



Bayesian Analysis of Regression Models Using Instrumental Variables: A Case Study (Iranian Rural Households Income and Expenditure Data)

Omid Akhgari and Mousa Golalizadeh*

Tarbiat Modares University

Received: 10/9/2016

Approved: 5/3/2017

Abstract. The instrumental variable (IV) regression is a common model in econometrics and other applied disciplines. This model is one of the proper candidate in dealing with endogeneity phenomenon which occurs in analyzing the multivariate regression when the errors are correlated with some covariates. One can consider IV regression as an special case of simultaneous equation models (SEM). There are some cases in which the normality assumption might not hold for the error term in these models and so the skew-normal distribution might be a suitable candidate. The present paper tackle the Bayesian inference based on Markov Chain Monte Carlo (MCMC) using this density for the error term while some instrumental variables are considered in the corresponding regression model. The proposed model is utilized to analysis the Iranian rural households income and expenditure collected in 2009.

Keywords. Instrumental variable; endogenous (exogenous) variable; bayesian inference; skew-normal distribution; Markov chain Monte Carlo logarithm.

MSC 2010: 93E24.

* Corresponding author

1 Introduction

The regression models are strong tools to represent relationship among variables in many applied disciplines. To fit these models using the ordinary least squares or maximum likelihood approaches require particular assumptions. They include: normality of errors, having constant variance for errors, the independency among errors. Apart from these crucial assumptions, it is supposed that the errors are independent from the explanatory variables. There are numerous cases in real life problems that this assumption does not hold. In these situations, the accuracy of the proposed regression model and relevant statistical inference are misleading (Cleveland, 1979). The cases in which the explanatory variables are correlated with errors in the model are known as endogeneity and the corresponding variables are called endogenous (Bowden and Turkington, 1984; Ebbes, 2004). Generally, the endogeneity might be the consequences of the following outcomes:

1. measurement errors
2. autoregressive errors
3. autocorrelation errors
4. lack of simultaneous equation models
5. omitted variables

Wooldridge (2002) believed that first four cases can be an special scenario of fifth. Pearl (2000) showed that in the case of endogeneity, the estimates of the parameters in most of the econometric (particularly microeconomic) problems are inconsistent. As pointed out by Lopes and Polson (2014), the researchers are not only interested on the estimate of the regression parameters but also the directions of those casual effects based upon the variables. Generally, the endogenous explanatory variables lead to the biased estimators and inconsistent estimators derived from the OLS (Ordinary Least Squares) method. To obtain unbiased and consistent estimate, the methods called the IV (Instrumental Variable) regression and Heckman corrected (Davidson and Mackinnon, 1993; Heckman, 1979), among others, has been suggested. However, regression based upon IV is more common than others. Another method, leading to the consistent estimators, was suggested by Imbnes and Angrist (1994), in which the estimators of the instrumental variables are used to predict endogenous variables. Despite its popularity,

the numbers of variables and their types are main objects to establish a proper IV models (Ruud, 2000).

As proposed by Ebbes et al. (2005) the behavioral instrumental variables have the similar role as the explanatory variables, in being uncorrelated with errors. The methods based on instrumental variable usually deliver worse results whenever the correlation between covariates and error is ignored. This, in turn, indicates low quality of the instrumental variables. An important question that arises in many practical problems is: “How can one find the effective instrumental variables for a particular model?”. As argued by Lopes and Polson (2014), generally, there is no clear answer to this question. However, additional data collection may be considered as a solution, but this, inevitably, involves extra costs. Jin and Maheu (2016) believed that choosing the instrumental variables are mostly dependent on their availability. Unfortunately, analytical methods based on instrumental variables is highly sensitive to the such selection and validation variables (Donald and Newey, 2001; Stock et al., 2002). Due to this drawback, the results of regression models using the instrumental variables are very different. Davidson and MacKinnon (1993) studied this issue in more details.

Another issue that may arise on studying the IV regression in real life application is deviation of the response variable from the normal distribution or generally, appearance of asymmetry among the data. One may argue that a set of transformations can be employed to make data relatively symmetric and so invoking the normal density would be possible. But it should be noted that such transformations are not always available. Or, despite availability of the those conversions, the ultimate interpretations of the estimates for parameters and models are not straightforward. One of the possible solutions in these circumstances is to use some asymmetric densities and then follow available statistical inference procedures to estimate the parameters. The skew-normal is one of the important distributions proposed to tackle the asymmetric feature in the density of data. The univariate and multivariate skew-normal distributions were proposed by Azzalini (1986) and Azzalini and Dalla valle (1996), respectively, to treat such cases. Later on, Azzalini and Capitanio (1999) studied further properties of these densities. Several generalizations of these were presented by Balakrishnan (2002), Genton (2004), Gupta et al. (2004) and Arellano-valle et al. (2010). Recently, Azzalini and Regoli (2012) investigated other interesting properties of skew-symmetric distribution.

In present study, we invoke the skew-normal distribution, instead of com-

monly used normal density, for modelling the errors in the IV regression. Derivation of the estimate of parameters in these models through the frequentist inference framework involves solving some complicated equations and hence invoking usual numerical approaches are inevitable. However, the Bayesian inference paradigm has trivial settings and, thanks to availability of many computational tools in this view, one can tackle the statistical inference easily. Accordingly, we confine ourselves to this approach in the current paper. The literature on Bayesian view to this problem has a long list. Dreze (1976) initialized the Bayesian analysis of the IV regression model. For a long time, the literature on Bayesian analysis of the IV regression model was primarily concerned with the numerical computation of integrals. See, for instance, Kloek and van Dijk (1978), Bauwens (1984), Steel (1991), Geweke (1996) and Bauwens and van Dijk (1989), among others. Other more recent papers include Kleibergen and Zivot (2003), Lopes and Polson (2014), Hansen and Kozbur (2014), Kobayashi (2016) and Jin and Maheu (2016).

The remainder of this paper is organized as follows. A brief review of the IV regression models and SEM (Simultaneous Equation Model) are presented in Section 2. Then, a Bayesian approach to treat the IV regression models with normal as well as skew-normal distributions for the errors are given in Sections 3. Analyzing real life data, related to the Iranian households income and expenditure in year 2009, is presented in Section 4. General conclusions are provided at the end.

2 Brief Review on Instrumental Variable Regression

Following Ebbes et al. (2005), the general IV regression model is written as

$$y_i = \beta_0 + \beta_1 Z_{1i} + \cdots + \beta_k Z_{ki} + \gamma Y_i + \epsilon_i, \quad i = 1, \dots, g, \quad (1)$$

where y_i is the dependent (endogenous) variable, ϵ_i is the error term, Z_{ji} for $j = 1, \dots, k$ are k exogenous regressors, which are uncorrelated with ϵ_i and Y_i is endogenous variable, which is correlated with ϵ_i , i.e.

$$\begin{aligned} \text{Cov}(Z_{ji}, \epsilon_i) &= 0, & j = 1, \dots, k, \quad i = 1, \dots, g, \\ \text{Cov}(Y_i, \epsilon_i) &\neq 0. \end{aligned} \quad (2)$$

The endogenous variables can usually be modeled with instrumental variables (say H 's), which are uncorrelated with ϵ_i 's but correlated with the error terms u_i 's. Therefore, these instrumental variables explain part of the variability in the endogenous regressors. This implies that the instrument variables cannot have a direct effect on y_i . In other words, the instrument variables H are exogenous. In terms of notation, we mathematically have

$$\begin{aligned} y_i &= \beta_0 + \beta_1 Z_{1i} + \cdots + \beta_k Z_{ki} + \gamma Y_i + \epsilon_i, \\ Y_i &= \gamma_0 + \gamma_1 H_{1i} + \cdots + \gamma_s H_{si} + u_i, \end{aligned} \quad (3)$$

in which, for $l = 1, \dots, s$, and $i = 1, \dots, g$,

$$\begin{aligned} \text{Cov}(H_{li}, \epsilon_i) &= 0, \\ \text{Cov}(H_{li}, u_i) &= 0. \end{aligned} \quad (4)$$

As can be seen, the model (3) might be an especial case of Simultaneous Equation Model (SEM), a class of useful models comprehensively treated in Hansen and Kozbur (2014). It is worth mentioning that the SEM formula with g equations for i th endogenous variable can be written as

$$y_i = X_i \beta_i + u_i = Z_i \beta_{1i} + Y_i \beta_{2i} + u_i, \quad i = 1, \dots, g, \quad (5)$$

where the error term u_i 's and endogenous variable y_i 's are $n \times 1$ column vectors and X_i 's are the $n \times k_i$ matrices of explanatory variables. Note that X_i 's are partitioned as $\begin{bmatrix} Z_i & Y_i \end{bmatrix}$, i.e. Z_i 's constitute the $n \times k_{1i}$ matrices of exogenous variables and Y_i 's include the $n \times k_{2i}$ matrices of endogenous variables where $k_i = k_{1i} + k_{2i}$. Accordingly, the vector β_i 's are partitioned as $\begin{bmatrix} \beta_{1i} & \beta_{2i} \end{bmatrix}$.

Obviously, to define the model (3) explicitly, some further assumptions are necessary. For $t, s = 1, \dots, n$, and $i, j = 1, \dots, g$, they are given by

$$\begin{aligned} E(u_i u_i^T) &= \sigma_{ii} I_n \\ E(u_{ti} u_{tj}) &= \sigma_{ij} \quad \forall t \quad \& \quad \forall i, j \\ E(u_{ti} u_{sj}) &= 0 \quad \forall t \neq s, \end{aligned} \quad (6)$$

where u_{ti} is t th error in i th equation and σ_{ij} is (i, j) th component of the covariance matrix (Σ) for the error term; called the contemporaneous covariance matrix. See, e.g. Ebbes et al. (2005) for more details.

Let $n \times g$ matrix U can, properly, be filled up with the entities u_{ti} . Then, the $g \times g$ matrix Σ , is derived via the following equality:

$$E(U^T U) = \Sigma. \quad (7)$$

The model (5) can also be written in a single expression via stacking vectors and setting matrix X_i , block-diagonally. More precisely, one can write equalities in (5) in a single new equality as

$$y_{\bullet} = X_{\bullet} \beta_{\bullet} + u_{\bullet}, \quad (8)$$

where y_{\bullet} , u_{\bullet} and β_{\bullet} are the $ng \times 1$ vectors of y_i 's, u_i 's, and β_i 's, respectively. Furthermore, X_{\bullet} is an $ng \times k$ matrix, defined in such way that equalities (5) can be derived accordingly and where $k = \sum_{i=1}^g k_i$.

Interestingly, let $Y = [y_1, \dots, y_g]$ be an $n \times g$ matrix with each column constituting y_i 's. Then, the model (5) can also be rewritten as

$$Y\Gamma = WB + U, \quad (9)$$

where Γ is a $g \times g$ non-singular matrix, B is a $k \times g$ matrix of coefficients for $n \times g$ design matrix W and U ; a $n \times g$ matrix, representing the error term. Note that each equation of model (5) is equivalent to the corresponding column in model (9). Hence, i th equation in model (5) can be written as follows

$$[y_i \ Y_i] \begin{pmatrix} 1 \\ -\beta_{2i} \end{pmatrix} = Z_i \beta_{1i} + u_i, \quad i = 1, \dots, g. \quad (10)$$

Note that the columns of Y_i in the model (10) are the same as the columns of Y in model (9). Also, the columns of Z_i play the same role as columns of W .

Now, let us assume that there are g endogenous variables in the model (9). Then, we can write

$$\begin{aligned} Y &= WB\Gamma^{-1} + U\Gamma^{-1} \\ &= WB\Gamma^{-1} + V, \end{aligned} \quad (11)$$

where $V = U\Gamma^{-1}$. We call this expressions Restricted Reduced Form (RRF). Not that, in contrast to this model, one is occasionally encountered with

Unrestricted Reduced Form (URF) equations, given by

$$Y = W\Pi + V, \quad (12)$$

where Π is unrestricted matrix with dimensions $k \times g$. Now, suppose i th equation in the model (12) is given by the equality $y_i = \omega\pi_i + e_i$. In other words, y_i 's and π_i 's are, in fact, i th columns of matrices Y and Π , respectively. Recalling, the model (11), its i th column can be written as

$$y_i = WB\gamma^i + \nu_i, \quad (13)$$

where γ^i is a g -vector of i th column of Γ^{-1} and ν_i is i th column of V . By stacking the columns of the equation (13), we derive the following equality

$$\begin{aligned} y_{\bullet} &= (I_g \otimes WB)\gamma^{\bullet} + \nu_{\bullet} \\ &= (I_g \otimes W)\pi_{\bullet} + \nu_{\bullet} \\ &= W_{\bullet}\pi_{\bullet} + \nu_{\bullet}, \end{aligned} \quad (14)$$

where γ^{\bullet} is a g^2 -vector after stacking vectors γ^i , ν_{\bullet} is an ng -vector filled with ν_i 's, W_{\bullet} is a $ng \times ng$ matrix and π_{\bullet} is a kg -vector constituting π_i 's. It is seen that π_i 's are the columns of Π where $\Pi = B\Gamma^{-1}$.

Based upon the equalities (14), to derive the distribution y_{\bullet} is equivalent to determine the density of ν_{\bullet} . First, note that we can write

$$\nu_i = U\gamma^i = \sum_{j=1}^g u_j \gamma^{ji}, \quad i = 1, \dots, g \quad (15)$$

where γ^{ji} is j th row and i th columns of Γ^{-1} . By stacking ν_i in (15) vertically, it is straightforward to show that

$$\nu_{\bullet} = ((\Gamma^T)^{-1} \otimes I_n)u_{\bullet}. \quad (16)$$

Using the equalities in (7) and (8), we have $\text{Var}(u_{\bullet}) = \Sigma \otimes I_n$. Hence, the covariance matrix of ν_{\bullet} is given by

$$\begin{aligned}
\text{Var}(\nu_{\bullet}) &= E(\nu_{\bullet}\nu_{\bullet}^T) \\
&= ((\Gamma^T)^{-1} \otimes I_n)(\Sigma \otimes I_n)(\Gamma^{-1} \otimes I_n) \\
&= ((\Gamma^T)^{-1}\Sigma\Gamma^{-1}) \otimes I_n \\
&= \Omega \otimes I_n,
\end{aligned} \tag{17}$$

where $\Omega = (\Gamma^T)^{-1}\Sigma\Gamma^{-1}$.

Now, based upon the information given so far, we can express the density of y_{\bullet} , explicitly. Let ν_{\bullet} follows the multivariate normal density with the mean vector zero (with length ng) and the covariance matrix $\Omega \otimes I_n$. Recalling the model (14), the probability density function of y_{\bullet} , say $f(y_{\bullet})$, is given by

$$f(y_{\bullet}) = C \exp \left\{ -\frac{1}{2}(y_{\bullet} - W_{\bullet}\pi_{\bullet})^T(\Omega^{-1} \otimes I_n)(y_{\bullet} - W_{\bullet}\pi_{\bullet}) \right\}, \tag{18}$$

where the normalizing constant turns to be $C = (2\pi)^{-\frac{ng}{2}} |\Omega \otimes I_n|^{-\frac{1}{2}}$. It can be shown that

$$(y_{\bullet} - W_{\bullet}\pi_{\bullet}) = (\Gamma^T \otimes I_n)^{-1}(y_{\bullet} - X_{\bullet}\beta_{\bullet}). \tag{19}$$

Now, plugging the right hand side of the equality in (19) into the exponent of the density in (18) and then recalling the equality $\Omega = (\Gamma^T)^{-1}\Sigma\Gamma^{-1}$, we are able to write

$$f(y_{\bullet}) = (2\pi)^{-\frac{ng}{2}} |\Sigma|^{-\frac{1}{2}} |\Gamma| \exp \left\{ -\frac{1}{2}(y_{\bullet} - X_{\bullet}\beta_{\bullet})^T(\Sigma^{-1} \otimes I_n)(y_{\bullet} - X_{\bullet}\beta_{\bullet}) \right\},$$

which is used to make statistical inference on the parameters in the case of the multivariate normal density assumption for the error term.

Now, suppose the error term follows the skew-normal distribution. In terms of mathematical notations, we write

$$\nu_{i\bullet} = (\nu_{i1}, \dots, \nu_{ig})^T \sim SN_g(0_g, \Omega, \lambda), \quad i = 1, \dots, n. \tag{20}$$

Following Azzalini and Dalla valle (1996), the density function of $\nu_{i\bullet}$ is given by

$$f_{\nu_{i\bullet}}(\nu_{i\bullet}) = 2\phi_g(0_g, \Omega) \Phi_1(\lambda^T \omega^{-1} \nu_{i\bullet}), \quad i = 1, \dots, n \tag{21}$$

where $\phi_g(0_g, \Omega)$ is the g -dimensional normal density with mean zero-vector and covariance matrix Ω . Here, $\Phi(\cdot)$ is cumulative distribution function of the standard normal, and λ is g -vector of shape parameter. Moreover, ω is a diagonal matrix whose components are the square root of the corresponding diagonal entries of the covariance matrix Ω . To make statistical inference using the response in the IV regression models, one just needs to adopt those components in the normal density case. More details on this issue can be found in Azzalini and Capitanio (1999).

3 Bayesian Inference Based on Normal and Skew-Normal Density

Recall the model (9), indexed with i , as,

$$Y_i \Gamma = W_i B + U_i, \quad i = 1, \dots, n \quad (22)$$

where Y_i is a $1 \times g$ vector, i.e. its columns are i th observations of endogenous variables. Further, Γ is a $g \times g$ matrix, W_i is a $1 \times k$ vector of exogenous variables, B is a $k \times g$ matrix of coefficients and U_i is a $1 \times g$ vector from g errors of simultaneous equations. It is observed that the model has three parameters, set in a vector, say $\theta = (\beta_\bullet, \Sigma, \Gamma)^T$.

By stacking columns of model (22), a modified version of the model is

$$y_{i\bullet} = W_{i\bullet} \pi_\bullet + \nu_{i\bullet}, \quad i = 1, \dots, n \quad (23)$$

where $y_{i\bullet}$ is a g -vector constituting the endogenous variables, $W_{i\bullet}$ is a $g \times gk$ block-diagonal matrix including exogenous variables, π_\bullet is a gk -vector of coefficients and ν_\bullet is a g -vector of the model errors. Considering the multivariate normal distribution for the errors in (23), we can write

$$\nu_{i\bullet} \sim N(0_g, \Omega), \quad i = 1, \dots, n \quad (24)$$

where $\Omega = (\Gamma^T)^{-1} \Sigma \Gamma^{-1}$. Then, $y_{i\bullet}$ will also follow the multivariate normal distribution as

$$y_{i\bullet} \sim N(W_{i\bullet} \pi_\bullet, \Omega), \quad i = 1, \dots, n, \quad (25)$$

where $\Omega = (\Gamma^T)^{-1} \Sigma \Gamma^{-1}$. Mathematically, the probability density function of

$y_{i\bullet}$ is given by

$$\begin{aligned} f(y_{i\bullet}|\pi_{\bullet}, \Sigma, \Gamma) &= |\Omega|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (y_{i\bullet} - W_{i\bullet}\pi_{\bullet})^T \Omega^{-1} (y_{i\bullet} - W_{i\bullet}\pi_{\bullet}) \right\} \\ &= |\Sigma|^{-\frac{1}{2}} |\Gamma| \exp \left\{ -\frac{1}{2} (y_{i\bullet} - X_{i\bullet}\beta_{\bullet})^T \Sigma^{-1} (y_{i\bullet} - X_{i\bullet}\beta_{\bullet}) \right\}. \end{aligned} \quad (26)$$

Here, it is expected that to derive the estimate of parameters through the frequentist inference methodology is equivalent to solve a numbers of complicated non-linear equations and so one should inevitably follow some numerical approaches. Nevertheless, thanks to availability of many computational tools in this view, the Bayesian inference framework sounds to have trivial steps and provide us more flexibility on estimating the parameters.

To follow the Bayesian paradigm, one first needs to specify the prior distributions for the parameters in the model. In the absence of prior information and in order to guarantee reasonable posterior distribution, one can adopt proper but diffuse priors (Liseo and Loperfido, 2006). Then, statistical inferences are made based upon the posterior densities of those parameters. It can be done either directly through sampling from densities or evaluating the posteriors using the full conditional distributions and employing the Markov Chain Monte Carlo (MCMC) sampling procedures. For this latter situation, one might need to invoke the Gibbs or Metropolis-Hastings algorithms. Davidson and MacKinnon (1993) can be consulted for more details on these subjects.

As mentioned already, the parameters of our model are set in θ . Recalling their features, we can consider the following independent priors for the components in θ :

$$\begin{aligned} \beta_{\bullet} &\sim N(\beta_{\circ}, \Sigma_{\beta_{\circ}}) \\ \Sigma &\propto |\Sigma|^{-\frac{g+1}{2}} \\ \Gamma &\sim MN_{g,g}(\Gamma_{\circ}, D, C), \end{aligned} \quad (27)$$

where β_{\circ} is a column k -vector; $\Sigma_{\beta_{\circ}}$ a $k \times k$ matrix and MN stands for the matrix normal distribution with the location parameter Γ_{\circ} (an $g \times g$ matrix), two scale parameters D and C (both are positive definite real $g \times g$ matrices). See, e.g. Genton (2004) for more on the MN , and its interesting properties.

Then, the joint posterior distribution of the parameters comes from the

expression

$$Pr(\beta_{\bullet}, \Sigma, \Gamma | D) \propto L(D | \beta_{\bullet}, \Sigma, \Gamma) \pi_1(\beta_{\bullet}) \pi_2(\Sigma) \pi_3(\Gamma). \quad (28)$$

Since the joint posterior distribution does not have a closed form here, one should derive the full conditional distributions of each component in order to make any statistical inference about the parameters. They are given below. First, the full conditional distribution of β_{\bullet} is given by

$$\beta_{\bullet} | (\Sigma, \Gamma, D) \sim N(\tilde{\beta}_{\bullet}, \tilde{\Sigma}), \quad (29)$$

where

$$\begin{aligned} \tilde{\beta}_{\bullet} &= \tilde{\Sigma} \left(\sum_{i=1}^n (X_{i\bullet}^T \Sigma^{-1} y_{i\bullet} + \Sigma_{\beta_o}^{-1} \beta_o) \right) \\ \tilde{\Sigma} &= \left(\sum_{i=1}^n (X_{i\bullet}^T \Sigma^{-1} X_{i\bullet}) + \Sigma_{\beta_o}^{-1} \right)^{-1}. \end{aligned} \quad (30)$$

Second, the full conditional distribution of Σ turns out to be

$$\Sigma | (\beta_{\bullet}, \Gamma, D) \sim IW(n, R) \quad (31)$$

where

$$R = \sum_{i=1}^n (y_{i\bullet} - X_{i\bullet} \beta_{\bullet})(y_{i\bullet} - X_{i\bullet} \beta_{\bullet})^T. \quad (32)$$

Finally, the full conditional distribution of Γ is expressed as

$$Pr(\Gamma | (\beta_{\bullet}, \Sigma, D)) \propto L(D | \beta_{\bullet}, \Sigma, \Gamma) \pi_3(\Gamma). \quad (33)$$

As expected, the full conditional distribution of Γ does not have a closed form. Clearly, one can invoke Metropolis-Hastings algorithm to sample from it and make relevant statistical inference. We shall see this in next section.

Now, we can set up the Bayesian inference when the error term follows the skew-normal distribution. According to Arellano-Valle et al. (2010) and based upon stochastic representation for the skew-normal random variable, we can write $y_{i\bullet} = \lambda \odot |z_0| + z_1$ where \odot denotes hadamard product, $z_0 \sim N(0_g, I_g)$, $z_1 \sim N(X_{i\bullet} \beta_{\bullet}, \Sigma)$ and λ is the shape parameter. Moreover, it is

assumed that z_0 and z_1 are independent. Suppose the components of $\theta_2 = (\beta_\bullet, \Sigma, \Gamma, \lambda, z_0)^T$ are independent a priori, and let us assume the following prior distributions for each component:

$$\begin{aligned}\beta_\bullet &\sim N(\beta_\circ, \Sigma_{\beta_\circ}), \quad \Sigma \propto |\Sigma|^{-\frac{g+1}{2}}, \\ \lambda &\sim N(\lambda_0, \Lambda_0), \quad \Gamma \sim MN_{g,g}(\Gamma_\circ, D, C).\end{aligned}\quad (34)$$

Here, the join posterior of all parameters also doesn't have a closed form. So we cannot compute the join posterior analytically. To turn around this problem, we benefit from the stochastic representation of the skew-normal distribution. Then, there is a hope that we could drive the full conditional distributions more easily. Below, we provide them in turn. It worths mentioning that the order they appear is exactly the same as the corresponding stages on updating the parameters in an MCMC sampling algorithm.

First, we have $\beta_\bullet | (\Sigma, \lambda, |z_0|, \Gamma, D) \sim N(\beta_*, \Sigma_{\beta_*})$ where

$$\begin{aligned}\Sigma_{\beta_*} &= \left(\sum_{i=1}^n X_{i\bullet}^T \Sigma^{-1} X_{i\bullet} + \Sigma_{\beta_\circ}^{-1} \right) \\ \beta_* &= \Sigma_{\beta_*} \left(\sum_{i=1}^n X_{i\bullet}^T \Sigma^{-1} y_{i\bullet} - X_{i\bullet}^T \Sigma^{-1} \Gamma^{-1} (\lambda \odot |z_0|) + \Sigma_{\beta_\circ}^{-1} \beta_\circ \right).\end{aligned}\quad (35)$$

At the second stage, we update Σ using

$$\Sigma | (\beta_\bullet, \lambda, \Gamma, |z_0|, D) \sim IW(R, n), \quad (36)$$

where

$$R = \sum_{i=1}^n \Gamma^T (y_{i\bullet} - [\lambda \odot |z_0| + W_{i\bullet} \pi_\bullet]) (y_{i\bullet} - [\lambda \odot |z_0| + W_{i\bullet} \pi_\bullet])^T \Gamma.$$

Next, the full conditional distributions of λ is given by

$$\lambda | (\beta_\bullet, \Sigma, \Gamma, |z_0|, D) \sim N(\lambda_*, \Sigma_{\lambda_*}), \quad (37)$$

where $\lambda_* = \Sigma_{\lambda_*} (Z^* \Gamma \Sigma^{-1} \sum_{i=1}^n (y_{i\bullet} - X_{i\bullet} \beta_\bullet) + \Sigma_{\lambda_\circ}^{-1} \lambda_\circ)$ and Z^* is an $g \times g$ diagonal matrix whose components are filled with the elements of vector $|z_0|$ and $\Sigma_{\lambda_*} = (\Sigma_{\lambda_\circ}^{-1} + n Z^* \Gamma \Sigma^{-1} \Gamma^T Z^*)^{-1}$.

The fourth step involves driving the density of $|z_0|$. It is given by

$$|z_0| \left| \left(\beta_{\bullet}, \Sigma, \lambda, D \right) \right. \sim TN_g \left(T_*, \Sigma_{T_*}, (0, +\infty) \right), \quad (38)$$

where $TN_g(\mu, \Sigma; (a, b))$ stands for the multivariate truncated normal distribution $N_g(\mu, \Sigma)$ lying within the interval (a, b) , $-\infty \leq a < b \leq +\infty$. Also $T_* = \Sigma_{T_*} (Z^* \Gamma \Sigma^{-1} \sum_{i=1}^n (y_{i\bullet} - X_{i\bullet} \beta_{\bullet}))$ and $\Sigma_{T_*} = (I_g + n \Lambda \Gamma \Sigma^{-1} \Gamma^T \Lambda)^{-1}$ where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_g)$.

Finally, the full conditional density of Γ is derived using the expression

$$Pr(\Gamma | \beta_{\bullet}, \Sigma, \lambda, z_0, D) \propto L(y_{i\bullet} | z_0, \pi_{\bullet}, \Omega, \lambda) \pi_3(\Gamma). \quad (39)$$

Similar to the case of normal assumption for the errors, there is not a closed form for the full conditional density of Γ . Hence, one should follow the Metropolis-Hastings algorithm to draw sample from it.

We are interested on applying the models and procedure discussed in this paper in a real life data. For this purpose, we considered the Iranian rural households income and expenditure data collected in year 2009. More details along with initial analysis and making inference on the parameters are given in the next section.

4 Real Application

A brief description of Iranian **rural household income and expenditure** survey would be nice to mention here. Our data includes about 13345 rural households from 32 provinces. Here, the main goal is on analyzing the effects of some variables on gross expenditure and income. In this study, both of these are considered as endogenous variables and other covariates are set as exogenous. A brief description of the considered variables are provided in Table 1.

As an initial analysis, we are interested on checking the normality assumption for the endogenous variables. Although they are going to be considered as bivariate, we used the Kolmogorov-Smirnov (KS) test for each one separately. As shall be seen, this conducts us to consider alternative density rather than normal. Based upon applying KS test for each variable, normality assumption has not been confirmed since we get P-value < 0.05 in both cases. Hence, one cannot consider the bivariate normal distribution for these two endogenous variables. As a geometric visualization tool, we

Table 1. Brief description of variables utilized in the model (40).

| Variable names | Abbreviation signs | Variable Type | Codes |
|--|--------------------|---------------|-------------------|
| Expenditure | GH | Quantitative | - |
| Households income | D | Quantitative | - |
| Family size | C_1 | Quantitative | - |
| Number of literate in household | C_2 | Quantitative | - |
| Number of employees in household | C_3 | Quantitative | - |
| Number of people with income | C_4 | Quantitative | - |
| Age | A | Quantitative | - |
| Floor 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 |
| Agriculture self-employment income | D_1 | Quantitative | - |
| Non-agriculture self-employment income | D_2 | Quantitative | - |
| Miscellaneous income | D_3 | Quantitative | - |
| Non-monetary other Incomes | D_4 | Quantitative | - |

provided the Figure 1, including the contour plot as well as the Q-Q plot of the income and gross expenditure. They are all confirming that densities of our endogenous variables deviate from the normal distribution. Although the normal density assumption has been violated, we keep considering it for our subsequent analysis at least to make a basis for comparison with the alternative density, precisely the skew-normal distribution.

Based upon a general view, the following IV regression model was utilized to express the inter-relationship between the income and gross expenditure variables:

$$\begin{aligned}
 GH &= \beta_{0_{GH}} + \sum_{i=1}^4 \beta_{C_i} C_i + \sum_{i=1}^5 \beta_{B_i} B_i + \beta_A A + \gamma D + \epsilon_1 \\
 D &= \beta_{0_D} + \sum_{i=1}^4 \beta_{D_i} D_i + \epsilon_2,
 \end{aligned} \tag{40}$$

where GH and D are endogenous while other are exogenous variables.

We are interested on making Bayesian inference on the parameters appeared in the model (40) to fit this real data. To do this, we used the OpenBUGS software for all computational tasks required in this paper. For all

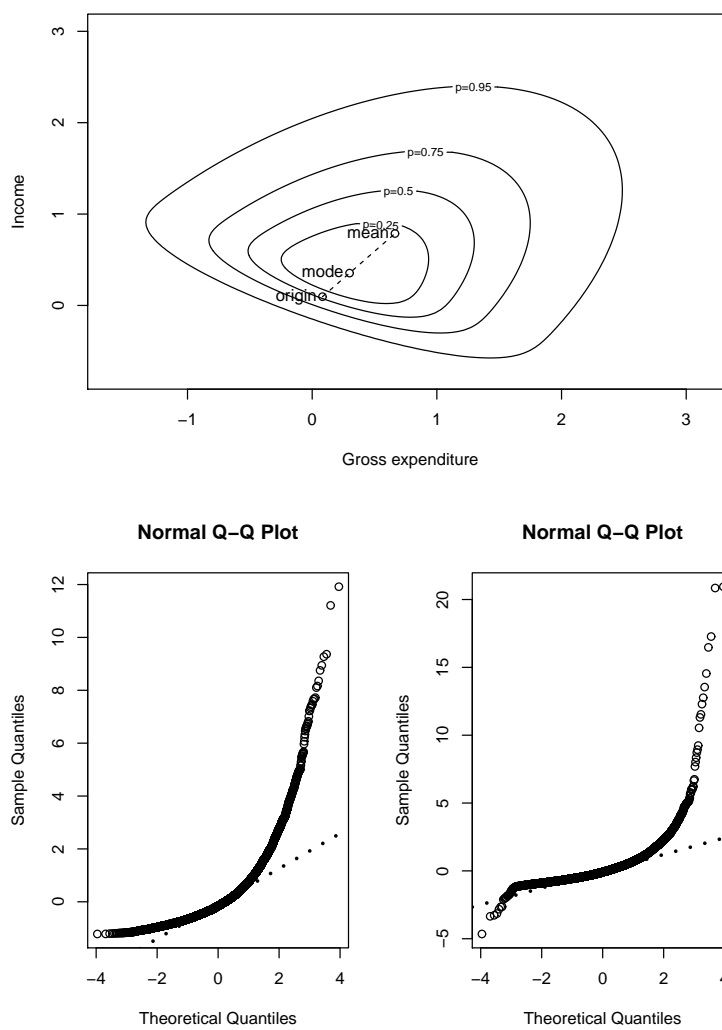


Figure 1. The contour plot of income and gross expenditure along with the Q-Q plot for the each variable in the model (40).

scenarios, we fixed the number of MCMC simulation on 100000 runs, with a burn-in 50000, and choosing each 50th generated sample after burn-in, leading to 1000 samples.

The results on fitting the model (40) are reported in Table 2. It includes two panels. The left and right panels are providing the results based upon normal and skew-normal distributions, abbreviated as N and SN, respectively, for the errors, respectively. Furthermore, in each panel, the results, from top to bottom, are the estimates of the parameters for the first and second expressions in the models (40). We should emphasize that the covariance matrix for the error term constitutes the components σ_{11} , σ_{12} and σ_{22} . Moreover, the shape parameter vector λ (for the skew-normal case) has two parameters namely λ_1 and λ_2 . Hence, their estimates are given at the bottom of the panels.

According to the first panel of Table 2, the family size, number of literate and employees in household have direct effect on the total family expenditures. However, the number of people in the household having income, regardless to which resources they earn, has a negative effect on total family expenditures. This also happens for the age and infrastructure area. As expected, the usage of luxury facilities i.e. private car, internet and mobile as well as the income have direct effect on the family total expenditures.

Based on the second part of Table 2, the agriculture self-employment, non-agriculture self-employment, miscellaneous and non-monetary other incomes have direct effect on the household total income. Some interesting remarks from Table 2 can also be overlooked. The values for both intercepts in the model (40) coming from skew-normal assumption is more than those corresponding values from normality case.

Let suppose that one is going to select the better model among two competing distributions normal and skew-normal for errors in the model (40). The common Bayesian criterion to use is Bayes Factor (BF). In addition, the logarithm of pseudomarginal likelihood (LPML) is another valuable criterion in this case. It is derived from predictive considerations, particularly Conditional Predictive Ordinate (CPO), and leads to pseudo Bayes factors for choosing an optimal models (Geisser and Eddy, 1979). As pointed out by Gelfand and Dey (1994), it is popular mainly due in part to its relative ease of computation making the LPML an stable estimate from the MCMC outputs. Both these criteria applied for the model (40) in assuming both normal and skew-normal distributions for the error term are reported in Table 3.

As seen, the value of LPML for skew-normal is greater than corresponding

Table 2. The result of fitting the IV regression model (40) based upon the skew-normal and normal distributions assumption.

| Parameter | N-MCMC | | SN-MCMC | |
|------------------|--------|--------|---------|-------|
| | Est | Sd | Est | Sd |
| $\beta_{0_{GH}}$ | -1.042 | 0.030 | -1.448 | 0.019 |
| β_{C_1} | 0.039 | 0.005 | 0.023 | 0.003 |
| β_{C_2} | 0.046 | 0.006 | 0.043 | 0.004 |
| β_{C_3} | 0.043 | 0.007 | 0.040 | 0.004 |
| β_{C_4} | -0.019 | 0.009 | 0.020 | 0.005 |
| β_{B_1} | 0.003 | 0.001 | 0.001 | 0.001 |
| β_{B_2} | 0.465 | 0.015 | 0.286 | 0.009 |
| β_{B_3} | 0.494 | 0.032 | 0.284 | 0.017 |
| β_{B_4} | -0.008 | 0.011 | 0.040 | 0.007 |
| β_{B_5} | 0.199 | 0.015 | 0.173 | 0.009 |
| β_A | 0.004 | 0.0003 | 0.001 | 0.001 |
| γ_D | 0.279 | 0.007 | 0.094 | 0.003 |
| β_{0_D} | 0.000 | 0.004 | -0.515 | 0.001 |
| β_{D_1} | 0.522 | 0.004 | 0.562 | 0.001 |
| β_{D_2} | 0.488 | 0.004 | 0.501 | 0.002 |
| β_{D_3} | 0.374 | 0.004 | 0.377 | 0.001 |
| β_{D_4} | 0.031 | 0.004 | 0.030 | 0.002 |
| σ_{11} | 0.643 | 0.007 | 0.020 | 0.001 |
| σ_{12} | 0.141 | 0.004 | 0.001 | 0.001 |
| σ_{22} | 0.315 | 0.003 | 0.004 | 0.002 |
| λ_1 | - | - | 1.169 | 0.007 |
| λ_2 | - | - | 0.763 | 0.003 |

Table 3. The performance of competing models according to the criteria BF and LPML.

| Model | BF | LPML |
|-------------|-------|-----------|
| Normal | 0.872 | -44856.07 |
| Skew-Normal | 1.148 | -20851.6 |

value for normal. Moreover, the ratio of BF for the skew-normal in comparison to the normal model is greater than one; indicating superiority of the former one. As a general results, we can claim that the model (40) in which the endogenous variables are distributed as skew-normal fits the data better than normal density. The interested reads might prefer to consider other densities such as T-student or skewed-t for further investigations. However, their mathematical computations need to be full developed. We left them for our future research.

In order to check the correlation among the parameters using the MCMC outputs, we plotted the posterior distribution for some of the bivariate parameters. It has been done along with the common quantitative criteria (not reported here). Parts of the graph are provided in the Figure 2. As seen, they are performing differently in terms of correlation among MCMC samples. However, as a general results, we can assert that the parameters are independent.

Furthermore, we are interested on performing a sensitivity analysis for some parameters in the model. To do this, we changed the variance of the prior of β_i 's for $i = 1, \dots, 16$, using the $N(0, 10^k)$ for $k = 1, \dots, 6$. To keep the paper in standard size, we just reported those for the six parameters chosen randomly. The procedure to draw the MCMC samples was the same as explained earlier in this section. The Figure 3 shows the results of the estimates for those parameters versus the index for the level of variance for the prior densities. As seen, there are promising convergence for almost all parameters for $k > 3$. So, we can claim that our proposed estimators are not sensitive to change the variance of the prior densities. Obviously, taking into account other prior densities and follow the sensitivity analysis needs further investigations. We aim to do it in our future research.

5 Conclusion

We described the statistical inference about the parameters in considering both normal and skew-normal distributions for the error in the IV regression models in this paper. We only paid attention to the Bayesian inference here. As an application, the IV regression model was fitted on Iranian rural households income and expenditure data collected in 2009. In order to select the best model, we considered common model selection criteria such as BF and LPML. Based upon our analysis in this paper, when the endogenous variables are not symmetric in nature, it is recommended to consider some skew

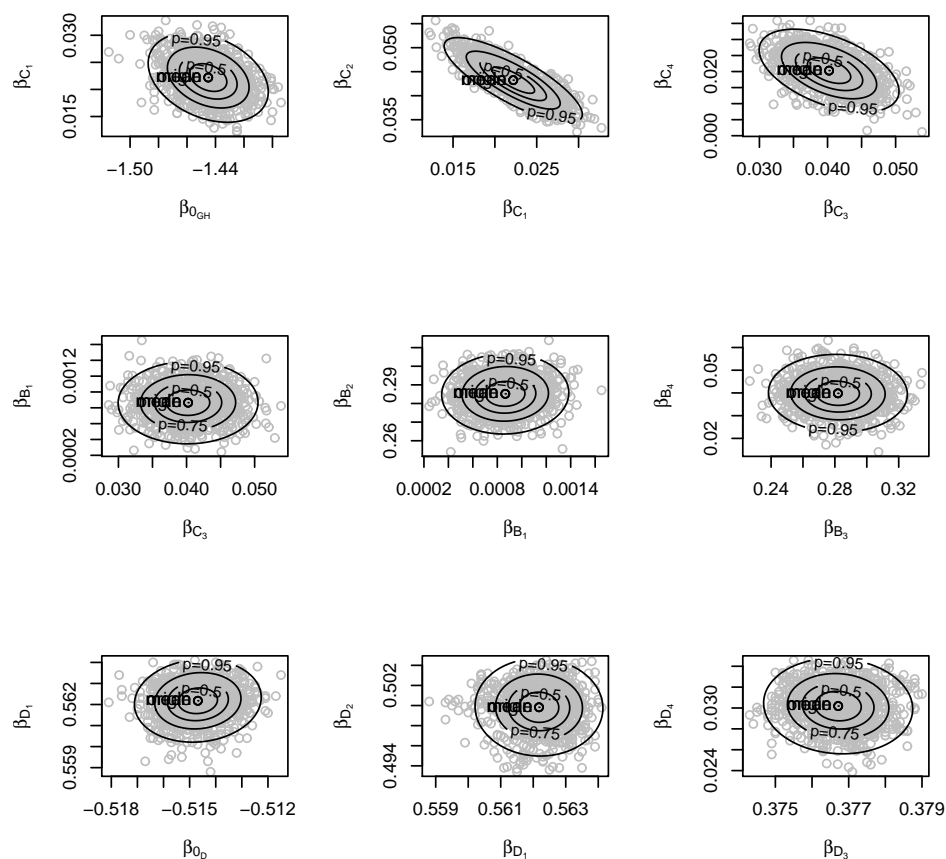


Figure 2. The empirical posterior distributions for some bivariate parameters in the proposed model (40).

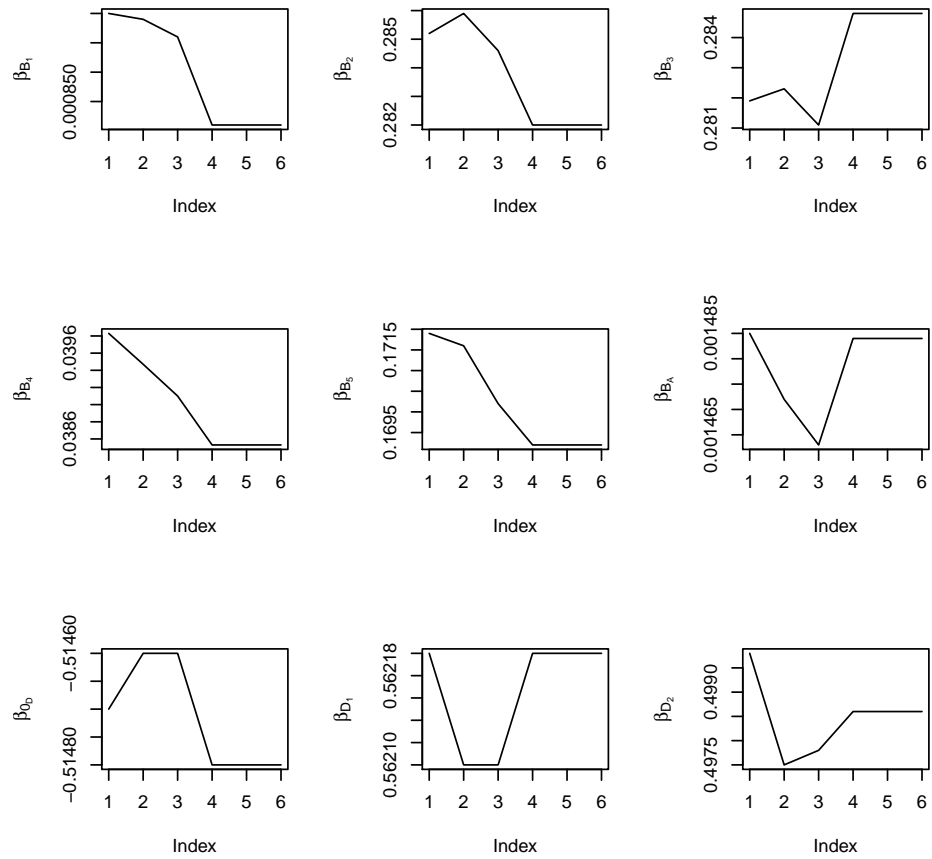


Figure 3. Assessing the sensitivity of estimators for some parameters of the proposed model (40) according to difference prior densities **based on skew-normal assumption for $k=1, \dots, 6$.**

distributions. We showed that the skew-normal distribution is performing better than normal while using the IV regression model for the Iranian rural households income and expenditure data in 2009.

In future study, we aim to develop proper multilevel models when there exist some endogenous variables among data. In particular, we are going to investigate invoking the instrumental variables in a multilevel model where the errors follow the skew-normal distributions. Moreover, evaluating sensitivity analysis for those model while following the Bayesian methodology is worth to study.

References

- 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.
- Azzalini, A. (1986). Further Results on a Class of Distributions Which Includes the Normal Ones. *Statistica*, **46**, 199-208.
- Azzalini, A. and Dalla Valle, A. (1996). The Multivariate Skew-Normal Distribution. *Biometrika*, **83**, 715-726.
- Azzalini, A. and Capitanio A. (1999). Statistical Application of the Multivariate Skew-Normal Distribution., *Journal of the Royal Statistical Society B*, **61**, 579-602.
- Azzalini, A. and Regoli, G. (2012). Some Properties of Skew-Symmetric Distributions, *Annals of the Institute of Statistical Mathematics*, **64**, 857-879.
- Balakrishnan, N. (2002). Discussion of Skewed Multivariate Models Related to Hidden Truncation and/or Selective Reporting. *Test*, **11**, 37-39.
- Bauwens, L. (1984). *Bayesian Full Information Analysis of the Simultaneous Equation Models Using Monte-Carlo Integration*. Springer-Verlag, Berlin.
- Bauwens, L. and Van Dijk, H.K. (1989). *Bayesian Limited Information Analysis*. In J.J. Gabszewicz, et. al., editor, *Economic Decision-Making: Games, Econometrics and Optimisation*. North-Holland Publishing Co., Amsterdam.
- Bowden, S. and Turkington, D.A. (1984). *Instrumental Variables*. Oxford University Press, New York.
- Cleveland, W.S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, **74**, 829-836.
- Davidson, R. and MacKinnon, J.G. (1993). *Estimation and Inference in Econometrics*. Oxford University Press, New York.

- Donald, S.G. and Newey, W.K. (2001). Choosing the Number of Instruments. *Econometrica*, **69**, 1161-1169.
- Dreze, J.H. (1976). Bayesian Limited Information Analysis of the Simultaneous Equations Model. *Econometrica*, **44**, 1045-1075.
- Ebbes, P. (2004). *Latent Instrumental Variables: A New Approach to Solve for Endogeneity*. Ph.D. Thesis, University of Groningen.
- Ebbes, P., Wedel, M., Baockenholt, U. and Steerneman, A.G.M. (2005). Solving and Testing for Regressor-Error (in) Dependence When no Instrumental Variables are Available: With New Evidence for the effect of Education on Income. *Quantitative Marketing and Econometrics*, **3**, 365-392.
- Geisser, S. and Eddy, W.F. (1979). A Predictive Approach to Model Selection. *Journal of American Statistical Association*, **74**, 153-160.
- 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.
- Genton, G.G. (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. Chapman & Hall/CRC, Boca Raton.
- Gupta, A.K., Gonzalez-faraz, G. and Domnguez-molina, J.A. (2004). A Multivariate Skew-Normal Distribution. *Journal of Multivariate Analysis*, **89**, 181-190.
- Geweke, J. (1996). Bayesian Reduced Rank Regression in Econometrics. *Journal of Econometrics*, **75**, 121-146.
- Hansen, C. and Kozbur, D. (2014). Instrumental Variables Estimation with Many Weak Instruments Using Regularized Jackknife Instrumental Variables Estimator (Jive). *Journal of Econometrics*, **182**, 290-308.
- Heckman, J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, **47**, 153-161.
- Imbens, G. and Angrist, J. (1994). Identification and Estimation of Local Average Treatment Effect. *Econometrica*, **62**, 467-476.
- Jin, X. and Maheu, J.M. (2016). Bayesian Semiparametric Modeling of Realized Covariance Matrices. *Journal of Econometrics*, **192**, 19-39.
- Kleibergen, F. and Zivot, E. (2003). Bayesian and Classical Approaches to Instrumental Variable Regression. *Journal of Econometrics*, **114**, 2-72.
- Kloek, T. and Van Dijk, H.K. (1978). Bayesian Estimates of Equation System Parameters : An Application of Integration by Monte-Carlo. *Econometrica*, **44**, 345-351.

- Kobayashi, G. (2016). Bayesian Endogenous Tobit Quantile Regression. *Bayesian Analysis*, DOI:10.1214/16-BA996.
- Liseo, B. and Loperfido, N. (2006). A Note on Reference Priors for the Scalar Skew-Normal Distribution. *Journal of Statistical Planning and Inference*, **136**, 373-389.
- Lopes, H.F. and Polson, N.G. (2014). Bayesian Instrumental Variables: Priors and Likelihoods. *Econometric Reviews*, **33**, 100-121.
- Pearl, J. (2000). *Causality*. Cambridge University Press, New York.
- Ruud, P.A. (2000). *An Introduction to Econometric Theory*. Oxford University Press, New York.
- Steel, M.F.J. (1991). A Bayesian Analysis of Simultaneous Equation Models by Combining Recursive Analytical and Numerical Approaches. *Journal of Econometrics*, **48**, 335-351.
- Stock, J.H., Wright, J.H. and Yogo, M. (2002). A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business and Economic Statistics*, **20**, 518-529.
- Wooldridge, J.M. (2002). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge.

Omid Akhgari

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

Mousa Golalizadeh

Department of Statistics,
Faculty of Mathematical Sciences,
Tarbiat Modares University,
Tehran, Iran.
email: golalizadeh@modares.ac.ir

