Journal of the Nigerian Association of Mathematical Physics
Volume 45 (March, *2018 Issue*), pp181 – 186

DECOMPOSED TRIANGULAR MATRICES ON THE ESTIMATORS OF UNEQUAL OBSERVATIONS OF SEEMINGLY UNRELATED REGRESSION MODEL

Olaomi O. E[1], Okewole D. M.[2] and Alaba O. O.[3,a]

[1]**Department of Statistics, University of Ibadan, Nigeria.**
[2]**Department of Mathematical Sciences, Redeemers University, Ede, Nigeria.**
[3]**Department of Statistics, University of Ibadan, Nigeria.**

Abstract

The Seemingly Unrelated Regression model is a generalisation of a linear regression model that is linked by contemporaneously correlated error terms. This work examined the effects of unequal sample observations on the partitioned triangular matrix in seemingly unrelated regression using the Cholesky method of decomposition. The partitioned upper and lower non-singular triangular matrices was to establish contemporaneous relationship among equations. through their errors. A Monte Carlo experiment was performed on a three-equation model with sample sizes n = 20, 25, 40, 50, 60, 75, 80 and 100 for equal observations and T = 20, 40, 60 and 80, with extra observations E = 5, 10, 15 and 20 respectively for the unequal observations. The Root Mean Square Error (RMSE) was used to adjudge the performance of the estimators when unequal observations are considered. Unequal observations had little or no effect on both the upper and lower triangular matrices in this experiment. It was also observed that RMSE of SUR estimator was lower than that of OLS estimator for both equal and unequal number of observations, irrespective of the triangular matrix used. The upper triangular matrix had higher RMSE values than the lower triangular matrix for SUR and OLS estimators.

Keywords: Extra Observations, Triangular matrices, Root Mean Square Error, Ordinary Least Squares, Seemingly Unrelated Regression.

1.0 Introduction

Seemingly Unrelated Regressions (SUR) is a generalisation of a linear regression model that consists of several regression equations, each having its own dependent variable and potentially different sets of exogenous explanatory variables. The SUR system as proposed in [1] comprises of several individual relationships linked by correlated disturbances. Taking cognizance of such correlation leads to efficient estimates of the coefficients and standard errors. The model can be estimated equation-by-equation using Ordinary Least Squares (OLS). Such estimates are consistent, however generally not as efficient as the SUR estimator, which results to Feasible Generalized Least Squares with a specified variance-covariance matrix. There are two main motivations for use of SUR. When contemporaneous correlations between the disturbances are high and the explanatory variables in different equations considered are uncorrelated, then efficiency would be attained.

Two important cases when SUR is in fact equivalent to OLS are when the error terms are uncorrelated between the equations and when each equation contains exactly the same set of regressors on the right-hand side. The SUR estimator is efficient as it takes into account the covariance structure of the errors. Univariate analysis may result in inefficient estimates of the covariate effects [1, 2].

Corresponding Author: Alaba O.O., Email: oluwayemisioyeronke@yahoo.com, Tel: +2348052172452

The variance-covariance matrix is a Hermittan positive-definite non-singular symmetric matrix usually partitioned by the Cholesky decomposition. The Cholesky decomposition partitions the variance-covariance matrix into an upper and lower triangular matrices. Researchers have worked on equal observations for partitioned variance- covariance matrix [3]. Attempt is made to investigate the scenario when observations are not equal. This work is basically to investigate the effect(s) of unequal sample observations on partitioned triangular matrices in seemingly unrelated regression model. The rest of the paper is organised as follows: Section 2 illustrates the theory behind the methodology followed by the design of the simulation experiment in section 3. Analysis and discussion of results are presented in sections 4 and 5. In Section 6 some concluding remarks are provided.

2. Materials and Methods

The system of 'm' seemingly unrelated regression equations can be stacked in two equivalent compact matrix forms. It can be expressed as a multiple linear regression model:

$$y_i = X_i\beta_i + \varepsilon_i \;, \quad i = 1, 2, \ldots, k \qquad (1)$$

Where

$y = (y_1', \ldots, y_k')'$ is the nk x 1 response vector.

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix}, \quad X = \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_k \end{pmatrix},$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}, \text{ and } \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{pmatrix} \qquad (2)$$

X is the nk x p structured design matrix, with $p = \sum_{j=1}^{m} P_j$, $\beta = (\beta_1', \ldots, \beta_k')'$ is the kp x 1 parameter vector. It is assumed that the error vectors are contemporaneously correlated.

$\varepsilon = (\varepsilon_1', \ldots, \varepsilon_k')'$ is the error vector with

$$E(\varepsilon\varepsilon') = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & X_{kk} \end{pmatrix} \qquad (3)$$

Eq (1) can be estimated separately using the Ordinary Least Squares (OLS) estimator.

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'y \qquad (4)$$

The results will ignore the covariance structure of the errors. A more efficient estimator is the Generalized Least Squares (GLS) estimator, which is used to estimate a k x k covariance matrix of the disturbances.

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}y) \qquad (5)$$

$$= (X'(\hat{\Sigma}^{-1} \otimes I_n)X)^{-1}(X'(\hat{\Sigma}^{-1} \otimes I_n)y)$$

An arbitrary three-equation SUR model with correlated errors and unequal number of observations is specified as follows:

$$\begin{aligned} y_1 &= \beta_{10} + \beta_{11}X_{11} + \beta_{12}X_{12} + \varepsilon_1, && n=1,\ldots,T \\ y_2 &= \beta_{20} + \beta_{21}X_{21} + \beta_{22}X_{22} + \varepsilon_2, && n=1,\ldots,T \qquad (6) \\ y_3 &= \beta_{30} + \beta_{31}X_{31} + \beta_{32}X_{32} + \varepsilon_3, && n=1,\ldots,T+E \end{aligned}$$

The T observations on each of the equations are assumed to match in `time', but there are E extra observations on the third equation. The positive definite 3 x 3 variance-covariance matrix is defined by

$$\Omega = Cov(\varepsilon) = \begin{pmatrix} \sigma_{11}I_T & \sigma_{12}I_T & \sigma_{13}I_T \\ \sigma_{21}I_T & \sigma_{22}I_T & \sigma_{23}I_T \\ \sigma_{31}I_{T+E} & \sigma_{32}I_{T+E} & \sigma_{33}I_{T+E} \end{pmatrix} \qquad (7)$$

Where Ω_{ij} are elements of Σ, and

$$\Omega^{-1} = \Sigma^{-1} \otimes I_T = \begin{pmatrix} \sigma^{11} I_T & \sigma^{12} I_T & \sigma^{13} I_T \\ \sigma^{21} I_T & \sigma^{22} I_T & \sigma^{23} I_T \\ \sigma^{31} I_{T+E} & \sigma^{32} I_{T+E} & \sigma^{33} I_{T+E} \end{pmatrix} \quad (8)$$

The estimator $\hat{\Sigma}$ under study is defined as in [4] as:

Estimator $E_1 = (\hat{\sigma}_{ij})$, which is necessarily non negative definite, given by

$$\hat{\sigma}_{11} = \frac{n_1 s_{111} + n_2 s_{113}}{n_1 + n_3}$$
$$\hat{\sigma}_{12} = \hat{\sigma}_{21} = \hat{s}_{123} \quad (9)$$
$$\hat{\sigma}_{22} = \frac{n_2 s_{22} + n_3 s_{223}}{n_2 + n_3}$$

Estimator $E_2 = \sigma_{ij}{}^*$

$$\sigma^*_{11} = \hat{\sigma}_{11}$$
$$\sigma^*_{22} = \hat{\sigma}_{22} \quad (10)$$
$$r = \sqrt{\hat{\sigma}_{11} \hat{\sigma}_{22}}$$

3. Design of the Simulation Experiment

The specified three-equation model with arbitrary values given for its parameters is as follows:

$$y_1 = 70 + 15X_{11} + 8X_{12} + \varepsilon_1$$
$$y_2 = 60 + 10X_{21} + 7X_{22} + \varepsilon_2 \quad (11)$$
$$y_3 = 65 + 12X_{31} + 10X_{32} + \varepsilon_3$$

Under the assumptions that $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ are *i.i.d* with $N(0,\Sigma)$; where $N = 1, 2, \ldots, T$ for the first and second equations and $N = T + E$ for the third equation. The Monte Carlo method was used to simulate the explanatory variables (X_{11}, X_{12}, X_{21}, X_{22}, X_{31}, X_{32}) from a Uniform distribution for various sample sizes N=20, 25, 40, 50, 60, 75, 80 and 100. The data were generated following the steps below;

The vectors of the X's independent regressors were generated by drawing X_{11}, X_{12}, X_{21}, X_{22}, X_{31}, X_{32} from a Uniform (0, 1) distribution. The $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ are series of random normal deviates which were standardized and appropriately transformed to ensure that the disturbance terms are contemporaneously correlated and have specific variance-covariance matrices $\hat{\Sigma}$ estimated in the model.

From definition, $\hat{\Sigma}$ is a positive definite matrix, therefore, there exists a non- singular triangular matrix P such that

$$\hat{\Sigma} = PP' \quad (12)$$

The estimated variance-covariance matrix $\hat{\Sigma}$ was obtained and later decomposed into upper and lower non-singular triangular matrices such that $\hat{\Sigma} = PP'$

$$PP' = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ 0 & p_{22} & p_{23} \\ 0 & 0 & p_{33} \end{pmatrix} \begin{pmatrix} p_{11} & 0 & 0 \\ p_{21} & p_{22} & 0 \\ p_{31} & p_{32} & p_{33} \end{pmatrix} \quad (13)$$
$$= \begin{pmatrix} 1 & 0.7 & 0.6 \\ 0.7 & 1 & 0.7 \\ 0.6 & 0.7 & 1 \end{pmatrix}$$

The variance-covariance decomposition is

$$P=\begin{pmatrix} 0.697 & 0.392 & 0.6 \\ 0 & 0.714 & 0.7 \\ 0 & 0 & 1 \end{pmatrix} \tag{14}$$

For the upper triangular matrix, and;

$$P'=\begin{pmatrix} 0.697 & 0 & 0 \\ 0.392 & 0.714 & 0 \\ 0.6 & 0.7 & 1 \end{pmatrix} \tag{15}$$

For the lower triangular matrix.

Strong contemporaneous relationship among the three sets of equations through the error terms, was established by,

$$U = P^* \varepsilon = (\varepsilon_1^*, \varepsilon_2^*, \varepsilon_3^*); \tag{16}$$

Such that for upper triangular matrix we have

$$\mathrm{U} = \mathrm{P}\ {*}\ \varepsilon = \begin{pmatrix} 0.697\varepsilon_1 + 0.392\varepsilon_2 + 0.6\varepsilon_3 \\ 0.714\varepsilon_2 + 0.7\varepsilon_3 \\ \varepsilon_3 \end{pmatrix} \tag{17}$$

And for lower triangular matrix, we have;

$$\mathrm{U} = \mathrm{P'}\ {*}\ \varepsilon = \begin{pmatrix} 0.697\varepsilon_1 \\ 0.392\varepsilon_1 + 0.714\varepsilon_2 \\ 0.6\varepsilon_1 + 0.7\varepsilon_2 + \varepsilon_3 \end{pmatrix} \tag{18}$$

Since certain values have been assigned to the structural parameters, we pre-determined the coefficients in the Monte Carlo experiment. The experiment was carried out for each of the sample sizes. Programs were executed using the STATA software package following the above procedures. The known parameters were then estimated for both equal and unequal number of observations using SUR and OLS estimation methods. The performances of both estimators were compared.

4.0 Discussion of Results

The results of the Root Mean Square Error (RMSE) and R-squared (R^2) of OLS and SUR estimators were presented using the two triangular matrices. For equal number of observations, the RMSE values of upper and lower triangular matrices for SUR when n=25, are: 1.849, 1.4631, 0.8821; and 0.8920, 1.2629 and 2.5207 while OLS estimators are given as 1.8639, 1.5367, 1.1556 and 0.8939, 1.3186 and 2.6439 respectively. When n=50, the SUR RMSE gave: 1.6995, 1.4744, 1.1372 and 0.7366, 1.1504 and 2.3482, while OLS estimators gave: 1.7447, 1.4942, 1.1505, and 0.755, 1.1747 and 2.3615 for upper and lower triangular matrices respectively. When n=75, the SUR RMSE values of upper and lower triangular matrices are: 1.4858, 1.2744, 0.9076 and 0.7319, 1.0971 and 2.0124 respectively; while OLS estimators are given as 0.7282, 1.1176, and 2.0286. When n=100, the SUR RMSE values for upper and lower triangular matrices are 1.5723, 1.3831, 1.0843 and 0.7009, 1.007 and 2.1685 respectively, while OLS estimators gave 1.5927, 1.4027, 1.0849, and 0.7114, 1.0938 and 2.1578.

Also, for unequal number of observations, where N=T+E the RMSE values of upper and lower triangular matrices for SUR when T=20 and E=5, are: 1.9263, 1.5575, 1.1128; and 0.9435, 1.3707 and 2.6605 while OLS estimators gave 1.9271, 1.6660, 1.1556 and 0.9482, 1.4678 and 2.8363 respectively. When T=40 and E=10, the SUR RMSE gave: 1.8305, 1.5703, 1.2140 and 0.7939, 1.2287 and 2.5316, while OLS estimators gave: 1.8305, 1.5703, 1.2140, and 0.8168, 1.2610 and 2.5353 for upper and lower triangular matrices respectively. When N=60 and E=15, the SUR RMSE values of upper and lower triangular matrices are: 1.5547, 1.3114, 0.9334 and 0.7447, 1.1172 and 2.0890 respectively; while OLS estimators are given as 1.5809, 1.3443, and 0.9536. When T=80 and E=20, the SUR RMSE values for upper and lower triangular matrices are 1.5628, 1.3569, 1.0454 and 0.7060, 1.0848 and 2.1237 respectively, while OLS estimators gave 1.5861, 1.3796, 1.0849, and 0.7185, 1.1039 and 2.1578.

Table 1: Upper Triangular Matrix with Equal Observations

		RMSE			R^2		
N	Estimators	y_1	y_2	y_3	y_1	y_2	y_3
25	SUR	1.8495	1.4631	0.8821	0.8770	0.8821	0.9471
	OLS	1.8639	1.5367	1.1556	0.8901	0.8856	0.9473
50	SUR	1.6995	1.4744	1.1372	0.8605	0.8880	0.9047
	OLS	1.7447	1.4942	1.1505	0.8619	0.8918	0.9083
75	SUR	1.4858	1.2744	0.9076	0.9151	0.8844	0.9601
	OLS	1.5129	1.2920	0.9175	0.9155	0.8859	0.9608
100	SUR	1.5723	1.3831	1.0843	0.9061	0.8534	0.9514
	OLS	1.5927	1.4207	1.0849	0.9066	0.8538	0.9528

Table 2: Lower Triangular Matrix with Equal Observations

		RMSE			R^2		
N	Estimators	y_1	y_2	y_3	y_1	y_2	y_3
25	SUR	0.8920	1.2629	2.5207	0.9644	0.9089	0.7865
	OLS	0.8939	1.3186	2.6439	0.9686	0.9126	0.7925
50	SUR	0.7366	1.1504	2.3482	0.9716	0.9292	0.6884
	OLS	0.7550	1.1747	2.3615	0.9719	0.9306	0.7038
75	SUR	0.7139	1.0971	2.0124	0.9784	0.9134	0.8381
	OLS	0.7282	1.1176	2.0286	0.9784	0.9138	0.8421
100	SUR	0.7009	1.0779	2.1685	0.9795	0.9086	0.8444
	OLS	0.7114	1.0938	2.1578	0.9795	0.9087	0.8506

Table 3: Upper Triangular Matrix with Unequal Observations

		RMSE			R^2		
N	Estimators	y_1	y_2	y_3	y_1	y_2	y_3
T = 20	SUR	1.9263	1.5575	1.1128	0.8745	0.8856	0.9468
E = 5	OLS	1.9271	1.6660	1.1556	0.8933	0.8887	0.9473
T = 40	SUR	1.8305	1.5703	1.2140	0.8284	0.8845	0.8832
E = 10	OLS	1.8826	1.6122	1.2364	0.8272	0.8874	0.9083
T = 60	SUR	1.5547	1.3114	0.9334	0.9121	0.8731	0.9549
E = 15	OLS	1.5809	1.3443	0.9536	0.9137	0.8734	0.9552
T = 80	SUR	1.5628	1.3569	1.0454	0.9084	0.8578	0.9554
E = 20	OLS	1.5861	1.3796	1.0849	0.9092	0.8585	0.9576

Table 4: Lower Triangular Matrix with Unequal Observations

N	Estimators	RMSE y_1	RMSE y_2	RMSE y_3	R^2 y_1	R^2 y_2	R^2 y_3
T = 20 E = 5	SUR	0.9435	1.3707	2.6605	0.9602	0.9058	0.7765
	OLS	0.9482	1.4678	2.8363	0.9659	0.9082	0.7841
T = 40 E = 10	SUR	0.7939	1.2287	2.5316	0.9642	0.9259	0.6292
	OLS	0.8168	1.2610	2.5353	0.9649	0.9278	0.6560
T = 60 E = 15	SUR	0.7447	1.1172	2.0890	0.9773	0.9068	0.8179
	OLS	0.7622	1.1426	2.1289	0.9774	0.9070	0.8203
T = 80 E = 20	SUR	0.7060	1.0848	2.1237	0.9796	0.9083	0.8532
	OLS	0.7185	1.1039	2.1578	0.9797	0.9086	0.8506

5.0 Summary of Results

It was observed that for equal and unequal sample observations, the RMSE of the decomposed lower triangular matrix are lower than that of the decomposed upper triangular matrix for both SUR and OLS estimators except for equation y_3. The SUR estimators were more efficient than OLS estimators. For both the decomposed upper and lower triangular matrix, the unequal sample observations had little or no effect on SUR estimator. Also, it was observed as expected that as the sample size increases, the RMSE values reduced considerably for both equal and unequal number of observations in the decomposed upper and lower triangular matrices.

6.0 Conclusion

The main focus of this paper is to examine the effects of unequal sample observations on decomposed triangular matrices in seemingly unrelated regression. This study found out that the unequal sample observations had no notable effect on both the upper and lower triangular matrices in the estimation of seemingly unrelated regressions. We also found out that the lower triangular matrix of the decomposed variance-covariance matrix performed better than that of the upper triangular matrix.

References

[1]. Zellner A. (1962): An Efficient Method of Estimating Seemingly Unrelated Regression Equations and Tests of Aggregation Bias, *Journal of the American Statistical Association*, 57 (298): 348-368; 500-509.

[2]. Bartels, R. and Fiebig, D.G. (1991): A Simple Characterization of Seemingly Unrelated Regressions Model in which OLS is BLUE. The American Statistician 45(2): 137-140.

[3]. Alaba O.O., Olubusoye O.E., and Oyebisi O.O (2013): Cholesky Decomposition of Variance-Covariance Matrix Effect on the Estimators of Seemingly Unrelated Regression Model; *Journal of Science Research* Vol 12: 371-380.

[4]. Srivastava J. N. and Zaatar M. K. (1973): A Monte Carlo Comparison of Four Estimators of the Dispersion Matrix of a Bivariate Normal Population, Using Incomplete Data, *Journal of the American Statistical Association*, 68:341, 180-183.