Tests for Aggregation Bias in Seemingly Unrelated Regression with Unequal Observations

Alaba, O. & Akinrelere, O.
Department of Statistics
University of Ibadan
Ibadan, Nigeria
E-mails: oluwayemisioyeronke@yahoo.com, akinrelereoyinlade@gmail.com

ABSTRACT

The presence of aggregation bias in seemingly unrelated regression with unequal number of observations was considered. The cholesky method of decomposition was used to partition the variance-covariance matrix into the upper and lower triangular matrices. A Monte Carlo experiment was performed on a two-equation model with sample sizes n = 20, 40, 60 and 80 with extra observations of E = 5, 10, 15 and 20 respectively for the unequal observations. It was observed that the RMSE of SUR estimator is lower than that of the OLS estimator for both equal and unequal observations on the two triangular matrices. The coefficient of the weighted predictors is not equal to zero, at different number of observations considered, which implies the presence of aggregation bias in the system of equations.

Keywords: Seemingly Unrelated Regression, Aggregation Bias, Unequal observations, Cholesky Decomposition.

iSTEAMS Proceedings Reference Format
Alaba, O. & Akinrelere, O. (2019): Tests for Aggregation Bias in Seemingly Unrelated Regression with Unequal Observations. Proceedings of the 18th iSTEAMS Multidisciplinary Cross-Border Conference, University of Ghana, Legon, Accra, Ghana. 28th – 30th July, 2019
Pp 239-248 www.isteams.net - DOI Affix - https://doi.org/ 10.22624/AIMS/iSTEAMS-2019/V18N1P27

1. INTRODUCTION

A multiple regression describes the behaviour of variable based on set of ex- planatory variables. In a system of linear multiple regression equations, each equation illustrate some economics situation. To examine a system of simul- taneous equation model in which one or more of the explanatory variables are endogenous. In situation where none of the variables in the system are simultaneous, there may be interactions between the individual equations if the random error components are related with equations correlated to each other. The equations may be linked through the jointness of the distribution of the error terms, such behaviour reflect the Seeming Unrelated Regression equations Davidson and Mackinnon, (1993). The SUR proposed by Zellner, (1962) is a generalisation of linear regres- sion model consisting of several regression model, each having its own depen- dent variable and potentially different sets of exogenous variables.

Modelling the relationship between individual behaviour and aggregate statistic from both levels can be used for parameter estimation. The use of aggregation structure is to examine the micro econometric estimation problems. Aggre- gation implies the link between the economics interactions at the micro and macro levels, which is the expected difference between effects for group and the individual. If there is no confounding then the difference is a combination of confounding and aggregation bias. Greene, (2003). Schmidt, (1977) investigated estimation of SUR with unequal number of observations. He opined that except when the disturbances are very highly correlated across equations, there does not seem to be much of an advantage in using the extra observations to estimate.

Consistency and efficiency of SUR tested in variance to the OLS estimator in the presence of atypical ob- servation (outlier) at varying percentage interval of outliers, it was discovered that the SUR estimator gives a better performance than the OLS estimator. The asymptotic efficiency of SUR estimator was maintained as the sample size increased. Adepoju and Akinwumi, (2017). Bogoev and Sergi, (2012) investigated whether there are heterogeneities and asymmetries in the size and speed of the adjustment of lending rates to changes in the cost of the funds rate. They reported the presence of aggre- gation bias implying that the empirical studies based on aggregation data may provide biased results. The linearly aggregated demand functions are subjected to aggregation bias if aggregate demand is a function of the distri- bution of the expenditure across consumers as well as the level of aggregate expenditures, Deaton and Muellbauer, (1980)

The data availability for estimating multi-equation models are often in- complete in the sense that some equations have observations for a longer period than others. The study test for aggregation bias in a SUR model with unequal sample observation through the decomposition of its variance- covariance matrix into upper and lower triangular matrices. The rest of the paper is organized as follows: Section 2 presented the methodology, followed by result presentation presented in section 3. Section 4 presented the discussion and conclusion presented in section 5.

2. METHODOLOGY

2.1 Model

Consider a system of regression equation models:

$$Y_i = X_i\beta_i + \varepsilon_i \quad 1 = 1, 2, \cdots \tag{1}$$

The system in equation 1 can be written as:

$$\begin{array}{ccccc} y_1 & = & X_1\beta_1 & + & \varepsilon_1 \\ y_2 & = & X_2\beta_2 & + & \varepsilon_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y_N & = & X_N\beta_N & + & \varepsilon_N \end{array} \tag{2}$$

where y_i is $N \times 1$ vector of observation on the i^{th} dependent variables, X_i is a $N \times K$ matrix of explanatory variables, β_i is a $K \times 1$ vector of regression parameters and ε_i is $N \times 1$ vector of random error components.
The system in equation 2 can be written in matrix form as:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & \cdots & 0 \\ 0 & X_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ X_{N1} & X_{N2} & \cdots & X_{NN} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix} \tag{3}$$

The disturbance vector in equation 1 is assumed to have the following variance-covariance matrix:

$$\Sigma = V(\varepsilon_i) = \begin{pmatrix} \sigma_{11}I & \sigma_{12}I & \cdots & \sigma_{1N}I \\ \sigma_{21}I & \sigma_{22}I & \cdots & \sigma_{2N}I \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{N1}I & \sigma_{N2}I & \cdots & \sigma_{NN}I \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_{NN} \end{pmatrix} \otimes I \tag{4}$$

$$= \Sigma \otimes I_T \tag{5}$$

2.2 Seeming Unrelated Regression with Unequal Observation

Consider a set of two regression models in equation 1 which can be written in stacked form as:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \tag{6}$$

There are N observations on the first equation and N + E observations on the second equation. N is assumed to be in time and E is the extra observations on the second equation. Based on this, Y_1 will be of the dimension $N \times 1$ and Y_2 will be of the dimensional matrix $(N + E) \times 1$. The variance-covariance matrix is given as:

$$\begin{array}{ccccc} y_1 & = & X_1\beta_1 & + & \varepsilon_1 \\ y_2 & = & X_2\beta_2 & + & \varepsilon_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y_N & = & X_N\beta_N & + & \varepsilon_N \end{array} \tag{2}$$

where y_i is $N \times 1$ vector of observation on the i^{th} dependent variables, X_i is a $N \times K$ matrix of explanatory variables, β_i is a $K \times 1$ vector of regression parameters and ε_i is $N \times 1$ vector of random error components.
The system in equation 2 can be written in matrix form as:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & \cdots & 0 \\ 0 & X_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ X_{N1} & X_{N2} & \cdots & X_{NN} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix} \tag{3}$$

The disturbance vector in equation 1 is assumed to have the following variance-covariance matrix:

$$\Sigma = V(\varepsilon_i) = \begin{pmatrix} \sigma_{11}I & \sigma_{12}I & \cdots & \sigma_{1N}I \\ \sigma_{21}I & \sigma_{22}I & \cdots & \sigma_{2N}I \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{N1}I & \sigma_{N2}I & \cdots & \sigma_{NN}I \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_{NN} \end{pmatrix} \otimes I \tag{4}$$

$$= \Sigma \otimes I_T \tag{5}$$

2.2 Seeming Unrelated Regression with Unequal Observation

Consider a set of two regression models in equation 1 which can be written in stacked form as:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \tag{6}$$

There are N observations on the first equation and N + E observations on the second equation. N is assumed to be in time and E is the extra observations on the second equation. Based on this, Y_1 will be of the dimension $N \times 1$ and Y_2 will be of the dimensional matrix $(N + E) \times 1$. The variance-covariance matrix is given as:

γ is the weighted average of ($\beta_{11} + \beta_{12}$) and ($\beta_{21} + \beta_{22}$) with weights:

$$w_1(t) = \left[\frac{x_{11}(t) + x_{12}(t)}{x_{11}(t) + x_{12}(t) + x_{21}(t) + x_{22}(t)}\right] \tag{14}$$

and

$$1 - w_1(t) = \left[\frac{x_{21}(t) + x_{22}(t)}{x_{11}(t) + x_{12}(t) + x_{21}(t) + x_{22}(t)}\right] \tag{15}$$

Then introduced the weights in equations 14 and 15 into equation 13, if the coefficient of the weighted predictors is equal to zero, it means that no aggregation bias is present

2.4 Simulation Study

The study considered a system of SUR equations having two distinct linear equations:

$$\begin{aligned} Y_1 &= 45 + 35X_{11} + 15X_{12} + U_1 \\ Y_2 &= 30 + 20X_{21} + 40X_{22} + U_2 \end{aligned} \tag{16}$$

Definite positive variance-covariance matrix considered is defined as follows:

$$\Sigma_{2\times 2} = \begin{bmatrix} \sigma_{12} & \sigma_{11} \\ \sigma_{22} & \sigma_{21} \end{bmatrix} = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix} \tag{17}$$

Decomposing the variance-covariance matrix in equation 17, we have:

$$\Sigma = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix} = \begin{bmatrix} 0.714 & 0.7 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.714 & 0 \\ 0.7 & 1 \end{bmatrix} \tag{18}$$

The random series for the upper triangular matrix is:

$$\begin{aligned} \varepsilon_{1t} &= 0.714U_1 + 0.7U_2 \\ \varepsilon_{2t} &= U_2 \end{aligned} \tag{19}$$

while the random disturbance series for lower triangular matrix is:

$$\begin{aligned} \varepsilon_{1t} &= 0.714U_1 \\ \varepsilon_{2t} &= 0.7U_1 + U_2 \end{aligned} \tag{20}$$

The vectors of the explanatory variables were generated from uniform distribution U(0,1) and error term generated from standard normal distribution N(0,1) for sample sizes 20, 40, 60 and 80 replicated 10000 times. Four different number of extra observations were used on the second model of equation 16, that is E = 5, 10, 15, and 20.

3. RESULT PRESENTATION

The results of the Root Mean Square Error (RMSE), R2, Absolute bias, standard error and Probability value of SUR and OLS estimators were pre- sented using the two triangular matrices shown in Tables 1 to 5. The results presented the lower and upper triangular matrices for unequal number of observations and a test for aggregation. Table 1 shows the RMSE and R2 re- sults with unequal observations for the lower triangular matrix. The RMSE of SUR and OLS estimator when n = 20 at y1 are: 0.6925 and 0.6980, while at y2 are: 1.5574 and 1.5889 respectively. The result reported that the RMSE and R2 of the SUR estimator is more efficient than the OLS estimator for the two model across the sample size considered. It is also observed that the RMSE and R2 of y1 is lower compared to y2 with extra observations for each of the sample size considered.

Table 2 shows the upper triangular matrix of RMSE and R2 value of SUR and OLS estimators for each of the sample size considered. When n = 20 at y1 the RMSE value of SUR and OLS are: 1.2926 and 0.9854 while at y2 the RMSE value are 0.9959 and 0.9940 respectively. It is observed that at y1 the RMSE and R2 value is greater than the RMSE and R2 value of y2 except at sample size 80 where we had a reverse order. Table 3 shows the parameter estimate with unequal observations at lower triangular matrix. When n = 25, 50, 75 and 100 the standard error of β10 are: 0.2115, 0.1737, 0.1295 and 0.1210 respectively. At β22, the standard error are: 0.8759, 0.7084, 0.5956 and 0.5601 respectively. It is observed that as the sample observations increases, the standard error decreases. Table 4 shows the parameter estimate with unequal observations at upper triangular matrix. when n = 25, 50, 75 and 100 the standard error of β10 are: 0.2115, 0.2667, 0.2209 and 0.2119 respectively. At β22, the standard error are: 0.6583, 0.4652, 0.3777 and 0.3554 respectively. It is observed that as the sample sizes increases, the standard error decreases inconsistently. Table 5 shows the test for aggregation bias that the coefficients of the weighted predictors are not equal to zero (That is β11 = β12 = β21 = β22) at different number of observations considered. This suggested the presence of aggregation bias in the system of equations.

Table 1: Simulated Result of RMSE with Unequal Observations (Lower Tri- angular Matrix)

	Sample size	**SUR**		**OLS**	
Eqn	**N**	**RMSE**	R^2	**RMSE**	R^2
y1	T = 20	0.6925	0.9956	0.6980	0.9957
y2	T = 20, E = 5	1.5574	0.9848	1.5889	0.9853
y1	T = 40	0.6496	0.9965	0.6657	0.9965
y2	T = 40, E = 10	1.5102	0.9842	0.9869	0.9865
y1	T = 60	0.5410	0.9974	0.5526	0.9975
y2	T = 60, E = 15	1.1748	0.9929	1.4410	0.9902
y1	T = 80	0.4189	0.9986	0.4486	0.9986
y2	T = 80, E = 20	1.0417	0.9943	1.1493	0.9931

Table 2: Simulated Result of RMSE with Unequal Observations (Upper Tri- angular Matrix)

	Sample size	**SUR**		**OLS**	
Eqn	**N**	**RMSE**	R^2	**RMSE**	R^2
y1	T = 20	1.2926	0.9851	1.3048	0.9854
y2	T = 20, E = 5	0.9959	0.9937	1.008	0.9940
y1	T = 40	1.2416	0.9872	1.2706	0.9873
y2	T = 40, E = 10	0.9826	0.9932	0.9682	0.9946
y1	T = 60	0.9471	0.9922	0.9682	0.9925
y2	T = 60, E = 15	0.8198	0.9965	0.9462	0.9957
y1	T = 80	0.8491	0.9942	0.9062	0.9944
y2	T = 80, E = 20	0.8728	0.9959	0.8638	0.9960

4. DISCUSSION

The lower triangular matrix performed better than the upper triangular ma- trix as the RMSE of the lower triangular matrix were generally smaller than that of the upper triangular matrix. Alaba *et al*, (2013). The study shows that there was gain in efficiency in the SUR estimator as it performed better than the OLS estimators. It is shown that the standard error decreases as the sample size increases. Adepoju and Akinwumi, (2017) It is observed that y_1 performed better in terms of efficiency for the lower

Table 3: Simulated Result of Parameter Estimate with Unequal Observations (Lower Triangular Matrix)

	SUR						OLS					
N = T + E = 25												
Parameters	β10	β11	β12	β20	β21	β22	β10	β11	β12	β20	β21	β22
Estimate	45.0461	34.7155	15.319	29.7183	20.4187	40.8514	45.0080	34.5963	15.5212	29.3894	20.5401	41.1781
ABIAS	0.0461	0.2845	0.3190	0.2818	0.4187	0.8514	0.0080	0.4037	0.5212	0.6106	0.5401	1.1701
Std. Error	0.2115	0.2738	0.2891	0.4581	0.7154	0.6859	0.2554	0.3416	0.3563	0.5171	0.7936	0.8759
N = T + E = 50												
Estimate	44.8999	35.0761	15.1021	29.9544	19.6852	40.1844	44.6893	35.4205	15.2144	29.4829	21.1692	39.8975
ABIAS	0.1000	0.07614	0.1021	0.0456	0.3148	0.1844	0.3108	0.4205	0.2144	0.5171	1.1692	0.1025
Std. Error	0.1737	0.1949	0.2062	0.3531	0.5278	0.4283	0.2466	0.3006	0.3148	0.5221	0.8566	0.7084
N = T + E = 75												
Estimate	44.9568	35.0904	15.0655	30.2573	19.7299	39.5897	45.0317	35.0401	14.9701	30.0935	19.9380	39.5254
ABIAS	0.0433	0.0904	0.0655	0.2573	0.2701	0.4103	0.0317	0.0401	0.0299	0.0935	0.0620	0.4747
Std. Error	0.1295	0.1406	0.1469	0.2895	0.3157	0.3187	0.2358	0.3113	0.3250	0.4240	0.5957	0.5956
N = T + E = 100												
Estimate	44.9691	35.0609	14.9457	29.7172	20.0025	40.3090	44.7168	35.4017	15.1073	29.4370	20.0958	40.8756
ABIAS	0.0309	0.0609	0.0543	0.2828	0.0025	0.3090	0.2832	0.4017	0.1073	0.5630	0.0958	0.8756
Std. Error	0.1210	0.1445	0.1255	0.2646	0.2832	0.3028	0.2090	0.2994	0.2596	0.4187	0.5506	0.5601

Table 4: Simulated Result of Parameter Estimate with Unequal Observations (Upper Triangular Matrix)

	SUR						OLS					
N = T + E = 25												
Parameters	β10	β11	β12	β20	β21	β22	β10	β11	β12	β20	β21	β22
Estimate	45.1807	34.5949	15.4224	30.1698	19.8516	40.1882	45.0683	34.3536	15.9046	29.5959	20.2452	40.9584
ABIAS	0.1801	0.40511	0.4224	0.1698	0.1484	0.1882	0.0683	0.6464	0.9046	0.4041	0.2452	0.9584
Std. Error	0.2115	0.4071	0.4375	0.4581	0.4344	0.4108	0.5159	0.6899	0.7197	0.3886	0.5965	0.6583
N = T + E = 50												
Estimate	44.6879	35.2049	15.2049	30.0167	19.7735	40.0062	44.3470	35.8809	15.3577	29.8357	20.4563	39.8209
ABIAS	0.3121	0.2049	0.2049	0.0167	0.2265	0.0062	0.6530	0.8809	0.3577	0.1643	0.4563	0.1791
Std. Error	0.2667	0.2841	0.3013	0.2167	0.3072	0.2483	0.4322	0.5266	0.5515	0.3428	0.5625	0.4652
N = T + E = 75												
Estimate	44.874	35.1303	15.0718	30.0944	19.8158	39.7873	45.0994	34.9351	14.8322	30.0515	19.8944	39.6383
ABIAS	0.1260	0.1303	0.0718	0.0944	0.1842	0.2127	0.0994	0.0649	0.1678	0.0515	0.1056	0.3617
Std. Error	0.2209	0.2172	0.2269	0.1699	0.1667	0.1682	0.4500	0.5941	0.6203	0.2688	0.3777	0.3777
N = T + E = 100												
Estimate	44.8220	35.1797	14.9626	29.8216	20.0425	40.1017	44.3768	35.7359	15.2990	29.5481	20.1045	40.6723
ABIAS	0.1779	0.1797	0.0374	0.1785	0.0425	0.1017	0.6234	0.7359	0.2990	0.4519	0.1045	0.6723
Std. Error	0.2119	0.2410	0.2093	0.1583	0.1609	0.1720	0.3907	0.5596	0.4853	0.2656	0.3493	0.3554

triangular matrix while equation y_2 performed better for the upper triangular matrix as a result of the decomposition.

Table 5: Simulated Result of Test for Aggregation Bias

N		Lower Triangular Matrix				Upper Triangular Matrix			
	Parameter	Coefficient	Std. Error	t-statistic	P-value	Coefficient	Std. Error	t-statistic	P-value
T = 20, E = 5	Intercept	73.6468	1.0152	72.5473	1.1530e-23	73.6737	1.2045	61.1649	2.46e-22
	Weighted X	1.0168	0.00981	103.5985	1.9193e-26	1.0172	0.0116	87.3469	4.12e-25
T = 40, E = 10	Intercept	73.6737	1.2045	61.1649	2.46e-22	73.5048	1.0349	71.0284	5.13e-42
	Weighted X	1.0172	0.0116	87.3469	4.12e-25	1.0121	0.0091	111.1125	2.3e-49
T = 60, E = 15	Intercept	74.9964	1.0360	72.3923	1.44e-58	75.3043	1.0742	70.1055	9.1e-58
	Weighted X	0.9999	0.0094	105.9661	4.31e-68	0.9962	0.0098	101.8096	4.33e-67
T = 80, E = 20	Intercept	73.8605	0.8870	83.2733	5.75e-78	73.6978	0.9030	81.6150	2.71e-77
	Weighted X	1.0093	0.0081	124.6182	1.61e-91	1.0104	0.0082	122.5321	5.97e-91

5. CONCLUSION

There is presence of aggregation bias in SUR model with unequal numbers of observations. The RMSE value of the lower triangular matrix is smaller compared with the upper triangular matrix of the decomposed variance- covariance matrix. The SUR estimator is efficient than the OLS estimator

REFERENCES

1. Adepoju A.A., and Akinwumi A.O., (2017). Effects of Atypical Observations on the Estimation of Seemingly Unrelated Regression Model. Journal of Mathematical Sciences and Applications, 5(2) 30-35
2. Alaba O.O., Olubusoye O.E., and Oyebisi O.O., (2013). Cholesky Decompo- sition of Variance-Covariance Matrix Effect on the Estimators of Seemingly Unrelated Regression Model. Journal of Science Research, 12, 371-380
3. Bogoev, J. and Sergi, S. B. (2012). Investigating Aggregation Bias in the case of the Interest Rate Pass-through. Intereconomics, 47, 361
4. Davidson, R. and MacKinnon, J. (1993). Estimation and Inference in Econo- metrics, Oxford University Press.
5. Deaton, A., & Muellbauer, J., (1980). Economics and Consumer Behaviour. New York: Cambridge University Press.
6. Greene, W.H. (2003). Econometric Analysis. 5th Edition Upper Saddle River. New Jersey, 339 - 360.
7. Gujarati, D.N. (1995). Basic Econometrics. Third Edition, McGraw Hill Pp 497- 499
8. Schmidt, P., (1977). Estimation of a Linear SUR Model with Unequal Num- bers of Observations. J. Econometrics, 5, 365 - 377
9. Zellner, A. (1962). An efficient Method of Estimating Seemingly Unrelated Regression Equations and Test for Aggregation Bias. Journal of the Ameriican Statistical Association, 57, 348-368.