Hierarchical Bayesian Analysis of the Seemingly Unrelated Regression and Simultaneous Equations Models Using a Combination of Direct Monte Carlo and Importance Sampling Techniques

Tomohiro Ando* and Arnold Zellner[†]

Abstract. Computationally efficient simulation methods for hierarchical Bayesian analysis of the seemingly unrelated regression (SUR) and simultaneous equations models (SEM) are proposed and applied. These methods combine a direct Monte Carlo (DMC) approach and an importance sampling procedure to calculate Bayesian estimation and prediction results, namely, Bayesian posterior densities for parameters, predictive densities for future values of variables and associated moments, intervals and other quantities. The results obtained by our approach are compared to those yielded by use of MCMC techniques. Finally, we show that our algorithm can be applied to the Bayesian analysis of state space models.

Keywords: Bayesian estimation and Prediction, Direct Monte Carlo, Hierarchical Priors Importance sampling, Markov Chain Monte Carlo

1 Introduction

The seemingly unrelated regression (SUR) model with fixed parameters was introduced by Zellner (1962, 1963) who used a generalized least squares approach. A Bayesian estimation approach for the SUR model was first introduced by Zellner (1971), followed later by various other techniques, e.g., the likelihood approach (Fraser et al., 2005), Bayesian analyses, the Bayesian method of moments, a Direct Monte Carlo approach (Zellner and Ando, 2008a) and so on. Applications of Markov-Chain Monte Carlo (MCMC) methodology to the SUR model, under various assumptions, have been conducted in many studies, including Percy (1992, 1996), Chib and Greenberg (1995), and Smith and Kohn (2000).

As well as the SUR model, the simultaneous equations model (SEM) has been widely employed to analyze the behavior of economies and other multivariate systems; see e.g., Zellner and Chen (2002), Aliprantis et al. (2007), Kibambe and Zellner (2007), and the references given therein.

---

*Graduate School of Business Administration, Keio University, Tokyo, Japan, mailto:andoh@kbs.keio.ac.jp

[†]Booth School of Business, University of Chicago, Chicago, IL, http://faculty.chicagobooth.edu/arnold.zellner/more/index.htm

In this paper we analyze SUR models and SEMs using a hierarchical Bayesian approach. One of the approaches for estimating these models is via MCMC simulations. Using MCMC, one can approximate the posterior densities for each of the parameters and construct a Bayesian predictive density for future observations that is useful for prediction. Thanks to recent advancement of computer technology, Bayesian analyses using MCMC techniques have become widely utilized. However the use of MCMC methods involves some problems. First, the length of burn-in period has to be determined. Second, we have to use an appropriate proposal density so that the MCMC algorithms have appropriate acceptance rates. Therefore, there are many papers that investigate acceptance rates for MCMC procedures; see, e.g. Roberts and Rosenthal (2001). Third, there is no universal rule for determining the number of MCMC samples to employ. Furthermore, we still have to check the convergence of MCMC algorithms. Although many methods have been proposed (Geweke (1992), Raftery and Lewis (1992, 1995), Heidelberger and Welch (1983), Schruben (1982), Gelman and Rubin (1992), Brooks and Gelman (1997), and Zellner and Min (1995)), there is no guarantee that the MCMC algorithms produce samples from the desired posterior distributions in a finite run.

Recently, Zellner and Ando (2008a, 2008b) developed a new efficient Bayesian estimation approach based on a direct Monte Carlo (DMC) approach (see e.g., Geweke (2005) for a general definition of a DMC algorithm) for the Bayesian analysis of the SUR model and SEM. The difference between our paper and Zellner and Ando (2008a, 2008b) is that in the former, instead of a Jeffreys's prior, a hierarchical prior on the coefficients is used for the SUR model and similarly for the SEM. This modification makes the inference problem more complicated. Unfortunately, the algorithms in Zellner and Ando (2008a, 2008b) are not applicable for the hierarchical Bayesian analysis of the SUR model and the SEM.

Herein we develop new efficient Bayesian estimation and prediction procedures that do not involve the computational problems of MCMC. The method combines a DMC and an importance sampling procedure. We show that use of the developed method permits easy computation of posterior densities of the parameters and predictive densities for future values. In a similar manner, Zellner and Ando (2008c) combined a DMC approach and an importance sampling procedure for Bayesian analysis of Student-t SUR models using a diffuse prior. Zellner and Ando (2008c) also showed that inequality restrictions can be incorporated in the approach. The main differences between Zellner and Ando (2008c) and our paper are that (a) we use a hierarchical prior on the coefficients, instead of a diffuse prior and (b) we treat both the SUR and SEM models, while Zellner and Ando (2008c) considered only the SUR model.

As pointed out by a referee, with importance sampling within DMC one can make use of i.i.d. sampling and Laws of Large numbers and Central Limit Theorems to analyze the accuracy of the results. Because the draws from our procedure are i.i.d., we do not need various types of MCMC convergence criteria based on statistical testing procedures, e.g., Geweke (1992)'s comparison of the equivalence of means calculated by draws from first part and the last part of output samples. We can easily perform other statistical tests on variances and other moments, where such tests are much more complicated for MCMC due to the autocorrelation in MCMC output.

The structure of the remainder of this paper is as follows. In section 2, we briefly review the SUR model and several Bayesian model estimation procedures. Section 3 presents an efficient estimation procedure for the SUR model. Also, our algorithm permits us to impose restrictions on the parameters' ranges. In Section 4, we provide an overview of the SEM and then show how our approach can be applied to the SEM. Numerical studies are conducted in Section 5 and 6. For comparative purposes the performance of our algorithm is compared to to that of an MCMC algorithm. We also point out that our approach can be applied to the Bayesian analysis of state space models. Section 7 concludes.

2 Preliminaries: SUR Model and Several Inference Procedures

2.1 Overview of SUR Model

The linear SUR model involves a set of regression equations with cross-equation parameter restrictions and correlated error terms having differing variances. Algebraically, the SUR model is given by:

$$\boldsymbol{y}_j = X_j \boldsymbol{\beta}_j + \boldsymbol{u}_j, \; j = 1, ..., m, \;\; \text{with} \;\; E[\boldsymbol{u}_i \boldsymbol{u}_j'] = \left\{ \begin{array}{ll} \omega_{ij} I, & (i \neq j) \\ \omega_i^2 I, & (i = j) \end{array} \right. , \tag{1}$$

Here $\boldsymbol{y}_j$ and $\boldsymbol{u}_j$ are $n \times 1$ vectors, X_j is a $n \times p_j$ matrix of rank p_j of observations, and $\boldsymbol{\beta}_j$ is a p_j-dimensional coefficient vector. The domains of parameter values are given as follows: $-\infty < \beta_{jr} < \infty$, $(r = 1, ..., p_j, j = 1, ..., m)$, $-\infty < \omega_{ij} < \infty$, $(i, j = 1, ..., m, i \neq j)$ and $0 < \omega_j < \infty$, $(j = 1, ..., m)$. As shown in (1), the equations have different independent variables and variances. Also, the model permits error terms in different equations to be correlated. In matrix form, the model can be expressed as $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{u}$, $\boldsymbol{u} \sim N(\boldsymbol{0}, \Omega \otimes I)$, where $N(\boldsymbol{\mu}, \Sigma)$ denotes the normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ, $\otimes$ is the tensor product, Ω is an $m \times m$ symmetric matrix with diagonal elements $\{\omega_1^2, ..., \omega_m^2\}$, and the off-diagonal ijth elements are ω_{ij}, $\boldsymbol{y}' = (\boldsymbol{y}_1', ..., \boldsymbol{y}_m')$, $X = \text{diag}\{X_1, ..., X_m\}$, $\boldsymbol{\beta}' = (\boldsymbol{\beta}_1', ..., \boldsymbol{\beta}_m')$ and $\boldsymbol{u}' = (\boldsymbol{u}_1', ..., \boldsymbol{u}_m')$.

The normal likelihood function is

$$L(\boldsymbol{y}|\boldsymbol{\beta}, \Omega) = \frac{1}{(2\pi)^{nm/2}|\Omega|^{n/2}} \exp\left[-\frac{1}{2}\text{tr}\left\{ R\Omega^{-1} \right\} \right],$$

where "tr" denotes the trace of a matrix, $|\Omega| = \det(\Omega)$ is the value of the determinant of Ω, the ijth element of the $m \times m$ matrix $R = (r_{ij})$ is $r_{ij} = (\boldsymbol{y}_i - X_i\boldsymbol{\beta}_i)'(\boldsymbol{y}_j - X_j\boldsymbol{\beta}_j)$.

The problem is how to estimate the model parameters. The maximum likelihood estimates of $\boldsymbol{\beta}$ and Ω are obtained by maximizing the likelihood function. Zellner (1962, 1963) considered the parameter estimation problem from the frequentist points of view. If Ω is known, a parameter estimate can be obtained by using the generalized least squares (GLS) approach, say $\hat{\boldsymbol{\beta}}$. In practice, however, $\hat{\boldsymbol{\beta}}$ depends on Ω that is usually unknown and thus "feasible" generalized least squared estimates have been

proposed. The ordinary least squares residuals for each equation can be used to estimate Ω consistently. Furthermore, the maximum likelihood estimates of $\boldsymbol{\beta}$ and Ω can be obtained by using an iterative SUR approach.

In the following sections we briefly review some past Bayesian studies.

2.2   Markov chain Monte Carlo Approach

Currently, one of the most widely used methods for calculating an approximation to the posterior of the SUR model is the MCMC approach that is described and applied in many recent Bayesian econometrics and statistics texts. Also, Zellner (1971), Press (1972), Box and Tiao (1973), Percy (1992), and Srivastava and Giles (1987) derived and studied the posterior distributions of the parameters of the normal SUR model.

In the absence of prior knowledge, use of Bayesian analysis with noninformative priors is very common in practice. One of the most widely used noninformative priors, introduced by Jeffreys (1946, 1961), is Jeffreys's invariant prior:

$$\pi_1(\boldsymbol{\beta}, \Omega) = \pi_1(\boldsymbol{\beta})\pi_1(\Omega) \propto |\Omega|^{-\frac{m+1}{2}}, \tag{2}$$

which is proportional to the square root of the determinant of Fisher information matrix. One of the advantages of the use of Jeffreys's prior is that it is invariant under any one-to-one reparameterization of the model.

The joint posterior density function is then given by Bayes' theorem as

$$g(\boldsymbol{\beta}, \Omega|D) \propto |\Omega|^{-(n+m+1)/2} \exp\left[-\frac{1}{2}\text{tr}\left\{R\Omega^{-1}\right\}\right].$$

The conditional posteriors $g(\boldsymbol{\beta}|\Omega, D)$ and $g(\Omega|\boldsymbol{\beta}, D)$ are

$$g(\boldsymbol{\beta}|\Omega, D) = N(\hat{\boldsymbol{\beta}}, \hat{\Omega}_\beta) \quad \text{and} \quad g(\Omega|\boldsymbol{\beta}, D) = IW(R, n), \tag{3}$$

where $IW(\cdot, \cdot)$ denotes the inverse Wishart distribution, and

$$\hat{\boldsymbol{\beta}} = \{X'(\Omega^{-1} \otimes I)X\}^{-1}X'(\Omega^{-1} \otimes I)\boldsymbol{y},$$
$$\hat{\Omega}_\beta = (X'(\Omega^{-1} \otimes I)X)^{-1},$$

Note that the conditional posteriors of $\boldsymbol{\beta}$ and Ω depend upon each other. Because the conditional posterior densities $\boldsymbol{\beta}|\Omega, \boldsymbol{y}$ and $\Omega|\boldsymbol{\beta}, \boldsymbol{y}$ are available, the standard SUR model is also amenable to a 2-block Gibbs sampling formulation; see e.g., Percy (1992).

Sometimes one has prior information regarding values of the coefficient vectors $\boldsymbol{\beta}$. In this case, one can use a normal prior for $\boldsymbol{\beta}$,

$$\pi_2(\boldsymbol{\beta}, \Omega) = \pi_2(\boldsymbol{\beta})\pi_2(\Omega), \quad \text{with} \quad \pi_2(\boldsymbol{\beta}) = N(\boldsymbol{\beta}_0, A_\beta^{-1}), \quad \pi_2(\Omega) \propto |\Omega|^{-\frac{m+1}{2}} \tag{4}$$

which leads to the following conditional posteriors

$$g(\boldsymbol{\beta}|\Omega, D) = N(\bar{\boldsymbol{\beta}}, \bar{\Omega}_\beta) \quad \text{and} \quad g(\Omega|\boldsymbol{\beta}, D) = IW(R, n),$$

with

$$\bar{\boldsymbol{\beta}} = (X'(\Omega^{-1} \otimes I)X + A_\beta)^{-1}(X'(\Omega^{-1} \otimes I)X\hat{\boldsymbol{\beta}} + A_\beta \boldsymbol{\beta}_0),$$
$$\bar{\Omega}_\beta = (X'(\Omega^{-1} \otimes I)X + A_\beta)^{-1}.$$

As well as in the above case, we have to use a numerical approach to compute marginal posterior densities, moments, etc.

Recently, Zellner and Ando (2008a) developed a new model estimation procedure based on a DMC method, building on earlier work by Zellner and Chen (2002). This DMC procedure can be applied to simultaneous equations models and many others. In the next section, we review their DMC approach.

2.3   A Direct Monte Carlo Approach

Zellner and Ando (2008a) derived a direct Monte Carlo procedure for Bayesian analysis of the SUR model. In their paper, the standard SUR model (1) is reformulated as follows:

$$\begin{cases} \boldsymbol{y}_1 = X_1\boldsymbol{\beta}_1 + \boldsymbol{e}_1 \equiv Z_1\boldsymbol{b}_1 + \boldsymbol{e}_1, \\ \boldsymbol{y}_j = X_j\boldsymbol{\beta}_j + \sum_{l=1}^{j-1} \rho_{jl}(\boldsymbol{y}_l - X_l\boldsymbol{\beta}_l) + \boldsymbol{e}_j \equiv Z_j\boldsymbol{b}_j + \boldsymbol{e}_j, \quad j = 2, ..., m, \end{cases} \quad (5)$$

where the $n \times (p_j + j - 1)$ matrices Z_j are functions of $\boldsymbol{\beta}_{j-1}, ..., \boldsymbol{\beta}_1$, and

$$E[\boldsymbol{e}_i\boldsymbol{e}_j'] = \begin{cases} O, & (i \neq j) \\ \sigma_i^2 I, & (i = j) \end{cases}, \quad \text{and} \quad \Sigma = \text{diag}\{\sigma_1^2, ..., \sigma_m^2\},$$

where O is a zero matrix.

Readers may wonder why we transform the standard SUR model (1) into (5). The reason is that we can draw the posterior samples directly from the joint posterior distribution as described below. It is true that one can use MCMC to draw posterior samples. However, as pointed out in Section 1, the use of MCMC usually involves many complicated decisions to be made by MCMC users. We thus consider the transformed SUR model (5) that permits use of a DMC approach.

Zellner, et al (1988), Zellner and Chen (2002) and Zellner and Ando (2008b) considered this transformation in the context of simultaneous equations modeling. A similar transformation was considered in Fraser et al. (2005). Note that the diagonal elements of Ω and Σ are different.

Zellner and Ando (2008a) pointed to the capability of transforming from the parameters of the transformed model in (5) back to the parameters of the original formulation in equation (1). There is a one to one relation between the parameters of the SUR model in (1) and those of the transformed model in (5). The likelihood function of the parameters $\boldsymbol{\theta}' = (\boldsymbol{b}_1', ..., \boldsymbol{b}_m', \sigma_1^2, ..., \sigma_m^2)'$ is

$$L(\boldsymbol{y}|\boldsymbol{b}, \Sigma) = \prod_{j=1}^m \frac{1}{(2\pi\sigma_j^2)^{n/2}} \exp\left[-\frac{(\boldsymbol{y}_j - Z_j\boldsymbol{b}_j)'(\boldsymbol{y}_j - Z_j\boldsymbol{b}_j)}{2\sigma_j^2}\right].$$

In contrast to the standard model (1), we can decompose the likelihood function thanks to $E[\boldsymbol{e}_i \boldsymbol{e}_j'] = O, (i \neq j)$. The prior density function specified in (2) expressed in terms of $\{\boldsymbol{b}, \Sigma\}$ is

$$\pi(\boldsymbol{b}, \Sigma) \propto |\Omega(\boldsymbol{b}, \Sigma)|^{-\frac{m+1}{2}} |J| = \prod_{j=1}^{m}(\sigma_j^2)^{-\frac{m+1}{2}} \times \prod_{j'=1}^{m-1}(\sigma_{j'}^2)^{m-j'} = \prod_{j=1}^{m}(\sigma_j^2)^{\frac{m-2j-1}{2}}, \qquad (6)$$

where $|J|$ is a Jacobian factor. The joint posterior density of parameters is then

$$\pi(\boldsymbol{\theta}|D) \propto \prod_{j=1}^{m}(\sigma_j)^{-(n-m+2j+1)} \exp\left[-\frac{(\boldsymbol{y}_j - Z_j\boldsymbol{b}_j)'(\boldsymbol{y}_j - Z_j\boldsymbol{b}_j)}{2\sigma_j^2}\right],$$

which is equivalent to the conditional normal inverse-gamma posterior

$$\pi(\boldsymbol{b}_j|\boldsymbol{b}_{j-1}, ..., \boldsymbol{b}_1, \sigma_j^2, D) = N\left(\hat{\boldsymbol{b}}_j, \sigma_j^2(Z_j'Z_j)^{-1}\right),$$
$$\pi(\sigma_j^2|\boldsymbol{b}_{j-1}, ..., \boldsymbol{b}_1, D) = IG\left(\hat{\gamma}_j/2, \hat{\nu}_j/2\right),$$

where for $j = 1, ..., m$, $IG(\cdot, \cdot)$ denotes the inverse Gamma distribution, and

$$\hat{\boldsymbol{b}}_j = \left(Z_j'Z_j\right)^{-1} Z_j'\boldsymbol{y}_j,$$
$$\hat{\gamma}_j = \left(\boldsymbol{y}_j - Z_j\hat{\boldsymbol{b}}_j\right)' \left(\boldsymbol{y}_j - Z_j\hat{\boldsymbol{b}}_j\right),$$
$$\hat{\nu}_j = n - m - p_j + j + 1.$$

Then a direct Monte Carlo sampling procedure for Bayesian analysis of the standard SUR model (Zellner and Ando (2008a)) is given as follows.

A direct Monte Carlo sampling procedure:

Step 1 (initialization). Fix the order of a set of m equations. Set the number of samples N to be generated. Set $j = 1$. Generate $\sigma_1^{2(k)}$, $k = 1, ..., N$, and insert the drawn values in $\pi(\boldsymbol{b}_1|\sigma_1^2, D)$. Then make a draw $\boldsymbol{b}_1^{(k)}$ from $\pi(\boldsymbol{b}_1|\sigma_1^{2(k)}, D)$, for $k = 1, ..., N$.

Step 2 Increase the iteration index j by one $j \rightarrow (j+1)$. Draw $\sigma_j^{(k)}$ from the conditional inverse gamma density $\pi(\sigma_j^2|\boldsymbol{b}_{j-1}^{(k)}, ..., \boldsymbol{b}_1^{(k)}, D)$, and then generate $\boldsymbol{b}_j^{(k)}$ from $\pi(\boldsymbol{b}_j|\boldsymbol{b}_{j-1}^{(k)}, ..., \boldsymbol{b}_1^{(k)}, \sigma_j^{(k)}, D)$, for $k = 1, ..., N$.

Step 3 Repeat Step 2 sequentially until $j = m$.

3 Hierarchical Bayesian Analysis of SUR Models

3.1 Hierarchical Bayesian SUR Modeling

Here, we again employ the transformed model (5). The likelihood function of the parameters $\boldsymbol{\theta}' = (\boldsymbol{b}_1', ..., \boldsymbol{b}_m', \sigma_1^2, ..., \sigma_m^2)'$ is

$$L(\boldsymbol{y}|\boldsymbol{b}, \Sigma) \;\; = \;\; \prod_{j=1}^{m} \frac{1}{(2\pi\sigma_j^2)^{n/2}} \exp\left[-\frac{(\boldsymbol{y}_j - Z_j\boldsymbol{b}_j)'(\boldsymbol{y}_j - Z_j\boldsymbol{b}_j)}{2\sigma_j^2} \right].$$

In this paper, we shall consider the informative prior specification (4) for $\boldsymbol{\beta}$ in the original model and introduce a hierarchical prior structure:

$$\pi(\boldsymbol{\beta}, \Omega, \boldsymbol{\lambda}) = \pi(\boldsymbol{\beta}|\boldsymbol{\lambda})\pi(\Omega)\pi(\boldsymbol{\lambda}). \tag{7}$$

Here we use a normal prior for $\boldsymbol{\beta}$, and a diffuse prior for Ω as given in Section 2.3, and a gamma prior $G(a_0, b_0)$ for $\pi(\lambda_j)$:

$$\pi(\boldsymbol{\beta}|\boldsymbol{\lambda}) \;\; \propto \;\; \exp\left\{ -\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\Gamma(\boldsymbol{\lambda}, A)(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \right\},$$

$$\pi(\Omega) \;\; \propto \;\; |\Omega|^{-(m+1)/2},$$

$$\pi(\boldsymbol{\lambda}) \;\; = \;\; \prod_{j=1}^{m} \pi(\lambda_j) \propto \prod_{j=1}^{m} (\lambda_j)^{a_0-1} \exp\{-\lambda_j/b_0\},$$

where $\boldsymbol{\beta}' = (\boldsymbol{\beta}_1', ..., \boldsymbol{\beta}_m')$, $\boldsymbol{\beta}_0' = (\boldsymbol{\beta}_{01}', ..., \boldsymbol{\beta}_{0m}')$ and

$$\Gamma^{-1}(\boldsymbol{\lambda}, A) \;\; = \;\; \begin{pmatrix} \lambda_1 A_{11} & \sqrt{\lambda_1\lambda_2}A_{12} & \cdots & \sqrt{\lambda_1\lambda_m}A_{1m} \\ \sqrt{\lambda_2\lambda_1}A_{21} & \lambda_2 A_{22} & \cdots & \sqrt{\lambda_2\lambda_m}A_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\lambda_m\lambda_1}A_{m1} & \sqrt{\lambda_m\lambda_2}A_{m2} & \cdots & \lambda_m A_{mm} \end{pmatrix}$$

$$\equiv \;\; \begin{pmatrix} \Gamma_{11} & \Gamma_{12} & \cdots & \Gamma_{1m} \\ \Gamma_{21} & \Gamma_{22} & \cdots & \Gamma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{m1} & \Gamma_{m2} & \cdots & \Gamma_{mm} \end{pmatrix}$$

is the covariance matrix of $\boldsymbol{\beta}$. Here λ_j are scale parameters for the β_j and A_{ij} is a pre-specified matrix. The specification of the hierarchical prior, including A_{ij} and $\boldsymbol{\beta}_{0j}$ might come from various sources of information, e,g., economic theory, knowledge of biology, past studies, intuition, and so on.

For convenience, we shall decompose this prior structure by using the following

transformation:

$$
\begin{cases}
\boldsymbol{\beta}_1 = \boldsymbol{\beta}_{01} + \boldsymbol{w}_1, \\
\boldsymbol{\beta}_2 = \boldsymbol{\beta}_{02} + \Gamma_{21}\Gamma_{11}^{-1}(\boldsymbol{\beta}_1' - \boldsymbol{\beta}_{01}')' + \boldsymbol{w}_2, \\
\boldsymbol{\beta}_3 = \boldsymbol{\beta}_{03} + \Gamma_{3,1\sim2}\Gamma_{1\sim2,1\sim2}^{-1}(\boldsymbol{\beta}_1' - \boldsymbol{\beta}_{01}', \boldsymbol{\beta}_2' - \boldsymbol{\beta}_{02}')' + \boldsymbol{w}_3, \\
\quad\vdots \\
\boldsymbol{\beta}_m = \boldsymbol{\beta}_{0m} + \Gamma_{m,1\sim m-1}\Gamma_{1\sim m-1,1\sim m-1}^{-1}(\boldsymbol{\beta}_1' - \boldsymbol{\beta}_{01}', ..., \boldsymbol{\beta}_{m-1}' - \boldsymbol{\beta}_{0m-1}')' + \boldsymbol{w}_m,
\end{cases} \tag{8}
$$

with

$$
\begin{cases}
\mathrm{Var}(\boldsymbol{w}_1) = \Gamma_{11} \equiv \lambda_1 W_1, \\
\mathrm{Var}(\boldsymbol{w}_2) = \Gamma_{22} - \Gamma_{12}\Gamma_{11}^{-1}\Gamma_{21} \equiv \lambda_2 W_2, \\
\mathrm{Var}(\boldsymbol{w}_3) = \Gamma_{33} - \Gamma_{3,1\sim2}\Gamma_{1\sim2,1\sim2}^{-1}\Gamma_{1\sim2,3} \equiv \lambda_3 W_3, \\
\quad\vdots \\
\mathrm{Var}(\boldsymbol{w}_m) = \Gamma_{m,m} - \Gamma_{m,1\sim m-1}\Gamma_{1\sim m-1,1\sim m-1}^{-1}\Gamma_{1\sim m-1,m} \equiv \lambda_m W_m
\end{cases} \tag{9}
$$

and

$$
\mathrm{Cov}\left(\boldsymbol{w}_i, \boldsymbol{w}_j | \Gamma, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_m\right) = O,
$$

which follows from direct verification. Here the $p_j \times p_j$ matrices Γ_{jj} are the j-th block diagonal elements of $\Gamma(\boldsymbol{\lambda}, A)$, and $\Gamma_{j,1\sim k-1}$ is given as $= (\Gamma_{j,1}, ..., \Gamma_{j,k-1})$. Also, $W_1 = A_{11}$, $W_2 = A_{22} - A_{12}A_{11}^{-1}A_{21}$, and so on. Then we have

$$
\pi(\boldsymbol{\beta}_j | \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_{j-1}, \lambda_j) \quad \propto \quad \exp\left\{-\frac{\lambda_j}{2}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_{0j}^*)' W_j (\boldsymbol{\beta}_j - \boldsymbol{\beta}_{0j}^*)\right\},
$$

with

$$
\boldsymbol{\beta}_{0j}^* = \boldsymbol{\beta}_{0j} + \Gamma_{j,1\sim j-1}\Gamma_{1\sim j-1,1\sim j-1}^{-1}(\boldsymbol{\beta}_1' - \boldsymbol{\beta}_{01}', \boldsymbol{\beta}_2' - \boldsymbol{\beta}_{02}', ..., \boldsymbol{\beta}_{j-1}' - \boldsymbol{\beta}_{0j-1}')'.
$$

This orthogonal transformation is seen to depend on the ordering of the equations. However, our algorithm is order invariant as shown in the next section.

Under this prior specification, we transform the prior (7) to that for $\{\boldsymbol{b}, \Sigma, \boldsymbol{\gamma}\}$ as follows. First, there is a one to one mapping between the first p_j elements of $\boldsymbol{b}_j$ and $\boldsymbol{\beta}_j$. Defining $\lambda_j = \gamma_j / \sigma_j^2$, the corresponding prior for $\boldsymbol{b}_j$ becomes

$$
\begin{aligned}
\pi(\boldsymbol{b}_j | \boldsymbol{b}_1, ..., \boldsymbol{b}_{j-1}, \sigma_j^2, \gamma_j) &= (2\pi)^{-(p_j - j + 1)/2}(\gamma_j/\sigma_j^2)^{(p_j - j + 1)/2}|D_j|_+^{1/2} \\
&\quad \times \exp\left\{-\frac{\gamma_j}{2\sigma_j^2}(\boldsymbol{b}_j - \boldsymbol{b}_{0j})' D_j (\boldsymbol{b}_j - \boldsymbol{b}_{0j})\right\},
\end{aligned}
$$

in which $|D|_+$ is the product of $(p - j + 1)$ nonzero eigenvalues of D_j and

$$
D_j \quad = \quad \begin{pmatrix} W_j & O_{p_j, j-1} \\ O'_{p_j, j-1} & O_{j-1, j-1} \end{pmatrix}
$$

with $O_{p_j,j-1}$ being a $p_j \times (j-1)$ dimensional zero matrix, and $\boldsymbol{b}_{0j} = (\boldsymbol{\beta}_{0j}^{*\prime}, 0, ..., 0)'$. The transformation of the prior $\pi(\Omega)\pi(\boldsymbol{\lambda})$ is then

$$\pi(\Sigma, \boldsymbol{\gamma}) \quad \propto \quad |\Sigma|^{-(m+1)/2} \left(\frac{\gamma_j}{\sigma_j^2}\right)^{a_0-1} \exp\left\{-\frac{\gamma_j}{\sigma_j^2 b_0}\right\} \times |J_m|,$$

and $|J_m|$ is a Jacobian term for the transformation given by,

$$|J_m| = (\sigma_1^2)^{m-1} \times (\sigma_2^2)^{m-2} \cdots \times (\sigma_{m-1}^2)^1 \times (\sigma_1^2 \times \cdots \times \sigma_m^2)^{-1} = \prod_{j=1}^{m}(\sigma_j^2)^{m-j-1}.$$

For example, the Jacobian of the transformation from $\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \omega_{11}^2, \omega_{12}^2, \omega_{12}, \lambda_1, \lambda_2\}$ to $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \sigma_1^2, \sigma_2^2, \gamma_1, \gamma_2\}$ is

$$|J_2| = \begin{vmatrix} \frac{\partial\boldsymbol{\beta}_1'}{\partial\boldsymbol{b}_1} & \frac{\partial\boldsymbol{\beta}_2'}{\partial\boldsymbol{b}_1} & \frac{\partial\omega_{12}}{\partial\boldsymbol{b}_1} & \frac{\partial\omega_1^2}{\partial\boldsymbol{b}_1} & \frac{\partial\omega_2^2}{\partial\boldsymbol{b}_1} & \frac{\partial\lambda_1}{\partial\boldsymbol{b}_1} & \frac{\partial\lambda_2}{\partial\boldsymbol{b}_1} \\ \frac{\partial\boldsymbol{\beta}_1'}{\partial\boldsymbol{b}_2} & \frac{\partial\boldsymbol{\beta}_2'}{\partial\boldsymbol{b}_2} & \frac{\partial\omega_{12}}{\partial\boldsymbol{b}_2} & \frac{\partial\omega_1^2}{\partial\boldsymbol{b}_2} & \frac{\partial\omega_2^2}{\partial\boldsymbol{b}_2} & \frac{\partial\lambda_1}{\partial\boldsymbol{b}_2} & \frac{\partial\lambda_2}{\partial\boldsymbol{b}_2} \\ \frac{\partial\boldsymbol{\beta}_1'}{\partial\sigma_1^2} & \frac{\partial\boldsymbol{\beta}_2'}{\partial\sigma_1^2} & \frac{\partial\omega_{12}}{\partial\sigma_1^2} & \frac{\partial\omega_1^2}{\partial\sigma_1^2} & \frac{\partial\omega_2^2}{\partial\sigma_1^2} & \frac{\partial\lambda_1}{\partial\sigma_1^2} & \frac{\partial\lambda_2}{\partial\sigma_1^2} \\ \frac{\partial\boldsymbol{\beta}_1'}{\partial\sigma_2^2} & \frac{\partial\boldsymbol{\beta}_2'}{\partial\sigma_2^2} & \frac{\partial\omega_{12}}{\partial\sigma_2^2} & \frac{\partial\omega_1^2}{\partial\sigma_2^2} & \frac{\partial\omega_2^2}{\partial\sigma_2^2} & \frac{\partial\lambda_1}{\partial\sigma_2^2} & \frac{\partial\lambda_2}{\partial\sigma_2^2} \\ \frac{\partial\boldsymbol{\beta}_1'}{\partial\gamma_1} & \frac{\partial\boldsymbol{\beta}_2'}{\partial\gamma_1} & \frac{\partial\omega_{12}}{\partial\gamma_1} & \frac{\partial\omega_1^2}{\partial\gamma_1} & \frac{\partial\omega_2^2}{\partial\gamma_1} & \frac{\partial\lambda_1}{\partial\gamma_1} & \frac{\partial\lambda_2}{\partial\gamma_1} \\ \frac{\partial\boldsymbol{\beta}_1'}{\partial\gamma_2} & \frac{\partial\boldsymbol{\beta}_2'}{\partial\gamma_2} & \frac{\partial\omega_{12}}{\partial\gamma_2} & \frac{\partial\omega_1^2}{\partial\gamma_2} & \frac{\partial\omega_2^2}{\partial\gamma_2} & \frac{\partial\lambda_1}{\partial\gamma_2} & \frac{\partial\lambda_2}{\partial\gamma_2} \end{vmatrix}$$

$$= \begin{vmatrix} I & O & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ O & \begin{pmatrix} I \\ \boldsymbol{0} \end{pmatrix} & \begin{pmatrix} \boldsymbol{0} \\ \sigma_1^2 \end{pmatrix} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0}' & \boldsymbol{0}' & \rho_{21} & 1 & \rho_{21}^2 & -\gamma_1/(\sigma_1^2)^2 & 0 \\ \boldsymbol{0}' & \boldsymbol{0}' & 0 & 0 & 1 & 0 & -\gamma_2/(\sigma_2^2)^2 \\ \boldsymbol{0}' & \boldsymbol{0}' & 0 & 0 & 0 & 1/(\sigma_1^2) & 0 \\ \boldsymbol{0}' & \boldsymbol{0}' & 0 & 0 & 0 & 0 & 1/(\sigma_2^2) \end{vmatrix}$$

$$= \quad \sigma_1^2 \times (\sigma_1^2 \times \sigma_2^2)^{-1}.$$

In the same way, we can derive the Jacobian term for any dimension m.

Then, the transformed prior for Σ and $\boldsymbol{\gamma}$ is

$$\pi(\Sigma, \boldsymbol{\gamma}) \quad \propto \quad \prod_{j=1}^{m} \left[(\sigma_j^2)^{\frac{m-1}{2}+a_0-j-1}(\gamma_j)^{a_0-1} \exp\left\{-\frac{\gamma_j}{\sigma_j^2 b_0}\right\}\right].$$

The conditional posterior density of parameters of $\{\boldsymbol{b}, \Sigma\}$, given the values of $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_m)'$, $j = 1, ..., m$, is

$$\pi(\boldsymbol{b}_j|\boldsymbol{b}_{j-1}, ..., \boldsymbol{b}_1, \sigma_j^2, \boldsymbol{\lambda}, D) = N\left(\hat{\boldsymbol{b}}_j, \sigma_j^2(Z_j'Z_j + \gamma_j D_j)^{-1}\right), \tag{10}$$

$$\pi(\sigma_j^2|\boldsymbol{b}_{j-1}, ..., \boldsymbol{b}_1, \boldsymbol{\lambda}, D) = IG\left(\hat{a}_j/2, \hat{h}_j/2\right), \tag{11}$$

for $j = 1, ..., m$, and

$$\hat{\boldsymbol{b}}_j = \left(Z'_j Z_j + \gamma_j D_j \right)^{-1} \left[\left(Z'_j Z_j \right)^{-1} \tilde{\boldsymbol{b}} + \lambda_j D_j \boldsymbol{b}_0 \right],$$

$$\tilde{\boldsymbol{b}}_j = \left(Z'_j Z_j \right)^{-1} Z'_j \boldsymbol{y}_j,$$

$$\hat{a}_j = n - m - 2a_0 + 2j + 1,$$

$$\hat{h}_j = \gamma_j / b_0 + (n - p_j + j - 1)\hat{s}_j^2 + \left(\hat{\boldsymbol{b}}_j - \boldsymbol{b}_{0j} \right)' \left(Z'_j Z_j + \gamma_j D_j \right)^{-1} \left(\hat{\boldsymbol{b}}_j - \boldsymbol{b}_{0j} \right)$$

$$\hat{s}_j^2 = \left(\boldsymbol{y}_j - Z_j \tilde{\boldsymbol{b}}_j \right)' \left(\boldsymbol{y}_j - Z_j \tilde{\boldsymbol{b}}_j \right) / (n - p_j + j - 1).$$

Given a value of $\boldsymbol{\lambda}$, we can generate the conditional posterior samples of $\{\boldsymbol{b}^{(k)}, \Sigma^{(k)}\}$ for $k = 1, ..., N$ from $\pi(\boldsymbol{b}, \Sigma | \boldsymbol{y}, \boldsymbol{\lambda})$. The problem is, whatever the form of prior for $\boldsymbol{\lambda}$, to our knowledge, an analytical expression for the marginal posterior density of $\boldsymbol{\lambda}$ can not be obtained. In this paper, we shall employ an importance sampling procedure, which is a general technique for establishing the properties of a particular distribution, while only having samples generated from a different distribution rather than the distribution of interest.

From Bayes' theorem, we have

$$\pi(\boldsymbol{b}, \Sigma, \boldsymbol{\lambda} | \boldsymbol{y}) \propto L(\boldsymbol{y} | \boldsymbol{b}, \Sigma)\pi(\boldsymbol{b}, \Sigma, \boldsymbol{\lambda}).$$

If we generate the values $\boldsymbol{\lambda}^{(k)}$ for $k = 1, ..., N$ from a density $\pi^S(\boldsymbol{\lambda})$, and then generate the conditional posterior samples of $\{\boldsymbol{b}, \Sigma\}$ from $\pi(\boldsymbol{b}, \Sigma | \boldsymbol{\lambda}, \boldsymbol{y})$, the samples $\{\boldsymbol{b}^{(k)}, \Sigma^{(k)}, \boldsymbol{\lambda}^{(k)}\}$ obtained are generated from the following density: $\pi(\boldsymbol{b}, \Sigma | \boldsymbol{\lambda}, \boldsymbol{y})\pi^S(\boldsymbol{\lambda})$. As an importance sampling density function $\pi^S(\boldsymbol{\lambda})$, we can use the inverse gamma density. Also, we can use an inverse gamma prior density for the elements of $\boldsymbol{\lambda}$. In each of the densities, a gamma prior density with parameter 1 and 0.0001 was used. To sample $\{\boldsymbol{b}, \Sigma, \boldsymbol{\lambda}\}$ from the joint posterior density $\pi(\boldsymbol{b}, \Sigma, \boldsymbol{\lambda} | \boldsymbol{y})$, we shall use an importance sampling technique. Taking $\pi(\boldsymbol{b}, \Sigma | \boldsymbol{\lambda}, \boldsymbol{y})\pi^S(\boldsymbol{\lambda})$ as a proposal distribution, $\pi(\boldsymbol{b}, \Sigma, \boldsymbol{\lambda} | D)$ as the actual joint posterior distribution, the importance weight is given as

$$w_k = \frac{\pi(\boldsymbol{b}^{(k)}, \Sigma^{(k)}, \boldsymbol{\lambda}^{(k)} | \boldsymbol{y})}{\pi(\boldsymbol{b}^{(k)}, \Sigma^{(k)} | \boldsymbol{\lambda}^{(k)}, \boldsymbol{y})\pi^S(\boldsymbol{\lambda}^{(k)})}.$$

Therefore, we can generate joint posterior samples $\{\boldsymbol{b}^{(k)}, \Sigma^{(k)}, \boldsymbol{\lambda}^{(k)}\}$ directly. Note that the numerator for the weights is an unnormalized posterior, which is why we need to normalize the weights.

One issue with importance sampling is the behavior of the weights. It is most desirable that they be bounded, but they should at least have finite variances so the Central Limit Theorem and the delta-method can be used. It is difficult to show this theoretically. However, if it is true, one can see numerically that the average squared weights appear to settle down to a constant value. This is easier to check than looking for convergence in MCMC. We have checked this property for our algorithm using the

simulation setting in Section 5.1. As a result of 100 repetitions, we have found that the average squared weight appears to settle down to a constant value.

The marginal posterior density function $\pi(\boldsymbol{\lambda}|\boldsymbol{y})$ is approximated as

$$\pi(\boldsymbol{\lambda}|\boldsymbol{y}) \approx \sum_{k=1}^{N} \delta(\boldsymbol{\lambda} = \boldsymbol{\lambda}^{(k)}) \times v_k,$$

where $v_k = w_k/(\sum_{l=1}^{N} w_l)$ are the normalized importance weights, and $\delta(\boldsymbol{\lambda} = \boldsymbol{\lambda}_i)$ is the conditional indicator density function for $\boldsymbol{\lambda} = \boldsymbol{\lambda}_i$. Therefore, we can easily sample $\boldsymbol{\lambda}$ from the marginal posterior density. Given the posterior samples of the degrees of freedom parameter $\boldsymbol{\lambda}^{(k)}$ ($k = 1, ..., N$), we can also generate the conditional posterior samples of $\{\boldsymbol{b}^{(k)}, \Sigma^{(k)}\}$ from $\pi(\boldsymbol{b}, \Sigma|\boldsymbol{y}, \boldsymbol{\lambda}^{(k)})$ for $k = 1, ..., N$.

A Monte Carlo sampling procedure:

Step 1 (initialization). Fix the order of a set of m equations. Set the number of samples N to be generated. Generate $\boldsymbol{\lambda}^{(k)}$ from $\pi^{(S)}(\boldsymbol{\lambda})$, $k = 1, ..., N$. Set $j = 1$.

Step 2 Generate $\sigma_1^{2(k)}$ from (11), $k = 1, ..., N$, and insert the drawn values in $\pi(\boldsymbol{b}_1|\sigma_1^2, \boldsymbol{\lambda}, \boldsymbol{y})$. Then make a draw $\boldsymbol{b}_1^{(k)}$ from $\pi(\boldsymbol{b}_1|\sigma_1^{2(k)}, \boldsymbol{\lambda}, \boldsymbol{y})$ in (10), for $k = 1, ..., N$.

Step 3 Increase the iteration index j by one $j \rightarrow j + 1$. Draw $\sigma_j^{(k)}$ from the conditional inverse gamma density $\pi(\sigma_j^2|\boldsymbol{b}_{j-1}^{(k)}, ..., \boldsymbol{b}_1^{(k)}, \boldsymbol{\lambda}, \boldsymbol{y})$, and then generate $\boldsymbol{b}_j^{(k)}$ from $\pi(\boldsymbol{b}_j|\boldsymbol{b}_{j-1}^{(k)}, ..., \boldsymbol{b}_1^{(k)}, \sigma_j^{(k)}, \boldsymbol{\lambda}, \boldsymbol{y})$, for $k = 1, ..., N$.

Step 4 Repeat Step 3 sequentially until $j = m$.

Step 5 Sample from the generated samples $\{\boldsymbol{b}^{(k)}, \Sigma^{(k)}, \boldsymbol{\lambda}^{(k)}\}$ by using the importance sampling technique.

3.2 Some Remarks

We investigate whether the developed algorithm is invariant to the ordering of the set of equations. Consider the $m = 2$ equation SUR model and fix the ordering of equations as 1 and then 2. In this case, the transformed prior for $\boldsymbol{\beta}_2$ is

$$\pi(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1, \lambda_2) \propto \exp\left\{ -\frac{\lambda_2}{2}(\boldsymbol{\beta}_2 - \boldsymbol{\beta}_{02}^*)'W_2(\boldsymbol{\beta}_2 - \boldsymbol{\beta}_{02}^*) \right\},$$

with $W_2 = A_{22} - A_{12}A_{11}^{-1}A_{21}$. We can see that the matrix W_2 does not contain $\boldsymbol{\lambda}$, other than λ_2. Thus, under $m = 2$, the transformation of the original prior $\pi(\boldsymbol{\beta}|\boldsymbol{\lambda})$ in (7) to the transformed prior $\pi(\boldsymbol{\beta}_j|\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_{j-1}, \lambda_j)$ is order invariant.

Next we consider the case when $m = 3$ and fix the ordering of equations as 1, 2 and then 3. We already know that the equations 1 and 2 are order invariant. First, note that, for general matrices A, B, C, and D with $a \times a$, $a \times b$, $b \times a$, $b \times b$, we have the

following formula

$$\left(\begin{array}{cc} A & B \\ C & D \end{array} \right)^{-1} = \left(\begin{array}{cc} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ (D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{array} \right).$$

Using this formulae, the transformed prior for $\boldsymbol{\beta}_3$ is

$$\pi(\boldsymbol{\beta}_3|\boldsymbol{\beta}_2, \boldsymbol{\beta}_1, \lambda_3) \propto \exp\left\{ -\frac{\lambda_3}{2}(\boldsymbol{\beta}_3 - \boldsymbol{\beta}_{03}^*)'W_3(\boldsymbol{\beta}_3 - \boldsymbol{\beta}_{03}^*) \right\},$$

where each element of W_3 is a function of the pre-determined matrices A_{ij}, $i, j = 1, 2, 3$. This can be easily verified. Thus, we again see that the transformation of the original prior $\pi(\boldsymbol{\beta}|\boldsymbol{\lambda})$ in (7) to the transformed prior $\pi(\boldsymbol{\beta}_j|\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_{j-1}, \lambda_j)$ is order invariant. In the same way, we can proof the order invariance for any m.

Also, Zellner and Ando (2008a) investigated the relationship between Ω in the original SUR model and Σ in the transformed model (9). When we make an inference about Ω based on the generated samples $\{\Sigma^{(k)}, \boldsymbol{b}^{(k)}; k = 1, ..., N\}$ from a direct Monte Carlo sampling procedure, we can use the following recursive relations between Ω and Σ:

$$\omega_1^2 = \sigma_1^2,$$
$$\omega_j^2 = \sum_{k=1}^{j-1} \rho_{jk}^2 \omega_k^2 + \sum_{k,l=1, \ k<l}^{j-1} \rho_{jk}\rho_{jl}\omega_{lk} + \sigma_j^2, \quad (j \neq 1), \tag{12}$$
$$\omega_{ji} = \sum_{k=1, k\neq i}^{j-1} \rho_{jk}\omega_{ki} + \rho_{ji}\omega_i^2, \quad (j \neq 1).$$

There is a one to one mapping between the first p_j elements of $\boldsymbol{b}_j$ and $\boldsymbol{\beta}_j$. Also, we use $\lambda_j = \gamma_j/\sigma_j^2$. Using these equations, the posterior samples for the original hierarchical Bayesian SUR models can easily be obtained. Note that we transformed the original prior $\pi(\boldsymbol{\beta}, \Omega, \boldsymbol{\lambda})$ to $\pi(\boldsymbol{b}, \Sigma, \boldsymbol{\gamma})$ so that the transformation is invariant to the ordering of equations. After generating the posterior samples from the transformed model, which is order invariant, we just transform them back to the original parameter space.

Economic applications of SUR models frequently involve inequality restrictions on the coefficients. To express such restrictions, let us define a feasible region for the coefficients $\boldsymbol{b}_j$ by the inequality constraints, denoted by S_{b_j}, and define the indicator function

$$I_{S_{b_j}}(\boldsymbol{b}_j) = \left\{ \begin{array}{ll} 1, & (\boldsymbol{b}_j \in S_{b_j}) \\ 0, & (\boldsymbol{b}_j \notin S_{b_j}) \end{array} \right. .$$

Using the results of Zellner and Ando (2008c), the inequality restrictions can be incorporated in the above analysis.

When one wants to select the best model from among a set of candidate models, model selection criteria are available. Because we used an improper prior for the SUR model, the calculation of the marginal likelihood and posterior odds has some difficulties.

However, some model selection criteria recently have been developed, including deviance information criteria (DIC), Spiegelhalter et al., (2002), Bayesian predictive information criterion (BPIC, Ando (2007), and so on. The BPIC criterion is useful not only for selecting an optimal combination of the predictors, but this criterion is also applicable to select an optimal ordering of the equations when we employ order variant priors.

4 Hierarchical Bayesian Analysis of the Simultaneous Equations Model

The Simultaneous Equations Model (SEM), that incorporates instantaneous feedback relationships, was put forward many years ago and has been widely employed in econometric studies. In this section we provide a hierarchical Bayesian analysis of the SEM.

4.1 Overview of Simultaneous Equations Model

Consider the following m equation SEM:

$$Y\Gamma = XB + U, \tag{13}$$

where $Y = (\boldsymbol{y}_1, ..., \boldsymbol{y}_m)$ is an $n \times m$ matrix of observations on m endogenous variables, the $m \times m$ nonsingular matrix Γ is a matrix coefficient for the endogenous variables, $X = (\boldsymbol{x}_1, ..., \boldsymbol{x}_p)$ is an $n \times p$ matrix of observations on the p predetermined variables, the $p \times m$ matrix $B = (\boldsymbol{b}_1, ..., \boldsymbol{b}_m)$ is the coefficient matrix for the predetermined variables, and $U = (\boldsymbol{u}_1, ..., \boldsymbol{u}_m)$ is the $n \times m$ error matrix. It is known that some restrictions on the parameters are needed for model identification. In this paper, we assume that appropriate restrictions are imposed on the SEM structure to provide identification.

Multiplying both sides of (13) by Γ^{-1}, the unrestricted reduced form equations are

$$Y = X\Pi + V, \tag{14}$$

where $\Pi = B\Gamma^{-1} = (\boldsymbol{\pi}_1, ..., \boldsymbol{\pi}_m)$ is a $p \times m$ reduced form coefficient matrix, and $V = U\Gamma^{-1} = (\boldsymbol{v}_1, ..., \boldsymbol{v}_m)$ is the reduced form error matrix. In this paper, the n rows of V, $\boldsymbol{v}_i$ $(i = 1, ..., n)$, are assumed to be independently drawn from a multivariate normal distribution with zero mean vector and $m \times m$ positive definite covariance matrix Ω, $\boldsymbol{v}_i \sim N(\boldsymbol{0}, \Omega)$, where

$$\Omega = \begin{pmatrix} \sigma_1^2 & \boldsymbol{\omega}_1' \\ \boldsymbol{\omega}_1 & \Omega_1 \end{pmatrix}.$$

The problem is how to estimate the unknown parameters in the restricted model (13). In the next section, after reviewing the work of Zellner, Bauwens and van Dijk (1988) that developed a DMC algorithm for one equation of a SEM and Zellner and Ando (2008b) for m equations of a SEM, we develop hierarchical Bayesian analysis of simultaneous equations model.

4.2   Hierarchical Bayesian Simultaneous Equations Model

A single identified equation of a SEM is given in (15), and the reduced form equation for Y_1 is given in (16),

$$\boldsymbol{y}_1 = Y_1\boldsymbol{\gamma}_1 + X_1\boldsymbol{b}_1 + \boldsymbol{u}_1, \tag{15}$$

$$Y_1 = X\Pi_1 + V_1, \tag{16}$$

where $V_1 \sim N(\boldsymbol{0}, \Omega_1)$. Noting that the m-multivariate normal density of $(u_{1i}, \boldsymbol{v}'_{1i})'$ can be expressed as a conditional normal density of u_{1i} given a value of $\boldsymbol{v}_{1i}$ and a marginal multivariate normal density of $\boldsymbol{v}_{1i}$, Zellner, Bauwens and van Dijk (1988) derived $u_{1i}|\boldsymbol{v}_{1i} \sim N(\boldsymbol{v}'_{1i}\boldsymbol{\eta}_1, \sigma_1^2 - \boldsymbol{\omega}'_1\Omega_1^{-1}\boldsymbol{\omega}_1)$ with $\boldsymbol{\eta}_1 = \Omega_1^{-1}\boldsymbol{\omega}_1$ and $\boldsymbol{v}_{1i} \sim N(\boldsymbol{0}, \Omega_1)$. Then performing the transformation of random variables, $\boldsymbol{u}_1 = V_1\boldsymbol{\eta}_1 + \boldsymbol{\varepsilon}_1$ from $\boldsymbol{u}_1|V_1$ to $\boldsymbol{y}_1|Y_1$, one obtains

$$\boldsymbol{y}_1 = Y_1\boldsymbol{\gamma}_1 + X_1\boldsymbol{b}_1 + V_1\boldsymbol{\eta}_1 + \boldsymbol{\varepsilon}_1,$$

$$Y_1 = X\Pi_1 + V_1,$$

where $X = (X_1, X_0)$ and $(\varepsilon_{1i}, \boldsymbol{v}'_i)'$ $i = 1, ..., n$ are independent random drawings from a multivariate normal distribution with mean zero and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{\varepsilon_1}^2 & \boldsymbol{0}' \\ \boldsymbol{0} & \Omega_1 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 - \boldsymbol{\omega}'_1\Omega_1^{-1}\boldsymbol{\omega}_1 & \boldsymbol{0}' \\ \boldsymbol{0} & \Omega_1 \end{pmatrix}.$$

To develop m equations of the SEM, we next present the structural equation for $\boldsymbol{y}_2$ and the reduced form equation for Y_2,

$$\boldsymbol{y}_2 = Y_2\boldsymbol{\gamma}_2 + X_2\boldsymbol{b}_2 + \boldsymbol{u}_2, \tag{17}$$

$$Y_2 = X\Pi_2 + V_2,$$

where $V_2 \sim N(\boldsymbol{0}, \Omega_2)$. Again, noting that the $(m-1)$-multivariate normal density of $(u_{2i}, \boldsymbol{v}'_{2i})'$ can be expressed as a conditional normal density of u_{2i} given a value of $\boldsymbol{v}_{2i}$ and a marginal $(m-2)$-multivariate normal density of $\boldsymbol{v}_{2i}$, we have $u_{2i}|\boldsymbol{v}_{2i} \sim N(\boldsymbol{v}'_{2i}\boldsymbol{\eta}_2, \sigma_2^2 - \boldsymbol{\omega}'_2\Omega_2^{-1}\boldsymbol{\omega}_2)$ with $\boldsymbol{\eta}_2 = \Omega_2^{-1}\boldsymbol{\omega}_2$ and $\boldsymbol{v}_{2i} \sim N(\boldsymbol{0}, \Omega_2)$. Performing the transformation of random variables from $\boldsymbol{u}_2|V_2$ to $\boldsymbol{y}_2|Y_2$ and from V_2 to Y_2, we obtain;

$$\boldsymbol{y}_2 = Y_2\boldsymbol{\gamma}_2 + X_2\boldsymbol{b}_2 + V_2\boldsymbol{\eta}_2 + \boldsymbol{\varepsilon}_2,$$

$$Y_2 = X\Pi_2 + V_2,$$

where $(\varepsilon_{2i}, \boldsymbol{v}'_{2i})'$ $i = 1, ..., n$ are independent random drawings from a multivariate normal distribution with mean zero and covariance matrix

$$\begin{pmatrix} \sigma_{\varepsilon_2}^2 & \boldsymbol{0}' \\ \boldsymbol{0} & \Omega_2 \end{pmatrix} = \begin{pmatrix} \sigma_2^2 - \boldsymbol{\omega}'_2\Omega_2^{-1}\boldsymbol{\omega}_2 & \boldsymbol{0}' \\ \boldsymbol{0} & \Omega_2 \end{pmatrix}.$$

Then we have

$$\Sigma = \begin{pmatrix} \sigma_1^2 - \boldsymbol{\omega}'_1\Omega_1^{-1}\boldsymbol{\omega}_1 & \boldsymbol{0}' \\ \boldsymbol{0} & \Omega_1 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 - \boldsymbol{\omega}'_1\Omega_1^{-1}\boldsymbol{\omega}_1 & & \boldsymbol{0}' \\ \boldsymbol{0} & \sigma_2^2 - \boldsymbol{\omega}'_2\Omega_2^{-1}\boldsymbol{\omega}_2 & \boldsymbol{0}' \\ & \boldsymbol{0} & \Omega_2 \end{pmatrix}.$$

We can recursively express $\boldsymbol{y}_j$ $(j = 3, 4, ...m)$ in the same way. Then it can be easily shown that a general SEM can be transformed through linear operations to assume a recursive form, that is,

$$
\begin{aligned}
\boldsymbol{y}_1|Y_1 &= Y_1\boldsymbol{\gamma}_1 + X_1\boldsymbol{b}_1 + V_1\boldsymbol{\eta}_1 + \boldsymbol{\varepsilon}_1 \equiv Z_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1, \\
\boldsymbol{y}_2|Y_2 &= Y_2\boldsymbol{\gamma}_2 + X_2\boldsymbol{b}_2 + V_2\boldsymbol{\eta}_2 + \boldsymbol{\varepsilon}_2 \equiv Z_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2, \\
&\vdots \\
\boldsymbol{y}_{m-1}|Y_{m-1} &= Y_{m-1}\boldsymbol{\gamma}_{m-1} + X_{m-1}\boldsymbol{b}_{m-1} + V_{m-1}\boldsymbol{\eta}_{m-1} + \boldsymbol{\varepsilon}_{m-1} \\
&\equiv Z_{m-1}\boldsymbol{\beta}_{m-1} + \boldsymbol{\varepsilon}_{m-1}, \\
\boldsymbol{y}_m &= X_m\boldsymbol{b}_m + \boldsymbol{\varepsilon}_m \equiv Z_1\boldsymbol{\beta}_m + \boldsymbol{\varepsilon}_m,
\end{aligned} \tag{18}
$$

with the $\boldsymbol{\varepsilon}_j$'s uncorrelated and having differing variances:

$$
E[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j'] = \left\{ \begin{array}{ll} O, & (i \neq j) \\ \sigma_{\varepsilon_j}^2 I, & (i = j) \end{array} \right. , \quad \text{and} \quad \Sigma = \text{diag}\{\sigma_{\varepsilon_1}^2, ..., \sigma_{\varepsilon_m}^2\},
$$

with $\sigma_{\varepsilon_j}^2 = \sigma_j^2 - \boldsymbol{\omega}_j'\Omega_j^{-1}\boldsymbol{\omega}_j$. Each Y_j in the model is $Y_1 = (\boldsymbol{y}_2, ..., \boldsymbol{y}_m)$, $Y_2 = (\boldsymbol{y}_3, ..., \boldsymbol{y}_m)$,..., $Y_{m-1} = \boldsymbol{y}_m$, $Y_m = (0)$, $\boldsymbol{\gamma}_1 = (\gamma_{12}, ..., \gamma_{1m})'$ $\boldsymbol{\gamma}_2 = (\gamma_{23}, ..., \gamma_{2m})'$,..., $\boldsymbol{\gamma}_{m-1} = \gamma_{m-1,m}$, $\boldsymbol{\gamma}_m = (0)$. The matrices Z_j are $Z_1 = (Y_1, X_1, V_1)$, $(n \times \{p_1 + 2(m-1)\})$,..., $Z_m = X_m$, $(n \times p_m)$. The V_j's in the model are given by $V_1 = Y_1 - X\Pi_1$, $V_2 = Y_2 - X\Pi_2$, ,..., $V_{m-1} = Y_{m-1} - X\Pi_{m-1}$.

Since the transformed model (18) has the same form as that of the transformed SUR model (5), considered in Section 3, we can directly apply our previous algorithm to this model. For other work on transforming linear structural models to fully recursive forms, see Basmann (1965) and Spanos (1986, p.610 ff).

As pointed out by a referee, it is well known that the unrestricted SEM model is not identified and thus the posterior with a flat prior is also flat in regions of the parameter space. As a result the posterior is improper and a Gibbs sampler will get stuck in an absorbing state. The existence of an absorbing state would be a pathological occurrence in any situation and could occur whether there is a proper limiting distribution or not. Kleibergen and van Dijk (1998) have shown that for a three equation SEM there exists an issue of existence of moments. Kleibergen and van Dijk (1998) pointed out that a solution is to use informative priors. To avoid a setup that causes an absorbing state to occur, our informative prior approach thus might be useful from this perspective.

A referee also pointed out that even if the model is identified through zero restrictions on the parameters then there is an issue of existence of the posterior with a flat prior. In SEM context, one faces the issue of reduced rank (See Kleibergen and van Dijk (1998)) where singular values may exist and the transformation is not anymore one-to-one. We point out that if the considered SEM does not have a one-to-one mapping from the original model to the transformed model, our approach is not applicable.

5   Simulation results

In order to assess the performance of the proposed procedure, we present numerical results based on simulated data and a real data application. All calculations were performed on Microsoft Windows XP, Pentium-R 2.0 GHz, running R version 2.50. Each of the random draws from the probability density functions (normal, inverse-gamma, and inverse-Wishart) were generated by using the elementary functions incorporated in R. R is freely available and is a specialized programming language for statistical data analysis. Readers may obtain R software at http://cran.r-project.org/.

5.1   Simulation settings

We compare the properties of our Bayesian model estimation procedures and those of a MCMC procedure. We simulate data sets from an $m = 2$ dimensional SUR model. Without loss of generality in the model structure, we set $p_j = 2$, $j = 1, 2$ in model (1), which gives a simple two equation SUR model. This model can be written as follows:

$$\left(\begin{array}{c} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{array} \right) = \left(\begin{array}{cc} X_1 & O \\ O & X_2 \end{array} \right) \left(\begin{array}{c} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{array} \right) + \left(\begin{array}{c} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \end{array} \right),$$

for $i = 1, ..., n$, where $\boldsymbol{y}_j$ and $\boldsymbol{u}_j$ are $n \times 1$ vectors, X_j is an $n \times 2$ matrix and $\boldsymbol{\beta}_j$ is a 2-dimensional vector. Each element of Ω is set to be

$$\Omega = \left(\begin{array}{cc} \omega_1^2 & \omega_{12} \\ \omega_{21} & \omega_2^2 \end{array} \right) = \left(\begin{array}{cc} 0.1 & -0.05 \\ -0.05 & 0.2 \end{array} \right).$$

The covariate matrices X_j $j = 1, 2$ were generated from a uniform density over the interval $(-5, 5)$. The coefficient vector was set to be $\boldsymbol{\beta}_1 = (3, -2)'$ and $\boldsymbol{\beta}_2 = (2, 1)'$. This enabled the generation of simulated response observations. In this simulation we set the number of observations to be $n = 100$.

For the simulated data set, we calculated the posterior density using two methods, our proposed algorithm and an MCMC approach. We shall consider the informative prior specification for $\boldsymbol{\beta}$ in the original model and assume the independence of $\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_m$ in (7):

$$\pi(\boldsymbol{\beta}, \Omega, \boldsymbol{\lambda}) = \pi(\boldsymbol{\beta}|\Omega, \boldsymbol{\lambda})\pi(\Omega)\pi(\boldsymbol{\lambda}) = \left[\prod_{j=1}^{m} \pi(\boldsymbol{\beta}_j|\lambda_j) \right] \pi(\Omega) \left[\prod_{j=1}^{m} \pi(\lambda_j) \right],$$

and use a normal prior for $\boldsymbol{\beta}_j$, $N(\boldsymbol{\beta}_{0j}, \lambda_j A_{\beta_j})$, $j = 1, ..., m$.

$$\pi(\boldsymbol{\beta}_j|\lambda_j) \quad \propto \quad \exp \left\{ -\frac{\lambda_j}{2}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_{0j})' A_j (\boldsymbol{\beta}_j - \boldsymbol{\beta}_{0j}) \right\}.$$

In the MCMC approach, a combination of the Metropolis-Hastings and Gibbs sampling algorithms is used. The details of the MCMC algorithm are as follows:

Step 1 Initialize $\boldsymbol{\beta}$, Ω and $\boldsymbol{\lambda}$.

Step 2 Sample the coefficient parameter from $\boldsymbol{\beta}|\Omega, \boldsymbol{\lambda}, D$.

Step 3 Sample the covariance matrix parameter from $\Omega|\boldsymbol{\beta}, D$.

Step 4 Sample the variance parameter from $\boldsymbol{\lambda}|\boldsymbol{\beta}, \Omega, D$.

Step 5 Repeat Step 2, 3 and Step 4 for a sufficiently long time.

In Steps 2 and 3, Gibbs sampling algorithms are employed. To sample $\boldsymbol{\lambda}$, at the k-th iteration, we generate a candidate $\boldsymbol{\lambda}^{(k)}$ from the proposal density $g(\boldsymbol{\lambda})$. We sampled all the $\boldsymbol{\lambda}$ vector at once using a multivariate proposal distribution. The proposed candidate is accepted with the probability α given by

$$\alpha = \min\left\{1, \frac{h(\boldsymbol{\beta}^{(k)}, \Omega^{(k)}, \boldsymbol{\lambda}^{(k)})/g(\boldsymbol{\lambda}^{(k)})}{h(\boldsymbol{\beta}^{(k)}, \Omega^{(k)}, \boldsymbol{\lambda}^{(k-1)})/g(\boldsymbol{\lambda}^{(k-1)})}\right\},$$

with

$$h(\boldsymbol{\beta}, \Omega, \boldsymbol{\lambda}) = L(\boldsymbol{y}|\boldsymbol{\beta}, \Omega) \times \prod_{j=1}^{m} \pi(\boldsymbol{\beta}_j|\lambda_j)\pi(\lambda_j).$$

The proposal density of $\boldsymbol{\lambda}$ is specified as $g(\boldsymbol{\lambda}) = \prod_{j=1}^{m} g(\lambda_j)$, where $g(\lambda_j)$ is an inverse gamma with parameters 5 and 0.005 so that an acceptance rate is around 30%. This value was determined taking account of the autocorrelation in the output of the MCMC procedure.

To save computational time, the initial values of the parameters $\boldsymbol{\beta}$ and Ω are chosen to be generalized least squares estimates. The initial values of λ_j were generated from the above mentioned proposal distribution.

In our application, we generated 10,000 posterior samples using our approach. The total number of MCMC iterations is chosen to be 11,000, of which the first 1,000 iterations are discarded. It is necessary to check whether generated posterior samples are taken from a stationary distribution. We assessed the convergence of MCMC simulation by calculating the convergence diagnostic (CD) test statistics (Geweke (1992)). Geweke's (1992) CD test statistic is useful to test the equality of the means of the first and latter part of a Markov chain. The CD test statistic has an asymptotically standard normal distribution. To check the convergence, one might use the result of Heyde and Johnstone (1979). However, we here used Geweke's (1992) CD test. All the results we report in this paper are based on samples that have passed the Geweke's (1992) convergence test at the "conventional" significance level of 5% for all parameters. We along with others did not check the power of this test. Also, there was no evidence of lack of convergence based on an examination of trace plots.

5.2 Results

Table 1 reports posterior means, modes and standard deviations as well as credible intervals for the model's parameters. The effective sample size and the convergence di-

agnostic (CD) test statistics (Geweke (1992)) of the MCMC algorithm are also reported. Using the posterior draws for each of the parameters, we calculated these statistics. The 95% credible intervals are computed using the 2.5th and 97.5th percentiles of the posterior samples.

DMC-IS approach

	TV	Mean	Mode	SDs	95%CIs	
β_{11}	3.00	2.9869	2.9833	0.0507	2.8827	3.0839
β_{12}	-2.00	-2.0318	-2.0333	0.0516	-2.1369	-1.9332
β_{21}	2.00	2.0063	2.0233	0.0749	1.8767	2.1684
β_{22}	1.00	1.0121	1.0276	0.0668	0.8951	1.1622
ω_1^2	0.10	0.1073	0.1101	0.0164	0.0824	0.1467
ω_{12}	-0.05	-0.0566	-0.0618	0.0181	-0.1009	-0.0297
ω_2^2	0.20	0.2206	0.2282	0.0352	0.1684	0.3048
$\lambda_1 \times 10^5$	–	43.7947	109.024	98.6957*	11.6144	371.9998
$\lambda_2 \times 10^5$	–	67.7915	146.6019	120.4277*	12.8755	457.2539

MCMC algorithm

	TV	Mean	Mode	SDs	95%CIs		CD	INEFs
β_{11}	3.00	2.9873	2.9819	0.0494	2.8849	3.0792	0.2584	2.3444
β_{12}	-2.00	-2.0353	-2.0307	0.0502	-2.1281	-1.9309	1.3290	2.8372
β_{21}	2.00	2.0266	2.0235	0.0707	1.8815	2.1613	-0.7933	2.0874
β_{22}	1.00	1.0108	1.0284	0.0633	0.9037	1.1531	-0.1509	2.9153
ω_1^2	0.10	0.1049	0.1055	0.0155	0.0793	0.1397	-0.7965	2.8420
ω_{12}	-0.05	-0.0554	-0.0592	0.0167	-0.0957	-0.0295	-0.4144	3.1602
ω_2^2	0.20	0.2059	0.2068	0.0301	0.1565	0.2740	1.7044	3.9600
$\lambda_1 \times 10^5$	–	11.3709	13.6154	4.7979	7.4735	25.9415	-0.0714	12.4020
$\lambda_2 \times 10^5$	–	10.1274	13.5254	4.8362	6.7511	25.1868	-0.5841	18.5921

Table 1: Simulated data: Summary of the parameter estimates for the proposed algorithm and for the MCMC algorithm, including the posterior means, the posterior modes, the posterior standard deviations (SDs), and 95% credible intervals (95%CIs). For MCMC results, the inefficiency factors (INEFs; the integrated autocorrelation time), and Geweke's (1992) convergence diagnostic test statistic (CD) are also calculated. Our algorithm combines a direct Monte Carlo approach and an importance sampling procedure (DMC-IS approach), as explained in the text. The true values (TV) of the parameters are also provided. * Note the differences in these moments. Magnitudes of SDs/Mean from MCMC is smaller than those from DMC-IS approach. Based on the common use of MCMC output, these statistics are calculated based on an i.i.d assumption. Since the MCMC samples are autocorrelated (as shown in 1 (d)), these statistics should be calculated taking account of the autocorrelation. To take account of this matter, some use just each seventh draw of the MCMC output in their calculations.
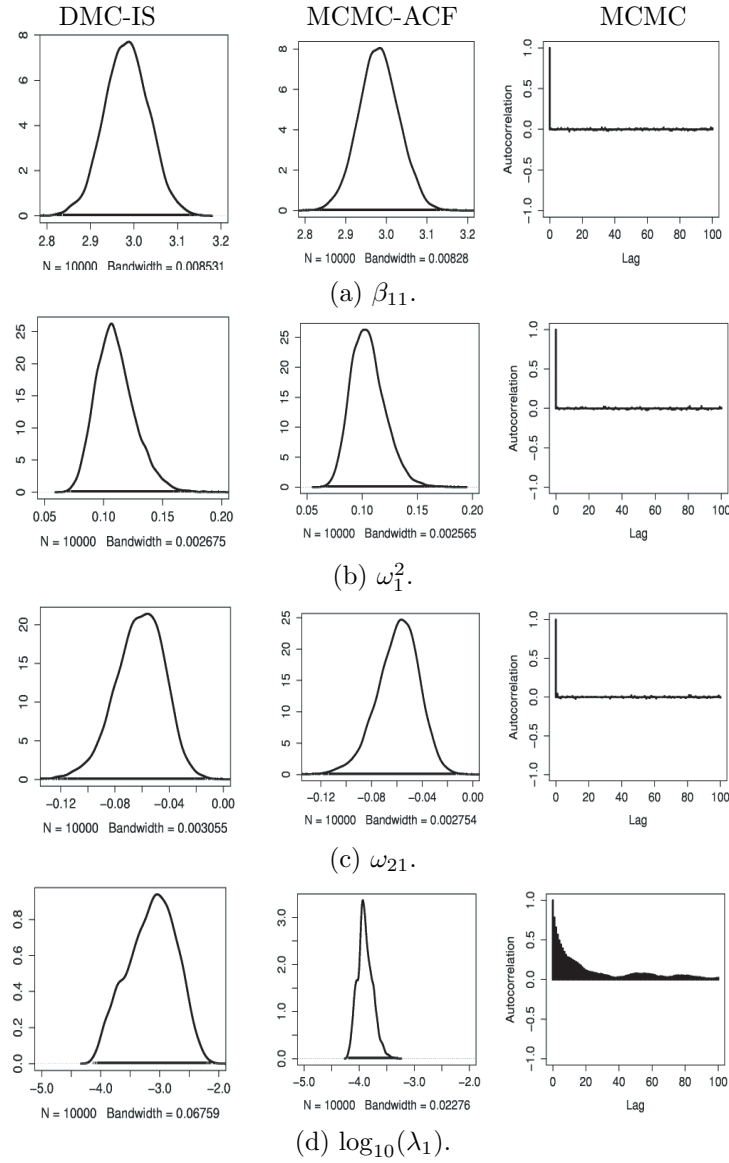
Figure 1: Estimated posterior densities for data simulated from the models in Section 5.1. (a) β_{11}, (b) ω_1^2, (c) ω_{21}, (d) $\log_{10}(\lambda_1)$. The results are for the proposed algorithm that combines a direct Monte Carlo approach and an importance sampling procedure (DMC-IS) (Left), the MCMC algorithm (Center), and the autocorrelation (ACF) plot for each of the parameters estimated by MCMC algorithm (Right).

Estimated predictive density for y_1 based on the DMC-IS method.

Estimated predictive density for y_1 based on the MCMC method.

Estimated predictive density for y_2 based on the DMC-IS method.

Estimated predictive density for y_2 based on the MCMC method.
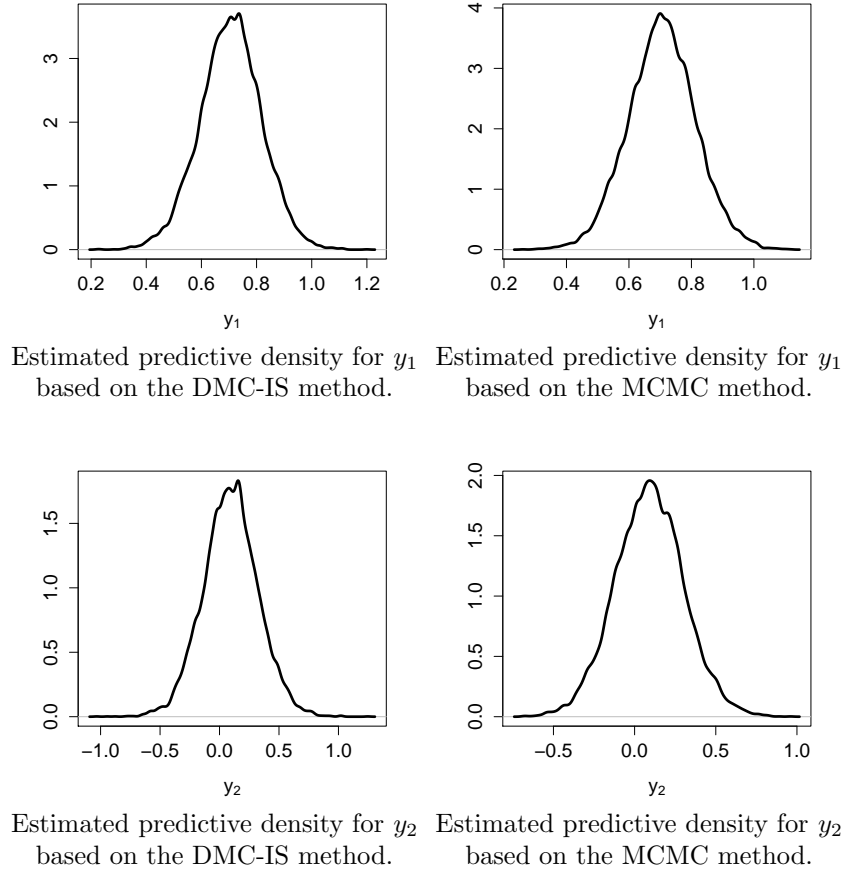
Figure 2: Simulated data: Estimated predictive density based on the proposed method and MCMC method. The proposed algorithm that combines a direct Monte Carlo approach and an importance sampling procedure (DMC-IS). The density is evaluated at the points $\boldsymbol{x}_1 = (0.1, -0.2)'$ and $\boldsymbol{x}_2 = (0.2, -0.3)'$.

A concept related to MCMC convergence is the inefficiency factor that is useful to measure the efficiency of the MCMC sampling algorithm. A large value of inefficiency factor indicates that we need a large number of MCMC simulations. The effective sample size, the number of MCMC output L divided by the inefficiency factor, is useful to measure the efficiency of the MCMC sampling algorithm.

As shown in Table 1, the calculated inefficiency factors for λ_1 and λ_2 exceeds 10 in value. This implies that the generated MCMC output available for analyses of posterior distributions is less than 10% of the total MCMC generated output.

It can be seen that the estimation results of our proposed method for the parameters appear quite reasonable. For instance, the true model is estimated with reasonably accurate results based on our proposed method. The 95% credible intervals include the true parameter values. Figure 1 shows the estimated posterior densities for each of the model parameters. Figure 1 also provides plots of the autocorrelation functions for the parameter draws obtained from the MCMC algorithms. The autocorrelation functions were calculated using 10,000 samples. We can clearly see that the samples from the Gibbs sampling algorithm are autocorrelated, while those from our proposed approach are theoretically and practically zero. This is one of the clear advantages of our approach.

Figure 2 shows the estimated predictive density based on our method. By using the posterior samples $\{\boldsymbol{\beta}^{(k)}, \Omega^{(k)}, \boldsymbol{\lambda}^{(k)}; k = 1, ..., N\}$, given a value of $\boldsymbol{x}$, the predictive density can be approximated as

$$\int f(y|\boldsymbol{x}, \boldsymbol{\beta}, \Omega) g(\boldsymbol{\beta}, \Omega, \boldsymbol{\lambda}|D) d\boldsymbol{\beta} d\Omega d\boldsymbol{\lambda} \approx \frac{1}{N} \sum_{k=1}^{N} f(y|\boldsymbol{x}, \boldsymbol{\beta}^{(k)}, \Omega^{(k)}).$$

The density is evaluated at the point $\boldsymbol{x}_1 = (0.1, -0.4)'$ and $\boldsymbol{x}_2 = (0.2, -0.3)'$. Because the actual predictive density is not known and thus we have no bench-mark against which to compare it, we compared the estimated predictive density with the true sampling density of $\boldsymbol{y} = (y_1, y_2)'$ given $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. We found that the estimated predictive density is very close to the true density.

6   Real data analysis

In this section, we apply our method to real data. Here we use the same prior settings as used in the previous section.
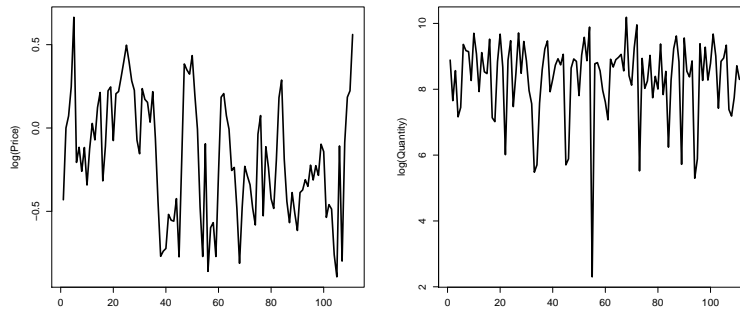
6.1   Fulton fish market data

As our application, we report the results of an analysis of the demand for fish. We apply our method to obtain an estimate of the price elasticity of demand. Observations on price and quantity of fresh whiting sold in the Fulton fish market (Graddy (1995), Chernozhukov and Hansen (2008)) over the five month period from December 2, 1991 to May 8, 1992 are used. The aggregated data on price and quantity are collected each day. The price is measured as the average daily price and the quantity is measured by the total amount of fish sold each day. The number of observations, namely, the number of days the market was open over the sample period, is $n = 111$. Figure 3 provides a plot of the data.

Following Chernozhukov and Hansen (2008), we consider the following demand and

supply model:

$$\log Q_t = b_{q0} + b_{q1} \log P_t + u_{1t},$$

$$\log P_t = b_{p0} + b_{p1} W_{1t} + b_{p2} W_{2t} + v_{2t}, \tag{19}$$

where W_{1t} and W_{2t} are two different instrumental variables that capture weather conditions at sea. W_{1t} is a dummy variable, *Stormy*, which indicates wave height greater than 4.5 ft and wind speed greater than 18 knots, and W_{2t} is also a dummy variable, *Mixed*, indicating wave height greater than 3.8 ft and wind speed greater than 13 knots. See the original paper for further discussion of the equations in (19).

(a) log(Daily Price)   (b) log(Daily Quantity)
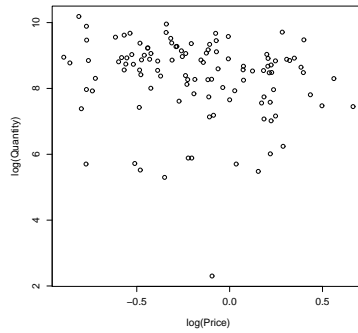
(c) log(Daily Price) and log(Daily Quantity)

Figure 3: Demand for fish data. The data contain observations on (a) price and (b) quantity of fresh whiting sold in the Fulton fish market in New York over the five month period from December 2, 1991 to May 8, 1992. The price is measured as the average daily price and the quantity as the total amount of fish sold that day. In total, the sample consists of 111 observations for the days in which the market was open over the sample period. Figure (c) shows the relationship between the price and quantity.

	Mean	Mode	SDs	95%CIs	
b_{p0}	8.5527	8.612	0.2824	7.9790	9.0720
b_{p1}	-0.5302	-0.5251	0.4186	-1.5486	0.2455
b_{p2}	-0.3974	-0.3311	0.3082	-1.2885	0.1181
b_{q1}	-0.7969	-0.8881	1.1236	-3.2708	1.7682
b_{q0}	6.7523	7.3118	9.3794	-14.4819	26.2981
ω_{11}^2	1.5112	1.5368	0.2147	1.1438	1.9131
ω_{12}	1.3926	1.2765	1.6754	-2.6440	5.0220
ω_{22}^2	1.1532	3.0544	5.7868	0.1990	16.2084
$\lambda_1 \times 10^5$	37.2639	85.3016	133.143	21.2622	298.7509
$\lambda_2 \times 10^5$	39.4343	89.4008	114.4483	22.2346	327.3828

Table 2: Real data application 1: Analysis of fish market data. Summary of the parameter estimates from our algorithm. The posterior means, modes, standard deviations, and 95% credible intervals (95%CIs) are calculated. Following Chernozhukov and Hansen (2008), we consider the demand and supply model: $\log Q_t = b_{q0} + b_{q1} \log P_t + u_{1t}$, $\log P_t = b_{p0} + b_{p1}W_{1t} + b_{p2}W_{2t} + v_{2t}$, where W_{1t} and W_{2t} are two different instrumental variables that capture weather conditions at sea. W_{1t} is a dummy variable, *Stormy*, which indicates wave height greater than 4.5 ft and wind speed greater than 18 knots, and W_{2t} is also a dummy variable, *Mixed*, indicating wave height greater than 3.8 ft and wind speed greater than 13 knots.

	DMC-IS[a]	DMC[b]	OLS	2SLS	MELO	BMOM
b_{q1}	-0.7969	-0.6930	-0.0484	-0.8750	-0.4272	-0.4143
b_{q0}	6.7523	5.5832	0.2273	7.1062	3.3786	3.2712

Table 2 reports the posterior means, posterior standard deviations, and 95% posterior intervals using our algorithm. Based on 10,000 draws for each of the parameters, we calculated the posterior means, the standard deviations, and 95% posterior intervals. The 95% posterior intervals are estimated using the 2.5th and 97.5th percentiles of the drawn posterior samples. As we expected, price and quantity are inversely related. Although the 95% posterior intervals contain positive values of b_{q1}, the posterior mean of b_{q1} is negative.

Table 3: Real data application 1: Analysis of fish market data. Summary of the parameter estimates for various algorithms. Point estimates from each method are reported. DMC-IS; our algorithm combines a direct Monte Carlo approach and an importance sampling procedure, as explained in the text. DMC; direct Monte Carlo in Zellner and Ando (2008b), OLS; ordinary least squares, 2SLS; two-stage least squares, BMOM; Bayesian method of moments, and MELO; minimum expected loss estimate. The number of draws from our DMC is 10,000. The results, except for our method, are from Zellner and Ando (2008b). See Zellner (1994) for the definitions of BMOM and MELO. a, b: Posterior means are used.

We also compared the estimation results with those produced by several other methods, namely, DMC, OLS, 2SLS, BMOM and MELO (See Zellner (1994)). Estimation results are reported in Table 3. OLS estimates have very different values compared to those provided by alternative methods. Finally, we would like to point out that our algorithm can be used for the residual analysis. Using the posterior density for the realized error terms, we can check the distributional assumptions made in the analysis of the equation in (19).

6.2   Point of Sales data analysis of brand in a soy source category.

In this section, we apply our method to an analysis of POS data for 3 soy source brands. Daily sales data were collected at stores of a Japanese supermarket chain, that contain information about the daily price P, the daily number of visitors to the store V, the execution/non-execution of display promotion D, $D = 1$ execution $D = 0$ otherwise, the execution/non-execution of advertisement A, $A = 1$ execution $A = 0$ otherwise and holiday indicator H, $H = 1$ holiday $H = 0$ otherwise. We use $n = 986$ data for the period of 1999 to 2002.

To this data, we fit the $m = 3$ equation SEM:

$$y_{j,t} = \beta_{j0} + \beta_{j1}V_t + \beta_{j2}H_t + \sum_{j<k}^{3} \gamma_k y_{k,t} + \sum_{k=1}^{3} \beta_{jk}^P P_{j,t} + \sum_{k=1}^{3} \beta_{jk}^D D_{j,t} + \sum_{k=1,3} \beta_{jk}^A A_{j,t} + u_{jt},$$

$j = 1, ..., 3$, where the predictors are $y_{j,t}$: log-daily sales volume, V_t: the daily number of visitors to the store at time t, H_t: the holiday indicator, $P_{j,t}$: the daily log-price for brand j at time t, $D_{j,t}$: the daily display promotion activity for brand j at time t, and A_t: the daily advertisement activity for brand j at time t, respectively. Note that since there is no advertising activity for brand $j = 2$, we can not estimate β_{j2}^A for $j = 1, 2, 3$.

Table 4 reports the coefficient estimation results that contain much useful information. First, as we expected, the posterior means of price elasticity for each brand β_{jj}^P are negative. However, the 95% posterior interval for brand 2 contains 0. It indicates that the price cut might not be effective for brand 2. In fact, brand 2 is an everyday low price brand. Therefore, consumers may not respond to the price cuts because the price is already very low. Second, the estimated cross-price elasticities β_{jk}^P ($j \neq k$) exhibit

usual results, namely, these brands are competitive. Third, the display promotion elasticity for brand $j = 3$, β_{33}^D, indicates that the display promotion activity increases sales. Also, the display promotion activity for brand $j = 3$ affects the sales of brand $j = 1$ rather than that of brand $j = 2$. Therefore, brands 1 and 3 are competitive. It also indicates that the brand 2 is relatively free from competition. Fourth, advertisement activity for brand 2 and 3 boosts sales, while it does not for sales of brand 1, and that may encourage the store manager to consider other ways of promoting brand 1 sales.

	Mean	Mode	SDs	95%CIs	
Brand 1: β_{10}	-1.334	-0.759	0.985	-1.335	1.774
β_{11}	0.257	0.274	0.038	0.235	0.360
β_{11}^P	-2.706	-2.809	0.183	-3.206	-2.706
β_{12}^P	3.816	3.853	0.179	3.489	4.514
β_{13}^P	-0.181	-0.208	0.155	-0.642	0.052
β_{11}^D	1.094	1.067	0.057	0.878	1.099
β_{12}^D	0.015	0.000	0.031	-0.100	0.015
β_{13}^D	-0.260	-0.237	0.055	-0.298	-0.070
β_{11}^A	1.255	1.188	0.13	0.789	1.255
β_{13}^A	-0.060	-0.075	0.059	-0.211	0.058
Brand 2: β_{20}	-0.397	-0.256	0.298	-0.512	0.744
β_{21}	0.361	0.363	0.023	0.320	0.456
β_{21}^P	0.317	0.294	0.121	-0.044	0.431
β_{22}^P	-0.039	-0.022	0.152	-0.246	0.423
β_{23}^P	0.220	0.211	0.029	0.158	0.257
β_{21}^D	-0.015	-0.001	0.049	-0.032	0.130
β_{22}^D	0.268	0.265	0.017	0.217	0.281
β_{23}^D	0.025	0.02	0.02	-0.048	0.031
β_{21}^A	0.044	0.027	0.076	-0.168	0.167
β_{23}^A	0.214	0.184	0.058	0.052	0.217
Brand 3: β_{30}	0.585	0.603	0.313	-0.234	1.258
β_{31}	0.248	0.170	0.318	-0.461	0.881
β_{31}^P	1.752	1.680	0.498	0.283	2.564
β_{32}^P	1.453	1.605	0.605	0.577	3.240
β_{33}^P	-2.823	-2.961	0.312	-3.760	-2.709
β_{31}^D	-0.852	-0.828	0.143	-1.044	-0.529
β_{32}^D	-0.013	-0.033	0.185	-0.406	0.455
β_{33}^D	0.924	0.892	0.067	0.734	0.924
β_{31}^A	-0.967	-0.887	0.215	-1.281	-0.347
β_{33}^A	0.902	0.879	0.136	0.615	1.243

Table 4: Real data analysis 2: Summary of the parameter estimates for the proposed algorithm. The posterior means, posterior modes, standard deviations (SDs), and 95% credible intervals (95%CIs) are presented below. This is a summary table for the coefficients β and estimation results for other parameters are reported in Table 5. We fit the $m = 3$ equation SEM: $y_{j,t} = \beta_{j0} + \beta_{j1}V_t + \beta_{j2}H_t + \sum_{j<k}^{3} \gamma_k y_{k,t} + \sum_{k=1}^{3} \beta_{jk}^{P} P_{j,t} + \sum_{k=1}^{3} \beta_{jk}^{D} D_{j,t} + \sum_{k=1,3} \beta_{jk}^{A} A_{j,t} + u_{jt}$, $j = 1, ..., 3$, where $y_{j,t}$: log-daily sales volume, V_t: the daily number of visitors to the store at time t, H_t: the holiday indicator, $P_{j,t}$: the daily log-price for brand j at time t, $D_{j,t}$: the daily display promotion activity for brand j at time t, and A_t: the daily advertisement activity for brand j at time t, respectively. Note that since there is no advertisement activity for brand $j = 2$, we can not estimate β_{j2}^{A} for $j = 1, 2, 3$.

	Mean	Mode	SDs	95%CIs	
ω_1^2	0.720	0.717	0.012	0.686	0.733
ω_2^2	0.189	0.194	0.012	0.189	0.231
ω_3^2	0.838	1.006	0.460	0.838	1.781
ω_{12} (ω_{21})	0.062	0.078	0.035	0.062	0.185
ω_{13} (ω_{31})	-0.332	-0.338	0.190	-0.788	0.285
ω_{32} (ω_{23})	0.033	-0.002	0.179	-0.392	0.427
γ_{12}	-0.015	-0.032	0.045	-0.171	-0.015
γ_{13}	0.652	0.650	0.131	0.308	0.907
γ_{23}	-0.248	-0.065	0.821	-2.041	1.668
λ_1	1.092	1.121	0.071	1.049	1.338
λ_2	7.988	7.330	1.101	4.605	7.989
λ_3	1.294	1.280	0.107	1.015	1.507

Table 5: Real data analysis 2: Summary of the parameter estimates for the proposed algorithm. The posterior means, posterior modes, standard deviations (SDs), and 95% credible intervals (95%CIs) are presented below. The estimation results for the coefficient parameters are reported in Table 4 We fit the $m = 3$ equation SEM: $y_{j,t} = \beta_{j0} + \beta_{j1}V_t + \beta_{j2}H_t + \sum_{j<k}^{3} \gamma_k y_{k,t} + \sum_{k=1}^{3} \beta_{jk}^{P} P_{j,t} + \sum_{k=1}^{3} \beta_{jk}^{D} D_{j,t} + \sum_{k=1,3} \beta_{jk}^{A} A_{j,t} + u_{jt}$, $j = 1, ..., 3$, where $y_{j,t}$: log-daily sales volume, V_t: the daily number of visitors to the store at time t, H_t: the holiday indicator, $P_{j,t}$: the daily log-price for brand j at time t, $D_{j,t}$: the daily display promotion activity for brand j at time t, and A_t: the daily advertisement activity for brand j at time t, respectively. Note that since there is no advertisement activity for brand $j = 2$, we can not estimate β_{j2}^{A} for $j = 1, 2, 3$.

Table 5 reports the remaining parameter estimation results. From the variance estimation results, we can see that the sales of brand 2 are more stable than those of others. About the parameter estimates for the endogenous variables, we see interesting results. The 95% posterior interval for γ_{13} is positive, while that for γ_{12} is negative. It

indicates that when sales of brand $j = 2$ are large, the sales of other brands are small. However, the posterior mean value of γ_2 is very small. From this, we can conclude that the positioning of brand 2 is different from those of brands 1 and 3.

7   Discussion and Conclusions

We developed a computationally efficient method for applying Bayesian inference techniques in analyses of SUR models and SEM with hierarchical priors. In particular, we developed a new algorithm by combining direct Monte Carlo (DMC) and importance sampling approaches to compute various quantities of interest.

There are many directions for further research. First, we assumed a standard linear relationship in each equation of the SUR model and the SEM. It is clear that our approach can be applied to more complex variants of the models, for example those involving use of regression splines, B-splines, kernel bases, wavelet bases, and so on. It is straightforward to apply our method in Bayesian analyses of these variants of our models.

The developed algorithm can be applied to various statistical models. For example, we can apply our algorithm in Bayesian analyses of state space models. The state space model consists of two stochastic components: an observation equation and a system equation:

$$\left\{ \begin{array}{ll} \text{Observation equation}: & \boldsymbol{y}_t = Z_t\boldsymbol{\beta}_t + \boldsymbol{u}_t \quad \boldsymbol{u}_t \sim N(\boldsymbol{0}, \Omega), \\[2mm] \text{System equation}: & \boldsymbol{\beta}_{t+1} = T_t\boldsymbol{\beta}_t + \boldsymbol{\eta}_t \quad \boldsymbol{\eta}_t \sim N(\boldsymbol{0}, Q), \end{array} \right.$$

with $\boldsymbol{\beta}_1 \sim N(\boldsymbol{\beta}_{01}, Q_0)$. Here $\boldsymbol{y}_t = (y_{1t}, ..., y_{mt})'$ is the p-dimensional vector, $\boldsymbol{\beta}_t = (h_{1t}, ..., h_{qt})'$ is the q-dimensional vector, The main focus concerns how to construct these two equations so that the model captures the true structure governing the time series of $\boldsymbol{y}_t$. See Kitagawa (1987), Putnam and Quintana (1994), Kitagawa and Gersch (1996), Quintana and Putnam (1996), West and Harrison (1997), Durbin and Koopman (2001) for further discussion and applications of state space models.

The model can be represented in a general matrix form:

$$\begin{aligned} \boldsymbol{y}_j &= X_j\boldsymbol{\beta}_j + \boldsymbol{u}_j, \ \boldsymbol{u}_j \sim N(\boldsymbol{0}, \omega_j^2), \quad E[\boldsymbol{u}_i\boldsymbol{u}_j'] = \left\{ \begin{array}{ll} \omega_{ij}I, & (i \neq j) \\ \omega_i^2 I, & (i = j) \end{array} \right. \\ \boldsymbol{\beta} &= T(\boldsymbol{\beta}_0 + R\boldsymbol{\eta}) = T\boldsymbol{\beta}_0 + T\boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim N(\boldsymbol{0}, Q), \end{aligned}$$

with $\boldsymbol{y}_j = (y_{j1}', ..., y_{jn}')'$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', ..., \boldsymbol{\beta}_n')'$, $\boldsymbol{\eta} = (\boldsymbol{\eta}_1', ..., \boldsymbol{\eta}_n')'$, $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1', ..., \boldsymbol{\varepsilon}_n')'$. $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_{01}', \boldsymbol{0}', ..., \boldsymbol{0}')'$, and

$$T = \begin{pmatrix} I_m & O & O & O & \cdots & O \\ T_1 & I_m & O & O & \cdots & O \\ T_2T_1 & T_2 & I_m & O & \cdots & O \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ T_n\cdots T_1 & T_n\cdots T_2 & T_n\cdots T_3 & \cdots & I_m \end{pmatrix}.$$

As we can see, the transformed model has the same form as that of the SUR model with an informative prior. Therefore it can be transformed into the transformed SUR model, considered in Section 3, and thus we can directly apply our algorithm to analyze it. Jungbacker and Koopman (2005) proposed an importance sampling procedure for state space models. A natural extension of our method is to the case where a Kalman filter is used.

Also, there is an interesting connection to the issue of weak instruments, which is very important in econometrics. Hoogerheide et al. (2007) showed how one may be close to singularity (reduced rank possibility of the parameter matrix) when the instruments are very weak. Then one faces the difficulty in finding a good candidate or importance distribution. How to deal with this topic is a very interesting problem that we may consider in future research.

Also, Conley et al (2008) used a more general error structure for Bayesian instrumental variable analysis. Although further investigation is needed, one may incorporate the DMC approach into their framework. We plan to study this model further and to analyze it in a future paper.

References

Aliprantis, C.D., Barnett, W.A., Cornet, B. and Durlauf, S. 2007. The interface between econometrics and economic theory. *Journal of Econometrics* 136: 325–724.

Ando, T. 2007. Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. *Biometrika* 94: 443–458.

Basmann, R. 1965. A note on the statistical testability of "explicit casual chains" against the class of "interdependent" models. *Journal of the American Statistical Association* 60: 1080–1093.

Box, G. E. P. and Tiao, G. C. 1973. *Bayesian Inference in Statistical Analysis.* Cambridge, MA: Addison-Wesley.

Brooks, S. P. and Gelman, A. 1997. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7: 434–455.

Chernozhukov, V. and Hansen, C. 2008. Instrumental variable quantile regression: A robust inference approach, *Journal of Econometrics* 142: 379–398.

Chib, S. and Greenberg, E. 1995. Hierarchical Analysis of SUR Models with Extensions to Correlated Serial Errors and Time-Varying Parameter Models. *Journal of Econometrics* 68: 339–360.

Conley, T., Hansen, C., McCulloch, R. and Rossi, P. 2008. A semi-parametric Bayesian approach to the instrumental variable problem, *Journal of Econometrics* 144: 276–305.

Durbin, J. and Koopman, S. 2001. *Time series analysis by state space methods.* Oxford: Oxford University Press.

Fraser, D. A. S., Rekkasb, M. and Wong, A. 2005. Highly accurate likelihood analysis for the seemingly unrelated regression problem. *Journal of Econometrics* 127: 17–33.

Gelman, A. and Rubin, D. B. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7: 457–511.

Geweke, J. 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: *Bayesian Statistics*, vol. 4, 169–193. Oxford: Clarendon Press.

Geweke, J. 2005. *Contemporary Bayesian Econometrics and Statistics*. New York: Wiley.

Graddy, K. 1995. Testing for Imperfect Competition at the Fulton Fish Market. *RAND Journal of Economics* 26: 75–92.

Heidelberger, P. and Welch P. D. 1983. Simulation run length control in the presence of an initial transient. *Operations Research* 31: 1109–1144.

Hoogerheide, L. F., Kaashoek, J. F. and van Dijk, H. K. 2007. On the shape of posterior densities and credible sets in instrumental variable regression models with reduced rank: an application of flexible sampling methods using neural networks. *Journal of Econometrics* 139: 154–180.

Heyde, C. C. and Johnstone, I. M. 1979. On Asymptotic Posterior Normality for Stochastic Processes *Journal of the Royal Statistical Society* B41: 184–189.

Jeffreys, H. 1946. An Invariant Form for the Prior Probability in Estimation Problems. *Proceedings of the Royal Society of London* A196: 453–461.

Jeffreys, H. 1961. *Theory of Probability* (3rd ed.). Oxford: Oxford University Press.

Jungbacker, B. and Koopman S. J. 2005. On Importance Sampling for State Space Models. Tinbergen Institute Discussion Paper No. 05–117.

Kibambe, J. and Zellner, A. 2007. Some Policy Experiments Using A Marshallian Macroeconometric Model: Case of South Africa, H. G. B. Alexander Research Foundation Graduate School of Business University of Chicago, Working Paper

Kleibergen, F. R. and van Dijk, H. K. 1998. Bayesian simultaneous equations analysis using reduced rank structures, *Econometric Theory* 14: 701–743.

Putnam, B. and Quintana, J. 1994. New Bayesian statistical approaches to estimating and evaluating models of exchange rates determination. *Proceedings of the ASA Section on Bayesian Statistical Science*. American Statistical Association.

Percy, D. F. 1992. Predictions for Seemingly Unrelated Regressions. *Journal of the Royal Statistical Society* B54: 243–252.

Percy, D. F. 1996. Zellner's Influence on Multivariate Linear Models. In: *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner*, 203–214. New York: John Wiley and Sons.

Press, S. J. 1972. *Applied Multivariate Analysis*. New York: Holt, Rinehart and

Winston.

Quintana, J. and Putnam, B. 1996. Debating currency markets efficiency using multiple-factor models. *Proceedings of the ASA Section on Bayesian Statistical Science.* American Statistical Association.

Raftery, A. E. and Lewis, S. M. 1992. One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo. *Statistical Science* 7: 493–497.

Roberts, G.O. and Rosenthal J.S. 2001. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science* 16: 351–367

Schruben, L. W. 1982. Detecting initialization bias in simulation experiments, *Operations Research* 30: 569–590.

Smith, M. and Kohn, R. 2000. Nonparametric Seemingly Unrelated Regression. *Journal of Econometrics* 98: 257–282.

Spanos, A. 1986. *Statistical Foundations of Econometric Modelling.* Cambridge: Cambridge University Press.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. 2002. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society* B64, 583–639.

Srivastava, V. K. and Giles, D. E. A. 1987. *Seemingly Unrelated Regression Equations Models.* New York: Dekker.

West, M. and Harrison, P.J. 1997. *Bayesian Forecasting and Dynamic Models.* New York: Springer-Verlag.

Zellner, A. 1962. An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias. *Journal of the American Statistical Association* 57: 348–368.

Zellner, A. 1963. Estimators for seemingly unrelated regression equations: some exact finite sample results. *Journal of the American Statistical Association* 58: 977-992.

Zellner, A. 1971. *An introduction to Bayesian inference in econometrics.* New York: Wiley.

Zellner, A. 1994. Bayesian method of moments (BMOM) analysis of mean and regression models. In: *Prediction and modelling honoring Seymour Geisser*, 61–74. New York: Springer.

Zellner, A. and Ando, T. 2008a. A direct Monte Carlo approach for Bayesian analysis of the seemingly unrelated regression model. H.G.B. Alexander Research Foundation, Graduate School of Business, University of Chicago.

Zellner, A. and Ando, T. 2008b. A direct Monte Carlo approach for Bayesian analysis of the simultaneous equation model. H.G.B. Alexander Research Foundation, Graduate School of Business, University of Chicago.

Zellner, A. and Ando, T. 2008c. Bayesian and Non-Bayesian analysis of the seemingly unrelated regression model with Student-t errors and its application for forecasing. *International Journal of Forecasting*, to appear.

Zellner, A., Bauwens, L. and Van Dijk, H. K. 1988. Bayesian specification analysis and estimation of simultaneous equation models using Monte Carlo Methods. *Journal of Econometrics* 38: 39–72.

Zellner, A. and Chen, B, 2002. Bayesian Modeling of Economies and Data Requirements. *Macroeconomic Dynamics* 5: 673–700.

Zellner, A. and Min, C. K. 1995. Gibbs Sampler Convergence Criteria. *Journal of the American Statistical Association* 90: 921–927.

Acknowledgments