Journal of Statistical Theory and Applications
Vol. 20(1), March 2021, pp. 97–110
DOI: https://doi.org/10.2991/jsta.d.210126.002; ISSN 1538-7887
https://www.atlantis-press.com/journals/jsta

On Seemingly Unrelated Regression Model with Skew Error

Omid Akhgari[1], Mousa Golalizadeh[2,*]

[1]Department of Statistics, Amin University, Tehran, Iran
[2]Department of Statistics, Tarbiat Modares University, Tehran, Iran

ARTICLE INFO

Article History
Received 07 Oct 2018
Accepted 05 Jan 2021

Keywords
Seemingly unrelated regression
Endogenous variable
Exogenous variable
Skew-normal distribution

ABSTRACT

Sometimes, invoking a single causal relationship to explain dependency between variables might not be appropriate particularly in some economic problems. Instead, two jointly related equations, where one of the explanatory variables is endogenous, can represent the actual inheritance inter-relationship among variables. Such typical models are called simultaneous equation models of which the seemingly unrelated regression (SUR) models is a special case. Substantial progress has been made regarding the statistical inference on estimating the parameters of these models in which errors follow a normal distribution. But, less research was devoted to a case that the distributions of the errors are asymmetric. In this paper, statistical inference on the parameters for the SUR models, assuming the skew-normal density for errors, is tackled. Moreover, the results of the study are compared with those of other naive methodologies. The proposed model is utilized to analyze the income and expenditure of Iranian rural households in the year 2009.

1. INTRODUCTION

Most linear regression models rely on the relationship between a dependent variable to one or more explanatory variables. The main objective in treating these models is estimating and predicting the average value of dependent variables subject to some explanatory variables. But in many cases, particularly in some economic problems, the causal relationship represented by a single equation is not appropriate. The drawback of such single models is twofold. Mainly, not only does the response variable depends on the explanatory variables, but the response variable also determines some of the explanatory variables. Generally, it can be argued that there are simultaneous or two-sided relationships between the response and some of the explanatory variables in these cases. Hence, to separate the variables as explanatory and dependent does not make sense in real-life circumstances. In these situations, the number of equations will, naturally, be more than one. Precisely, there is an equation for every endogenous or dependent variable. Generally, following Haavelmo [1] when the dependent variable of a particular model is an explanatory variable, one should use the simultaneous equations models (SEMs). The particular case of these models is called the seemingly unrelated regression (SUR) model.

Evidence shows that Zellner [2] was the pioneer researcher to estimate the parameters of the SUR model using the generalized least square method. The history of the frequent approach to such models was somewhat low. But, there were much research on following the Bayesian approach. The application of the Bayesian approach in the SUR model was first proposed by Zellner [3]. Afterward, other methods for estimating parameters were used, including the maximum likelihood method [4], Bayesian moment and direct Monte Carlo method [5].The MCMC application in the SUR model has appeared in many studies under various assumptions. To name some we can mention, for example, Percy [6], Chib and Greenberg [7], and Smith and Kohn [8]. Recently, Zellner and Ando [5] and also Zellner *et al.* [9] have investigated the estimation of the parameters in the SUR model using a hierarchical Bayes approach through the direct Monte Carlo and importance sampling techniques.

Another important aspect of the SUR models, which was and is worth to study, refers to the type of distribution considered for the error term. It is quite common to assume the normal density for this case. But, there are numerous examples in which the empirical distribution of variables often exhibits asymmetric structure and so the normal distribution can no longer be used in these cases. In these situations, some transformations may be used to make the distribution of data to, relatively, follow normal density. However, such transformations have their own drawbacks, including the biase of the estimator [10].Using asymmetric distributions possessing the same characteristics as normal distribution, has recently received significant attention in the literature. The skew-normal distribution is one of the important distributions proposed to tackle the asymmetric feature of data. Historically, the univariate skew-normal distribution was advocated by Azzalini [11]. Then, Azzalini and Dalla valle [12] proposed the multivariate skew-normal distribution. Azzalini and Capitano [13] further studied the properties of this density. Several generalizations of this distribution have been presented by Balakrishnan [14], Genton [15], Gupta

*Corresponding author. Email: golalizadeh@modares.ac.ir

et al. [16], and Arellanovalle *et al.* [17]. Recently, Azzalini and Regoli [18] have investigated some other properties of the skew-symmetric distribution. As a new line of research, we consider the SUR model allowing the error in the model to follow the skew-normal distribution. The estimation of parameters using the maximum likelihood methodology is also treated. Intensive simulation studies are conducted to evaluate the proposed methods. Application of the model to real-life data is also given.

The present paper is organized as follows: A brief review of the SUR model is presented in Section 2. Then, a likelihood-based approach to estimate the parameters with the skew distribution for the errors in the SUR models is discussed in Section 3. The simulation study as well as the analysis of the real data, related to the Iranian rural households income and expenditure on in the year 2009, are presented in Section 4. General conclusions are provided at the end. The proofs for some of the results are given in the Appendix.

2. SUR MODEL

Suppose X_t is an $n \times k_t$ matrix of explanatory variables and β_t a column vector of parameters with the length k_t. Furthermore, suppose there are g equations corresponding with g endogenous variables, a column vector with the length n, indicated by $y_1, \ldots, y_g$. Hence, the t-th equation of a linear simultaneous system can be written as

$$y_t = X_t\beta_t + u_t, \quad t = 1, \ldots, g, \tag{2.1}$$

where

$$\begin{aligned} &E(u_{ti}) = 0 \quad Var(u_{ti}) = \sigma_{tt} \\ &Cov(u_{ti}, u_{si}) = \sigma_{ts}, \quad t, s = 1, 2, ..., g \quad i = 1, 2, ..., n. \end{aligned} \tag{2.2}$$

Let us assume that, g-vectors $y_{i\bullet}$ and $u_{i\bullet}$ consist of y_{ti} and u_{ti}, respectively, stacked vertically for fixed t. Accordingly, the k-vector $\beta_\bullet$ is formed by stacking β_i vertically. Then, the matrix $X_{t\bullet}$ will be of dimension $g \times k$, where $k = \sum_{t=1}^{g} k_t$. In fact, it is a block-diagonal matrix with diagonal blocks X_{ti} also for fixed t with rank $1 \times k_t$. In short, the notations can be summarized as follows:

$$y_{i\bullet} = \begin{pmatrix} y_{1i} \\ \vdots \\ y_{gi} \end{pmatrix}_{g\times 1} \quad u_{i\bullet} = \begin{pmatrix} u_{1i} \\ \vdots \\ u_{gi} \end{pmatrix}_{g\times 1} \quad \beta_\bullet = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}_{k\times 1} \quad X_{i\bullet} = \begin{pmatrix} X_{1i} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & X_{gi} \end{pmatrix}_{g\times k}. \tag{2.3}$$

Based on this notation model (2.1) can be rewritten as

$$y_{i\bullet} = X_{i\bullet}\beta_\bullet + u_{i\bullet}, \quad i = 1, \ldots, n, \tag{2.4}$$

Note that as a common assumption, we now consider $u_{i\bullet} \sim N(0_g, \Sigma)$ where $\Sigma = \left\{\sigma_{ts}\right\}_{g\times g}$.

In the present study, we aim to estimate the parameters of this SUR model. This can be achieved via many parametric and nonparametric estimating procedures including 2SLS[1], 3SLS[2], GMM[3], LIML[4] and FIML,[5] Anderson and Rubin [19], Theil [20], and Davidson and Mackinnon [21]. In this paper we focus on FIML according to normal and skew-normal errors assumption. Moreover, a number of important statistical features pertaining to these models are provided.

Based upon the information provided so far, we can write down the likelihood function to estimate the parameters. As is common, it is preferred to use the logarithm of the likelihood, in which we write it as $l(\beta_\bullet, \Sigma)$, in our problem. It is given by

$$l(\beta_\bullet, \Sigma) = -\frac{ng}{2}\log 2\pi - \frac{n}{2}\log|\Sigma| - \frac{1}{2}\sum_{t=1}^{n}[(y_{t\bullet} - X_{t\bullet}\beta_\bullet)^T\Sigma^{-1}(y_{t\bullet} - X_{t\bullet}\beta_\bullet)], \tag{2.5}$$

[1] Two-stage Least Square

[2] Three-stage Least Square

[3] Generalized Method of Moments

[4] Limited Information Maximum Likelihood

[5] Fully Information Maximum Likelihood

and should be maximized to obtain the FIML estimators. It is quite straightforward to show (see, e.g. Anderson and Rubin [19]) that the maximum likelihood estimators of the parameters are given by

$$\begin{aligned}\hat{\beta}_{\bullet} &= \left[\sum_{t=1}^{n} X_{t\bullet}^{T}\Sigma^{-1}X_{t\bullet}\right]^{-1}\left[\sum_{t=1}^{n} X'_{t\bullet}\Sigma^{-1}y_{t\bullet}\right]\\ \hat{\Sigma} &= \tfrac{1}{n}\sum_{t=1}^{n}[(y_{t\bullet}-X_{t\bullet}\beta_{\bullet})(y_{t\bullet}-X_{t\bullet}\beta_{\bullet})^{T}].\end{aligned} \tag{2.6}$$

Moreover, via invoking a simple computation, it can be shown that

$$Var(\hat{\beta}_{\bullet}) = \left[\sum_{t=1}^{n}, X_{t\bullet}^{T}, \Sigma^{-1}, X_{t\bullet}\right]^{-1}. \tag{2.7}$$

So far, the estimators have been calculated based on the assumption of normality for the error. However, if the distribution of errors is asymmetric, such as specifically skew-normal then to obtain the estimators are not as trivial as seen above. To treat this, we first briefly review the skew-normal distribution in the subsequent section. Then, the FIML estimators of the parameters are obtained under such assumption, while the model includes endogenous variables.

3. SUR MODELED WITH SKEW-NORMAL DISTRIBUTION

We first recall the definition and a few key properties of the skew-normal distribution, as given by Azzalini and Dalla Valle [12]. Suppose Z is a k-dimensional random variable, then it follows the multivariate skew-normal distribution if it is continuous with density function

$$2\phi_k(z;\Psi)\Phi(\lambda^T z), (z\epsilon\mathbb{R}^k), \tag{3.1}$$

where $\phi_k(z;\Psi)$ is the k-dimensional normal density with zero mean vector and correlation matrix Ψ being of full rank, $\Phi(.)$ is the cumulative distribution function of the k-dimensional standard normal, and λ is a k-dimensional column vector with constant values. To show this in short form, it is common to write $Z \sim SN_k(0_k, \Psi, \lambda)$.

The parameter λ plays a key role in representing the main features of density in (3.1). Since it controls the skewness of density, it is usually referred to as shape parameter or, also, skewness control. This density function is skewed to the right (left) for positive (negative) values of λ. When $\lambda = 0$, the distribution function (3.1) reduces to $N(0_k, \Psi)$, where 0_q is a zero vector of length q.

Location and scale parameters can be also added to the skew-normal density of Z given in (3.1). Let us write

$$Y = \xi + \omega Z, \tag{3.2}$$

where $\xi = \left(\xi_1, ..., \xi_k\right)^T$, and $\omega = \text{diag}(\omega_1, ..., \omega_k)$, are location and scale parameters, respectively. Note that components of ω are assumed to be all positive. The density function of Y is then given by

$$2\phi_k(y-\xi;\Omega)\Phi(\lambda^T\omega^{-1}(y-\xi)), \tag{3.3}$$

where $\Omega = \omega\Psi\omega^T = \omega\Psi\omega$ represents the covariance matrix of Y. We use the standard notation $Y \sim SN_k(\xi, \Omega, \lambda)$ to indicate that Y follows the density function in (3.3). To have a general graphical view of this density, we provided some plots for particular values of the parameters in (3.3). The Figure 1 shows the contour plots of bivariate skew-normal density and the histogram of each variable for a bivariate skew-normal density. Now, we are in a position to concentrate on the estimators in an SUR model under the skew-normal distribution for the error term. Consider the model (2.4), with altering the index i to t, where

$$u_{t\bullet} = (u_{t1}, ..., u_{tg})^T \sim SN_g(0_g, \Sigma, \lambda), \qquad t = 1, ..., n. \tag{3.4}$$

Now, suppose one is interested in the estimator of parameters in this model through the maximum likelihood approach. Then, corresponding logarithm of the likelihood function, say $\ell = l(\lambda, \beta_{\bullet}, \Sigma)$, which is given by

$$\begin{aligned}\ell = l(\lambda, \beta_{\bullet}, \Sigma) &= n\log 2 - \tfrac{ng}{2}\log(2\pi) - \tfrac{n}{2}\log|\Sigma|\\ &-\tfrac{1}{2}\sum_{t=1}^{n}[(y_{t\bullet}-X_{t\bullet}\beta_{\bullet})^T\Sigma^{-1}(y_{t\bullet}-X_{t\bullet}\beta_{\bullet})] + \sum_{t=1}^{n}\log[\Phi_1(\lambda^T\Sigma^{-1/2}u_{t\bullet})],\end{aligned} \tag{3.5}$$

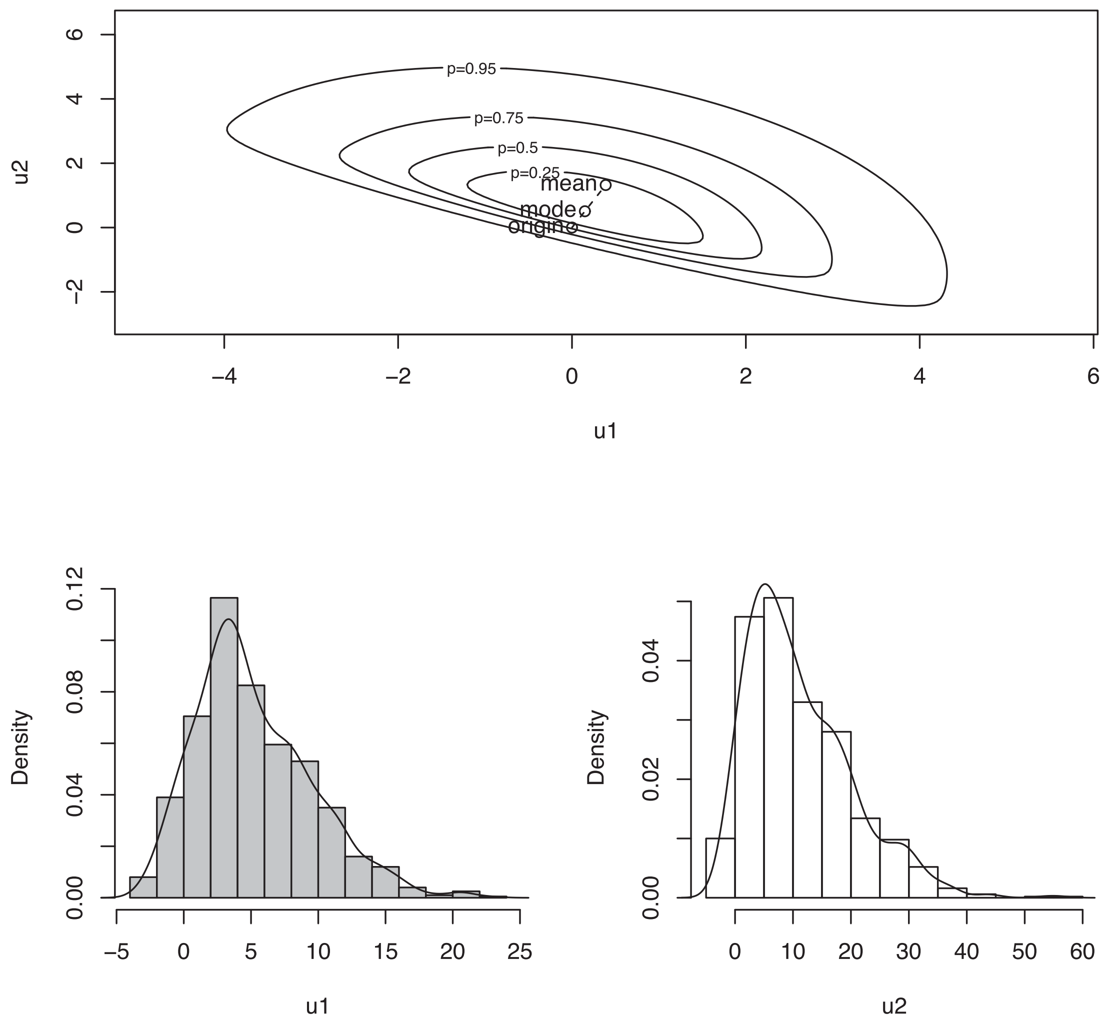

Figure 1 Contour plot of bivariate skew-normal density functions when $\xi^T = (0, 0)$, $\lambda^T = (4, 7)$, $\sigma_{11} = 4$, $\sigma_{12} = -1$, and $\sigma_{22} = 5.5$ plotted in upper panel. Also the marginal histogram of each variable are provided in the lower panel.

needs to be maximized. If we regard $\eta = \Sigma^{-1/2}\lambda$ as a new parameter, instead of λ, it results in splitting the parameters in (3.5) in the following sense: for fixed β and η, maximization of l with respect to Σ is equivalent to maximizing the analogous function for normal density for fixed β, which has a well-known solution (see, e.g. Mardia *et al.* [22]) given by

$$\hat{\Sigma}(\beta_{\bullet}) = V(\beta_{\bullet}) = \frac{1}{n}\sum_{t=1}^{n}(y_{t\bullet} - X_{t\bullet}\beta_{\bullet})(y_{t\bullet} - X_{t\bullet}\beta_{\bullet})^T. \tag{3.6}$$

By substituting this estimation into the expression in (3.5), one will obtain

$$l^*(\eta, \beta_{\bullet}) = C - \frac{n}{2}\log|V(\beta_{\bullet})| - \frac{ng}{2} + \sum_{t=1}^{n}\log\zeta_0(\eta^T u_{t\bullet}), \tag{3.7}$$

where $\zeta_0(x) = \log(2\Phi(x))$ and $x \sim N(0, 1)$. Now, to get the estimators for the rest of the parameters, one needs to maximize $l^*(\eta, \beta_{\bullet})$, which is, in fact, the profile likelihood function [23], with respect to η and $\beta_{\bullet}$. To do so, the partial derivatives of $l^*(\eta, \beta_{\bullet})$ with respect to η and $\beta_{\bullet}$

can be written, respectively, as

$$\begin{aligned}
\frac{\partial l^*(\eta,\beta_\bullet)}{\partial\eta} &= \sum_{t=1}^{n} u_{t\bullet}\zeta_1\left(\eta^T u_{t\bullet}\right) = \sum_{t=1}^{n}\left(y_{t\bullet}-X_{t\bullet}\beta_\bullet\right)\zeta_1\left[\eta'\left(y_{t\bullet}-X_{t\bullet}\beta_\bullet\right)\right] \\
\frac{\partial l^*(\eta,\beta_\bullet)}{\partial\beta_\bullet} &= -\frac{n}{2}\frac{\partial\log|V(\beta_\bullet)|}{\partial\beta_\bullet} - \sum_{t=1}^{n} X_{t\bullet}^T\eta\zeta_1\left[\eta^T\left(y_{t\bullet}-X_{t\bullet}\beta_\bullet\right)\right] \\
&= -\frac{n}{2}\begin{pmatrix} \operatorname{tr}\left(V^{-1}\frac{\partial V}{\partial\beta_1}\right) \\ \operatorname{tr}\left(V^{-1}\frac{\partial V}{\partial\beta_2}\right) \\ \vdots \\ \operatorname{tr}\left(V^{-1}\frac{\partial V}{\partial\beta_k}\right)\end{pmatrix} - \sum_{t=1}^{n} X_{t\bullet}^T\eta\zeta_1\left[\eta^T\left(y_{t\bullet}-X_{t\bullet}\beta_\bullet\right)\right],
\end{aligned} \tag{3.8}$$

where $\zeta_1(x) = \phi(x)/\Phi(x)$. As seen, one cannot derive some closed solutions (estimators) from the equations in (3.8). Hence, some numerical maximization procedures need to be implemented for this purpose. There are numerous literature for such numerical computations. See, for example, Robert and Casella [24]. A common approach is to follow the quasi-Newton algorithm. To do so, we are required to get the second derivatives of the expression in (3.7). They are given as follows:

$$\begin{aligned}
\frac{\partial^2 l^*(\eta,\beta_\bullet)}{\partial\eta\partial\eta^T} &= \sum_{t=1}^{n}\left(y_{t\bullet}-X_{t\bullet}\beta_\bullet\right)\left(y_{t\bullet}-X_{t\bullet}\beta_\bullet\right)^T\zeta_2\left[\eta^T\left(y_{t\bullet}-X_{t\bullet}\beta_\bullet\right)\right] \\
\frac{\partial^2 l^*(\eta,\beta_\bullet)}{\partial\beta_\bullet^T\partial\beta_\bullet} &= -\frac{n}{2}\begin{pmatrix} \operatorname{tr}\left(V^{-1}\frac{\partial^2 V}{\partial\beta_1^2}\right) & \operatorname{tr}\left(V^{-1}\frac{\partial^2 V}{\partial\beta_2\partial\beta_1}\right) & \dots & \operatorname{tr}\left(V^{-1}\frac{\partial^2 V}{\partial\beta_k\partial\beta_1}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{tr}\left(V^{-1}\frac{\partial^2 V}{\partial\beta_1\partial\beta_k}\right) & \operatorname{tr}\left(V^{-1}\frac{\partial^2 V}{\partial\beta_2\partial\beta_k}\right) & \dots & \operatorname{tr}\left(V^{-1}\frac{\partial^2 V}{\partial\beta_k^2}\right)\end{pmatrix} \\
&\quad +\frac{n}{2}\begin{pmatrix} \operatorname{tr}\left(V^{-1}\frac{\partial V}{\partial\beta_1}V^{-1}\frac{\partial V}{\partial\beta_1}\right) & \operatorname{tr}\left(V^{-1}\frac{\partial V}{\partial\beta_2}V^{-1}\frac{\partial V}{\partial\beta_1}\right) & \dots & \operatorname{tr}\left(V^{-1}\frac{\partial V}{\partial\beta_k}V^{-1}\frac{\partial V}{\partial\beta_1}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{tr}\left(V^{-1}\frac{\partial V}{\partial\beta_1}V^{-1}\frac{\partial V}{\partial\beta_k}\right) & \operatorname{tr}\left(V^{-1}\frac{\partial V}{\partial\beta_2}V^{-1}\frac{\partial V}{\partial\beta_k}\right) & \cdots & \operatorname{tr}\left(V^{-1}\frac{\partial V}{\partial\beta_k}V^{-1}\frac{\partial V}{\partial\beta_k}\right)\end{pmatrix} \\
&\quad + \sum_{t=1}^{n} X_{t\bullet}^T\eta\eta^T X_{t\bullet}\zeta_2\left[\eta^T\left(y_{t\bullet}-X_{t\bullet}\beta_\bullet\right)\right] \\
\frac{\partial^2 l^*(\eta,\beta_\bullet)}{\partial\beta_\bullet^T\partial\eta} &= -\sum_{t=1}^{n}\left\{X_{t\bullet}^T\eta(y_{t\bullet}-X_{t\bullet}\beta_\bullet)^T\zeta_2\left[\eta^T\left(y_{t\bullet}-X_{t\bullet}\beta_\bullet\right)\right] + X_{t\bullet}^T\zeta_1\left[\eta^T\left(y_{t\bullet}-X_t\cdot\beta_\bullet\right)\right]\right\}, \\
\frac{\partial^2 l^*(\eta,\beta_\bullet)}{\partial\eta^T\partial\beta_\bullet} &= \left(\frac{\partial^2 l^*(\eta,\beta_\bullet)}{\partial\beta_\bullet^T\partial\eta}\right)^T,
\end{aligned} \tag{3.9}$$

where $\zeta_2(x) = -\zeta_1(x)[x+\zeta_1(x)]$. If Υ is the parameter of interest, using the gradient of the function in which this parameter appears, the quasi-Newton algorithm apply as

$$\Upsilon_{(k+1)} = \Upsilon_{(k)} - (\nabla^2 f)^{-1}_{(k)}(\nabla f)_{(k)}, \tag{3.10}$$

where the indices are used to show the value of the estimator at corresponding stage and (ignoring the index)

$$\nabla f = \begin{pmatrix} \dfrac{\partial l^*(\eta,\beta_\bullet)}{\partial\eta} \\ \dfrac{\partial l^*(\eta,\beta_\bullet)}{\partial\beta_\bullet}\end{pmatrix}, \quad \nabla^2 f = \begin{pmatrix} \dfrac{\partial^2 l^*(\eta,\beta_\bullet)}{\partial\beta_\bullet\partial\beta_\bullet^T} & \dfrac{\partial^2 l^*(\eta,\beta_\bullet)}{\partial\beta_\bullet^T\partial\eta} \\ \dfrac{\partial^2 l^*(\eta,\beta_\bullet)}{\partial\eta^T\partial\beta_\bullet} & \dfrac{\partial^2 l^*(\eta,\beta_\bullet)}{\partial\eta\partial\eta^T}\end{pmatrix}. \tag{3.11}$$

We conduct some simulation studies using model (2.1) along with normal and skew-normal distributions in the following section. Moreover, we investigate the application of these methods in real-life data.

4. SIMULATION STUDIES AND APPLICATION

Here, we outline our simulation study to evaluate the performance of the parameters estimation for the SUR models given in Section 2. Suppose we have the following model:

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

To further identification of this model, we need to indicate a distribution for $(u_1, u_2)^T$. To start, let us assume $u = (u_1, u_2)^T \sim N(0_2, \Sigma)$, y_1 and y_2 are endogenous variables and z_1, x_1 and x_2 are exogenous variables. To compare this model with an alternative, we also consider the case in which $u = (u_1, u_2)^T \sim SN(0_2, \Sigma, \lambda)$.

We fix the parameter in our simulation studies as $\beta_{\bullet} = (6, -3, -4, 9, 3, -2)^T$, $\lambda = (2, 3)^T$ and

$$\Sigma = \begin{pmatrix} 12 & -2 \\ -2 & 11 \end{pmatrix}. \tag{4.2}$$

To initiate our simulation studies, we take the sample size equal to 1000, in which using two equations in (4.1) ends up with the total observations 2000. Then, we generate data for 1000 times from skew-normal distribution. Thereafter, the model was fitted by both maximum likelihood approaches (normal and skew-normal assumptions) as described in previous sections. Particularly, the parameters were estimated based upon either equations in (2.6) and (3.10), depending on the distribution considered for the errors in the model.

The results gained from our simulation studies for both the normal and skew-normal cases are given in Table 1. As seen, the table is partitioned into two parts. The three left- hand sides panels are related to the results coming from the normal assumption and the rest on the right belong to the skew-normal assumption both for error term. The distributions are indicated by N (Normal) and SN (Skew-Normal). Furthermore, the table includes estimate, standard deviation (SD), and effect size (ES).

Based on the results in Table 1, the estimates for β_1, β_2, γ_1, and γ_2 have small ES in both cases. The ES for the intercept is high regardless of which distribution is considered for the error term. However, it is higher in the normal model compared to the skew-normal case. Overall, the estimates in the SN case are closer to the real value of parameters before conducting the simulation. In general, when response variables follow a skew-normal distribution in the SUR model, the methods relied on the skew-normal density for the error leads to more accurate estimation than the normal density case.

One notes that the likelihood ratio test for the null hypothesis $\lambda = 0$ can be considered as a criterion for a comparison in whether or not the skew-normal distribution should be considered. This test is given by

$$2\left\{\ell(\hat{\beta}_{\bullet}, \hat{\Sigma}, \hat{\lambda}) - \ell(\hat{\mu}, \hat{\Omega}, 0)\right\}, \tag{4.3}$$

where $\hat{\beta}_{\bullet}$, $\hat{\Sigma}$, and $\hat{\lambda}$ denote the MLE under the assumption of skew-normality (shorten as SN-ML) and $\hat{\mu}$ and $\hat{\Omega}$, are MLE under the assumption of normality (shorten as N-ML) for the errors. Following Casella and Berger [25], the expression (4.3) follows χ^2_{df} where df is the difference on the dimensions of parameter in the alternative and null hypotheses. The logarithm of the likelihood and AIC criterion for both methods appear in Table 2. As it can be seen, the logarithm of the likelihood for the SN-ML is higher than that of N-ML. Moreover, the

Table 1 The result of SUR model fitted according to the skew-normal and normal assumptions.

	N-ML			SN-ML		
Parameter	**Estimate**	**SD**	**ES**	**Estimate**	**SD**	**ES**
β_0	9.552	0.602	3.552	5.736	0.313	0.264
β_1	−3.001	0.018	0.001	−3.003	0.011	0.003
β_2	−4.002	0.067	0.002	−3.982	0.023	0.018
γ_0	18.11	0.751	9.11	8.711	0.451	0.289
γ_1	3.007	0.063	0.007	3.007	0.029	0.007
γ_2	−2.013	0.068	0.013	−1.969	0.037	0.031
σ_{11}	20.55	4.849	8.55	12.47	1.059	0.474
σ_{22}	41.42	5.543	30.42	11.25	3.377	0.258
σ_{12}	−9.57	1.414	7.577	−1.72	1.175	0.28
λ_1	–	–	–	2.210	0.691	0.210
λ_2	–	–	–	3.211	1.080	0.211

Table 2 Criteria to compare two methods of model parameters estimate.

Criteria	N-ML	SN-ML
AIC	13143.32	13035.198
Log likelihood	−6560.66	−6508.599

Table 3 Description of variables utilized in model (4.4).

Variable Names	Abbreviation Signs	Variable Type	Coding
Households 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	–
Nonagriculture self-employment income	D_2	Quantitative	–
Miscellaneous income	D_3	Quantitative	–
Non-monetary other incomes	D_4	Quantitative	–

AIC criterion for the SN-ML is less than that of the N-ML. Therefore, SN-ML outperforms N-ML in this study which means that, in comparison with the N-ML distribution, using the skew-normal density for the error term in the SUR model (3.10), leads to an improvement on the accuracy and bias of the estimators. Here, the likelihood ratio test statistics was $LRT = 2\left\{\ell(\hat{\beta}_{\bullet}, \hat{\Sigma}, \hat{\lambda}) - \ell(\hat{\mu}, \hat{\Omega}, 0)\right\} = 119.48$ with $df = 2$. Hence, the test is significant at 0.05 level; therefore, it can be stated that the skew parameters (λ) is not zero. This supports our initial assumption on considering the skew-normal distribution for the error terms.

We were interested in applying the proposed model in this paper in real-life data. To do this, we used the Iranian rural households income and expenditure data collected in the year 2009. It includes 13345 families from 32 provinces. In the present paper, the main goal is a survey effects of some variables on Iranian rural households income and expenditure. In this study, these two variables are considered as endogenous variables and other covariates are set as exogenous. Based on a general view and also consulting experts in the Statistical Center of Iran, the following SUR was utilized to express the inter-relationship between rural households income and expenditure in Iran:

$$\begin{aligned} GH &= \beta_0 + \sum_{i=1}^{4} \beta_{C_i} C_i + \sum_{i=1}^{5} \beta_{B_i} B_i + \beta_A A + \epsilon_1 \\ D &= \gamma_0 + \sum_{i=1}^{4} \gamma_{D_i} D_i + \epsilon_2. \end{aligned} \tag{4.4}$$

A general description of the considered variables is provided in Table 3. Figures 2–4 present a geometric display of two important variables.

To initiate the analysis, the validity of the normality assumption for the response variables should be tested. We used the Kolmogorov–Smirnov (KS) test statistics for this purpose. The results of the KS test was significant with p-value < 0.05, rejecting the null hypothesis; assuming the normality density. To have a visual inspection of the density, the Q-Q plot of the households income and expenditure are also drawn in Figure 5. They show the departure of univariate normal distribution for both variables. The contour plot in Figure 5 also demonstrates a departure from the bivariate normal distribution. It can be argued that some transformations, such as logarithm, to make density

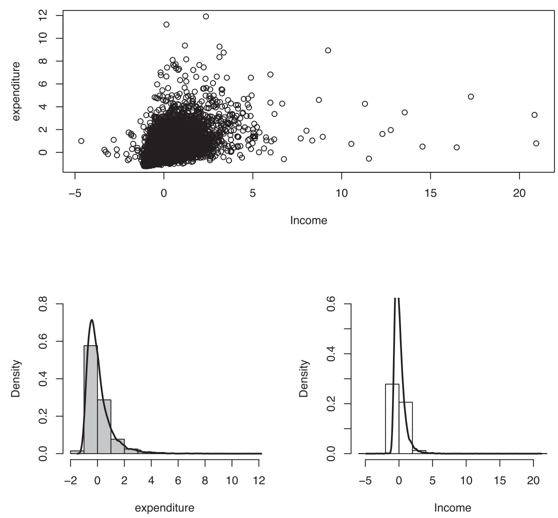

Figure 2 | The scatter plot of Iranian rural households income and expenditure. Also the marginal histogram of each variable are provided in the lower panel.

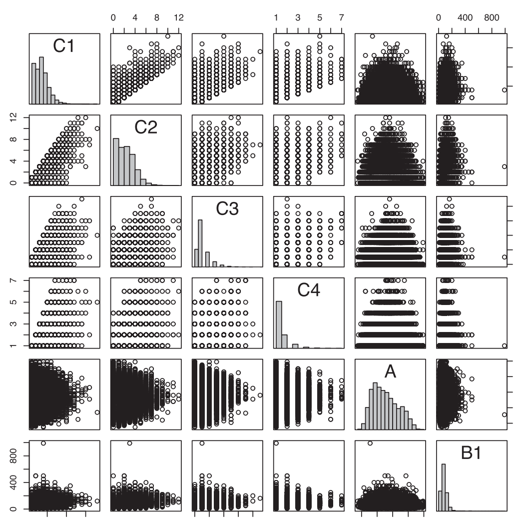

Figure 3 | The pairs plot of quantitative variables described in Table 3.

normal is appropriate. However, the income variable includes some negative values and so we are not allowed to utilize this transformation. Instead, we preferred to use the skew-normal distribution for the errors and attempted to model the rural households income and expenditure in Iran based upon this methodology. Nonetheless, to have a basement for our further comparison, the normal distribution was also considered for the errors in this example.

The results from employing aforementioned models for our example are appeared in Table 4. As seen, it includes three panels. The first (second) panel shows the results for the first (second) equation of the model (4.4). Confining ourselves only to those significant estimates of the parameters at %5 level, the results for the normal and skew-normal densities are provided in both panels. The last panel shows the estimation for the components of the covariance matrix and shape parameters. A test was conducted to check whether or not the skewness parameter (λ) is equal to zero. This led to $LRT = 2\left\{\ell(\hat{\beta}_{\bullet}, \hat{\Sigma}, \hat{\lambda}) - \ell(\hat{\mu}, \hat{\Omega}, 0)\right\} = -24385.1 - (-24444.4) = 59.3$ with $df = 2$. Since the test was significant at 0.05 level, we accept that the skew parameter is not zero, and using the skew-normal MLE is more effective than the normal MLE.

Based on the results given in the first panel of Table 4, using facilities (including the Internet, gas, and mobile), has a direct effect on family households expenditure in Iran. In other words, using these facilities can increase family households expenditure. It is also seen that, family size, number of literate, employees, and people with income in household and age have direct link with family households expenditure.

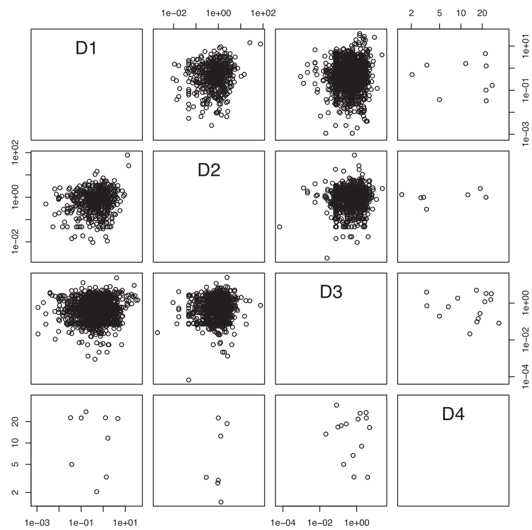

Figure 4 | The pairs plot of quantitative variables described in Table 3.

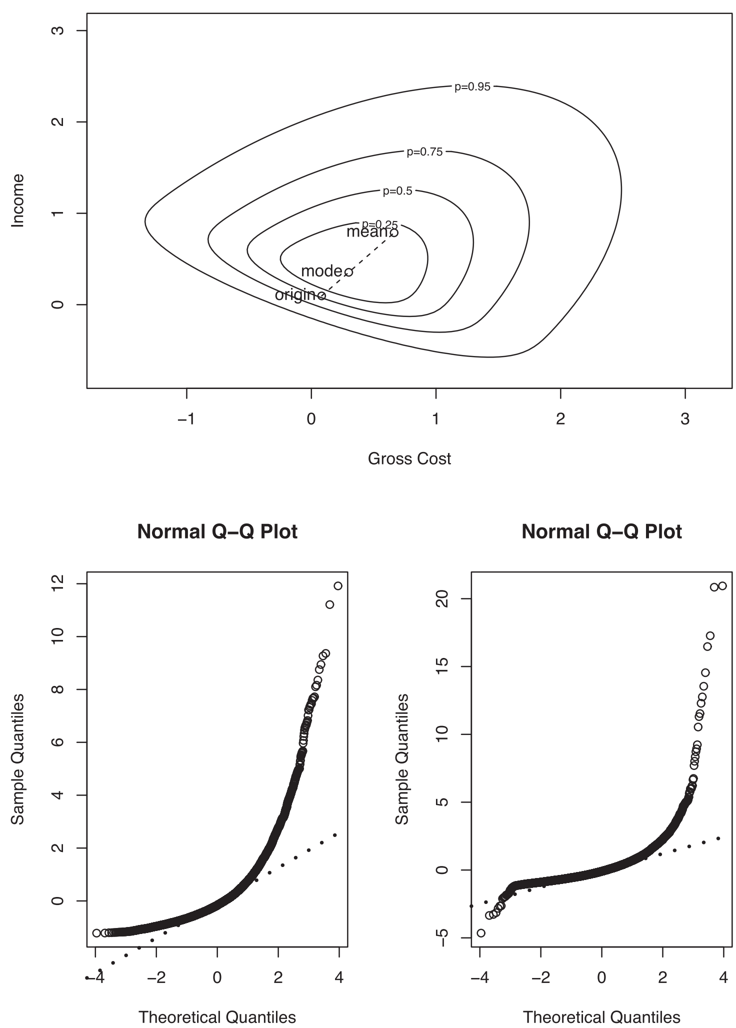

Figure 5 | The contour plot of rural households income and expenditure along with the Q-Q plot for each variable.

Table 4 The result of fitting the seemingly unrelated regression (SUR) model in (4.4) considering the skew-normal and normal distributions assumption for the response in the Iranian rural households income and expenditure data on year 2009.

	Estimation		Std.error	
Parameter	N-ML	SN-ML	N-ML	SN-ML
β_0	−1.50	−1.34	0.047	0.006
β_{C_1}	0.036	0.040	0.009	0.001
β_{C_2}	0.082	0.046	0.010	0.002
β_{C_3}	0.106	0.059	0.011	0.004
β_{C_4}	0.061	−0.024	0.014	0.004
β_{B_1}	0.003	0.003	0.0006	0.0001
β_{B_2}	0.004	0.002	0.0002	0.0005
β_{B_3}	0.649	0.531	0.024	0.013
β_{B_4}	0.689	0.490	0.051	0.032
β_{B_5}	0.064	0.031	0.018	0.0085
β_A	0.276	0.137	0.025	0.0064
γ_0	0	−0.103	0.025	0.0038
γ_{D_1}	0.544	0.499	0.013	0.0039
γ_{D_2}	0.503	0.487	0.013	0.0039
γ_{D_3}	0.412	0.352	0.013	0.0039
γ_{D_4}	0.033	0.030	0.013	0.0038
σ_{11}	0.656	0.051	0.011	0.009
σ_{21}	0.084	0.009	0.009	0.001
σ_{22}	0.315	0.018	0.008	0.004
λ_1	–	1.181	–	0.104
λ_2	–	0.869	–	0.097

Moreover, regarding the second panel of Table 4, the agriculture self-employment, non-agriculture self-employment, miscellaneous income, and non-monetary other incomes have direct effect on the family incomes.

5. CONCLUSION

There are some examples of encountering with data having an asymmetric histogram. Considering some skew-normal distributions is usually a solution to construct a model. The problem will be harder if one should take SEMs into account. Confining to the SUR model, which is a particular case of SEM, we discussed the method of estimation for the parameters of this model in this paper. Here, the response variables were following the skew-normal distribution. Performance of the proposed method has been compared with an alternative case in which the normal density is incorrectly assumed for the error. Then, we applied the methods discussed in this paper on real data. Results shown superiority of our approach to other methods relied on normal distribution for the error. There is still room to extend the model in this paper. One of the possible options is to investigate the performance of the Bayesian approach on the SUR model with skew-normal assumption for the error term. Moreover, to check how other skew distributions such as skew-t density works on the SUR models worth to study.

CONFLICTS OF INTEREST

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article

Funding Statement

Receiving support from the Center of Excellence in Analysis of Spatio Temporal Correlated Data at Tarbiat Modares University.

REFERENCES

1. T. Haavelmo, Econometrica. 11 (1943), 1–12.
2. A. Zellner, J. Am. Stat. Assoc. 58 (1963), 977–992.
3. A. Zellner, An Introduction to Bayesian Inference in Econometrics, John Wiley and Sons, New York, NY, USA, 1971.
4. D.A.S. Fraser, M. Rekkasb, A. Wong, J. Econom. 127 (2005), 17–33.
5. A. Zellner, T. Ando, Bayesian Anal. 5 (2010), 65–96.
6. D.F. Percy, J. R. Stat. Soc. Ser. B. 54 (1992), 243–252.
7. S. Chib, E. Greenberg, J. Econ. 68 (1995), 339–360.
8. M. Smith, R. Kohn, J. Econ. 98 (2000), 257–282.
9. A. Zellner, T. Ando, N. Basturk, L. Hoogerheide, H. Van Dijk, Econom. Rev. 33 (2014), 3–35.
10. D. Warton, F. Hui, Ecol. Soc. Am. 92 (2011), 3–10.
11. A. Azzalini, Statistica. 46 (1986), 199–208.
12. A. Azzalini, A. Dalla Valle, Biometrika. 83 (1996), 715–726.
13. A. Azzalini, A. Capitanio, J. R. Stat. Soc. B. 61 (1999), 579–602.
14. N. Balakrishnan, Test. 11 (2002), 37–39.
15. G.G. Genton, Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality, Chapman and Hall, Boca Raton, FL, USA, 2004.
16. A.K. Gupta, G. Gonzlez-faras, J.A.A. Domnguez-molina, J. Multivar. Anal. 89 (2004), 181–190.
17. R.B. Arellano-Valle, M.A. Cortes, H.W. Gomez, Commun. Stat. Theory Methods. 39 (2010), 912–922.
18. A. Azzalini, G. Regoli, Ann. Inst. Stat. Math. 64 (2012), 857–879.
19. T.W. Anderson, H. Rubin, Ann. Math. Stat. 20 (1949), 46–63.
20. H. Theil, Bull. Int. Stat. Inst. 34 (1954), 122–129.
21. R. Davidson, J.G. MacKinnon, Estimation and Inference in Econometrics, Oxford University Press, New York, NY, USA, 1993.
22. K.V. Mardia, J.T. Kent, J.M. Bibby, Multivariate Analysis, Academic Press, New York, NY, USA, 1979.
23. A. Raue,C. Kreutz,T. Maiwald,J. Bachmann,M. Schilling,U. Klingmuller, J. Timmer, Bioinformatics. 25 (2009), 1923–1929.
24. C. Robert, G. Casella, Monte Carlo Statistical Methods, Springer-Verlag, New York, NY, USA, 2004.
25. G. Casella, R. Berger, Statistical Inference, second ed., Duxbury Press, Pacific Grove, CA, USA, 2002.

APPENDIX

Theorem: For any fixed $(p \times p)$ matrix $A > 0$.

$$f(\Sigma) = |\Sigma|^{-n/2} \exp\left\{-\frac{1}{2} tr\Sigma^{-1}A\right\} \tag{5.1}$$

is maximized over $\Sigma > 0$ by $\Sigma = n^{-1}A$, and so $f(n^{-1}A) = |n^{-1}A|^{-n/2} e^{-\frac{np}{2}}$.

In Equation (3.8), $\frac{\partial V}{\partial \beta_j}$ is determined as follows:

Suppose $A_i = y_{ti} - X_{ti}\beta_i^*$ is the i-th observation from i-th equation and β_i^* is k_i-vector and X_{ti}^T is a k_i-vector. Also consider

$$A_t^* = \begin{pmatrix} A_1 & A_2 & \dots & A_g \\ 0 & A_2 & \dots & A_g \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_g \end{pmatrix}_{g\times g}, \beta_{\bullet} = \begin{pmatrix} \beta_1^* \\ \vdots \\ \beta_g^* \end{pmatrix}_{k\times 1}, X_{t\bullet} = \begin{pmatrix} X_{t1} & 0 & \dots & 0 \\ 0 & X_{t2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_{tg} \end{pmatrix}_{g\times k} \tag{5.2}$$

where $k = \sum_{i=1}^{g} k_i$. Here, the main goal is to get the derivative of V with respect to j-th parameter of $\beta_{\bullet}$, that is β_j (for $j = 1, \dots, k$). Therefore, we define k-vector whose that its j-th element is 1 and the other ones are all zero. Similarly, we determine β_j a g-vector in which its the i-th element is 1 and the other ones are zero. Since $X_{t(j)}^*$ is only appears in i-th equation in a particular manner, we define:

$$a = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}_{k\times 1}, \quad b = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}_{g\times 1}. \tag{5.3}$$

Hence, $X_{t(j)}^* = b^T X_{t\bullet} a$ where $X_{t(j)}^*$ is the corresponding variable to β_j. The last step for determining the derivative of V is to set the matrix C_{t_j} as

$$C_{t_j} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -X_{t(j)}^* & \dots & 0 & \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}_{g\times g} A_t^*. \tag{5.4}$$

As it can be seen, the first column and i-th row of C_{t_j} is equal to $-X_{t(j)}^*$. Finally, for all other observations, the corresponding derivative is given as:

$$\frac{\partial V}{\partial \beta_j} = \frac{1}{n} \sum_{t=1}^{n} (C_{t_j} + C_{t_j}^T). \tag{5.5}$$

As a general rule, the Hessian matrix is required if one is interested in utilizing the quasi-Newton algorithm. The relevant derivatives to construct such a matrix are as follows:

$$\begin{aligned} \frac{\partial^2 \ell^*}{\partial \beta_{\bullet}^T \partial \eta} &= \frac{\partial}{\partial \beta_{\bullet}^T} \left[\sum_{t=1}^{n} \left(y_t - X_{t\bullet}\beta_{\bullet}\right) \zeta_1 \left(\eta^T \left(y_t - X_{t\bullet}\beta_{\bullet}\right)\right) \right] \\ &= \sum_{t=1}^{n} \left[\frac{\partial}{\partial \beta_{\bullet}^T} y_t \zeta_1 \left(\eta^T \left(y_t - X_{t\bullet}\beta_{\bullet}\right)\right) - \frac{\partial}{\partial \beta_{\bullet}^T} X_{t\bullet}\beta_{\bullet}\zeta_1 \left(\eta^T \left(y_t - X_{t\bullet}\beta_{\bullet}\right)\right) \right] \\ &= \sum_{t=1}^{n} \left[-X_{t\bullet}^T \eta y_t^T \xi_2 \left(\eta^T \left(y_t - X_{t\bullet}\beta_{\bullet}\right)\right) - X_{t\bullet}^T \zeta_1 \left(\eta^T \left(y_t - X_{t\bullet}\beta_{\bullet}\right)\right) + X_{t\bullet}^T \eta \beta_{\bullet}^T X_{t\bullet}^T \zeta_2 \left(\eta^T \left(y_t - X_t \bullet \beta_{\bullet}\right)\right) \right] \\ &= -\sum_{t=1}^{n} \left[X_{t\bullet}^T \eta \left(y_t - X_{t\bullet}\beta_{\bullet}\right)^T \zeta_2 \left(\eta^T \left(y_t - X_{t\bullet}\beta_{\bullet}\right)\right) + X_{t\bullet}^T \zeta_1 \left(\eta^T \left(y_t - X_{t\bullet}\beta_{\bullet}\right)\right) \right]. \end{aligned} \tag{5.6}$$

Notice that we used the property $\left(\frac{\partial^2 \ell^*}{\partial \eta^T \partial \beta_\bullet}\right)^T = \frac{\partial^2 \ell^*}{\partial \beta_\bullet^T \partial \eta}$. The second derivative of ℓ^* subject to η is straightforward. However, the computation of $\frac{\partial^2 \ell^*}{\partial \beta_\bullet \partial \beta_\bullet^T}$ is too tough. To obtain this derivative, we applied formula (5.5) to get:

$$
\begin{aligned}
\frac{\partial^2 \ell^*}{\partial \beta_\bullet \partial \beta_\bullet^T} &= \frac{\partial}{\partial \beta_\bullet}\left\{-\frac{n}{2}\begin{pmatrix} \mathrm{tr}\left(V^{-1}\frac{\partial V}{\partial \beta_1}\right) \\ \vdots \\ \mathrm{tr}\left(V^{-1}\frac{\partial V}{\partial \beta_k}\right)\end{pmatrix}^T - \sum_{t=1}^n \eta^T X_{t\bullet}\zeta_1\left[\eta'\left(y_{t\bullet} - X_{t\bullet}\beta_\bullet\right)\right]\right\} \\
&= -\frac{n}{2}\frac{\partial}{\partial \beta_\bullet}\begin{pmatrix} \mathrm{tr}\left(V^{-1}\frac{\partial V}{\partial \beta_1}\right) \\ \vdots \\ \mathrm{tr}\left(V^{-1}\frac{\partial V}{\partial \beta_k}\right)\end{pmatrix}^T + \sum_{t=1}^n X_{t\bullet}^T \eta\eta^T X_{t\bullet}\zeta_2\left(\eta^T\left(y_t - X_{t\bullet}\beta_\bullet\right)\right) \\
&= -\frac{n}{2}\begin{pmatrix} \frac{\partial}{\partial \beta_1}\mathrm{tr}\left(V^{-1}\frac{\partial V}{\partial \beta_1}\right) & \frac{\partial}{\partial \beta_2}\mathrm{tr}\left(V^{-1}\frac{\partial V}{\partial \beta_1}\right) & \cdots & \frac{\partial}{\partial \beta_k}\mathrm{tr}\left(V^{-1}\frac{\partial V}{\partial \beta_1}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial \beta_1}\mathrm{tr}\left(V^{-1}\frac{\partial V}{\partial \beta_k}\right) & \frac{\partial}{\partial \beta_2}\mathrm{tr}\left(V^{-1}\frac{\partial V}{\partial \beta_k}\right) & \cdots & \frac{\partial}{\partial \beta_k}\mathrm{tr}\left(V^{-1}\frac{\partial V}{\partial \beta_k}\right)\end{pmatrix} \\
&\quad + \sum_{t=1}^n X_{t\bullet}^T \eta\eta^T X_t \bullet \zeta_2\left(\eta^T\left(y_t - X_t \bullet \beta_\bullet\right)\right) \\
&= -\frac{n}{2}\begin{pmatrix} \mathrm{tr}\left(V^{-1}\frac{\partial^2 V}{\partial \beta_1^2}\right) & \mathrm{tr}\left(V^{-1}\frac{\partial^2 V}{\partial \beta_2 \partial \beta_1}\right) & \cdots & \mathrm{tr}\left(V^{-1}\frac{\partial^2 V}{\partial \beta_k \partial \beta_1}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{tr}\left(V^{-1}\frac{\partial^2 V}{\partial \beta_1 \partial \beta_k}\right) & \mathrm{tr}\left(V^{-1}\frac{\partial^2 V}{\partial \beta_2 \partial \beta_k}\right) & \cdots & \mathrm{tr}\left(V^{-1}\frac{\partial^2 V}{\partial \beta_k^2}\right)\end{pmatrix} \\
&\quad + \frac{n}{2}\begin{pmatrix} \mathrm{tr}\left(V^{-1}\frac{\partial V}{\partial \beta_1}V^{-1}\frac{\partial V}{\partial \beta_1}\right) & \mathrm{tr}\left(V^{-1}\frac{\partial V}{\partial \beta_2}V^{-1}\frac{\partial V}{\partial \beta_1}\right) & \cdots & \mathrm{tr}\left(V^{-1}\frac{\partial V}{\partial \beta_k}V^{-1}\frac{\partial V}{\partial \beta_1}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{tr}\left(V^{-1}\frac{\partial V}{\partial \beta_1}V^{-1}\frac{\partial V}{\partial \beta_k}\right) & \mathrm{tr}\left(V^{-1}\frac{\partial V}{\partial \beta_2}V^{-1}\frac{\partial V}{\partial \beta_k}\right) & \cdots & \mathrm{tr}\left(V^{-1}\frac{\partial V}{\partial \beta_k}V^{-1}\frac{\partial V}{\partial \beta_k}\right)\end{pmatrix} \\
&\quad + \sum_{t=1}^n X_{t\bullet}^T \eta\eta^T X_{t\bullet}\zeta_2\left[\eta^T\left(y_{t\bullet} - X_{t\bullet}\beta_\bullet\right)\right].
\end{aligned}
\tag{5.7}
$$

On getting (5.7), we employed the following equality in which F is a non-singular matrix:

$$
\frac{\partial^2 \log|F|}{\partial x_i \partial x_j} = \frac{\partial \mathrm{tr}(F^{-1}\frac{\partial F}{\partial x_j})}{\partial x_i} = \mathrm{tr}\left(F^{-1}\frac{\partial^2 F}{\partial x_i \partial x_j}\right) - \mathrm{tr}\left(F^{-1}\frac{\partial^2 F}{\partial x_i \partial x_j}\right)
\tag{5.8}
$$

The components of the second matrix in the last expression (5.7) are determined using (5.5). Assuming β_j is a member of i-th equation in the SUR, we have:

$$
B = C_{t_j} + C_{t_j}^T = \begin{pmatrix} 0 & \dots & -X_{t(j)}^* A_1 & \dots & 0 \\ 0 & \dots & -X_{t(j)}^* A_2 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ -X_{t(j)}^* A_1 & \dots & -2X_{t(j)}^* A_i & \dots & -X_{t(j)}^* A_g \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & -X_{t(j)}^* A_g & \dots & 0 \end{pmatrix},
\tag{5.9}
$$

where all of the arrays equal zero except i-th row and column. The main diagonal of the favorite matrix β_j is a member of i-th equation and so

$$\frac{\partial^2 V}{\partial \beta_j^2} = \begin{pmatrix} 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 2X^*_{t(j)} & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 \end{pmatrix}, \qquad j = 1, \dots, k. \tag{5.10}$$

If both β_j and β_l are members of i-th equation in a SUR, then; we have:

$$\frac{\partial^2 V}{\partial \beta_j \beta_l} = \begin{pmatrix} 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 2X^*_{t(j)}X^*_{t(l)} & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 \end{pmatrix}, \qquad j, l = 1, \dots, k, \tag{5.11}$$

where all of the arrays are zero except the element in the (i, i) position. If β_j is a member of i-th equation and β_l is a member of m-th equation where $i \neq m$, then; we have:

$$\frac{\partial^2 V}{\partial \beta_j \beta_l} = \begin{pmatrix} 0 & \dots & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \dots & \dots & X^*_{t(j)}X^*_{t(l)} & \dots & 0 \\ 0 & X^*_{t(j)}X^*_{t(l)} & \dots & \dots & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \dots & \dots & 0 & \dots & 0 \end{pmatrix}, \qquad j, l = 1, \dots, k \tag{5.12}$$

where all of the arrays are zero except (i, m)-th and (m, i)-th components.