Bayesian Stochastic Search for VAR Model Restrictions
January 2005, Revised November 2006 [1]

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

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

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

Abstract

We propose a Bayesian stochastic search approach to selecting restrictions for Vector Autoregressive (VAR) models. For this purpose, we develop a Markov Chain Monte Carlo (MCMC) algorithm that visits high posterior probability restrictions on the elements of both the VAR regression coefficients and the error variance matrix. Numerical simulations show that stochastic search based on this algorithm can be effective at both selecting a satisfactory model and improving forecasting performance. To illustrate the potential of our approach, we apply our stochastic search to VAR modelling of inflation transmission from Producer Price Index (PPI) components to the Consumer Price Index (CPI).

KEY WORDS: Bayesian VAR, Stochastic Search, Markov Chain Monte Carlo.

JEL CLASSIFICATIONS: C11, C15, C32.

[1]George's research was supported by NSF grants DMS-0130819 and DMS-0605102. Ni's research was supported by a grant from the MU Research Council and by Hong Kong Research Grant Council grant CUHK4128/03H. Sun's research was supported by the National Science Foundation grants SES-0095919 and SES-0351523, and NIH grant R01-CA100760. We are grateful for the generous suggestions of the reviewers and for the helpful comments received when earlier versions of the paper were presented at the Federal Reserve Bank of Kansas City, the 2004 Joint Statistical Meeting, and the 2004 Meeting of the Midwest Econometrics Group.

1 Introduction

Econometricians constantly face the challenge of building models that are both general enough to analyze complicated problems and restrictive enough to yield sharp finite sample inferences. This task can sometimes be accomplished by using an overparametrized model, namely a model with more parameters than can be reasonably estimated with the data at hand, and then seeking plausible restrictions that allow adequate inference. We propose a Bayesian stochastic search approach for selecting such restrictions for Vector Autoregression (VAR) modeling settings.

In a VAR model, the variable of interest $\boldsymbol{y}_t$, a p-dimensional column vector, is determined by

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

for $t = 1, \cdots, T$, where the exogenous variable $\boldsymbol{z}_t$ is an h-dimensional vector, the lag L is a known positive integer, the regression coefficients $\boldsymbol{C}$ and $\boldsymbol{A}_j$ are $h \times p$ and $p \times p$ unknown matrices, the error terms $\boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_T$ are independent identically distributed $N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$, and the covariance matrix $\boldsymbol{\Sigma}$ is an unknown $p \times p$ positive definite matrix. VAR models have been widely used for characterizing multivariate time series data and for macroeconomic forecasting. The absence of restrictions on the regression coefficients and the covariance matrix in (1) results in a very large number of parameters relative to the data at hand. With a limited number of observations such "over-parameterization" typically has adverse consequences on the precision of inference and the reliability of prediction.

To obtain restrictions on over-parameterized models, several approaches based on researchers' knowledge have been employed in the econometrics literature. For example, in traditional simultaneous equations (SE) settings, economists have used economic theory to justify parameter restrictions, restrictions often considered "incredible" and yielding unsatisfactory results, (see Sims (1980)). In fact, the checkered performance of large scale SE models has led to the development and popularity of VAR models.

For VAR modeling, the Bayesian approach can be used to incorporate knowledge about parameter values via prior distributions. Indeed, information about the pattern of macroeconomic data and on structural economic models has been utilized for VAR prior elicitation and model restrictions. For example, Doan et al. (1984) suggested a Minnesota prior that shrinks the VAR parameters towards a random walk model. Although such priors may be suitable in some cases, their effectiveness may sometimes be limited. McNees (1986) compared the forecasting performance of Litterman's Bayesian VAR with some popular SE models across various small sample macroeconomic problems and found that no single approach was superior across all settings.

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)) have proposed using identified VARs of the form

$$\boldsymbol{y}_t'\boldsymbol{A}_0 = \boldsymbol{z}_t'\boldsymbol{C}\boldsymbol{A}_0 + \sum_{j=1}^{L} \boldsymbol{y}_{t-j}'\boldsymbol{A}_j\boldsymbol{A}_0 + \boldsymbol{\epsilon}_t'\boldsymbol{A}_0, \tag{2}$$

where $\boldsymbol{A}_0$ is such that the covariance of structural shocks $\boldsymbol{\epsilon}_t'\boldsymbol{A}_0$ is the identity $\boldsymbol{I}$ and so satisfies

$$\boldsymbol{\Sigma} = \boldsymbol{A}_0^{-1\prime}\boldsymbol{A}_0^{-1}. \tag{3}$$

Such identified Bayesian VARs set particular elements of the decomposition of the precision matrix to zero on the ground of formal or informal economic theory (see Sims (1986) for justifications of such approaches). In contrast, SE models restrict $\boldsymbol{A}_0$ and $\boldsymbol{A}_j$ but not the error covariance structure. However, both approaches are based on an implicit assumption that the relevant restrictions are known. At least for some economic problems, current theoretical knowledge does not warrant such confidence.

An alternative to restriction preselection in over-parametrized models is to use restrictions that are supported by the data itself. This may be accomplished by using a comprehensive hierarchical model that treats each possible set of restrictions as a distinct submodel, and then uses priors to describe the uncertainty across all submodels. Under such a setup, the posterior distribution will then increase the prior weights on those restrictions that are best supported by the data. Although it is typically not feasible to exhaustively calculate all submodel posterior probabilities, many of the higher posterior probability submodels can be found by stochastic search using MCMC algorithms. This is precisely the approach we propose for the selection of VAR model restrictions. This approach is fundamentally different from previous VAR modeling approaches because it does not apriori rule out submodels (of the VAR under consideration). Instead, it allows for the comparison of submodels based on the data.

When the number of parameters in a VAR model for macroeconomic data is large relative to the number of observations, asymptotic theory will typically be an unreliable guide for the finite sample estimation properties. An advantage of the Bayesian approach is that it produces finite sample inferences on the parameters of interest while simultaneously allowing for both model and parameter uncertainty. This is especially important when model uncertainty is a more serious concern than parameter uncertainty, as is often the case.

Models such as (2) are useful for testing existing theory and for discovering patterns in data that

may motivate new theory. In many applications, we expect our approach to suggest that many, if not most, coefficients are close to zero, thereby enabling researchers to focus on the more realistic submodels. And the relative probabilities of the selected submodels will provide them with further guidance. As will be seen, our approach allows the flexibility to consider either $\boldsymbol{A}_0$ restrictions or VAR regression coefficient restrictions, or both. In the VAR framework, Hsiao (1981) proposed a frequentist procedure to reduce the lag length of one of the variables in a bi-variate VAR. Such lag selection is easily accommodated as a special case of our approach.

The main challenge for restriction selection in our setting lies in the large number of potential submodels. For example, the variable selection problem for univariate regression with h explanatory variables involves comparing 2^h competing submodels, an infeasible computation even for moderate h. In this context, George & McCulloch (1993) proposed SSVS (Stochastic Search Variable Selection), a Bayesian MCMC stochastic search algorithm that greatly reduced the amount of computation - see George (2000) for a survey of recent developments in the area. Our approach, an extension of SSVS for VAR restriction selection, faces additional challenges. First, it involves consideration of all possible restrictions on the off-diagonal elements of $\boldsymbol{A}_0$. Second, for a model with p endogenous variables, h exogenous variables (including a constant) and L lags, there are $2^{(h+Lp)p+p(p-1)/2}$ competing submodels, which can be many more than the 2^h univariate regression submodels. Furthermore, serial correlation among the VAR variables biases the MLE in finite samples, making selection even more difficult. The key to tackling these restriction selection challenges is the development of priors and algorithms that allow for efficient simulation of MCMC samples from the posterior.

To illustrate that our approach has the potential to select effective restrictions in settings where a VAR model is adequate, we simulate numerical examples in which we generate data from known submodels. We then compare the performance on this simulated data of the MLE, the benchmark Bayesian VARs without restriction selection (i.e., conventional Bayesian VARs), and the Bayesian VARs using our restriction selection approach. In some of these examples, our approach achieves a very high success rate in selecting the exactly correct data-generating model. And even when the correct model is not selected, our approach still obtained improvements in terms of the MSE of point forecasts. We also find that the Kullback-Liebler distance is substantial between the predictive density averaged over the visited models and that of the benchmark VAR.

SSVS is but one approach in a voluminous theoretical and empirical statistical literature on Bayesian model selection, starting with Jeffreys (1961) who proposed the use of posterior odds for model selection and the use of correction factors to mitigate the dangers of chance selection with

multiple alternatives. References to many of the vast number of developments since Jeffreys' work, can be found in Zellner et al. (2001) and Clyde & George (2004).

It should be emphasized that our proposed restriction selection approach for overparametrized VARs is only meant for those settings where a restricted VAR would ultimately be a satisfactory model for the data. With their strong assumptions of linearity, fixed parameters, constant error term covariance and normality of errors, VAR models may be inadequate for economic problems where flexibilities such as nonlinearity and time varying parameters are needed. Indeed, the forecasting performance of VARs has not been so good on some economic problems as shown McNees (1986). At the very least, the adequacy of a VAR model should be checked at the outset. As pointed out by Zellner & Palm (1975), the adequacy of a formulation such as (2) can be determined by checking the impled forms of the transfer functions and final equations using the data, a generalization of the well-known univariate Box-Jenkins model identification procedures. An alternative developed by Zellner and Palm is the SEMTSA (Structural Econometric Modelling Time Series Analysis) approach, which connects the univariate time series ARIMA (Box-Jenkins) model and multivariate structural models for estimation and forecast using cross-country and cross-sectoral data. The SEMTSA approach starts with a multivariate ARMA model that include exogenous and endogenous variables and restricts the structure of the individual process of the endogenous variables based on economic theories. Such a macroeconomic framework includes additional features such as indicators that are good predictors of turning points of business cycles and building blocks based on Marshallian equilibrium setting of disaggregated sectoral data. As shown in Zellner & Palm (2004), the SEMTSA framework can deliver better forecasting performance than unrestricted VARs.

Although we have used the VAR framework to illustrate our approach, the main ideas of Bayesian stochastic search selection restriction can be extended to more complicated fameworks such as SEMTSA. Although submodel prior probabilities in structural model settings might be more grounded in economic theory than in VAR settings, for example a large prior probability of including a price parameter in a demand equation, the technical aspects of our approach would be the same. Indeed, the MCMC stochastic search for high probability restrictions in a SEMTSA formulation would proceed in the same was as in our VAR formulation. Though straightforward in principle, such extensions to more sophisticated frameworks is left for future work.

The paper is organized as follows. Section 2 defines notation and our hierarchical priors. Section 3 derives conditional posterior distributions and lays out the Bayesian algorithms. Section 4 reports simulation results of numerical examples. Section 5 applies our method to a real problem. Section 6

offers concluding remarks.

2 The Model and Prior

2.1 Likelihood and decomposition of the precision matrix

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

$$\boldsymbol{Y} \;\; = \;\; \boldsymbol{X}\boldsymbol{\Phi} + \boldsymbol{\epsilon}, \tag{4}$$

where

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{y}_1' \\ \vdots \\ \boldsymbol{y}_T' \end{pmatrix}, \; \boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1' \\ \vdots \\ \boldsymbol{x}_T' \end{pmatrix}, \; \boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{C} \\ \boldsymbol{A}_1 \\ \vdots \\ \boldsymbol{A}_L \end{pmatrix}, \; \boldsymbol{\epsilon} = \begin{pmatrix} \boldsymbol{\epsilon}_1' \\ \vdots \\ \boldsymbol{\epsilon}_T' \end{pmatrix}. \tag{5}$$

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

Conditional on the initial value $\boldsymbol{x}_1$, which we assume throughout to be available, the likelihood function of $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$ is

$$f(\boldsymbol{Y} \mid \boldsymbol{\Phi}, \boldsymbol{\Sigma}) \;\; \propto \;\; |\boldsymbol{\Sigma}|^{-T/2} etr\Big\{-\frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\Phi})\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\Phi})'\Big\}, \tag{6}$$

where $etr(\boldsymbol{A}) \equiv exp(Trace(\boldsymbol{A}))$. Letting

$$\boldsymbol{S}(\boldsymbol{\Phi}) = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\Phi})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\Phi}), \tag{7}$$

a commonly used estimate of $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$ is the maximum likelihood estimate (MLE):

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

Here $\boldsymbol{S}(\widehat{\boldsymbol{\Phi}}_M)$ is the sum of squares of MLE residuals. We assume that the sample size T is large enough so that the MLEs of $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$ exist with probability one.

Identifying restrictions for a VAR model may be obtained by restricting the elements of $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$. Restrictions on $\boldsymbol{\Sigma}$ are typically obtained by restricting the elements of a re-parametrization of $\boldsymbol{\Sigma}$, and this is how we shall proceed. Motivated by condition (3) of the identified VAR, our focus will be on restricting the elements of the $p \times p$ upper-triangular matrix $\boldsymbol{\Psi}$ satisfying

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Psi}\boldsymbol{\Psi}'. \tag{9}$$

Note that the elements of the precision matrix $\mathbf{\Sigma}^{-1}$ ($\mathbf{\Sigma}^{-1} = (\sigma^{ij})$) are natural quantities of interest because $-\sigma^{ij}/\sqrt{\sigma^{ii}\sigma^{jj}}$ 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. Letting ψ_{ij} denote the $(i,j)th$ entry of $\mathbf{\Psi}$, we assume without loss of generality that the diagonal elements satisfy $\psi_{ii} > 0$. Such a matrix $\mathbf{\Psi}$ without restriction on ψ_{ij} is called normalized and just-identified by Sims & Zha (1998).

Substantial progress has been made on Bayesian VAR modeling by imposing restrictions on the $\mathbf{\Psi}$ matrix. Sims & Zha (1998), and Waggoner & Zha (2003) proposed Bayesian analysis based on normal priors on $\mathbf{\Psi}$ and $\mathbf{\Phi}$. In each study, the authors assumed a given form for $\mathbf{\Psi}$ disallowing for the possibility that alternative models could be supported by the data. Further, their approach did not restrict attention to normalized $\mathbf{\Psi}$, making MCMC posterior evaluation more difficult. Waggoner & Zha (2003) proposed a Gibbs sampler to simulate the columns of $\mathbf{\Psi}$ from an orthonormal basis. In this paper, we focus on the selection of normalized VAR models for two reasons. First, such a structure induces analytically convenient conditional posteriors forms. Second, and more importantly, it assures that selected models are globally identified. An arbitrary $\mathbf{\Psi}$ with $p(p-1)/2$ or more restrictions does not necessarily have a one-to-one mapping with the $\mathbf{\Sigma}$ matrix via (9). Due to the nonlinear nature of the mapping (9), one can only derive local rank conditions for identification of $\mathbf{\Psi}$. 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 $\mathbf{\Sigma}$ to $\mathbf{\Psi}$ may still be non-unique. An example of multiple mappings from $\mathbf{\Sigma}$ to $\mathbf{\Psi}$ given by Bekker & Pollock (1986) is shown to satisfy the local identifying condition by Amisano & Giannini (1997). Once we restrict the $\mathbf{\Psi}$ matrix to be upper-triangular, the mapping (9) is unique. Noting that $\mathbf{\Psi}$ satisfies (3), we use $\boldsymbol{A}_0$ and $\mathbf{\Psi}$ interchangeably throughout the rest of the paper.

An alternative re-parametrization for $\mathbf{\Sigma}$ that has been succesfully exploited by Pourahmadi (1999), Smith & Kohn (2002) among others is the Cholesky factorization

$$\mathbf{\Sigma}^{-1} = \mathbf{\Gamma}'\mathbf{\Lambda}^{-1}\mathbf{\Gamma} \tag{10}$$

where $\mathbf{\Gamma}$ is a lower triangular matrix with 1's on the diagonal and $\mathbf{\Lambda}$ is a diagonal matrix of the eigenvalues of $\mathbf{\Sigma}$. In particular, Smith & Kohn (2002), who developed an alternative approach to finding parsimony in $\mathbf{\Sigma}^{-1}$, used the likelihood function to the power of the reciprocal of the sample size as the prior on the elements of the Cholesky decomposition of the covariance matrix. Their impressive approach allows for simpler posterior computation than does our approach. However, as will be seen, our element-wise normal-inverse gamma prior is more flexible and allows researchers

to incorporate elementwise prior information on the decomposed covariance matrix (e.g., based on identification schemes of a structural VAR). Yet other approaches to finding parsimony can be obtained by rewriting the covariance decomposition as a recursive regression model, [2] see Pourahmadi (1999). The details and advantages of such approaches are given in an appendix.

As will be seen, our MCMC stochastic search algorithm for restrictions on $\mathbf{\Sigma}^{-1}$ has some appealing properties. By construction, every simulated $\mathbf{\Sigma}^{-1}$ is positive definite. The simulation is iterative in the column components of $\boldsymbol{A}_0$. Under our priors, all conditional posteriors are standard distributions so that our MCMC can be carried out entirely with Gibbs sampling algorithms that appear to be rapidly convergent.

2.2 Priors

In the traditional multivariate regression literature, Bayesian analysis of VAR models typically focuses on $\mathbf{\Phi}$ or $\mathbf{\Sigma}$ as a whole, without component-wise specifications. For example, for priors on $\mathbf{\Sigma}$, 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 $\mathbf{A_0}$. 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 the components of $\mathbf{A_0}$ that utilizes the structure of the matrix and allows for easier Bayesian computation.

Let $n = (h + Lp)p$, the total number of unknown regression coefficients. Denote $\boldsymbol{\phi} = vec(\mathbf{\Phi}) = (\phi_1, \phi_2, \cdots, \phi_n)'$. For $j = 2, \cdots, p$, let $\boldsymbol{\eta}_j = (\psi_{1j}, \cdots, \psi_{j-1,j})'$. Write $\boldsymbol{\eta} = (\boldsymbol{\eta}_2', \cdots, \boldsymbol{\eta}_p')'$ and $\boldsymbol{\psi} = (\psi_{11}, \cdots, \psi_{pp})'$. We now propose hierarchical priors for $(\boldsymbol{\phi}, \boldsymbol{\eta}, \boldsymbol{\psi})$. These priors are controlled by hyperparameters for which we suggest default settings when prior information is unavailable. Throughout we use $(\cdot \mid \cdot)$ to denote conditional distribution and $[\cdot \mid \cdot]$ to denote conditional density.

(i) Priors on $\boldsymbol{\phi} = vec(\mathbf{\Phi}) = (\phi_1, \phi_2, \cdots, \phi_n)'$. Of the $n = (h + Lp)p$ elements in $\boldsymbol{\phi}$, assume that m are subject to restriction and the $n - m$ others are always included in the model. The prior for the elements that are included in every model is

$$\boldsymbol{\phi}_{non} \quad \sim \quad N_{n-m}(\boldsymbol{\phi}_{non}, \boldsymbol{M}_{non}), \tag{11}$$

where $\boldsymbol{\phi}_{non}, \boldsymbol{M}_{non}$ are hyperparameters. Natural choices for these hyperparameters are $\boldsymbol{\phi}_{non} = 0$ and $\boldsymbol{M}_{non} = cI$ with c large to reduce prior influence on $\boldsymbol{\phi}_{non}$.

[2] We thank the associate edior and a referee for pointing this out.

Denoting the m elements considered for restriction by $\boldsymbol{\phi}_m = (\phi_{s1}, \cdots, \phi_{sm})$, let $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \cdots, \gamma_m)$ be a vector of 0-1 variables and let $\boldsymbol{D} = diag(h_1, \cdots, h_m)$ where

$$h_i \;\; = \;\; \begin{cases} \tau_{0i}, & \text{if } \gamma_i = 0, \\ \tau_{1i}, & \text{if } \gamma_i = 1 \end{cases}$$

with preselected constants $\tau_{0i} < \tau_{1i}$. Letting $\boldsymbol{R}$ be a preselected correlation matrix, the prior we consider for $\boldsymbol{\phi}_m$, conditional on $\boldsymbol{\gamma}$, is

$$(\boldsymbol{\phi}_m \mid \boldsymbol{\gamma}) \sim N_m(\mathbf{0}, \mathbf{DRD}), \tag{12}$$

Note that under this prior, each element of $\boldsymbol{\phi}_m$ has the distribution

$$(\phi_{si} \mid \gamma_i) \;\; \sim \;\; (1-\gamma_i)N(0, \tau_{0i}^2) + \gamma_i N(0, \tau_{1i}^2). \tag{13}$$

The prior (12) is controlled by the hyperparameters $(\tau_{0i}, \tau_{1i}), i = 1, \ldots, m$ and $\boldsymbol{R}$. For selecting (τ_{0i}, τ_{1i}), which controls the mixture variances in (13), the basic idea is to set τ_{0i} small and τ_{1i} large so that ϕ_{si} will be restricted to be small when $\gamma_i = 0$ and unrestricted when $\gamma_i = 1$. In this way, $\boldsymbol{\gamma}$ identifies restrictions on $\boldsymbol{\phi}$. Although the choice $\tau_{0i} = 0$ forces the hard restriction $\phi_{si} = 0$, it may be more realistic to use a small nonzero value of τ_{0i}, which also accommodates the possibility of a small nonzero ϕ_{si} that is unimportant and can be ignored. A rough strategy is to set τ_{0i} such that the effect of ϕ_{si} on Y would be inconsequential when $|\phi_{si}| \leq 3\tau_{0i}$, and to set τ_{1i} so that $N(0, \tau_{1i}^2)$ would allocate substantial probability to all reasonable values of ϕ_{si}. Alternatively, a default semiautomatic approach would be to set $\tau_{ki} = c_k \hat{\sigma}_{\phi_{si}}$, where $\hat{\sigma}_{\phi_{si}}$ is the standard error associated with the unconstrained least squares estimate of ϕ_{si}, with values of $c_0 < c_1$ such as $c_0 = 1/10$ and $c_1 = 10$. These recommendations follow the considerations of George & McCulloch (1993) and George & McCulloch (1997) who provide further guidance for such choices. Finally, a natural default choice for $\boldsymbol{R}$ is $\boldsymbol{I}$, under which the components of $\boldsymbol{\phi}_m$ are apriori independent.

Combining the priors for $\boldsymbol{\phi}_{non}$ and $\boldsymbol{\phi}_m$, which we treat as independent of each other, we have

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

For the unrestricted elements of $\boldsymbol{\phi}$, the corresponding elements of $\boldsymbol{\phi}_0^{(\gamma)}$ and $\boldsymbol{M}^{(\gamma)}$ are given by (11). For the other elements of $\boldsymbol{\phi}$, the corresponding elements of $\boldsymbol{\phi}_0^{(\gamma)}$ and $\boldsymbol{M}^{(\gamma)}$ are given by (12).

(ii) Priors on $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \cdots, \gamma_m)$. We assume the elements of $\boldsymbol{\gamma}$ are independent Bernoulli $p_i \in (0, 1)$ random variables so that

$$P(\gamma_i = 1) = p_i, \;\; P(\gamma_i = 0) = 1 - p_i, \;\; i = 1, \ldots, m. \tag{14}$$

For each i, p_i reflects the prior belief that ϕ_{si} should be restricted. In the absence of such prior information a natural default choice is $p_i \equiv .5$. Alternatively one may prefer to set these p_i larger to guard against chance restriction selection due to multiplicities in the spirit of Jeffreys (1961, pp 253–255).

(iii) Priors on $\boldsymbol{\eta} = (\psi_{12}, \psi_{13}, \psi_{23}, \cdots, \psi_{p-1,p})'$. For $j = 2, \cdots, p$, let $\boldsymbol{\omega}_j = (\omega_{1j}, \cdots, \omega_{j-1,j})'$ be a vector of 0-1 variables, and let $\boldsymbol{D}_j = diag(h_{1j}, \cdots, h_{j-1,j})$ where

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

with preselected constants $\kappa_{0ij} < \kappa_{1ij}$. Letting $\boldsymbol{R}_j$ be a preselected $(j-1) \times (j-1)$ correlation matrix, the prior we consider for $\boldsymbol{\eta}_j$, conditional on $\boldsymbol{\omega}_j$, is

$$(\boldsymbol{\eta}_j \mid \boldsymbol{\omega}_j) \quad \overset{ind}{\sim} \quad N_{j-1}(\mathbf{0},\ \boldsymbol{D}_j \boldsymbol{R}_j \boldsymbol{D}_j), \quad \text{for } j = 2, \cdots, p. \tag{15}$$

Under this prior, each element of $\boldsymbol{\eta}_j$ has distribution

$$(\psi_{ij} \mid \omega_{ij}) \quad \sim \quad (1 - \omega_{ij}) N(0, \kappa_{0ij}^2) + \omega_{ij} N(0, \kappa_{1ij}^2), \text{ for } i = 1, \cdots, j-1. \tag{16}$$

The prior (15) is controlled by the hyperparameters $(\kappa_{0ij}, \kappa_{1ij}), i = 1, \ldots, m$, and $\boldsymbol{R}_j, j = 2, \ldots, p$. For selecting $(\kappa_{0ij}, \kappa_{1ij})$, which controls the mixture variances in (16), the basic idea is to set κ_{0ij} small and κ_{1ij} large so that ψ_{ij} will be restricted to be small when $\omega_{ij} = 0$ and unrestricted when $\omega_{ij} = 1$. In this way, $\boldsymbol{\omega}$ identifies restrictions on $\boldsymbol{\eta}$. Settings for $(\kappa_{0ij}, \kappa_{1ij})$ can be chosen using similar considerations for setting (τ_{0i}, τ_{1i}) as discussed in (ii) above. Finally, a natural default choice for $\boldsymbol{R}_j$ is $\boldsymbol{I}_{(j-1)\times(j-1)}$, under which the components of $\boldsymbol{\eta}_j$ are apriori independent.

(iv) Priors on $\boldsymbol{\omega} = (\boldsymbol{\omega}_2', \cdots, \boldsymbol{\omega}_p')'$. We assume the elements of $\boldsymbol{\omega}$ are independent Bernoulli $q_{ij} \in (0, 1)$ random variables so that

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

For each (i, j), q_{ij} reflects the prior belief that ψ_{ij} should be restricted. In the absence of such prior information a natural default choice is $q_{ij} \equiv .5$. Alternatively one may prefer to set these q_{ij} larger to guard against chance restriction selection due to multiplicities in the spirit of Jeffreys (1961, pp 253–255).

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

$$[\psi_{ii}] = \frac{2 b_i^{a_i}}{\Gamma(a_i)} \psi_{ii}^{2(a_i - 1)} \exp(-b_i \psi_{ii}^2), \quad \text{for } \psi_{ii} > 0. \tag{18}$$

In the absence of prior information about ψ_{ii}, a natural default choice is to render this prior noninfluential with the hyperparameters (a_i, b_i) set to small values such as $(.01, .01)$.

3 Stochastic Search for VAR Restrictions

Combining the likelihood function with a particular prior density defines a posterior probability distribution over the unknown parameters and the restriction indices $\boldsymbol{\gamma}$ and $\boldsymbol{\omega}$. The marginal posterior distributions of $\boldsymbol{\gamma}$ and $\boldsymbol{\omega}$ thus summarize all post-data restriction uncertainty. Although exhaustive calculation of these posteriors is not feasible even in moderately sized problems, it is possible to simulate a Markov chain sample from these posteriors using MCMC methods. Because higher posterior probability values will appear more often in such samples, the simulation of these Markov chains is effectively a stochastic search for the higher posterior probability restrictions, i.e., those that are more highly supported by the data. To carry out this MCMC search in our setting, we propose an extension of the Gibbs sampling approach of George & McCulloch (1993). This simply entails sequential substitution sampling from the full conditional posterior distributions of all the unknown parameters and restriction indices. We now proceed to describe these conditional distributions and the resulting MCMC algorithms. We first discuss stochastic search algorithms for $\boldsymbol{\Phi}$ restrictions and for $\boldsymbol{\Psi}$ restrictions separately, and then consider them together.

3.1 Restriction search for $\boldsymbol{\Psi}$

In some (but not all) VAR modelling applications, economists begin with reasonably good prior knowledge of the nature of the VAR regression coefficient matrix $\boldsymbol{\Phi}$, and so are interested only in selecting restrictions for $\boldsymbol{\Psi}$. For instance, macroeconomic time series such as stock price indices are known to exhibit approximately random walk behavior (which is a basis of the well-known Minnesota prior on $\boldsymbol{\Phi}$). Incorporating such prior information may improve the Bayesian inference on $\boldsymbol{\Phi}$ as well as $\boldsymbol{\Psi}$.

When restrictions are not considered for $\boldsymbol{\Phi}$, we let the prior on $\boldsymbol{\phi} = vec(\boldsymbol{\Phi})$ be

$$\boldsymbol{\phi} \;\sim\; N(\boldsymbol{\phi}_0, \boldsymbol{\Xi}_0) \tag{19}$$

instead of the mixture prior described in Section 2.2. The restriction index $\boldsymbol{\gamma}$ is thus ignored in this setting. The conditional posterior distribution of $\boldsymbol{\phi}$ given $(\boldsymbol{\Psi}; \boldsymbol{Y})$ under this prior is

$$(\boldsymbol{\phi} \mid \boldsymbol{\Psi}; \boldsymbol{Y}) \sim N(\widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\Xi}}), \tag{20}$$

where

$$\widehat{\boldsymbol{\Xi}} = \Big\{\boldsymbol{\Psi}\boldsymbol{\Psi}' \otimes (\boldsymbol{X}'\boldsymbol{X}) + \boldsymbol{\Xi}_0^{-1}\Big\}^{-1}, \tag{21}$$

$$\widehat{\boldsymbol{\phi}} = \widehat{\boldsymbol{\Xi}}\Big\{\boldsymbol{\Psi}\boldsymbol{\Psi}' \otimes (\boldsymbol{X}'\boldsymbol{X})\widehat{\boldsymbol{\phi}}_M + \boldsymbol{\Xi}_0^{-1}\boldsymbol{\phi}_0\Big\}. \tag{22}$$

To derive the other conditional distributions under the remaining Section 2.2 priors, note that the likelihood function (6) of $(\boldsymbol{\phi}, \boldsymbol{\Psi})$ can also be rewritten as

$$f(\boldsymbol{Y} \mid \boldsymbol{\phi}, \boldsymbol{\Psi}) \propto |\boldsymbol{\Psi}|^T etr\Big\{-\frac{1}{2}\boldsymbol{\Psi}'\boldsymbol{S}(\boldsymbol{\Phi})\boldsymbol{\Psi}\Big\}, \tag{23}$$

where $\boldsymbol{S}(\boldsymbol{\Phi})$ is given by (7) and write $\boldsymbol{S}(\boldsymbol{\Phi}) = (s_{ij})$. For $j = 2, \cdots, p$, define $\boldsymbol{s}_j = (s_{1j}, \cdots, s_{j-1,j})'$. Let $\boldsymbol{S}_j$ be the upper-left $j \times j$ submatrix of $\boldsymbol{S}(\boldsymbol{\Phi})$. Define $v_1 = s_{11}$ and $v_i = |\boldsymbol{S}_i|/|\boldsymbol{S}_{i-1}|$ for $i = 2, \cdots, p$. It is well known that $v_i = s_{ii} - \boldsymbol{s}_i'\boldsymbol{S}_{i-1}^{-1}\boldsymbol{s}_i > 0$ for $i = 2, \cdots, p$. Thus (23) equals

$$f(\boldsymbol{Y} \mid \boldsymbol{\phi}, \boldsymbol{\Psi}) \propto \prod_{i=1}^{p} \psi_{ii}^T \exp\Big[-\frac{1}{2}\Big\{\sum_{i=1}^{p}\psi_{ii}^2 v_i + \sum_{j=2}^{p}(\boldsymbol{\eta}_j + \psi_{jj}\boldsymbol{S}_{j-1}^{-1}\boldsymbol{s}_j)'\boldsymbol{S}_{j-1}(\boldsymbol{\eta}_j + \psi_{jj}\boldsymbol{S}_{j-1}^{-1}\boldsymbol{s}_j)\Big\}\Big]. \tag{24}$$

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

Fact 1 *(a) For given $(\boldsymbol{\phi}, \boldsymbol{\omega}, \boldsymbol{\psi}; \boldsymbol{Y})$, the posterior distributions of $\boldsymbol{\eta}_2, \cdots, \boldsymbol{\eta}_p$ are independent, and*

$$(\boldsymbol{\eta}_j \mid \boldsymbol{\phi}, \boldsymbol{\omega}, \boldsymbol{\psi}; \boldsymbol{Y}) \sim N_{j-1}(\boldsymbol{\mu}_j, \boldsymbol{\Delta}_j), \tag{25}$$

where

$$\boldsymbol{\mu}_j = -\psi_{jj}\{\boldsymbol{S}_{j-1} + (\boldsymbol{D}_j\boldsymbol{R}_j\boldsymbol{D}_j)^{-1}\}^{-1}\boldsymbol{s}_j, \tag{26}$$

$$\boldsymbol{\Delta}_j = \{\boldsymbol{S}_{j-1} + (\boldsymbol{D}_j\boldsymbol{R}_j\boldsymbol{D}_j)^{-1}\}^{-1}. \tag{27}$$

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

$$(\psi_{ii}^2 \mid \boldsymbol{\phi}, \boldsymbol{\omega}; \boldsymbol{Y}) \sim gamma(a_i + \frac{1}{2}T, B_i), \tag{28}$$

where

$$B_i = \begin{cases} b_1 + \frac{1}{2}s_{11}, & if \;\; i = 1, \\ b_i + \frac{1}{2}\Big\{s_{ii} - \boldsymbol{s}_i'[\boldsymbol{S}_{i-1} + (\boldsymbol{D}_i\boldsymbol{R}_i\boldsymbol{D}_i)^{-1}]^{-1}\boldsymbol{s}_i\Big\}, & if \;\; i = 2, \cdots, p. \end{cases} \tag{29}$$

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

$$u_{ij1} = [\boldsymbol{\eta}_j \mid \boldsymbol{\omega}_{(-ij)}, \omega_{ij} = 1]q_{ij}, \tag{30}$$

$$u_{ij2} = [\boldsymbol{\eta}_j \mid \boldsymbol{\omega}_{(-ij)}, \omega_{ij} = 0](1 - q_{ij}). \tag{31}$$

If $\boldsymbol{R}_j = \boldsymbol{I}_{j-1}$ *in prior (15), for fixed* $j = 2, \cdots, p$ *and given* $(\boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\omega}_k, k \neq j; \boldsymbol{Y})$, $\omega_{1j}, \cdots, \omega_{j-1,j}$ *are independent.* $(\omega_{ij} \mid \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\omega}_k, k \neq j; \boldsymbol{Y}) \overset{ind}{\sim}$ *Bernoulli* $(\tilde{u}_{ij1}/(\tilde{u}_{ij1} + \tilde{u}_{ij2}))$, *where*

$$\tilde{u}_{ij1} = \frac{1}{\kappa_{1ij}} \exp\Big(-\frac{\psi_{ij}^2}{2\kappa_{1ij}^2}\Big) q_{ij}, \tag{32}$$

$$\tilde{u}_{ij2} = \frac{1}{\kappa_{0ij}} \exp\Big(-\frac{\psi_{ij}^2}{2\kappa_{0ij}^2}\Big) (1 - q_{ij}). \tag{33}$$

The proof of Fact 1 is in the appendix. Combining the above conditional distributions, the MCMC stochastic search algorithm for sampling from the full posterior distribution $[\boldsymbol{\psi}, \boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{\omega}; Y]$ is obtained as follows. Given initial values $\boldsymbol{\psi}_0, \boldsymbol{\eta}_0, \boldsymbol{\phi}_0, \boldsymbol{\omega}_0$, the kth cycle is obtained from $\boldsymbol{\psi}_{(k-1)}$, $\boldsymbol{\eta}_{(k-1)}$, $\boldsymbol{\phi}_{(k-1)}$, $\boldsymbol{\omega}_{(k-1)}$ by sequentially simulating the steps:

1. draw $(\boldsymbol{\psi}_{(k)} \mid \boldsymbol{\phi}_{(k-1)}, \boldsymbol{\omega}_{(k-1)}; \boldsymbol{Y})$ from the gamma distribution (28);
2. draw $(\boldsymbol{\eta}_{(k)} \mid \boldsymbol{\psi}_{(k)}, \boldsymbol{\phi}_{(k-1)}, \boldsymbol{\omega}_{(k-1)}; \boldsymbol{Y})$ from the normal distribution (25);
3. draw $(\boldsymbol{\omega}_{(k)} \mid \boldsymbol{\eta}_{(k)})$ from the Bernoulli distribution (30);
4. draw $(\boldsymbol{\phi}_{(k)} \mid \boldsymbol{\Sigma}_{(k)}, \boldsymbol{\omega}_{(k)}; \boldsymbol{Y})$ from the normal distribution (20), where $\boldsymbol{\Sigma}_{(k)}$ is computed from $\boldsymbol{\psi}_{(k)}$ and $\boldsymbol{\eta}_{(k)}$.

3.2 Restriction search for $\boldsymbol{\Phi}$

We now consider stochastic search for restrictions only on the regression coefficients $\boldsymbol{\Phi}$. A special example of such restrictions in applications is the specification of VAR lags. Researchers 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 familiar bias-variance tradeoff: if the specified lag is too short the results may be very misleading, but if the specified lag is too long the large number of parameters may result in erratic estimates. Furthermore, researchers may find it desirable in some cases to include only some of the lag coefficients $\boldsymbol{A}_j$ $(1 \leq j \leq L)$, or to include only some of the elements in $\boldsymbol{A}_j$ but not the entire matrix. Our restriction research allows for more than just lag specification, and can meet such needs.

In a study related to the selection of regression coefficients, Brown et al. (1998) extended the approach of George & McCulloch (1993) to a multivariate regression setting. However, their multivariate model has a practical focus different from VARs. They also treated the covariance matrix $\boldsymbol{\Sigma}$ as a nuisance parameter and integrated it out with respect to a conjugate normal-inverse Wishart prior.

We now proceed to obtain the conditional distributions for the Gibbs sampler in this setting. When restrictions are not considered for $\boldsymbol{\Psi}$, we let the prior on $\boldsymbol{\eta}$ be

$$\boldsymbol{\eta}_j \overset{ind}{\sim} N_{j-1}(\mathbf{0},\ \boldsymbol{\Omega}_{j-1}), \quad \text{for } j = 2, \cdots, p \tag{34}$$

instead of the mixture prior described in Section 2.2. Note that the restriction index $\boldsymbol{\omega}$ is thus ignored in this context. The conditional posterior of $(\boldsymbol{\eta}_j \mid \boldsymbol{\psi}_{jj}, \boldsymbol{\phi}; \boldsymbol{Y})$ is a normal distribution

$$N(-\psi_{jj}(\boldsymbol{S}_{j-1} + \boldsymbol{\Omega}_{j-1}^{-1})^{-1}\boldsymbol{s}_{j-1,j},\ (\boldsymbol{S}_{j-1} + \boldsymbol{\Omega}_{j-1}^{-1})^{-1}).$$

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

$$f(\boldsymbol{Y} \mid \boldsymbol{\phi}, \boldsymbol{\Psi}) \ \propto\ |\boldsymbol{\Psi}|^T \exp\Big[-\frac{1}{2}(\boldsymbol{\phi} - \widehat{\boldsymbol{\phi}}_M)'\{\boldsymbol{\Psi}\boldsymbol{\Psi}' \otimes (\boldsymbol{X}'\boldsymbol{X})\}(\boldsymbol{\phi} - \widehat{\boldsymbol{\phi}}_M) - \frac{1}{2}tr\{\boldsymbol{\Psi}\boldsymbol{\Psi}'\boldsymbol{S}(\widehat{\boldsymbol{\Phi}}_M)\}\Big]. \tag{35}$$

The full conditional posteriors of $\boldsymbol{\phi}$, $\boldsymbol{\eta}$, $\boldsymbol{\psi}$, $\boldsymbol{\gamma}$ using the remaining Section 2.2 priors are obtained as follows.

Fact 2 *(a) The conditional posterior distribution*

$$(\boldsymbol{\phi} \mid \boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\psi}; \boldsymbol{Y}) \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Delta}), \tag{36}$$

where

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

(b) Denote $\boldsymbol{\gamma}_{(-i)} = (\gamma_1, \cdots, \gamma_{i-1}, \gamma_{i+1}, \cdots, \gamma_m)$. *Then under prior (12),*

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

where

$$u_{i1} = [\boldsymbol{\phi} \mid \boldsymbol{\gamma}_{(-i)}, \gamma_i = 1]p_i, \tag{37}$$

$$u_{i2} = [\boldsymbol{\phi} \mid \boldsymbol{\gamma}_{(-i)}, \gamma_i = 0](1 - p_i). \tag{38}$$

If $\boldsymbol{R} = \boldsymbol{I}_m$ *in prior (12), then given* $(\boldsymbol{\phi}, \boldsymbol{\eta}_j, \psi_{ii}, 1 \le i \le p; \boldsymbol{Y})$, $\gamma_i \overset{ind}{\sim}$ *Bernoulli* $(\tilde{u}_{i1}/(\tilde{u}_{i1} + \tilde{u}_{i2}))$, *where*

$$\tilde{u}_{i1} = \frac{1}{\tau_{0i}} \exp\Big(-\frac{\phi_i^2}{2\tau_{0i}^2}\Big)p_i, \tag{39}$$

$$\tilde{u}_{i2} = \frac{1}{\tau_{1i}} \exp\Big(-\frac{\phi_i^2}{2\tau_{1i}^2}\Big)(1 - p_i). \tag{40}$$

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

Combining the above conditional distributions, the MCMC stochastic search algorithm for sampling from the full posterior distribution $[\boldsymbol{\psi}, \boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{\gamma}; Y]$ is obtained as follows. Given initial values $\boldsymbol{\psi}_0, \boldsymbol{\eta}_0, \boldsymbol{\phi}_0, \boldsymbol{\gamma}_0$, the kth cycle is obtained from $\boldsymbol{\psi}_{(k-1)}, \boldsymbol{\eta}_{(k-1)}, \boldsymbol{\phi}_{(k-1)}, \boldsymbol{\gamma}_{(k-1)}$ by sequentially simulating the steps:

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

$$(\psi_{ii}^2 \mid \boldsymbol{\phi}; \boldsymbol{Y}) \sim gamma(a_i + \frac{1}{2}T, B_i), \tag{41}$$

where

$$B_i \;=\; \begin{cases} b_1 + \frac{1}{2}s_{11}, & \text{if } \; i = 1, \\ b_i + \frac{1}{2}\Big\{ s_{ii} - \boldsymbol{s}'_{i-1,i}(\boldsymbol{S}_{i-1} + \boldsymbol{\Omega}_{i-1}^{-1})^{-1}\boldsymbol{s}_{i-1,i} \Big\}, & \text{if } \; i = 2, \cdots, p; \end{cases} \tag{42}$$

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

$$(\boldsymbol{\eta}_i \mid \psi_{ii}; \boldsymbol{S}_i) \;\; \propto \;\; N(-\psi_{ii}(\boldsymbol{S}_{i-1} + \boldsymbol{\Omega}_{i-1}^{-1})^{-1}\boldsymbol{s}_{i-1,i},\; (\boldsymbol{S}_{i-1} + \boldsymbol{\Omega}_{i-1}^{-1})^{-1}); \tag{43}$$

3. draw $(\boldsymbol{\phi}_{(k)} \mid \boldsymbol{\gamma}_{(k-1)}, \boldsymbol{\Sigma}_{(k)}; \boldsymbol{Y})$ from normal distribution (36) 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 (37).

3.3 Restriction search for $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$

We now consider joint restriction search for both $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$. For the MCMC sampling algorithm, we need the full conditional distributions for $\boldsymbol{\theta} \equiv (\boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{\psi})$ given $\boldsymbol{Y}$ which can be obtained as before. Note that the joint posterior of $(\boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\omega}, \boldsymbol{\psi})$ is

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

These conditional distributions and the MCMC algorithm are given in the appendix.

The stochastic search results for VAR restrictions can be used in a variety of ways. To begin with, the most frequently visited restrictions tend to have higher posterior probability and hence are promising candidates for further analysis. For these candidates, the method proposed by Chib (1995)

can be used to calculate marginal likelihoods and posterior odds ratios. If each set of restrictions is viewed as a VAR identification scheme, then odds ratios provide a natural test of identifying schemes. As opposed to the commonly used F tests for $\mathbf{\Phi}$ identification, odds ratios for the most frequently visited restrictions offer a promising Bayesian alternative.

The stochastic search results can also be used to obtain improved VAR forecasts as compared for example to the benchmark VAR forecasts that impose no restrictions $\mathbf{\Phi}$ and $\mathbf{\Psi}$. In Section 4, we compare these estimates in simulated examples. For evaluation, we use average MSE to compare point forecasts and Kullback-Liebler divergence to compare predictive density estimates. The technical details of average MSE and Kullback-Liebler divergence are given in the appendix.

4 Simulated Numerical Examples

We use two simulated numerical examples to illustrate the performance of the stochastic search algorithm. For each example, we simulate 100 samples of size $T = 50$, and for each sample we conduct stochastic search by simulating a Markov Chain of $50,000$ cycles (after $10,000$ burn-in cycles). Simulation results change little when the Markov Chain length was reduced to $20,000$ cycles, suggesting that the Markov chains converge rather quickly. The stochastic search is evaluated in two ways, the effectiveness in finding the true data generating model and the forecast performance.

The stochastic search for restrictions is conducted in two steps. First, for each sample we run the appropriate MCMC search algorithm laid out earlier in the paper. We report averages over all the runs over all the samples of the restriction indices $\boldsymbol{\gamma}$, $\boldsymbol{\omega}$ and the parameter matrices $\mathbf{\Psi}$, $\mathbf{\Phi}$. Three types of estimators are compared: (1) MLE; (2) UB, which is based on the benchmark unrestricted Bayesian VAR; (3) RB, which is obtained by averaging the restricted Bayesian VAR models that are visited over the stochastic search. Our second criterion is predictive performance. We report the relative performance of average MSE of the three estimates listed above. We also report the Kullback-Liebler difference between the UB predictive density and the RB predictive density described in the appendix.

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

$$\mathbf{\Phi} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \text{ and } \mathbf{\Psi} = \begin{pmatrix} 1 & .5 & .5 & .5 & .5 & .5 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{44}$$

In all Bayesian models, the intercept term is assigned a normal prior with a zero mean and a variance of 25. For the UB estimators, the Minnesota prior (which is centered at a random walk on each variable) is applied to $\mathbf{\Phi}$ with the covariance matrix of the prior being diagonal and with variance of 25. For the restriction priors from Section 2.2 for RB, the hyperparameters were set at $p_i = .5$, $q_{ij} = 0.5$, $a = b = .01$, $\boldsymbol{R} = \boldsymbol{I}$, $\boldsymbol{R}_j = \boldsymbol{I}_j$, $(\tau_{0i}, \tau_{1i}) = (0.1, 5)$ and $(\kappa_{0ij}, \kappa_{1ij}) = (0.1, 5)$.

The average of the MLEs for $\mathbf{\Psi}$ and $\mathbf{\Phi}$ over the 100 samples are

$$\widehat{\mathbf{\Phi}}_{MLE} = \begin{pmatrix} 1.28 & .76 & .74 & 1.04 & .92 & .80 \\ .80 & .12 & .13 & .12 & .13 & .13 \\ .04 & .75 & .02 & .05 & .01 & .05 \\ .04 & .04 & .74 & .04 & .04 & .02 \\ .02 & .03 & .04 & .74 & .04 & .06 \\ .06 & .02 & .03 & .02 & .74 & .02 \\ .04 & .05 & .04 & .04 & .03 & .72 \end{pmatrix}, \widehat{\mathbf{\Psi}}_{MLE} = \begin{pmatrix} 1.85 & .38 & .38 & .39 & .39 & .38 \\ .00 & 1.20 & -.16 & -.15 & -.13 & -.18 \\ .00 & .00 & 1.16 & -.18 & -.16 & -.18 \\ .00 & .00 & .00 & 1.15 & -.18 & -.22 \\ .00 & .00 & .00 & .00 & 1.10 & -.24 \\ .00 & .00 & .00 & .00 & .00 & 1.05 \end{pmatrix}.$$

Although the Minnesota prior on $\mathbf{\Phi}$ shrinks the posterior of the parameter matrix towards random walk models, such shrinkage is limited due to the large variance of the prior. Thus, as we would expect, the values for the MLEs and the UB estimates are similar. As we will see, the forecasting performance of the UB estimator is also is similar to that of the MLE.

We next turn to the joint restriction search for $\mathbf{\Psi}$ and $\mathbf{\Phi}$. The average of the restriction indices $\boldsymbol{\gamma}$ and $\boldsymbol{\omega}$ over all stochastic searches over all 100 samples are

$$\widehat{\gamma} = \begin{pmatrix} 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 0.04 & .05 & .05 & .05 & .05 \\ .04 & 1.00 & .04 & .04 & .04 & .04 \\ .04 & .05 & 1.00 & .05 & .04 & .04 \\ .04 & .04 & .04 & 1.00 & .04 & .05 \\ .04 & .04 & .04 & .04 & 1.00 & .04 \\ .04 & .04 & .04 & .04 & .04 & 1.00 \end{pmatrix}, \widehat{\boldsymbol{\omega}} = \begin{pmatrix} 1.00 & .58 & .55 & .53 & .52 & .50 \\ .00 & 1.00 & .08 & .08 & .07 & .08 \\ .00 & .00 & 1.00 & .07 & .08 & .08 \\ .00 & .00 & .00 & 1.00 & .08 & .10 \\ .00 & .00 & .00 & .00 & 1.00 & .09 \\ .00 & .00 & .00 & .00 & .00 & 1.00 \end{pmatrix}.$$

These $\widehat{\gamma}$ and $\widehat{\boldsymbol{\omega}}$ averages illustrate how often the stochastic search is visiting the correct restrictions. In Table 1, we report how often over all cycles and all samples the visited restriction indices completely

Table 1: The frequency of matching the exactly correct restrictions

$\mathbf{\Phi}$ Restrictions	$\mathbf{\Psi}$ Restrictions	$\mathbf{\Phi}$ and $\mathbf{\Psi}$ Restrictions
30.9%	1.7%	0.6%

matched every element of the true model. For the joint restriction search for $\mathbf{\Psi}$ and $\mathbf{\Phi}$, a perfect match was obtained about 0.6% of the time. Although this may appear small, it is quite reasonable if we note that (with intercepts always included) the number of parameters subject to selection is $6 \times 6 + 6(6-1)/2 = 51$ yielding 2^{51} restriction candidates. Thus, the chance of randomly picking the true model is much less than 0.6%. Table 1 indicates that in this example, matching the correct $\mathbf{\Psi}$ index is less likely than matching the correct $\mathbf{\Phi}$ index. Since the MLE of $\mathbf{\Psi}$ is quite biased, the posteriors of some of the off-diagonal elements of $\mathbf{\Psi}$ are not centered at the true values. When we impose no restrictions on the regression coefficients $\mathbf{\Phi}$, the chance that the correct $\mathbf{\Psi}$ is matched out of the 2^{15} possible models is 1.7%, which is relatively large. However, with all elements of $\mathbf{\Psi}$ included, the stochastic search matches the correct index for $\mathbf{\Phi}$ 30.9% of the time, which is remarkable given that there are 2^{36} restriction candidates.

For each sample, the average RB posterior mean estimates of $\mathbf{\Psi}$ and $\mathbf{\Phi}$ are

$$\widehat{\mathbf{\Phi}} = \begin{pmatrix} 1.11 & .95 & .93 & 1.12 & 1.05 & .98 \\ .87 & .06 & .07 & .06 & .07 & .06 \\ .03 & .80 & .03 & .04 & .02 & .04 \\ .02 & .04 & .81 & .04 & .04 & .03 \\ .02 & .03 & .03 & .80 & .03 & .04 \\ .03 & .03 & .03 & .03 & .81 & .03 \\ .02 & .04 & .03 & .03 & .03 & .79 \end{pmatrix} \text{ and } \widehat{\mathbf{\Psi}} = \begin{pmatrix} 1.09 & .42 & .40 & .41 & .39 & .38 \\ .00 & 1.08 & -0.04 & -0.03 & -0.03 & -0.04 \\ .00 & .00 & 1.08 & -0.03 & -0.03 & -0.03 \\ .00 & .00 & .00 & 1.11 & -0.03 & -0.04 \\ .00 & .00 & .00 & .00 & 1.10 & -0.04 \\ .00 & .00 & .00 & .00 & .00 & 1.10 \end{pmatrix}.$$

We next examine the one-step ahead forecasting MSE improvement over the MLE of the UB and RB posterior mean estimates of $\mathbf{\Phi}$. Table 2 shows that there is little improvement by the UB estimates over the MLE. This result is expected given the diffuse nature of the UB prior. The RB estimator performs substantially better than the UB estimator, suggesting that good restrictions were often selected. Note that a 66% improvement means that the forecast error of restricted VAR is only one-third of that of the unrestricted VAR. This is consistent with the previous observation that our selection procedure often closely reflected the correct restriction indices.

Our second measure of predictive performance, the KL distance between the UB predictive density and the RB predictive density is reported in Table 3 for the 100 samples. The consistently large

Table 2: Forecast improvement over the MLE

	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$
UB	8.5%	3.7%	5.2%	2.9%	6.3%	6.6%
ARB	65.5%	53.5%	61.5%	46.5%	67.3%	62.4%

Table 3: KL distance between the UB and ARB predictive densities

Mean	SD	Min	Max
44.36	10.66	20.15	71.64

KL distance shows that the RB predictive density is substantially better. The stochastic search restrictions have overcome over-fitting and resulted in better predictive performance.

We now consider stochastic search for restrictions for $\mathbf{\Phi}$ and $\mathbf{\Psi}$ separately. One of our goals is to shed light on whether the restriction search for $\mathbf{\Psi}$ influences the restriction search for $\mathbf{\Phi}$ and vice versa. Using the same data as before, but considering restriction only for $\mathbf{\Phi}$, the average of the stochastic search values of the restriction index $\boldsymbol{\gamma}$ are almost the same as what we previously observed. Again using the same data, but now considering restriction only for $\mathbf{\Psi}$, then the averages of the $\mathbf{\Phi}$ posterior mean estimates are close to that of the MLE. Furthermore, the average restriction index $\boldsymbol{\omega}$ under restriction search for $\mathbf{\Psi}$ only is not as good as under the joint search. This suggests that restriction of the regression coefficients $\mathbf{\Phi}$ helps in selecting the correct restrictions for $\mathbf{\Psi}$.

Example 2 In Example 1, we saw that when restrictions are appropriately imposed on $\mathbf{\Phi}$, stochastic search procedures improve on benchmark UB procedures. In this example, we see that even when selection is ineffective in selecting the correct restrictions on $\mathbf{\Phi}$, the stochastic search results can still be used to obtain substantial forecasting improvements. Consider a four-variable VAR with two lags

and parameters

$$\mathbf{\Phi} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ .45 & 0 & 0 & 0 \\ 0 & .50 & 0 & 0 \\ 0 & 0 & .55 & 0 \\ 0 & 0 & 0 & .60 \\ .40 & 0 & 0 & 0 \\ 0 & .40 & 0 & 0 \\ 0 & 0 & .40 & 0 \\ 0 & 0 & 0 & .40 \end{pmatrix} \text{ and } \mathbf{\Psi} = \begin{pmatrix} 1 & .5 & .5 & .5 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{45}$$

Proceeding as we did in Example 1, the average of the MLEs of $\mathbf{\Psi}$ and $\mathbf{\Phi}$ are

$$\widehat{\mathbf{\Phi}}_{MLE} = \begin{pmatrix} 2.32 & 2.03 & 1.72 & -.25 \\ .31 & .02 & .04 & .04 \\ .01 & .36 & .02 & -.02 \\ -.03 & .03 & .44 & .08 \\ -.03 & .01 & .05 & .54 \\ .27 & .04 & .03 & .12 \\ .02 & .25 & .01 & .02 \\ .03 & .01 & .26 & .05 \\ .04 & .01 & .03 & .40 \end{pmatrix} \text{ and } \widehat{\mathbf{\Psi}}_{MLE} = \begin{pmatrix} 1.56 & .43 & .46 & .45 \\ 0 & 1.09 & -.22 & -.17 \\ 0 & 0 & 1.07 & -.25 \\ 0 & 0 & 0 & 1.02 \end{pmatrix}.$$

The average of the restriction indices $\boldsymbol{\gamma}$ and $\boldsymbol{\omega}$ are

$$\widehat{\gamma} = \begin{pmatrix} 1.00 & 1.00 & 1.00 & 1.00 \\ .87 & .06 & .06 & .06 \\ .05 & 0.87 & .05 & .05 \\ .04 & .05 & .93 & .06 \\ .03 & .03 & .04 & .97 \\ .15 & .05 & .06 & .07 \\ .05 & .13 & .05 & .05 \\ .04 & .05 & .16 & .05 \\ .03 & .03 & .04 & .42 \end{pmatrix} \text{ and } \widehat{\boldsymbol{\omega}} = \begin{pmatrix} 1.00 & .53 & .56 & .49 \\ .00 & 1.00 & .08 & .06 \\ .00 & .00 & 1.00 & .07 \\ .00 & .00 & .00 & 1.00 \end{pmatrix}.$$

From these averages, we can see that the restrictions on $\mathbf{\Phi}$ visited by the stochastic search often incorrectly exclude the lag 2 coefficients, although the restrictions on $\mathbf{\Psi}$ seem to accurately reflect the nature of the true data-generating model. Consequently frequencies of selecting the data generating parameter $\mathbf{\Phi}$ in Table 4 are substantially lower than their Table 1 counterparts.

Table 4: The frequency of matching the exactly correct restrictions

$\mathbf{\Phi}$ Restrictions	$\mathbf{\Psi}$ Restrictions	$\mathbf{\Phi}$ and $\mathbf{\Psi}$ Restrictions
0.02%	14.1%	0%

The averages of the posterior means of $\mathbf{\Psi}$ and $\mathbf{\Phi}$ are

$$\widehat{\mathbf{\Psi}} = \begin{pmatrix} 1.03 & .48 & .51 & .48 \\ .00 & 1.04 & -0.02 & -0.02 \\ .00 & .00 & 1.08 & -0.03 \\ .00 & .00 & .00 & 1.06 \end{pmatrix} \text{ and } \widehat{\mathbf{\Phi}} = \begin{pmatrix} 2.16 & 2.20 & 1.87 & .41 \\ .41 & .02 & .02 & .03 \\ .01 & .44 & .02 & -.01 \\ -.00 & .01 & .50 & .04 \\ .00 & .01 & .04 & .59 \\ .20 & .01 & .02 & .06 \\ .01 & .16 & .01 & .00 \\ .01 & .02 & .21 & .02 \\ .00 & .02 & .04 & .38 \end{pmatrix}.$$

In spite the fact that the restrictions on $\mathbf{\Phi}$ were not selected as well as in Example 1, Table 5 shows that improvement in one-step ahead forecasting MSE by the RB estimators over the MLE and UB estimators is still substantial here. Furthermore, Table 6 shows that the KL distance between the UB and RB predictive densities is also substantial, although not as quite as large as in Table 3.

Table 5: Forecast improvement over the MLE

	$i = 1$	$i = 2$	$i = 3$	$i = 4$
UB	3.4%	5.3%	1.5%	4.1%
RB	40.2%	47.8%	19.6%	44.5%

Table 6: KL distance between the UB and RB predictive densities

Mean	SD	Min	Max
16.35	5.28	3.25	29.26

5 An Empirical Analysis of PPI to CPI Inflation Transmission

We now apply the stochastic search approach to an empirical inquiry. The issue at hand is how Producer Price Index (PPI) inflation is passed through different stages of production to Consumer Price Index (CPI) inflation. The dynamics of inflation pass-through is a complicated matter. Until now, the most common tool for examining the problem has been unrestricted VAR. As illustrated by our numerical examples, restricted VAR models via stochastic search can provide sharper inference and substantially improve forecasting over unrestricted VARs.

There are several explanations why PPI inflation of crude materials may lead CPI inflation. Blomberg and Harris (1995) argued that the commodity market for crude materials is competitive when prices quickly respond to new information on future demand and supply. In contrast, the prices of final goods and consumer prices respond more slowly because of market rigidities such as menu costs or long-term contractual arrangements. The existence of commodity markets makes it possible 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, while PPI does not. PPI includes capital equipment while 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. And even if 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 in a longer delay in changes of aggregate output prices and consumer prices, (see Mattey (1981)). In addition to the cumulative delays, there are also complications in whether and how price changes in a particular component of PPI are passed to a similar component of CPI. It is also possible that because of effects such as shocks in transportation costs, the CPI for a given component may change while the PPI component does not. This all makes the overall response of CPI to PPI inflation unclear. The theoretical ambiguities of PPI to CPI inflation transmission make it a very interesting empirical question.

Existing empirical studies show that VAR estimates of the response of CPI inflation to commodity price inflation are generally unstable over sample periods. Blomberg & Harris (1995) and Furlong & Ingenito (1996) found that commodity price inflation in the 1970s and early 1980s led to CPI

inflation, but this did not happen afterwards. Clark (1995) showed that the extent to which PPI inflation predicts CPI inflation varies through time. Weinhagen (2002) studied price transmission through various stages of production. He found that for the sample period of 1974 to 1989, inflation of the PPI for crude materials and intermediate goods were good predictors of CPI inflation, but for the period of 1990 to 2001 only inflation of the PPI for finished goods predicts CPI inflation. These studies used benchmark VAR models without considering the restriction selection issue. It is widely acknowledged that CPI inflation data shows an upward bias and that industry-level PPI and CPI relationship may be nonlinear. VAR may be inadequate to model the nonlinear relationship and it is useful for to consider more flexible modelling alternatives in future research.

In the following, we limit our scope to restrictions of a VAR of seven-variables, 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 of foodstuff and feeds (FF), PPI of crude materials for processing (CM), PPI of intermediate materials and supplies (IM), PPI of finished consumer goods (FC), CPI for urban consumers (CPI), unemployment rate (UNEMP) and Federal Funds Rates (FFR). The unemployment rate is a measure of the state of the economy, and the Federal Funds rates capture the stance of monetary policy. Motivated by the findings in the literature on the change of pattern of inflation transmission, we will compare the results of the two sample periods, the first sample period is from January 1969 to December 1980, the second sample period is from January 1981 to August 2005.

To allow for long delays in inflation transmission we use a VAR with a lag length of 12. We assume that the intercept term is always included. The total number of parameters in $\mathbf{\Psi}$ is 28 and in $\mathbf{\Phi}$ is 595. Given the limited number of sample observations, there is potential for substantial overfitting, so restriction selection may be particularly valuable here. Given the large number of parameters, we ran the MCMC for $20,000$ cycles excluding $10,000$ burn-in iterations. The MLEs of the components of $\boldsymbol{\phi}$ (the vectorized regression coefficient $\mathbf{\Phi}$) are plotted in Figure 1 and the posterior means of the restriction indexes on the regression coefficients $\boldsymbol{\gamma}$ in Figure 2. The posterior means and standard deviations of the components of $\boldsymbol{\phi}$ are plotted in Figures 3 and 4 for the two samples.

The left column of Figure 1 plots the MLEs of the 85 parameters of $\boldsymbol{\phi}$ in each equation for the sample 1969 to 1980. The right column plots the same set of parameters for the sample 1981 to 2005. These two columns show markedly different patterns, indicating very different MLE estimates for the two sample periods.

Figure 2 plots the model selection indexes of the regression coefficients for the two sample periods.

The first parameter, the intercept is selected with probability one in all equations by design. For the 1969-1980 sample, in the PPI FF and CM equations, the first lag of CPI and the fourth lag of FC are selected with high frequencies. Figure 3 shows that these estimates are negative. The difference between the restricted and unrestricted VARs $\mathbf{\Phi}$ lies in the size of the coefficients. The former is larger than the latter but often switch signs after each lag, showing a sign of overfitting. For the sample 1969-1980, the model selection index of the CPI equation show that most of the lags of the variables in the model are not useful to predicting CPI, except for the ninth lag of CPI. This suggests that most variations in CPI for the sample period are unexpected. The inflation rate of PPIs are not useful predictors of CPI inflation because the former are quite erratic while the latter is quite stable. This conclusion differs from the earlier literature on the transmission of PPI shocks to CPI, which are usually based on impulse responses of unrestricted VARs. The impulse responses are nonlinear functions of the decomposition of the precision matrix, $\mathbf{\Psi}$, and the regression coefficients. The unrestricted VAR estimates of regression coefficients often show alternating signs from one lag to another. For example, for the 1969-1980 sample the posterior mean (posterior standard deviation) of the CPI equation of the first and second lags of the PPIs: the first lag of FF .076(.043), the second lag of FF $-$.029(.044); first lag of CM $-$.091(.059), the second lag of CM .040(.059); first lag of ITM $-$.018(.059), the second lag of ITM .010(.058). The estimates of these coefficients in the restricted VAR are in much smaller scale. This pattern holds for both the pre 1980 and the post 1980 samples. There is a strong correlation between the unexpected CPI inflation and shocks to some components of PPI inflation. We will discuss this link in the $\mathbf{\Psi}$ matrix later. In the ITM equation, the second lag of CPI is selected about half of the time and the estimate is positive.

The posterior mean of the restriction index matrix $\boldsymbol{\gamma}$, the MLE and Bayesian estimate of $\boldsymbol{\phi}$ all indicate that there is limited pass-through of PPI inflation of foodstuff and feeds and PPI inflation of crude materials to PPI of finished goods and CPI. There is also limited pass-through of PPI inflation of intermediate materials to PPI of finished goods. For the sample 1969-1980 the estimates of $\mathbf{\Phi}$ show that the sum of the first four lags of FF in the CPI equation is .018 and the sum of the first four lags of FC in the CPI equation is .105.

The UNEMP equation is close to a random walk, with the first lag of UNEMP being the only important regressor. The FFR equation contains its own lag and the first lag of UNEMP, as well as distant lags of CPI. This pattern suggests that for the earlier sample, variations in FFR was largely endogenous in response to past fluctuations in unemployment rate and CPI. The notion of endogenous FFR is confirmed by Figure 3, which shows that in the FFR equation the first lag of UNEMP has a large negative estimate, the third lag of ITM as well as the fifth to seventh lags of

CPI have large and positive estimates. FFR is persistent, as indicated by its positive own lag.

While the pattern that the restricted VARs differs from unrestricted VARs holds for the 1981-2005 sample and the 1969-1980 sample, there are some differences in the estimates of restricted VARS for the two samples. Comparison of the right columns in Figures 4 and 3 show that the regression coefficients in the CF and CM equations are smaller for the $1981-2005$ sample than the $1969-1980$ sample. The estimated FFR equation is also different for the two samples. In the $1969-1980$ sample, FFR is negatively correlated with the lag of UNEMP, for the $1981-2005$, FFR is much closer to a random walk, and is not as strongly correlated with the lags of other variables.

For the $1969-1980$ period, the posterior mean of the restriction index matrix $\boldsymbol{\omega}$ is

$$\widehat{\boldsymbol{\omega}} = \begin{pmatrix} & \text{FF} & \text{CM} & \text{IM} & \text{FC} & \text{CPI} & \text{UNEMP} & \text{FFR} \\ \text{FF} & 1.000 & \mathbf{1.000} & .067 & .146 & .041 & .031 & .026 \\ \text{CM} & & 1.000 & \mathbf{0.357} & \mathbf{.790} & .039 & .034 & .028 \\ \text{IM} & & & 1.000 & .195 & \mathbf{.968} & .117 & .069 \\ \text{FC} & & & & 1.000 & \mathbf{.594} & .063 & .060 \\ \text{CPI} & & & & & 1.000 & \mathbf{.303} & .130 \\ \text{UNEMP} & & & & & & 1.000 & \mathbf{.645} \\ \text{FFR} & & & & & & & 1.000 \end{pmatrix}.$$

Most of the upper-triangular elements of the index $\boldsymbol{\omega}$ are close to zero. The significant (larger than 0.3 and in boldface) contemporaneous correlations are: FF with CM, CM with IM, IM with FC, IM and FC with CPI, CPI with UNEMP, and UNEMP with FFR. The corresponding posterior mean (posterior standard deviation in parentheses) of $\boldsymbol{\Psi}$ is

$$\widehat{\boldsymbol{\Psi}} = \begin{pmatrix} .295 & -1.739 & -0.067 & -.067 & .060 & .056 & .020 \\ (.019) & (.136) & (.112) & (.150) & (.077) & (.065) & (.063) \\ & 2.397 & -.233 & -.475 & -.022 & .000 & .004 \\ & (.183) & (.163) & (.228) & (.095) & (.086) & (.082) \\ & & 2.421 & -.139 & -1.054 & .081 & -.036 \\ & & (.186) & (.226) & (.339) & (.181) & (.127) \\ & & & 2.985 & -.533 & -.015 & -.002 \\ & & & (.232) & (.466) & (.122) & (.120) \\ & & & & 6.579 & -.337 & -.014 \\ & & & & (.567) & (.619) & (.288) \\ & & & & & 4.043 & -.746 \\ & & & & & (.315) & (.634) \\ & & & & & & 1.335 \\ & & & & & & (.110) \end{pmatrix}.$$

Consistent with the restriction selection result, most of the elements of $\boldsymbol{\Psi}$ are close to zero except for the elements with substantial probability of being selected. We use $\boldsymbol{\epsilon} = (\epsilon_{FF}, \epsilon_{CM}, \epsilon_{IM}, \epsilon_{FC}, \epsilon_{CPI}, \epsilon_{UNEMP}, \epsilon_{FFR})'$ to denote the VAR residuals (which can be interpreted as unexpected inflations)

and $\boldsymbol{u}$ to denote the structural shock vector. The covariance of $\boldsymbol{u}$ is the identity matrix. The point estimates of the $\boldsymbol{\Psi}$ matrix give restrictions on how regression residuals relate to one another. For instance, the fifth column of $\boldsymbol{\Psi}$ implies that $6.579\epsilon_{CPI} = -.060\epsilon_{CFF} + .022\epsilon_{CM} + 1.054\epsilon_{IM} + .533\epsilon_{FC} + u_{CPI}$, suggesting unexpected CPI inflation roughly equals the sum of one-sixth of that of intermediate materials PPI, one-twelveth of that of finished goods PPI, and an idiosyncratic shock with a standard deviation of about .152%.

In comparison to the posterior mean estimator of $\boldsymbol{\Psi}$, the MLE of $\boldsymbol{\Psi}$ shows much larger coefficients, which gives a much different picture of the relationship among unexpected inflations from the Bayesian estimates. For example, the fifth column of $\boldsymbol{\Psi}$ of the MLE is

$$8.096\epsilon_{CPI} = -1.898\epsilon_{FF} - 1.789\epsilon_{CM} + 4.189\epsilon_{IM} + 4.592\epsilon_{FC} + u_{CPI}.$$

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 the unexpected CPI inflation with the unexpected PPI inflation of foodstuff and feeds and with unexpected PPI inflation of crude materials but a positive one with the unexpected PPI inflation of intermediate materials and finished goods. These contemporaneous relationships are hard to explain given the literature discussed earlier in this section. The Bayesian estimates obtained through restriction averaging make more economic sense.

For the sample 1981 to 2005, the estimates of the contemporaneous relationship of VAR variables are as follows. The posterior mean of the restriction index matrix $\boldsymbol{\omega}$ is

$$\widehat{\boldsymbol{\omega}} = \begin{pmatrix} & \text{FF} & \text{CM} & \text{IM} & \text{FC} & \text{CPI} & \text{UNEMP} & \text{FFR} \\ \text{FF} & 1.000 & 0.062 & .022 & .024 & .023 & .020 & .025 \\ \text{CM} & & 1.000 & \mathbf{.324} & .027 & .021 & .023 & .019 \\ \text{IM} & & & 1.000 & \mathbf{1.000} & \mathbf{.935} & .074 & \mathbf{.408} \\ \text{FC} & & & & 1.000 & \mathbf{1.000} & .090 & .060 \\ \text{CPI} & & & & & 1.000 & \mathbf{.997} & \mathbf{.416} \\ \text{UNEMP} & & & & & & 1.000 & \mathbf{1.000} \\ \text{FFR} & & & & & & & 1.000 \end{pmatrix}.$$

The main difference between the pre and post 1981 sample is that post 1981 the shock in FF is no longer correlated with unexpected CM, and unexpected FFR is much more strongly correlated with unexpected shocks in other variables, IM, CPI, and UNEMP. This is in contrast to the random walk pattern of the FFR regression coefficients in the post 1981 sample. An explanation to the change in the lag and contemporaneous estimates is that the FFR has become the target variable of monetary policy since the early 1980s and has responded more quickly to unexpected changes in the economy,

whereas in the early sample its variations reflects more of passive adjustments to changes in other macroeconomic variables. As in the pre-1981 sample, the following post-1981 posterior mean of $\mathbf{\Psi}$ (with posterior standard deviation in parentheses) is quite different from the MLE.

$$\widehat{\mathbf{\Psi}} = \begin{pmatrix} \underset{(0.021)}{0.477} & \underset{(.030)}{-.144} & \underset{(.031)}{.029} & \underset{(.031)}{-.044} & \underset{(.032)}{.023} & \underset{(.030)}{.014} & \underset{(.031)}{-.058} \\ & \underset{(.015)}{.336} & \underset{(.025)}{-.249} & \underset{(.027)}{-.074} & \underset{(.029)}{.016} & \underset{(.024)}{.032} & \underset{(.029)}{.009} \\ & & \underset{(.141)}{2.995} & \underset{(.227)}{-2.386} & \underset{(.317)}{-.867} & \underset{(.121)}{-.050} & \underset{(.294)}{-.287} \\ & & & \underset{(.146)}{3.138} & \underset{(.234)}{-1.241} & \underset{(.125)}{.081} & \underset{(.115)}{.027} \\ & & & & \underset{(.369)}{7.686} & \underset{(.461)}{-1.815} & \underset{(.567)}{-.431} \\ & & & & & \underset{(.097)}{2.204} & \underset{(.261)}{-4.610} \\ & & & & & & \underset{(.096)}{2.034} \end{pmatrix}.$$

The existing literature that are based on MLE or unrestricted VARs show erratic coefficients in PPI and CPI inflation equations. In comparison, the Bayesian estimates based on stochastic search are much more stable. The empirical application of the stochastic search approach yields additional insights to the relationship between PPI and CPI inflations.

6 Concluding Remarks

Vector Autoregressive models have been widely used for macroeconomic forecasting and policy analysis. It is recognized that the conventional unconstrained VARs are over-parameterized. To remedy this problem, we propose a Bayesian stochastic search for VAR model restrictions. For this purpose, we develop an MCMC algorithm that visits high posterior probability elements of both the VAR regression coefficients matrix and error variance matrix. Numerical simulations show that the stochastic search algorithm is quite effective in visiting the data generating model in some cases and can yield improved predictions even when it does not.

To illustrate the important practical potential of our approach, we apply the method for selection among models of inflation transmission from the PPI of crude materials to the PPI of finished goods and the CPI. We find that for the sample period from 1969 to 2005, the Bayesian estimates of a conventional unconstrained VAR and a restricted VAR via stochastic search depict different dynamics relationship of PPI and CPI inflations as well as different contemporaneous relationships of unexpected inflations.

A number of methodological issues remain. As is typical of stochastic search applications, the number of restrictions actually visited by the MCMC simulation is only a small portion of all possible restrictions. Because of the enormous number of potential restrictions for a typical VAR model, the general issue of mixing should be investigated. Although the similar results we obtained over many different runs are comforting, our limited investigations has not really shown that our stochastic search will avoid getting stuck in a posterior mode. As we noted in the introduction, VAR models may be too restrictive for some economic data. It is useful to examine the performance of the proposed stochastic search method in more complicated models with time-varying regression coefficients and variance parameters such as those examined by West & Harrison (1989) and applied by Quintana & Putnam (1996).

7 Appendix: Technical Issues

7.1 An alternative framework for restriction selection for Ψ

The VAR model of p-column vector $\boldsymbol{y}_t$ is

$$\boldsymbol{y}_t' = \boldsymbol{z}_t'\boldsymbol{C} + \sum_{j=1}^{L} \boldsymbol{y}_{t-j}'\boldsymbol{A}_j + \boldsymbol{\epsilon}_t', \tag{46}$$

for $t = 1, \cdots, T$, where $\boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_T$ are independent identically distributed $N_p(\mathbf{0}, \boldsymbol{\Sigma})$. To decompose $\boldsymbol{\Sigma}$ in the framework of Pourahmadi (1999), consider the following regression:

$$\epsilon_{tj} \quad = \quad \sum_{k=1}^{j-1} \theta_{jk}\epsilon_{tk} + \xi_{tj} \tag{47}$$

where $j = 1, \cdots, p$ and $t = 1, \cdots, T$, and where $\mathbb{E}(\boldsymbol{\xi}_t\boldsymbol{\xi}_t') = \boldsymbol{\Lambda} = diag(\lambda_1, \cdots, \lambda_p)$. Writing (47) in matrix form we obtain

$$\boldsymbol{\Gamma}\boldsymbol{\epsilon}_\mathbf{t} = \boldsymbol{\xi}_t, \tag{48}$$

where $\boldsymbol{\epsilon}_t = (\epsilon_{t1}, \cdots, \epsilon_{tp})'$ and

$$\boldsymbol{\Gamma} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -\theta_{21} & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ -\theta_{p1} & -\theta_{p2} & \cdots & 1 \end{pmatrix}. \tag{49}$$

From (48) it follows that lower triangular matrix $\boldsymbol{\Gamma}$ diagonalizes $\boldsymbol{\Sigma}$.

$$\boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}' = \boldsymbol{\Lambda}. \tag{50}$$

The idea of the regression based decomposition of the covariance matrix is similar to the approach taken in this paper, but there is a difference between the two approaches. Regression (47) involves the errors of the VAR, which are not directly observable and need to be computed by through regression coefficients $\boldsymbol{C}, \boldsymbol{A}_1, \cdots, \boldsymbol{A}_L$ in each MCMC cycle. If we conduct our SSVS restriction selection of (47), we will develop a recursive structure of conditional posterior similar to what we did in this paper. In particular, the conditional posterior column vectors of $\boldsymbol{\Gamma}'$ are $f(\boldsymbol{\Gamma}'|\boldsymbol{Y}, \boldsymbol{\theta})$, ($\boldsymbol{\theta}$ represents the other parameters in the model) can be rewritten as $\prod_j f(\boldsymbol{\gamma}_j|\boldsymbol{S}_j, \boldsymbol{\lambda}_j)$, ($\boldsymbol{S}_j$ represents the jth principle minor of $\boldsymbol{S}$, the sq of MLE errors). The selection on $\boldsymbol{\Psi}$ in this paper is more straightforward than conducting our SSVS restriction selection on (47) for $\boldsymbol{\Gamma}'$.

7.2 Proof of Fact 1.

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

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

where $\boldsymbol{\mu}_j$ is defined in (26) and $\boldsymbol{\Delta}_j$ in (27). Part (a) is obvious. For part (b), the case when $i = 1$ clearly holds; when $i = 2, \cdots, p$ the result follows by integrating out $\boldsymbol{\eta}_j$. □

7.3 Algorithm with restrictions on both Φ and Ψ

Combining all the conditional distributions, the following MCMC stochastic search algorithm for sampling from the full posterior distribution $(\boldsymbol{\psi}, \boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\omega}; Y)$ is obtained. Given initial values $\boldsymbol{\psi}_0, \boldsymbol{\eta}_0, \boldsymbol{\phi}_0, \boldsymbol{\gamma}_0, \boldsymbol{\omega}_0$, the kth cycle is obtained from $\boldsymbol{\psi}_{(k-1)}, \boldsymbol{\eta}_{(k-1)}, \boldsymbol{\phi}_{(k-1)}, \boldsymbol{\gamma}_{(k-1)}, \boldsymbol{\omega}_{(k-1)}$ by sequentially simulating the steps:

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

$$
(\psi_{ii}^2 \mid \boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\omega}; \boldsymbol{Y}) \sim gamma(a_i + \frac{1}{2}T, B_i), \tag{51}
$$

where

$$B_i = \begin{cases} b_1 + \frac{1}{2}s_{11}, & \text{if } i = 1, \\ b_i + \frac{1}{2}\Big\{s_{ii} - \boldsymbol{s}_i'[\boldsymbol{S}_{i-1} + (\boldsymbol{D}_i\boldsymbol{R}_i\boldsymbol{D}_i)^{-1}]^{-1}\boldsymbol{s}_i\Big\}, & \text{if } i = 2, \cdots, p. \end{cases} \tag{52}$$

2. draw $(\boldsymbol{\eta}_{(k)} \mid \boldsymbol{\psi}_{(k)}, \boldsymbol{\phi}_{(k-1)}, \boldsymbol{\gamma}_{(k-1)}, \boldsymbol{\omega}_{(k-1)}; \boldsymbol{Y})$ from normal distribution (53);

$$(\boldsymbol{\eta}_j \mid \boldsymbol{\phi}, \boldsymbol{\gamma}; \boldsymbol{\omega}, \boldsymbol{\psi}; \boldsymbol{Y}) \sim N_{j-1}(\boldsymbol{\mu}_j, \boldsymbol{\Delta}_j), \tag{53}$$

where

$$\boldsymbol{\mu}_j = -\psi_{jj}\{\boldsymbol{S}_{j-1} + (\boldsymbol{D}_j\boldsymbol{R}_j\boldsymbol{D}_j)^{-1}\}^{-1}\boldsymbol{s}_j, \tag{54}$$

$$\boldsymbol{\Delta}_j = \{\boldsymbol{S}_{j-1} + (\boldsymbol{D}_j\boldsymbol{R}_j\boldsymbol{D}_j)^{-1}\}^{-1}. \tag{55}$$

3. draw $(\boldsymbol{\omega}_{(k)} \mid \boldsymbol{\eta}_{(k)}, \boldsymbol{\psi}_{(k)}, \boldsymbol{\phi}_{(k-1)}, \boldsymbol{\gamma}_{(k-1)}, \boldsymbol{\omega}_{(k-1)}; \boldsymbol{Y})$ from Bernoulli distribution (30).

4. draw $(\boldsymbol{\phi}_{(k)} \mid \boldsymbol{\gamma}_{(k-1)}, \boldsymbol{\Sigma}_{(k)}, \boldsymbol{\omega}_{(k)}; \boldsymbol{Y})$ from normal distribution (56), where $\boldsymbol{\Sigma}_{(k)}$ is computed from $\boldsymbol{\psi}_{(k)}$ and $\boldsymbol{\eta}_{(k)}$

$$(\boldsymbol{\phi} \mid \boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{\psi}; \boldsymbol{Y}) \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Delta}), \tag{56}$$

where

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

5. draw $(\boldsymbol{\gamma}_{(k)} \mid \boldsymbol{\phi}_{(k)}, \boldsymbol{\psi}_{(k)}, \boldsymbol{\eta}_{(k)}, \boldsymbol{\omega}_{(k)}; \boldsymbol{Y})$ from Bernoulli distribution (37).

7.4 Average MSE for point forecasts

A standard frequentist criterion for forecast evaluation is the average quadratic deviation of the forecast from the actual value. The j-step-ahead forecast error at period T can be decomposed into two orthogonal parts:

$$\boldsymbol{y}_{T+j} - \widehat{\boldsymbol{y}}_{T+j} \mid \widehat{\boldsymbol{\Phi}} = (\boldsymbol{y}_{T+j} - \widehat{\boldsymbol{y}}_{T+j} \mid \boldsymbol{\Phi}) + (\widehat{\boldsymbol{y}}_{T+j} \mid \boldsymbol{\Phi} - \widehat{\boldsymbol{y}}_{T+j} \mid \widehat{\boldsymbol{\Phi}}),$$

where $\widehat{\boldsymbol{y}}_{T+j} \mid \boldsymbol{\Phi}$ and $\widehat{\boldsymbol{y}}_{T+j} \mid \widehat{\boldsymbol{\Phi}}$ are the forecasts conditional on observations up to period T. They can be calculated from their corresponding VAR models by setting the error term to zero after period T.

The first term in the right-hand-side above is the sampling error, and the second term is the forecasting error attributable to the deviation of the estimates from the true parameters. When the

true parameters are known, as is the case in our Section 4 simulation examples, this second term can be calculated for alternative estimators. For the comparison of forecasting procedures it suffices to compare the mean squared errors (MSE) of the second term since the first term does not depend on $\widehat{\boldsymbol{\Phi}}$. For N sample estimates $\widehat{\boldsymbol{\Phi}}^{(1)}, \ldots, \widehat{\boldsymbol{\Phi}}^{(N)}$, the average one-step-ahead MSE is

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

7.5 Kullback-Liebler divergence for predictive densities

In VAR modelling applications, a predictive density provides a more comprehensive forecast than simple point estimates. For a VAR model in matrix form (4), namely $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\Phi} + \boldsymbol{\epsilon}$, this is the problem of predicting a future observation $\boldsymbol{Z}$ based on historical data $\boldsymbol{Y}$ by an estimate of its density $f(\boldsymbol{Z} \mid \boldsymbol{Y}, \boldsymbol{\Phi}, \boldsymbol{\Sigma})$. Here $\boldsymbol{Z}$ is a $g \times p$ matrix where g is the forecasting horizon. Let $\boldsymbol{\Phi} = (\boldsymbol{c}, \boldsymbol{B}_1, \cdots \boldsymbol{B}_L)$ and $\boldsymbol{z}_j'$ be the jth row of $\boldsymbol{Z}$. If $g \geq L$, then under (4)

$(\boldsymbol{z}_1 \mid \boldsymbol{\Phi}, \boldsymbol{\Sigma}; \boldsymbol{Y}) \sim N_p(\boldsymbol{c} + \boldsymbol{B}_1 \boldsymbol{y}_T + \boldsymbol{B}_2 \boldsymbol{y}_{T-1} + \cdots + \boldsymbol{B}_L \boldsymbol{y}_{T-L+1}, \boldsymbol{\Sigma})$;

$(\boldsymbol{z}_2 \mid \boldsymbol{z}_1, \boldsymbol{\Phi}, \boldsymbol{\Sigma}; \boldsymbol{Y}) \sim N_p(\boldsymbol{c} + \boldsymbol{B}_1 \boldsymbol{z}_1 + \boldsymbol{B}_2 \boldsymbol{y}_T + \cdots + \boldsymbol{B}_L \boldsymbol{y}_{T-L+2}, \boldsymbol{\Sigma})$;

$\cdots$,

$(\boldsymbol{z}_g \mid \boldsymbol{z}_{g-1}, \cdots, \boldsymbol{z}_1, \boldsymbol{\Phi}, \boldsymbol{\Sigma}; \boldsymbol{Y}) \sim N_p(\boldsymbol{c} + \boldsymbol{B}_1 \boldsymbol{z}_{g-1} + \boldsymbol{B}_2 \boldsymbol{z}_{g-2} + \cdots + \boldsymbol{B}_L \boldsymbol{z}_{g-L}, \boldsymbol{\Sigma})$.

The multi-step prediction of $\boldsymbol{Z}$ conditional on the model, parameters, and data can be written in a hierarchical fashion. The case when $g < L$ can be derived analogously. Thus the predictive density $f(\boldsymbol{Z} \mid \boldsymbol{\Phi}, \boldsymbol{\Sigma}; \boldsymbol{Y}) = f(\boldsymbol{z}_g, \boldsymbol{z}_{g-1}, \cdots, \boldsymbol{z}_1 \mid \boldsymbol{\Phi}, \boldsymbol{\Sigma}; \boldsymbol{Y})$ can be obtained via the iterated expression

$$\begin{aligned} f(\boldsymbol{z}_g, \boldsymbol{z}_{g-1}, \cdots, \boldsymbol{z}_1 \mid \boldsymbol{\Phi}, \boldsymbol{\Sigma}; \boldsymbol{Y}) = f(\boldsymbol{z}_g \mid \boldsymbol{z}_{g-1}, \cdots, \boldsymbol{z}_{g-L}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}; \boldsymbol{Y}) f(\boldsymbol{z}_{g-1} \mid \boldsymbol{z}_{g-2}, \cdots, \boldsymbol{z}_{g-L+1}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}; \boldsymbol{Y}) \\ \cdots f(\boldsymbol{z}_{L+1} \mid \boldsymbol{z}_L \cdots, \boldsymbol{z}_1, \boldsymbol{\Phi}, \boldsymbol{\Sigma}; \boldsymbol{Y}) \cdots f(\boldsymbol{z}_2 \mid \boldsymbol{z}_1, \boldsymbol{\Phi}, \boldsymbol{\Sigma}, \boldsymbol{Y}) f(\boldsymbol{z}_1 \mid \boldsymbol{\Phi}, \boldsymbol{\Sigma}; \boldsymbol{Y}). \end{aligned}$$

Letting $\pi(\boldsymbol{\Sigma}, \boldsymbol{\Phi} \mid M_k)$ be the prior distribution on $\boldsymbol{\Sigma}, \boldsymbol{\Phi}$ under model M_k, the Bayes predictive density estimate given M_k is given by

$$\pi(\boldsymbol{Z} \mid M_k, \boldsymbol{Y}) = \int \pi(\boldsymbol{Z} \mid M_k, \boldsymbol{\Sigma}, \boldsymbol{\Phi}; \boldsymbol{Y}) \pi(\boldsymbol{\Sigma}, \boldsymbol{\Phi} \mid M_k, \boldsymbol{Y}) d(\boldsymbol{\Sigma}, \boldsymbol{\Phi}). \tag{57}$$

And for model uncertainty further reflected by a prior $\pi(M_k)$, $k = 1, \ldots, K$, the Bayes predictive density estimate is

$$\pi(\boldsymbol{Z} \mid \boldsymbol{Y}) = \sum_{k=1}^{K} \pi(\boldsymbol{Z} \mid M_k, \boldsymbol{Y}) \pi(M_k \mid \boldsymbol{Y}). \tag{58}$$

The integration for marginal likelihood in (57) is costly, and given the number of candidate models, the evaluation of the predictive density over the entire model space will not be feasible even in

moderate sized problems. Fortunately, we can approximate the predictive density $\pi(\boldsymbol{Z} \mid \boldsymbol{Y})$ in (58) by the average of the conditional predictive densities visited by our MCMC stochastic search, namely

$$\hat{\pi}(\boldsymbol{Z} \mid \boldsymbol{Y}) \quad = \quad \frac{1}{A} \sum_{i=1}^{A} \pi(\boldsymbol{Z} \mid M_{i(k)}, \boldsymbol{\Sigma}_{i(k)}, \boldsymbol{\Phi}_{i(k)}; \boldsymbol{Y}), \tag{59}$$

where $i = 1, \cdots, A$ are the MCMC cycles. (The notation $M_{i(k)}$ means that model k is visited in the ith cycle and $(\boldsymbol{\Sigma}_{i(k)}, \boldsymbol{\Phi}_{i(k)})$ the parameter $(\boldsymbol{\Sigma}, \boldsymbol{\Phi})$ of model k in the ith cycle). Variation in the number of MCMC cycles A gives rise to very similar results, indicating that our approximation is reasonably good.

It follows from the basic results of Aitchison (1975), that the Bayes predictive density $\pi(\boldsymbol{Z} \mid \boldsymbol{Y})$ minimizes the Bayes risk under Kullback-Leibler loss

$$E_\pi E_Z KL(f; \hat{f}) = E_\pi E_Z \int \log\{f(\boldsymbol{Z} \mid \boldsymbol{\Phi}, \boldsymbol{\Sigma}; \boldsymbol{Y}) / \hat{f}(\boldsymbol{Z} \mid \boldsymbol{Y})\} f(\boldsymbol{Z} \mid \boldsymbol{\Phi}, \boldsymbol{\Sigma}, \boldsymbol{Y}) d\boldsymbol{Z},$$

where the expectation here is over $\boldsymbol{Z}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}$. This fact has used by Madigan & Raftery (1994) and others to argue that under a prior which reflects model uncertainty it is wiser to average the predictive density over the model space than to condition on any single model.

Finally, for our comparison of density predictors for $\boldsymbol{Z}$ in Section 4, we used the Kullback-Liebler divergence between the predictive density obtained through averaging restricted models, $\pi_{RB}(\boldsymbol{Z} \mid \boldsymbol{Y})$, and the predictive density obtained with the single unrestricted model, $\pi_{UB}(\boldsymbol{Z} \mid \boldsymbol{Y})$,

$$KL(\pi_{RB}; \pi_{UB}) = \int \log\{\pi_{RB}(\boldsymbol{Z} \mid \boldsymbol{Y}) / \pi_{UB}(\boldsymbol{Z} \mid \boldsymbol{Y})\} \pi_{RB}(\boldsymbol{Z} \mid \boldsymbol{Y}) d\boldsymbol{Z}.$$

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.

Brown, P. J., Vannucci, M. & Fearn, T. (1998), 'Multivariate Bayesian variable selection and prediction', *Journal of the Royal Statistical Society, Series B, Methodological* **60**, 627–641.

Chib, S. (1995), 'Marginal likelihood from the Gibbs output', *Journal of the American Statistical Association* **90**, 1313–1321.

Clark, T. E. (1995), 'Do producer prices lead consumer prices?', *Federal Reserve Bank of Kansas City Economic Review* **Third Quarter**, 25–39.

Clyde, M. & George, E. (2004), 'Model uncertainty', *Statistical Science* **19**(1), 81–94.

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**, 1304–1308.

George, E. I. & McCulloch, R. E. (1993), 'Variable selection via Gibbs sampling', *Journal of the American Statistical Association* **88**, 881–889.

George, E. I. & McCulloch, R. E. (1997), 'Approaches to bayesian variable selection', *Statistica Sinica* **7**, 339–379.

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.

Jeffreys, H. (1961), *Theory of Probability*, Oxford: Clarendon Press.

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, Working Paper Series 113* .

McNees, S. K. (1986), 'Forecasting accuracy of alternative techniques: A comparison of U.S. macroeconomic forecasts (C/R: p16-23)', *Journal of Business and Economic Statistics* **4**, 5–15.

Pourahmadi, M. (1999), 'Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation', *Biometrika* **86**, 677–690.

Quintana, J. M. & Putnam, B. H. (1996), Debating currency markets efficiency using dynamic multiple-factor models, *in* 'ASA Proceedings of the Section on Bayesian Statistical Science', pp. 55–80.

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.

Smith, M. & Kohn, R. (2002), 'Parsimonious covariance matrix estimation for longitudinal data', *Journal of the American Statistical Association* **97**(460), 1141–1153.

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.

West, M. & Harrison, J. (1989), *Bayesian forecasting and dynamic models*, Springer-Verlag Inc.

Yang, R. & Berger, J. O. (1994), 'Estimation of a covariance matrix using the reference prior', *The Annals of Statistics* **22**, 1195–1211.

Zellner, A., Keuzenkamp, H. & McAleer, M. e. (2001), *Simplicity, Inference and Modelling: Keeping It Sophisticatedly Simple*, Cambridge: Cambridge University Press.

Zellner, A. & Palm, F. C. (1975), 'Time series analysis of structural monetary models of the u.s. economy', *Sankya, Series C* **37**, 12–56.

Zellner, A. & Palm, F. C. (2004), *The Structural Econometric Time Series Analysis Approach*, Cambridge: Cambridge University Press.

Zellner, A. & Tiao, G. C. (1964), 'Bayesian analysis of the regression model with autocorrelated errors', *Journal of the American Statistical Association* **59**, 763–778.

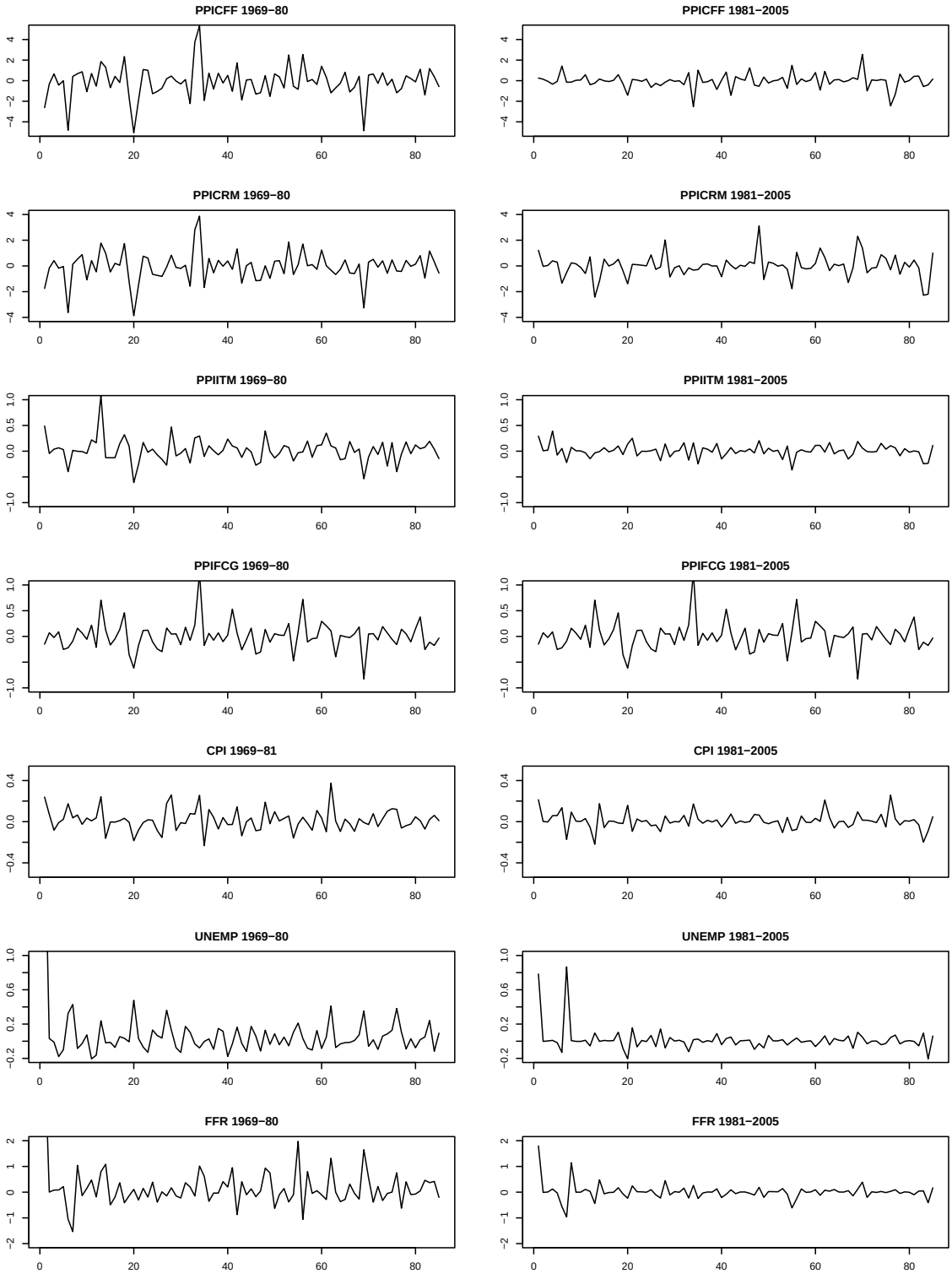

Figure 1: the MLE of regression coefficient $\phi = vec(\mathbf{\Phi})$ of the seven variable VAR of 12 lags. The left column pertains to the sample 1969-1980. The right column pertains to the sample 1981-2005.

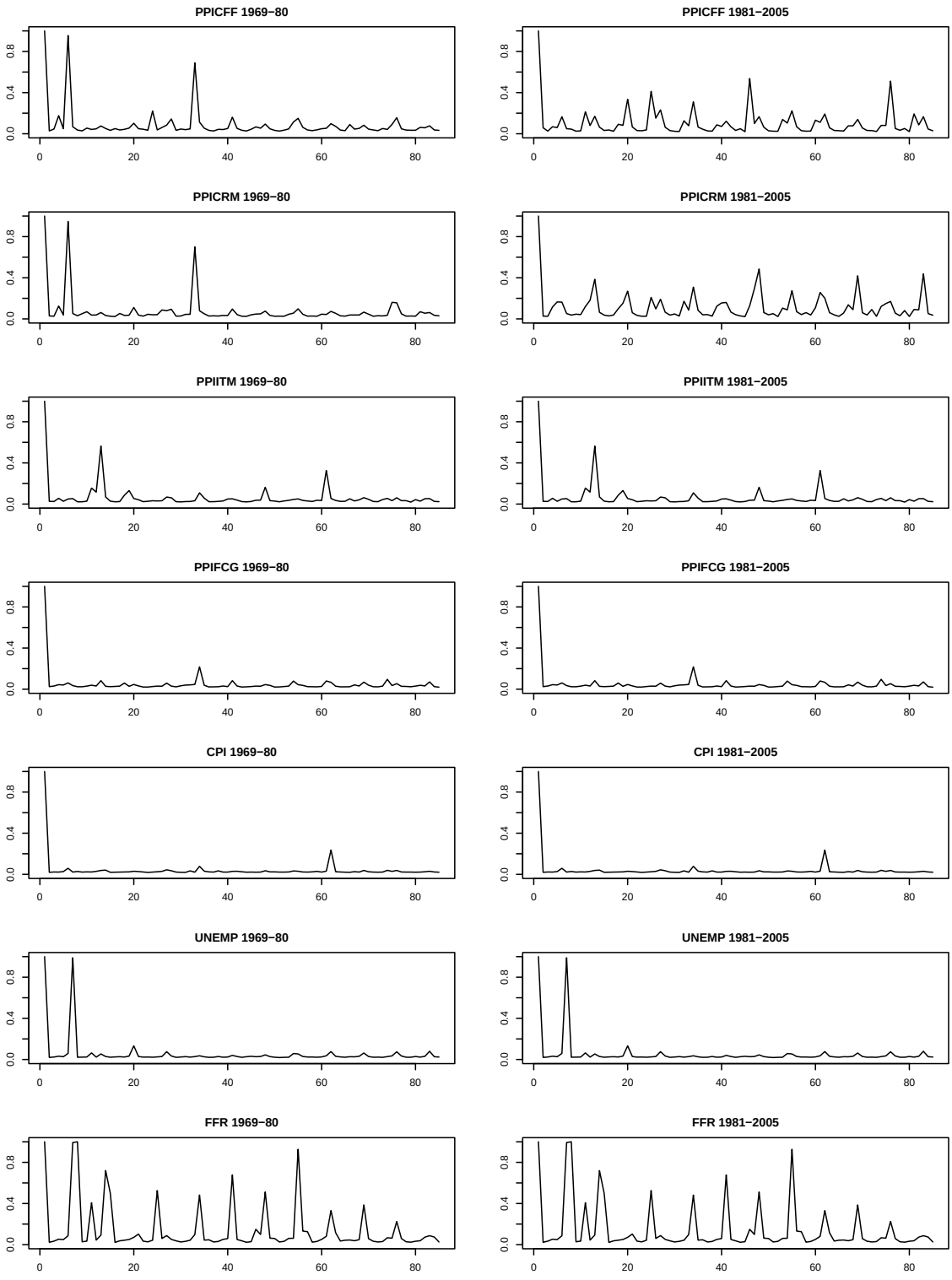

Figure 2: Posterior mean of the model selection index corresponding to ϕ. The left column pertains to the sample 1969-1980. The right column pertains to the sample 1981-2005.

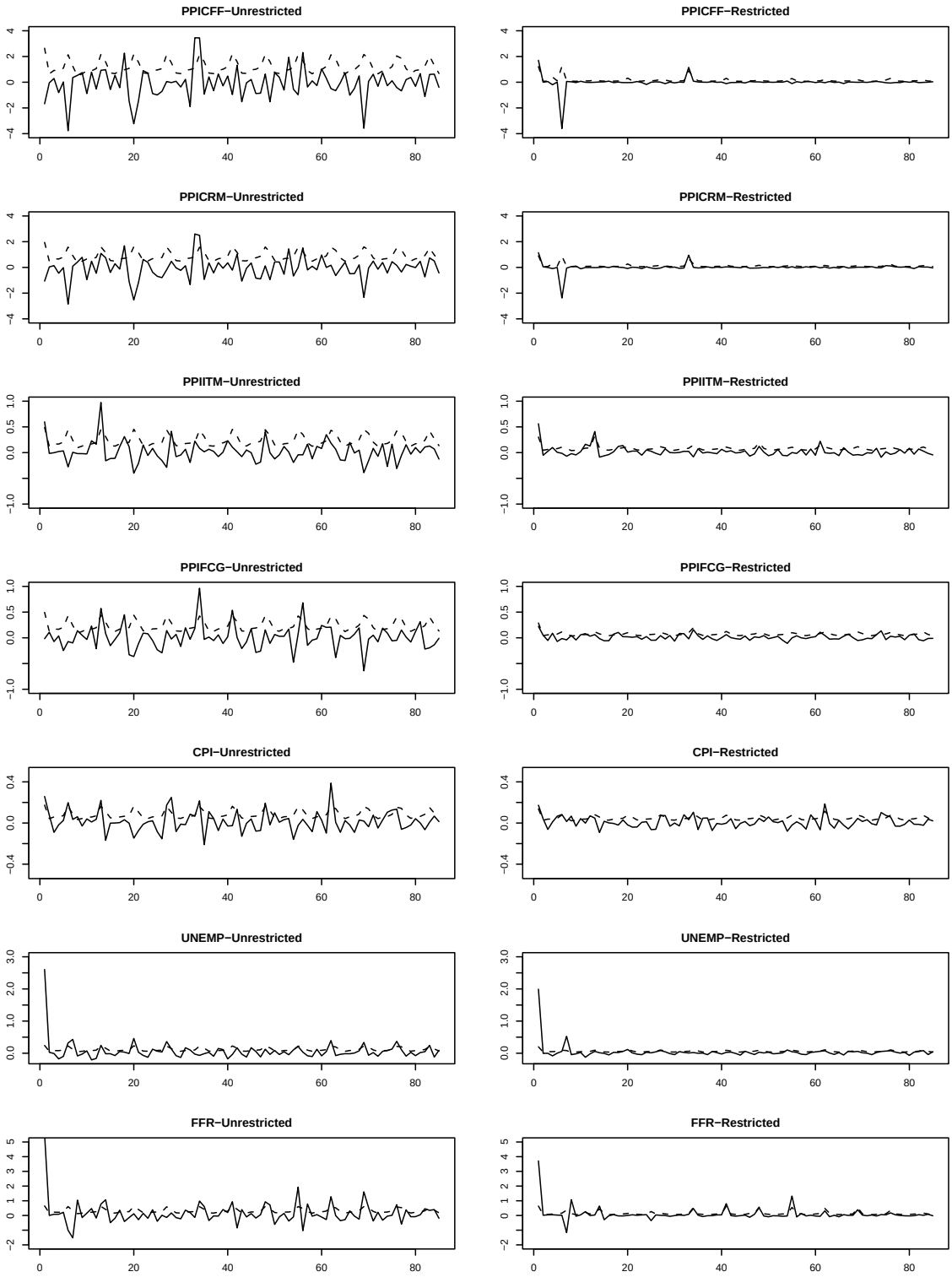

Figure 3: Plot of $\phi = vec(\mathbf{\Phi})$, sample: 1969:1-1980:12. The left column: parameters of unrestricted VAR. The right column: restricted VAR. In each graph in either column, the solid line is the posterior mean of the parameters corresponding to the equation, the dashed line the posterior standard deviation of parameters.

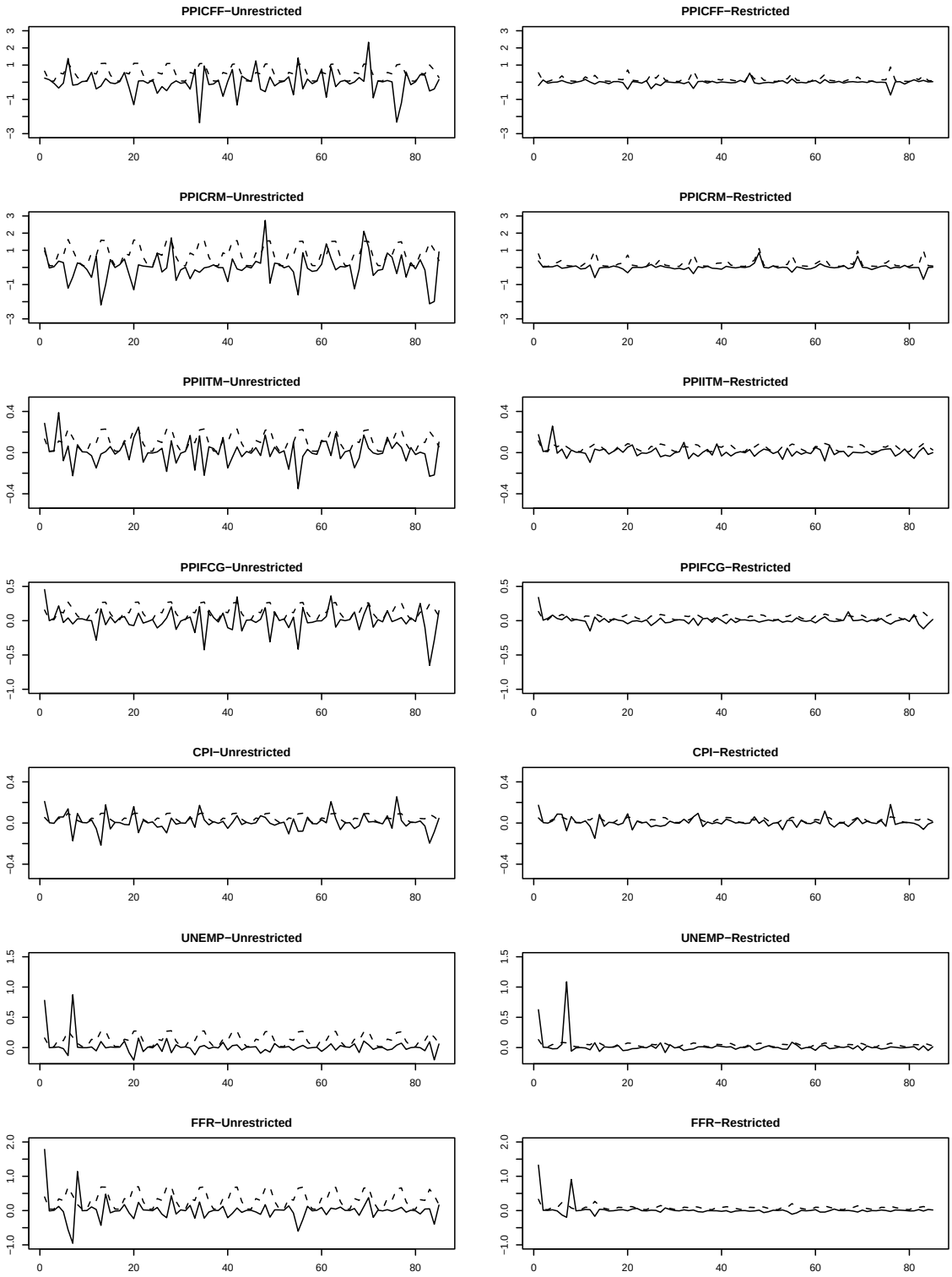

Figure 4: Plot of $\boldsymbol{\phi} = vec(\boldsymbol{\Phi})$, sample: 1981:1-2005:8. The left column: parameters of unrestricted VAR. The right column: restricted VAR. In each graph in either column, the solid line is the posterior mean of the parameters corresponding to the equation, the dashed line the posterior standard deviation of parameters.