

Stochastic Search Model Selection for Restricted VAR Models

Edward George, Department of Statistics, Wharton School, University of Pennsylvania

Dongchu Sun, Department of Statistics, University of Missouri-Columbia

Shawn Ni, Department of Economics, University of Missouri-Columbia

Abstract

In this paper, we conduct Bayesian stochastic variable selection of Vector Autoregressive (VAR) models. We develop a Markov Chain Monte Carlo (MCMC) algorithm that selects the elements of both VAR regression coefficients matrix and error variance matrix. Numerical simulations show that stochastic search algorithm is quite effective in selecting the correct model in some cases and always results in better performance in forecast even when it does not succeed in selecting the correct model. We apply the method to examine VAR models of inflation transmission from Producer Price Index (PPI) of crude materials to PPI of finished goods and Consumer Price Index(CPI). We find that for the sample period from 1969 to 2003, the MLE and Bayesian estimates give different contemporaneous relationships of unexpected inflations in the identified VAR. Regarding the VAR regression coefficients, while neither MLE nor the Bayesian estimates show much delayed crude material price inflation pass-through to the PPI of finished goods and CPI, the MLE shows sizeable delayed feedback of PPI of finished goods and CPI inflation to PPI of crude materials, the corresponding Bayesian estimates are in smaller magnitude.

1 Introduction

Econometricians constantly face the challenge of building models that are flexible enough for economic analysis yet making the number of parameters small enough to allow for sharp inference from limited data. In the traditional simultaneous equations (SE) framework, economists impose restrictions on parameters under the justifications of economic theories. In large scale models, the restrictions are in many cases considered “incredible” (see Sims (1980)) and often yield unsatisfactory results. Vector Autoregressive (VAR) models, became widely used after the seminal work of Sims (1972,1980) as alternatives to the SE models for time series data. A p -dimensional column variable of interest, \mathbf{y}_t , is determined by

$$\mathbf{y}'_t = \mathbf{z}'_t \mathbf{C} + \sum_{j=1}^L \mathbf{y}'_{t-j} \mathbf{A}_j + \boldsymbol{\epsilon}'_t, \quad (1)$$

for $t = 1, \dots, T$, where \mathbf{z}_t is a h -dimensional vector of exogenous variable, L is a known positive integer, \mathbf{C} is a $h \times p$ unknown matrix, \mathbf{A}_j is an unknown $p \times p$ matrix, $\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_T$ are independent identically distributed $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ errors, and $\boldsymbol{\Sigma}$ is an unknown $p \times p$ positive definite matrix.

The lack of restriction on parameters in VAR models results in “over-parameterization” that leads to imprecise estimation and difficulties in frequentist finite sample inference. It is also known that over-parametrization hampers the predictive performance. The Bayesian approach, with the researcher’s knowledge on the parameters reflected in the priors, has been shown useful for finite sample inferences. Several approaches have been developed for prior elicitation and model restrictions.

One approach of prior elicitation is based on the general pattern of macroeconomic data. Doan et al. (1984) suggested a Minnesota prior that shrinks the parameters towards the random walk model. Applying VAR model (1) for policy analysis raises an additional issue. (1) is in a reduced form that may be consistent with a variety of “structural” models. This is because right-multiplying equation (1) by any full rank matrix gives rise to an equation of the same form of (1). The second approach of prior elicitation and model restrictions is to incorporate the information of structural economic models into VARs. Ingram & Whiteman (1994) and DeJong et al. (2000), among others, developed methods of using dynamic general equilibrium models for prior elicitation. Several authors (e.g., Sims (1986), Blanchard (1989), Gordon & Leeper (1994), and Sims & Zha (1998)) proposed restrictions on the covariance matrix $\boldsymbol{\Sigma}$ or equivalently on the contemporaneous relationship among variables without restricting the lag coefficients. Specifically, the identification

of VAR models is through a mapping of reduce-form error to structural shocks $\boldsymbol{\epsilon}'_t \mathbf{A}_0 = \mathbf{u}'_t$, with certain elements of \mathbf{A}_0 being set at zero. The covariance matrix of structural shocks is assumed to be identity. The identified VAR can be written as

$$\mathbf{y}'_t \mathbf{A}_0 = \mathbf{z}'_t \mathbf{C} \mathbf{A}_0 + \sum_{j=1}^L \mathbf{y}'_{t-j} \mathbf{A}_j \mathbf{A}_0 + \boldsymbol{\epsilon}'_t \mathbf{A}_0. \quad (2)$$

In the identified VAR literature $\boldsymbol{\Sigma}$ is estimated from the reduced-form VAR. The identification concerns conditions under which the structural parameters in \mathbf{A}_0 are recovered from $\boldsymbol{\Sigma}$ using

$$\boldsymbol{\Sigma} = \mathbf{A}_0^{-1'} \mathbf{A}_0^{-1}. \quad (3)$$

The identified VAR models differ from the traditional simultaneous equation (SE) models. Instead of setting the covariance matrix at identity, SE models impose restrictions on the matrix \mathbf{A}_0 and regression coefficients in model (2) to eliminate a large number of parameters.

The third approach to obtain sharper results is using the data sample under study to further restrict the model. This exercise constitutes model selection, a topic studied by a rich statistical literature. In the VAR framework, Hsiao (1981) offered a frequentist procedure to reduce the lag length of one of the variables in a bi-variate VAR.

In the present paper we propose a data sample based method of Bayesian VAR model selection through stochastic search. Our approach is fundamentally different from the those discussed above. The existing approaches restrict certain elements of the covariance matrix (or matrix \mathbf{A}_0) and/or regression coefficients to zero on the ground of formal or informal economic theory, based on an implicit assumption that there is only one or at most a few relevant models and we know them all. At least for some economic problems current theoretical development does not warrant such confidence. In this study we take a different perspective on model selection. We do not impose zero restrictions on any coefficient, meaning that we do not rule out any model a priori. Instead, we develop a method that allows for comparison of all possible models. The approach gives data much greater power in selecting models. We expect that in most applications our model selection results would show that many, even most, of the coefficients may turn out to be close to zero. Researchers can then focus on all potentially useful models, with the number of models considerably narrowed down and the probability of each model being true calculated by the data driven model selection. The regression models such as (2) are not only useful for confirming or rejecting existing theories, they can be used as a powerful vehicle of discovering new theories. Our approach includes variable lag selection as a special case. The added flexibility in selecting regression coefficients may

be practically useful. For instance, if certain seasonal variable is dependent more on the the N th lag but not much on lags shorter than N . The routine lag selection may result in inclusion of all N lags, consequently too many parameters, or if N is very large, selection of an incorrect model with lags 1 to J ($J < N$) for a finite sample. Our model selection avoids this difficulty.

The key of the data driven model selection is to be able to compare a large number of models. Variable selection problem in a univariate model with m explanatory variables involves comparing 2^m number of competing models. Even for moderate size m , the amount of computation is prohibitive. In the context of univariate model, George & McCulloch (1993) proposed a Bayesian MCMC stochastic search algorithm that greatly reduced the amount of computation. George (2000) provided a survey on recent developments in the area. In the VAR setting the number of elements to be selected is quite large. For a VAR model with p variables, h exogenous variables (including a constant) and L lags there are $2^{(Lp+h)p + \frac{p(p-1)}{2}}$ competing models. Clearly, comparing each of the all possible models is infeasible. In this study, we extend the George & McCulloch (1993) approach to VAR models. There are two substantial difference between selection of univariate models and VARs. First, selection of VARs concern off-diagonal elements of matrix \mathbf{A}_0 . Second, the regression coefficients are elements of a matrix instead of a vector. These extensions require analysis of new priors and development of new algorithms that allow for efficient simulation of posteriors. An additional difficulty for selecting the correct model is due to the serial correlation of the VAR variables that renders MLE biased in the finite sample.

We simulate a number of numerical examples in which we generate data from known models and compare the sample statistics of MLE, benchmark Bayesian VARs without model selection (i.e., conventional Bayesian VARs), and Bayesian VARs with stochastic search. The examples show that our stochastic search approach is effective. In some numerical examples, our approach achieves very high success rate in selecting the data-generating model. Even when stochastic search does not yields the correct model, it is still useful in improving predictions in terms of the frequentist MSE of point forecast. We also find that the Kullback-Liebler distance between predictive density averaged from all models visited in MCMC and that of the benchmark VAR is substantial.

An important advantage of the Bayesian model selection approach is that while conducting model selection, it produces finite sample inferences on parameters of interest. This advantage is important because it is very difficult to conduct finite sample inferences with frequentist methods. For example, in practice, the numbers of parameters in VAR models are typically quite large relative

to the observations of macroeconomic data, which makes asymptotic theory an unreliable guidance for finite sample properties of VAR estimators. Bayesian method is effective for joint finite sample inference of VAR models and parameters given each model.

The paper is organized as follows. Section 2 defines the notations and discusses priors. Section 3 derives conditional posterior distributions and lays out the Bayesian algorithms. Section 4 discusses criteria of evaluating the performance of model selection. Section 5 reports simulation results. Section 6 applies our method to a real problem. Section 7 offers concluding remarks.

2 The Model and Prior

2.1 Likelihood and Decomposition of the Precision Matrix

Define $\mathbf{x}'_t = (\mathbf{z}'_t, \mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-L})$. We then rewrite (1) in the familiar matrix form

$$\mathbf{Y} = \mathbf{X}\Phi + \boldsymbol{\epsilon}, \quad (4)$$

where

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}'_1 \\ \vdots \\ \mathbf{y}'_T \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}'_1 \\ \vdots \\ \mathbf{x}'_T \end{pmatrix}, \quad \Phi = \begin{pmatrix} \mathbf{c} \\ \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_L \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \boldsymbol{\epsilon}'_1 \\ \vdots \\ \boldsymbol{\epsilon}'_T \end{pmatrix}. \quad (5)$$

Here \mathbf{Y} and $\boldsymbol{\epsilon}$ are $T \times p$ matrices, Φ is a $(h + Lp) \times p$ matrix, \mathbf{x}'_t is a $1 \times (h + Lp)$ row vector, and \mathbf{X} is a $T \times (h + Lp)$ matrix of observations. The likelihood function of (Φ, Σ) is

$$f(\mathbf{Y} | \Phi, \Sigma) \propto |\Sigma|^{-T/2} \text{etr} \left\{ -\frac{1}{2} (\mathbf{Y} - \mathbf{X}\Phi) \Sigma^{-1} (\mathbf{Y} - \mathbf{X}\Phi)' \right\}. \quad (6)$$

Write

$$\mathbf{S}(\Phi) = (\mathbf{Y} - \mathbf{X}\Phi)' (\mathbf{Y} - \mathbf{X}\Phi). \quad (7)$$

A commonly used estimate of (Φ, Σ) is the maximum likelihood estimate (MLE):

$$\hat{\Phi}_M = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}, \quad \hat{\Sigma}_M = \frac{1}{T} \mathbf{S}(\hat{\Phi}_M). \quad (8)$$

Here $\mathbf{S}(\hat{\Phi}_M)$ is the residual sum of squares. We assume that the sample size T is large enough so that the MLEs of Φ and Σ exist with probability one.

Due to the fact that a covariance matrix or precision matrix is non-negative definite, re-parameterization is usually needed for its estimation or simulation. Leonard & Hsu (1992) and Yang & Berger (1994) simulated the matrix logarithm of the covariance matrix. Pourahmadi (1999) provided an iterative estimation formula of the covariance matrix from iid observations. Recently, Wong et al. (2002) conducted Bayesian estimation of the Cholesky decomposition of the precision matrix Σ^{-1} , where Σ^{-1} was decomposed as a product of a diagonal matrix and an upper-triangular matrix with unit diagonal elements: $\Sigma^{-1} = (\sigma^{ij}) = (B_i B_j \lambda_{ij}) = \mathbf{B}\mathbf{A}\mathbf{B}$. Here \mathbf{B} is $\text{diag}(B_1, \dots, B_p)$, and \mathbf{A} is the partial correlation matrix with unity diagonal elements. They place gamma priors on B_i^2 and mixture of discrete and constant priors on the upper triangular elements of the partial correlation matrix. The simulation algorithms for the \mathbf{B} matrix and \mathbf{A} matrix both involve MCMC Metropolis-Hastings steps. A drawback of their simulation routine is that many of the simulated Σ^{-1} are not positive definite and must be discarded.

Our research on the decomposition of the precision matrix Σ^{-1} ($\Sigma^{-1} = (\sigma^{ij})$) is motivated by the condition of the identified VAR (3). We consider decomposition of not only the just-identified but the over-identified VAR as well. The elements of the precision matrix, σ^{ij} , are often important since $-\sigma^{ij}/\sqrt{\sigma^{ii}\sigma^{jj}}$ is a partial correlation coefficient. For example, if $\sigma^{12} = 0$, the first two components of the errors are independent given the rest of the components.

We decompose the precision matrix as

$$\Sigma^{-1} = \Psi_p \Psi_p', \quad (9)$$

where Ψ_p is the $p \times p$ upper-triangular matrix with ψ_{ij} as its (i, j) th entry, so $\psi_{ij} = 0$ for $i > j$. Note that there is no restriction on off-diagonal elements ψ_{ij} , but ψ_{ii} should never change sign, so we can assume that $\psi_{ii} > 0$. Such a matrix Ψ_p without restriction on ψ_{ij} is called normalized and just-identified by Sims & Zha (1998).

Much progress has been made on Bayesian VAR models that have restrictions on the Ψ matrix. Sims & Zha (1998), Sims & Zha (1998), and Waggoner & Zha (2003) proposed Bayesian analysis based on normal priors on Ψ and Φ . In each study, the authors assumed a given form of Ψ and did not consider assessing the possibility that the data supported alternative models. On the other hand, their approach is not restricted to normalized Ψ , which generally makes the MCMC more difficult. Waggoner & Zha (2003) suggested a Gibbs sampler that simulates each column of Ψ from an orthonormal basis that is orthogonal to other columns of Ψ . In this paper, we focus on the selection of normalized VAR models for two reasons. First, such a structure grants nice analytical

form of conditional posteriors. Second, and more importantly, it assures that the selected models are globally identified. An arbitrary Ψ with $p(p-1)/2$ or more restrictions does not necessarily have an one-to-one mapping with the Σ matrix via (9). Due to the nonlinear nature of the mapping (9), one can only derive local rank conditions for identification of Ψ . Verifying the validity of the conditions for parameters in the posterior space is difficult. Furthermore, even when the local rank condition is satisfied, the mapping from Σ to Ψ may still not be unique. An example of multiple mappings from Σ to Ψ given by Bekker & Pollock (1986) is shown to satisfy the local identifying condition by Amisano & Giannini (1997). Once we restrict the Ψ matrix to be upper-triangular, the mapping (9) is unique.

The covariance matrix simulated by our MCMC procedure has two good properties. First, by construction, every simulated Σ^{-1} is positive definite. Second, the simulation is iterative in the column components of \mathbf{A}_0 . Most of the conditional posteriors are standard distributions. The MCMC simulation does not involve Metropolis algorithms and is expected to be efficient.

2.2 Priors

In the traditional multivariate regression literature, Bayesian analysis of VAR models typically focuses on Φ or Σ as a whole, without component-wise specifications. For example, for priors on Σ , the most common ones are the informative Wishart and the noninformative Jeffreys prior (See Geisser (1965), Zellner & Tiao (1964)) or reference prior of Yang & Berger (1994). In the identified VAR model, restrictions are placed on the components of the \mathbf{A}_0 matrix. Sims & Zha (1998) assigned a multivariate normal prior to the vectorized \mathbf{A}_0 matrix and simulated the posteriors. In the following, we propose an alternative approach of imposing priors on components of \mathbf{A}_0 that utilizes the structure of the matrix and allows for easier Bayesian computation.

Let $m = (Lp + h)p$, the total number of unknown regression coefficients. Denote $\phi = \text{vec}(\Phi) = (\phi_1, \phi_2, \dots, \phi_m)'$. For $j = 2, \dots, p$, let $\eta_j = (\psi_{1j}, \dots, \psi_{j-1,j})'$. Write $\eta = (\eta'_2, \dots, \eta'_p)'$ and $\psi = (\psi_{11}, \dots, \psi_{pp})'$. We now propose hierarchical priors for (ϕ, η, ψ) . Here and in the following, we use $(\cdot | \cdot)$ to denote conditional distribution and $[\cdot | \cdot]$ to denote conditional density.

(i) Priors of ϕ . Assume that m elements in ϕ are subject to selection and the rest $(h + Lp) \times p$ elements are always included in the model. The prior for the elements of ϕ that are included in every model is

$$\phi_{non} \sim N_{(h+Lp) \times p - m}(\phi_{non}, \mathbf{M}_{non}), \quad (10)$$

where $\boldsymbol{\phi}_{non}, \mathbf{M}_{non}$ are hyperparameters. Now we turn to the m elements of $\boldsymbol{\phi}$, $\boldsymbol{\phi}_m = (\phi_{s1}, \dots, \phi_{sm})$, considered for selection. Let γ_i be independent Bernoulli (p_i) random variables for fixed $p_i \in (0, 1)$. For given $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_m)$, denote $\mathbf{D} = \text{diag}(h_1\tau_1, \dots, h_m\tau_m)$ and

$$h_i = c_i^{\gamma_i} = \begin{cases} 1, & \text{if } \gamma_i = 0, \\ c_i, & \text{if } \gamma_i = 1. \end{cases}$$

Here τ_i are small and c_i are large constants. Conditional on the model index $\boldsymbol{\gamma}$, the prior for the m elements of $\boldsymbol{\psi}$ for selection is

$$(\boldsymbol{\phi}_m | \boldsymbol{\gamma}) \sim N_m(\mathbf{0}, \mathbf{DRD}), \quad (11)$$

where \mathbf{R} is a known correlation matrix. When $\mathbf{R} = \mathbf{I}_m$, the m elements of $\boldsymbol{\psi}$ to be selected have the conditional prior

$$(\phi_{si} | \gamma_i) \stackrel{ind}{\sim} (1 - \gamma_i)N(0, \tau_i^2) + \gamma_i N(0, c_i^2 \tau_i^2). \quad (12)$$

Combining these independent priors we have

$$(\boldsymbol{\phi} | \boldsymbol{\gamma}) \sim N(\boldsymbol{\phi}_0^{(\boldsymbol{\gamma})}, \mathbf{M}^{(\boldsymbol{\gamma})}).$$

For elements of $\boldsymbol{\phi}$ that are included in every model, the corresponding elements $\boldsymbol{\phi}^{(\boldsymbol{\gamma})}$ and $\mathbf{M}^{(\boldsymbol{\gamma})}$ are given by (10). For other elements of $\boldsymbol{\phi}$ the corresponding elements $\boldsymbol{\phi}^{(\boldsymbol{\gamma})}$ and $\mathbf{M}^{(\boldsymbol{\gamma})}$ are given by (11).

(ii) Priors of $\boldsymbol{\gamma}$. We assume the independent Bernoulli priors for $\boldsymbol{\gamma}$:

$$P(\gamma_i = 1) = p_i, P(\gamma_i = 0) = 1 - p_i, \text{ for } i = 1, \dots, m. \quad (13)$$

(iii) Priors of $\boldsymbol{\eta}$ $= (\psi_{12}, \psi_{13}, \psi_{23}, \dots, \psi_{p-1,p})'$. Let ω_{ij} have independent Bernoulli (q_{ij}) random variables for $1 \leq i < j \leq p$. For $j = 2, \dots, p$, set $\boldsymbol{\omega}_j = (\omega_{1j}, \dots, \omega_{j-1,j})'$. For given $\boldsymbol{\omega}_j$, denote $\mathbf{D}_j = \text{diag}(h_{1j}\kappa_{1j}, \dots, h_{j-1,j}\kappa_{j-1,j})$, where

$$h_{ij} = d_{ij}^{\omega_{ij}} = \begin{cases} 1, & \text{if } \omega_{ij} = 0, \\ d_{ij}, & \text{if } \omega_{ij} = 1. \end{cases}$$

Here κ_{ij} are small and d_{ij} are large constants. We assume that

$$(\boldsymbol{\eta}_j | \boldsymbol{\omega}_j) \stackrel{ind}{\sim} N_{j-1}(\mathbf{0}, \mathbf{D}_j \mathbf{R}_j \mathbf{D}_j), \text{ for } j = 2, \dots, p. \quad (14)$$

where \mathbf{R}_j is a $(j-1) \times (j-1)$ known correlation matrix. When $\mathbf{R}_j = \mathbf{I}_{j-1}$,

$$(\psi_{ij} | \omega_{ij}) \stackrel{ind}{\sim} (1 - \omega_{ij})N(0, \kappa_{ij}^2) + \omega_{ij}N(0, d_{ij}^2 \kappa_{ij}^2), \text{ for } i = 1, \dots, j-1. \quad (15)$$

(iv) **Priors of $\boldsymbol{\psi}$** $= (\psi_{11}, \dots, \psi_{pp})'$. Assume that $\psi_{ii}^2 \stackrel{ind}{\sim}$ gamma (a_i, b_i) distributions. Here (a_i, b_i) are positive constants. So for $i = 1, \dots, p$, ψ_{ii} has the density

$$[\psi_{ii}] = \frac{2b_i^{a_i}}{\Gamma(a_i)} \psi_{ii}^{2(a_i-1)} \exp(-b_i \psi_{ii}^2), \quad \text{for } \psi_{ii} > 0. \quad (16)$$

(v) **Priors of $\boldsymbol{\omega}$** . We assume the independent Bernoulli priors for $\boldsymbol{\omega}$:

$$P(\omega_{ij} = 1) = q_{ij}, \quad P(\omega_{ij} = 0) = 1 - q_{ij}, \quad i = 1, \dots, p, \quad j = 1, \dots, p-1. \quad (17)$$

3 Bayesian Model Selections

We first consider model selections of $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ in separation, and then consider joint selection of $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$.

3.1 Selection of $\boldsymbol{\Phi}$

We first discuss model selection pertaining to regression coefficients $\boldsymbol{\Phi}$. A special example of selection of $\boldsymbol{\Phi}$ in applications is specification of VAR lags. Researcher often discover that quantities of interest change dramatically when the VAR lag is increased from L to $L+1$. The sensitivity of VAR estimates to lag specification highlights the restriction implied by the VAR lag specification: if the lag is specified too short the results may be very misleading, but if the lag is specified long then the large number of parameters may result in erratic estimates. In addition, in some cases, researchers may find it desirable to include in the models only some of the lag coefficients \mathbf{A}_j ($1 \leq j \leq L$) instead all of them, or to include some of the elements in \mathbf{A}_j but not the entire matrix. In these cases, a more flexible model selection approach is needed. The selection we are considering is more general than the lag specification. It is also an extension of the approach by (George & McCulloch (1993)) to VAR model.

For $\widehat{\boldsymbol{\Phi}}_M$ given in (8), denote $\widehat{\boldsymbol{\phi}}_M = \text{vec}(\widehat{\boldsymbol{\Phi}}_M)$. The likelihood function (6) of $(\boldsymbol{\phi}, \boldsymbol{\Psi})$ can be rewritten as

$$f(\mathbf{Y} \mid \boldsymbol{\phi}, \boldsymbol{\Psi}) \propto |\boldsymbol{\Psi}|^T \exp \left[-\frac{1}{2} (\boldsymbol{\phi} - \widehat{\boldsymbol{\phi}}_M)' \{ \boldsymbol{\Psi} \boldsymbol{\Psi}' \otimes (\mathbf{X}' \mathbf{X}) \} (\boldsymbol{\phi} - \widehat{\boldsymbol{\phi}}_M) - \frac{1}{2} \text{tr} \{ \boldsymbol{\Psi} \boldsymbol{\Psi}' \mathbf{S}(\widehat{\boldsymbol{\Phi}}_M) \} \right]. \quad (18)$$

For model selection concerning $\boldsymbol{\Phi}$ only, all elements of the index matrix $\boldsymbol{\omega}$ are unity. The computation of conditional posteriors of $\boldsymbol{\phi}$, $\boldsymbol{\eta}$, $\boldsymbol{\psi}$, $\boldsymbol{\gamma}$ is as follows.

Fact 1 (a) *The conditional posterior distribution*

$$\boldsymbol{\phi} \mid (\boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\psi}; \mathbf{Y}) \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Delta}), \quad (19)$$

where

$$\begin{aligned} \boldsymbol{\mu} &= \{(\boldsymbol{\Psi}\boldsymbol{\Psi}') \otimes (\mathbf{X}'\mathbf{X}) + (\mathbf{M}^{(\boldsymbol{\gamma})})^{-1}\}^{-1} \{(\boldsymbol{\Psi}\boldsymbol{\Psi}') \otimes (\mathbf{X}'\mathbf{X})\} \widehat{\boldsymbol{\phi}}_M + (\mathbf{M}^{(\boldsymbol{\gamma})})^{-1} \boldsymbol{\phi}_0^{(\boldsymbol{\gamma})}; \\ \boldsymbol{\Delta} &= \{(\boldsymbol{\Psi}\boldsymbol{\Psi}') \otimes (\mathbf{X}'\mathbf{X}) + (\mathbf{M}^{(\boldsymbol{\gamma})})^{-1}\}^{-1} \end{aligned}$$

(b) Denote $\boldsymbol{\gamma}_{(-i)} = (\gamma_1, \dots, \gamma_{i-1}, \gamma_{i+1}, \dots, \gamma_m)$. Then under prior (11)

$$(\gamma_i \mid \boldsymbol{\phi}, \boldsymbol{\gamma}_{(-i)}, \boldsymbol{\eta}, \boldsymbol{\psi}; \mathbf{Y}) \sim \text{Bernoulli}(u_{i1}/(u_{i1} + u_{i2})),$$

where

$$u_{i1} = [\boldsymbol{\phi} \mid \boldsymbol{\gamma}_{(-i)}, \gamma_i = 1] p_i, \quad (20)$$

$$u_{i2} = [\boldsymbol{\phi} \mid \boldsymbol{\gamma}_{(-i)}, \gamma_i = 0] (1 - p_i). \quad (21)$$

If $\mathbf{R} = \mathbf{I}_m$ in prior (11), then for given $(\boldsymbol{\phi}; \boldsymbol{\eta}_j, \psi_{ii}, 1 \leq i \leq p; \mathbf{Y})$, $\gamma_i \stackrel{\text{ind}}{\sim} \text{Bernoulli}(\tilde{u}_{i1}/(\tilde{u}_{i1} + \tilde{u}_{i2}))$, where

$$\tilde{u}_{i1} = \frac{1}{c_i} \exp\left(-\frac{\phi_i^2}{2c_i^2 \tau_i^2}\right) p_i, \quad (22)$$

$$\tilde{u}_{i2} = \exp\left(-\frac{\phi_i^2}{2\tau_i^2}\right) (1 - p_i). \quad (23)$$

Proof. Part (a) is standard. For part (b), note that as in George & McCulloch (1993), the conditional distribution $(\gamma_i \mid \boldsymbol{\phi}, \boldsymbol{\gamma}_{(-i)}, \boldsymbol{\eta}, \boldsymbol{\psi}; \mathbf{Y})$ is not dependent on \mathbf{Y} due to the hierarchial structure of the model parameters. \square

Let the prior on $\boldsymbol{\eta}$ be

$$\boldsymbol{\eta}_j \stackrel{\text{ind}}{\sim} N_{j-1}(\mathbf{0}, \boldsymbol{\Omega}_{j-1}), \quad \text{for } j = 2, \dots, p. \quad (24)$$

The posterior of $\boldsymbol{\phi}, \boldsymbol{\Psi}, \boldsymbol{\gamma}$ can be draw from the following procedure.

1. draw $[\boldsymbol{\psi}_{(k)} \mid \boldsymbol{\phi}_{(k-1)}, \mathbf{Y}]$ from the gamma distribution

$$(\psi_{ii}^2 \mid \boldsymbol{\phi}; \mathbf{Y}) \sim \text{gamma}(a_i + \frac{1}{2}T, B_i), \quad (25)$$

where

$$B_i = \begin{cases} b_1 + \frac{1}{2}s_{11}, & \text{if } i = 1, \\ b_i + \frac{1}{2}\{s_{ii} - \mathbf{s}'_{i-1,i}(\mathbf{S}_{i-1} + \boldsymbol{\Omega}_{i-1}^{-1})^{-1}\mathbf{s}_{i-1,i}\}, & \text{if } i = 2, \dots, p; \end{cases} \quad (26)$$

2. draw $[\boldsymbol{\eta}_{(k)} \mid \boldsymbol{\psi}_{(k)}, \boldsymbol{\phi}_{(k-1)} \mathbf{Y}]$ from normal distribution

$$(\boldsymbol{\eta}_i \mid \psi_{ii}; \mathbf{S}_i) \propto N(-\psi_{ii}(\mathbf{S}_{i-1} + \boldsymbol{\Omega}_{i-1}^{-1})^{-1} \mathbf{s}_{i-1,i}, (\mathbf{S}_{i-1} + \boldsymbol{\Omega}_{i-1}^{-1})^{-1}); \quad (27)$$

3. draw $[\boldsymbol{\phi}_{(k)} \mid \boldsymbol{\gamma}_{(k-1)}, \boldsymbol{\Sigma}_{(k)}; \mathbf{Y}]$ from normal distribution (19) where $\boldsymbol{\Sigma}_{(k)}$ is computed from $\boldsymbol{\psi}_{(k)}$ and $\boldsymbol{\eta}_{(k)}$;

4. draw $[\boldsymbol{\gamma}_{(k)} \mid \boldsymbol{\phi}_{(k)}]$ from Bernoulli distribution (20).

3.2 Model Selection of Identifying Restrictions on $\boldsymbol{\Psi}$

It often the case that for identification VARs economists are more interested in selecting models for $\boldsymbol{\Psi}$ and not much concerned about restrictions in the regression coefficient $\boldsymbol{\Phi}$. Focusing on $\boldsymbol{\Psi}$ greatly reduced the total number of candidate models. In many applications, researchers have fairly good knowledge on the nature of VAR regression coefficients. For instance, many macroeconomic time series such as GDP and stock prices exhibit random walk type of behavior (which is the base of the well-known Minnesota prior on $\boldsymbol{\Phi}$). Incorporating such prior information may improve the Bayesian inference on $\boldsymbol{\Phi}$ as well as $\boldsymbol{\Psi}$. We now conduct this more restricted model selection.

Let the prior on $\boldsymbol{\phi}$ be

$$\boldsymbol{\phi} \sim N(\boldsymbol{\phi}_0, \boldsymbol{\Xi}_0). \quad (28)$$

The conditional posterior distribution of $\boldsymbol{\phi}$ given $(\boldsymbol{\Psi}_p; \mathbf{Y})$ is

$$(\boldsymbol{\phi} \mid \boldsymbol{\Psi}_p; \mathbf{Y}) \sim N(\widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\Xi}}), \quad (29)$$

where

$$\widehat{\boldsymbol{\Xi}} = \left\{ \boldsymbol{\Psi} \boldsymbol{\Psi}' \otimes (\mathbf{X}' \mathbf{X}) + \boldsymbol{\Xi}_0^{-1} \right\}^{-1}, \quad (30)$$

$$\widehat{\boldsymbol{\phi}} = \widehat{\boldsymbol{\Xi}} \left\{ \boldsymbol{\Psi} \boldsymbol{\Psi}' \otimes (\mathbf{X}' \mathbf{X}) \widehat{\boldsymbol{\phi}}_M + \boldsymbol{\Xi}_0^{-1} \boldsymbol{\phi}_0 \right\}. \quad (31)$$

To derive other conditional distributions, note that the likelihood function (6) of $(\boldsymbol{\phi}, \boldsymbol{\Psi})$ can also be rewritten as

$$f(\mathbf{Y} \mid \boldsymbol{\phi}, \boldsymbol{\Psi}) \propto |\boldsymbol{\Psi}|^T \text{etr} \left\{ -\frac{1}{2} \boldsymbol{\Psi}' \mathbf{S}(\boldsymbol{\Phi}) \boldsymbol{\Psi} \right\}, \quad (32)$$

where $\mathbf{S}(\boldsymbol{\Phi})$ is given by (7). Write $\mathbf{S}(\boldsymbol{\Phi}) = (s_{ij})$. For $j = 2, \dots, p$, define $\mathbf{s}_j = (s_{1j}, \dots, s_{j-1,j})'$. Let \mathbf{S}_j be the upper-left $j \times j$ submatrix of $\mathbf{S}(\boldsymbol{\Phi})$. Define $v_1 = s_{11}$ and $v_i = |\mathbf{S}_i|/|\mathbf{S}_{i-1}|$ for $i = 2, \dots, p$.

It is well known that $v_i = s_{ii} - \mathbf{s}'_i \mathbf{S}_{i-1}^{-1} \mathbf{s}_i > 0$ for $i = 2, \dots, p$. Then (32) equals to

$$f(\mathbf{Y} | \boldsymbol{\phi}, \boldsymbol{\Psi}) \propto \prod_{i=1}^p \psi_{ii}^T \exp \left[-\frac{1}{2} \left\{ \sum_{i=1}^p \psi_{ii}^2 v_i + \sum_{j=2}^p (\boldsymbol{\eta}_j + \psi_{jj} \mathbf{S}_{j-1}^{-1} \mathbf{s}_j)' \mathbf{S}_{j-1} (\boldsymbol{\eta}_j + \psi_{jj} \mathbf{S}_{j-1}^{-1} \mathbf{s}_j) \right\} \right]. \quad (33)$$

This expression allows us to derive the conditional posterior of $\boldsymbol{\Psi}$.

Fact 2 (a) For given $(\boldsymbol{\phi}, \boldsymbol{\omega}, \boldsymbol{\psi}; \mathbf{Y})$, the posterior distributions of $\boldsymbol{\eta}_2, \dots, \boldsymbol{\eta}_p$ are independent, and

$$(\boldsymbol{\eta}_j | \boldsymbol{\phi}, \boldsymbol{\omega}, \boldsymbol{\psi}; \mathbf{Y}) \sim N_{j-1}(\boldsymbol{\mu}_j, \boldsymbol{\Delta}_j), \quad (34)$$

where

$$\boldsymbol{\mu}_j = -\psi_{jj} \{ \mathbf{S}_{j-1} + (\mathbf{D}_j \mathbf{R}_j \mathbf{D}_j)^{-1} \}^{-1} \mathbf{s}_j, \quad (35)$$

$$\boldsymbol{\Delta}_j = \{ \mathbf{S}_{j-1} + (\mathbf{D}_j \mathbf{R}_j \mathbf{D}_j)^{-1} \}^{-1}. \quad (36)$$

(b) For given $(\boldsymbol{\phi}, \boldsymbol{\omega}; \mathbf{Y})$, the posterior distributions of $\psi_{11}^2, \dots, \psi_{pp}^2$ are independent and

$$(\psi_{ii}^2 | \boldsymbol{\phi}, \boldsymbol{\omega}; \mathbf{Y}) \sim \text{gamma}(a_i + \frac{1}{2}T, B_i), \quad (37)$$

where

$$B_i = \begin{cases} b_1 + \frac{1}{2}s_{11}, & \text{if } i = 1, \\ b_i + \frac{1}{2} \{ s_{ii} - \mathbf{s}'_i [\mathbf{S}_{i-1} + (\mathbf{D}_i \mathbf{R}_i \mathbf{D}_i)^{-1}]^{-1} \mathbf{s}_i \}, & \text{if } i = 2, \dots, p. \end{cases} \quad (38)$$

(c) For $j = 2, \dots, p$ and $i = 1, \dots, j-1$, denote $\boldsymbol{\omega}_{(-ij)} = (\omega_{1j}, \dots, \omega_{i-1,j}, \omega_{i+1,j}, \dots, \omega_{j-1,j})'$. For given $(\boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\omega}_{(-ij)}, \boldsymbol{\omega}_k, k \neq j; \mathbf{Y})$, $\omega_{ij} \sim \text{Bernoulli}(u_{ij1}/(u_{ij1} + u_{ij2}))$, where

$$u_{ij1} = [\boldsymbol{\eta}_j | \boldsymbol{\omega}_{(-ij)}, \omega_{ij} = 1] q_{ij}, \quad (39)$$

$$u_{ij2} = [\boldsymbol{\eta}_j | \boldsymbol{\omega}_{(-ij)}, \omega_{ij} = 0] (1 - q_{ij}). \quad (40)$$

If $\mathbf{R}_j = \mathbf{I}_{j-1}$ in prior (14), for fixed $j = 2, \dots, p$ and given $(\boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\omega}_k, k \neq j; \mathbf{Y})$, $\omega_{1j}, \dots, \omega_{j-1,j}$ are independent. $(\omega_{ij} | \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\omega}_k, k \neq j; \mathbf{Y}) \stackrel{\text{ind}}{\sim} \text{Bernoulli}(\tilde{u}_{ij1}/(\tilde{u}_{ij1} + \tilde{u}_{ij2}))$, where

$$\tilde{u}_{ij1} = \frac{1}{d_{ij}} \exp\left(-\frac{\psi_{ij}^2}{2d_{ij}^2 \kappa_{ij}^2}\right) q_{ij}, \quad (41)$$

$$\tilde{u}_{ij2} = \exp\left(-\frac{\psi_{ij}^2}{2\kappa_{ij}^2}\right) (1 - q_{ij}). \quad (42)$$

Proof. For $i = 2, \dots, p$, the conditional posterior density of $(\psi_{ii}^2, i = 1, \dots, p; \boldsymbol{\eta}_j, j = 2, \dots, p)$ given $(\boldsymbol{\phi}, \boldsymbol{\omega}; \mathbf{Y})$ is given by

$$\begin{aligned}
& [\psi_{ii}^2, i = 1, \dots, p; \boldsymbol{\eta}_j, j = 2, \dots, p \mid \boldsymbol{\phi}, \boldsymbol{\omega}; \mathbf{Y}] \\
& \propto \left\{ \prod_{i=1}^p (\psi_{ii}^2)^{a_i + \frac{T}{2} - 1} \right\} \exp \left[-\frac{1}{2} \left\{ \sum_{i=1}^p \psi_{ii}^2 (w_i + 2b_i) \right\} \right] \exp \left\{ -\frac{1}{2} \sum_{i=2}^p \boldsymbol{\eta}'_i \mathbf{D}_i \mathbf{R}_i \mathbf{D}_i \boldsymbol{\eta}_i \right\} \\
& \times \exp \left\{ -\frac{1}{2} \sum_{i=2}^p (\boldsymbol{\eta}_i + \psi_{ii} \mathbf{S}_{i-1}^{-1} \mathbf{s}_i)' \mathbf{S}_{i-1} (\boldsymbol{\eta}_i + \psi_{ii} \mathbf{S}_{i-1}^{-1} \mathbf{s}_i) \right\} \\
& = \left\{ \prod_{i=1}^p (\psi_{ii}^2)^{a_i + \frac{T}{2} - 1} \right\} \exp \left\{ -\sum_{i=1}^p \psi_{ii}^2 B_i - \frac{1}{2} \sum_{j=2}^p (\boldsymbol{\eta}_j - \boldsymbol{\mu}_j)' \boldsymbol{\Delta}_j^{-1} (\boldsymbol{\eta}_j - \boldsymbol{\mu}_j) \right\},
\end{aligned}$$

where $\boldsymbol{\mu}_j$ is defined in (35) and $\boldsymbol{\Delta}_j$ in (36). Part (a) is obvious. For part (b), the case when $i = 1$ holds clearly; when $i = 2, \dots, p$ the result follows by integrating out $\boldsymbol{\eta}_j$. \square

There is no need to be concerned with the model selection index $\boldsymbol{\gamma}$. The Gibbs MCMC sampling procedure is modified as follows. Suppose in cycle k , we have $\boldsymbol{\psi}_{(k-1)}, \boldsymbol{\eta}_{(k-1)}, \boldsymbol{\phi}_{(k-1)}, \boldsymbol{\omega}_{(k-1)}$ given. Then we

1. draw $[\boldsymbol{\psi}_{(k)} \mid \boldsymbol{\phi}_{(k-1)}, \boldsymbol{\omega}_{(k-1)}, \mathbf{Y}]$ from the gamma distribution (25);
2. draw $[\boldsymbol{\eta}_{(k)} \mid \boldsymbol{\psi}_{(k)}, \boldsymbol{\phi}_{(k-1)}, \boldsymbol{\omega}_{(k-1)}, \mathbf{Y}]$ from the normal distribution (34);
3. draw $[\boldsymbol{\omega}_{(k)} \mid \boldsymbol{\eta}_{(k)}]$ from the Bernoulli distribution (39);
4. draw $[\boldsymbol{\phi}_{(k)} \mid \boldsymbol{\Sigma}_{(k)}, \boldsymbol{\omega}_{(k)}; \mathbf{Y}]$ from the normal distribution (29), where $\boldsymbol{\Sigma}_{(k)}$ is computed from $\boldsymbol{\psi}_{(k)}$ and $\boldsymbol{\eta}_{(k)}$.

3.3 Joint Model Selection of Φ and Ψ

We now consider joint selection of Φ and Ψ . To conduct MCMC simulations, we need derive the full conditional distributions for $\boldsymbol{\theta} \equiv (\boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{\psi})$ for given data \mathbf{Y} , where $\boldsymbol{\omega} = (\boldsymbol{\omega}'_2, \dots, \boldsymbol{\omega}'_p)'$.

Note that the joint posterior of $(\boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\omega}, \boldsymbol{\psi})$ is then

$$[\boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\eta}; \boldsymbol{\gamma}, \boldsymbol{\omega}; \mathbf{Y}] \propto f(\mathbf{Y} \mid \boldsymbol{\phi}, \boldsymbol{\Psi}) [\boldsymbol{\phi} \mid \boldsymbol{\gamma}] [\boldsymbol{\gamma}] [\boldsymbol{\eta} \mid \boldsymbol{\omega}] [\boldsymbol{\omega}] [\boldsymbol{\psi}].$$

The full conditional distributions are given as follows.

Based on the above discussion we design the following Gibbs MCMC sampling procedure. Suppose in cycle k , we have $\boldsymbol{\psi}_{(k-1)}, \boldsymbol{\eta}_{(k-1)}, \boldsymbol{\phi}_{(k-1)}, \boldsymbol{\gamma}_{(k-1)}, \boldsymbol{\omega}_{(k-1)}$ given. Then we

1. draw $[\boldsymbol{\psi}_{(k)} \mid \boldsymbol{\phi}_{(k-1)}, \boldsymbol{\gamma}_{(k-1)}, \boldsymbol{\omega}_{(k-1)}, \mathbf{Y}]$ from the gamma distribution

$$(\psi_{ii}^2 \mid \boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\omega}; \mathbf{Y}) \sim \text{gamma}(a_i + \frac{1}{2}T, B_i), \quad (43)$$

where

$$B_i = \begin{cases} b_1 + \frac{1}{2}s_{11}, & \text{if } i = 1, \\ b_i + \frac{1}{2}\{s_{ii} - \mathbf{s}'_i[\mathbf{S}_{i-1} + (\mathbf{D}_i\mathbf{R}_i\mathbf{D}_i)^{-1}]^{-1}\mathbf{s}_i\}, & \text{if } i = 2, \dots, p. \end{cases} \quad (44)$$

2. draw $[\boldsymbol{\eta}_{(k)} \mid \boldsymbol{\psi}_{(k)}, \boldsymbol{\phi}_{(k-1)}, \boldsymbol{\gamma}_{(k-1)}, \boldsymbol{\omega}_{(k-1)}, \mathbf{Y}]$ from normal distribution (45);

$$(\boldsymbol{\eta}_j \mid \boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\omega}, \boldsymbol{\psi}; \mathbf{Y}) \sim N_{j-1}(\boldsymbol{\mu}_j, \boldsymbol{\Delta}_j), \quad (45)$$

where

$$\boldsymbol{\mu}_j = -\psi_{jj}\{\mathbf{S}_{j-1} + (\mathbf{D}_j\mathbf{R}_j\mathbf{D}_j)^{-1}\}^{-1}\mathbf{s}_j, \quad (46)$$

$$\boldsymbol{\Delta}_j = \{\mathbf{S}_{j-1} + (\mathbf{D}_j\mathbf{R}_j\mathbf{D}_j)^{-1}\}^{-1}. \quad (47)$$

3. draw $[\boldsymbol{\omega}_{(k)} \mid \boldsymbol{\eta}_{(k)}, \boldsymbol{\psi}_{(k)}, \boldsymbol{\phi}_{(k-1)}, \boldsymbol{\gamma}_{(k-1)}, \boldsymbol{\omega}_{(k-1)}, \mathbf{Y}]$ from Bernoulli distribution (39).

4. draw $[\boldsymbol{\phi}_{(k)} \mid \boldsymbol{\gamma}_{(k-1)}, \boldsymbol{\Sigma}_{(k)}, \boldsymbol{\omega}_{(k)}; \mathbf{Y}]$ from normal distribution (48), where $\boldsymbol{\Sigma}_{(k)}$ is computed from $\boldsymbol{\psi}_{(k)}$ and $\boldsymbol{\eta}_{(k)}$

$$\boldsymbol{\phi} \mid (\boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{\psi}; \mathbf{Y}) \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Delta}), \quad (48)$$

where

$$\boldsymbol{\mu} = \{(\boldsymbol{\Psi}\boldsymbol{\Psi}') \otimes (\mathbf{X}'\mathbf{X}) + (\mathbf{M}^{(\gamma)})^{-1}\}^{-1}\{(\boldsymbol{\Psi}\boldsymbol{\Psi}') \otimes (\mathbf{X}'\mathbf{X})\}\widehat{\boldsymbol{\phi}}_M + (\mathbf{M}^{(\gamma)})^{-1}\boldsymbol{\phi}_0^{(\gamma)};$$

$$\boldsymbol{\Delta} = \{(\boldsymbol{\Psi}\boldsymbol{\Psi}') \otimes (\mathbf{X}'\mathbf{X}) + (\mathbf{M}^{(\gamma)})^{-1}\}^{-1}$$

5. draw $[\boldsymbol{\gamma}_{(k)} \mid \boldsymbol{\phi}_{(k)}, \boldsymbol{\psi}_{(k)}, \boldsymbol{\eta}_{(k)}, \boldsymbol{\omega}_{(k)}, \mathbf{Y}]$ from Bernoulli distribution (20).

4 Evaluating Stochastic Search Results

The results of stochastic search for the VAR model can be analyzed in a variety of ways. First the as is in general, stochastic search uncovers a set of most frequently visited models that good candidates for further analysis. For the most frequently visited models, we can calculate the posterior odds ratios. As we noted earlier, since an identifying scheme of VAR can be viewed as testing whether a particular set of elements in $\boldsymbol{\Psi}$ are zero, our odds ratio result will provide a test of one identifying scheme over an alternative. The hypothesis tests on $\boldsymbol{\Psi}$ may be done with all elements of VAR regression coefficients $\boldsymbol{\Phi}$ included. As for hypothesis testing concerning $\boldsymbol{\Phi}$, in current practise hypothesis testing is addressed by F tests. With all elements in $\boldsymbol{\Psi}$ included, odds

ratio tests for frequently visited models offers a promising Bayesian alternative that enjoys a very important advantage over the traditional F test.

The second way of using the results of the stochastic search is for improving the VAR forecasts over the benchmark model that includes all elements of Φ and Ψ . For forecast, we propose two possibilities. One is select a single promising model based on the search result and use it in place of the benchmark model. Another is to use the average of the forecast of all models visited in the MCMC. In terms of the evaluation of the competing forecast, we suggest the following criteria. We use frequentist MSE to value point forecast and the Kullback-Liebler divergence to evaluate the predictive density.

4.1 Improvement of Frequentist MSE of Forecast

One of the frequentist criteria for evaluating the estimators in numerical examples is the forecasting error attributable to the deviation of estimates from the true parameters. The j -step-ahead forecasting at period can be decomposed into two orthogonal parts:

$$\mathbf{y}_{T+j} - \hat{\mathbf{y}}_{T+j}|\hat{\Phi} = (\mathbf{y}_{T+j} - \hat{\mathbf{y}}_{T+j}|\Phi) + (\hat{\mathbf{y}}_{T+j}|\Phi - \hat{\mathbf{y}}_{T+j}|\hat{\Phi}),$$

where $\hat{\mathbf{y}}_{T+j}|\Phi$ and $\hat{\mathbf{y}}_{T+j}|\hat{\Phi}$ are the forecasts conditional on observations up to period T . They can be calculated from the VAR by setting the error term to zero after period T .

The first term in the right-hand-side is the sampling error, the second term is the forecasting error attributable to the deviation of estimates from the true parameters. Since the true parameters are known for numerical examples, the second term can be calculated for alternative estimators, and the mean squared errors (MSE) of the second term can be compared.

The frequentist average of the one-step-ahead MSEF for N samples is

$$\hat{\mathbb{E}}(\Phi - \hat{\Phi})' \mathbf{x}'_T \mathbf{x}_T (\Phi - \hat{\Phi}) = \frac{1}{N} \sum_{n=1}^N (\Phi - \hat{\Phi}^{(n)})' \mathbf{x}_T^{(n)'} \mathbf{x}_T^{(n)} (\Phi - \hat{\Phi}^{(n)}).$$

With model selection on matrix Ψ only, the regression coefficient matrix Φ is unrestricted. The informative prior applied to Φ help to reduce forecast errors if the the prior reflect the nature of data generating parameters. With model selection on both Φ and Ψ , the additional restrictions on Φ may further improve the forecast if the “correct” model is visited frequently in the MCMC simulation.

4.2 Predictive Density Through Model Averaging

In applications of VAR models, the predictive density maybe of interests besides the point estimate of the forecast. A wealth of literature in forecasting shows that under the model uncertainty, a well-advised approach for predictive density is averaging predictive density over the model space. Consider the VAR in the matrix form

$$\mathbf{Y} = \mathbf{X}\Phi + \epsilon.$$

Suppose one needs to predict a future observation \mathbf{Z} with density $g(\mathbf{Z} | \mathbf{Y}, \Phi, \Sigma)$ based on historical data \mathbf{Y} . Here \mathbf{Z} is an $m \times p$ matrix and depends on \mathbf{Y} . (Σ_k, Φ_k) represent the parameters under model M_k , where certain elements of the matrices are restricted. Define $\Phi'_k = (\mathbf{c}'_k, \mathbf{B}'_{1(k)}, \dots, \mathbf{B}'_{L(k)})$. Let \mathbf{z}'_j be the j th row of \mathbf{Z} . Under VAR model k , \mathbf{Y} follows model (4) and if $m \geq L$, then

$$\begin{aligned} \mathbf{z}_1 | (\Phi_k, \Sigma_k, \mathbf{Y}) &\sim N_p(\mathbf{c}'_k + \mathbf{B}'_{1(k)}\mathbf{y}_T + \mathbf{B}'_{2(k)}\mathbf{y}_{T-1} + \dots + \mathbf{B}'_{L(k)}\mathbf{y}_{T-L+1}, \Sigma_k); \\ \mathbf{z}_2 | (\mathbf{z}_1, \Phi_k, \Sigma_k, \mathbf{Y}) &\sim N_p(\mathbf{c}'_k + \mathbf{B}'_{1(k)}\mathbf{z}_1 + \mathbf{B}_2(\mathbf{k})'\mathbf{y}_T + \dots + \mathbf{B}'_{L(k)}\mathbf{y}_{T-L+2}, \Sigma_k); \\ &\dots, \\ \mathbf{z}_m | (\mathbf{z}_{m-1}, \dots, \mathbf{z}_1, \Phi_k, \Sigma_k, \mathbf{Y}) &\sim N_p(\mathbf{c}'_k + \mathbf{B}'_{1(k)}\mathbf{z}_{m-1} + \mathbf{B}'_{2(k)}\mathbf{z}_{m-2} + \dots + \mathbf{B}'_{L(k)}\mathbf{z}_{m-L}, \Sigma_k). \end{aligned}$$

The multi-step prediction of \mathbf{Z} conditional on the model, parameters, and data can be written in a hierarchal fashion. The case when $m < L$ can be derived analogously.

The predictive density $\pi(\mathbf{Z} | \Phi, \Sigma, \mathbf{Y}) = \pi(\mathbf{z}_m, \mathbf{z}_{m-1}, \dots, \mathbf{z}_1 | \Phi, \Sigma, \mathbf{Y})$ can be evaluated by using the fact that $\pi(\mathbf{z}_m, \mathbf{z}_{m-1}, \dots, \mathbf{z}_1 | \Phi, \Sigma, \mathbf{Y}) = \pi(\mathbf{z}_m, | \mathbf{z}_{m-1}, \dots, \mathbf{z}_{m-L}, \Phi, \Sigma, \mathbf{Y})\pi(\mathbf{z}_{m-1} | \mathbf{z}_{m-2}, \dots, \mathbf{z}_{m-L+1}, \Phi, \Sigma, \mathbf{Y}) \dots \pi(\mathbf{z}_{L+1} | \mathbf{z}_L, \dots, \mathbf{z}_1, \Phi, \Sigma, \mathbf{Y})\pi(\mathbf{z}_2 | \mathbf{z}_1, \Phi, \Sigma, \mathbf{Y})\pi(\mathbf{z}_1 | \Phi, \Sigma, \mathbf{Y})$.

For a set of models M_k ($k = 1, \dots, N$). The prediction density is

$$\begin{aligned} \pi(\mathbf{Z} | \mathbf{Y}) &= \sum_{k=1}^N \pi(\mathbf{Z} | M_k, \mathbf{Y})\pi(M_k | \mathbf{Y}) \\ &= \frac{\sum_{k=1}^N \int \pi(\mathbf{Z} | M_k, \mathbf{Y})\pi(\mathbf{Y} | M_k, \Sigma, \Phi)\pi(\Sigma, \Phi | M_k)\pi(M_k)d(\Sigma, \Phi)}{\sum_{j=1}^N \int \pi(\mathbf{Y} | M, \Sigma, \Phi)\pi(\Sigma, \Phi | Mj)\pi(Mj)d(\Sigma, \Phi)} \end{aligned}$$

The George-McCulloch stochastic search algorithm produces posterior conditional on model index M_k . The integration for marginal likelihood is costly. Given the number of candidate models the evaluation may not be feasible to study the entire model space. We want to only consider the likely candidate models. We use simulated models to compute the forecast density function

$$\pi(\mathbf{Z} | \mathbf{Y}) = \sum_{k \in I} \pi(\mathbf{Z} | M_k, \mathbf{Y})\hat{\pi}(M_k | \mathbf{Y})$$

$$= \sum_{k \in I} \int \pi(\mathbf{Z} | M_k, \boldsymbol{\Sigma}, \boldsymbol{\Phi}, \mathbf{Y}) \pi(\boldsymbol{\Sigma}, \boldsymbol{\Phi} | M_k, \mathbf{Y}) d(\boldsymbol{\Sigma}, \boldsymbol{\Phi}) \hat{\pi}(M_k | \mathbf{Y}),$$

where I is the set of models visited in MCMC, $\hat{\pi}(M_k | \mathbf{Y})$ is the frequency model M_k is visited in the MCMC. This is approximated by summing the predictive density of \mathbf{Z} over MCMC cycles $i = 1, \dots, A$ in the George-McCulloch algorithm, where posteriors of the models and posteriors of parameters conditional on the models are jointly simulated.

$$\hat{\pi}(\mathbf{Z} | \mathbf{Y}) = \frac{1}{A} \sum_i \pi(\mathbf{Z} | M_{i(k)}, \boldsymbol{\Sigma}_{i(k)}, \boldsymbol{\Phi}_{i(k)}, \mathbf{Y}),$$

where $M_{i(k)}$ ($k \in I$) represents the notion that model k is visited in the i th MCMC cycle and $(\boldsymbol{\Sigma}_{i(k)}, \boldsymbol{\Phi}_{i(k)})$ the parameter $(\boldsymbol{\Sigma}, \boldsymbol{\Phi})$ of model k in the i th MCMC cycle.

To evaluate the prediction of \mathbf{Z} , we modify Aitchison (1975)'s goodness-of-fit of prediction by the Kullback-Liebler divergence between $\pi(\mathbf{Z} | \mathbf{Y})$ and its estimate $s(\mathbf{Z} | \mathbf{Y})$, as the loss function

$$KL(\pi; s) = \int \log \left\{ \int \pi(\mathbf{Z} | \mathbf{Y}, M_k) \pi(M_k | \mathbf{Y}) dM_k / s(\mathbf{Z} | \mathbf{Y}) \right\} \pi(\mathbf{Z} | \mathbf{Y}) d\mathbf{Z}.$$

Madigan & Raftery (1994) show that the predictions through averaging is better than that of any single model in the candidate pool in terms of the entropy loss. If $s(\mathbf{Z} | \mathbf{Y}) = \pi(\mathbf{Z} | \mathbf{Y}, M_k)$ for any given k then $KL(\pi; s) > 0$.

The predictive density for a given model M_k is

$$\begin{aligned} \pi(\mathbf{Z} | \mathbf{Y}, M_k) &= \int \pi(\mathbf{Z} | M_k, \boldsymbol{\Sigma}, \boldsymbol{\Phi}, \mathbf{Y}) \pi(\boldsymbol{\Sigma}, \boldsymbol{\Phi} | M_k, \mathbf{Y}) d(\boldsymbol{\Sigma}, \boldsymbol{\Phi}) \\ &\approx \frac{1}{A} \sum_{j=1}^A \pi(\mathbf{Z} | M_k, \boldsymbol{\Sigma}_j, \boldsymbol{\Phi}_j, \mathbf{Y}), \end{aligned}$$

where $\boldsymbol{\Sigma}_j$ and $\boldsymbol{\Phi}_j$ ($j = 1, \dots, A$) are drawn from posterior conditional \mathbf{Y} and model k .

The K-L distance can be calculated by drawing \mathbf{Z}_n ($n = 1, \dots, D$) from $\pi(\mathbf{Z} | \mathbf{Y})$ and calculate

$$\frac{1}{D} \sum_{n=1}^D \log \frac{\hat{\pi}(\mathbf{Z}_n | \mathbf{Y})}{\hat{\pi}(\mathbf{Z}_n | M_k, \mathbf{Y})} = \frac{1}{N} \sum_{n=1}^D \log \frac{\sum_{i=1}^A \pi(\mathbf{Z}_n | M_{i(k)}, \boldsymbol{\Sigma}_{i(k)}, \boldsymbol{\Phi}_{i(k)}, \mathbf{Y})}{\sum_{j=1}^A \pi(\mathbf{Z}_n | M_k, \boldsymbol{\Sigma}_j, \boldsymbol{\Phi}_j, \mathbf{Y})}. \quad (49)$$

Here $\hat{\pi}(\mathbf{Z}_n | \mathbf{Y})$ and $\hat{\pi}(\mathbf{Z}_n | M_k, \mathbf{Y})$ are approximations of predictive densities.

The algorithm for computing the K-L loss of predictive density is as follows.

(1) Draw prediction $\mathbf{Z}_n \sim \pi(\mathbf{Z}_n | M_{i(k)}, \boldsymbol{\Sigma}_{i(k)}, \boldsymbol{\Phi}_{i(k)}, \mathbf{Y})$ in stochastic search MCMC ($n = 1, \dots, D$). Save \mathbf{Z}_n , ($n = 1, \dots, D$). Continue running the MCMC. For each draw of prediction \mathbf{Z}_n calculate $\sum_{i=1}^A \pi(\mathbf{Z}_n | M_{i(k)}, \boldsymbol{\Sigma}_{i(k)}, \boldsymbol{\Phi}_{i(k)}, \mathbf{Y})$ ($i = 1, \dots, A$).

(2) For a given model M_k , simulate posterior of $(\boldsymbol{\Sigma}_j, \boldsymbol{\Phi}_j) | M_k, \mathbf{Y}$, $j = 1, \dots, N$. For each \mathbf{Z}_n compute the value of $\sum_{j=1}^A \pi(\mathbf{Z}_n | \boldsymbol{\Sigma}_j, \boldsymbol{\Phi}_j, M_k, \mathbf{Y})$. Then compute K-L distance via (49).

Regardless of the models, the normality of the conditional posterior allows for easy evaluation at a predictive value \mathbf{Z}_n of the densities in numerator (averaged across models) and denominator (with a fixed model) of (49).

5 Numerical Examples

We use a couple of numerical examples to illustrate the performance of the stochastic search algorithm. For each example parameter setting, we simulate 100 samples and for each data sample conduct stochastic search model selection with Markov Chain of 50000 cycles (after 10000 burn-in runs). Simulation results are little changed when the Markov Chain length is reduced to 20000 suggesting that the Markov chains converge rather quickly.

The performance of stochastic search is evaluated in two ways, the effectiveness in finding the true data generating model and the performance in forecast. We now discuss each evaluation in detail.

The model selection via stochastic search is conducted in two-steps. First, for each data sample we first run the MCMC algorithm laid out earlier in the paper. We report frequentist averages of the posterior mean of the model index matrix $\boldsymbol{\gamma}$ and $\boldsymbol{\omega}$, as well as those of parameters $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$. Because the VAR parameters $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ are jointly simulated with model selection indexes $\boldsymbol{\gamma}$ and $\boldsymbol{\omega}$, the posterior mean of $\boldsymbol{\Phi}$ is the average of $\boldsymbol{\Phi}$ over different models.

Second, we offer a parsimonious method in choosing a model as an average of stochastic search. For empirical researchers, a main aim of model stochastic search is the selection of most probable models. For a given sample, one may document the most frequently visited models in MCMC simulations and calculate posterior odds of the frequently visited models. To avoid the cost of computing marginal likelihood for the posterior odds, we propose the following simplifying procedure for selecting the restricted model based on MCMC simulations. We simply delete (i.e., always impose a tight prior centered at 0 on) the elements in $\boldsymbol{\gamma}$ or $\boldsymbol{\omega}$ if the corresponding posterior mean of the first round search is below a certain threshold (say 0.25). There is only one restricted model for each sample. We then simulate the restricted model and report the posterior mean of parameters of interest as well as the performance of forecast. The frequentist average of the restrictions and other quantities of interest are then reported.

Four estimates are compared. (1) MLE; (2) estimates of benchmark Bayesian VAR where all

parameters in Φ and Ψ are included (with diffuse prior); (3) estimates of Bayesian VAR under stochastic search (step-one in model selection); (4) estimates of selected Bayesian VAR with restrictions obtained based on the search result (step-two).

Our second criterion is predictive performance. Following the discussion in Section 4, we report the relative performance of frequentist average of MSE of the four estimates listed above. We also report the Kullback-Liebler divergence between the predictive densities of the benchmark Bayesian VAR with all parameters of Φ and Ψ included and that of the Bayesian VAR under stochastic search.

Example 1 Consider a six-variable VAR with one lag and with parameters

$$\Phi = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \text{ and } \Psi = \begin{pmatrix} 1 & .5 & .5 & .5 & .5 & .5 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (50)$$

The sample size $T = 50$. We fix a subset of hyper-parameters of the priors. $p_i = .5$, $q_{ij} = 0.5$, $a = b = .01$. The intercept term of every equation is assumed to be always included. The prior on the intercepts is normal with zero mean and a large variance of 25.0. With these hyper-parameters fixed, we compare results under various values of hyper-parameters τ_i , κ_{ij} , c_i , and d_{ij} in the following: $\tau_i = 0.1$, $\kappa_{ij} = 0.1$, $c_i = 50.0$, $d_{ij} = 50.0$. The Minnesota prior (which implies random walk of each variable) is applied to Φ with the covariance matrix of the prior being diagonal and with variance of 25.0.

The frequentist average of MLEs of Ψ and Φ are

$$\hat{\Phi}_{MLE} = \begin{pmatrix} 1.283 & 0.762 & 0.739 & 1.037 & 0.917 & 0.804 \\ 0.795 & 0.118 & 0.126 & 0.124 & 0.129 & 0.131 \\ 0.039 & 0.745 & 0.024 & 0.049 & 0.014 & 0.053 \\ 0.042 & 0.042 & 0.735 & 0.038 & 0.038 & 0.019 \\ 0.023 & 0.032 & 0.044 & 0.735 & 0.043 & 0.058 \\ 0.060 & 0.016 & 0.027 & 0.020 & 0.739 & 0.022 \\ 0.035 & 0.045 & 0.042 & 0.039 & 0.034 & 0.719 \end{pmatrix}.$$

$$\hat{\Psi}_{MLE} = \begin{pmatrix} 1.848 & 0.379 & 0.380 & 0.388 & 0.389 & 0.380 \\ 0.000 & 1.195 & -0.158 & -0.149 & -0.132 & -0.179 \\ 0.000 & 0.000 & 1.163 & -0.184 & -0.157 & -0.177 \\ 0.000 & 0.000 & 0.000 & 1.150 & -0.176 & -0.218 \\ 0.000 & 0.000 & 0.000 & 0.000 & 1.101 & -0.242 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.050 \end{pmatrix}.$$

To set a benchmark for model comparison, we first simulate the posterior without excluding any elements in Φ and Ψ . Given the fact that the prior on Φ is quite diffuse, with all elements of Ψ and Φ included the posteriors are determined by the sample information.

$$\hat{\Phi}_{Mean} = \begin{pmatrix} 1.317 & 0.715 & 0.691 & 0.968 & 0.855 & 0.743 \\ 0.800 & 0.114 & 0.123 & 0.121 & 0.126 & 0.128 \\ 0.039 & 0.746 & 0.023 & 0.048 & 0.013 & 0.052 \\ 0.041 & 0.042 & 0.739 & 0.039 & 0.038 & 0.019 \\ 0.021 & 0.034 & 0.044 & 0.739 & 0.044 & 0.060 \\ 0.059 & 0.016 & 0.026 & 0.021 & 0.742 & 0.022 \\ 0.034 & 0.045 & 0.043 & 0.038 & 0.034 & 0.722 \end{pmatrix}.$$

Although the Minnesota prior on Φ shrinks the posterior of the parameter matrix to the mean, such shrinkage is limited due to the large variance of the prior. Comparison of the frequentist average of the posterior mean of Φ for the benchmark VAR with that of the MLE shows that they are quite similar. Furthermore, we will report that the forecasting performance of the benchmark Bayes estimator (the posterior mean) is similar to that of the MLE.

5.1 Joint Model Selection on Φ and Ψ

In the following, we report results for joint model selection for Ψ and Φ . The frequentist averages of the model selection indices through stochastic search are

$$\hat{\omega} = \begin{pmatrix} 1.000 & 0.575 & 0.551 & 0.531 & 0.520 & 0.502 \\ 0.000 & 1.000 & 0.079 & 0.081 & 0.074 & 0.078 \\ 0.000 & 0.000 & 1.000 & 0.069 & 0.083 & 0.082 \\ 0.000 & 0.000 & 0.000 & 1.000 & 0.079 & 0.096 \\ 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.085 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 \end{pmatrix}.$$

and

$$\hat{\gamma} = \begin{pmatrix} 1.000 & 1.000 & 1.000 & 1.000 & 1.000 & 1.000 \\ 1.000 & 0.043 & 0.051 & 0.046 & 0.047 & 0.047 \\ 0.036 & 0.999 & 0.039 & 0.042 & 0.042 & 0.044 \\ 0.036 & 0.047 & 0.999 & 0.045 & 0.040 & 0.043 \\ 0.037 & 0.040 & 0.041 & 0.999 & 0.042 & 0.045 \\ 0.036 & 0.042 & 0.039 & 0.041 & 0.999 & 0.040 \\ 0.038 & 0.044 & 0.040 & 0.041 & 0.043 & 0.998 \end{pmatrix}.$$

The restricted VAR model is obtained by imposing tight priors with zero mean on elements of Φ and Ψ that have model indexes smaller than 0.25. We report the frequency at which the index matrixes γ and ω are above the cutoff level, i.e., the corresponding elements in Φ and Ψ are not restricted by the tight zero-mean prior.

$$\begin{pmatrix} 1.000 & 1.000 & 1.000 & 1.000 & 1.000 & 1.000 \\ 1.000 & 0.000 & 0.010 & 0.010 & 0.040 & 0.010 \\ 0.000 & 1.000 & 0.010 & 0.000 & 0.010 & 0.030 \\ 0.000 & 0.030 & 1.000 & 0.020 & 0.000 & 0.020 \\ 0.000 & 0.000 & 0.000 & 1.000 & 0.020 & 0.010 \\ 0.010 & 0.010 & 0.010 & 0.000 & 1.000 & 0.000 \\ 0.000 & 0.010 & 0.010 & 0.020 & 0.010 & 1.000 \end{pmatrix}.$$

Clearly, the MCMC algorithm is very effective in picking the correct restrictions for Φ . The frequentist average of the restrictions on Ψ is

$$\begin{pmatrix} 1.000 & 0.770 & 0.790 & 0.730 & 0.690 & 0.730 \\ * & 1.000 & 0.050 & 0.050 & 0.040 & 0.030 \\ * & * & 1.000 & 0.020 & 0.030 & 0.060 \\ * & * & * & 1.000 & 0.030 & 0.070 \\ * & * & * & * & 1.000 & 0.060 \\ * & * & * & * & * & 1.000 \end{pmatrix}.$$

The frequentist averages of the posterior mean of Ψ and Φ through the stochastic search are quite similar to those under the restricted models. So we only report the latter:

$$\begin{aligned} \hat{\Psi} &= \begin{pmatrix} 1.088 & 0.499 & 0.490 & 0.495 & 0.467 & 0.480 \\ 0.000 & 1.099 & -0.029 & -0.024 & -0.021 & -0.029 \\ 0.000 & 0.000 & 1.101 & -0.022 & -0.023 & -0.033 \\ 0.000 & 0.000 & 0.000 & 1.131 & -0.024 & -0.032 \\ 0.000 & 0.000 & 0.000 & 0.000 & 1.111 & -0.028 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.125 \end{pmatrix}, \\ \hat{\Phi} &= \begin{pmatrix} 1.106 & 0.925 & 0.937 & 1.135 & 1.097 & 0.997 \\ 0.880 & 0.051 & 0.062 & 0.056 & 0.064 & 0.059 \\ 0.025 & 0.807 & 0.027 & 0.035 & 0.021 & 0.043 \\ 0.022 & 0.040 & 0.810 & 0.041 & 0.034 & 0.030 \\ 0.019 & 0.031 & 0.030 & 0.799 & 0.034 & 0.041 \\ 0.031 & 0.029 & 0.030 & 0.031 & 0.809 & 0.030 \\ 0.022 & 0.036 & 0.036 & 0.034 & 0.032 & 0.792 \end{pmatrix}. \end{aligned}$$

The averages of the indices γ and ω in matrix form illustrate the ability of the selecting the right model. A perfect match of every elements in γ with Ψ and ω with Φ would indicate successful selection of the true model. For the Ψ and Φ matrices, the average of joint perfect match is about 0.6%. It may appear to be small but is quite reasonable if we recognize the fact that the number of parameters subject to test (with the intercepts always included) is $6 \times 6 + 6(6 - 1)/2 = 51$. So there are 2^{51} candidate models. The chance that the right model is selected is quite small. The table indicates that for the given example of data-generating model, picking correct Ψ is more difficult

than picking correct Φ . As shown above, the MLE of Ψ is quite biased. As a result, the posteriors of some of the off-diagonal elements of Ψ are not centered at the true values. When we impose no restrictions on the regression coefficients Φ , the chance that the correct Ψ is picked out of 2^{15} models is 1.7%, which is quite large. With all elements of Ψ included, the stochastic search picks the correct model for Φ out of 2^{36} candidate models 30.9% of the time, which is remarkable.

Table 1. The frequency of picking the correct model in the MCMC

<i>Hyper – parameter set</i>	<i>selectΦ</i>	<i>selectΨ</i>	<i>SelectΦ&Ψ</i>
	30.9%	1.7%	0.6%

In the following, we examine improvement of MSE of one-step forecast by using the following posterior mean of MCMC as estimator for Φ in a number of cases. (1) We estimate Ψ and Φ without going through model selection, i.e., all elements are included by setting γ and the upper-triangular ω to unity. If the priors on Ψ and Φ are diffuse then the difference between the posterior mean of Φ and the MLE should be small. Consequently, one should not expect much improvement in forecasting errors by the Bayesian estimates over the MLE. (2) We also report the improvement of posterior mean of Φ with stochastic search on Ψ and Φ . The resultant estimate of Φ for a given sample is the average over all models visited in MCMC. If the correct model or models very close to the correct models are frequently visited, then the stochastic search should improve over the MLE. (3) For each sample, we select a restricted model based on the posterior mean of the model selection index. As discussed earlier, we delete elements in Ψ and Φ that have model selection indexes less than a threshold value (here it is set at 0.25). We use the posterior mean of Φ of the restricted model as the estimator and report the corresponding MSE of forecast. The difference between estimates in (2) and (3) is that (3) concerns a single model, if the restrictions are not correctly placed then the estimate may behave less well than (2), the opposite is true if the restrictions are more correctly imposed than most of the models in (2).

Table 2. Improvement of forecast accuracy over MLE over 100 samples

<i>Hyper – parameter set</i>	<i>i = 1</i>	<i>i = 2</i>	<i>i = 3</i>	<i>i = 4</i>	<i>i = 5</i>	<i>i = 6</i>
(1)	8.5%	3.7%	5.2%	2.9%	6.3%	6.6%
(2)	65.5%	53.5%	61.5%	46.5%	67.3%	62.4%
(3)	64.5%	52.8%	59.7%	46.1%	67.8%	62.4%

(1)All elements of Ψ and Φ are included. (2)Frequentist Average of posterior mean in stochastic search. (3)Frequentist Average of restricted models.

The table shows that without going through model selection, there is little improvement by the Bayesian estimates over the MLE. This result is expected given the diffuse nature of the prior employed on Φ . Overall the 100 samples, the posterior mean of average over all visited models and that of the restricted models make comparable improvement over the MLE, suggesting that the stochastic search often results in picking the correct model. This is hardly surprising, given the earlier reported posterior mean of the model selection index γ and the frequentist average of the restrictions reflect very closely the true data-generating model.

Our second measure of predictive performance of stochastic search, the K-L divergence of the predictive densities over predictive performance by model averaging is reported in the following table.

Table 3. K-L distance of the predictive densities over 100 samples

<i>mean</i>	<i>std.</i>	<i>min</i>	<i>max</i>
44.358	10.655	20.152	71.638

The large K-L divergence shows that predictive density from model averaging is substantially different from that of benchmark VAR model that does not go through model selection. The restrictions implied in model selection overcomes over fitting and results in better predictive performance.

5.2 Sensitivity to Priors

At this point, some discussion on the hyper-parameters is in order. We conducted investigation on how the success rate of model selection depends on the hyper-parameters of the priors, c , d , κ , and τ ; as well as on the true parameters in Ψ and Φ . In general, the smaller κ and τ make the model selection indices more sensitive to the corresponding parameter values. The hyper-parameters c and d affect the model selection in a more complicated way.

The reported results are based on the setting hyper-parameters $c=d=50$. If these parameters are reduced, for example to $c=d=10$, then the distinction between the excluded and included variables is not as large. Consequently, the chance that the correct model being picked is reduced. On the other hand, the reduced prior variance may lead to better inference if the prior correctly reflect the nature of the data-generating parameters. The researchers must consider the trade-off in prior elicitation.

Regarding the cutoff threshold, the value of 0.25 works almost perfectly for selection of Φ but

is too large for Ψ . If the cutoff restriction of 0 elements for Ψ is reduced to 0.15, the restrictions are much closer to the true model:

$$\begin{pmatrix} 1.000 & 0.910 & 0.860 & 0.860 & 0.810 & 0.730 \\ * & 1.000 & 0.100 & 0.050 & 0.100 & 0.070 \\ * & * & 1.000 & 0.070 & 0.080 & 0.090 \\ * & * & * & 1.000 & 0.080 & 0.090 \\ * & * & * & * & 1.000 & 0.110 \\ * & * & * & * & * & 1.000 \end{pmatrix}.$$

There is not much improvement to MSE of predictions since the latter pertains to Φ only. In the following we give detailed comparison of the joint model selection for Ψ and Φ with separate searches of the parameters.

5.3 Model Selection on Φ or Ψ Only

We now examine model selection separately with respect to Ψ and Φ . The question we seek to answer is whether selection of Ψ influences the selection of Φ and visa verse. We do so by comparing the posterior of model selection indices and parameters of Φ (Ψ) with all elements of Ψ (Φ) included, then compare the results with those from jointed selection reported earlier.

The frequentist average of the posterior mean of model selection index γ from the first step MCMC is almost the same as that with the joint selection of Ψ and Φ .

$$\hat{\gamma} = \begin{pmatrix} 1.000 & 1.000 & 1.000 & 1.000 & 1.000 & 1.000 \\ 1.000 & 0.041 & 0.054 & 0.045 & 0.051 & 0.049 \\ 0.033 & 0.999 & 0.042 & 0.043 & 0.043 & 0.050 \\ 0.033 & 0.063 & 0.998 & 0.053 & 0.042 & 0.050 \\ 0.033 & 0.042 & 0.044 & 0.999 & 0.049 & 0.049 \\ 0.035 & 0.042 & 0.043 & 0.045 & 0.999 & 0.041 \\ 0.032 & 0.048 & 0.043 & 0.044 & 0.042 & 0.998 \end{pmatrix}.$$

For each sample, we take a second step and eliminate the elements in Φ that has the posterior mean smaller than 0.25. The elements not eliminated is represented by 1. The frequentist average of these restrictions is

$$\begin{pmatrix} 1.000 & 1.000 & 1.000 & 1.000 & 1.000 & 1.000 \\ 1.000 & 0.000 & 0.010 & 0.010 & 0.040 & 0.010 \\ 0.000 & 1.000 & 0.010 & 0.000 & 0.010 & 0.030 \\ 0.010 & 0.040 & 1.000 & 0.030 & 0.000 & 0.030 \\ 0.000 & 0.000 & 0.000 & 1.000 & 0.020 & 0.010 \\ 0.010 & 0.010 & 0.010 & 0.000 & 1.000 & 0.000 \\ 0.000 & 0.010 & 0.010 & 0.020 & 0.000 & 1.000 \end{pmatrix}.$$

It is clear that on average the restrictions are very close to the true model. Under the restrictions, the frequentist average of the posterior mean of Ψ and Φ for the restricted models are

$$\hat{\Psi} = \begin{pmatrix} 1.087 & 0.536 & 0.517 & 0.515 & 0.522 & 0.496 \\ 0.000 & 1.104 & -0.050 & -0.030 & -0.028 & -0.056 \\ 0.000 & 0.000 & 1.112 & -0.047 & -0.020 & -0.036 \\ 0.000 & 0.000 & 0.000 & 1.150 & -0.025 & -0.056 \\ 0.000 & 0.000 & 0.000 & 0.000 & 1.141 & -0.044 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.164 \end{pmatrix},$$

$$\hat{\Phi} = \begin{pmatrix} 1.095 & 0.933 & 0.948 & 1.140 & 1.059 & 0.996 \\ 0.882 & 0.050 & 0.061 & 0.055 & 0.062 & 0.057 \\ 0.025 & 0.801 & 0.029 & 0.036 & 0.022 & 0.044 \\ 0.022 & 0.043 & 0.806 & 0.044 & 0.036 & 0.032 \\ 0.019 & 0.032 & 0.032 & 0.796 & 0.035 & 0.042 \\ 0.029 & 0.029 & 0.032 & 0.031 & 0.808 & 0.031 \\ 0.022 & 0.039 & 0.037 & 0.035 & 0.031 & 0.790 \end{pmatrix}.$$

The averages of the indices γ in matrix form illustrate the ability of the selecting the right model. A perfect match of every elements in γ with Φ would indicate selection of the exact true model. The average rate of perfect match is over thirty percent, even when the Ψ matrix is unrestricted.

The data-generating parameters are unchanged. However, in this example, Φ matrix is assumed be unrestricted. Model selection concerns only the Ψ matrix. Under the hyper-parameters $\tau_i = 0.1$, $\kappa_{ij} = 0.1$, $c_i = 50.0$, $d_{ij} = 50.0$ the frequentist average of model selection index for Ψ is

$$\hat{\omega} = \begin{pmatrix} 1.000 & 0.548 & 0.520 & 0.510 & 0.492 & 0.474 \\ 0.000 & 1.000 & 0.086 & 0.081 & 0.076 & 0.086 \\ 0.000 & 0.000 & 1.000 & 0.071 & 0.083 & 0.088 \\ 0.000 & 0.000 & 0.000 & 1.000 & 0.079 & 0.104 \\ 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.086 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 \end{pmatrix}.$$

Now, for each sample we take a second step and exclude the upper-triangular elements in Ψ that has the posterior mean smaller than 0.25. The elements not excluded is represented by 1. The frequentist average of these restrictions is

$$\begin{pmatrix} 1.000 & 0.720 & 0.720 & 0.680 & 0.660 & 0.660 \\ * & 1.000 & 0.060 & 0.050 & 0.040 & 0.040 \\ * & * & 1.000 & 0.010 & 0.040 & 0.050 \\ * & * & * & 1.000 & 0.030 & 0.080 \\ * & * & * & * & 1.000 & 0.080 \\ * & * & * & * & * & 1.000 \end{pmatrix}.$$

On average the restrictions are not as close to the true model as for the regression coefficients.

Under the restrictions, the frequentist averages of the posterior mean of Ψ and Φ are

$$\hat{\Psi} = \begin{pmatrix} 1.099 & 0.474 & 0.458 & 0.475 & 0.447 & 0.440 \\ 0.000 & 1.095 & -0.035 & -0.024 & -0.020 & -0.036 \\ 0.000 & 0.000 & 1.096 & -0.021 & -0.026 & -0.032 \\ 0.000 & 0.000 & 0.000 & 1.128 & -0.024 & -0.039 \\ 0.000 & 0.000 & 0.000 & 0.000 & 1.109 & -0.038 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.123 \end{pmatrix},$$

$$\hat{\Phi} = \begin{pmatrix} 1.310 & 0.714 & 0.692 & 0.971 & 0.865 & 0.746 \\ 0.799 & 0.115 & 0.123 & 0.121 & 0.126 & 0.128 \\ 0.039 & 0.747 & 0.023 & 0.048 & 0.013 & 0.052 \\ 0.041 & 0.042 & 0.739 & 0.039 & 0.038 & 0.020 \\ 0.021 & 0.034 & 0.044 & 0.739 & 0.044 & 0.059 \\ 0.059 & 0.016 & 0.026 & 0.021 & 0.742 & 0.022 \\ 0.034 & 0.045 & 0.043 & 0.038 & 0.034 & 0.722 \end{pmatrix}.$$

The frequentist average for Φ is close to that of the MLE. On average, the model selection index ω under search for Ψ alone is not as good as that under the joint search. This indicates that restricting the regression coefficients helps in selecting the correct model for Ψ .

Example 2

In the first example, we find that when Φ are correctly selected the stochastic search improves the predictive performance over the benchmark model where no model selection is conducted. In this example we show that even when model selection is not very effective in selecting the Φ matrix stochastic search still yields substantial improvement in forecasting.

Consider a four-variable VAR with two lags and with parameters

$$\Phi = \begin{pmatrix} 1 & 0.5 & 0.5 & 0.5 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \text{ and } \Psi = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0.45 & 0 & 0 & 0 \\ 0 & 0.50 & 0 & 0 \\ 0 & 0 & 0.55 & 0 \\ 0 & 0 & 0 & 0.60 \\ 0.40 & 0 & 0 & 0 \\ 0 & 0.40 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.40 & 0 \\ 0 & 0 & 0 & 0.40 \end{pmatrix}. \quad (51)$$

The frequentist average of MLEs of Ψ and Φ are

$$\hat{\Phi}_{MLE} = \begin{pmatrix} 2.319 & 2.034 & 1.718 & -0.246 \\ 0.313 & 0.016 & 0.042 & 0.043 \\ 0.011 & 0.360 & 0.019 & -0.018 \\ -0.032 & 0.028 & 0.441 & 0.075 \\ -0.032 & 0.012 & 0.049 & 0.541 \\ 0.273 & 0.041 & 0.034 & 0.122 \\ 0.018 & 0.251 & 0.013 & 0.019 \\ 0.031 & 0.005 & 0.259 & 0.045 \\ 0.039 & 0.009 & 0.025 & 0.401 \end{pmatrix}.$$

$$\hat{\Psi}_{MLE} = \begin{pmatrix} 1.561 & 0.425 & 0.455 & 0.447 \\ 0 & 1.091 & -0.217 & -0.173 \\ 0 & 0 & 1.072 & -0.247 \\ 0 & 0 & 0 & 1.022 \end{pmatrix}.$$

We use the same set of priors as in the previous example. $\tau_i = 0.1$, $\kappa_{ij} = 0.1$, $c_i = 50.0$, $d_{ij} = 50.0$. The frequentist average of the model selection indices are

$$\hat{\gamma} = \begin{pmatrix} 1.000 & 1.000 & 1.000 & 1.000 \\ 0.871 & 0.056 & 0.064 & 0.058 \\ 0.047 & 0.874 & 0.050 & 0.049 \\ 0.041 & 0.049 & 0.930 & 0.056 \\ 0.030 & 0.031 & 0.036 & 0.965 \\ 0.148 & 0.051 & 0.055 & 0.067 \\ 0.045 & 0.130 & 0.047 & 0.047 \\ 0.038 & 0.051 & 0.163 & 0.053 \\ 0.031 & 0.031 & 0.036 & 0.416 \end{pmatrix},$$

$$\hat{\omega} = \begin{pmatrix} 1.000 & 0.533 & 0.556 & 0.490 \\ 0.000 & 1.000 & 0.078 & 0.062 \\ 0.000 & 0.000 & 1.000 & 0.071 \\ 0.000 & 0.000 & 0.000 & 1.000 \end{pmatrix}.$$

The average of restrictions on Φ and Ψ are

$$\begin{pmatrix} 1.000 & 1.000 & 1.000 & 1.000 \\ 1.000 & 0.020 & 0.040 & 0.030 \\ 0.020 & 1.000 & 0.030 & 0.020 \\ 0.010 & 0.010 & 1.000 & 0.040 \\ 0.000 & 0.000 & 0.000 & 1.000 \\ 0.340 & 0.020 & 0.020 & 0.050 \\ 0.030 & 0.240 & 0.020 & 0.030 \\ 0.010 & 0.030 & 0.430 & 0.020 \\ 0.000 & 0.000 & 0.000 & 0.780 \end{pmatrix}, \begin{pmatrix} 1.000 & 0.800 & 0.850 & 0.840 \\ * & 1.000 & 0.090 & 0.050 \\ * & * & 1.000 & 0.050 \\ * & * & * & 1.000 \end{pmatrix}.$$

The restrictions on Ψ reflect the nature of the true data-generating model quite accurately, while those of Φ incorrectly exclude the lag 2 coefficients at a high frequency. Consequently the frequency that the exactly correct Φ matrix being selected is lower than the previous example.

Table 4. The frequency of picking the correct model

Hyper – parameterset	select Φ	select Ψ	Select Φ & Ψ
$\tau_i = 0.1, \kappa_{ij} = 0.1, c_i = 50.0, d_{ij} = 50.0$.02%	14.1%	0%

The frequentist averages of the posterior mean of Ψ and Φ are

$$\hat{\Psi} = \begin{pmatrix} 1.034 & 0.476 & 0.513 & 0.480 \\ 0.000 & 1.041 & -0.022 & -0.021 \\ 0.000 & 0.000 & 1.076 & -0.028 \\ 0.000 & 0.000 & 0.000 & 1.064 \end{pmatrix},$$

$$\hat{\Phi} = \begin{pmatrix} 2.159 & 2.201 & 1.872 & 0.414 \\ 0.411 & 0.015 & 0.018 & 0.025 \\ 0.006 & 0.444 & 0.021 & -0.006 \\ -0.001 & 0.014 & 0.497 & 0.042 \\ 0.000 & 0.012 & 0.036 & 0.588 \\ 0.198 & 0.012 & 0.022 & 0.059 \\ 0.012 & 0.161 & 0.009 & 0.002 \\ 0.010 & 0.015 & 0.212 & 0.022 \\ 0.004 & 0.015 & 0.036 & 0.380 \end{pmatrix}.$$

With $\tau_i = 0.1, \kappa_{ij} = 0.1, c_i = 50.0, d_{ij} = 50.0$, the improvement in terms of frequentist MSE by either model averaging or the selected model over the benchmark model or MLE is still substantial.

Table 5. Improvement of forecast accuracy over MLE: Select both Φ and Ψ

Hyper – parameterset	$i = 1$	$i = 2$	$i = 3$	$i = 4$
(1)	3.4%	5.3%	1.5%	4.1%
(2)	40.2%	47.8%	19.6%	44.5%
(3)	34.7%	37.1%	31.7%	43.8%

(1)All elements of Ψ and Φ are included. (2)Frequentist Average of posterior mean in stochastic search. (3)Frequentist Average of restricted models.

The next table shows that the Kullback-Liebler divergence between the predictive density through model averaging and that of the benchmark model is still quite large, but not as large as that in the previous example.

Table 6. K-L distance of the predictive densities over 100 samples

<i>mean</i>	<i>std.</i>	<i>min</i>	<i>max</i>
16.352	5.281	3.254	29.256

6 An Empirical Analysis on transmission of PPI to CPI Inflation

We now apply the stochastic search algorithm to an empirical inquiry. The issue at hand is how Producer Price Index (PPI) inflation is passed through different stages of production and to inflation of Consumer Price Index (CPI). Uncovering the dynamic pattern of inflation transmission will help to better predict and prepare for the consequences of an external shock in price of crude materials.

The dynamics of inflation pass-through is a complicated matter. Until now, the most common tool for examining the problem is unrestricted VAR. As we have shown through numerical examples, restricted VAR models through stochastic search provide sharper description of the data generating model and can substantially improve the forecast over the unrestricted VARs.

There are several reasons for PPI inflation of crude materials to lead CPI inflation. Blomberg and Harris (1995) argued that the commodity market for crude materials are competitive where prices quickly respond to new information on future demand and supply. In contrast, the prices of final goods and consumer prices respond more slowly due to market rigidities such as menu costs or long-term contractual arrangements. The existence of commodity market makes it possible to for investors to hedge against inflation, bidding up commodity prices when investors anticipate a rise in inflation. But the pass-through of crude material price inflation to CPI inflation is not definitive for a number of reasons.

First, CPI and PPI differ in construction. CPI includes services, imports, distribution costs, and sales taxes whereas the PPI excludes them; the PPI includes capital equipment while the CPI does not. Second, shifts in relative demand of commodities and final goods and tightening of monetary policy in response to commodity price inflation may weaken the link between the commodity price inflation and CPI inflation. Even the inflation pass-through from crude materials to CPI does occur, it may be a lengthy process. When input price increases, firms and industries may adjust prices after short delays, but these short lags cumulate through a chain of production, resulting a longer delay in changes of aggregate output prices and consumer prices (see Matthey (1981)). In addition to the cumulative delays, there are also complications in whether and how price change in a particular component of the PPI is passed to a similar component of the CPI. There exist

possibilities that due to shocks in the costs to transportation, the CPI for a given component may change even though the PPI for the same component does not, making the overall response of CPI to PPI inflation unclear. The theoretical ambiguities make the transmission of PPI to CPI inflation an interesting empirical question.

The existing empirical studies show that VAR estimates on the responses of CPI inflation to commodity price inflation are generally unstable over sample periods. Blomberg & Harris (1995) and Furlong & Ingenito (1996) found that in the 1970s and early 1980s inflation in commodity prices leads to CPI inflation, but for the more recent years it is no longer true. Clark (1995) showed whether inclusion of PPI inflation reduces predictive errors of CPI inflation depends on sample periods. Weinhagen (2002) studied price transmission through various stages of production. He found that for the sample period of 1974 to 1989, inflation of PPI of crude materials and intermediate goods are good predictors of CPI inflation. But for the period of 1990 to 2001 only inflation of PPI of finished goods predicts CPI inflation. These studies used benchmark VAR models without considering the model selection issue. They can not address questions such as “What’s the probability a shock in PPI inflation of crude material be passed to PPI of intermediate materials within the same month?” or “What’s the probability a shock in PPI inflation of finished consumer goods be passed to CPI in the next month, two months and six months?”

In the following, we estimate a five-variable VAR including the monthly PPI and CPI inflation of the U.S. economy. Our primary interest lies in the contemporaneous and lag transmission of PPI inflation in various stages of the supply chain. The variables are PPI foodstuff and feeds (FF), PPI crude materials for processing (CM), PPI intermediate materials and supplies (IM), PPI finished consumer goods (FC), and CPI for urban consumers (CPI). The sample period is from January 1969 to August 2003.

To allow for long delays in inflation transmission we use a VAR with lag length of 12. The total number of parameters in Ψ is 15 and that in Φ is 305. Given the sample observations of just over 400 periods, there is a potential for over fitting, rendering model selection particularly valuable. Given the large number of parameters, we choose the MCMC of 100,000 with 10,000 burn-in runs.

The posterior mean of model selection index matrix ω is

$$\hat{\omega} = \begin{pmatrix} & \text{FF} & \text{CM} & \text{IM} & \text{FC} & \text{CPI} \\ \text{FF} & 1.000 & 0.024 & 0.022 & 0.021 & 0.025 \\ \text{CM} & 0.000 & 1.000 & 0.022 & 0.023 & 0.019 \\ \text{IM} & 0.000 & 0.000 & 1.000 & 1.000 & 0.037 \\ \text{FC} & 0.000 & 0.000 & 0.000 & 1.000 & 0.032 \\ \text{CPI} & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 \end{pmatrix}.$$

Most of the upper-triangular elements of the index ω are close to zero, except for that from PPI of intermediate materials to finished goods. The elements close to zero suggest no contemporaneous transmission of inflation. The corresponding posterior mean of Ψ is

$$\hat{\Psi} = \begin{pmatrix} 0.504 & -0.056 & 0.040 & 0.019 & 0.059 \\ * & 0.348 & -0.045 & 0.048 & 0.003 \\ * & * & 2.469 & -1.457 & -0.030 \\ * & * & * & 2.762 & 0.009 \\ * & * & * & * & 6.063 \end{pmatrix}.$$

Consistent with the model selection result, most of the elements of Ψ are close to zero except for $\psi_{3,4}$. We denote $\epsilon = (\epsilon_{FF}, \epsilon_{CM}, \epsilon_{IM}, \epsilon_{FC}, \epsilon_{CPI})'$ to denote the VAR residuals (which can be interpreted as unexpected inflations), and $\mathbf{u} = (u_{FF}, u_{CM}, u_{IM}, u_{FC}, u_{CPI})'$ to denote the structural shock vector. The covariance of \mathbf{u} is the identity matrix. From (2), the point estimates of the Ψ matrix give restrictions on how unexpected inflation of different price indexes relate to one another. For instance, the fourth column of Ψ implies that $2.762\epsilon_{FC} = -.019\epsilon_{FF} - .048\epsilon_{CM} + 1.457\epsilon_{IM} + u_{FC}$, suggesting unexpected inflation in PPI of finished goods equals one half of that in PPI of intermediate materials and supplies plus a idiosyncratic shock with a standard deviation of about 0.36%. In comparison to the posterior mean estimator of Ψ , the MLE of Ψ shows much larger coefficients, which gives much different relationship among unexpected inflations as the Bayesian estimates do. For the fourth column of Ψ , the MLE implies $2.595\epsilon_{FC} = -.134\epsilon_{FF} - .482\epsilon_{CM} + 1.481\epsilon_{IM} + u_{FC}$.

$$\hat{\Psi}_{MLE} = \begin{pmatrix} 0.571 & -0.043 & 0.180 & 0.134 & 0.824 \\ 0.000 & 0.385 & -0.571 & 0.482 & 0.172 \\ 0.000 & 0.000 & 3.057 & -1.481 & -0.259 \\ 0.000 & 0.000 & 0.000 & 2.595 & -0.068 \\ 0.000 & 0.000 & 0.000 & 0.000 & 6.561 \end{pmatrix}.$$

For the fifth column the MLE is $6.561\epsilon_{CPI} = -.824\epsilon_{FF} - .172\epsilon_{CM} + .259\epsilon_{IM} + .068\epsilon_{FC} + u_{CPI}$, compared to the Bayesian estimates $6.063\epsilon_{CPI} = .059\epsilon_{FF} - .003\epsilon_{CM} + .030\epsilon_{IM} - .009\epsilon_{FC} + u_{CPI}$.

Clearly the Bayesian estimates suggest that the unexpected CPI inflation is close to a structural shock, while the MLE shows a strong negative correlation between unexpected CPI inflation with the unexpected Feeds and crude materials but a positive one with the intermediate materials. These contemporaneous relationships are not plausible given the discussion earlier in this section. The Bayesian estimates make more economic sense.

As noted earlier, the dynamics of the inflation pass through is theoretically ambiguous. We assume that the intercept term is always included.

The posterior mean of model selection indexes on the regression coefficients γ for lag 1 to 12 are given by

c	1.000	1.000	1.000	1.000	1.000						
<i>lag1</i>	0.367	0.024	0.020	0.020	0.020	<i>lag2</i>	0.027	0.039	0.019	0.020	0.019
	0.021	0.337	0.019	0.020	0.020		0.027	0.027	0.020	0.020	0.019
	0.060	0.082	0.607	0.035	0.026		0.846	0.072	0.324	0.036	0.022
	0.166	0.235	0.029	0.365	0.022		0.065	0.098	0.022	0.024	0.020
	0.295	0.652	0.036	0.055	0.594		0.110	0.323	0.035	0.105	0.024
<i>lag3</i>	0.022	0.023	0.020	0.022	0.019	<i>lag4</i>	0.026	0.034	0.020	0.020	0.019
	0.026	0.048	0.029	0.037	0.020		0.023	0.023	0.020	0.020	0.020
	0.145	0.070	0.060	0.051	0.026		0.581	0.096	0.024	0.040	0.024
	1.000	0.094	0.027	0.023	0.051		0.076	0.180	0.022	0.023	0.023
	0.295	0.286	0.030	0.031	0.027		0.108	0.144	0.155	0.071	0.022
<i>lag5</i>	0.022	0.026	0.021	0.020	0.021	<i>lag6</i>	0.065	0.027	0.020	0.020	0.020
	0.024	0.027	0.019	0.020	0.019		0.946	0.024	0.020	0.020	0.020
	0.212	0.154	0.028	0.031	0.021		0.190	0.098	0.022	0.023	0.021
	0.110	0.088	0.031	0.024	0.020		0.140	0.146	0.022	0.027	0.020
	0.116	0.141	0.033	0.032	0.023		0.105	0.176	0.032	0.045	0.028
<i>lag7</i>	0.031	0.023	0.020	0.021	0.019	<i>lag8</i>	0.023	0.023	0.020	0.020	0.019
	0.150	0.025	0.020	0.020	0.020		0.045	0.025	0.021	0.019	0.020
	0.172	0.071	0.023	0.028	0.021		0.086	0.102	0.026	0.023	0.025
	0.063	0.074	0.027	0.024	0.020		0.604	0.310	0.022	0.025	0.020
	0.132	0.948	0.029	0.036	0.023		0.126	0.320	0.035	0.062	0.028
<i>lag9</i>	0.025	0.023	0.020	0.020	0.019	<i>lag10</i>	0.023	0.024	0.020	0.020	0.020
	0.022	0.041	0.020	0.021	0.019		0.022	0.044	0.020	0.020	0.020
	0.070	0.089	0.028	0.023	0.022		0.069	0.077	0.023	0.033	0.021
	0.065	0.094	0.058	0.026	0.021		0.717	0.078	0.029	0.025	0.021
	0.108	0.150	0.029	0.039	0.097		0.415	0.335	0.042	0.035	0.022
<i>lag11</i>	0.033	0.023	0.020	0.020	0.020	<i>lag12</i>	0.023	0.038	0.020	0.020	0.020
	0.022	0.028	0.020	0.021	0.019		0.046	0.025	0.019	0.020	0.019
	0.534	0.071	0.027	0.026	0.022		0.058	0.072	0.022	0.028	0.021
	0.065	0.087	0.028	0.026	0.021		0.131	0.086	0.023	0.027	0.023
	0.126	0.154	0.030	0.042	0.045		0.098	0.188	0.034	0.028	0.056

The posterior mean of Φ is given by

c	0.224	0.307	0.013	0.029	0.073						
<i>lag1</i>	0.229	0.004	0.003	-0.007	0.006	<i>lag2</i>	-0.059	0.087	0.009	0.013	0.004
	-0.007	0.213	0.013	0.008	-0.002		-0.070	-0.058	-0.010	-0.007	0.003
	0.021	-0.026	0.319	0.039	0.070		-0.735	-0.019	0.236	0.082	-0.038
	0.123	-0.180	0.004	0.204	0.040		0.007	0.056	-0.013	-0.031	0.018
	-0.295	1.149	-0.040	0.077	0.308		-0.019	-0.414	-0.039	0.121	-0.032
<i>lag3</i>	0.013	-0.009	0.022	0.030	0.007	<i>lag4</i>	-0.050	-0.073	0.018	-0.005	0.006
	0.060	0.113	0.089	0.113	-0.002		0.029	0.015	-0.008	0.008	-0.009
	0.103	-0.009	0.122	0.110	0.066		-0.481	0.041	-0.029	-0.090	-0.049
	1.421	-0.049	0.051	-0.001	0.133		-0.025	0.129	0.007	-0.022	0.050
	-0.289	0.325	0.001	0.016	0.056		-0.038	0.022	0.150	-0.094	-0.003
<i>lag5</i>	0.005	0.029	-0.019	-0.020	-0.010	<i>lag6</i>	-0.136	0.042	0.012	0.012	0.011
	0.037	-0.045	-0.006	0.006	-0.008		0.398	0.030	-0.016	-0.001	0.004
	0.160	0.108	0.060	0.065	0.029		0.142	0.051	0.009	-0.003	-0.003
	-0.078	0.031	-0.071	-0.037	-0.005		-0.101	0.098	-0.006	0.047	-0.004
	0.029	-0.001	-0.027	-0.019	-0.017		0.022	-0.116	0.032	0.066	0.064
<i>lag7</i>	0.073	-0.013	0.005	-0.005	0.000	<i>lag8</i>	-0.025	0.007	0.021	0.033	0.000
	-0.180	0.014	-0.003	0.005	-0.009		0.111	-0.027	0.012	0.006	0.004
	-0.125	-0.003	-0.023	-0.059	-0.018		-0.042	0.049	-0.041	-0.008	0.063
	-0.027	0.011	-0.051	0.020	0.003		0.476	0.256	0.006	0.045	0.019
	-0.057	-2.241	0.011	-0.040	0.006		0.062	0.389	0.039	0.086	0.058
<i>lag9</i>	0.053	0.014	0.024	-0.007	-0.004	<i>lag10</i>	-0.026	-0.015	0.003	0.015	0.001
	-0.017	-0.089	-0.014	-0.026	-0.006		0.001	0.097	0.001	0.006	0.005
	0.035	-0.041	-0.056	-0.026	-0.033		0.000	0.027	0.020	0.074	-0.029
	0.029	-0.046	0.123	0.037	-0.010		-0.545	-0.025	-0.064	-0.026	0.028
	-0.017	0.064	0.015	-0.054	0.162		-0.437	0.399	-0.067	0.047	0.010
<i>lag11</i>	0.089	-0.009	0.015	-0.009	-0.007	<i>lag12</i>	-0.031	-0.089	0.011	0.023	0.005
	-0.016	-0.043	0.021	0.024	-0.001		0.113	-0.035	-0.001	-0.004	-0.008
	0.394	0.015	-0.059	0.036	0.027		-0.012	-0.024	0.020	0.059	-0.016
	0.014	-0.040	0.063	-0.049	-0.005		-0.098	-0.043	0.006	-0.043	-0.044
	0.064	-0.086	0.022	0.062	0.110		0.027	0.143	0.053	-0.012	-0.129

In comparison, the MLE of Φ is

c	0.279	0.350	-0.002	0.010	0.073						
<i>lag1</i>	0.149	-0.007	0.005	-0.002	0.006	<i>lag2</i>	-0.039	0.122	0.012	0.014	0.004
	-0.009	0.101	0.011	0.007	-0.002		-0.071	-0.070	-0.010	-0.009	0.003
	-0.027	0.244	0.282	0.102	0.076		-0.916	-0.398	0.320	0.139	-0.039
	0.326	-0.704	-0.056	0.045	0.038		0.216	0.473	-0.065	-0.084	0.024
	-0.655	1.618	-0.026	0.173	0.248		0.157	-1.935	-0.065	0.226	-0.035
<i>lag3</i>	0.031	0.044	0.025	0.032	0.007	<i>lag4</i>	-0.043	-0.106	0.024	-0.001	0.007
	0.061	0.150	0.087	0.113	-0.002		0.029	0.085	0.003	0.022	-0.008
	0.314	0.032	0.175	0.156	0.064		-0.807	-0.210	-0.067	-0.143	-0.048
	1.226	-0.336	0.035	-0.018	0.139		0.407	0.629	0.023	-0.038	0.062
	-0.897	0.899	0.048	0.102	0.072		-0.199	-0.500	0.275	-0.119	0.006
<i>lag5</i>	0.013	-0.026	-0.022	-0.028	-0.011	<i>lag6</i>	-0.214	0.107	0.007	0.006	0.011
	0.019	-0.038	-0.005	0.010	-0.009		0.387	0.045	-0.017	0.000	0.003
	0.575	0.510	0.079	0.087	0.023		0.918	0.436	0.032	-0.001	-0.007
	-0.403	0.264	-0.114	-0.107	0.003		-0.662	0.378	-0.016	0.031	0.002
	0.056	-0.650	-0.155	-0.157	-0.036		0.124	-0.646	0.087	0.182	0.079
<i>lag7</i>	0.103	-0.027	0.004	-0.004	0.000	<i>lag8</i>	-0.027	0.064	0.024	0.036	0.001
	-0.166	-0.023	-0.004	0.006	-0.010		0.108	-0.056	0.014	0.010	0.004
	-0.573	-0.198	-0.054	-0.097	-0.015		-0.681	-0.085	-0.056	-0.023	0.079
	0.077	0.058	-0.074	0.043	-0.001		0.819	0.960	0.045	0.093	0.013
	-0.067	-2.834	-0.055	-0.143	0.003		0.660	1.371	0.189	0.257	0.063
<i>lag9</i>	0.098	0.091	0.025	-0.007	-0.005	<i>lag10</i>	-0.015	-0.038	0.005	0.015	0.001
	-0.009	-0.231	-0.013	-0.024	-0.006		0.033	0.140	-0.002	0.001	0.004
	0.222	-0.431	-0.128	-0.088	-0.037		0.120	0.587	0.081	0.134	-0.036
	0.058	0.003	0.177	0.070	-0.013		-0.813	-0.292	-0.080	-0.042	0.032
	0.136	0.614	-0.043	-0.216	0.200		-1.285	0.907	-0.127	0.023	0.015
<i>lag11</i>	0.121	-0.004	0.011	-0.011	-0.008	<i>lag12</i>	-0.057	-0.156	0.008	0.020	0.005
	-0.015	-0.113	0.022	0.027	-0.001		0.153	-0.016	0.000	-0.001	-0.008
	0.817	0.146	-0.069	0.036	0.022		0.032	-0.095	0.045	0.117	-0.018
	-0.183	-0.651	0.026	-0.093	-0.005		-0.553	-0.455	-0.009	-0.063	-0.048
	0.438	-0.611	0.034	0.122	0.141		0.532	1.007	0.099	-0.028	-0.172

The posterior mean of model selection index matrix γ , the MLE and Bayesian estimate of Φ all indicate that there is little pass-through of PPI inflations of foodstuff & feeds and crude materials to PPI of finished goods and CPI. There is limited pass-through of PPI inflation of intermediate

materials to PPI of finished goods. The estimates of Φ show that there is moderate pass-through of PPI of finished goods to CPI with the lags of 1 to 4 months.

The difference between the MLE of Φ and the Bayes estimate lies in the equations of PPI inflation in foodstuff & feeds and PPI of crude materials. The MLE shows that in the equation of the PPI of crude materials the coefficients of PPI of finished goods and CPI are quite sizable, but often switch signs after each lag. The Bayesian estimates are much smaller at each lag. For instance, the sum of the 12 lag coefficients of CPI inflation variable in the equation of foodstuff & feeds and crude materials is -0.760 for MLE and -0.367 for the Bayesian estimates, but the sum of the absolute values of the coefficients is 13.592 for the MLE and only 5.349 for the Bayesian estimates.

Consistent with the large difference in point estimates of Φ matrix obtained for the Bayesian estimates under stochastic search and MLE, the K-L divergence of averaged predictive density over models visited by the stochastic search algorithm and the benchmark model that includes all elements in Ψ and Φ is 136.344 , suggesting a large difference between the two predictive densities.

7 Concluding Remarks

Vector Autoregressive models have been widely used for macroeconomic forecasting and policy analysis. It has been recognized that while the conventional approach of leaving the VAR parameters unconstrained avoids the erroneous restrictions, results in over-parametrization. In this paper, we conduct Bayesian stochastic variable selection of VAR models. We develop a Markov Chain Monte Carlo (MCMC) algorithm that selects the elements of both VAR regression coefficients matrix and error variance matrix. Numerical simulations show that stochastic search algorithm is quite effective in selecting the correct model in some cases and always results in better performance in forecast even when it does not succeed in selecting the correct model. We apply the method in selection among models of inflation transmission from Producer Price Index (PPI) of crude materials to PPI of finished goods and Consumer Price Index(CPI). We find that the MLE and Bayesian estimates give different contemporaneous relationships of unexpected inflations in the identified VAR and substantial different estimates of the feedback of PPI of finished goods and CPI inflation to PPI of crude materials.

A number of methodological issues remain. As in any applications of stochastic search, the number of models actually visited in MCMC is a small portion of all possible models. Due to the

fact that a typical VAR contains a larger number of parameters than a typical univariate regression model, researchers should be concerned with the possibility that MCMC is trapped at a small region of the model space. The problem is not addressed in complete satisfaction even similar results are obtained from MCMC runs of different length. Another general methodological problem concerns the effect of priors. Some comparison of simulation results are made under different prior choice, but computation cost prevents the comparison from being exhaustive. Empirical Bayesian approach of prior elicitation may be a useful alternative.

References

- Aitchison, J. (1975), ‘Goodness of prediction fit’, *Biometrika* **62**, 547–554.
- Amisano, G. & Giannini, C. (1997), *Topics in structural var econometrics, second edition*, Berlin: Springer-Verlag.
- Bekker, P. A. & Pollock, D. S. G. (1986), ‘Identification of linear stochastic models with covariance restrictions’, *Journal of Econometrics* **31**, 179–208.
- Blanchard, O. J. (1989), ‘A traditional interpretation of macroeconomic fluctuations’, *American Economic Review* pp. 1146–1164.
- Blomberg, B. S. & Harris, E. S. (1995), ‘The commodity-consumer price connection: Fact or fable’, *Federal Reserve Bank of San Francisco Economic Policy Review* **October**, 21–38.
- DeJong, D. N., Ingram, B. F. & Whiteman, C. H. (2000), ‘A Bayesian approach to dynamic macroeconomics’, *Journal of Econometrics* **98**(2), 203–223.
- Doan, T., Litterman, R. B. & Sims, C. A. (1984), ‘Forecasting and conditional projection using realistic prior distributions’, *Econometric Reviews* **3**(1), 1–100.
- Furlong, F. & Ingenito, R. (1996), ‘Commodity prices and inflation’, *Federal Reserve Bank of San Francisco Economic Review* **2**, 27–47.
- Geisser, S. (1965), ‘Bayesian estimation in multivariate analysis’, *The Annals of Mathematical Statistics* **36**, 150–159.
- George, E. I. (2000), ‘The variable selection problem’, *Journal of the American Statistical Association* **95**(452), 1304–1308.
- George, E. I. & McCulloch, R. E. (1993), ‘Variable selection via Gibbs sampling’, *Journal of the American Statistical Association* **88**, 881–889.
- Gordon, D. B. & Leeper, E. M. (1994), ‘The dynamic impacts of monetary policy: An exercise in tentative identification’, *Journal of Political Economy* **102**, 1228–1247.
- Hsiao, C. (1981), ‘Autoregressive modelling and money-income causality detection’, *Journal of Monetary Economics* **7**, 85–106.
- Ingram, B. F. & Whiteman, C. H. (1994), ‘Supplementing the Minnesota prior-forecasting macroeconomic time series using real business cycle model priors’, *Journal of Monetary Economics* **34**, 497–510.
- Leonard, T. & Hsu, J. S. J. (1992), ‘Bayesian inference for a covariance matrix’, *The Annals of Statistics* **20**, 1669–1696.
- Madigan, D. & Raftery, A. E. (1994), ‘Model selection and accounting for model uncertainty in graphical models using Occam’s window’, *Journal of the American Statistical Association* **89**, 1535–1546.
- Mattey, J. P. (1981), ‘The cumulation of individual responses to input price changes’, *Board of Governors of the Federal Reserve System Economic Activity Section Working Paper Series 113*.

- Pourahmadi, M. (1999), ‘Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation’, *Biometrika* **86**, 677–690.
- Sims, C. A. (1980), ‘Macroeconomics and reality’, *Econometrica* **48**, 1–48.
- Sims, C. A. (1986), ‘Are forecast models usable for policy analysis?’, *Quarterly Review of Federal Reserve Bank of Minneapolis* **Winter**, 2–16.
- Sims, C. A. & Zha, T. (1998), ‘Bayesian methods for dynamic multivariate models’, *International Economic Review* **39**, 949–968.
- Waggoner, D. F. & Zha, T. (2003), ‘A gibbs sampler for structural vector autoregressions’, *Journal of Economic Dynamics and Control* **28**, 349–366.
- Weinhagen, J. (2002), ‘An empirical analysis of price transmission by stage of process’, *Monthly Labor Review* **November**, 3–11.
- Wong, F., Carter, C. & Kohn, R. (2002), Efficient estimation of covariance selection models, in ‘Bayesian statistics 4. Proceedings of the Seventh Valencia International Meeting’, pp. 182–217.
- Yang, R. & Berger, J. O. (1994), ‘Estimation of a covariance matrix using the reference prior’, *The Annals of Statistics* **22**, 1195–1211.
- Zellner, A. & Tiao, G. C. (1964), ‘Bayesian analysis of the regression model with autocorrelated errors’, *Journal of the American Statistical Association* **59**, 763–778.