Journal of Modern Mathematics and Statistics 5 (1): 9-12, 2011
ISSN: 1994-5388

Effect of Multicollinearity and the Sensitivity of the Estimation Methods in Simultaneous Equation Model

D.A. Agunbiade
Department of Mathematical and Computer Sciences,
Fountain University, Osogbo, Osun State, Nigeria

Abstract: This study has investigated the effect of multicollinearity and the sensitivity of three different estimation methods of a Simultaneous equation econometric model. A Monte Carlo approach was employed for a three equation just identified model. The anlaysis for sample size n = 100 replication, R = 200 revealed the preference order of 3SLS, 2SLS and OLS using both mean and bias estimation criteria.

Key words: Multicollinearity, Monte Carlo, Simultaneous equation model, estimation methods, Nigeria

INTRODUCTION

Multicillinearity is one of the known associated problems with the best linear regression model. When the measured explanatory variables are too highly inter-corelated to allow precise analysis of their individual effects, we have multicollinearity. This problem which is peculiar in single equation estimation is addressed here by the introduction of simultaneous equation model. However, this problem may still exist in individual equations of the Simultaneous equation model.

To this end, we have employed a Monte Carlo approach to under study the comparative performance of three estimation methods under various degrees of correlation among the three specified exogenous variables.

The use of Monte Carlo approach has been supported by Wagnar (1958), Parker (1972), Hendry (1976), Johnston (1984) and Kutsoyiannis (2001), the approach has found extensive use in the fields of operational research, nuclear physics and a host of others where there are varieties of problem beyond the available tools of theoretical mathematics.

Studies on estimation under multicollinearity effects of simultaneous models revealed that a high degree of multicollinearity among the explanatory variables has a disastrous effect on estimation of the coefficients, β by the OLS (Mishra, 2004). This method was considered by Pleli and Tankovic (2005) as naive approach because the estimators are biased and inconsistent (2SLS, ILS) and full-information approach (3SLS, FIML). Adenomon and Fesojaiye (2008) merely compared the Seemingly Unrelated Regression (SUR) with the OLS technique and confirmed the superiority of the SUR estimator to the OLS estimators. In the opinion of Ayinde (2007), where he compared OLS with some GLS estimators, he observed that with increasing replications OLS estimator is preferred in estimating all the model parameters at all levels of correlation. However, this opinion negates that of Pleli and Tankovic (2005) in which they advised an econometrician to avoid the use of naive approach (OLS) in estimating the parameters of a system of simultaneous equations.

MATERIALS AND METHODS

Presentation of the model design: The following model with three structural equations are designed and presented thus:

$$\begin{aligned} y_{1t} &= \beta_{13}y_{3t} + \gamma_{11}x_{2t} + \gamma_{12}x_{2t} + u_{1t} \\ y_{2t} &= \beta_{21}y_{1t} + \gamma_{21}x_{1t} + \gamma_{23}x_{3t} + u_{2t} \\ y_{3t} &= \beta_{32}y_{2t} + \gamma_{31}x_{1t} + \gamma_{32}x_{2t} + u_{3t} \end{aligned} \quad (1)$$

Where, y_{1t}-y_{3t} are the endogenous variables and x_{1t}-x_{3t} are the exogenous or predetermined variables. The u_{1t}-u_{3t} are the random disturbance terms which are assumed to be independently and identically normally distributed with zero means and finite matrix, i.e., U NID $(0, \sum)$.

Procedure for data generation: In order to ensure that the data conform to all the specifications of the econometric model, the data series are generated as follows:

- The vector $[(x_{1t}$-$x_{3t}),\ t = 1, 2, \ldots, N)]$ of independent or exogenous variables are randomly and uniformly generated using Kmenta and Joseph (1963) and Kmenta (1971) to ensure they conform to the following scenario:

   - When the exogenous variables (x_{1t}-x_{3t}) selected are relatively highly negatively correlated ($\rho x_i x_j$<-0.05). This is considered as the incidence of high negative multicollinearity and so referred to as category I
   - When the exogenous variables (x_{1t}-x_{3t}) selected are feebly negatively or positively correlated ($-0.05<\rho x_i x_j \leq +0.05$). These are considered as the incidence of feebly negative or positive multicollinearity and henceforth referred to as category II
   - When the exogenous variables (x_{1t}-x_{3t}) selected are relatively highly positively correlated ($\rho x_i x_j$>+0.05). This is considered as the incidence of high positive multicollinearity and thus referred to as category III
- Three corresponding mutually independent N(0, 1) sequences say $\{(\xi_{1t}-\xi_{3t})$: t = 1, 2, …, T$\}$ are generated and are manipulated as follows to ensure that the disturbance terms are distributed as N(0, $\sum$) and inter temporarily independent. Let:

$$\xi t = (\xi_{1t}, \xi_{2t}, \xi_{3t})^1 \quad (2)$$

Then by construction:

$$E(\xi_t) = 0, \text{cov}\ (\xi_t, \xi_{2t})^1 = \delta_{tt}{}^1\ I \quad (3)$$

Consider now the random vectors:

$$u_t = P\ \xi\ t$$

By construction, the vectors $\{u_t$: 1, 2, …, T$\}$ have the properties:

$$E(u_t) = 0 \text{ and } \sum = \text{cov}\ (u_t\ u_{t1}) = \Omega \otimes 1t$$

Where, $\otimes$ is the kronecker product or direct product. In this fashion, the desire error terms having the prescribed variance covariance matrix is obtained.

- According to this approach, specific values have to be assigned to the structural parameters:

$$\delta.1 = \begin{bmatrix} \beta_{13} \\ \delta_{11} \\ \delta_{12} \end{bmatrix}\ \delta.2 = \begin{bmatrix} \beta_{21} \\ \delta_{21} \\ \delta_{23} \end{bmatrix}\ \delta.3 = \begin{bmatrix} \beta_{32} \\ \delta_{31} \\ \delta_{32} \end{bmatrix} \quad (4)$$

Say:

$$\delta^0 = \begin{bmatrix} \delta^0.1 \\ \delta^0.2 \\ \delta^0.3 \end{bmatrix} \quad (5)$$

putting;

$$\beta^0 = \begin{bmatrix} 1 & 0 & -\beta^0_{13} \\ -\beta^0_{21} & 1 & 0 \\ 0 & -\beta^0_{32} & 1 \end{bmatrix} \quad (6)$$

and;

$$C^0 = \begin{bmatrix} -\delta^0_{11} & -\delta^0_{12} & 0 \\ -\delta^0_{21} & 1 & -\delta^0_{23} \\ 0 & -\delta^0_{32} & -\delta^0_{33} \end{bmatrix}$$

The reduced form of the system can be written as:

$$(y_{1t}, y_{2t}, y_{3t}) = (x_{1t}, x_{2t}, x_{3t})\ C_0\beta_0 + (u_{1t}, u_{2t}, u_{3t})\ \beta_0 \quad (7)$$

Since, all quantities on the right hand side of Eq. 7 are known, vectors of the determined variables $\{(y_{1t}-y_{3t})$: t = 1, 2, …, T$\}$ are then generated.

At the end of this 3rd step, the data set $\{(y_{1t}-y_{3t}, x_{1t}-x_{3t})$: t = 1, 2, …, T$\}$ have been generated by a model describe in Eq. 1 with fixed parameter vector δ^0 and error terms that are inter temporarily independent and distributed as N (0, $\sum$), the matrix $\sum$ being known.

The next step at this stage is to determine the known parameters as if they were not known using different methods and comparing the performances or sensitivity of the estimators.

RESULTS AND DISCUSSION

In this study, the samples were restricted to size n = 100, replicated in 200 times (R = 200) for an upper triangular matrix P to determine the effects of multicollinearity and the sensitivity of the estimation methods in a simultaneous equation model.

Meanwhile, we have employed Ordinary Least Square (OLS), 2 Stage Least Square (2SLS) and 3 Stage Least Square (3SLS) as the estimation methods. We have equally chosen the mean of estimates and bias of estimates as criteria for evaluating the performance of the estimators (Table 1-10). Table 1-10 have essentially shown the true parameters values, average estimated values, standard derivation (sample), standard deviation (population) and bias of estimates for the estimation methods employed at the three different levels of correlation desired. However, Table 10 has revealed a true

Table 1: Mean estimates and bias estimates of OLS for category I

Replication	Equation 1			Equation 2			Equation 3		
(assumed parameters)	$\beta_{13}(1.8)$	$\gamma_{11}(0.2)$	$\gamma_{12}(1.2)$	$\beta_{21}(1.5)$	$\gamma_{21}(2.5)$	$\gamma_{23}(2.1)$	$\beta_{32}(0.9)$	$\gamma_{31}(0.4)$	$\gamma_{32}(3.3)$
Average	1.30580	0.815587	0.173486	1.167873	-1.855278	-1.5828290	0.651204	0.086216	-1.463089
Standard deviation (S)	0.02377	0.757848	0.461011	0.023115	0.468098	0.6738338	0.010459	0.397089	0.544163
Standard deviation (δ)	0.22845	0.750213	0.456377	0.022882	0.463394	0.6670650	0.010354	0.393098	0.538694
Bias θ-θ	0.49420	0.615587	0.332127	0.332127	4.355278	3.6282900	0.248796	0.313784	0.538694

Table 2: Mean estimates and bias estimates of OLS for category II

Replication	Equation			Equation 2			Equation 3		
(assumed parameters)	$\beta_{13}(1.8)$	$\gamma_{11}(0.2)$	$\gamma_{12}(1.2)$	$\beta_{21}(1.5)$	$\gamma_{21}(2.5)$	$\gamma_{23}(2.1)$	$\beta_{32}(0.9)$	$\gamma_{31}(0.4)$	$\gamma_{32}(3.3)$
Average	1.2858280	1.003114	0.352331	1.158210	-1.469831	-1.3386396	0.688071	0.156861	-1.490858
Standard deviation (S)	0.0381470	0.538728	0.582215	1.509700	0.374653	0.6877080	0.146083	0.336979	0.370397
Standard deviation (δ)	0.0377630	0.533313	0.573630	0.149453	0.370870	0.6087970	0.144615	0.333592	0.366674
Bias θ-θ	0.5141472	0.803114	0.847669	0.341790	3.969831	3.4386390	0.211929	0.243139	4.790858

Table 3: Mean estimates and bias estimates of OLS for category III

Replication	Equation			Equation 2			Equation 3		
(assumed parameters)	$\beta_{13}(1.8)$	$\gamma_{11}(0.2)$	$\gamma_{12}(1.2)$	$\beta_{21}(1.5)$	$\gamma_{21}(2.5)$	$\gamma_{23}(2.1)$	$\beta_{32}(0.9)$	$\gamma_{31}(0.4)$	$\gamma_{32}(3.3)$
Average	0.252677	1.021620	0.818742	1.179460	-1.3813396	-0.730514	0.829480	0.196253	-1.4653030
Standard deviation (S)	0.019996	0.541854	0.322532	0.015027	0.3508910	0.350891	0.008384	0.315887	0.3106688
Standard deviation (δ)	0.019795	0.536408	0.319281	0.148760	0.3473650	0.303241	0.008299	0.312713	0.3075650
Bias θ-θ	0.547323	0.821162	0.381258	0.320533	3.8813400	2.830514	0.203747	0.203747	4.7653030

Table 4: Mean estimates and bias estimates of 2SLS for category I

Replication	Equation 1			Equation 2			Equation 3		
(assumed parameters)	$\beta_{13}(1.8)$	$\gamma_{11}(0.2)$	$\gamma_{12}(1.2)$	$\beta_{21}(1.5)$	$\gamma_{21}(2.5)$	$\gamma_{23}(2.1)$	$\beta_{32}(0.9)$	$\gamma_{31}(0.4)$	$\gamma_{32}(3.3)$
Average	1.294396	0.158902	-2.423510	1.0911190	-4.471467	-3.444422	0.0616542	0.381535	-1.804354
Standard deviation (S)	0.739101	2.672572	1.675328	4.3401560	11.157992	14.908293	0.4452180	3.140689	1.677272
Standard deviation (δ)	0.731673	2.645889	1.658490	4.2965260	11.045849	14.758458	0.4407400	3.109124	1.660414
Bias θ-θ	0.500644	0.500644	1.422351	0.4088800	6.971467	5.544220	0.2836570	0.018456	5.103540

Table 5: Mean estimates and bias estimates of 2SLS for category II

Replication	Equation 1			Equation 2			Equation 3		
(assumed parameters)	$\beta_{13}(1.8)$	$\gamma_{11}(0.2)$	$\gamma_{12}(1.2)$	$\beta_{21}(1.5)$	$\gamma_{21}(2.5)$	$\gamma_{23}(2.1)$	$\beta_{32}(0.9)$	$\gamma_{31}(0.4)$	$\gamma_{32}(3.3)$
Average	1.640204	0.480456	-1.7041474	1.0419980	-3.681072	3.693921	0.695659	-3.349145	-1.417137
Standard deviation (S)	1.704093	3.678219	12.8161100	2.2719090	3.663477	26.761766	0.410703	4.083402	3.926510
Standard deviation (δ)	1.686966	3.641251	12.0591790	2.2490753	3.626657	26.492790	0.406576	4.042362	3.887047
Bias θ-θ	0.139796	-0.280456	2.9041740	0.4580020	6.181072	-1.593920	0.204341	0.749145	4.717137

Table 6: Mean estimates and bias estimates of 2SLS for category III

Replication	Equation 1			Equation 2			Equation 3		
(assumed parameters)	$\beta_{13}(1.8)$	$\gamma_{11}(0.2)$	$\gamma_{12}(1.2)$	$\beta_{21}(1.5)$	$\gamma_{21}(2.5)$	$\gamma_{23}(2.1)$	$\beta_{32}(0.9)$	$\gamma_{31}(0.4)$	$\gamma_{32}(3.3)$
Average	0.881720	4.337829	-0.436278	10.427220	-29.260718	13.5667350	-0.325689	9.790342	0.375723
Standard deviation (S)	5.317600	26.725574	4.542259	49.294466	111.060445	97.1472590	4.910694	48.045510	10.282373
Standard deviation (δ)	5.288130	26.437169	4.496608	48.799000	109.944231	96.1709160	4.861338	47.562630	10.179030
Bias θ-θ	0.910827	-4.137829	1.636278	-8.927220	31.760718	-11.4667351	1.225689	-9.390342	2.924276

Table 7: Mean estimates and bias estimates of 3SLS for category I

Replication	Equation 1			Equation 2			Equation 3		
(assumed parameters)	$\beta_{13}(1.8)$	$\gamma_{11}(0.2)$	$\gamma_{12}(1.2)$	$\beta_{21}(1.5)$	$\gamma_{21}(2.5)$	$\gamma_{23}(2.1)$	$\beta_{32}(0.9)$	$\gamma_{31}(0.4)$	$\gamma_{32}(3.3)$
Average	1.4068826	0.9539760	0.31608797	2.019120840	-2.025479680	-1.61586120	0.6933164	0.282234300	-1.26668940
Standard deviation (S)	0.1473646	0.8460730	0.77223580	4.245016043	1.567934977	1.29267927	0.1348603	0.560533400	0.78079900
Standard deviation (δ)	0.3838809	0.9198223	0.79053920	2.060343671	1.211583663	1.13915726	0.3672330	0.748687790	0.88362838
Bias θ-θ	0.3931174	0.9538760	0.88391200	0.519120840	+14.525479680	+3.71515860	+0.2066836	+0.117774566	+4.56668940

Table 8: Mean estimates and bias estimates of 3SLS for category II

Replication	Equation 1			Equation 2			Equation 3		
(assumed parameters)	$\beta_{13}(1.8)$	$\gamma_{11}(0.2)$	$\gamma_{12}(1.2)$	$\beta_{21}(1.5)$	$\gamma_{21}(2.5)$	$\gamma_{23}(2.1)$	$\beta_{32}(0.9)$	$\gamma_{31}(0.4)$	$\gamma_{32}(3.3)$
Average	1.2853120	1.31760338	0.83690662	1.35556580	-2.07043428	-0.84533510	0.71395782	0.0505858460	-1.468029960
Standard deviation (S)	0.5930807	1.22568430	1.34879670	0.31196705	1.80624893	3.35819275	0.12583460	0.6604403484	1.137981235
Standard deviation (δ)	0.3136930	1.10710627	1.16137707	0.88854011	1.34367607	1.83253724	0.35471739	0.7774339100	1.066762040
Bias θ-θ	0.5146880	1.11760338	+0.36349338	+0.14443420	+4.57043430	+2.94533510	+0.18604218	+0.3464145400	+4.768029960

Table 9: Mean estimates and bias estimates of 3SLS for category III

Replication (assumed parameters)	Equation 1			Equation 2			Equation 3		
	$\beta_{13}(1.8)$	$\gamma_{11}(0.2)$	$\gamma_{12}(1.2)$	$\beta_{21}(1.5)$	$\gamma_{21}(2.5)$	$\gamma_{23}(2.1)$	$\beta_{32}(0.9)$	$\gamma_{31}(0.4)$	$\gamma_{32}(3.3)$
Average	1.376075560	1.21673120	0.14233106	1.329072570	12114618.173	-1.340326030	0.6925229008	0.2874163000	11.484284858
Standard deviation (S)	1.260599570	6.67945450	2.93736270	4.759215553	8588680.3050	1.260352022	0.1707178212	1.1295949100	1.017881670
Standard deviation (δ)	1.122764256	2.58446406	1.70387360	0.689870780	2930.6416000	1.122654008	0.4131801317	1.1011344402	1.090121900
Bias θ-θ	0.423924400	1.01673120	1.01673120	-1.057668940	+0.576689400	+3.440326000	+0.2074709920	+1.1125836920	+4.784284858

Table 10: Average of parameter estimates for data set (N = 100, R = 200)

Replication	Equation 1			Equation 2			Equation 3		
	$\beta_{13}(1.8)$	$\gamma_{11}(0.2)$	$\gamma_{12}(1.2)$	$\beta_{21}(1.5)$	$\gamma_{21}(2.5)$	$\gamma_{23}(2.1)$	$\beta_{32}(0.9)$	$\gamma_{31}(0.4)$	$\gamma_{32}(3.3)$
OLS	1.3058	0.8156	0.1735	1.1679	-1.8553	-1.5829	0.6512	-0.0862	-1.5225
2SLS	1.3522	0.1589	0.3486	1.0911	-4.7140	-3.4444	0.6163	0.3735	-1.8044
3SLS	1.4069	0.9540	-0.3233	2.0193	-2.0253	-1.6152	0.6934	0.2822	-1.3169

state of the comparative performance of those estimates. On the overall, the following order of preference of estimators can be deduced: 3SLS>2SLS>OLS.

CONCLUSION

This study has investigated the sensitivity of three different estimators to the problem of multicollinearity among the exogenous variables in a Simultaneous equation model. A three equation just identified model was used for the Monte Carlo analysis. We simulated sample of n = 100 for a replication of R = 200 for an upper triangular matrix for three exogenous variables which were later used to generate corresponding three endogenous variables. Three levels of correlation on three different estimation methods were considered. The results obtained were analyzed and estimated using the mean of estimates and its bias. Although, the study did not produce identical estimates as the assumed parameter but some estimates are quite close, especially in 2 and 3SLS. The presence of multicollinearity may accounts for the situation. The bias estimate also confirmed the preference of this order of estimation methods 3SLS>2SLS>OLS.

REFERENCES

Adenomon, M.O. and M.O Fesojaiye, 2008. A Comparison of SUR and OLS estimating techniques. Proceedings of the 32nd Annual Conference of Nigerian Statistical Association held at Trade Fair Complex, Ekiti, Nigeria.

Ayinde, K., 2007. A comparative study of the performance of the OLS and some GLS estimators when stochastic regressors are both collinear and correlated with error terms. J. Math. Stat., 3: 196-200.

Hendry, D.F., 1976. The structure of simultaneous equation estimators. J. Econometr., 4: 51-88.

Johnston, J., 1984. Econometrics Methods. 3rd Edn., McGraw-Hill Book Company, New York, USA., ISBN-10: 0-07-032685 pp: 568.

Kmenta, J. and M.E. Joseph, 1963. A Monte Carlo study of alternative estimates of the coobb-douglas production function. Econometrica, 31: 363-390.

Kmenta, J., 1971. Element of Econometrics. Macmillan Publishing Company, New York.

Kutsoyiannis, A., 2001. Theory of Econometrics. 2nd Edn., Macmillan Publishing Company, London.

Mishra, S.K., 2004. Estimation under Multicollinearity: Application of restricted Liu and maximum entropy estimators to the Portland cement dataset. http://mpra.ub.uni-muenchen.de/1809/.

Parker, J.B., 1972. The use of the Monte Carlo methods for solving large-scale problems in Neutronics. J. R. Stat. Soc. Am., 135: 16-43.

Pleli, N. and S. Tankovic, 2005. Simultaneous equation econometric model: Choosing an estimation method. Kennedyjev Trg 6, Zegreb 10000.

Wagnar, H.M., 1958. A Monte Carlo study of estimates of simultaneous linear structural equations. Econometrica, 26: 117-133.