REVSTAT – Statistical Journal
Volume 18, Number 4, October 2020, 531–551

ON BAYESIAN ANALYSIS OF SEEMINGLY UNRELATED REGRESSION MODEL WITH SKEW DISTRIBUTION ERROR

Authors: Omid Akhgari
– Department of Statistics, Faculty of Mathematical Sciences, Tarbiat Modares University, Tehran, Iran
o.akhgari@modares.ac.ir

Mousa Golalizadeh
– Department of Statistics, Faculty of Mathematical Sciences, Tarbiat Modares University, Tehran, Iran
golalizadeh@modares.ac.ir

Received: May 2017 Revised: January 2018 Accepted: July 2018

Abstract:

- The simultaneous equation models (SEMs) are one of the standard statistical tools for analyzing multivariate regression when the errors are correlated with some covariates. A particular version of the SEMs is the Seemingly Unrelated Regression (SUR) models which consist of several regression equations with errors being correlated across the equations. There are many occasions in which the normality assumption for the error term might not hold in these models. Although transforming the error to comply with the normal density is a solution, the interpretation of the estimators for the parameters and the associated model might not be straightforward. However, taking into account the skew-normal distribution for the error might, sometimes, be a good alternative. In this paper such scenario is considered as well as a Bayesian framework to estimate the parameters, with a brief review of frequentist methodology. The full conditional posterior densities are derived and relevant statistical inferences are provided. A simulation study is conducted to evaluate the performance of the proposed method. Also, the utilized model is applied to fit relevant equations on Iran gross and income data collected in the year 2009.

Keywords:

- *Simultaneous Equation Model; skew-normal distribution; Bayesian inference; Markov chain Monte Carlo; gross and income.*

AMS Subject Classification:

- 62J20, 62J99.

1. INTRODUCTION

There are many examples in which a single equation can represent a causal relationship among variables. However, there is a case in which individual expression may not cover the desired effect or produce estimates with weak statistical properties. There are examples from many scientific fields such as econometrics where single equations are not enough. In such cases, Simultaneous Equations Models (SEMs) can appropriately represent a joint relationship among variables. The interested readers can, for example, consult [19] for more details on this topic.

Regarding the assumptions of the general structure of any linear model, the predictors are not only fixed but also independent of the error term in the model. However, there are numerous real-life examples in which some of the covariates in a model are correlated with the error term. According to [30] and [11], such variables are called endogenous.

One of the particular cases of SEMs is SUR models, first proposed by [31]. He also explained the procedures to estimate the parameters of these models using the generalized least square method. Amazingly, the literature on treating such models from the frequentist point of view is scarce. For instance, the well known maximum likelihood method of estimation for the parameters of the SUR was only tackled by [13]. However, the popularity of Bayesian approach was more than expected. To name some, we can refer to [28], [21], [29] and [12]. Historically, Bayesian statistical inference on the SUR model was first proposed by [32]. Also, Bayesian moment and direct Monte Carlo method were followed by [33]. Most literature shows that popular MCMC sampling technique was the central theme of study to treat the SUR model . The references include [24], [10], and [27]. Also, [35] proposed the implementation of hierarchical Bayes approach in this model using direct Monte Carlo and importance sampling techniques. Recently, [26] studied the topic of variable selection in the SUR models.

Another important aspect of the SUR models refers to a way one considers a distribution for the error term. It is quite common to choose it as normal. But, there are numerous examples in which the empirical density of response is often asymmetric in practice. One of the procedures to overcome this problem is to utilize some transformations. It might induce relatively normal distribution for the transformed response. However, this strategy has some drawbacks. First, the estimators are usually bias. Secondly, there is lack of proper interpretation for the estimators of the parameters based on the transformed response. Using some asymmetric distributions, which not only possess the same properties as the normal distribution but also can overcome the deficiencies mentioned above, has recently received considerable attention in the literature. The skew-normal density, initially proposed by [3], is one of the well-known distributions to tackle the asymmetric feature of the data. [5] have also discovered the properties of the multivariate skew normal distribution. Later on, [4] studied further features of this density. [7], [17], [18] and [2], among others, provided several generalizations of this distribution. Recently, [6] investigated some other properties of the skew-symmetric distribution.

Most of the research conducted to estimate parameters of a SUR is focused on the case in which the distribution of the variable under investigation is normal. Instead, in this paper we consider the skew-normal distribution for the errors in the SUR and propose procedures to estimate the parameters using the Bayesian methodology. We also conduct some intensive simulation studies to evaluate the methods suggested in this article. Moreover, we show an application of the model in this paper on real-life data.

To present our results, we organized the paper as follows. First, a brief review of the Seemingly Unrelated Regression (SUR) and a Bayesian approach to treating the SUR model with normal distribution for the errors are presented in Section 2. Then, a Bayesian approach to treating the SUR model with skew-normal distribution for the errors is given in Section 3. The simulation study for evaluating the proposed models and the analyses of real-life data, related to the gross and income in the year 2009 in Iran, for illustration purpose, are presented in Sections 4 and 5. General conclusions are provided at the end. The proofs for some theoretical results are sketched in Appendix.

2. BAYESIAN INFERENCE ON PARAMETERS OF A SUR MODEL WITH ERROR DISTRIBUTED AS NORMAL

Econometric analysis of the linear models are usually classified into two scenarios which identify based on the numbers of the equations used to express the relationship among the variables. In the single equation methodology, a dependent variable is typically modeled as a function of one or more covariates. In many situations, such single equation may not cover the desired effect or may even produce estimates with poor statistical properties. The methods of SUR model have been proposed to eliminate the shortcomings obstacles involved in the former methodology. Statistical inference based on the normal response in this model is the object of the current section.

Let assume we aim to estimate the parameters of a SUR model. This objective can commonly be achieved via many parametric and nonparametric estimating procedures based on the frequentist inference including OLS[1], IOLS[2], FGLS[3], IGLS[4] and ML[5]. See, for example, [28] for more details on this topic. However, there are some problems to implement the ML method of estimation in a SUR model. First, there are not usually some explicit expressions for the estimators of the parameters. This fact leads, in turn, to a high cost of analytical computations to solve corresponding normal equations. Secondly, if there is any initial subjective information about the parameters it cannot be directly utilized in the frequentist inference methodology. To overcome these two problems, one can follow a Bayesian approach instead. This section describes the procedure to perform such inference along with general notations used throughout the current paper.

Suppose there are g equations with g endogenous variables associating with $y_1, y_2, ..., y_g$. Specifically, for $i = 1, 2, ..., g$, suppose $X^{(i)}$ is an $n \times k_i$ matrix of explanatory variables and $\beta^{(i)}$ is a k_i-vector of parameters. Then, the i-th equation of a linear simultaneous system can

[1] Ordinary Least Square
[2] Iterative Ordinary Least Square
[3] Feasible Generalized Least Squares
[4] Iterative Generalized Least Squares
[5] Maximum Likelihood

be written as

$$y_{ti} = \sum_{l=1}^{k_i} x_{tl}^{(i)} \beta_l^{(i)} + u_{ti} = X_{t\bullet}^{(i)} \boldsymbol{\beta}_{t\bullet}^{(i)} + u_{ti}, \quad t = 1, 2, ..., n, \tag{2.1}$$

where

$$X_{t\bullet}^{(i)} = (x_{t1}, x_{t2}, ..., x_{tk_i}),$$
$$\boldsymbol{\beta}_{t\bullet}^{(i)} = (\beta_1^{(i)}, \beta_2^{(i)}, ..., \beta_{k_i}^{(i)})^\mathsf{T},$$

and

$$\begin{aligned} &E(u_{ti}) = 0, \qquad \mathrm{Var}(u_{ti}) = \sigma_{ii}, \\ &\mathrm{Cov}(u_{ti}, u_{tj}) = \sigma_{ij}, \quad i, j = 1, 2, ..., g, \quad t = 1, 2, ..., n. \end{aligned} \tag{2.2}$$

Let us define, for fixed t, the g-vectors $y_{t\bullet}$ and $u_{t\bullet}$ consist of the y_{ti}'s and the u_{ti}'s, respectively, for $i = 1, ..., g$. Accordingly, the k-vector $\beta_\bullet$ is formed by stacking the $\boldsymbol{\beta}_{t\bullet}^{(i)}$ vertically. The matrix of $X_{t\bullet}$ is of dimension $g \times k$ and is defined to be a block-diagonal matrix with diagonal blocks $X_{t\bullet}^{(i)}$ also for fixed t, $k = \sum_{i=1}^g k_i$. Precisely, our new notations can be summarized as follows:

$$y_{t\bullet} = \begin{pmatrix} y_{t1} \\ y_{t2} \\ \vdots \\ y_{tg} \end{pmatrix}_{g\times 1}, \qquad u_{t\bullet} = \begin{pmatrix} u_{t1} \\ u_{t2} \\ \vdots \\ u_{tg} \end{pmatrix}_{g\times 1}, \qquad \beta_\bullet = \begin{pmatrix} \boldsymbol{\beta}_{t\bullet}^{(1)} \\ \boldsymbol{\beta}_{t\bullet}^{(2)} \\ \vdots \\ \boldsymbol{\beta}_{t\bullet}^{(g)} \end{pmatrix}_{k\times 1},$$

$$X_{t\bullet} = \begin{pmatrix} X_{t\bullet}^{(1)} & 0 & \dots & 0 \\ 0 & X_{t\bullet}^{(2)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_{t\bullet}^{(g)} \end{pmatrix}_{g\times k}. \tag{2.3}$$

Hence, the linear simultaneous system (2.1) is rewritten as follows

$$y_{t\bullet} = X_{t\bullet}\beta_\bullet + u_{t\bullet}, \quad t = 1, 2, ..., n. \tag{2.4}$$

Based on the assumption for the first two moments of u's, let us consider the normal distribution for them. Then, following the new notations, $u_{t\bullet} \sim N(0_g, \Sigma)$ where

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1g} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2g} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{g1} & \sigma_{g2} & \dots & \sigma_{gg} \end{pmatrix}. \tag{2.5}$$

Now, recalling the expression (2.4) and distribution $u_{t\bullet}$, the likelihood function for the parameters $(\beta_\bullet, \Sigma)$, provide the data including those available in $y_\bullet$ and $X_\bullet$, represented by D, leads to

$$L\big((\beta_\bullet, \Sigma)|D\big) = \frac{1}{(2\pi)^{ng/2}|\Sigma|^{n/2}} \exp\Big\{ -\frac{1}{2} \operatorname{tr}\left(V\Sigma^{-1}\right)\Big\}, \tag{2.6}$$

where 'tr' denotes the trace of matrix and V is a $g \times g$ matrix given by

$$V = \sum_{t=1}^n (y_{t\bullet} - X_{t\bullet}\beta_\bullet)(y_{t\bullet} - X_{t\bullet}\beta_\bullet)^\mathsf{T}.$$

Now, suppose one prefers to follow a Bayesian methodology to estimate the parameters of the SUR model (2.4). As is common, one first should determine priors for the parameters. Both the noninformative and informative priors can be used here. Let us assume a uniform prior for $\beta_\bullet$ and Jeffreys prior for Σ, independent of each other [20]. Then, we have our joint prior, say $\pi_1(\,.\,)$, as

$$\pi_1(\beta_\bullet, \Sigma) = \pi(\beta_\bullet)\pi(\Sigma) \propto |\Sigma|^{-\frac{g+1}{2}}. \tag{2.7}$$

The joint posterior density function is then given by Bayes' theorem, i.e.

$$\pi(\beta_\bullet, \Sigma|D) \propto |\Sigma|^{-(n+g+1)/2} \exp\big[-\frac{1}{2}\,\mathrm{tr}\{V\Sigma^{-1}\}\big]. \tag{2.8}$$

Now, it is straightforward to compute the full conditional posterior distribution $\pi(\beta_\bullet|\Sigma, D)$ and $\pi(\Sigma|\beta_\bullet, D)$. They are given by

$$\begin{aligned} \beta_\bullet|(\Sigma, D) &\sim N(\hat{\beta}_\bullet, \hat{\Sigma}_{\beta_\bullet}) \\ \Sigma|(\beta_\bullet, D) &\sim IW(V, n), \end{aligned} \tag{2.9}$$

where

$$\begin{aligned} \widehat{\beta}_\bullet &= \widehat{\Sigma}_{\beta_\bullet}\big(\sum_{t=1}^{n} X_{t\bullet}^{\mathsf{T}}\Sigma^{-1}y_{t\bullet}\big), \\ \widehat{\Sigma}_{\beta_\bullet} &= \big[\sum_{t=1}^{n} X_{t\bullet}^{\mathsf{T}}\Sigma^{-1}X_{t\bullet}\big]^{-1}, \end{aligned} \tag{2.10}$$

and $IW(\cdot,\cdot)$ denotes the inverse Wishart distribution. As seen, both full conditional posterior distributions have closed forms. Hence, the standard SUR model is also amenable to a 2-block Gibbs sampling formulation. See, for example, [34], for more details.

In some circumstances, one might prefer an informative prior for the parameters $\beta_\bullet$. In such case, it is common to consider the normal density. Precisely, let assume $\beta_\bullet \sim N(\beta_\circ, A_{\beta_\bullet}^{-1})$. Further, suppose the same prior as before has been considered for Σ, i.e. $\pi\big(\Sigma\big) \propto \big|\Sigma\big|^{-\frac{g+1}{2}}$, independently from $\beta_\bullet$. Then, the joint posterior distribution has a closed form in this case as well. However, the conditional posterior distributions have relatively different structures. In particular, it can be shown that

$$\begin{aligned} \beta_\bullet|(\Sigma, D) &\sim N(\overline{\beta}_\bullet, \overline{\Sigma}_{\beta_\bullet}), \\ \Sigma|(\beta_\bullet, D) &\sim IW(V, n), \end{aligned} \tag{2.11}$$

where

$$\begin{aligned} \overline{\beta}_\bullet &= \overline{\Sigma}_{\beta_\bullet}\big[\big(\sum_{t=1}^{n} X_{t\bullet}^{\mathsf{T}}\Sigma^{-1}y_{t\bullet}\big) + A_{\beta_\bullet}\beta_\circ\big], \\ \overline{\Sigma}_{\beta_\bullet} &= \big[\big(\sum_{t=1}^{n} X_{t\bullet}^{\mathsf{T}}\Sigma^{-1}y_{t\bullet}\big) + A_{\beta_\bullet}\big]^{-1}. \end{aligned} \tag{2.12}$$

So far, the full conditional posterior distributions were derived using the assumption of the normal distribution of the errors. In the next section, we assume that the error term follows the skew-normal distribution and compute the posterior density and full conditional distributions.

It worths to mention here that one of the possible procedure to draw samples from the posterior density of the parameters is to follow an MCMC algorithm. Particularly, if the full conditional distributions of relevant parameters are available in closed forms a Gibbs sampling algorithm could be employed to draw samples from corresponding densities. The literature shows that such view to the SUR models [32], [25], [8], [29] and [24] has already investigated this topic.

3. BAYESIAN INFERENCE ON SUR MODELS USING THE SKEW-NORMAL DENSITY FOR ERROR

To consider a normal density for the distribution of the error while utilizing a SUR model is a standard procedure to make statistical inference. However, this assumption might not hold in some real-life example and so corresponding statistical inferences might not lead to feasible results. Instead, to use skew-normal distribution for the density of error is an alternative option. Having said that, to recall ML method of estimation is then one of the conventional parametric statistical inference methods to consider. However, similar to the situation mentioned in the case of considering the normal distribution for error (see initial discussions in Section 2), there are some problems to implement this method as well. Hence, we outline the Bayesian statistical inference on the parameters of a SUR model with the error comes from skew-normal in this section. Moreover, some important statistical features of this strategy are also highlighted.

To start, let us first briefly review the properties of a SUR model under assumption of an skew-normal density for the error in the model (2.4). Specifically, let write

$$u_{t\bullet} = \big(u_{t1}, ..., u_{tg}\big)^{\mathsf{T}} \sim SN\big(0_g,\ \Sigma,\ \lambda\big), \qquad t = 1, ..., n. \tag{3.1}$$

Following [3], the distribution of $u_{t\bullet}$, for $t = 1, ..., n$, is given by

$$f_{U_{t\bullet}}\big(u_{t\bullet}\big) = 2\phi_g\big(u_{t\bullet}; 0_g,\ \Sigma\big)\ \Phi_g\big(\lambda^{\mathsf{T}}\omega^{-1}u_{t\bullet}\big), \tag{3.2}$$

where $\phi_g(u_{t\bullet}; 0, \Sigma)$ is the g-dimensional normal density with zero mean vector and covariance matrix Σ, $\Phi_g(\cdot)$ is the cumulative distribution function of the standard normal density, and λ is a g-dimensional vector with constant values. Here, ω is a diagonal matrix whose components are the square root of the corresponding covariance matrix Σ. Now, we can write down either the likelihood function of the parameters or its logarithm. We prefer the later one, denoted here by $l(\lambda, \beta_\bullet, \Sigma)$, which is given by

$$\begin{aligned} l(\lambda, \beta_\bullet, \Sigma) \;=\; & n\ \log 2 - \frac{ng}{2}\ \log(2\pi) - \frac{n}{2}\ \log|\Sigma| \\ & - \frac{1}{2}\sum_{t=1}^{n}\big[(y_{t\bullet} - X_{t\bullet}\beta_\bullet)^{\mathsf{T}}\Sigma^{-1}(y_{t\bullet} - X_{t\bullet}\beta_\bullet)\big] + \sum_{t=1}^{n}\log\ \Phi_1\big(\lambda^{\mathsf{T}}\omega^{-1}u_{t\bullet}\big). \end{aligned} \tag{3.3}$$

If one is going to estimate the parameters directly using (3.3), there exist some problems. The main drawbacks are lack of convergence in employing any likelihood-based numerical algorithm such pseudo-Newton and the high cost of computations. To circumvent these issues, we propose to follow the Bayesian methodology instead.

Here, an integral part of specifying a Bayesian paradigm is the selection of some prior distributions for all unknown parameters, i.e., $\theta = (\beta_\bullet, \Sigma, \lambda)$. In the absence of prior information and to guarantee to have feasible properties for the posterior, we adopt proper but diffuse priors. Suppose elements of θ are independent a priori, and the following priors have been considered

$$\begin{aligned} &\beta_\bullet \sim N\big(\beta_\circ, \Sigma_{\beta_\circ}\big), \quad \pi(\Sigma) \propto |\Sigma|^{-\frac{g+1}{2}}, \\ &\lambda \sim N\big(\lambda_0, \Lambda_0\big), \quad z_0 \sim N\big(0_g, I_g\big). \end{aligned} \tag{3.4}$$

Then, the join posterior of all parameters is given by

$$\begin{aligned} \pi(\beta_\bullet, \Sigma, \lambda|y) \;\propto\; & \phi_k\big(\beta_\bullet; \beta_\circ, \Sigma_{\beta_\circ}\big) \\ & \times\; \phi_g\big(\lambda; \lambda_0, \Lambda_0\big) \\ & \times\; |\Sigma|^{-\frac{g+1}{2}} \\ & \times\; 2^n \prod_{t=1}^{n} \phi_g\big(y_{t\bullet}; X_{t\bullet}\beta_\bullet, \Sigma\big)\Phi_1(\lambda^\mathsf{T}\omega^{-1}(y_{t\bullet} - X_{t\bullet}\beta_\bullet)). \end{aligned} \tag{3.5}$$

As seen, this expression doesn't have a closed form, so we can not compute the join posterior analytically. To turn around this problem, we use the stochastic representation of the skew-normal distribution (see [1]), i.e. $y_{t\bullet} = \lambda \odot |z_0| + z_1$ where $\odot$ denotes Hadamard product, $z_0 \sim N\big(0_g, I_g\big)$, $z_1 \sim N\big(X_{\bullet t}\beta_\bullet, \Sigma\big)$. Moreover, it is assumed that z_0 and z_1 are independent. Then, it is expected that we could drive the full conditional distributions for each parameters. Below, we provide them in turn. More details on computing those expressions are given in Appendix. Note that we write the full conditional distribution in an deliberator order. The reason to do so is when one is going to update samples from corresponding densities for each parameter in an MCMC sampling algorithm the same order should be followed.

First, we have

$$\beta_\bullet|\big(\Sigma, \lambda, |z_0|, D\big) \sim N\big(\widetilde{\beta}_\bullet, \widetilde{\Sigma}_{\beta_\bullet}\big),$$

where $\widetilde{\beta}_\bullet = \widetilde{\Sigma}_{\beta_\bullet}\big(\sum_{t=1}^n X_{t\bullet}^\mathsf{T}\Sigma^{-1}y_{t\bullet} + \Sigma_{\beta_\circ}^{-1}\beta_\circ - \sum_{t=1}^n X_{t\bullet}^\mathsf{T}\Sigma^{-1}\Lambda|z_0|\big)$ and $\widetilde{\Sigma}_{\beta_\bullet} = \big(\sum_{t=1}^n X_{t\bullet}^\mathsf{T}\Sigma^{-1}X_{t\bullet} + \Sigma_{\beta_\circ}^{-1}\big)^{-1}$.

Secondly, we have

$$\Sigma|\big(\beta_\bullet, \lambda, |z_0|, D\big) \sim IW(R, n), \tag{3.6}$$

where $R = \sum_{t=1}^n \big(y_{t\bullet} - [\lambda \odot |z_0| + X_{t\bullet}\beta_\bullet]\big)\big(y_{t\bullet} - [\lambda \odot |z_0| + X_{t\bullet}\beta_\bullet]\big)^\mathsf{T}$.

Next, the full conditional distributions of λ is given by

$$\lambda|\big(\beta_\bullet, \Sigma, |z_0|, D\big) \sim N\big(\widetilde{\lambda}, \widetilde{\Lambda}\big), \tag{3.7}$$

where $\widetilde{\Lambda} = \big(nZ_0^*\Sigma^{-1}Z_0^* + \Lambda_0^{-1}\big)^{-1}$, $\widetilde{\lambda} = \widetilde{\Lambda}(\sum_{t=1}^n Z_0^*\Sigma^{-1}y_{t\bullet} - \sum_{t=1}^n Z_0^*\Sigma^{-1}X_{t\bullet}\beta_\bullet + \Lambda_0^{-1}\lambda_0)$, and Z_0^* is an $g \times g$ diagonal matrix whose components are filled with elements of vector $|z_0|$.

Finally, at the last step, the density of $|z_0|$ should be derived. It is straightforward to show that

$$|z_0| \;\Big|\Big(\beta_\bullet, \Sigma, \lambda, D\Big) \sim TN\Big(\widetilde{z}_0, \Psi_{z_0}, (0, +\infty)\Big), \tag{3.8}$$

where $TN\Big(\mu, \Sigma, (a,b)\Big)$ stands for the multivariate truncated normal distribution $N(\mu,\Sigma)$ lying within the interval $(a,b), -\infty \leq a < b \leq +\infty$. Also $\Psi_{z_0} = \big(I_g + n\Delta\Sigma_{-1}\Delta\big)^{-1}$ and $\widetilde{z}_0 = \Psi_{z_0}\big(\sum_{t=1}^{n}\Delta\Sigma^{-1}[y_{t\bullet} - X_{t\bullet}^{\mathsf{T}}\beta_\bullet]\big)$ where $\Delta = \mathrm{diag}(\lambda_1, ..., \lambda_g)$.

Now, we are at a position to conduct some simulation studies to evaluate the proposed models.

4. SIMULATION STUDIES

Here, we outline our simulation studies to evaluate the procedure in estimating the parameters of the SUR models given in Sections 2 and 3. Suppose the following simultaneous model is given:

$$\begin{cases} y_1 = \beta_0 + \beta_1 z_1 + \beta_2 x_1 + u_1, \\ y_2 = \gamma_0 + \gamma_1 z_1 + \gamma_2 x_2 + u_2. \end{cases} \tag{4.1}$$

The assumptions imposed for this model are the same as those proposed in Sections 2 and 3. Moreover, $u = (u_1, u_2)^{\mathsf{T}} \sim N(0, \Sigma)$ where y_1 and y_2 are endogenous variables. In addition, we assume variables z_1, x_1 and x_2 are exogenous. When we switch to the scenario in which the error terms follow the asymmetric distribution, we assume $u = (u_1, u_2)^{\mathsf{T}} \sim SN(0_2, \Sigma, \lambda)$ where λ is shape parameter vector.

To generate data from the model with equations in (4.1), we do need to fix the parameters. We are writing them all together either in regression equations or explicit expressions. They are given as follows:

$$\begin{aligned} y_1 &= 4 - 3z_1 - 4x_1 + u_1, \\ y_2 &= 7 + 3z_1 - 2x_2 + u_2, \end{aligned} \tag{4.2}$$

$$\begin{aligned} u &= \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \sim SN(0_2, \Sigma, \lambda), \\ \Sigma &= \begin{pmatrix} 3 & -1 \\ -1 & 4 \end{pmatrix}, \quad \lambda = \begin{pmatrix} 4 \\ 7 \end{pmatrix}. \end{aligned}$$

The sample size for each equation is fixed at 1000 cases. Consequently, due to having two equations in (4.1), the total number of available data is 2000. The Bayesian inference is conducted using the proper priors given in (3.4) with the hyperparameters fixed on some specific values. Particularly, we consider

$$\begin{aligned} \beta_\bullet \sim N(0_2, 100I_2), \qquad \lambda \sim N(0_2, 100I_2), \\ \Sigma \propto \begin{vmatrix} 0.01 & 0 \\ 0 & 0.01 \end{vmatrix}^{-\frac{1}{2}}. \end{aligned}$$

As mentioned earlier, one can follow the Gibbs sampling algorithm to draw samples and then estimate the parameters of the models while employing an MCMC algorithm. To do this, we fixed the simulated sample size at 100,000 iterations for each chain. Convergence of the

MCMC algorithm was confirmed by the Gelman and Rubin convergence measure [16], but not reported here. To get independent samples, the burn-in was set on 25,000 iterations for each chain and the last 75,000 iterations were used to make statistical inference on parameters. Then, with taking each 50-th observation, we were ultimately left with 1,500 samples. The summarized results are presented in Table 1.

As can be seen, the Table 1 includes two parts. The results using the normal and skew-normal distributions assumption for the errors are shown in the left and right panels, respectively. The regression coefficients and covariance elements are also estimated. As the values in the left panel of the table show, both intercepts for each equation of the model (4.1) are overestimated. This phenomenon is also the case for the elements of the covariance matrix. The other coefficients are estimated relatively as good as expected.

Table 1: The estimate of parameters and other measures after fitting the SUR model (4.1) under the assumptions of the skew-normal (right panel) and normal (left panel) distributions for the errors.

Parameter	N-MCMC					SN-MCMC				
	Mean	Sd	2.5%	97.5%	ES	Mean	Sd	2.5%	97.5%	ES
β_0	9.38	0.334	8.722	10.01	5.38	3.875	0.326	3.284	4.521	0.125
β_1	−3.009	0.017	−3.043	−2.973	0.009	−2.986	0.015	−3.027	−2.97	0.014
β_2	−3.981	0.024	−4.026	−3.936	0.009	−3.983	0.021	−4.029	−3.948	0.017
γ_0	16.79	0.562	15.7	17.86	9.79	7.510	0.391	5.619	7.139	0.510
γ_1	3.05	0.033	2.981	3.111	0.05	3.006	0.020	2.987	3.065	0.006
γ_2	−1.934	0.039	−2.009	−1.856	0.066	−2.000	0.025	−2.027	−1.927	0.000
σ_{11}	20.25	0.883	18.52	22.01	17.25	3.504	0.572	2.522	4.798	0.504
σ_{12}	1.123	1.215	−1.307	3.549	2.123	−1.298	0.643	−2.436	0.056	0.298
σ_{22}	74.53	3.334	68.23	81.09	70.53	3.211	0.769	1.839	4.887	0.789
λ_1	—	—	—	—	—	6.498	0.250	6.303	7.451	2.498
λ_2	—	—	—	—	—	12.136	0.371	12.01	12.43	5.136

The results using the skew-normal Bayesian approach is given on the right panel of Table 1. As seen, relatively small values of effect size (ES), defined as absolute bias, indicate that all parameters are well estimated. To consider all measures, a general result is that taking into account the skew-normal distribution instead of normal for the errors does better a job of fitting the model (4.1) while employing a Bayesian approach to making statistical inference.

To evaluate performance of the method in more details, we iterated the procedure of generating the data from model (4.1) with the sample size of $n = 1000$ for the numbers of 50 times and then computed the Mean Squared Error (MSE) criterion, i.e.

$$\mathrm{MSE}(\hat{\alpha}) = \frac{1}{50}\sum_{i=1}^{50}(\hat{\alpha}_i - \alpha_{\mathrm{True}})^2. \tag{4.3}$$

The results (not shown here) have confirmed that the MSE criterion in estimating the parameters using the skew-normal Bayesian approach is very close to zero, but this was not the case for the normal distribution assumption for the error in the model (4.1). It means that the estimators derived from the skew-normal case are relatively more accurate and precise than the normal assumption.

5. REAL APPLICATION

We are interested in applying the proposed models in this paper on real-life data. To do this, we used the cost and income data collected on year 2009 in Iran. There are about 13,345 families from 32 provinces. Here, the main goal is on survey effects of some variables on gross cost (GH) and income (D). In this study, both of these quantities are considered as endogenous variables and other covariates are set as exogenous. A general description of the considered variables are reported in Table 2. Also, Figures 1, 2 and 3 (upper panel) provide some geometric displays of some exogenous and two endogenous variables, i.e. GH, D.

Table 2: A general description of variables utilized in real application.

Variable names	Abbreviation signs	Variable Type	Codes
Gross cost	GH	Quantitative	—
Income	D	Quantitative	—
Family size	C_1	Quantitative	—
Number of literate	C_2	Quantitative	—
Number of employees	C_3	Quantitative	—
Number of people with income	C_4	Quantitative	—
Age	A	Quantitative	—
Location Area	B_1	Quantitative	—
Private car	B_2	Qualitative	1: Use, 0: Nonuse
Internet	B_3	Qualitative	1: Use, 0: Nonuse
Gas	B_4	Qualitative	1: Use, 0: Nonuse
Mobile	B_5	Qualitative	1: Use, 0: Nonuse
Incomes of agricultural free businesses	D_1	Quantitative	—
Incomes of nonfarm free businesses	D_2	Quantitative	—
Other Incomes	D_3	Quantitative	—
Other Non-monetary Incomes	D_4	Quantitative	—

To initiate a statistical analysis based upon a common linear regression model, we are concerned about the accuracy of considering the normality assumption for the response variables, i.e. GH, and D, here. We used the Kolmogorov- Smirnov (KS) test for this purpose. Based on this test, normality assumption has not been confirmed for both endogenous variables with the p-value < 0.05. Further, to have visual tools, the quantile-quantile plots for each of the income and gross cost were also drawn. They appeared on the lower panel of Figure 3. As seen, both plots are confirming a lack of the normal distributions fitting for each variable. Moreover, the contour plot based on these endogenous variables, which is appeared on the upper panel Figure 3, also shows a departure from the bivariate normal distribution assumption. It might be argued here that a logarithm transformation of the endogenous variables might solely lead to a better fit of normality assumption for these variables. However, based on our investigation (not reported here), we did not reach to such conclusion. Hence, we preferred to invoke some asymmetric densities, particularly skew-normal distribution, to proceed our analysis. However, we also utilized the normal distribution for the endogenous variables to make a comparison, similar to our simulation studies.

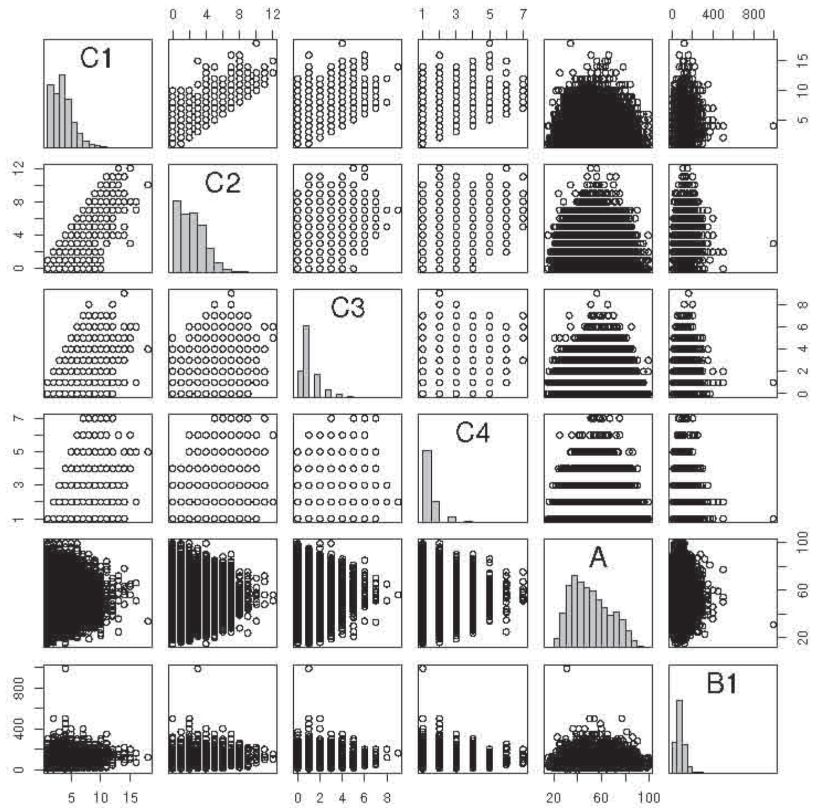


Figure 1: The pairs plot of quantitative variables described in Table 2.

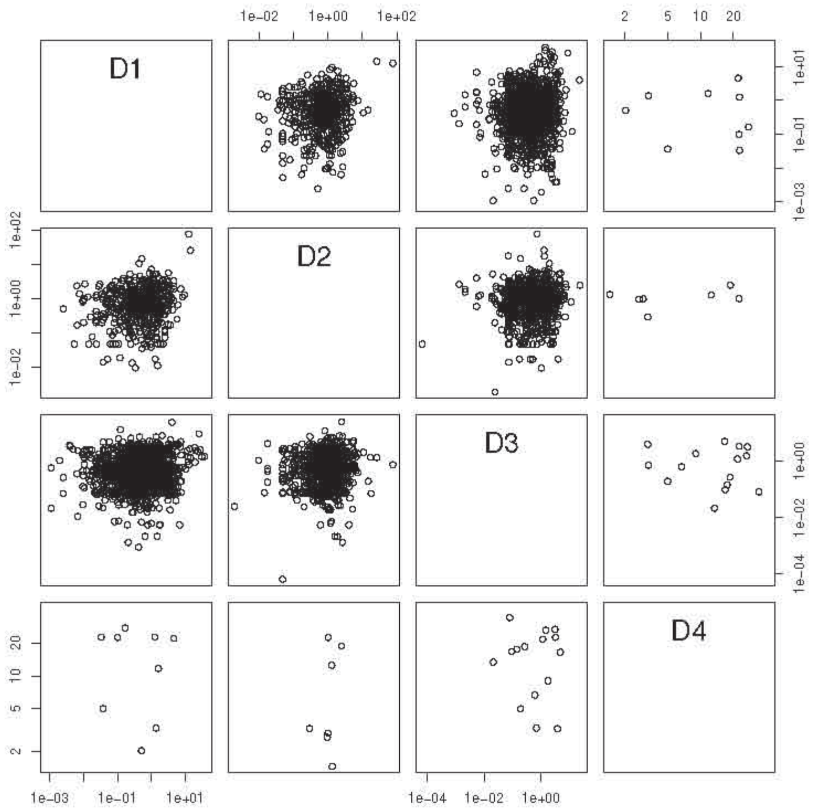
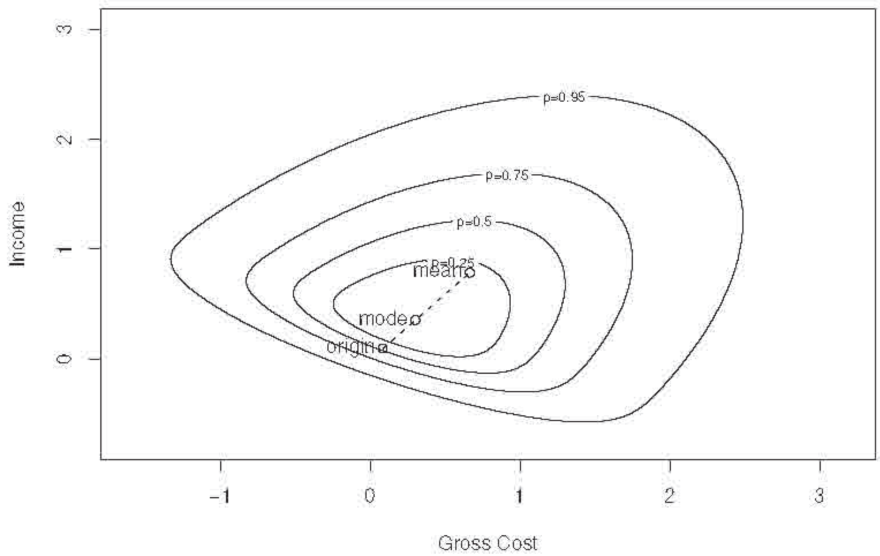


Figure 2: The pairs plot for several types of incomes described in Table 2.



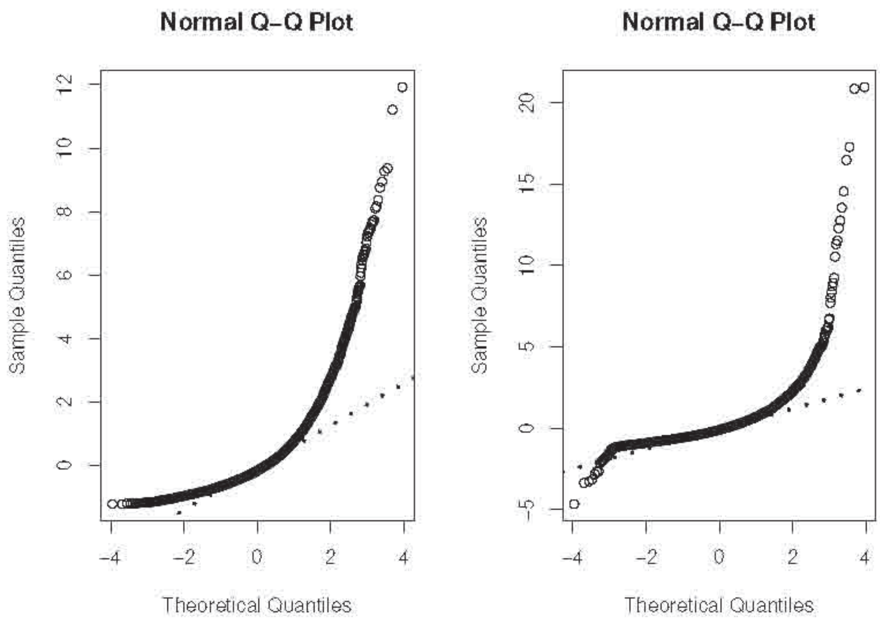


Figure 3: The contour plot of gross cost (GH) against income (D) (upper panel) along with the quantile-quantile plot (lower panel) for each of them. Lack of normal distribution fitting, either jointly or marginally, using endogenous variables are apparent from both plots.

Based upon a general view of the data and also after consulting the subjects with some econometrics experts in Statistical Center of Iran, the following SUR model was utilized to express the inter-relationship between endogenous and exogenous variables:

$$
\begin{aligned}
GH &= \beta_0 + \sum_{i=1}^{4} \beta_{C_i} C_i + \sum_{i=1}^{5} \beta_{B_i} B_i + \beta_A A + \epsilon_1, \\
D &= \gamma_0 + \sum_{i=1}^{4} \gamma_{D_i} D_i + \epsilon_2.
\end{aligned} \tag{5.1}
$$

This model has been fitted through both frequentist and Bayesian approaches as well as under the assumption of the normal (N) and skew-normal (SN) distributions for the errors. Table 3 shows the estimates along with standard errors of the estimates. As seen, the table is divided in three parts. The first row panel represents the quantities mentioned above for the parameters of the first equation in (5.1). Similarly, those for the second equation appear in the second row panel. Finally, the last row panel constitutes the estimates and their standard errors for the components of the covariance matrix of the errors in (5.1) as well as those values for the skewness parameters, if they are required. The important point to emphasize is that we have only reported those estimates which were significant at %5 level. Hence, one does not see some of the coefficients from the SUR model (5.1) in Table 3.

Table 3: The estimates along with standard errors of the estimates for the parameters of the SUR model fitted through both frequentist and Bayesian approaches as well as under assumption of the normal (N) and skew-normal (SN) distributions for the errors in (5.1) for the Iranian cost and income data collected in year 2009.

Parameter	Bayesian				Frequentist			
	Estimation		Standard error		Estimation		Standard error	
	N	SN	N	SN	N	SN	N	SN
β_0	-1.338	-1.581	0.030	0.017	-1.50	-1.34	0.047	0.006
β_{C_1}	0.048	0.026	0.006	0.003	0.036	0.040	0.009	0.001
β_{C_2}	0.056	0.045	0.006	0.004	0.082	0.046	0.010	0.002
β_{C_3}	0.070	0.047	0.007	0.004	0.106	0.059	0.011	0.004
β_{C_4}	-0.015	0.014	0.009	0.005	0.061	-0.024	0.014	0.004
β_{B_1}	0.006	0.001	0.001	0.001	0.003	0.003	0.0006	0.0001
β_{B_2}	0.004	0.002	0.001	0.001	0.004	0.002	0.0002	0.0005
β_{B_3}	0.531	0.304	0.016	0.009	0.649	0.531	0.024	0.013
β_{B_4}	0.509	0.281	0.032	0.018	0.689	0.490	0.051	0.032
β_{B_5}	0.011	0.049	0.012	0.007	0.064	0.031	0.018	0.0085
β_A	0.236	0.180	0.016	0.009	0.276	0.137	0.025	0.0064
γ_0	0	-0.514	0.004	0.0001	0	-0.103	0.025	0.0038
γ_{D_1}	0.500	0.562	0.004	0.001	0.544	0.499	0.013	0.0039
γ_{D_2}	0.467	0.498	0.004	0.002	0.503	0.487	0.013	0.0039
γ_{D_3}	0.352	0.375	0.004	0.001	0.412	0.352	0.013	0.0039
γ_{D_4}	0.030	0.030	0.004	0.002	0.033	0.030	0.013	0.0038
σ_{11}	0.671	0.018	0.007	0.009	0.656	0.051	-	-
σ_{21}	0.129	0.001	0.004	0.002	0.084	0.009	-	-
σ_{22}	0.317	0.003	0.003	0.001	0.315	0.018	-	-
λ_1	-	1.194	-	0.007	-	1.181	-	-
λ_2	-	0.764	-	0.003	-	0.869	-	-

Based upon results given at first row panel in Table 3, the number of literate, employees and family size have a direct effect on total family gross cost. The usage of facilities, including Private car, internet, gas and mobile, also has the positive impact on family gross cost. In other words, the utilization of these services leads to an increase in family gross cost. However, if the families are not using these items still there is an increment on cost too. The rationale behind this surprising result comes from the SUR model in which those families would then pay for other luxuries items. Finally, the age of the people who are in charge of the family cost and also the area where the families live both leada to the positive effect on the gross cost.

Now, let us analyze the result at the second row panel in Table 3. As seen, the incomes from the agricultural and non-farm free businesses, other incomes and non-monetary gains have direct effect on the family incomes. Furthermore, other non-monetary earning incomes have less effect on the family incomes subject to other variables. Also, the effect of the incomes from agricultural free businesses on the family incomes are high.

Some interesting results appear in the second row panel of Table 3. First, it is related to the estimate of the intercept (γ_0). Unlike the case for the skew-normal distribution, its estimation is zero when assuming a normal density for the errors in the SUR model (5.1). Second, as seen, the estimates for the components of the covariance matrix based on normality assumption are somewhat bigger than those in the skew-normal case. Albeit, this needs more considerations.

So far, the reader probably discovered a proper strategy to fit the SUR model (5.1) based on the presented results. However, we are interested in selecting one of two methodologies and distributional assumption for the errors through utilizing a sensible statistical measure. There are several methods to choose the appropriate model among two possible candidates while employing either Bayesian or frequentist statistical inferences methodologies. It is usually accepted among statisticians that the Bayes factor criterion is a proper measure to compare the performance of different candidate models while implementing a Bayesian methodology. However, in utilizing the frequentist methodology, the researchers consider the log likelihood and AIC criteria. To compare two candidate models L_1 and L_2 the Bayes factor is represented by a quantity which is simply a ratio (see [22]). Precisely, suppose $\pi(L_1)$ and $\pi(L_2)$ are priors for two models. Then, given the data D, the Bayes factor of model L_1 w.r.t L_2 is written as

$$B_{12} = \frac{Pr(D|L_1)}{Pr(D|L_2)} = \frac{\frac{\pi(L_1|D)}{\pi(L_2|D)}}{\frac{\pi(L_1)}{\pi(L_2)}}, \tag{5.2}$$

where $Pr(D|L) = \int f(D|L,\theta)\pi(\theta|D)d\theta$ and $\theta = (\beta, \gamma, \Sigma, \lambda)$. However, because we are not able to compute the joint posterior analytically this criteria cannot be employed here. [23] proposed a method when there is no closed form for the posterior density. Following them, if $\{\theta\}_{i=1}^{m}$ are samples from the posterior distribution of $\pi(\theta|D, L)$, we can write:

$$f^{(j+1)}(D|L) = \frac{\frac{km}{1-k} + \sum_{j=1}^{m} \frac{f(D|\theta^{(j)},L)}{kf^{(j)}(D|L)+(1-k)f(D|\theta^{(j)},L)}}{\frac{km}{(1-k)f^{(j)}(D|L)} + \sum_{j=1}^{m} \frac{1}{kf^{(j)}(D|L)+(1-k)f(D|\theta^{(j)},L)}}, \tag{5.3}$$

where k is a small value being in the interval $(0, 1)$. To derive this quantity, we repeated our analysis till achieving a reasonable convergence. In some small-scale numerical experiments, we have discovered that the quantity (5.3) performed well for k as small as 0.01.

The logarithm of pseudo-marginal likelihood (LPML) is another criterion to select between two candidate models (see [14]). It is derived from predictive considerations, particularly Conditional Predictive Ordinate (CPO), and leads to pseudo-Bayes factors for choosing an optimal model. It is popular mainly due in part to its relative ease of computation making the LPML a stable estimate base on the samples derived from any MCMC algorithm. Following the [15] and assuming availability of the samples $\theta^{(1)}, ..., \theta^{(s)}$, obtained from corresponding posterior, the i-th CPO_i and $LPML$ are, respectively, estimated as

$$\frac{1}{CPO_i} = \frac{1}{s}\sum_{k=1}^{s} \frac{1}{f_i(y_i|\theta^{(k)}, M)}, \tag{5.4}$$

and

$$LPML = \sum_{i=1}^{n} \log(CPO_i). \tag{5.5}$$

The LPML and BF quantities for both cases (N and SN) are reported in Table 4. As seen, the value of LPML for SN is greater than that for N. Moreover, the ratio of BF for SN in compare with the N model is relatively bigger and so indicating the superiority of SN again. Although there are some debates on using these criteria under the frequentist view, we also reported the estimates of the parameters for both N and SN cases just to have a basis for seeing difference on utilizing two methodologies.

Following the results gained in analyzing this example, our recommendation is to consider a skew-normal rather than the normal density for the error while using a SUR model to analyze the Iran gross and income data collected in year 2009.

Table 4: The performance criteria of the SUR model (5.1) fitted using the skew-normal distribution on the Iranian cost and income data collected in year 2009.

Model	Bayesian		Frequentist	
	BF	LPML	AIC	Log likelihood
N	0.917	−59274.01	80925.15	−40443.57
SN	1.091	−34585.61	−124982.3	62512.15

6. CONCLUSION

When dealing with simultaneous relationship among variables, the SUR model is a particular case of SEM. The frequentist inference utilized for the SUR model under the skew-normal assumption or the error is very time consuming and also challenging to tackle. Hence, a Bayesian inference implemented in the SUR model under skew-normal distribution assumption for errors is developed in this paper. Regarding the model selection, the BF and LPML criteria had some superiorities in choosing better model using our real data set as well as in our simulation studies. Based on the results in this paper we can stat when data are not symmetric, the SUR model accompanied with considering a skew-normal distribution for the error performs well on fitting the data at least in comparison with invoking the normal density.

In future study, we aim to investigate how the endogenous variables can improve the estimation of the parameters in the SEMs while the errors follow the skew-normal distributions. Moreover, to plug in these structures into multilevel models is the way we are going to extend our current research.

A. APPENDIX

According to formula (2.8), full conditional posterior is given by

$$\begin{aligned}
g(\beta_\bullet|\Sigma,D) &\propto L(D|\beta_\bullet,\Sigma)\pi_1(\beta_\bullet) \propto \exp\big\{-1/2\ \operatorname{tr}(V\Sigma^{-1})\big\}\\
&\propto \exp\Big\{-1/2\sum_{t=1}^{n}(y_{t\bullet}-X_{t\bullet}\beta_\bullet)^{\mathsf T}\Sigma^{-1}(y_{t\bullet}-X_{t\bullet}\beta_\bullet)\Big\}\\
&\propto \exp\Big\{-1/2\sum_{t=1}^{n}\big[-\beta_\bullet^{\mathsf T}X_{t\bullet}^{\mathsf T}\Sigma^{-1}+\beta_\bullet^{\mathsf T}X_{t\bullet}^{\mathsf T}\Sigma^{-1}X_{t\bullet}\beta_\bullet-y_{t\bullet}^{\mathsf T}\Sigma^{-1}X_{t\bullet}\beta_\bullet\big]\Big\}\\
&\propto \exp\Big\{-1/2\big[\beta_\bullet-\big(\sum_{t=1}^{n}X_{t\bullet}^{\mathsf T}\Sigma^{-1}X_{t\bullet}\big)^{-1}\big(\sum_{t=1}^{n}X_{t\bullet}^{\mathsf T}\Sigma^{-1}y_{t\bullet}\big)\big]^{\mathsf T}\big[\sum_{t=1}^{n}X_{t\bullet}^{\mathsf T}\Sigma^{-1}X_{t\bullet}\big]\\
&\qquad \big[\beta_\bullet-\big(\sum_{t=1}^{n}X_{t\bullet}^{\mathsf T}\Sigma^{-1}X_{t\bullet}\big)^{-1}\big(\sum_{t=1}^{n}X_{t\bullet}^{\mathsf T}\Sigma^{-1}y_{t\bullet}\big)\big]\Big\}.
\end{aligned}\tag{A.1}$$

So $\beta_\bullet$ is multivariate normal with mean $\big(\sum_{t=1}^{n}X_{t\bullet}^{\mathsf T}\Sigma^{-1}X_{t\bullet}\big)^{-1}\big(\sum_{t=1}^{n}X_{t\bullet}^{\mathsf T}\Sigma^{-1}y_{t\bullet}\big)$ and covariance matrix $\big(\sum_{t=1}^{n}X_{t\bullet}^{\mathsf T}\Sigma^{-1}X_{t\bullet}\big)^{-1}$. The full conditional posterior is derived as follows:

$$g(\Sigma|\beta_\bullet,D) \propto L(D|\beta_\bullet,\Sigma)\pi_1(\Sigma) \propto |\Sigma|^{-(n+g+1)/2}\exp\big[-1/2\ \operatorname{tr}\{V\Sigma^{-1}\}\big].\tag{A.2}$$

It is straightforward to check that this expression is proportional to an inverse Wishart distribution with degrees of freedom n and scale covariance V.

Based on the prior density $\beta_\bullet \sim N(\beta_\circ, A_{\beta_\circ}^{-1})$ and $\pi(\Sigma)\propto|\Sigma|^{-\frac{g+1}{2}}$, the posterior distribution can easily be computed. However, it doesn't have a closed form. Instead, the full conditional posterior can be obtained using the expressions

$$\begin{aligned}
g(\beta_\bullet|\Sigma,D) &\propto L(D|\beta_\bullet,\Sigma)\pi_2(\beta_\bullet)\\
&\propto |\Sigma|^{-\frac{n}{2}}\exp\Big\{-1/2\ \Big[\sum_{t=1}^{n}(y_{t\bullet}-X_{t\bullet}\beta_\bullet)^{\mathsf T}\Sigma^{-1}(y_{t\bullet}-X_{t\bullet}\beta_\bullet)\\
&\quad +(\beta_\bullet-\beta_\circ)^{\mathsf T}A_{\beta_\bullet}(\beta_\bullet-\beta_\circ)\Big]\Big\}\\
&\propto \exp\Big\{-1/2\ \Big[\beta_\bullet^{\mathsf T}\Big\{\big(\sum_{t=1}^{n}X_{t\bullet}^{\mathsf T}\Sigma^{-1}X_{t\bullet}\big)+A_{\beta_\bullet}\Big\}\beta_\bullet\\
&\quad -2\beta_\bullet^{\mathsf T}\Big\{\big(\sum_{t=1}^{n}X_{t\bullet}^{\mathsf T}\Sigma^{-1}y_{t\bullet}\big)+A_{\beta_\bullet}\beta_\circ\Big\}\Big]\Big\}\\
&\propto \exp\Big\{-1/2\ (\beta_\bullet-\overline{\beta}_\bullet)^{\mathsf T}\overline{\Sigma}_{\beta_\bullet}^{-1}(\beta_\bullet-\overline{\beta}_\bullet)\Big\}.
\end{aligned}\tag{A.3}$$

Consequently $\beta_\bullet$ given (Σ,D) is multivariate normal with mean

$$\overline{\beta}_\bullet = \overline{\Sigma}_{\beta_\bullet}\Big[\big(\sum_{t=1}^{n}X_{t\bullet}^{\mathsf T}\Sigma^{-1}y_{t\bullet}\big)+A_{\beta_\bullet}\beta_\circ\Big]$$

and covariance matrix

$$\overline{\Sigma}_{\beta_\bullet} = \Big[\Big(\sum_{t=1}^{n} X_{t\bullet}^{\mathsf{T}} \Sigma^{-1} X_{t\bullet}\Big) + A_{\beta_\bullet}\Big]^{-1}.$$

Similarity, the full conditional posterior $\Sigma|(\beta_\bullet, D)$ is given by

$$\pi(\Sigma|\beta_\bullet, D) \propto L(D|\beta_\bullet, \Sigma)\pi_2(\Sigma) \propto |\Sigma|^{-(n+g+1)/2} \exp\big[-1/2 \ \mathrm{tr}\{V\Sigma^{-1}\}\big]. \tag{A.4}$$

Seeing similarity of this expression to (A.2), the full conditional distribution $\Sigma|(\beta_\bullet, D)$ is inverse Wishart with degrees of freedom n and scale covariance V.

Recall: The probability density function for the random matrix X $(n \times p)$ that follows the matrix normal distribution $MN_{n\times p}\big(M, U, V\big)$ has the form

$$p\big(X|M, U, V\big) = \frac{\exp\big(-\frac{1}{2}\,\mathrm{tr}\,\big[V^{-1}(X - M^{\mathsf{T}})U^{-1}(X - M)\big]\big)}{(2\pi)^{np/2}|V|^{n/2}|U|^{p/2}}, \tag{A.5}$$

where M is $n \times p$, U is $n \times n$ and V is $p \times p$ matrices. Note that the matrix normal is linked to the multivariate normal distribution in the following way:

$$X \sim MN_{n\times p}\big(M, U, V\big) \tag{A.6}$$

if and only if

$$\mathrm{Vec}(X) \sim N\big(\mathrm{Vec}(M)_{np}, V \otimes U\big), \tag{A.7}$$

where $\mathrm{Vec}(M)$ denotes the vectorization of M.

Suppose that $y_{t\bullet} \sim SN\big(X_{t\bullet}\beta_\bullet\big)$. According to stochastic representations of multivariate skew-normal distribution (see [1]), we have

$$y_{t\bullet} = \lambda \odot |z_0| + z_1, \tag{A.8}$$

where $\odot$ denotes Hadamard product, $z_0 \sim N\big(0_g, I_g\big)$, $z_1 \sim N\big(X_{\bullet t}\beta_\bullet, \Sigma\big)$ and also z_0 and z_1 are independent. Thus,

$$\begin{pmatrix} z_0 \\ z_1 \end{pmatrix} \sim N\Big(\begin{pmatrix} 0 \\ X_{t\bullet}\beta_\bullet \end{pmatrix}, \begin{pmatrix} I_g & 0 \\ 0 & \Sigma \end{pmatrix}\Big). \tag{A.9}$$

Thus, the conditional distribution $y_{t\bullet}$ given z_0 leads to

$$y_{t\bullet}|z_0 \sim N\big(\lambda \odot |z_0| + X_{t\bullet}\beta_\bullet, \Sigma\big). \tag{A.10}$$

The full conditional posterior distribution of all parameters are determined based on

(A.10). So, for $\beta_\bullet|\big(\Sigma,\lambda,z_0,D\big)$, we have

$$\begin{aligned}
\pi\big(\beta_\bullet|\Sigma,\lambda,z_0,D\big) &\propto L(y_{t\bullet}|z_0,\Sigma,\lambda,\beta_\bullet)\pi(\beta_\bullet)\\
&\propto |\Sigma|^{-\frac{n}{2}}\exp\Big\{-\frac{1}{2}\sum_{t=1}^{n}\big(y_{t\bullet}-[\lambda\odot|z_0|+X_{t\bullet}\beta_\bullet]\big)^{\mathsf{T}}\Sigma^{-1}\\
&\quad\big(y_{t\bullet}-[\lambda\odot|z_0|+X_{t\bullet}\beta_\bullet]\big)\Big\}\\
&\quad\times|\Sigma_{\beta_\circ}|^{-\frac{1}{2}}\exp\big\{(\beta_\bullet-\beta_\circ)^{\mathsf{T}}\Sigma_{\beta_\circ}^{-1}(\beta_\bullet-\beta_\circ)\big\}\\
&\propto \exp\Big\{-\frac{1}{2}\big[\beta_\bullet^{\mathsf{T}}(\sum_{t=1}^{n}X_\bullet^{\mathsf{T}}\Sigma^{-1}X_\bullet^{\mathsf{T}}+\Sigma_{\beta_\circ}^{-1})\beta_\bullet\\
&\quad-2\beta_\bullet^{\mathsf{T}}(\sum_{t=1}^{n}X_\bullet^{\mathsf{T}}\Sigma^{-1}y_\bullet-\sum_{t=1}^{n}X_\bullet^{\mathsf{T}}\Sigma^{-1}\lambda\odot|z_0|+\Sigma_{\beta_\circ}^{-1}\beta_\circ)\big]\Big\}\\
&\propto \exp\Big\{-\frac{1}{2}(\beta_\bullet-\widetilde{\beta}_\bullet)^{\mathsf{T}}\widetilde{\Sigma}_{\beta_\bullet}^{-1}(\beta_\bullet-\widetilde{\beta}_\bullet)\Big\}. \qquad \text{(A.11)}
\end{aligned}$$

Consequently, $\beta_\bullet|\big(\Sigma,\lambda,z_0,D\big)$ is multivariate normal with mean

$$\widetilde{\beta}_\bullet=\widetilde{\Sigma}_{\beta_\bullet}(\sum_{t=1}^{n}X_\bullet^{\mathsf{T}}\Sigma^{-1}y_\bullet-\sum_{t=1}^{n}X_\bullet^{\mathsf{T}}\Sigma^{-1}\lambda\odot|z_0|+\Sigma_{\beta_\circ}^{-1}\beta_\circ)$$

and covariance

$$\widetilde{\Sigma}_{\beta_\bullet}=(\sum_{t=1}^{n}X_\bullet^{\mathsf{T}}\Sigma^{-1}X_\bullet^{\mathsf{T}}+\Sigma_{\beta_\circ}^{-1})^{-1}.$$

Similarly, the full conditional posterior distribution $\Sigma|\big(\beta_\bullet,\lambda,z_0,D\big)$ is computed: i.e.

$$\begin{aligned}
\pi\big(\Sigma|\beta_\bullet,\lambda,z_0,D\big) &\propto L(y_{t\bullet}|z_0,\Sigma,\lambda,\beta_\bullet)\pi(\Sigma)\\
&\propto |\Sigma|^{-\frac{n}{2}}\exp\Big\{-\frac{1}{2}\sum_{t=1}^{n}\big(y_{t\bullet}-[\lambda\odot|z_0|+X_{t\bullet}\beta_\bullet]\big)^{\mathsf{T}}\Sigma^{-1}\\
&\quad\big(y_{t\bullet}-[\lambda\odot|z_0|+X_{t\bullet}\beta_\bullet]\big)\Big\}\\
&\quad\times|\Sigma|^{-\frac{g+1}{2}}\\
&\propto |\Sigma|^{-\frac{n+g+1}{2}}\exp\Big\{-\frac{1}{2}\operatorname{tr}(R\Sigma^{-1})\Big\}, \qquad \text{(A.12)}
\end{aligned}$$

where $R=\sum_{t=1}^{n}\big(y_{t\bullet}-[\lambda\odot|z_0|+X_{t\bullet}\beta_\bullet]\big)\big(y_{t\bullet}-[\lambda\odot|z_0|+X_{t\bullet}\beta_\bullet]\big)^{\mathsf{T}}$. So

$$\Sigma|\big(\beta_\bullet,\lambda,z_0,D\big)\sim IW(R,n),$$

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

Suppose

$$Z_0^*=\begin{pmatrix}|Z_{0_1}| & 0 & \dots & 0\\ 0 & |Z_{0_2}| & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & |Z_{0_g}|\end{pmatrix}_{g\times g}.$$

Now, we can write $\lambda^{\mathsf{T}} \odot |z_0|^{\mathsf{T}} = \lambda^{\mathsf{T}} Z_0^*$. Then, the full conditional $\lambda|\big(\beta_\bullet, \Sigma, z_0, D\big)$ is given by

$$
\begin{aligned}
\pi\big(\lambda|\Sigma, \beta_\bullet, z_0, D\big) &\propto L(y_{t\bullet}|z_0, \Sigma, \lambda, \beta_\bullet)\pi(\lambda) \\
&\propto |\Sigma|^{-\frac{n}{2}} \exp\Big\{-\frac{1}{2}\sum_{t=1}^{n}\big(y_{t\bullet} - [\lambda \odot |z_0| + X_{t\bullet}\beta_\bullet]\big)^{\mathsf{T}}\Sigma^{-1} \\
&\quad \big(y_{t\bullet} - [\lambda \odot |z_0| + X_{t\bullet}\beta_\bullet]\big)\Big\} \\
&\quad \times |\Lambda_0|^{-\frac{1}{2}} \exp\big\{(\lambda - \lambda_0)^{\mathsf{T}}\Lambda_0^{-1}(\lambda - \lambda_0)\big\} \\
&\propto \exp\Big\{-\frac{1}{2}\big[\lambda^{\mathsf{T}}(nZ_0^*\Sigma^{-1}Z_0^* + \Lambda_0^{-1})\lambda \\
&\quad - 2\lambda^{\mathsf{T}}\big(\sum_{t=1}^{n} Z_0^*\Sigma^{-1}y_{t\bullet} - \sum_{t=1}^{n} Z_0^*\Sigma^{-1}X_{t\bullet}\beta_\bullet + \Lambda_0^{-1}\lambda_0\big)\big]\Big\} \\
&\propto \exp\Big\{-\frac{1}{2}(\lambda - \widetilde{\lambda})^{\mathsf{T}}\widetilde{\Lambda}^{-1}(\lambda - \widetilde{\lambda})\Big\}. \qquad \text{(A.13)}
\end{aligned}
$$

As a result, the full conditional distribution $\lambda|\big(\beta_\bullet, \Sigma, z_0, D\big)$ is multivariate normal with mean

$$
\widetilde{\lambda} = \widetilde{\Lambda}\big(\sum_{t=1}^{n} Z_0^*\Sigma^{-1}y_{t\bullet} - \sum_{t=1}^{n} Z_0^*\Sigma^{-1}X_{t\bullet}\beta_\bullet + \Lambda_0^{-1}\lambda_0\big)
$$

and covariance

$$
\widetilde{\Lambda} = (nZ_0^*\Sigma^{-1}Z_0^* + \Lambda_0^{-1})^{-1}.
$$

Suppose $\Delta = \mathrm{diag}(\lambda_1, ..., \lambda_g)$, such that $\lambda \odot |z_0| = \Delta|z_0|$. Then, the full conditional distribution $|z_0|$ given $\big(\beta_\bullet, \Sigma, \lambda, D\big)$ is determined as

$$
\begin{aligned}
\pi\Big(|z_0|\;\Big|\Sigma, \beta_\bullet, \lambda, D\Big) &\propto L(y_{t\bullet}|z_0, \Sigma, \lambda, \beta_\bullet)\pi(|z_0|) \\
&\propto |\Sigma|^{-\frac{n}{2}} \exp\Big\{-\frac{1}{2}\sum_{t=1}^{n}\big(y_{t\bullet} - [\Delta|z_0| + X_{t\bullet}\beta_\bullet]\big)^{\mathsf{T}}\Sigma^{-1} \\
&\quad \big(y_{t\bullet} - [\Delta|z_0| + X_{t\bullet}\beta_\bullet]\big)\Big\} \\
&\quad \times \exp\Big\{-\frac{n}{2}|z_0|^{\mathsf{T}}I_g|z_0|\Big\} \\
&\propto \exp\Big\{-\frac{1}{2}\big[|z_0|^{\mathsf{T}}(n\Delta\Sigma^{-1}\Delta + I_g)|z_0| \\
&\quad - 2|z_0|^{\mathsf{T}}\big(\sum_{t=1}^{n}\Delta\Sigma^{-1}y_{t\bullet} - \sum_{t=1}^{n}\Delta\Sigma^{-1}X_{t\bullet}\beta_\bullet\big)\big]\Big\} \\
&\propto \exp\Big\{-\frac{1}{2}(|z_0| - \widetilde{z}_0)^{\mathsf{T}}\Psi_z^{-1}(|z_0| - \widetilde{z}_0)\Big\}. \qquad \text{(A.14)}
\end{aligned}
$$

Consequently the full conditional posterior distribution is $TN(\widetilde{z}_0, \Psi_z, (0, +\infty))$ where

$$
\widetilde{z}_0 = \Psi_z\big(\sum_{t=1}^{n}\Delta\Sigma^{-1}y_{t\bullet} - \sum_{t=1}^{n}\Delta\Sigma^{-1}X_{t\bullet}\beta_\bullet\big)
$$

and

$$
\Psi_z = (n\Delta\Sigma^{-1}\Delta + I_g).
$$

Here, $TN\big(\mu, \Sigma; (a, b)\big)$ stands for the multivariate truncated normal distribution $N(\mu, \Sigma)$ lying within the interval (a, b), $-\infty \le a < b \le +\infty$.

Hadmard Product:

For two matrices, A, B, of the same dimension, $m \times n$ the Hadamard product, $A \odot B$, is a matrix, of the same dimension as the operands, with elements given by $(A \odot B)_{i,j} = (A)_{i,j} \cdot (B)_{i,j}$, writing as [9]

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \odot \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} & a_{12}b_{12} \\ a_{21}b_{21} & a_{22}b_{22} \end{pmatrix}. \tag{A.15}$$

ACKNOWLEDGMENTS

The authors also acknowledge the valuable suggestions from the referees.

REFERENCES

[1] Arellano-Valle, R.B.; Bolfarine, H. and Lachos, V.H. (2007). Bayesian inference for skew-normal linear mixed models, *Journal of Applied Statistics*, **34**, 663–682.

[2] Arellano-Valle, R.B.; Cortes, M.A. and Gomez, H.W. (2010). An extension of the epsilon-skew-normal distribution, *Communications in Statistics Theory and Methods*, **39**, 912–922.

[3] Azzalini, A. (1986). Further results on a class of distribution which includes the normal ones, *Statistica*, **46**, 199–208.

[4] Azzalini, A. and Capitanio, A. (1999). Statistical application of the multivaritae skew-normal distribution, *Journal of the Royal Statistical Society*, *Series B*, **61**, 579–602.

[5] Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution, *Biometrika*, **83**, 715–726.

[6] Azzalini, A. and Regoli, G. (2012). Some properties of skew-symmetric distributions, *Annals of the Institute of Statistical Mathematics*, **64**, 857–879.

[7] Balakrishnan, N. (2002). Discussion of skewed multivariate models related to hidden truncation and/or selective reporting, *Test*, **11**, 37–39.

[8] Box, G.E.P. and Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Cambridge.

[9] Chandler, D. (1962). The norm of the Schur product operation, *Numerische Mathematik*, **4**, 343–344.

[10] 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.

[11] Epple, D. and Bennett, T.M. (2006). Simultaneous equation econometrics: the missing example, *Economic Inquiry*, **44**, 374–384.

[12] Fiebig, D. (2001). *Seemingly unrelated regression.* In: "Companion in Econometrics" (B. Baltagi, Ed.), Basil Blackwell, London.

[13] 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.

[14] Geisser, S. and Eddy, W.F. (1979). A predictive approach to model selection, *Journal of American Statistical Association*, **74**, 153–160.

[15] Gelfand, A.E. and Dey, D.K. (1994). Bayesian model choice: asymptotics and exact calculations, *Journal of Royal Statistical Society*, *Series B*, **56**, 501–514.

[16] Gelman, A. and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences, *Statistical Science*, **4**, 457–472.

[17] Genton, G.G. (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Chapman & Hall, Boca Raton.

[18] Gupta, A.K.; González-Farías, G. and Domínguez-Molina, J.A. (2004). A multivariate skew-normal distribution, *Journal of Multivariate Analysis*, **89**, 181–190.

[19] Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations, *Econometrica*, **11**, 1–12.

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

[21] Judge, G.G.; Griffiths, W.E.; Hill, R.C.; Lutkepohl, H. and Lee, T.C. (1985). *The Theory and Practice of Econometrics* (Second edition), John Wiley and Sons, New York.

[22] Lee, P.M. (2012). *Bayesian Statistics: An Introduction*, John Wiley and Sons, New York.

[23] Newton, M.A. and Raftery, A.E. (1994). Approximate Bayesian inference by the weighted likelihood bootstrap, *Journal of the Royal Statistical Society*, *Series B*, **56**, 3–48.

[24] Percy, D.F. (1992). Predictions for seemingly unrelated regressions, *Journal of the Royal Statistical Society*, *Series B*, **54**, 243–252.

[25] Press, S.J. (1972). *Applied Multivariate Analysis*, John Wiley and Sons, New York.

[26] Puelz, D.; Hahn, P.R. and Carvalho, C. (2017). Variable selection in seemingly unrelated regressions with random predictors, *Bayesian Analysis*, **12**, 969–989 (available at SSRN).

[27] Smith, M. and Kohn, R. (2000). Nonparametric seemingly unrelated regression, *Journal of Econometrics*, **98**, 257–282.

[28] Srivastava, V.K. and Dwivedi, T.D. (1979). Estimation of seemingly unrelated regression equations: a brief survey, *Journal of Econometrics*, **10**, 15–32.

[29] Srivastava, V.K. and Giles, D.E.A. (1987). *Seemingly Unrelated Regression Equations Models*, Dekker, New York.

[30] Theil, H. (1954). Estimation of parameters in econometric models, *Bulletin of the International Statistical Institute*, **34**, 122–129.

[31] Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests of aggregation bias, *Journal of American Statistical Association*, **57**, 500–509.

[32] Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*, John Wiley and Sons, New York.

[33] Zellner, A. and Ando, T. (2010a). Approach for Bayesian analysis of the seemingly unrelated regression mode, *Journal of Econometrics*, **159**, 33–45.

[34] Zellner, A. and Ando, T. (2010b). Bayesian and non-Bayesian analysis of the seemingly unrelated regression model with student-t errors, and its application for forecasting, *International Journal of Forecasting*, **26**, 413–434.

[35] Zellner, A.; Ando, T.; Basturk, N.; Hoogerheide, L. and Van Dijk, H. (2014). Bayesian analysis of instrumental variable models: acceptance-rejection within direct Monte Carlo, *Econometric Reviews*, **33**, 3–35.