibit notably lower predictive weights probabilities, and the nominal variables even lower weights. Some interesting patterns do, however, emerge if we consider

Figure 8 Variable inclusion, predictive weights

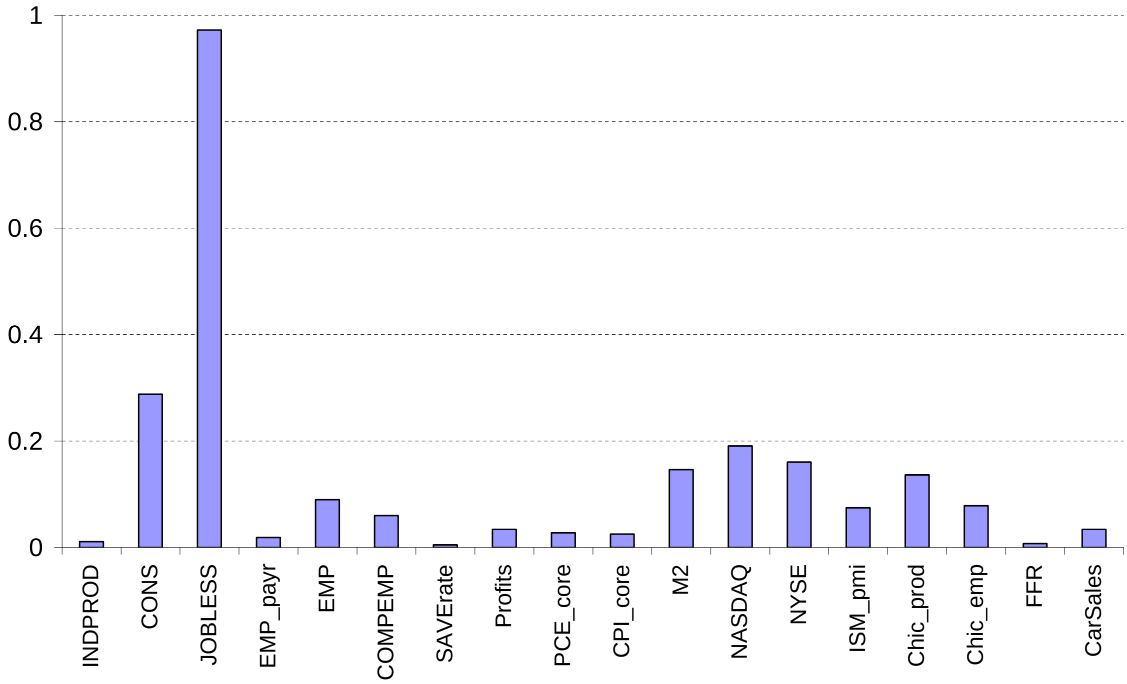


Table 4 Top 10 Models

Rank			**Variables**		**Pred weight**
1	GDP	JOBLESS			0.126
2	GDP	CONS	JOBLESS		0.093
3	GDP	JOBLESS	NASDAQ		0.059
4	GDP	JOBLESS	NYSE		0.044
5	GDP	JOBLESS	M2		0.037
6	GDP	JOBLESS	Chic_prod		0.034
7	GDP	CONS	JOBLESS	NASDAQ	0.033
8	GDP	CONS	JOBLESS	NYSE	0.030
9	GDP	JOBLESS	EMP		0.022
10	GDP	JOBLESS	M2	Chic_prod	0.021
	Equal weights for all specifications/models				0.001

The table presents the top ten models based on data from 1971:2 to 2007:2. The column Pred weight reports the predictive weight of each model.

groups of variables measuring the same underlying feature of the economy. Employment (EMP_payr, EMP, chic_EMP), stock prices (NASDAQ, NYSE), consumption (CONS), money supply (M2) and production (INDPROD, ISM_pmi, chic_PROD) are all important factors.

Table 4 presents posterior analysis for the top ten models, using the current data set. As a point of reference the table also gives the "posterior probability", 1/988, for an equal weighting scheme. Given the variable predictive weights it is not a surprise that the top ranked model consists of GDP and jobless claims.

6 Conclusions

This paper proposes to use weights based on the predictive likelihood for combining forecasts from dynamic multivariate forecasting models such as VAR-models. Our approach overcomes a basic difficulty with standard Bayesian forecast combination based on the marginal with multivariate forecasting models, that the marginal likelihood can change with the dimension of the model in ways that are unrelated to the forecasting performance for the variable of interest. This is achieved by considering the marginal predictive likelihood for the variable of interest rather than the joint predictive likelihood which suffers from the same problem.

The predictive likelihood is not available in closed form for forecasts at lead times greater than 1 and we propose simulation strategies for estimating the predictive likelihood. Our approach is completely general and does not rely on natural conjugate priors or the availability of closed form solutions for the posterior quantities. All that is required is the ability to simulate from the posterior distribution of the parameters and to simulate one step ahead forecasts. The approach is thus also well suited for non-linear forecasting models.

We evaluate the performance of the forecast combination procedure in a small Monte Carlo study and in an application to forecasting US GDP growth. Overall the forecast combinations perform very well. In the Monte Carlo study the forecast combination outperforms our benchmark autoregression by as much as 23% but does slightly worse for forecasts more than four quarters ahead.

References

Andersson, M. K. and Löf, M. (2007), 'The riksbank's new indicator procedures', *Economic Review* (1), 76–95.

Bernanke, B. S. and Boivin, J. (2003), 'Monetary policy in a data-rich environment', *Journal of Monetary Economics* **50**, 525–546.

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

Elliott, G., Granger, C. W. J. and Timmermann, A., eds (2006), *Handbook of Economic Forecasting*, Vol. 1, Elsevier.

Fernández, C., Ley, E. and Steel, M. F. J. (2001), 'Benchmark priors for Bayesian model averaging', *Journal of Econometrics* **100**, 381–427.

Galbraith, J. and Tkacz, G. (2006), How far can we forecast?: Forecast content horizons for some important macroeconomic time series, Working Paper 2006-13, Department of Economics, McGill University.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999), 'Bayesian model averaging: A tutorial (with discussion)', *Statistical Science* **14**, 382–417. Corrected version available at http://www.stat.washington.edu/www/research/online/hoeting1999.pdf.

Jacobson, T. and Karlsson, S. (2004), 'Finding good predictors for inflation: A bayesian model averaging approach', *Journal of Forecasting* **23**, 479–496.

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

Kapetanios, G., Labhard, V. and Price, S. (2007), Forecast combinations and the bank of england's suite of statistical forecasting models, Technical Report 323, Bank of England.

Koop, G. and Potter, S. (2004), 'Forecasting in dynamic factor models using Bayesian model averaging', *Econometrics Journal* **7**(2), 550–565.

Lütkepohl, H. (1993), *Introduction to Multiple Time Series Analysis*, 2nd edn, Springer-Verlag.

Min, C.-K. and Zellner, A. (1993), 'Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting and international growth rates', *Journal of Econometrics* **56**, 89–118.

Stock, J. H. and Watson, M. W. (2002*a*), 'Forecasting using principal components from a large number of predictors', *Journal of the American Statistical Association* **97**(460), 1167 – 1179.

Stock, J. H. and Watson, M. W. (2002*b*), 'Macroeconomic forecasting using diffusion indexes', *Journal of Business & Economic Statistics* **20**, 147–162.

Stock, J. H. and Watson, M. W. (2004), 'Combination forecasts of output growth in a seven-country data set', *Journal of Forecasting* **23**(6), 405 – 430.

Stock, J. H. and Watson, M. W. (2006), Forecasting with many predictors, *in* Elliott, Granger and Timmermann (2006), chapter 10.

Timmermann, A. (2006), Forecast combinations, *in* Elliott et al. (2006), chapter 4.

Appendices

A Data used for the US GDP forecasts

The data set consists of real, nominal and indicator type variables:

- GDP: National Income Account, Overall, Total, Constant Prices, SA (US Dept. of Commerce)
- INDPROD: Production, Overall, Total, SA (Federal Reserve)
- CONS: Personal Outlays, Overall, Total, Constant Prices, SA (US Dept. of Commerce)
- JOBLESS: Jobless claims, SA (US Dept. of Labor)
- EMP_payr: Employment, Overall, Nonfarm Payroll, Total, SA (Bureau of Labor Statistics)
- EMP: Civilian Employment, Business Cycles Indicators, SA (The Conference Board)
- COMPEMP: National Income Account, Compensation of Employees, Total, SA (The US Dept. of Commerce)
- SAVErate: Personal Savings, Rate, SA (Federal Reserve)
- Profits: National Income Account, Corporate Profits, with IVA and CCAdj, Total, SA (The US Dept. of Commerce)
- PCE_core: Price Index, PCE, Overall, Personal Consumption Expenditures less Food and Energy, SA (Bureau of Economic Analysis)
- CPI_core: Consumer Prices, All Items less Food and Energy, SA (Bureau of Labor Statistics)
- M2: Money Supply M2, SA (Federal Board of Governors)
- NASDAQ: Composite Index, Close (NASDAQ)
- NYSE: Composite Index, Close (NYSE)
- ISM_pmi: Business Surveys, ISM Manufacturing, PMI Total, SA (Institute for Supply Management)
- Chic_prod: Business Surveys, Chicago PMI, Production, SA (PMAC)
- Chic_emp: Business Surveys, Chicago PMI, Employment, SA (PMAC)
- FFR: Policy Rates, Fed Funds Effective Rate (Federal Reserve)
- CarSales:

– Car Sales, Domestic, SA (The US Dept. of Commerce)
– Car Sales, Imported, SA (The US Dept. of Commerce)
– Truck Sales, Domestic Light, SA (The US Dept. of Commerce)
– Truck Sales, Imported Light, SA (The US Dept. of Commerce)

B Monte Carlo Experiments

B.1 DGP 1

The DGP is

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

and the irrelevant variables are generated as

$$\begin{aligned} z_{1,t} &= 0.5y_{1,t-1} + 0.5z_{1,t-1} + e_{1,t} \\ z_{2,t} &= 0.5y_{2,t-1} + 0.5z_{2,t-1} + e_{2,t} \\ z_{3,t} &= 0.7z_{3,t-1} + e_{3,t} \\ z_{4,t} &= 0.2z_{4,t-1} + e_{4,t} \\ z_{5,t} &= e_t \end{aligned}$$

with $u_{i,t}$ and $e_{i,t}$ iid $N(0,1)$. $T = 100$ (not accounting for lag lengths) and an additional 12 observations are set aside for forecast evaluation. Model averaging and model selection over the 42 possible models with up to four variables. y_1 is always included in the model. The results are based on 100 Monte Carlo replicates.

Table B1 Predictive weights for variables, DGP 1, models estimated with lag length $p = 2$

Model prior, $\delta_k = 0.2$

h	hold out sample, $m = 30$			hold out sample, $m = 50$			hold out sample, $m = 70$		
	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$
1	0.79	0.17	4.71	0.90	0.17	5.41	0.92	0.15	6.11
2	0.67	0.18	3.68	0.75	0.19	4.06	0.83	0.19	4.28
3	0.51	0.18	2.78	0.60	0.19	3.13	0.65	0.20	3.30
4	0.42	0.19	2.26	0.48	0.20	2.40	0.49	0.20	2.47
8	0.31	0.19	1.57	0.32	0.18	1.77	0.28	0.20	1.40
$1-4$	0.76	0.17	4.38	0.83	0.19	4.27	0.79	0.19	4.10
$1-8$	0.70	0.18	3.81	0.76	0.19	4.00	0.66	0.18	3.76
$1,4,8$	0.76	0.17	4.49	0.83	0.16	5.25	0.76	0.16	4.68

Model prior, $\delta_k = 0.5$

h	hold out sample, $m = 30$			hold out sample, $m = 50$			hold out sample, $m = 70$		
	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$
1	0.88	0.31	2.79	0.96	0.31	3.09	0.96	0.28	3.48
2	0.80	0.33	2.43	0.86	0.32	2.64	0.89	0.32	2.74
3	0.68	0.35	1.96	0.74	0.34	2.17	0.76	0.31	2.45
4	0.60	0.36	1.67	0.66	0.35	1.88	0.63	0.32	1.96
8	0.49	0.37	1.32	0.49	0.35	1.38	0.40	0.32	1.27
$1-4$	0.85	0.30	2.89	0.88	0.29	3.03	0.84	0.26	3.25
$1-8$	0.78	0.28	2.80	0.82	0.26	3.15	0.71	0.22	3.27
$1,4,8$	0.85	0.29	2.88	0.89	0.27	3.33	0.82	0.23	3.64

$p(\cdot)$ denotes the predictive weight for the variable.

Table B2 Predictive weights for variables, DGP 1, models estimated with lag length $p = 2$ and updated posterior distributions for calculation of predictive weights

Model prior, $\delta_k = 0.2$

h	hold out sample, $m = 30$			hold out sample, $m = 70$		
	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$
1	0.78	0.17	4.60	0.97	0.14	6.84
4	0.39	0.18	2.16	0.56	0.19	3.00
$1-4$	0.77	0.17	4.39	0.92	0.15	6.35

Model prior, $\delta_k = 0.5$

h	hold out sample, $m = 30$			hold out sample, $m = 70$		
	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$
1	0.88	0.32	2.76	0.99	0.28	3.54
4	0.58	0.36	1.60	0.72	0.36	2.00
$1-4$	0.86	0.31	2.80	0.96	0.25	3.78

$p(\cdot)$ denotes the predictive weight for the variable.

Table B3 Predictive weights for variables, DGP 1, models estimated with lag length $p = 4$

Model prior, $\delta_k = 0.2$

h	hold out sample, $m = 30$			hold out sample, $m = 50$		
	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$
1	0.77	0.17	4.49	0.88	0.17	5.31
2	0.65	0.19	3.41	0.73	0.19	3.93
3	0.47	0.19	2.44	0.55	0.19	2.86
4	0.38	0.20	1.91	0.41	0.19	2.15
8	0.28	0.19	1.48	0.27	0.18	1.51
$1-4$	0.75	0.19	3.90	0.79	0.20	3.92
$1-8$	0.66	0.19	3.51	0.68	0.19	3.57
$1,4,8$	0.72	0.17	4.23	0.77	0.15	5.00

Model prior, $\delta_k = 0.5$

h	hold out sample, $m = 30$			hold out sample, $m = 50$		
	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$
1	0.88	0.32	2.71	0.94	0.31	3.07
2	0.79	0.35	2.26	0.83	0.35	2.40
3	0.65	0.37	1.76	0.71	0.36	1.97
4	0.57	0.37	1.52	0.59	0.36	1.67
8	0.48	0.37	1.30	0.45	0.34	1.33
$1-4$	0.85	0.31	2.77	0.86	0.29	2.92
$1-8$	0.76	0.28	2.70	0.75	0.26	2.92
$1,4,8$	0.83	0.31	2.72	0.83	0.26	3.21

$p(\cdot)$ denotes the predictive weight for the variable.

Table B4 Model selection, DGP 1, models estimated with lag length $p = 2$. Average posterior probability and proportion selected for true model.

Model prior, $\delta_k = 0.2$

h	hold out, $m = 30$		hold out, $m = 50$		hold out, $m = 70$	
	Weight	Selected	Weight	Selected	Weight	Selected
1	0.31	0.87	0.37	0.78	0.44	0.70
2	0.26	0.69	0.31	0.68	0.38	0.61
3	0.19	0.46	0.24	0.51	0.30	0.45
4	0.16	0.29	0.19	0.34	0.23	0.34
8	0.12	0.19	0.15	0.19	0.12	0.13
$1-4$	0.33	0.61	0.40	0.59	0.42	0.46
$1-8$	0.33	0.50	0.38	0.46	0.31	0.34
$1,4,8$	0.34	0.66	0.42	0.60	0.41	0.45

Model prior, $\delta_k = 0.5$

h	hold out, $m = 30$		hold out, $m = 50$		hold out, $m = 70$	
	Weight	Selected	Weight	Selected	Weight	Selected
1	0.08	0.20	0.10	0.20	0.18	0.39
2	0.07	0.18	0.09	0.25	0.17	0.34
3	0.06	0.17	0.08	0.21	0.15	0.32
4	0.05	0.18	0.08	0.18	0.15	0.26
8	0.05	0.25	0.08	0.19	0.10	0.15
$1-4$	0.13	0.28	0.22	0.32	0.33	0.37
$1-8$	0.19	0.30	0.29	0.35	0.28	0.28
$1,4,8$	0.14	0.31	0.24	0.38	0.34	0.40

Table B5 Model selection, DGP 1, models estimated with lag length $p = 2$ and updated posteriors for calculation of predictive weights. Average posterior probability and proportion selected for true model.

Model prior, $\delta_k = 0.2$

h	hold out, $m = 30$			hold out, $m = 70$	
	Weight	Selected		Weight	Selected
1	0.31	0.86		0.47	0.90
4	0.15	0.26		0.24	0.48
$1-4$	0.33	0.61		0.49	0.72

Model prior, $\delta_k = 0.5$

h	hold out, $m = 30$			hold out, $m = 70$	
	Weight	Selected		Weight	Selected
1	0.07	0.15		0.14	0.38
4	0.05	0.23		0.09	0.24
$1-4$	0.11	0.29		0.26	0.41

Table B6 Model selection, DGP 1, models estimated with lag length $p = 4$. Average posterior probability and proportion selected for true model.

Model prior, $\delta_k = 0.2$

h	hold out, $m = 30$		hold out, $m = 50$	
	Weight	Selected	Weight	Selected
1	0.31	0.82	0.37	0.77
2	0.24	0.60	0.30	0.61
3	0.18	0.39	0.23	0.42
4	0.15	0.22	0.17	0.28
8	0.11	0.13	0.14	0.15
$1-4$	0.30	0.53	0.38	0.50
$1-8$	0.31	0.47	0.34	0.37
$1, 4, 8$	0.32	0.65	0.39	0.57

Model prior, $\delta_k = 0.5$

h	hold out, $m = 30$		hold out, $m = 50$	
	Weight	Selected	Weight	Selected
1	0.08	0.17	0.11	0.26
2	0.07	0.12	0.09	0.22
3	0.06	0.11	0.08	0.18
4	0.05	0.14	0.07	0.21
8	0.05	0.22	0.08	0.24
$1-4$	0.12	0.22	0.23	0.36
$1-8$	0.19	0.28	0.29	0.36
$1, 4, 8$	0.13	0.27	0.24	0.33

Table B7 Forecast performance, RMSE relative to univariate AR(2), hold out sample $m = 30$, models estimated with lag length $p = 2$

Model prior, $\delta_k = 0.2$

	Predictive likelihood evaluated at horizon										
	Posterior distribution not updated for predictive weights								Updated posterior		
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$	1	4	$1-4$
1	0.965	0.970	0.966	0.968	0.969	0.967	0.974	0.975	0.960	0.971	0.966
2	0.876	0.890	0.902	0.922	0.945	0.886	0.881	0.876	0.871	0.925	0.874
3	0.896	0.907	0.929	0.936	0.957	0.893	0.895	0.888	0.894	0.941	0.892
4	0.935	0.941	0.952	0.957	0.964	0.930	0.931	0.924	0.935	0.962	0.930
5	0.978	0.983	0.988	0.991	0.996	0.984	0.985	0.980	0.977	0.989	0.979
6	0.965	0.972	0.976	0.979	0.991	0.971	0.973	0.968	0.964	0.978	0.968
7	0.967	0.970	0.973	0.979	0.989	0.969	0.972	0.970	0.965	0.978	0.966
8	0.986	0.987	0.988	0.989	0.997	0.989	0.987	0.986	0.986	0.987	0.988
9	0.996	0.996	0.996	0.997	0.997	0.998	0.999	0.997	0.996	0.996	0.999
10	0.985	0.986	0.986	0.991	0.996	0.990	0.990	0.988	0.984	0.989	0.987
11	0.997	0.998	0.999	1.003	1.003	1.001	1.002	1.001	0.996	1.001	0.999
12	1.006	1.008	1.008	1.008	1.008	1.008	1.008	1.008	1.006	1.007	1.009

Model prior, $\delta_k = 0.5$

	Predictive likelihood evaluated at horizon										
	Posterior distribution not updated for predictive weights								Updated posterior		
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$	1	4	$1-4$
1	0.961	0.962	0.959	0.957	0.953	0.961	0.969	0.967	0.956	0.959	0.959
2	0.869	0.873	0.882	0.897	0.911	0.873	0.870	0.864	0.861	0.896	0.858
3	0.887	0.893	0.907	0.912	0.931	0.884	0.887	0.879	0.883	0.915	0.880
4	0.926	0.929	0.936	0.940	0.947	0.921	0.925	0.917	0.924	0.945	0.920
5	0.973	0.978	0.982	0.985	0.988	0.976	0.980	0.975	0.970	0.985	0.973
6	0.959	0.965	0.969	0.971	0.981	0.963	0.967	0.962	0.958	0.970	0.961
7	0.963	0.965	0.967	0.971	0.980	0.964	0.967	0.963	0.961	0.970	0.961
8	0.985	0.986	0.986	0.987	0.995	0.987	0.985	0.984	0.985	0.987	0.988
9	0.996	0.997	0.997	0.998	0.998	0.999	0.999	0.997	0.997	0.999	1.001
10	0.985	0.986	0.985	0.991	0.994	0.987	0.987	0.986	0.985	0.990	0.986
11	0.998	0.999	0.999	1.003	1.003	0.999	0.999	0.999	0.998	1.003	0.999
12	1.008	1.009	1.008	1.010	1.010	1.009	1.007	1.007	1.007	1.009	1.009

Table B8 Forecast performance, RMSE relative to univariate AR(2), hold out sample $m = 50$, models estimated with lag length $p = 2$

Model prior, $\delta_k = 0.2$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	0.936	0.954	0.969	0.978	0.982	0.961	0.993	0.987
2	0.840	0.860	0.886	0.912	0.954	0.864	0.881	0.865
3	0.867	0.884	0.908	0.926	0.958	0.888	0.902	0.887
4	0.920	0.926	0.940	0.956	0.973	0.923	0.940	0.935
5	0.965	0.971	0.980	0.992	0.992	0.975	0.984	0.977
6	0.953	0.962	0.971	0.984	0.992	0.973	0.980	0.973
7	0.959	0.964	0.971	0.986	0.993	0.978	0.987	0.979
8	0.983	0.983	0.987	0.995	0.996	0.991	0.992	0.989
9	0.994	0.996	0.997	1.003	1.002	1.002	1.001	1.004
10	0.982	0.983	0.984	0.992	0.996	0.985	0.989	0.990
11	0.996	0.997	0.997	1.003	1.005	1.001	1.001	1.005
12	1.005	1.005	1.003	1.006	1.008	1.004	1.004	1.006

Model prior, $\delta_k = 0.5$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	0.941	0.949	0.957	0.968	0.971	0.960	0.989	0.980
2	0.844	0.856	0.870	0.888	0.923	0.864	0.875	0.863
3	0.865	0.874	0.893	0.908	0.937	0.882	0.895	0.879
4	0.913	0.916	0.929	0.942	0.958	0.913	0.931	0.925
5	0.965	0.968	0.974	0.987	0.985	0.970	0.979	0.975
6	0.952	0.958	0.963	0.977	0.983	0.968	0.974	0.969
7	0.957	0.961	0.966	0.980	0.987	0.974	0.980	0.975
8	0.982	0.982	0.986	0.994	0.992	0.989	0.987	0.987
9	0.993	0.996	0.997	1.003	1.002	1.001	0.998	1.001
10	0.981	0.983	0.985	0.991	0.993	0.984	0.986	0.989
11	0.995	0.997	0.997	1.003	1.005	1.000	0.998	1.003
12	1.005	1.005	1.005	1.007	1.009	1.005	1.003	1.006

Table B9 Forecast performance, RMSE relative to univariate AR(2), hold out sample $m = 70$, models estimated with lag length $p = 2$

Model prior, $\delta_k = 0.2$

	Predictive likelihood evaluated at horizon										
	Posterior distribution not updated for predictive weights								Updated posterior		
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$	1	4	$1-4$
1	0.930	0.966	0.963	0.988	1.000	0.965	0.990	0.972	0.943	0.987	0.955
2	0.836	0.845	0.860	0.904	0.957	0.845	0.856	0.853	0.836	0.913	0.836
3	0.863	0.874	0.900	0.925	0.960	0.880	0.906	0.888	0.866	0.923	0.866
4	0.923	0.923	0.939	0.954	0.959	0.937	0.952	0.945	0.920	0.949	0.919
5	0.961	0.965	0.972	0.988	0.986	0.972	0.988	0.985	0.965	0.982	0.964
6	0.954	0.953	0.965	0.981	0.993	0.961	0.980	0.974	0.953	0.975	0.952
7	0.959	0.956	0.966	0.990	0.984	0.967	0.977	0.970	0.958	0.982	0.961
8	0.983	0.983	0.990	0.997	0.986	0.985	0.985	0.983	0.982	0.989	0.982
9	0.996	0.999	1.004	1.006	0.998	1.003	1.005	1.002	0.994	1.000	0.999
10	0.985	0.986	0.990	0.994	0.993	0.987	0.988	0.987	0.981	0.989	0.984
11	0.997	1.001	1.004	1.005	1.003	1.003	1.003	1.002	0.994	1.002	0.997
12	1.005	1.005	1.008	1.011	1.010	1.005	1.008	1.006	1.003	1.008	1.004

Model prior, $\delta_k = 0.5$

	Predictive likelihood evaluated at horizon										
	Posterior distribution not updated for predictive weights								Updated posterior		
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$	1	4	$1-4$
1	0.935	0.970	0.964	0.980	0.985	0.961	0.986	0.960	0.951	0.971	0.960
2	0.839	0.846	0.852	0.878	0.929	0.836	0.848	0.843	0.841	0.882	0.837
3	0.865	0.871	0.887	0.905	0.937	0.870	0.899	0.880	0.866	0.898	0.862
4	0.920	0.920	0.932	0.942	0.946	0.930	0.947	0.937	0.915	0.937	0.911
5	0.962	0.964	0.971	0.985	0.983	0.967	0.984	0.979	0.962	0.979	0.961
6	0.955	0.951	0.960	0.975	0.986	0.955	0.973	0.968	0.952	0.968	0.950
7	0.960	0.957	0.965	0.985	0.976	0.961	0.973	0.967	0.957	0.977	0.959
8	0.985	0.984	0.989	0.995	0.982	0.983	0.984	0.983	0.982	0.991	0.983
9	0.997	1.002	1.005	1.007	0.997	1.003	1.004	1.003	0.994	1.003	0.999
10	0.985	0.988	0.993	0.996	0.992	0.988	0.987	0.987	0.981	0.991	0.982
11	0.997	1.003	1.006	1.005	1.004	1.003	1.001	1.001	0.995	1.004	0.997
12	1.006	1.006	1.009	1.011	1.011	1.006	1.006	1.005	1.003	1.010	1.005

Table B10 Forecast performance, RMSE relative to univariate AR(2), hold out sample $m = 30$, models estimated with lag length $p = 4$

Model prior, $\delta_k = 0.2$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	0.978	0.980	0.979	0.981	0.970	0.976	0.978	0.986
2	0.886	0.908	0.920	0.939	0.947	0.904	0.912	0.900
3	0.914	0.925	0.941	0.948	0.968	0.916	0.913	0.908
4	0.938	0.945	0.956	0.959	0.970	0.940	0.942	0.930
5	0.979	0.986	0.991	0.993	0.996	0.990	0.985	0.981
6	0.962	0.970	0.974	0.978	0.989	0.971	0.966	0.962
7	0.977	0.979	0.981	0.986	0.992	0.981	0.974	0.973
8	0.995	0.993	0.995	0.995	1.003	0.999	0.993	0.992
9	1.019	1.015	1.014	1.012	1.008	1.017	1.016	1.018
10	0.994	0.996	0.998	1.001	1.008	1.002	1.003	0.999
11	1.009	1.009	1.012	1.016	1.015	1.010	1.011	1.013
12	1.021	1.021	1.021	1.024	1.021	1.019	1.016	1.020

Model prior, $\delta_k = 0.5$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	0.972	0.971	0.975	0.970	0.959	0.974	0.975	0.981
2	0.886	0.896	0.910	0.919	0.920	0.897	0.904	0.894
3	0.907	0.914	0.925	0.926	0.945	0.907	0.903	0.900
4	0.931	0.935	0.942	0.945	0.952	0.932	0.933	0.922
5	0.975	0.979	0.984	0.985	0.985	0.981	0.978	0.974
6	0.959	0.964	0.967	0.969	0.977	0.964	0.961	0.958
7	0.975	0.975	0.975	0.978	0.984	0.976	0.971	0.967
8	0.997	0.997	0.997	0.995	1.003	0.999	0.995	0.993
9	1.022	1.020	1.019	1.017	1.015	1.022	1.021	1.021
10	0.996	0.996	0.996	1.001	1.003	0.998	0.999	0.996
11	1.011	1.010	1.011	1.015	1.015	1.010	1.012	1.012
12	1.023	1.023	1.023	1.025	1.024	1.022	1.019	1.021

Table B11 Forecast performance, RMSE relative to univariate AR(2), hold out sample $m = 50$, models estimated with lag length $p = 4$

Model prior, $\delta_k = 0.2$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	0.950	0.963	0.984	0.985	0.980	0.959	1.001	0.986
2	0.853	0.865	0.900	0.921	0.955	0.876	0.913	0.880
3	0.890	0.903	0.930	0.945	0.974	0.904	0.928	0.906
4	0.924	0.929	0.948	0.962	0.974	0.927	0.957	0.935
5	0.968	0.975	0.984	0.993	0.998	0.977	0.995	0.982
6	0.953	0.962	0.971	0.982	0.991	0.969	0.979	0.964
7	0.973	0.979	0.986	0.996	0.999	0.990	1.002	0.986
8	0.994	0.995	0.999	0.998	1.003	1.005	1.009	0.999
9	1.022	1.023	1.019	1.016	1.010	1.028	1.029	1.026
10	0.996	0.999	0.998	1.007	1.011	1.001	1.005	1.002
11	1.010	1.014	1.014	1.017	1.018	1.016	1.021	1.020
12	1.021	1.023	1.022	1.022	1.019	1.023	1.025	1.025

Model prior, $\delta_k = 0.5$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	0.952	0.960	0.971	0.970	0.970	0.952	0.985	0.980
2	0.861	0.870	0.885	0.900	0.925	0.873	0.904	0.882
3	0.889	0.897	0.913	0.925	0.952	0.895	0.917	0.899
4	0.918	0.923	0.935	0.945	0.959	0.923	0.952	0.931
5	0.967	0.974	0.976	0.985	0.993	0.975	0.992	0.978
6	0.952	0.960	0.962	0.973	0.983	0.964	0.973	0.957
7	0.969	0.975	0.978	0.990	0.995	0.982	0.992	0.979
8	0.995	0.997	0.999	0.999	1.005	1.003	1.005	0.998
9	1.021	1.025	1.022	1.023	1.020	1.026	1.029	1.026
10	0.997	1.000	1.000	1.006	1.009	1.001	1.004	1.002
11	1.010	1.015	1.014	1.018	1.021	1.015	1.021	1.019
12	1.022	1.025	1.025	1.025	1.024	1.023	1.024	1.024

B.2 DGP 2

The DGP is

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

and the irrelevant variables are generated as

$$\begin{aligned} z_{1,t} &= 0.5y_{1,t-1} + 0.5z_{1,t-1} + e_{1,t} \\ z_{2,t} &= 0.5y_{2,t-1} + 0.5z_{2,t-1} + e_{2,t} \\ z_{3,t} &= 0.7z_{3,t-1} + e_{3,t} \\ z_{4,t} &= 0.2z_{4,t-1} + e_{4,t} \\ z_{5,t} &= e_t \end{aligned}$$

with $u_{i,t}$ and $e_{i,t}$ iid $N(0,1)$. $T = 100$ (not accounting for lag lengths) and an additional 12 observations are set aside for forecast evaluation. Model averaging and model selection over the 42 possible models with up to four variables. y_1 is always included in the model. The results are based on 100 Monte Carlo replicates.

Table B12 Predictive weights for variables, DGP 2, models estimated with lag length $p = 2$

Model prior, $\delta_k = 0.2$

h	hold out sample, $m = 30$			hold out sample, $m = 50$		
	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$
1	0.86	0.18	4.67	0.95	0.16	5.80
2	0.83	0.19	4.26	0.93	0.19	4.82
3	0.62	0.21	2.95	0.79	0.23	3.42
4	0.45	0.24	1.87	0.58	0.27	2.13
8	0.35	0.25	1.40	0.37	0.29	1.28
$1-4$	0.89	0.20	4.38	0.94	0.20	4.60
$1-8$	0.84	0.24	3.46	0.86	0.24	3.56
$1,4,8$	0.84	0.23	3.62	0.91	0.25	3.68

Model prior, $\delta_k = 0.5$

h	hold out sample, $m = 30$			hold out sample, $m = 50$		
	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$
1	0.92	0.33	2.78	0.98	0.30	3.31
2	0.90	0.33	2.71	0.96	0.30	3.17
3	0.76	0.35	2.19	0.87	0.33	2.61
4	0.63	0.38	1.67	0.72	0.37	1.96
8	0.52	0.38	1.35	0.52	0.39	1.33
$1-4$	0.93	0.31	2.98	0.96	0.28	3.44
$1-8$	0.88	0.32	2.72	0.89	0.29	3.04
$1,4,8$	0.89	0.34	2.63	0.95	0.32	2.98

$p(\cdot)$ denotes the predictive weight for the variable.

Table B13 Predictive weights for variables, DGP 2, models estimated with lag length $p = 4$

Model prior, $\delta_k = 0.2$

h	hold out sample, $m = 30$			hold out sample, $m = 50$		
	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$
1	0.85	0.18	4.76	0.94	0.17	5.69
2	0.81	0.19	4.19	0.91	0.19	4.69
3	0.53	0.21	2.51	0.67	0.21	3.15
4	0.34	0.23	1.50	0.42	0.23	1.83
8	0.28	0.23	1.20	0.33	0.24	1.38
$1-4$	0.84	0.21	4.04	0.91	0.21	4.42
$1-8$	0.79	0.22	3.56	0.83	0.19	4.26
$1,4,8$	0.77	0.21	3.63	0.88	0.19	4.73

Model prior, $\delta_k = 0.5$

h	hold out sample, $m = 30$			hold out sample, $m = 50$		
	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$	$p(y_2)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$
1	0.91	0.32	2.82	0.97	0.31	3.17
2	0.89	0.33	2.72	0.95	0.31	3.08
3	0.68	0.36	1.92	0.79	0.33	2.37
4	0.54	0.37	1.44	0.60	0.37	1.62
8	0.45	0.37	1.23	0.49	0.37	1.33
$1-4$	0.89	0.31	2.83	0.94	0.28	3.34
$1-8$	0.85	0.30	2.86	0.88	0.24	3.62
$1,4,8$	0.84	0.31	2.68	0.93	0.27	3.43

$p(\cdot)$ denotes the predictive weight for the variable.

Table B14 Model selection, DGP 2, models estimated with lag length $p = 2$. Average posterior probability and proportion selected for true model.

Model prior, $\delta_k = 0.2$

h	hold out, $m = 30$		hold out, $m = 50$	
	Weight	Selected	Weight	Selected
1	0.32	0.85	0.39	0.77
2	0.29	0.73	0.34	0.63
3	0.20	0.51	0.27	0.52
4	0.14	0.24	0.18	0.31
8	0.10	0.14	0.12	0.19
$1-4$	0.33	0.55	0.37	0.49
$1-8$	0.30	0.44	0.35	0.41
$1,4,8$	0.30	0.60	0.35	0.47

Model prior, $\delta_k = 0.5$

h	hold out, $m = 30$		hold out, $m = 50$	
	Weight	Selected	Weight	Selected
1	0.07	0.15	0.11	0.28
2	0.07	0.27	0.10	0.23
3	0.06	0.17	0.09	0.26
4	0.04	0.12	0.07	0.21
8	0.04	0.13	0.07	0.19
$1-4$	0.11	0.27	0.19	0.27
$1-8$	0.15	0.24	0.26	0.31
$1,4,8$	0.10	0.19	0.19	0.29

Table B15 Model selection, DGP 2, models estimated with lag length $p = 4$. Average posterior probability and proportion selected for true model.

Model prior, $\delta_k = 0.2$

h	hold out, $m = 30$		hold out, $m = 50$	
	Weight	Selected	Weight	Selected
1	0.33	0.83	0.39	0.75
2	0.30	0.78	0.35	0.60
3	0.18	0.46	0.24	0.45
4	0.11	0.13	0.14	0.20
8	0.09	0.09	0.11	0.12
$1-4$	0.33	0.56	0.38	0.45
$1-8$	0.31	0.46	0.36	0.39
$1,4,8$	0.30	0.52	0.35	0.45

Model prior, $\delta_k = 0.5$

h	hold out, $m = 30$		hold out, $m = 50$	
	Weight	Selected	Weight	Selected
1	0.08	0.16	0.12	0.27
2	0.07	0.17	0.11	0.26
3	0.05	0.12	0.09	0.23
4	0.04	0.07	0.06	0.18
8	0.04	0.11	0.07	0.12
$1-4$	0.11	0.21	0.22	0.34
$1-8$	0.15	0.20	0.28	0.35
$1,4,8$	0.11	0.20	0.21	0.31

Table B16 Forecast performance RMSE relative to univariate AR(2), DGP 2, hold out sample $m = 30$, models estimated with lag length $p = 2$

Model prior, $\delta_k = 0.2$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	0.828	0.838	0.871	0.904	0.933	0.831	0.838	0.838
2	0.862	0.861	0.887	0.913	0.948	0.860	0.877	0.864
3	0.915	0.918	0.924	0.946	0.979	0.921	0.942	0.940
4	0.978	0.987	0.987	0.988	1.005	0.986	0.995	0.995
5	0.971	0.976	0.979	0.983	0.990	0.978	0.979	0.974
6	0.985	0.989	0.993	0.996	1.001	0.993	0.994	0.993
7	0.983	0.987	0.992	0.996	1.004	0.988	0.989	0.990
8	1.003	1.008	1.009	1.007	1.015	1.006	1.006	1.007
9	1.024	1.031	1.035	1.028	1.030	1.032	1.031	1.034
10	1.018	1.027	1.033	1.031	1.035	1.033	1.032	1.035
11	1.019	1.027	1.034	1.030	1.037	1.032	1.030	1.035
12	1.012	1.021	1.031	1.029	1.035	1.027	1.026	1.030

Model prior, $\delta_k = 0.5$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	0.831	0.839	0.857	0.881	0.902	0.840	0.831	0.838
2	0.849	0.845	0.862	0.882	0.916	0.852	0.868	0.856
3	0.911	0.913	0.918	0.935	0.966	0.915	0.940	0.936
4	0.977	0.983	0.983	0.985	1.001	0.983	0.993	0.994
5	0.970	0.973	0.974	0.978	0.984	0.976	0.978	0.974
6	0.981	0.984	0.986	0.989	0.996	0.987	0.992	0.990
7	0.983	0.984	0.989	0.991	0.998	0.985	0.988	0.988
8	1.003	1.006	1.006	1.005	1.013	1.003	1.004	1.006
9	1.026	1.031	1.034	1.028	1.031	1.031	1.031	1.033
10	1.021	1.027	1.032	1.031	1.035	1.031	1.032	1.034
11	1.023	1.028	1.034	1.029	1.037	1.030	1.030	1.035
12	1.016	1.022	1.029	1.027	1.032	1.025	1.025	1.029

Table B17 Forecast performance RMSE relative to univariate AR(2), DGP 2, hold out sample $m = 50$, models estimated with lag length $p = 2$

Model prior, $\delta_k = 0.2$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	0.827	0.846	0.872	0.897	0.937	0.832	0.834	0.831
2	0.850	0.850	0.882	0.916	0.946	0.850	0.869	0.855
3	0.911	0.909	0.913	0.933	0.952	0.913	0.912	0.912
4	0.974	0.978	0.979	0.984	0.999	0.986	0.987	0.987
5	0.968	0.971	0.973	0.973	0.992	0.972	0.978	0.983
6	0.981	0.984	0.988	0.988	1.004	0.987	0.990	0.997
7	0.978	0.980	0.985	0.985	1.003	0.986	0.989	0.993
8	0.997	1.000	1.002	1.003	1.012	1.006	1.011	1.012
9	1.018	1.024	1.030	1.025	1.034	1.026	1.032	1.039
10	1.015	1.021	1.027	1.026	1.038	1.025	1.031	1.041
11	1.014	1.020	1.027	1.025	1.045	1.025	1.033	1.043
12	1.008	1.016	1.025	1.026	1.048	1.019	1.029	1.040

Model prior, $\delta_k = 0.5$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	0.832	0.847	0.866	0.885	0.914	0.833	0.829	0.831
2	0.847	0.850	0.866	0.891	0.919	0.850	0.858	0.847
3	0.911	0.912	0.914	0.926	0.940	0.916	0.914	0.914
4	0.975	0.979	0.982	0.983	0.994	0.986	0.989	0.990
5	0.969	0.973	0.975	0.973	0.989	0.975	0.980	0.984
6	0.981	0.984	0.987	0.986	0.999	0.988	0.991	0.998
7	0.979	0.982	0.985	0.983	0.999	0.987	0.991	0.995
8	0.998	1.001	1.004	1.002	1.010	1.006	1.013	1.014
9	1.021	1.028	1.032	1.027	1.034	1.028	1.035	1.040
10	1.018	1.025	1.029	1.027	1.039	1.026	1.034	1.042
11	1.018	1.025	1.029	1.026	1.045	1.027	1.036	1.045
12	1.012	1.020	1.025	1.024	1.046	1.021	1.031	1.040

Table B18 Forecast performance RMSE relative to univariate AR(2), DGP 2, hold out sample $m = 30$, models estimated with lag length $p = 4$

Model prior, $\delta_k = 0.2$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	0.818	0.836	0.919	0.945	0.935	0.846	0.857	0.872
2	0.885	0.888	0.943	0.964	0.988	0.902	0.917	0.921
3	0.935	0.937	0.962	0.975	1.004	0.954	0.973	0.988
4	0.997	1.005	1.011	1.009	1.025	1.010	1.019	1.030
5	0.993	0.999	1.005	1.005	1.011	1.006	1.009	1.014
6	1.002	1.006	1.014	1.016	1.026	1.015	1.024	1.030
7	1.004	1.006	1.016	1.017	1.028	1.015	1.024	1.028
8	1.019	1.024	1.028	1.025	1.036	1.028	1.035	1.040
9	1.044	1.050	1.055	1.047	1.053	1.056	1.060	1.065
10	1.044	1.049	1.053	1.049	1.059	1.060	1.064	1.070
11	1.051	1.053	1.058	1.053	1.065	1.066	1.074	1.078
12	1.045	1.048	1.054	1.053	1.063	1.059	1.067	1.071

Model prior, $\delta_k = 0.5$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	0.816	0.825	0.891	0.920	0.905	0.837	0.844	0.865
2	0.872	0.864	0.909	0.924	0.948	0.887	0.899	0.897
3	0.933	0.934	0.961	0.966	0.997	0.951	0.961	0.982
4	1.001	1.005	1.012	1.007	1.025	1.014	1.016	1.030
5	0.995	0.999	1.004	1.002	1.010	1.009	1.007	1.014
6	1.003	1.005	1.011	1.011	1.022	1.016	1.021	1.026
7	1.006	1.007	1.014	1.012	1.025	1.016	1.020	1.025
8	1.023	1.025	1.028	1.024	1.037	1.028	1.034	1.039
9	1.054	1.056	1.058	1.051	1.059	1.062	1.064	1.067
10	1.053	1.055	1.056	1.051	1.063	1.064	1.066	1.071
11	1.062	1.062	1.062	1.057	1.070	1.071	1.077	1.079
12	1.054	1.055	1.056	1.053	1.065	1.064	1.068	1.071

Table B19 Forecast performance RMSE relative to univariate AR(2), DGP 2, hold out sample $m = 50$, models estimated with lag length $p = 4$

Model prior, $\delta_k = 0.2$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	0.807	0.823	0.901	0.927	0.933	0.825	0.835	0.822
2	0.875	0.874	0.924	0.968	0.978	0.884	0.893	0.874
3	0.942	0.949	0.963	0.974	0.987	0.966	0.958	0.954
4	1.000	1.010	1.011	1.010	1.020	1.017	1.005	1.013
5	0.990	0.998	1.006	0.999	1.007	1.001	0.989	0.997
6	1.004	1.009	1.016	1.014	1.013	1.015	1.000	1.009
7	1.006	1.008	1.016	1.013	1.015	1.020	1.001	1.010
8	1.022	1.025	1.023	1.022	1.024	1.030	1.025	1.027
9	1.044	1.049	1.050	1.046	1.039	1.053	1.040	1.046
10	1.041	1.044	1.046	1.043	1.039	1.049	1.037	1.044
11	1.047	1.047	1.049	1.048	1.040	1.054	1.038	1.046
12	1.041	1.043	1.048	1.050	1.043	1.050	1.032	1.041

Model prior, $\delta_k = 0.5$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	0.815	0.818	0.877	0.907	0.908	0.818	0.824	0.812
2	0.868	0.866	0.893	0.927	0.937	0.877	0.883	0.868
3	0.947	0.952	0.961	0.964	0.977	0.966	0.956	0.949
4	1.003	1.010	1.015	1.011	1.017	1.019	1.005	1.013
5	0.993	1.000	1.007	1.001	1.006	1.002	0.988	0.997
6	1.005	1.010	1.016	1.012	1.009	1.014	0.998	1.007
7	1.006	1.009	1.016	1.011	1.012	1.019	1.000	1.008
8	1.023	1.026	1.029	1.023	1.025	1.031	1.023	1.027
9	1.050	1.054	1.057	1.050	1.042	1.056	1.040	1.049
10	1.046	1.049	1.053	1.046	1.041	1.052	1.037	1.045
11	1.052	1.053	1.057	1.052	1.044	1.056	1.038	1.048
12	1.046	1.049	1.054	1.050	1.042	1.052	1.031	1.045

B.3 DGP 3

The DGP is

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

and the irrelevant variables are generated as

$$\begin{aligned} z_{1,t} &= 0.5 y_{1,t-1} + 0.5 z_{1,t-1} + e_{1,t} \\ z_{2,t} &= 0.5 y_{2,t-1} + 0.5 z_{2,t-1} + e_{2,t} \\ z_{3,t} &= 0.7 z_{3,t-1} + e_{3,t} \\ z_{4,t} &= 0.2 z_{4,t-1} + e_{4,t} \end{aligned}$$

with $u_{i,t}$ and $e_{i,t}$ iid $N(0,1)$. $T = 100$ (not accounting for lag lengths) and an additional 12 observations are set aside for forecast evaluation. Model averaging and model selection over the 57 possible models with up to five variables. y_1 is always included in the model. The results are based on 100 Monte Carlo replicates.

Table B20 Predictive weights for variables, DGP 3, models estimated with lag length $p = 2$

Model prior, $\delta_k = 0.2$

h	hold out sample, $m = 50$					hold out sample, $m = 70$				
	$p(y_2)$	$p(y_3)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$	$\frac{p(y_3)}{\max[p(z_i)]}$	$p(y_2)$	$p(y_3)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$	$\frac{p(y_3)}{\max[p(z_i)]}$
1	0.89	0.88	0.18	4.81	4.75	0.91	0.88	0.20	4.60	4.47
2	0.80	0.74	0.20	4.06	3.80	0.83	0.75	0.22	3.73	3.35
3	0.64	0.59	0.21	3.02	2.78	0.73	0.65	0.25	2.92	2.58
4	0.56	0.49	0.23	2.41	2.10	0.65	0.50	0.28	2.31	1.78
8	0.39	0.30	0.26	1.52	1.14	0.47	0.34	0.30	1.57	1.15
$1-4$	0.82	0.84	0.21	3.98	4.09	0.80	0.72	0.24	3.30	2.94
$1-8$	0.76	0.74	0.21	3.62	3.49	0.63	0.57	0.23	2.69	2.41
$1,4,8$	0.81	0.77	0.21	3.86	3.69	0.71	0.63	0.21	3.42	3.05

Model prior, $\delta_k = 0.5$

h	hold out sample, $m = 50$					hold out sample, $m = 70$				
	$p(y_2)$	$p(y_3)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$	$\frac{p(y_3)}{\max[p(z_i)]}$	$p(y_2)$	$p(y_3)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$	$\frac{p(y_3)}{\max[p(z_i)]}$
1	0.94	0.94	0.36	2.61	2.61	0.94	0.93	0.34	2.76	2.72
2	0.88	0.86	0.38	2.28	2.23	0.89	0.84	0.38	2.34	2.20
3	0.78	0.76	0.41	1.87	1.83	0.82	0.77	0.38	2.15	2.03
4	0.71	0.68	0.43	1.66	1.59	0.74	0.66	0.41	1.81	1.59
8	0.55	0.51	0.45	1.22	1.13	0.58	0.47	0.42	1.37	1.10
$1-4$	0.87	0.91	0.33	2.62	2.75	0.84	0.78	0.30	2.81	2.62
$1-8$	0.81	0.83	0.28	2.86	2.92	0.68	0.62	0.27	2.49	2.27
$1,4,8$	0.85	0.86	0.30	2.80	2.84	0.76	0.70	0.27	2.87	2.61

$p(\cdot)$ denotes the predictive weight for the variable.

Table B21 Predictive weights for variables, DGP 3, models estimated with lag length $p = 4$

Model prior, $\delta_k = 0.2$

h	hold out sample, $m = 50$				
	$p(y_2)$	$p(y_3)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$	$\frac{p(y_3)}{\max[p(z_i)]}$
1	0.89	0.86	0.17	5.16	5.03
2	0.77	0.71	0.19	3.98	3.65
3	0.63	0.56	0.20	3.19	2.82
4	0.53	0.44	0.22	2.44	2.03
8	0.32	0.29	0.25	1.25	1.16
$1-4$	0.79	0.78	0.20	3.88	3.82
$1-8$	0.62	0.64	0.19	3.35	3.45
$1,4,8$	0.72	0.75	0.21	3.49	3.63

Model prior, $\delta_k = 0.5$

h	hold out sample, $m = 50$				
	$p(y_2)$	$p(y_3)$	$\max[p(z_i)]$	$\frac{p(y_2)}{\max[p(z_i)]}$	$\frac{p(y_3)}{\max[p(z_i)]}$
1	0.94	0.94	0.34	2.76	2.74
2	0.86	0.84	0.38	2.26	2.21
3	0.77	0.75	0.39	1.99	1.93
4	0.70	0.66	0.40	1.74	1.64
8	0.49	0.51	0.41	1.18	1.22
$1-4$	0.86	0.86	0.31	2.79	2.81
$1-8$	0.69	0.76	0.25	2.79	3.08
$1,4,8$	0.78	0.86	0.29	2.71	3.01

$p(\cdot)$ denotes the predictive weight for the variable.

Table B22 Model selection, DGP 3, models estimated with lag length $p = 2$. Average posterior probability and proportion selected for true model.

Model prior, $\delta_k = 0.2$

h	hold out, $m = 50$		hold out, $m = 70$	
	Weight	Selected	Weight	Selected
1	0.39	0.73	0.41	0.62
2	0.28	0.56	0.30	0.43
3	0.17	0.31	0.20	0.24
4	0.10	0.14	0.12	0.13
8	0.04	0.04	0.06	0.08
$1-4$	0.37	0.49	0.35	0.40
$1-8$	0.30	0.31	0.20	0.21
$1,4,8$	0.33	0.49	0.24	0.24

Model prior, $\delta_k = 0.5$

h	hold out, $m = 50$		hold out, $m = 70$	
	Weight	Selected	Weight	Selected
1	0.14	0.32	0.22	0.36
2	0.13	0.27	0.20	0.25
3	0.10	0.22	0.17	0.28
4	0.08	0.20	0.12	0.22
8	0.05	0.10	0.07	0.06
$1-4$	0.25	0.34	0.33	0.37
$1-8$	0.27	0.30	0.21	0.19
$1,4,8$	0.24	0.32	0.24	0.27

Table B23 Model selection, DGP 3, models estimated with lag length $p = 4$. Average posterior probability and proportion selected for true model.

Hold out sample, $m = 50$

h	Model prior, $\delta_k = 0.2$		Model prior, $\delta_k = 0.5$	
	Weight	Selected	Weight	Selected
1	0.40	0.72	0.17	0.41
2	0.28	0.52	0.14	0.29
3	0.15	0.28	0.10	0.18
4	0.10	0.11	0.08	0.10
8	0.03	0.02	0.04	0.10
$1-4$	0.33	0.42	0.25	0.31
$1-8$	0.23	0.26	0.24	0.27
$1,4,8$	0.27	0.38	0.23	0.27

Table B24 Forecast performance RMSE relative to univariate AR(2), DGP 3, hold out sample $m = 50$, models estimated with lag length $p = 2$

Model prior, $\delta_k = 0.2$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	0.830	0.864	0.879	0.903	0.932	0.849	0.881	0.859
2	0.815	0.824	0.851	0.872	0.929	0.828	0.847	0.830
3	0.855	0.868	0.879	0.897	0.938	0.872	0.887	0.873
4	0.886	0.889	0.902	0.910	0.947	0.890	0.909	0.902
5	0.925	0.929	0.932	0.939	0.970	0.930	0.950	0.940
6	0.937	0.940	0.945	0.955	0.976	0.949	0.964	0.958
7	0.939	0.941	0.945	0.950	0.970	0.949	0.957	0.953
8	0.962	0.962	0.966	0.968	0.980	0.972	0.976	0.972
9	0.973	0.971	0.975	0.976	0.992	0.986	0.990	0.987
10	0.969	0.967	0.974	0.972	0.988	0.976	0.984	0.983
11	0.988	0.986	0.992	0.991	1.003	0.997	1.004	1.003
12	1.007	1.004	1.007	1.005	1.011	1.016	1.017	1.017

Model prior, $\delta_k = 0.5$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	0.819	0.844	0.847	0.854	0.882	0.833	0.855	0.839
2	0.814	0.820	0.832	0.840	0.882	0.821	0.831	0.824
3	0.853	0.865	0.868	0.880	0.914	0.865	0.880	0.867
4	0.883	0.885	0.893	0.898	0.929	0.886	0.906	0.898
5	0.924	0.928	0.927	0.931	0.958	0.929	0.946	0.935
6	0.936	0.940	0.940	0.949	0.967	0.946	0.961	0.954
7	0.938	0.941	0.941	0.946	0.960	0.944	0.953	0.948
8	0.966	0.967	0.968	0.969	0.974	0.973	0.977	0.970
9	0.978	0.978	0.978	0.980	0.991	0.988	0.993	0.986
10	0.971	0.970	0.974	0.974	0.984	0.975	0.984	0.979
11	0.992	0.989	0.994	0.995	1.003	0.996	1.003	1.000
12	1.013	1.011	1.013	1.013	1.015	1.016	1.018	1.016

Table B25 Forecast performance RMSE relative to univariate AR(2), DGP 3, hold out sample $m = 70$, models estimated with lag length $p = 2$

Model prior, $\delta_k = 0.2$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	0.821	0.826	0.849	0.846	0.921	0.842	0.892	0.888
2	0.815	0.830	0.871	0.860	0.946	0.845	0.878	0.884
3	0.858	0.863	0.876	0.875	0.944	0.870	0.897	0.897
4	0.902	0.898	0.910	0.909	0.941	0.905	0.915	0.921
5	0.929	0.916	0.928	0.929	0.958	0.926	0.939	0.941
6	0.934	0.927	0.934	0.939	0.972	0.932	0.947	0.948
7	0.940	0.931	0.940	0.946	0.966	0.945	0.950	0.950
8	0.962	0.955	0.962	0.968	0.982	0.959	0.967	0.968
9	0.978	0.972	0.975	0.979	0.987	0.979	0.978	0.980
10	0.975	0.968	0.970	0.978	0.989	0.978	0.981	0.982
11	0.990	0.987	0.991	0.996	1.001	0.994	0.992	0.995
12	1.008	1.006	1.007	1.013	1.015	1.010	1.009	1.013

Model prior, $\delta_k = 0.5$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	0.815	0.814	0.816	0.817	0.877	0.821	0.896	0.876
2	0.819	0.821	0.850	0.843	0.913	0.841	0.882	0.875
3	0.861	0.865	0.865	0.863	0.925	0.868	0.899	0.892
4	0.899	0.896	0.903	0.902	0.926	0.902	0.920	0.917
5	0.927	0.919	0.922	0.922	0.950	0.925	0.944	0.941
6	0.931	0.927	0.925	0.931	0.965	0.931	0.949	0.945
7	0.939	0.937	0.938	0.944	0.959	0.942	0.952	0.949
8	0.960	0.956	0.961	0.967	0.975	0.957	0.966	0.964
9	0.977	0.974	0.978	0.983	0.983	0.979	0.976	0.976
10	0.974	0.972	0.974	0.981	0.984	0.977	0.979	0.980
11	0.989	0.989	0.993	0.997	0.999	0.993	0.991	0.993
12	1.009	1.010	1.010	1.015	1.016	1.011	1.009	1.012

Table B26 Forecast performance RMSE relative to univariate AR(2), DGP 3, hold out sample $m = 50$, models estimated with lag length $p = 4$

Model prior, $\delta_k = 0.2$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	0.840	0.875	0.911	0.929	0.931	0.864	0.928	0.889
2	0.831	0.836	0.863	0.895	0.929	0.852	0.907	0.870
3	0.871	0.885	0.893	0.917	0.945	0.883	0.917	0.893
4	0.896	0.903	0.911	0.924	0.960	0.893	0.924	0.914
5	0.930	0.936	0.940	0.951	0.988	0.937	0.968	0.952
6	0.949	0.952	0.957	0.969	1.006	0.957	0.989	0.974
7	0.951	0.953	0.959	0.969	1.001	0.960	0.989	0.980
8	0.973	0.972	0.980	0.989	1.016	0.988	1.007	1.002
9	0.995	0.989	0.992	0.998	1.019	1.008	1.020	1.021
10	0.977	0.975	0.983	0.984	1.019	0.986	1.009	1.016
11	1.002	0.998	1.003	1.002	1.025	1.011	1.028	1.030
12	1.023	1.020	1.022	1.019	1.036	1.032	1.047	1.046

Model prior, $\delta_k = 0.5$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	0.826	0.846	0.873	0.883	0.891	0.847	0.910	0.868
2	0.837	0.836	0.851	0.864	0.897	0.845	0.888	0.856
3	0.873	0.880	0.881	0.899	0.923	0.877	0.906	0.882
4	0.900	0.902	0.906	0.918	0.947	0.893	0.920	0.905
5	0.934	0.938	0.937	0.944	0.974	0.936	0.960	0.941
6	0.950	0.953	0.949	0.962	0.997	0.953	0.983	0.964
7	0.952	0.953	0.953	0.963	0.993	0.956	0.979	0.968
8	0.977	0.977	0.979	0.987	1.012	0.985	1.003	0.994
9	1.002	0.999	0.997	1.005	1.023	1.007	1.020	1.018
10	0.980	0.979	0.983	0.985	1.017	0.984	1.004	1.008
11	1.007	1.003	1.005	1.005	1.025	1.009	1.024	1.026
12	1.029	1.027	1.028	1.027	1.040	1.031	1.044	1.044

B.4 DGP 4

The DGP is

$$y_t = 0.5y_{t-1} + 0.3y_{t-2} + u_t.$$

and the irrelevant variables are generated as

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

with $u_{i,t}$ and $e_{i,t}$ iid $N(0,1)$. $T = 100$ (not accounting for lag lengths) and an additional 12 observations are set aside for forecast evaluation. Model averaging and model selection over the 42 possible models with up to four variables. y_1 is always included in the model. The results are based on 100 Monte Carlo replicates.

Table B27 Predictive weights for irrelevant variables, DGP 4, models estimated with lag length $p = 2$

$\max[p(z_i)]$

	Model prior, $\delta_k = 0.2$			Model prior, $\delta_k = 0.5$		
h	$m = 30$	$m = 50$	$m = 70$	$m = 30$	$m = 50$	$m = 70$
1	0.19	0.21	0.23	0.37	0.38	0.36
2	0.19	0.23	0.24	0.39	0.37	0.35
3	0.20	0.25	0.23	0.38	0.38	0.37
4	0.20	0.24	0.25	0.39	0.38	0.38
8	0.22	0.24	0.27	0.38	0.37	0.34
$1-4$	0.22	0.24	0.23	0.35	0.33	0.29
$1-8$	0.22	0.21	0.24	0.31	0.27	0.27
$1,4,8$	0.20	0.21	0.23	0.35	0.32	0.30

$p(\cdot)$ denotes the predictive weight for the variable.

Table B28 Model selection, DGP 4, models estimated with lag length $p = 2$. Average posterior probability and proportion selected for true model.

Model prior, $\delta_k = 0.2$

h	hold out, $m = 30$		hold out, $m = 50$		hold out, $m = 70$	
	Weight	Selected	Weight	Selected	Weight	Selected
1	0.30	0.89	0.31	0.75	0.31	0.56
2	0.29	0.87	0.29	0.67	0.32	0.50
3	0.28	0.82	0.27	0.64	0.31	0.48
4	0.28	0.83	0.27	0.65	0.30	0.46
8	0.27	0.82	0.27	0.63	0.28	0.47
$1-4$	0.34	0.64	0.34	0.50	0.36	0.43
$1-8$	0.38	0.54	0.37	0.42	0.37	0.39
$1,4,8$	0.34	0.69	0.36	0.50	0.36	0.40

Model prior, $\delta_k = 0.5$

h	hold out, $m = 30$		hold out, $m = 50$		hold out, $m = 70$	
	Weight	Selected	Weight	Selected	Weight	Selected
1	0.04	0.16	0.05	0.18	0.09	0.19
2	0.04	0.18	0.06	0.25	0.11	0.24
3	0.04	0.18	0.05	0.21	0.12	0.30
4	0.04	0.19	0.05	0.23	0.12	0.25
8	0.04	0.24	0.06	0.19	0.12	0.24
$1-4$	0.09	0.30	0.16	0.26	0.25	0.29
$1-8$	0.17	0.38	0.26	0.28	0.31	0.32
$1,4,8$	0.08	0.28	0.17	0.27	0.25	0.31

Table B29 Forecast performance RMSE relative to univariate AR(2), DGP 4, hold out sample $m = 30$, models estimated with lag length $p = 2$

Model prior, $\delta_k = 0.2$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	0.999	0.992	0.994	0.993	0.990	0.995	0.990	0.989
2	1.000	1.000	0.998	0.998	1.001	1.000	1.003	1.000
3	0.991	0.991	0.993	0.992	0.994	0.991	0.991	0.992
4	0.994	0.992	0.992	0.993	0.995	0.989	0.990	0.995
5	0.994	0.994	0.991	0.991	0.995	0.989	0.995	0.995
6	0.999	0.999	0.998	0.999	1.004	0.995	0.999	1.003
7	1.004	1.006	1.004	1.005	1.005	1.006	1.008	1.008
8	1.007	1.009	1.007	1.006	1.007	1.008	1.010	1.010
9	1.006	1.007	1.006	1.006	1.008	1.006	1.009	1.009
10	1.003	1.003	1.003	1.002	1.004	1.002	1.006	1.006
11	1.002	1.002	1.001	1.000	1.002	0.999	1.003	1.003
12	0.999	1.000	0.999	0.999	0.999	0.999	1.003	1.002

Model prior, $\delta_k = 0.5$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	1.000	0.991	0.994	0.993	0.990	0.996	0.986	0.989
2	0.998	0.997	0.994	0.994	0.998	0.995	0.998	0.995
3	0.986	0.986	0.988	0.987	0.990	0.985	0.986	0.986
4	0.990	0.989	0.988	0.989	0.992	0.986	0.989	0.992
5	0.989	0.990	0.985	0.986	0.991	0.986	0.993	0.991
6	0.997	0.997	0.994	0.996	1.003	0.992	0.997	1.001
7	1.004	1.004	1.004	1.004	1.003	1.004	1.006	1.006
8	1.008	1.009	1.008	1.007	1.007	1.009	1.011	1.009
9	1.007	1.007	1.006	1.006	1.009	1.006	1.010	1.010
10	1.003	1.003	1.002	1.002	1.004	1.002	1.008	1.006
11	1.001	1.001	1.000	0.999	1.001	0.999	1.003	1.002
12	0.996	0.996	0.996	0.996	0.997	0.996	1.001	0.999

Table B30 Forecast performance RMSE relative to univariate AR(2), DGP 4, hold out sample $m = 50$, models estimated with lag length $p = 2$

Model prior, $\delta_k = 0.2$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	0.998	0.999	1.002	0.999	0.990	0.999	0.988	0.995
2	1.001	1.006	1.004	1.003	1.003	1.004	1.003	1.005
3	0.997	1.001	1.002	0.999	0.994	1.001	0.995	0.998
4	0.998	1.001	1.000	0.998	0.997	0.999	0.995	0.998
5	0.995	0.996	0.996	0.993	0.995	0.997	0.995	0.995
6	1.004	1.009	1.009	1.006	1.007	1.008	1.006	1.009
7	1.007	1.013	1.015	1.011	1.014	1.014	1.014	1.016
8	1.008	1.014	1.014	1.011	1.014	1.013	1.015	1.015
9	1.006	1.009	1.010	1.007	1.012	1.009	1.011	1.012
10	1.003	1.004	1.005	1.003	1.008	1.004	1.006	1.007
11	0.999	1.001	1.001	0.999	1.003	1.000	1.001	1.001
12	0.998	0.999	0.999	0.998	1.001	0.998	1.000	0.999

Model prior, $\delta_k = 0.5$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	1.000	1.000	1.005	1.000	0.990	1.001	0.990	0.996
2	0.998	1.004	1.002	0.999	1.000	1.003	1.001	1.002
3	0.992	0.995	0.995	0.993	0.989	0.998	0.992	0.996
4	0.995	0.997	0.995	0.995	0.995	0.996	0.993	0.997
5	0.991	0.993	0.993	0.990	0.992	0.995	0.992	0.995
6	1.001	1.005	1.006	1.003	1.006	1.005	1.005	1.008
7	1.006	1.012	1.013	1.011	1.013	1.012	1.013	1.015
8	1.009	1.014	1.014	1.012	1.015	1.013	1.014	1.015
9	1.006	1.009	1.010	1.008	1.012	1.009	1.011	1.011
10	1.002	1.005	1.005	1.004	1.008	1.004	1.006	1.006
11	0.998	1.000	1.000	0.999	1.002	0.999	1.001	1.000
12	0.995	0.996	0.997	0.996	1.000	0.996	0.998	0.997

Table B31 Forecast performance RMSE relative to univariate AR(2), DGP 4, hold out sample $m = 70$, models estimated with lag length $p = 2$

Model prior, $\delta_k = 0.2$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	0.990	0.992	0.990	0.990	0.995	0.990	0.988	0.988
2	1.002	1.001	0.996	1.000	0.992	1.009	1.005	0.998
3	0.995	0.998	0.996	0.995	0.995	0.997	1.003	1.004
4	0.991	0.990	0.991	0.991	0.991	0.991	0.992	0.993
5	0.991	0.989	0.988	0.986	0.982	0.993	0.988	0.986
6	1.008	1.007	1.006	1.006	1.002	1.012	1.014	1.010
7	1.009	1.011	1.014	1.011	1.000	1.013	1.016	1.010
8	1.009	1.011	1.012	1.009	1.001	1.010	1.011	1.007
9	1.009	1.009	1.012	1.010	1.005	1.013	1.015	1.009
10	1.004	1.001	1.003	1.002	0.999	1.003	1.006	1.002
11	1.001	0.998	1.001	1.001	0.999	1.002	1.006	1.002
12	0.999	0.996	0.999	0.997	0.997	0.998	1.001	0.999

Model prior, $\delta_k = 0.5$

	Predictive likelihood evaluated at horizon							
h	1	2	3	4	8	$1-4$	$1-8$	$1,4,8$
1	0.992	0.992	0.990	0.989	0.996	0.988	0.985	0.982
2	0.998	0.996	0.991	0.995	0.990	1.007	1.004	0.996
3	0.987	0.989	0.989	0.989	0.988	0.995	1.002	1.002
4	0.982	0.986	0.986	0.988	0.987	0.988	0.992	0.992
5	0.983	0.982	0.981	0.980	0.978	0.988	0.986	0.984
6	1.002	1.002	1.002	1.003	0.999	1.009	1.013	1.008
7	1.004	1.007	1.013	1.011	0.998	1.012	1.015	1.008
8	1.006	1.010	1.014	1.011	1.001	1.011	1.012	1.008
9	1.005	1.006	1.013	1.013	1.005	1.012	1.014	1.009
10	0.999	0.998	1.003	1.003	0.999	1.002	1.006	1.003
11	0.997	0.997	1.002	1.003	0.999	1.002	1.006	1.002
12	0.994	0.993	0.998	0.998	0.996	0.996	1.000	0.998