



## A Simulation Study of Bayesian Estimator for Seemingly Unrelated Regression under Different Distributional Assumptions

Ojo O. Oluwadare<sup>1\*</sup>, Owonipa R. Oluremi<sup>2</sup> and Enesi O. Lateifat<sup>2</sup>

<sup>1</sup>Department of Statistics, Federal University of Technology Akure, Nigeria.

<sup>2</sup>Department of Mathematical Sciences, Kogi State University, Anyigba, Nigeria.

### Authors' contributions

This work was carried out in collaboration among all authors. Author OOO designed the study, performed the statistical analysis, wrote the protocol, and wrote the first draft of the manuscript. Author ORO and Author EOL managed the analyses of the study. Author EOL managed the literature searches. All authors read and approved the final manuscript.

### Article Information

DOI: 10.9734/AJPAS/2020/v10i430251

#### Editor(s):

(1) Dr. Seemon Thomas, Mahatma Gandhi University, India.

(2) Dr. Ali Akgül, Siirt University, Turkey.

(3) Dr. S. M. Aqil Burney, University of Karachi, Pakistan.

#### Reviewers:

(1) Mohamed Reda Abonazel, Cairo University, Egypt.

(2) Zakariya Yahya Algamal, University of Mosul, Iraq.

Complete Peer review History: <http://www.sdiarticle4.com/review-history/64433>

Received: 10 November 2020

Accepted: 15 January 2021

Published: 25 January 2021

Original Research Article

## Abstract

This paper presents Bayesian analysis of Seemingly Unrelated Regression (SUR) model. An independent prior for parameters was used. The Bayesian method was compared with classical method of estimation to know the most efficient estimator under different distributional assumptions through a simulation study. In order to facilitate comparison among these estimators, Mean Squared Error (MSE) was considered as a criterion. Furthermore, based on the simulation, it was deduced that MSE of the Bayesian estimator is smaller than all the classical methods of estimation for SUR model while Normal distribution was considered as an ideal distribution in generation of disturbances in any simulation study.

Keywords: Bayesian; disturbance terms; independent prior; MSE; simulation.

AMS classification: 62J99; 62J20.

\*Corresponding author: E-mail: [ojoo@futa.edu.ng](mailto:ojoo@futa.edu.ng);

## **1 Introduction**

Seemingly unrelated regression (SUR) model is a kind of model that has recently gained popularity in different phenomena like finance, health and so many others. It is simply a generalization of linear regression model that involves more than one equation, and each of the equations has their own dependent variable and different sets of explanatory variables but all the errors in the model are contemporaneously correlated.

Many methods of estimation in SUR model had been proposed in classical inference. Method of estimating the parameters of SUR model was described by Zellner [1]. This method entails application of Aitken's Generalized Least Squares (GLS) to the entire system of equations. It was revealed that coefficients of the regression using the estimators are at least asymptotically more efficient than the equation by equation OLS method.

Zellner [2] also derived the finite sample distribution for parameters estimator in a two-equation case where the exogenous variables in different equations are orthogonal and the disturbances are normally distributed; he went further to compare the exact second-moment matrix of the estimator in this case with that of the Single-Equation Least-Squares (SELS).

An asymptotic series expansion of Edgeworth type of approximation to finite samples distribution function of SUR estimator of parameters was derived by Phillips [3]. This approximation was able to help to provide some evidence of finite sample behaviour of SUR estimator. The robust SUR estimator was also introduced by Koenker and Portnoy [4] based on M-estimators to solve the problem of outliers, but this procedure is not affine equivariant and does not take into consideration of the multivariate nature of the problem.

Hubert et al [5] proposed a fast algorithm called FastSUR for SUR model with the aid of simulation. This method performed well especially in the detection of outlier. A General Multivariate Chain Ladder (GMCL) method was proposed for claims reserving in non-insurance setting by Peremans et al [6]. This method was able to detect claims that have an abnormally large influence on reserve estimates.

The Bayesian inference in SUR model was first introduced with the work of Zellner [7]. In the work of Zellner [7], conditional posterior densities of SUR model parameters were derived. Although, when Zellner introduced Bayesian method of inference to SUR model, analytical methods for deriving posterior features of interest for individual SUR coefficients were not available.

However, analytical results become more visible with the works of Richard & Steel [8] and Kloek & van Dijk [9]. Also, with the introduction of numerical method, Markov Chain Monte Carlo (MCMC) by Gelfand and Smith [10] further boosted the applications of Bayesian inference in SUR model.

A Bayesian hierarchical SUR model was introduced by Smith and Kohn [11]. In this work, regression functions were represented as a linear combination of large number of terms using Bayesian hierarchical SUR framework. For the purpose of estimation, Markov Chain Monte Carlo (MCMC) procedure was also developed. It was observed that the method performs very well with the aid of both simulated and real data.

Bayesian hierarchical framework where regression function was used as a linear combination of large number of basis function, variance matrix of errors and set of predictors simultaneously in a SUR model was also demonstrated in the work of Ando [12]. In this work, a direct Monte Carlo technique was employed to solve the variable selection and model parameter estimation problems. The developed method was shown to be more computationally efficient than Smith and Kohn [11]'s method.

The SUR model was known to permits error terms in different equations to be correlated. However, many assume that the error terms are normally distributed, but some works have shown that errors can be assumed to have heavy tails distribution. Examples include Ng [15], Kowalski et al [13].

This work hereby presents a simulation study for SUR model with disturbances under different distributional assumptions using Bayesian method of estimation. This Bayesian estimation method will be compared with some classical estimators. In order to facilitate comparison among these estimation methods, the different distributional assumptions for the error terms of the model are; t, Gaussian and F will be utilized while MSE will be used as a performance criterion.

The remainder of the work is as follows. In section2, SUR model was reviewed. Section 3 presents the Bayesian estimation procedures for SUR model. Simulation study was conducted in section 4. For comparative purposes, the performance of the Bayesian and classical method of estimation are presented under different distributional assumptions in Section 5. Section 6 concludes.

## 2 Overview of Seemingly Unrelated Regression (SUR) Model

Consider the Seemingly Unrelated Regression (SUR) model given as:

$$y_{it} = X'_{ti} \alpha_i + \varepsilon_{ti} \tag{1}$$

where  $y_{it}$  is the observation on the  $i$ th dependent variable in equation  $i$ ,  $x_{ti}$  is a  $k$ -vector containing the  $t$ -th observation of the vector of explanatory variables in the  $i$ -th equation and  $\alpha_i$  is a  $k_i$ -vector of regression coefficients for the  $i$ -th equation while  $\varepsilon_{ti}$  is  $t$ -th value of the random error component associated with  $i$ -th of the model.

$$y_1 = X_1 \alpha_1 + \varepsilon_1$$

$$y_2 = X_2 \alpha_2 + \varepsilon_2$$

$$\vdots \quad \quad \quad \vdots$$

$$y_m = X_m \alpha_m + \varepsilon_m$$

The  $m$ -equations can be compactly expressed as:

$$y_i = X_i \alpha + \varepsilon_i \tag{2}$$

Where

$$y_i = (y_{1i}, \dots, y_{mi})'$$

$$X_i = \begin{pmatrix} X'_{1i} & 0 & \dots & 0 \\ 0 & X'_{2i} & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ 0 & \dots & \dots & X'_{mi} \end{pmatrix}$$

$$\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix}, \quad \varepsilon_i = (\varepsilon_{1i}, \varepsilon_{2i}, \dots, \varepsilon_{mi})'$$

And define  $K = \sum_{m=1}^M k_m$

We can as well now stack the observations together to become:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix}, \quad X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{pmatrix}$$

which can also be written as:

$$y = X\alpha + \varepsilon \tag{3}$$

Equation (3) is a familiar linear regression model.

The name of the model implies the equations of the model are independent of one another but have a contemporaneous covariance that is correlated through their error terms Zellner [1].

Therefore, the assumption of the variance-covariance matrix is given as:

$$E(\varepsilon) = 0$$

$$\begin{aligned} E(\varepsilon\varepsilon') &= \Sigma \otimes I_T = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \dots & \sigma_{1M}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \vdots & \sigma_{2M}^2 \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{M1}^2 & \sigma_{M2}^2 & \vdots & \sigma_{MM}^2 \end{pmatrix} \otimes I_T \\ &= \begin{pmatrix} \sigma_{11}^2 I_T & \sigma_{12}^2 I_T & \dots & \sigma_{1M}^2 I_T \\ \sigma_{21}^2 I_T & \sigma_{22}^2 I_T & \vdots & \sigma_{2M}^2 I_T \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{M1}^2 I_T & \sigma_{M2}^2 I_T & \vdots & \sigma_{MM}^2 I_T \end{pmatrix} \\ &= \varphi \end{aligned}$$

where  $\otimes$  is the kronecker product operator

$\varphi$  is  $(M T \times M T)$  positive definite variance covariance matrix.

With the above assumption, it can be seen that  $\varepsilon$  is  $N(0, \varphi)$ .

### 3 Bayesian Estimation Procedure

The likelihood function of SUR model given in (3) can be simply written as:

$$L(y | \alpha, \varphi) \propto \prod_{i=1}^m \frac{|H|}{(2\pi)^{nm/2}} \exp\{-\frac{1}{2} \text{tr } P\varphi^{-1}\} \tag{4}$$

where “tr” denotes the trace of matrix,  $|\varphi|$  is the value of determinant of  $\varphi$  and P is  $m \times m$  matrix given by:

$$P = (r_{ij}) = (y_j - x_{ij}\alpha_j)' (y_j - x_{ij}\alpha_j) \tag{5}$$

Priors can take any form and different forms of priors were used in literature in the estimation of SUR model. These include the use of recursive extended natural conjugate prior by Richard and Steel [8], Normal-Inverse Wishart prior by Akhgari and Gotalizadeh [14].

In this study, independent Normal-Wishart prior will be used and is given by:

$$P(\alpha, H) = P(\alpha) P(H) \tag{6}$$

Where

$$P(\alpha) = f_N(\alpha | \alpha^0, Q^0)$$

Also written as:

$$\frac{1}{(2\pi)^{k/2}} |Q^0|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\alpha - \alpha^0)' (Q^0)^{-1} (\alpha - \alpha^0) \right\} \tag{7}$$

and

$$P(H) = f_W(H | v^0, H^0) \tag{8}$$

This can also be written as

$$\frac{1}{C_W} |H|^{\frac{v^0-N-1}{2}} |\varphi|^{-\frac{v^0}{2}} \exp \left[ -\frac{1}{2} \text{tr}(\varphi^{-1} H) \right] \tag{9}$$

Where

$$C_W = 2^{\frac{v^0 N}{2}} \pi^{\frac{N(N-1)}{2}} \prod_{i=1}^N \Gamma\left(\frac{v^0+1-i}{C_W}\right)$$

N.B:  $f_N$  and  $f_W$  are respectively normal and wishart distribution functions.

Combining equation (4) with equations (7) and (9) and also ignore the terms that does not depends on both parameters  $\alpha$  and  $H$  we obtain:

$$P(\alpha, H | y) = \left\{ \exp \left[ \frac{1}{2} \left\{ \text{tr} P\varphi^{-1} + (\alpha - \alpha^0)' (Q^0)^{-1} (\alpha - \alpha^0) + \text{tr}(\varphi^{-1} H) \right\} \right] |H|^{\frac{v^0-n-1+nm}{2}} \right\} \tag{10}$$

It is observed that (10) does not looks like a well known distribution, we therefore obtain the posterior conditional for parameter  $\alpha$  by treating equation (10) as a function of  $\alpha$  by keeping  $H$  fixed, we thus have:

$$P(\alpha | y, H) = \exp \left\{ \frac{1}{2} (\alpha - \alpha^0)' (Q^0)^{-1} (\alpha - \alpha^0) \right\} \tag{11}$$

Hence, equation (11) is a kernel of multivariate normal distribution which can also be written as:

$$\alpha | y, H \sim N(\alpha^*, Q^*) \tag{12}$$

Also, the posterior for H conditional on  $\alpha$  is:

$$H | y, \alpha \sim W(v^*, H^*) \tag{13}$$

Where

$$Q^* = ((Q^0)^{-1} + \sum_{i=1}^n x_i' H x_i)^{-1}$$

$$v^* = n + v^0$$

$$H^* = [ (H^0)^{-1} + \sum_{i=1}^n (y_i - x_i \alpha)(y_i - x_i \alpha)' ]$$

The Bayesian estimator for  $\alpha$  is:

$$\alpha^* = ((Q^0)^{-1} \alpha^0 + \sum_{i=1}^n x_i' H y_i)^{-1}$$

**N.B:** The notations “0” under the parameters and “\*” over the parameters are the priors and posterior parameters, respectively.

## 4 Simulation Study

In order to compare the properties of Bayesian estimation procedures for SUR model with those of classical estimators, we simulate data sets from  $m=2$ , SUR model. Thus, the model can be simply written as:

$$\begin{aligned} y_1 &= 15 + 50x_{11} + \varepsilon_1 \\ y_2 &= 7 + 80x_{21} + \varepsilon_2 \end{aligned} \tag{14}$$

The regressors were generated from a uniform distribution for various sample sizes of 30, 100 and 700. Each elements of  $\varphi$  is set to be:

$$\varphi = \begin{pmatrix} 0.5 & -0.05 \\ -0.05 & 0.3 \end{pmatrix}$$

The disturbance terms,  $\varepsilon = (\varepsilon_1, \varepsilon_2)'$  are generated from three distributions given below:

Normal	$N(0, k\varphi)$	$N(0, k\varphi)$ , where $k = 1$ and $5$
T	$t(0, \varphi, 3)$	$t(0, \varphi, 5)$
F	$F(2,8)$	$F(5,12)$

The regressors and disturbance terms enabled the generation of response observations.

Number of replications:  $R = 10000$ , Burn-in period:  $R_1 = 1000$ .

In order to investigate the performance of the estimators for different distributional assumption, we will use the Mean Squared Error (MSE) and is given by:

$$MSE = \frac{\sum_{i=1}^R (\hat{\alpha}_i - \alpha_i)^2}{R} \tag{15}$$

### 4.1 Prior Specification

The prior hyperparameters are specified as:

$$v^0 = 4, (H^0)^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \alpha^0 = 0_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$Q^0 = 1000 I_4 = 1000 \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

## 5 Results and Discussion

This Section presents the results obtained from the simulation set-up under Section 4. In the results, MSE of Bayesian estimator is compared with SUR and OLS estimators using Normal, t and F-distributions for different data generation of disturbance terms as shown in Tables 1, 2, and 3. The distributions Normal, t, and F distributions are denoted by  $N(0, k, \varphi)$ ,  $t(0, v, \varphi)$  and  $F(v_1, v_2, 0, \varphi)$  respectively, where  $k = 1$  and  $5$ .

**Table 1. MSE of estimators under different distributional assumptions for sample size of 30**

	Estimators	N(0,1)	N(0,5)	t(0, $\varphi$ , 3)	t(0, $\varphi$ , 5)	F(2,8, $\varphi$ )	F(5,12, $\varphi$ )
<b>Equation 1</b>	Bayesian	0.0003	0.0037	0.0007	0.0008	0.3048	0.0669
	SUR	0.4658	0.8451	0.9451	0.4622	0.7471	0.6813
	OLS	0.4990	0.9054	1.0126	0.4879	0.7941	0.7284
<b>Equation 2</b>	Bayesian	0.0100	0.0037	0.0002	0.0683	0.3683	0.2166
	SUR	0.2236	0.8451	0.2447	0.4016	0.2777	0.5409
	OLS	0.2396	0.2899	0.4191	0.4302	0.2957	0.5792

In Table 1, it is observed that MSE of Bayesian estimator is smaller than all other estimators (SUR and OLS) for all the distributions used. Normal distribution with  $N(0,1)$  gives the smallest MSE for all the estimators. MSE of all the estimators with the use of F-distribution appears to be high.

**Table 2. MSE of estimators under different distributional assumptions for sample size of 100**

	Estimators	N(0,1)	N(0,5)	t(0, $\varphi$ , 3)	t(0, $\varphi$ , 5)	F(2,8, $\varphi$ )	F(5,12, $\varphi$ )
<b>Equation 1</b>	Bayesian	0.0656	0.0069	0.0114	0.0431	1.0839	0.7117
	SUR	0.5491	0.4253	0.8893	0.5508	0.9508	0.3403
	OLS	0.5604	0.4339	0.9074	0.4213	0.9698	0.3473
<b>Equation 2</b>	Bayesian	0.0013	0.0089	0.0044	0.0315	0.7398	0.4194
	SUR	0.3288	0.3566	0.6246	0.2923	1.1481	0.3491
	OLS	0.3355	0.3639	0.6373	0.2980	1.1714	0.3562

**Table 3. MSE of estimators under different distributional assumptions for sample size of 700**

	Estimators	N(0,1)	N(0,5)	t(0, $\varphi$ , 3)	t(0, $\varphi$ , 5)	F(2,8, $\varphi$ )	F(5,12, $\varphi$ )
<b>Equation 1</b>	Bayesian	0.0024	0.0007	0.0004	0.0015	1.0042	0.6947
	SUR	0.4940	0.5510	1.9690	0.8040	1.4089	197.8692
	OLS	0.4954	0.5525	1.9746	0.8063	1.4130	198.4436
<b>Equation 2</b>	Bayesian	0.0034	0.0007	0.0033	0.0004	0.2570	0.3275
	SUR	0.3368	0.2661	1.1079	0.4849	0.6559	0.3304
	OLS	0.3377	0.2668	1.1109	0.4863	0.6577	0.3314

Results from Tables 2 and 3 indicate that MSE of Bayesian estimator is smaller than other estimators for the distributions considered. Normal distribution gave small MSE compared to other distribution for all the estimators. However, it appears that as the sample size increases from 100 to 700, MSE also increases for both SUR and OLS estimators for all the distributions used while the MSE of Bayesian reduces consistently.

## 6 Conclusion

Disturbance terms of seemingly unrelated regression model are assumed to be correlated across equations. However, these disturbances are usually generated from Normal distribution in simulation study. But with the use of heavy tails distribution can increase the efficiency of estimators. In this work, the use of Normal, t, and F distribution were considered for generation of data for disturbances in the simulation study. In order to facilitate comparison between the Bayesian and classical estimators (seemingly unrelated regression and ordinary least squares estimators), we used mean squared error as criterion.

It was observed that Bayesian estimator consistently better than other estimators for all the distributional assumptions considered. The increase in sample sizes improves the efficiency of Bayesian estimator while it does not affect other estimators. The use of Normal distribution in the generation of disturbances gave smallest MSE for all the estimators compared to other distributions. Therefore, it is advised that Normal distribution should be used in simulation study in generation of disturbance terms in seemingly unrelated regression model while Bayesian estimator should be considered in estimation of parameters of seemingly unrelated regression model.

## Competing Interests

Authors have declared that no competing interests exist.

## References

- [1] Zellner A. An efficient method of estimating seemingly unrelated regression equations and test for aggregation bias. *Journal of American Statistical Association*. 1962;57:348-368.
- [2] Zellner A. Estimators for seemingly unrelated regression equations: Some finite sample results. *Journal of the American Statistical Association*. 1963;58:977-992.
- [3] Phillips PCB. An approximation to the finite sample distribution of Zellner's seemingly unrelated regression estimator. *Journal of Econometrics*. 1977;6:147-164.
- [4] Koenker R, Portnoy S. M-estimation of multivariate regressions. *Journal of the American Statistical Association*. 1990;85(412):1060–1068.
- [5] Hubert M, Verdonck T, Yorulmaz O. Fast robust SUR with economical and actuarial applications. *Statistical Analysis and Data Mining*. 2017;10(2):77–88.
- [6] Peremans, K. and S. Van Aelst. 2018. Robust Inference for Seemingly Unrelated Regression Models. *Journal of Multivariate Analysis to Appear*. arXiv:1801.04716.
- [7] Zellner A. An introduction to Bayesian inference in econometrics. New York: Wiley; 1971.
- [8] Richard JF, Steel MFJ. Bayesian analysis of systems of seemingly unrelated regression equations under a recursive extended natural conjugate prior density. *Journal of Econometrics*. 1988;38:7-37.
- [9] Kloek T, van Dijk HK. Bayesian estimates of equation system parameters: an application of integration by Monte Carlo. *Econometrica*. 1978;46:1-19.
- [10] Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*. 1990;85:398-409.
- [11] Smith M, Kohn R. Nonparametric seemingly unrelated regression. *J. Econom*. 2000;98:257–282.
- [12] Ando T. Bayesian variable selection for the seemingly unrelated regression models with a large number of predictors. *J. Japan Statist. Soc*. 2011;41(2):187-203.
- [13] Kowalski JR, Mendoza-Blanco XMT, Gleser LJ. On the difference in inference and prediction between the joint and independent t-error models for seemingly unrelated regressions, *Communications in Statistics. Part A - Theory and Methods*. 1999;28:2119–2140.
- [14] Akhgari O, Golalizadeh M. On Bayesian analysis of seemingly unrelated regression model with skew distribution error. *Statistical Journal*. 2019;1-24.
- [15] Ng VM. Robust Bayesian Inference for Seemingly Unrelated Regressions with Elliptical Errors. *Journal of Multivariate Analysis*. 2002;83:409–414.

© 2020 Oluwadare et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Peer-review history:

The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)

<http://www.sdiarticle4.com/review-history/64433>