Bayesian Estimation and Hypothesis Testing of Identified Normalized VAR Models

Shawn Ni, Department of Economics, University of Missouri, Columbia, MO 65211, USA
Dongchu Sun, Department of Statistics, University of Missouri, Columbia, MO 65211, USA

Summary

In this paper, Bayesian estimation and hypothesis testing are introduced for identified normalized Vector Autoregressive (VAR) models. A class of priors is proposed to take advantage of the structure of normalized VAR models. Efficient Markov Chain Monte Carlo algorithms are used for sampling from the posterior of the VAR parameters without using Metropolis algorithms. Marginal likelihoods are computed via Chib’s (1995) methods for hypothesis testing based on Bayes factors. Numerical simulations show that when the sample size is small, the commonly used Schwarz criterion for testing of competing VAR models deviates from the Bayes factor and is not as effective in selecting the correct model. The method developed in the study is applied to a real data set to test competing models on the macroeconomic, state-, and sector-specific effects of employment growth.

\footnote{The research was supported by a grant from the Research Board of the University of Missouri-System. Sun’s research was also supported by the National Science Foundation grants DMS-9972598 and SES-0095919, and a grant from the Missouri Department of Conservation. We thank Dr. Karon Speckman for proof-reading the paper.}
1 Introduction

Vector Autoregressive (VAR) models became widely employed for time series forecasting after the seminal work of Sims (1980). A VAR of a \( p \) dimensional column variable \( y_t, (t = 1, \ldots, T) \) is

\[
y'_t = c + \sum_{j=1}^{L} y'_{t-j} B_j + \epsilon'_t,
\]

for \( t = 1, \cdots, T \), where the VAR lag \( L \) is a known positive integer, intercept \( c \) is a \( 1 \times p \) unknown vector, \( B_j \) is an unknown \( p \times p \) matrix. The error terms \( \epsilon_1, \cdots, \epsilon_T \) are independent and identically distributed (iid) \( N_p(0, \Sigma) \), and covariance \( \Sigma \) is an unknown \( p \times p \) positive definite matrix. The reduced-form VAR above imposes no restrictions on the regression coefficients \( \Phi = (c', B'_1, \cdots, B'_L)' \) and the covariance matrix \( \Sigma \). While the unrestricted VAR model (1) is useful in summarizing historical data and making forecasts, it is considered by many (e.g., Sargent, 1984) as unfit for addressing policy questions. Policy analysis necessitates VAR identification, which, as eloquently put by Sims (1986), “is the interpretation of historically observed variation in data in a way that allows the variation to be used to predict the consequences of an action not yet undertaken.” Identification may be achieved using restrictions implied by dynamic general equilibrium models. Some Bayesian analysis along this line includes Ingram & Whiteman (1994) and DeJong et al. (2000). An alternative approach is to develop “identified VAR” models that retain much of the flexibility of the VAR model. Identification in this approach is made by incorporating restrictions on the covariance matrix \( \Sigma \) or equivalently on the contemporaneous relationship among variables without restricting the VAR regression coefficients (see, for example, Sims, 1986; Blanchard, 1989; Gordon & Leeper, 1994; and Sims & Zha, 1998.)

In an identified VAR, reduced-form residuals \( \epsilon_t \) are mapped to structural shocks \( u_t \) via \( \epsilon'_t A_0 = u'_t \). By assumption, the structural shocks \( u_t \) are iid. \( N(0, I_p) \). The identified VAR can be written as

\[
y'_t A_0 = cA_0 + \sum_{j=1}^{L} y'_{t-j} B_j A_0 + \epsilon'_t A_0.
\]

Equation (2) suggests an alternative way of viewing the \( A_0 \) matrix – it represents a contemporaneous relationship of variables in the structural equations. The identification concerns conditions under which the structural parameters in \( A_0 \) are recovered from \( \Sigma \) by

\[
\Sigma = A_0^{-1'} A_0^{-1}.
\]

For estimating \( A_0 \), the common practice is to use a two-step procedure. The first step is to obtain the maximum likelihood estimate (MLE) of \( \hat{\Sigma} \) from the unconstrained model (1). The second stage is to estimate \( A_0 \) by maximizing the likelihood \( l(\Sigma^{-1} | \hat{\Sigma}) = l'(A_0 | \hat{\Sigma}) \). The two-step procedure is
justified by the fact that the estimates of VAR regression coefficients do not depend on \( \hat{\Sigma} \), but the numerical solution to the maximization of \( l^*(A_0 | \hat{\Sigma}) \) is complicated by the nonlinearity of \( l^* \) as a function of the elements of \( A_0 \). The MLE results rely on an optimization routine and can be very unstable. In practice, the estimate of \( A_0 \) is often sensitive to perturbation in \( \hat{\Sigma} \). More importantly either in reduced form or identified form, frequentist finite sample inference of VARs cannot be conducted analytically, making it difficult to judge the statistical significance of the estimates. In recent years the Bayesian method that obtains finite sample inference through posterior simulation with no nonlinear maximization involved has become more common for VAR models.

Bayesian analysis combines information from the sample and the prior to form finite sample posterior distribution of parameters. Selecting priors for identified VARs is more difficult than for reduced-form models. For identified VARs, condition (3) relates the \( \Sigma^{-1} \) matrix to the \( A_0 \) matrix and restrictions are placed on the components of the \( A_0 \) matrix. The commonly employed priors (such as the prior in RATS software package) for the covariance matrix are functions of determinants or eigenvalues of \( \Sigma \). They do not allow for component-wise restrictions on the \( A_0 \) matrix. In many applications of identified VARs, the standard RATS prior is often used for finite sample distributions of impulse responses, ignoring the implicit restrictions on the \( \Sigma \) matrix. To correct the deficiencies of this “naive Bayesian procedure,” Sims & Zha (1998, 1999) proposed a set of full Bayesian routines for finite sample inference of identified VARs. They used normal priors for the vectorized \( A_0 \) and \( \Phi \). However, under this prior setting the conditional posterior of \( A_0 \) is a nonstandard distribution. Consequently, posterior simulation involves Metropolis algorithms, which can be costly for models of large number of parameters. Moreover, in this framework it is quite difficult to compute Bayes factors for testing competing VAR models. In the current identified VAR literature, hypothesis testing is based on the Schwarz criterion which is a good approximation of the logarithm of Bayes factor when the sample size is large.

In this study, we propose a class of priors that make full use of the special features of \( A_0 \) in normalized identified VARs. A normalized VAR has a upper-triangular \( A_0 \) that may be just- or over-identified. We suggest taking advantage of such a structure of \( A_0 \) when it is granted by economic theory. Our prior setting leads to more convenient inference for \( A_0 \) than a normal prior does. Under our prior setting, all conditional posteriors are standard distributions. This property has two

\[2\] It is obvious from (3) that matrix \( A_0 \) cannot be uniquely identified unless certain restrictions are imposed on it. In a VAR of \( p \) variables, the sample information used for estimating a \( p \times p \) matrix \( A_0 \) is contained in the estimate \( \hat{\Sigma} \), which has \( p(p+1)/2 \) distinct elements. A VAR is just- (over-) identified if exactly (more than) \( p(p-1)/2 \) independent restrictions are imposed on the \( A_0 \) matrix.
desirable consequences. First, we are able to compute the posterior of $A_0$ and $\Phi$ through conventional MCMC (Markov Chain Monte Carlo). Our MCMC algorithm does not involve Metropolis steps. The required length of the Markov Chain is generally longer when a Metropolis algorithm with a low acceptance rates is employed. Second, conditional posteriors in standard distribution permit calculation of marginal likelihood and/or the Bayes factor via an approach suggested by Chib (1995). With nonstandard conditional posteriors, computation of marginal likelihood is considerably more challenging. Being able to calculate the Bayes factor is important for Bayesian VAR users interested in testing competing models.

Section 2 of the paper lays out the framework for the Bayesian VAR model. Section 3 proposes a prior for normalized identified VARs and derives results on the posterior. Section 4 discusses computation of marginal likelihood and the Bayes factor. Section 5 contains numerical examples of Bayesian estimates based on the proposed prior. It shows that the Schwarz criterion is a poor approximate of the Bayes factor when the sample size is small. Section 6 reports an empirical application on employment growth of three states and three sectors. Section 7 offers concluding remarks.

2 Bayesian VAR Model

2.1 The VAR Model

We rewrite (1) in the familiar matrix form

$$Y = X\Phi + \epsilon,$$

(4)

where $x_t = (1, y_{t-1}', \ldots, y_{t-L}')$,

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_T \end{pmatrix}, \quad X = \begin{pmatrix} x_1 \\ \vdots \\ x_T \end{pmatrix}, \quad \Phi = \begin{pmatrix} c \\ B_1 \\ \vdots \\ B_L \end{pmatrix}, \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_T \end{pmatrix}.$$

Here $Y$ and $\epsilon$ are $T \times p$ matrices, $\Phi$ is a $(1 + Lp) \times p$ matrix, $x_t$ is a $1 \times (1 + Lp)$ row vector, and $X$ is a $T \times (1 + Lp)$ matrix of observations. Write

$$S(\Phi) = (Y - X\Phi)'(Y - X\Phi).$$

(5)

Then the likelihood function of $(\Phi, \Sigma)$ is

$$l(\Phi, \Sigma) \equiv f(Y \mid \Phi, \Sigma) \propto |\Sigma|^{-T/2}etr\left\{-\frac{1}{2}\Sigma^{-1}S(\Phi)\right\}.$$
Here $etr(A)$ is $\exp(\text{trace}(A))$ for a matrix $A$. A commonly used estimate of $(\Phi, \Sigma)$ is the MLE given by

$$(\hat{\Phi}_M, \hat{\Sigma}_M) = ((X'X)^{-1}X'Y, S(\hat{\Phi}_M)/T).$$ (7)

We assume that the sample size $T$ is large enough so that the MLEs of $\Phi$ and $\Sigma$ exist with the probability of one.

2.2 The Bayesian Approach and Prior Choice

Economic analysis of VAR models requires knowledge of the distributional properties of the estimates of parameters $(\Phi, \Sigma)$ and sometimes that of nonlinear functions of the estimates (e.g. impulse responses). Frequentist finite sample distributions of $\hat{\Phi}$ and $\hat{\Sigma}$ are not available in analytical form. Asymptotic theory is often a poor guidance of finite sample VAR inference for two reasons. First, because there are no restrictions on $\Phi$, a VAR of moderate size involves a large number of parameters. The size of the data set is often not large enough to justify the use of asymptotic theory. Furthermore, when nonlinear functions of the VAR estimates (such as impulse responses) are of interest, the asymptotic theory involves approximation of the nonlinear function, and the approximation becomes less precise the more nonlinear the function is (see Kilian, 1999). An alternative to asymptotic theory is the Bayesian approach, which combines information from the sample and the prior to form the finite sample posterior distribution of $(\Phi, \Sigma)$. As empirical literature in macroeconomics involves VARs with larger numbers of variables, the conflict between the growing sizes of the models and the limited availability of data will be sharper. This practical need motivates Bayesian analysis that improves finite sample inference techniques for macroeconomists.

2.3 Commonly Used Priors for VARs

The choice of prior has always been a contentious issue in Bayesian analysis. The most popular noninformative prior for $\Sigma$ is the Jeffreys prior $|\Sigma|^{-(p+1)/2}$ (See Geisser, 1965; and Tiao and Zellner, 1964). A modified version of the Jeffreys prior, $|\Sigma|^{-(L+1)p/2-1}$ is used in RATS (Regression Analysis of Time Series, a widely used software package). For estimating covariance matrix of an iid sample Yang & Berger (1994) proposed a reference prior in the spirit of Bernardo (1979) and Berger & Bernardo (1992), $\{[\Sigma]\prod_{1 \leq i < j \leq p}(d_i - d_j)]^{-1}$, where $d_1 > d_2 > \cdots > d_p$ are eigenvalues of $\Sigma$. These priors are the common choices for unrestricted covariance matrix. Sun & Ni (2002) and Ni & Sun (2003) proved the existence of posteriors of $(\Phi, \Sigma)$ under a large class of noninformative
priors, including the ones discussed above, and evaluated their performance in VAR settings. The
noninformative Jeffreys, RATS, and Yang-Berger priors are functions of determinants or eigenvalues
of $\Sigma$. They do not allow for introduction of component-wise information on the $A_0$ matrix and are
inappropriate for matrices with identifying restrictions.

It follows from (6) that the conditional likelihood function of $\Sigma$ for fixed $\Phi$ depends on $\Sigma^{-1} =
(\sigma_{ij}^2)$, the precision of matrix of the errors in VARs. Naturally, in Bayesian analysis one can set
a prior for $\Sigma^{-1}$ or re-parameterize the likelihood as a function of ($A_0$). In an identified VAR
model, restrictions are placed on the components of the $A_0$ matrix. Sims & Zha (1998) assigned a
multivariate normal prior to the vectorized $A_0$ matrix and simulated the posteriors. Their approach
applies to $A_0$ in general form. For certain problems, economic theory suggests that $A_0$ is upper-
triangular (i.e., normalized). In the following, we propose an alternative approach of imposing priors
on components of $A_0$ that utilizes the normalized structure of the matrix. Under plausible priors,
this approach yields closed-form expressions for the conditional posteriors of all VAR parameters for
the just-identified and over-identified models. We decompose the precision matrix as

$$\Sigma^{-1} = \Psi \Psi^\prime,$$

where $\Psi = \Psi_p$ is the $p \times p$ upper-triangular matrix with $\psi_{ij}$ as its $(i,j)th$ entry, so $\psi_{ij} = 0$ for $i > j$.
Here we add the subscript $p$ for convenience when we derive recursive formulas later. Moreover, $\Psi_p$
is used interchangeably with $A_0$. Note that there is no restriction on off-diagonal elements $\psi_{ij}$, but
$\psi_{ii}$ should never change signs, so we can assume that $\psi_{ii} > 0$. Such a matrix $\Psi_p$ with (without)
restrictions on $\psi_{ij}$ is called normalized and over- (just-) identified by Sims & Zha (1998).

Identification restrictions on $\Psi_p$ are related with the literature on sparse precision matrix. Elements
of zeros in a precision matrix $\Sigma^{-1}$ represent conditional independence of variables (see Wong
et al. 2002). Sparse precision matrices are important in small area estimation and spatial mod-
elling where random geographical effects depend only on neighboring regions (cf. Ghosh et al. 1999;
Gelfand et al. 2001; and Kim et al., 2001). The distribution with a degenerated banded precision
matrix (corresponding to banded $\Psi_p$ with some zeros on the diagonal elements) is a special case
of the partially informative normal by Sun et al. (1999) which has been used on spline smoothing
(cf. Wahba, 1985; and Speckman & Sun, 2003). Note however, although elements of zeros in an
over-identified $\Psi_p$ restrict the precision matrix, they do not necessarily imply that some elements of
the precision matrix are zero. In some cases, the over-identified parameterization of $\Psi_p$ is equivalent
to a type of conditional independence. Sun and Sun (2002) showed that the best invariance estimate
of the normal covariance matrix $\Sigma$ for some normalized and over-identified $\Psi_p$ is a Bayesian esti-
mate under the squared error loss or entropy loss with a noninformative prior that depends on the structure of the conditional independence.

3 Priors and Posteriors of Identified VAR Models

3.1 Normalized and Just-identified Parameterization

3.1.1 Priors for $\phi = \text{vec}(\Phi)$ and $\Psi_p$.

A common choice for priors on $\phi$ is normal with fixed mean $\phi_0$ and variance $\Xi_0$, i.e.,

$$
\phi \sim N(\phi_0, \Xi_0).
$$

To give the prior for $\Psi$, the Cholesky decomposition of $\Sigma^{-1}$ given by (8), we let $\psi_{p-1,p}$ denote the $(p-1) \times 1$ vector of the last column of $\Psi_p$ excluding $\psi_{pp}$. We have

$$
\Psi_1 = \psi_{11}, \quad \Psi_2 = \begin{pmatrix} \psi_{11} & \psi_{12} \\ 0 & \psi_{22} \end{pmatrix}, \quad \ldots, \quad \Psi_p = \begin{pmatrix} \Psi_{p-1} & \psi_{p-1,p} \\ 0 & \psi_{pp} \end{pmatrix}.
$$

The priors for off-diagonal elements are assumed to be independent multivariate normal distributions,

$$
\psi_{i-1,i} \overset{\text{iid}}{\sim} N_{i-1}(0, \Omega_{i-1}^{-1}), \quad \text{for } i = 2, \ldots, p.
$$

We have the following consideration on the prior for diagonal elements $\psi_{ii}$. Because $\psi_{ii} > 0$, it is equivalent to select a prior for $\psi_{ii}^2$. If a normal prior is used for $\psi_{ii}$, $\psi_{ii}^2$ would have a noncentral $\chi^2$ distribution with one degree of freedom. Motivated by Wong et al. (2002), we assume that

$$
\psi_{ii}^2 \overset{\text{iid}}{\sim} \text{gamma} (a_i, b_i), \quad \text{for } i = 1, 2, \ldots, p.
$$

The class of gamma distributions is quite general and more flexible than the $\chi^2$. Furthermore, it leads to a closed-form expression of the conditional posterior of $\Psi$, granting easy MCMC simulations. In comparison, using the normal distribution for $\psi_{ii}$ requires more extensive simulation, limiting the applicability of the approach to large-scale problems.

Eaton & Olkin (1987) considered the following setting. They let the prior of $\psi_{ii}^2$ be gamma with $a_i = b_i = 0$, and the prior for off-diagonal elements $\psi_{ij}$ be constant. Under such a prior, they find that an optimal equivariant estimate of the Cholesky decomposition of a covariance matrix from multivariate normal observations is a Bayesian estimator. Their prior setting is a special case of ours.
3.1.2 Conditional Posteriors of \((\phi, \Psi_p)\).

The full conditional distributions are useful for MCMC posterior simulations. Note that the likelihood function (6) of \((\phi, \Sigma)\) can be rewritten as

\[
l(\phi, \Sigma) = |\Sigma|^{-T/2} \exp \left[ -\frac{1}{2} (\phi - \hat{\phi}_M)' \{ \Sigma^{-1} \otimes (X'X) \} (\phi - \hat{\phi}_M) - \frac{1}{2} tr\{ \Sigma^{-1} S_M \} \right],
\]

where \(\hat{\phi}_M = vec(\hat{\Phi}_M)\), and \(S_M = S(\hat{\Phi}_M); (\hat{\Phi}_M, S(\hat{\Phi}_M))\) is given in (7). Combining (12) and (9), we know that the conditional posterior density of \(\phi\) given \((\Psi_p; Y)\) is

\[
[\phi \mid \Psi_p; Y] \propto \exp \left\{ -\frac{1}{2} (\phi - \hat{\phi}_M)' \{ \Sigma^{-1} \otimes (X'X) \} (\phi - \hat{\phi}_M) - \frac{1}{2} (\phi - \phi_0)' \Xi^{-1} (\phi - \phi_0) \right\}.
\]

Here and in the following, we use \((\cdot \mid \cdot)\) and \([\cdot \mid \cdot]\) to denote conditional distribution and conditional density, respectively. It is easy to show the following result.

**Theorem 1** The conditional posterior distribution of \(\phi\) given \((\Psi_p; Y)\) is

\[
(\phi \mid \Psi_p; Y) \sim N(\hat{\phi}, \hat{\Xi}),
\]

where

\[
\hat{\Xi} = \left\{ \Sigma^{-1} \otimes (X'X) + \Xi_0^{-1} \right\}^{-1}, \quad \hat{\phi} = \hat{\Xi} \left\{ \Sigma^{-1} \otimes (X'X) \hat{\phi}_M + \Xi_0^{-1} \phi_0 \right\}.
\]

Given \(\Phi\), we now use \(S = S_p\) to represent the covariance of VAR residuals \(S(\Phi)\), and \(S_i\) the upper-left \(i\) by \(i\) block of \(S(\Phi)\). The subscript \(i\) is included for the following recursive formulas. Define \(w_1 = s_{11}\) and \(w_i = |S_i|/|S_{i-1}|\) for \(i = 2, \ldots, p\). By definition, \(w_i = s_{ii} - s_{i-1,i} S_{i-1}^{-1} s_{i-1,i} > 0\) for \(i = 2, \ldots, p\). We then have a recursive formula,

\[
tr(\Psi_p'S_p\Psi_p) = tr(\Psi_{p-1}'S_{p-1}\Psi_{p-1}) + \psi_{pp}^2 w_p + g'S_{p-1}g,
\]

where \(g = \psi_{p-1,p} + \psi_{pp} S_{p-1}^{-1} s_{p-1,p}\). The likelihood function (6) can be written as

\[
\prod_{i=1}^{p} (\psi_{ii}^2)^{\frac{T}{2}} \exp \left( -\frac{1}{2} \left\{ \sum_{i=1}^{p} \psi_{ii}^2 w_i + \sum_{i=2}^{p} (\psi_{i-1,i} + \psi_{ii} S_{i-1}^{-1} s_{i-1,i})/S_{i-1}(\psi_{i-1,i} + \psi_{ii} S_{i-1}^{-1} s_{i-1,i}) \right\} \right).
\]

**Theorem 2** (a) For given \((\phi, \psi_{11}, \ldots, \psi_{pp}; Y)\), \(\psi_{i-1,i}, i = 2, \ldots, p\) are mutually independent and \((\psi_{i-1,i} \mid \phi, \psi_{i-1,i}, \ldots, \psi_{pp}; Y)\) depends only on \((\phi, \psi_{ii}; S_i)\);

\[
(\psi_{i-1,i} \mid \psi_{ii}; S_i) \propto N(h_i, (S_{i-1} + \Omega_{i-1})^{-1}).
\]
where  

\[ h_i = -\psi_{ii}(S_{i-1} + \Omega_{i-1})^{-1}s_{i-1,i}. \]  

(17)

(b) For given \((\phi; Y)\), the conditional posteriors of \(\psi_{11}^2, \ldots, \psi_{pp}^2\) are independent, and

\[ (\psi_{ii}^2 | \phi; Y) \sim \text{gamma}(a_i + \frac{1}{2}T, B_i), \]  

where

\[ B_i = \begin{cases} b_i + \frac{1}{2}s_{11}, & \text{if } i = 1, \\ b_i + \frac{1}{3}\{s_{ii} - s_{i-1,i}^{'}(S_{i-1} + \Omega_{i-1})^{-1}s_{i-1,i}\}, & \text{if } i = 2, \cdots, p. \end{cases} \]  

Proof. For the priors (10) and (11), the conditional posterior density of \((\psi_{ii}^2, i = 1, \cdots, p; \psi_{j-1,j}, j = 2, \cdots, p)\) given \((\phi; Y)\) is given by

\[
\begin{align*}
[\psi_{ii}^2, i = 1, \cdots, p; \psi_{j-1,j}, j = 2, \cdots, p | \phi, Y] &\propto \left\{ \prod_{i=1}^{p}(\psi_{ii}^2)^{a_i+\frac{T}{2}-1} \exp \left[ -\frac{1}{2} \left( \sum_{i=1}^{p} \psi_{ii}^2(w_i + 2b_i) \right) \right] \right\} \exp\left\{ -\frac{1}{2} \sum_{i=2}^{p} \psi_{i-1,i}^{'}\Omega_{i-1}\psi_{i-1,i} \right\} \\
&\times \exp\left\{ -\frac{1}{2} \sum_{i=2}^{p} (\psi_{i-1,i} + \psi_{ii}S_{i-1,i}^{-1}s_{i-1,i})^{'}S_{i-1}(\psi_{i-1,i} + \psi_{ii}S_{i-1,i}^{-1}s_{i-1,i}) \right\} \\
&= \left\{ \prod_{i=1}^{p}(\psi_{ii}^2)^{a_i+\frac{T}{2}-1} \right\} \exp\left\{ -\sum_{i=1}^{p} \psi_{ii}^2B_i - \frac{1}{2} \sum_{j=2}^{p} (\psi_{j-1,j} - h_j)^{'}(S_{j-1} + \Omega_{j-1})(\psi_{j-1,j} - h_j) \right\}.
\end{align*}
\]

The results hold immediately. \(\square\)

3.2 Normalized and Over-identified Parameterization

A desirable feature of the prior setting proposed in the previous section is that the posterior of \(A_0\) matrix can be derived in a similar fashion for over-identified and just-identified models. The result is due to the following fact.

Lemma 1 Consider an \(n\) dimensional vector \(\eta = (\eta_1, \cdots, \eta_n)^{'}\), while \(m\) of its elements, namely \(\eta_{k_1}, \cdots, \eta_{k_m}\), are zero. Let \(\Theta = (\theta_{ij})\) be an \(n \times n\) symmetric matrix. Define \(\bar{I}_{nm}\) as the remainder of the \(n \times n\) identity matrix \(I_n\) after deleting the \(m\) rows corresponding to the zero-entries in \(\eta\). Denote the \((n - m)\) dimensional vector \(\tilde{\eta} = \bar{I}_{nm}\eta\), and \((n - m) \times (n - m)\) matrix \(\tilde{\Theta} = \bar{I}_{nm}\Theta\bar{I}_{nm}^{'}\). Then

\[ \eta^{'}\Theta\eta = \tilde{\eta}^{'}\tilde{\Theta}\tilde{\eta}. \]

Proof. Note that

\[
\begin{align*}
\eta^{'}\Theta\eta &= \sum_{i=1,\cdots,n} \sum_{j=1,\cdots,n} \eta_i\theta_{ij}\eta_j = \sum_{i\neq k_1,\cdots, i\neq k_m} \sum_{j\neq k_1,\cdots, j\neq k_m} \eta_i\theta_{ij}\eta_j \\
&= \sum_{i=1,\cdots,n-m} \sum_{j=1,\cdots,n-m} \tilde{\eta}_i\tilde{\theta}_{ij}\tilde{\eta}_j = \tilde{\eta}^{'}\tilde{\Theta}\tilde{\eta}.
\end{align*}
\]
Let \( m_{i-1} \) be the number of non-zero entries in the vector \( \psi_{i-1,j} \) (so \( m_{i-1} \leq i - 1 \)). Construct a matrix \( \tilde{I}_{i-1} \) by deleting the rows of the identity matrix \( I_{i-1} \) corresponding to the zero-entries in \( \psi_{i-1,j} \). It is easy to see that \( \tilde{I}_{i-1}\tilde{I}_{i-1} = I_{m_{i-1}} \). Let \( \tilde{\psi}_{i-1,j} = \tilde{I}_{i-1}\psi_{i-1,j}, \tilde{S}_{i-1} = \tilde{I}_{i-1}S_{i-1}\tilde{I}_{i-1} \), and \( \tilde{s}_{i-1,j} = \tilde{I}_{i-1}s_{i-1,j} \). Following the lemma, with the presence of the zero entries in \( \psi_{i-1,j} \), we have

\[
(\psi_{i-1,j} + \psi_{ii}S_{ii}^{-1}s_{i-1,j})'S_{i-1}(\psi_{i-1,j} + \psi_{ii}S_{ii}^{-1}s_{i-1,j}) = (\tilde{\psi}_{i-1,j} + \psi_{ii}\tilde{S}_{ii}^{-1}\tilde{s}_{i-1,j})'\tilde{S}_{i-1}(\tilde{\psi}_{i-1,j} + \psi_{ii}\tilde{S}_{ii}^{-1}\tilde{s}_{i-1,j}) + \psi_{ii}^2(s_{i-1,j}\tilde{S}_{i-1}^{-1}s_{i-1,j} - \tilde{s}_{i-1,j}\tilde{S}_{i-1}^{-1}\tilde{s}_{i-1,j}).
\]

As for the just-identified model, we choose independent priors (9) for \( \phi \) and (11) for \( \psi_{11}, \ldots, \psi_{pp} \). Furthermore, we assume the following independent normal prior for \( \tilde{\psi}_{i-1,j} \),

\[
\tilde{\psi}_{i-1,j} \overset{\text{indep.}}{\sim} N_{m_{i-1}}(0, \tilde{\Omega}_{i-1}), \text{ for } i = 2, \ldots, p.
\]

The following theorem gives the full conditional distributions of \( \phi \) and \( \Psi_p \).

**Theorem 3**  
(a) The conditional posterior distribution of \( \phi \) given \( \Psi_p; Y \) is the same as (13).  
(b) For given \( \theta \), \( \psi_{11}, \ldots, \psi_{pp}; Y \), the conditional posterior of \( \tilde{\psi}_{i-1,j}, i = 2, \ldots, p \) are independent normal and \( (\tilde{\psi}_{i-1,j} | \phi, \psi_{11}, \ldots, \psi_{pp}; Y) \) depends only on \( (\phi, \psi_{ii}; S_i) \):

\[
(\tilde{\psi}_{i-1,j} | \psi_{ii}, S_i) \sim N(\tilde{h}_i, (\tilde{S}_{i-1} + \tilde{\Omega}_{i-1})^{-1}),
\]

where

\[
\tilde{h}_i = -\psi_{ii}(\tilde{S}_{i-1} + \tilde{\Omega}_{i-1})^{-1}\tilde{s}_{i-1,j}.
\]

(c) For given \( \phi; Y \), the conditional posteriors of \( \psi_{11}^2, \ldots, \psi_{pp}^2 \) are independent and \( (\psi_{ii}^2 | \phi; Y) \) follows gamma \( (a_i + \frac{1}{2}T, \tilde{B}_i) \), where

\[
\tilde{B}_i = \left\{ \begin{array}{ll}
    b_1 + \frac{1}{2}s_{i1}, & \text{if } i = 1, \\
    b_i + \frac{1}{2}\left(s_{ii} - \tilde{s}_{i-1,j}(\tilde{S}_{i-1} + \tilde{\Omega}_{i-1})^{-1}\tilde{s}_{i-1,j} \right), & \text{if } i = 2, \ldots, p.
\end{array} \right.
\]

**Proof.** Following the proof of Theorem 2, the conditional posterior density of \( (\psi_{ii}^2, i = 1, \ldots, p; \psi_{j-1,j}, j = 2, \ldots, p) \) given \( \phi; Y \) is

\[
\propto \left\{ \prod_{i=1}^p (\psi_{ii}^2)^{a_i + \frac{1}{2} - 1}\right\} \exp\left\{ -\sum_{i=1}^p \psi_{ii}^2\tilde{B}_i - \frac{1}{2} \sum_{j=2}^p (\tilde{\psi}_{j-1,j} - \tilde{h}_j)'(\tilde{S}_{j-1} + \tilde{\Omega}_{j-1})(\tilde{\psi}_{j-1,j} - \tilde{h}_j) \right\}.
\]

The results hold immediately.
3.3 Normalizable VAR

In some cases, the $A_0$ matrix may not be upper-triangular but can be normalized, either just-identified or over-identified, but we can normalize the VAR by switching the order of variables. Consider the following case:

$$ (\epsilon_{3t}, \epsilon_{1t}, \epsilon_{2t})A_0 = (u_{3t}, u_{1t}, u_{2t}), \quad \text{where } A_0 = \begin{pmatrix} \lambda_3 & 0 & 0 \\ \tau_2 & \lambda_1 & \tau_1 \\ 0 & 0 & \lambda_2 \end{pmatrix}. $$

Let $I_{ij}$ be the permutation matrix after switching the $i$th and $j$th rows of the identity matrix $I_3$. Define $C = I_{i3}I_{j3} = (c_{ij})$ with $c_{13} = c_{21} = c_{32} = 1$, and $c_{ij} = 0$ otherwise. It is easy to see that $CC' = I_3$. Then $\Psi = C'A_0C$ is an upper triangular matrix. The problem is equivalent to the normalized over-identified model based on the sum of squared residuals $\tilde{S} = C'SC$. This operation does not change the likelihood because neither the determinant nor the trace term changes. Consequently, by switching the order of variables, the model is transferred to a normalized VAR.

$$ (\epsilon_{1t}, \epsilon_{2t}, \epsilon_{3t})\Psi = (u_{1t}, u_{2t}, u_{3t}), \quad \text{where } \Psi = \begin{pmatrix} \lambda_1 & \tau_1 & \tau_2 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}. \quad (24) $$

3.4 An MCMC Algorithm for Posterior Simulation

Following the analytical form of the posterior of $\Psi_p$, we use a standard MCMC algorithm to sample from the joint posterior distribution $(\phi, \Sigma)$, without applying Metropolis algorithms. Metropolis algorithms permit simulation from nonstandard posteriors, but they may be inefficient when the acceptance rate is low (see Chib & Greenberg (1995) for a lucid survey). Suppose we have simulated $\Sigma_{k-1}$ and $\Phi_{k-1}$ from cycle $k - 1$. Then in cycle $k$ the following algorithm is used for computing the posterior under the proposed prior for the just-identified model.

Algorithm:

Step 1: Simulate $\Sigma_k$:
1a. Compute $S = S(\Phi_{k-1})$.
1b. For $i = 1, \ldots, p$, simulate $\xi_i \sim \text{gamma}(a_i + \frac{1}{2}T, B_i)$, where $B_i$ is given by (19).
1c. Simulate off-diagonal elements $\psi_{i-1,i}$ from (16).
1d. Compute $\Sigma_k = \Psi^{-1}_{p} \Psi^{-1}_{p} \Psi^{-1}_{p}$, where $\Psi_{p}$ is the upper triangular matrix, with diagonal element $\phi_{ii} = \sqrt{\xi_i}$, and the first $i - 1$ elements of the $ith$ column $\psi_{i-1,i}$ (from Step 1c).
Step 2: Simulate $\phi_k$ from $N(\tilde{\phi}_k, \tilde{\Xi}_k)$, where

$$
\tilde{\Xi}_k = \left[ (\Psi'_p \Psi_p) \otimes (X'X) + \Xi_0^{-1} \right]^{-1},
$$
$$
\tilde{\phi}_k = \tilde{\Xi}_k \left[ (\Psi'_p \Psi_p) \otimes (X'X) \tilde{\phi}_M + \Xi_0^{-1} \phi_0 \right].
$$

Our posterior analysis in Section 3.2 shows that the algorithm for over-identified models is similar to the one above.

4 Marginal Likelihood and Hypothesis Testing

Consider the problem of comparing two models, one of them (which may be just-identified or over-identified) is labelled as $M^1$, and an alternative model is labelled as $M^2$. Bayesian hypothesis testing or model selection is based on the posterior odds

$$
\pi(M^1) m^1(Y) \pi(M^2) m^2(Y),
$$

(25)

where $\pi(M^1)/\pi(M^2)$ is the prior odds. In the case that researchers assign unity to the prior odds, the test is determined by the Bayes factor of Model 1 versus Model 2, which is the ratio of marginal likelihoods,

$$
B_{12} = \frac{m^1(Y)}{m^2(Y)}.
$$

(26)

Assume that under Model $k$, the unknown parameter is $\theta^k$. Then the marginal likelihood of $Y$ under Model $M^k$ is

$$
m^k(Y) = \int [Y | \theta^k, M^k][\theta^k | M^k] d\theta^k.
$$

The difference between a test using the Bayes factor and a likelihood ratio test is that the former is based on the ratio of average likelihoods weighted by priors and the latter is based on the ratio of maximum likelihoods. Kass & Raftery (1995) provided a survey on Bayes factors and hypothesis testing.

The marginal likelihood can be calculated by integrating the posterior $\pi(\theta | Y)$ if the pdf of posterior is a standard distribution (see Zellner, 1971 for example of such cases). If the posterior is not a standard distribution, the marginal likelihood can be computed through numerical integration of the posterior. Unfortunately, the numerical integration is often difficult for large models. As Kass & Raftery (1995) pointed out, “most available software developed by numerical analyst is generally
so inefficient for these integrals that it is of little use.” A vast literature has been developed to improve the accuracy of computation of the marginal likelihood through numerical integration. For a few examples on the approaches and remaining difficulties, see Geweke (1989), McCulloch & Rossi (1991), Newton & Raftery (1994), Gelfand & Dey (1994), Carlin & Chib (1995), and Meng & Wong (1996).

We now illustrate the difficulty in estimating the marginal likelihood for the identified VAR. Without the loss of generality, assume Model 1 is a just-identified VAR. Under the prior (9)–(10), the posterior is not standard distribution but conditional posteriors are. In fact, the joint posterior of \((\phi, \Psi)\) is proportional to

\[
l(\phi, \Sigma) \exp \left[ -\frac{1}{2} (\phi - \phi_0)' \Xi^{-1} (\phi - \phi_0) \right] \pi(\Psi),
\]

where \(l(\phi, \Sigma)\) is given by (12) and \(\pi(\Psi)\) is the prior density of \(\Psi\). Integrating out the parameters \(\phi\), we can show that the marginal posterior for \(\Psi\) is

\[
\propto \pi(\Psi) | \Psi |^T | \Psi \Psi' \otimes (X'X) + \Xi^{-1} |^{1/2} \\
\exp \left\{ -\frac{1}{2} tr(\Psi \Psi' S_M) - \frac{1}{2} tr\{(\Psi \Psi' \otimes (X'X))^{-1} + \Xi_0\}^{-1} (\hat{\phi}_M - \phi_0)(\hat{\phi}_M - \phi_0)' \right\}.
\]

The above posterior of \(\Psi\) does not follow a standard distribution and integration with respect to \(\Psi\) cannot be conducted analytically. We have drawn \(\Psi, \phi\) from the posterior, which does not require knowing the normalizing constant. In applications, the dimension of parameters can easily be more than several hundreds. Despite increasing computer capacity and advances of numerical techniques, it is still desirable to avoid computing Bayes factors through numerical integrations.

One solution to the problem of calculating marginal likelihood without conducting massive numerical integration was proposed by Chib (1995). Chib suggested estimating marginal likelihood \(m(Y)\) from the likelihood \(l(Y | \theta)\), prior \(\pi(\theta)\), and posterior \(\pi(\theta | Y)\) at a given parameter value \(\theta^*\) based on the Bayes rule. \(\theta^*\) may be the posterior mean or MLE. The Bayes rule implies

\[
m(Y) = \frac{l(Y | \theta^*) \pi(\theta^*)}{\pi(\theta^* | Y)}.
\]

As suggested by Chib, taking the logarithm of the identity gives rise to

\[
\log \{m(Y)\} = \log \{l(Y | \theta^*)\} + \log \{\pi(\theta^*)\} - \log \{\pi(\theta^* | Y)\}.
\]

In our context, \(\theta = (\phi, \Psi)\). Suppose we don’t know the normalizing constant of the posterior \(\pi(\theta | Y)\) but we do for \((\Psi | Y)\) and conditional posterior \((\phi | \Psi; Y)\). We can rewrite the logarithm of joint
posterior of \((\phi^*, \Psi^*)\) as
\[
\log\{\pi(\phi^*, \Psi^* \mid Y)\} = \log\{\pi(\phi^* \mid \Psi^*, Y)\} + \log\{\pi(\Psi^* \mid Y)\}. \tag{27}
\]
The first term in the right hand side of (27) is simple, because the conditional posterior \(\pi(\phi \mid \Psi^*, Y)\) is normal. In fact,
\[
\pi(\phi^* \mid \Psi^*, Y) = (2\pi)^{-\frac{n(p+1)}{2}}|\Xi^*|^{-1/2}\exp\left\{-\frac{1}{2}(\phi^* - \hat{\phi}^*)^T\Xi^*(\phi^* - \hat{\phi}^*)\right\},
\]
where
\[
\hat{\phi}^* = \Xi^* \{\Psi^* \Psi^* \otimes (X'X) + \Xi_0^{-1}\}^{-1},
\]
\[
\Xi^* \equiv \{\Psi^* \Psi^* \otimes (X'X) + \Xi_0^{-1}\}^{-1}.
\]

The second term in the right hand side of (27) is not of a standard form but can be calculated by using the following relationship,
\[
\pi(\Psi^* \mid Y) = \int \pi(\Psi^* \mid \phi, Y)\pi(\phi \mid Y)d\phi.
\]
Therefore, given \(\phi^{(j)} (j = 1, \ldots, N)\) randomly sampled from the marginal posterior of \(\phi, (\Psi^* \mid Y)\) can be estimated by
\[
\frac{1}{N} \sum_{j=1}^{N} \pi(\Psi^{(j)} \mid Y, \phi^{(j)}) = \frac{1}{N} \sum_{j=1}^{N} \left\{\prod_{i=2}^{p} \pi(\phi^*_{i,i-1} \mid Y, \phi^{(j)}, (\psi_{ii}^2)^*)\right\}\left\{\prod_{j=1}^{p} \pi((\psi_{ii}^2)^* \mid Y, \phi^{(j)})\right\}. \tag{28}
\]
In our case here, define
\[
S^{(j)} = S(\Phi^{(j)}), \quad G_{ij} = (S^{(j)}_{i-1} + \Omega_{i-1})^{-1}, \quad B^{(j)} = b_i + \frac{1}{2} \left\{s^{(j)}_{i-1} - s^{(j)}_{i-1,i} G_{ij} s^{(j)}_{i-1,i}\right\}.
\]
It follows from (16) and (18) that
\[
\pi(\phi^*_{i,i-1} \mid Y, \phi^{(j)}, (\psi_{ii}^2)^*) = \frac{1}{(2\pi)^{m_{ii}^{-1}}|G_{ij}|^{1/2}}\exp\left\{-\frac{1}{2}(\phi^*_{i,i-1} + \psi_{ii}^* G_{ij} s^{(j)}_{i-1,i})^T G_{ij}^{-1}(\phi^*_{i,i-1} + \psi_{ii}^* G_{ij} s^{(j)}_{i-1,i})\right\}, \tag{29}
\]
\[
\pi((\psi_{ii}^2)^* \mid Y, \phi^{(j)}) = \frac{B^{(j)} A_i}{\Gamma(A_i)} ((\psi_{ii}^2)^*)^{A_i-1} \exp\left\{-((\psi_{ii}^2)^*)^{B^{(j)}}\right\}. \tag{30}
\]

The computation for the alternative Model 2, (assuming it is over-identified), is similar. Following the notations used in Section 3.2, we denote parameters in an over-identified model by the wiggle sign. For an over-identified model, (29) is changed to
\[
\pi(\phi^*_{i,i-1} \mid Y, \phi^{(j)}, (\psi_{ii}^2)^*) = \frac{1}{(2\pi)^{m_{ii}^{-1}}|G_{ij}|^{1/2}}\times\exp\left\{-\frac{1}{2}(\phi^*_{i,i-1} + \bar{\psi}_{ii}^* G_{ij} s^{(j)}_{i-1,i})^T G_{ij}^{-1}(\phi^*_{i,i-1} + \bar{\psi}_{ii}^* G_{ij} s^{(j)}_{i-1,i})\right\}, \tag{31}
\]
where \(m_i\) is the number of nonzero off-diagonal elements in column \(i\) of the upper-triangular matrix \(\Psi\).
4.1 Compare the Bayes Factor with the Schwarz Criterion

When Bayes factors are difficult to compute, researchers often use the Schwarz criterion for model selection. See, for example, Kass & Raftery (1995). The Schwarz criterion is defined as the log difference of the maximum likelihoods adjusted by the difference of the parameter numbers in the competing models and sample size. Suppose Model 1 (with MLE of parameter $\theta_1^*$) is over-identified with $q$ off-diagonal elements of $\Psi$ being set at zero and Model 2 (with MLE of parameter $\theta_2^*$) is just-identified. Then the Schwarz criterion is given by

$$S_{12} = \log\{l(Y \mid \theta_1^*)\} - \log\{l(Y \mid \theta_2^*)\} + \frac{q}{2} \log(T).$$

If the sample size $T$ is large enough, the Schwarz criterion $S_{12}$ is a good approximate of the logarithm of the Bayes factor $B_{12}$ in the sense $S_{12}/\log B_{12} \rightarrow 1$ as $T \rightarrow \infty$. In the following, we will examine finite sample performance of hypothesis testing based on the Bayes factors and the Schwarz criterion.

5 Numerical Examples

The hyper-parameters of the prior in the numerical example are as follows. The prior mean of the VAR regression coefficients, $\phi_0$ is set at its true value given below; its covariance $\Xi_0$ is set at $I$. The prior of the vector of off-diagonal elements of the $\Psi$ matrix is set to be quite disperse (in contrast to the zero constraints in the over-identified model) with its precision $\Omega$ being $0.1 \times I$. Unless specified otherwise, the hyperparameters in the gamma priors for diagonal elements of $\Psi$ are $a_i=0.1$ and $b_i=0.1$ for $i = 1, \ldots, p$.

Example 1 We first consider an identified VAR in which the $A_0$ matrix has a single zero constraint.

$$\Psi = \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & -0.5 & -0.5 \\ -0.5 & 1.25 & 0.25 \\ -0.5 & 0.25 & 1.25 \end{pmatrix}. \quad (32)$$

The data generating VAR has one lag. The intercept is set at zero. The lag regression coefficient matrix is the identity matrix, meaning that the VAR consists of random walk variables. For 1,000 sample replications with Markov Chain length 10,500 (with 500 burn-in runs), the MCMC computations take about one hour total on a 1.7 GHz Pentium4 PC. Simulation results are little changed when the Markov Chain length is reduced to 5,000 or the number of generated samples is reduced to 500, suggesting that the Markov chains converge rather quickly. In reporting the result we use
subscripts $M$ for MLE, $CM$ for constrained MLE, 1 for the posterior mean of the constrained $\Psi$ (over-identified model), and 2 for the posterior mean of the unconstrained $\Psi$ (just-identified model).

Although the constraint is placed on the $\Psi$ matrix, given the fact the likelihood is determined by $\Sigma = \Psi^{-1}\Psi^{-1}$, we report the frequentist average of $\Sigma$ instead of that of $\Psi$. The frequentist average of MLEs of $\Sigma$ over the 1,000 samples is

$$\hat{E}_0(\Sigma_M) = \begin{pmatrix} 0.848 & -0.419 & -0.422 \\ -0.419 & 1.052 & 0.209 \\ -0.422 & 0.209 & 1.073 \end{pmatrix}. $$

The frequentist average of $\Sigma$ indicates that the diagonal elements of MLE $\hat{\Sigma}_M$ have downward bias. Consistent with the finding in Ni & Sun (2003), the prior choice on $\Sigma$ makes little difference in estimation of $\Phi$. The frequentist average of constrained MLEs is

$$\hat{E}_0(\Sigma_{CM}) = \begin{pmatrix} 0.922 & -0.456 & -0.459 \\ -0.456 & 1.144 & 0.228 \\ -0.459 & 0.228 & 1.166 \end{pmatrix}. $$

To test the over-identified model (labelled as Model 1), we consider an alternative in which the regression coefficients $\Phi$ is the same as above but the $A_0$ matrix is just-identified (labelled as Model 2). The frequentist averages of posterior means under the two models are

$$\hat{E}_0(\Sigma_1) = \begin{pmatrix} 0.965 & -0.477 & -0.480 \\ -0.477 & 1.137 & 0.238 \\ -0.480 & 0.238 & 1.160 \end{pmatrix}, \quad \hat{E}_0(\Sigma_2) = \begin{pmatrix} 0.964 & -0.476 & -0.480 \\ -0.476 & 1.216 & 0.239 \\ -0.480 & 0.239 & 1.263 \end{pmatrix}. $$

For gauging the performance of competing estimates of the identified VAR, we consider the entropy loss function for $\Sigma$ and a quadratic loss for $\Phi$,

\begin{align*}
L_1(\hat{\Sigma}, \Sigma) &= tr(\hat{\Sigma}^{-1}\Sigma) - \log|\hat{\Sigma}^{-1}\Sigma| - p, \\
L_2(\hat{\Phi}, \Phi) &= tr((\hat{\Phi} - \Phi)'(\hat{\Phi} - \Phi)).
\end{align*}

Under the loss function $L_1 + L_2$, the Bayes estimator $(\Sigma, \Phi)$ is the posterior mean for the just-identified model. Deriving the Bayesian estimator of $\Sigma$ for the over-identified model is more complicated. We will evaluate the posterior mean instead.

Following the suggestion of Berger & Bernardo (1992), we report the estimated frequentist risks of the MLE and those of the posterior mean in Table 1a (with the standard errors in parentheses). The table shows that the constrained MLE dominates unconstrained MLE and the posterior mean under Model 1 dominates the constrained MLE. The table also indicates that the improvement in the $\Phi$-related frequentist average loss by using the normal prior is due to variance reduction.
Table 1a: Estimated frequentist risks $\hat{E}_0\{L_1(\cdot, \Sigma)\}$ of $\Sigma$ and $\hat{E}_0\{L_2(\cdot, \Phi)\}$ of $\Phi$. The parameters are $p = 3$, $L = 1$, $T = 50$, and $\Psi$ given in (32).

<table>
<thead>
<tr>
<th>$\hat{\Sigma}_M$</th>
<th>$\hat{\Sigma}_{CM}$</th>
<th>$\hat{\Sigma}_1$</th>
<th>$\hat{\Sigma}_2$</th>
<th>$\hat{\Phi}_M$</th>
<th>$\hat{\Phi}_1$</th>
<th>$\hat{\Phi}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.260</td>
<td>0.162</td>
<td>0.161</td>
<td>0.165</td>
<td>3.843</td>
<td>1.570</td>
<td>1.537</td>
</tr>
<tr>
<td>(0.185)</td>
<td>(0.131)</td>
<td>(0.128)</td>
<td>(0.120)</td>
<td>(5.136)</td>
<td>(1.041)</td>
<td>(1.008)</td>
</tr>
</tbody>
</table>

Turning to hypothesis testing, we generate 1,000 samples and compute the Bayes factor using the approach of Chib (1995). We calculate the values of the likelihood, prior, and posterior at the MLE. We compute the Bayes factor $B_{12}$, with 1 being the true data generating over-identified VAR and 2 being just-identified VAR. A Bayes factor greater than unity, or logarithm of a Bayes factor being positive suggests the data are in favor of Model 1 over Model 2. The strength of evidence is given by the size of the Bayes factor. Kass & Raftery (1995) suggested a guideline in interpreting Bayes factors: If the natural logarithm of the Bayes factor is between 0 and 1, the evidence against Model 2 is “not worth more than a bare mention”; between 1 and 3, the evidence is “positive”; between 3 and 5, the evidence is “strong”; and above 5, the evidence is “very strong.”

With Model 1 as the data-generating model, we plot in Figure 1(a) and 1(b) the pdf and cdf of frequentist distributions of natural logarithms of Bayes factors $\log(B_{12})$ for the 1,000 data samples. Sampling errors contribute to some incidence of Bayes factors that support the “wrong model.” The vast majority of the 1,000 data samples produce positive log Bayes factors, supporting the “right model.” However, very few samples produce Bayes factors that show “strong” support for the data-generating Model 1 over Model 2. We compare the test results based on Bayes factor with those based on the Schwarz criterion. The logarithm of Bayes factors is positive in 980 out of 1,000 samples, but is larger than 3 only in 265 samples; while the Schwarz criterion is positive in 878 out of 1,000 samples, none is larger than 3.

To examine the performance of the methods of testing under a different setting, we now simulate another 1,000 samples using a just-identified Model 2 as the data-generating model, with parameter $\Psi$ defined as

$$
\Psi = \begin{pmatrix}
1 & 0.5 & 0.5 \\
0 & 1 & 0.5 \\
0 & 0 & 1
\end{pmatrix}.
$$

The frequentist averages of losses with samples simulated from Model 2 are given in Table 1b.

Table 1b: Estimated frequentist risks $\hat{E}_0\{L_1(\cdot, \Sigma)\}$ of $\Sigma$ and $\hat{E}_0\{L_2(\cdot, \Phi)\}$ of $\Phi$. The parameters are $p = 3$, $L = 1$, $T = 50$, and $\Psi$ given in (35).
We plot the cdf and the pdf of the natural logarithm of the Bayes factor \( B_{12} \) of the 1,000 samples and the Schwarz criterion in Figure 1 (c) and (d). In 882 out of 1,000 cases, the logarithm of the Bayes factor \( B_{12} \) is negative and 684 of them smaller than \(-3\), supporting Model 2 against Model 1. The Schwarz criterion is negative in 900 samples and is smaller than \(-3\) in 608 samples. With the just-identified model being the data-generating model, the magnitudes of the Bayes factors are larger compared to those when the over-identified model is the data-generating model. The absolute value of the logarithm of Bayes factor is on average larger than the Schwarz criterion. The Bayes factor offers stronger support for the true model than the Schwarz criterion does.

With either Model 1 or Model 2 being the data-generating model and the other being the alternative, the tests based on Bayes factors are not very decisive. This may be due to the fact that in Example 1 the difference between constrained MLE, unconstrained MLE, and Bayesian estimates is not substantial because there are only three variables in the VAR and there is a single constraint. In the following, we consider a VAR with 6 variables and 10 constraints on the \( \Psi \) matrix.

**Example 2** Consider a six-variable VAR with one lag and with parameters

\[
\begin{bmatrix}
1 & .5 & .5 & .5 & .5 & .5 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}, \begin{bmatrix}
1 & -5 & -5 & -5 & -5 \\
-5 & 1.25 & .25 & .25 & .25 \\
-5 & .25 & 1.25 & .25 & .25 & .25 \\
-5 & .25 & .25 & 1.25 & .25 & .25 \\
-5 & .25 & .25 & .25 & 1.25 & .25 \\
-5 & .25 & .25 & .25 & .25 & 1.25
\end{bmatrix}.
\]

As in Example 1, the VAR variables are assumed random walks with no drifts. With this parameter setting, the simulated frequentist average of unconstrained and constrained MLE of \( \Sigma \) matrix both show a bias toward zero.

\[
\hat{E}_0(\hat{\Sigma}_M) = \begin{bmatrix}
0.748 & -0.379 & -0.375 & -0.375 & -0.375 & -0.381 \\
-0.379 & 0.955 & 0.201 & 0.190 & 0.190 & 0.202 \\
-0.375 & 0.201 & 0.948 & 0.194 & 0.184 & 0.190 \\
-0.375 & 0.190 & 0.194 & 0.937 & 0.191 & 0.193 \\
-0.375 & 0.190 & 0.184 & 0.191 & 0.941 & 0.184 \\
-0.381 & 0.202 & 0.190 & 0.193 & 0.184 & 0.955
\end{bmatrix}.
\]
\[
\hat{\Sigma}(\hat{\Sigma}) = \begin{pmatrix}
0.870 & -0.441 & -0.436 & -0.437 & -0.436 & -0.443 \\
-0.441 & 1.111 & 0.223 & 0.222 & 0.220 & 0.226 \\
-0.436 & 0.223 & 1.102 & 0.220 & 0.219 & 0.222 \\
-0.437 & 0.222 & 0.220 & 1.090 & 0.220 & 0.223 \\
-0.436 & 0.220 & 0.219 & 0.220 & 1.094 & 0.221 \\
-0.443 & 0.226 & 0.222 & 0.223 & 0.221 & 1.110
\end{pmatrix}.
\]

The posterior mean of \(\Sigma\) under the over-identified model with the correct restrictions on \(\Psi\) is
\[
\hat{\Sigma}(\hat{\Sigma}^1) = \begin{pmatrix}
0.917 & -0.465 & -0.461 & -0.461 & -0.461 & -0.466 \\
-0.465 & 1.055 & 0.235 & 0.235 & 0.233 & 0.239 \\
-0.461 & 0.235 & 1.048 & 0.233 & 0.232 & 0.235 \\
-0.461 & 0.235 & 0.233 & 1.036 & 0.233 & 0.235 \\
-0.461 & 0.233 & 0.232 & 0.233 & 1.041 & 0.233 \\
-0.466 & 0.239 & 0.235 & 0.235 & 0.233 & 1.054
\end{pmatrix}.
\]

The posterior mean of \(\Sigma\) under the just-identified model is
\[
\hat{\Sigma}(\hat{\Sigma}^2) = \begin{pmatrix}
0.917 & -0.465 & -0.460 & -0.461 & -0.461 & -0.466 \\
-0.465 & 1.188 & 0.248 & 0.234 & 0.233 & 0.249 \\
-0.460 & 0.248 & 1.203 & 0.240 & 0.227 & 0.234 \\
-0.461 & 0.234 & 0.240 & 1.210 & 0.237 & 0.237 \\
-0.461 & 0.233 & 0.227 & 0.237 & 1.237 & 0.226 \\
-0.466 & 0.249 & 0.234 & 0.237 & 0.226 & 1.273
\end{pmatrix}.
\]

Table 2a reports the frequentist average losses of competing estimators. It shows that the posterior means under the competing models are better than the unconstrained MLE but are not better than the constrained MLE. There are two reasons for this result. First, as we observed earlier, in the presence of constraints in \(\Psi\), the posterior mean of \(\Sigma\) is in general not the Bayesian estimator under the entropy loss. Second, the Bayesian estimator minimizes the posterior loss over the entire parameter space, rather than at the true parameter value.

Table 2a: Estimated frequentist risks \(\hat{\Sigma}(L(\cdot, \Sigma))\) of \(\hat{\Sigma}\) and \(\hat{\Phi}(L(\cdot, \Phi))\) of \(\hat{\Phi}\). The parameters are \(p = 6\), \(L = 1\), \(T = 50\), and \(\Psi\) given in (36).

<table>
<thead>
<tr>
<th>(\hat{\Sigma}_M)</th>
<th>(\hat{\Sigma}_{CM})</th>
<th>(\hat{\Sigma}_1)</th>
<th>(\hat{\Sigma}_2)</th>
<th>(\hat{\Phi}_M)</th>
<th>(\hat{\Phi}_1)</th>
<th>(\hat{\Phi}_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.403</td>
<td>0.455</td>
<td>0.561</td>
<td>0.707</td>
<td>18.251</td>
<td>4.832</td>
<td>4.615</td>
</tr>
<tr>
<td>(0.544)</td>
<td>(0.233)</td>
<td>(0.269)</td>
<td>(0.306)</td>
<td>(16.830)</td>
<td>(1.673)</td>
<td>(1.537)</td>
</tr>
</tbody>
</table>

The Bayes factor is positive in 990 out of 1,000 samples, and is larger than 3 in 980 samples, the Schwarz criterion is positive in 998 out of 1,000 samples and is larger then 3 in 981 samples. The high accuracy rates are in sharp contrast to Example 1. In Example 2, the over-identified model 1 is
substantially different from the just-identified Model 2. The Bayes factors easily reject the “wrong” model. Figure 1 (e) and (f) plot the natural logarithms of the Bayes factors of the 1,000 samples.

The Bayesian MCMC computation yields posterior of each elements of $\Sigma$ and $\Phi$. The posterior provide a better guide for finite sample inference than asymptotic theory does. The limit in space prevents us from reporting detailed results. In Figure 2 we plot the posterior mean of a single element $\sigma_{11}$ over the 1,000 samples. There is not much difference between the frequentist distributions of posterior means of the parameter $\sigma_{11}$ under Models 1 and 2 because both are generated by similar gamma distributions. The posterior means are quite different from the MLE. They show smaller downward bias (from the true value of unity) than the MLE and are more tightly distributed.

The posterior mean of $\Psi$ is influenced by the hyper-parameters of the priors. When we apply an improper gamma prior with $a_i=i-p$, for $i = 1, \ldots, p$, the diagonal elements of $\Sigma$ are larger and closer to the true parameter of 1.25. The frequentist average losses with respect to $\Sigma$ are also smaller.

Table 2b: Estimated frequentist risks $\hat{\mathbb{E}}_0\{L_1(\cdot, \Sigma)\}$ of $\hat{\Sigma}$ and $\hat{\mathbb{E}}_0\{L_2(\cdot, \Phi)\}$ of $\hat{\Phi}$. The parameters are $p = 6$, $L = 1$, $T = 50$, and $\Psi$ given in (36). In gamma prior for $\psi_{ii}$, $a_i=i-p$, $b_i=0.1$ for $i = 1, \ldots, 6$.

<table>
<thead>
<tr>
<th>$\hat{\Sigma}_M$</th>
<th>$\hat{\Sigma}_{CM}$</th>
<th>$\hat{\Sigma}_1$</th>
<th>$\hat{\Sigma}_2$</th>
<th>$\hat{\Phi}_M$</th>
<th>$\hat{\Phi}_1$</th>
<th>$\hat{\Phi}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.406</td>
<td>0.464</td>
<td>0.445</td>
<td>0.637</td>
<td>19.030</td>
<td>4.466</td>
<td>4.232</td>
</tr>
<tr>
<td>(0.525)</td>
<td>(0.240)</td>
<td>(0.219)</td>
<td>(0.238)</td>
<td>(17.590)</td>
<td>(1.511)</td>
<td>(1.404)</td>
</tr>
</tbody>
</table>

We conduct an exercise similar to that in Example 1 by generating data from the just-identified model (Model 2) with the following parameters for $\Psi$:

$$\Psi = \begin{pmatrix}
1 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\
0 & 1 & 0.5 & 0.5 & 0.5 & 0.5 \\
0 & 0 & 1 & 0.5 & 0.5 & 0.5 \\
0 & 0 & 0 & 1 & 0.5 & 0.5 \\
0 & 0 & 0 & 0 & 1 & 0.5 \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}.$$  \hfill (37)

Table 2c: Estimated frequentist risks $\hat{\mathbb{E}}_0\{L_1(\cdot, \Sigma)\}$ of $\hat{\Sigma}$ and $\hat{\mathbb{E}}_0\{L_2(\cdot, \Phi)\}$ of $\hat{\Phi}$. The parameters are $p = 6$, $L = 1$, $T = 50$, and $\Psi$ given in (37).

<table>
<thead>
<tr>
<th>$\hat{\Sigma}_M$</th>
<th>$\hat{\Sigma}_{CM}$</th>
<th>$\hat{\Sigma}_1$</th>
<th>$\hat{\Sigma}_2$</th>
<th>$\hat{\Phi}_M$</th>
<th>$\hat{\Phi}_1$</th>
<th>$\hat{\Phi}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.403</td>
<td>1.521</td>
<td>1.630</td>
<td>0.702</td>
<td>19.281</td>
<td>5.361</td>
<td>4.753</td>
</tr>
<tr>
<td>(0.544)</td>
<td>(0.250)</td>
<td>(0.292)</td>
<td>(0.303)</td>
<td>(17.280)</td>
<td>(1.896)</td>
<td>(1.503)</td>
</tr>
</tbody>
</table>
The posterior mean of $\Sigma$ under Model 2 dominates other estimates. We compute the Bayes factor $B_{12}$ using data generated from Model 2. Hypothesis testing based on the Bayes factors shows that for 880 out of 1,000 samples the natural logarithm of $B_{12}$ is smaller than $-3$, indicating that data favor Model 2. The frequentist pdf and cdf are plotted in Figure 1 (g) and (h). In comparison with the Schwarz criterion, the power of Bayes factor tests is stronger when the data are generated by the constrained (over-identified) model.

The relative performance of the Bayes factor vs. the Schwarz criterion depends on the sample size. When the sample size is reduced to 30 from 50, the Bayes factor becomes much different from the Schwarz criterion. In Figure 1 (i) and (j), the Bayes factors in 1,000 samples are larger than the Schwarz criterion. With the over-identified Model 1 being the data-generating model, in 930 out of 1000 samples the natural logarithm of $B_{12}$ is positive and in 880 samples is larger than 3, indicating that the data are in favor of Model 1. The Schwarz criterion is positive in 885 out of 1,000 samples and in 750 of them is larger than 3. With reduced sample size, the Schwarz criterion becomes less precise in approximating the logarithm of the Bayes factor and less effective in selecting the true model.

In summary, our numerical examples show that the relative performance of the testing approaches depends on the sample size and the similarity of the competing models. The examples considered in this paper involve moderate numbers of parameters. When identified VARs models with large numbers of parameters are subjects of testing, researchers should be aware of the fact that the Schwarz criterion should not be considered a close substitute for the Bayes factor. In such a case, as the following application indicates, the ability to compute the Bayes factor is a desirable property for an inferential approach.

6 An Empirical Application to Employment Growth

In search of fundamental causes of business cycles, economists have long recognized that economy-wide recessions may result from aggregate shocks that affect all sectors and regions of the economy as well as sector- and region-specific shocks (see Vining, 1947; and more recently Long & Plosser, 1987.) In the following, we apply the inferential techniques developed in this paper to a battery of identification schemes of VARs for employment growth.

Fluctuations in regional and industry employment during business cycles have been the subject of a rich empirical literature. Altonji & Ham (1998) examined 54 annual data series of 6 Canadian
provinces and 9 industries for regional and industry-specific factors in explaining employment growth. The large number of variables involved and limited available data necessitate numerous restrictions placed in their model. Norrbin & Schlagenhauf (1988) adopted a similar approach for U.S. regions and industries. Their use of quarterly instead of annual data affords more flexibility in their model than that of Altonji & Ham (1998), but it is still infeasible to apply VAR to the problem. Clark (1998) constructed an identified VAR to explore the roles of the U.S. national, state, and industry-specific shocks. He explored the 17 series employment data aggregated by 9 regions and by 8 industries which are far fewer than the corresponding 72 disaggregated series used by Norrbin & Schlagenhauf (1988), making it possible to estimate a VAR. In the study by Clark, interactions of employment growth aggregated by industry and region serve as the base for analysis of the source of volatility in employment growth.

The following analysis is different from the existing literature in two aspects. First, we use identified VARs on disaggregate data of employment growth in three sectors (manufacture, construction, and service) and in three states (Illinois, Missouri, and Arkansas). Our model is less restrictive than the ones used in the literature because we allow each employment series to be affected contemporaneously by macroeconomic, sector-specific, and cross-state and cross-sector shocks. Second and more importantly, we apply the Bayesian approach developed in this paper for inference and hypothesis testing. The literature cited above relies on asymptotic distribution of parameters and likelihood ratio tests for model selection. With macroeconomic and sector- and state-specific variables in the model, a VAR of this literature can easily involve a hundred parameters, rendering asymptotic theory unsuitable for finite sample inferences. The Bayesian VAR method in the proposed study is perfectly suited for such a problem.

We use consumption expenditure instead of industrial production as the state variable of macroeconomy because consumption decision reflects more information than the health of the industrial sector; moreover, it embodies expectation of economic agents on the future state of the economy. For this reason we place consumption on the top of the order for the VAR variables. We include in the model national employment growth in manufacture, construction, and service sectors. The ordering of contemporaneous effect across sectors is manufacture, construction, and service, reflecting the notion that manufacture jobs are most sensitive to business cycles. The order of sectors is the same for national and state-wide employment growth. The ordering across states is Illinois, Missouri, and Arkansas, under the assumption that the growth of job markets in larger states is asymmetrically influential on the neighboring states. The VAR variables (and the corresponding
residuals in the VAR) are represented as follows: \( c = \) per capita consumption expenditure growth, \( u_m = \) U.S. manufacture employment growth, \( u_c = \) U.S. construction employment growth, \( u_s = \) U.S. service employment growth, \( i_m = \) Illinois manufacture employment growth, \( i_c = \) Illinois construction employment growth, \( i_s = \) Illinois service employment growth, \( m_m = \) Missouri manufacture employment growth, \( m_c = \) Missouri construction employment growth, \( m_s = \) Missouri service employment growth, \( a_m = \) Arkansas manufacture employment growth, \( a_c = \) Arkansas construction employment growth, and \( a_s = \) Arkansas service employment growth.

All data series are collected from the Federal Reserve Bank of St Louis database and are measured in percentage terms. The monthly data sample is from 1982:1 to 2002:12, with 252 observations. We set the lag of the VAR at 1. We could make the lag length \( L \) part of the hypothesis testing exercise but choose instead to fix the lag length and focus on competing models of \( \Psi \) because the latter is the base for model identification. The correlation matrix of the 13 data series in the order described above (with the diagonal elements being standard deviations) is

\[
\begin{pmatrix}
.543 & .053 & .100 & .109 & .026 & -.029 & .076 & .052 & .010 & -.018 & .039 & .006 & -.050 \\
.006 & .016 & .351 & .073 & .031 & .302 & .094 & .120 & .450 & .091 & .081 & 1.56 & .256 \\
\end{pmatrix}
\]

The employment growth data series exhibit positive pairwise correlations across states and across sectors. It is obvious that one should not infer sector-specific effect from pairwise correlation of variables across sectors, because the correlation may stem from macroeconomic effects on the sectors. The identification of structural shocks from VAR residuals \( \epsilon \) in a model that allows for the cross-industry influence is \( \epsilon_t A_0 = u_t \). The prior mean for VAR regression coefficient \( B_1 \) is set at zero instead of identity because the data used are growth rates instead of levels of employment and the prior covariance is set at a very disperse \( 10 \times I \). The priors for the elements in \( \Psi \) (\( A_0 \)) are the same as that used in the numerical examples of the previous section.

We consider six competing identifying schemes. First, employment growth is affected by sectoral and macroeconomic shocks. Under this theory, all industries are affected contemporaneously by
economy-wide shocks, but there is no contemporaneous cross-industry or cross-state effect. The residuals of equations corresponding to industrial variables are correlated only to the extent through their correlation with the macroeconomic and industry-specific shocks. We call this identification Model 1. For the next model, we assume that in addition to the effects in Model 1, business cycles are also caused by within-state cross-sector shocks (e.g., employment growth in service in Missouri in a given month affects manufacture employment growth in Missouri). We call this identification Model 2. Next, in Model 3, in addition to the effects in Model 2, we further allow for within-industry but cross-state effect (i.e., the manufacture employment growth in Illinois is allowed to affect that in Missouri contemporaneously). In Model 4, we add contemporaneous cross-sector effects of neighboring states. In Model 5, we include national cross-sector effect to all states. Lastly, in Model 6, we allow for all cross state-cross industry effects, making the $\Psi (A_0)$ matrix just-identified. In summary, the following contemporaneous effects are allowed in the competing models.

- **Model 1**: Macro effects and sector effects.
- **Model 2**: Effects in Model 1 plus within-state cross-sector effects.
- **Model 3**: Effects in Model 2 plus within-sector-cross state effects.
- **Model 4**: Effects in Model 3 plus cross-sector of neighboring state effects.
- **Model 5**: Effects in Model 4 plus cross-sector national effects.
- **Model 6**: Includes all Macro, Industry, within state-cross industry, within industry-cross state, and cross state-cross industry effects.

Clearly, Model 6 is least restrictive and Model 1 is most restrictive. These models correspond to the following constraints on the $A_0$ matrix mapping the reduced-form residuals $\epsilon$ to structural shocks $u$. "*" represents parameters in $A_0$ unrestricted in all models. '1' represents parameters unrestricted in Model 1 in addition to parameters represented by '*'. '2' represents parameters unrestricted in Model 2 in addition to parameters represented by '*' and '1'. ... Finally, '6' represents parameters unrestricted in model 6 in addition to parameters represented by '*', '1', '2', '3', '4' and '5'.
To test the competing hypothesis, we calculate Bayes factors and the Schwarz criterion for Models 1 to 5 against Model 6. The results are summarized in Table 3.

Table 3. Results of Hypothesis Testing: $\log B_{i6}$ and $S_{i6}$, $i=1,2,3,4,5$

<table>
<thead>
<tr>
<th>Test</th>
<th>$i = 1$</th>
<th>$i = 2$</th>
<th>$i = 3$</th>
<th>$i = 4$</th>
<th>$i = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\log B_{i6}$</td>
<td>$-211.43$</td>
<td>$-202.61$</td>
<td>$-188.87$</td>
<td>$-181.75$</td>
<td>$-2.57$</td>
</tr>
<tr>
<td>$S_{i6}$</td>
<td>$56.64$</td>
<td>$56.92$</td>
<td>$73.20$</td>
<td>$46.68$</td>
<td>$8.48$</td>
</tr>
</tbody>
</table>

The Bayes factor tests recommend the just-identified Model 6. On the contrary, the Schwarz criterion suggests that every over-identified model considered is better than the just-identified model. In particular, the Schwarz criterion tests support Model 3. The result highlights the difference in the hypothesis testing approach.

We now explore implications of the model selection result. In applications of VAR models, macroeconomists often find impulse responses to structural innovations more economically intuitive than the VAR coefficients. A covariance stationary VAR implies $y_t = E_0 y_t + \sum_{j=0}^{t-1} \epsilon_{t-j} H_j$, where $H_0$ is the $p$ by $p$ identity matrix, and the impulse responses of $y_t$ to a shock $\epsilon_{t-j}$ from $j$ periods earlier is $H_j = \sum_{i=1}^{j} B_i H_{j-i}$, where $B_i = 0$ for $i$ larger than $L$. As discussed earlier, identification of the structural innovation vector $u_t$ from the VAR residual vector $\epsilon_t$ can be achieved by setting
\( \mathbf{u}_t' = \mathbf{c}_t \mathbf{A}_0 \). The impulse responses of \( \mathbf{y}_t \) to structural shocks \( \mathbf{u}_{t-j} \) are given by \( \mathbf{Z}_j = \mathbf{A}_0^{-1} \mathbf{H}_j \). By definition, impulse responses are nonlinear functions of \( (\mathbf{\Phi}, \mathbf{A}_0) \). The nonlinearity makes it difficult to derive frequentist inference but does not pose difficulties for Bayesian computations as long as posteriors of \( (\mathbf{\Phi}, \mathbf{A}_0) \) are available. The most common practice of reporting impulse responses by VAR users is based on RATS or other noninformative priors with a Cholesky decomposition of the \( \mathbf{\Sigma} \) matrix. This practice does not allow incorporation of economic theory with respect to specific elements of the \( \mathbf{A}_0 \) matrix. The Bayesian procedure proposed by Sims & Zha (1998,1999) was largely motivated to solve the problem. We will compare Bayesian inferences of the impulse responses based on competing restrictions on \( \mathbf{A}_0 \).

The impulse responses based on alternative identifying schemes shed light on how employment growth responds to macroeconomic as well as sector- and state-specific shocks. The limit in space does not permit us to report all of them. We focus on the posterior of impulse response of Illinois construction employment growth to a unit shock in national construction growth (representing sectoral effect) and a unit shock in Missouri manufacturer employment growth (representing cross-state and cross-sector effect), and that of Illinois service employment growth to a unit shock in Illinois construction employment growth (representing within-state and cross-sector effect) under Model 1 and Model 6. We compute the corresponding impulse based on the MLE of \( \mathbf{\Phi} \) and Cholesky decomposition of MLE of \( \mathbf{\Sigma} \) and plot it against the posterior mean and 10 and 90 percentile posterior bands under each model specification.

Figure 3 shows that the impulse responses exhibit similar shapes under competing models. However, the magnitude is quantitatively different. Based on the MLE, Illinois construction employment growth increases by .05% one month after a unit shock in Missouri manufacturer employment growth. Under Model 6, the posterior mean of the impulse response almost overlaps with that of MLE. With weaker restrictions, the posterior mean of the response is a much larger .09%. The difference between employment growth of .05% per month and .09% per month is about 5% per annum. Under Models 1, the response of Illinois service employment growth one month after a unit shock in construction employment growth is close to zero. Under Model 6, it is about .03%. Note that there is little difference in posteriors of VAR regression coefficients \( \mathbf{\Phi} \) under the competing assumptions on the \( \mathbf{A}_0 \) matrix. The difference in impulse responses stems from the difference in identification of state- and industry-specific shocks. The earlier results of Bayes factors suggest that analysis based on impulse responses based on Model 6 is advisable.
7 Concluding Remarks

In this paper, we propose a flexible prior setting that allows for incorporation of element-wise information to a triangular decomposition of precision matrix, with certain off-diagonal elements of the triangular decomposition \((A_0)\) being zero. We show that the prior setting is perfectly suited for Bayesian analysis of normalized identified VAR models, where component-wise restrictions on the \(A_0\) matrix are justified by economic theory. We show that the conditional posteriors of the elements of the \(A_0\) matrix are standard distributions, which allows for implementation of conventional MCMC posterior simulation and hypothesis testing using the algorithm outlined by Chib (1995). Our simulations show that the Schwarz criterion deviates substantially from the Bayes factor when the sample size is small and is not as effective in selecting the correct model. A VAR analysis of macroeconomic, state- and sector-specific shocks to employment growth suggests that the tests based on Bayes factors yield different results than tests based on the Schwarz criterion.

The present study can be extended in several ways. First, in the existing identified Bayesian VAR literature and in our discussion so far, all identification schemes impose restrictions that certain elements of \(A_0\) are zero on the grounds of formal or informal economic theory. The exercise can be viewed as an approach for model selection. But regarding the identification schemes, the common practice is to select one or at most a few competing models. An implicit assumption is that all relevant models are included in the pool of candidates for selection. For some economic problems, theoretical development does not warrant such confidence. A more prudent approach is to avoid imposing zero restrictions on any coefficient (that rule out certain models a priori) and instead to develop a method that allows comparison of all possible models. This alternative approach gives the data much greater power in selecting models. The key of the data-driven model selection is to be able to compare a large number of models. Every off-diagonal element in \(A_0\) can potentially be zero, which brings the total number of possible models to \(2^{p(p-1)/2}\). Even for a small \(p\), the amount of computation is prohibitive. For selection of variables in univariate models, George & McCulloch (1993) proposed a Bayesian MCMC stochastic search algorithm that greatly reduced the amount of computation. Extending their algorithm to VAR is one of our current research projects. Second, under more general prior settings, the conditional posteriors of identified VARs may not be standard distributions. In this case, posterior simulations may be done through Metropolis steps, but the algorithm of Chib (1995) for estimation of marginal likelihood is no longer applicable. Recently, Chib & Jeliazkov (2002) developed a method to estimate marginal likelihood from Metropolis-Hastings sampling without knowing the normalizing constant of the M-H proposal density. Posterior simulations via Metropolis-
Hastings algorithms require more lengthy Markov chains. An interesting research topic is to explore the efficiency of M-H simulations and conduct hypothesis testing using the Chib & Jeliazkov (2002) method for identified VARs.

References


Figure 1: Frequentist histograms and cdfs of the logarithm of the Bayes factor $\log B_{12}$ (solid line) and the Schwarz criterion $S_{12}$ (dotted line) for Examples 1 and 2. (a) and (b): Example 1, $p=3$, $T=50$, Model 1 is the true model; (c) and (d): Example 1, $p=3$, $T=50$, Model 2 is the true model; (e) and (f): Example 2, $p=6$, $T=50$, Model 1 is the true model; (g) and (h): Example 2, $p=6$, $T=50$, Model 2 is the true model; (i) and (j): Example 2, $p=6$, $T=30$, Model 1 is the true model.
Figure 2: Frequentist histograms of $\sigma_{11}$ in Example 2 (with data generated by Model 1): (a) MLE; (b) Model 1; (c) Model 2.
Figure 3: (a)(b)(c) are impulse responses under Model 1, (d)(e)(f) are under Model 6. Solid line=posterior mean, dashed line=MLE, dotted line=10 percentile, dotted-dashed line= 90 percentile.