Journal of Mathematical Sciences: Advances and Applications
Volume 15, Number 1, 2012, Pages 1-12

A NOTE ON THE EFFECT OF THE MULTICOLLINEARITY PHENOMENON OF A SIMULTANEOUS EQUATION MODEL

D. A. AGUNBIADE

Department of Mathematical Sciences
Olabisi Onabanjo University
Ago Iwoye
Nigeria
e-mail: bayoagunbiade@gmail.com

Abstract

This paper sought to examine both the intra and inter equation effects of multicollinearity on the exogeneous variables of a simultaneous equation model and also investigate the effect of the unsuspected dependence between random normal deviates used for generating the stochastic components of a simultaneous equation model. A Monte Carlo approach was used to achieve this by setting up a two-equation with five structural parameters simultaneous econometric model and applying the six different estimation techniques. It was found that while 2SLS and LIML produced identical results, the results of other techniques were at variants, which confirmed the likely presence of multicollinearity among the regressor variables.

1. Introduction

It is a common practice for researchers in undertaking Monte Carlo studies to resort to the use of randomly generated values of exogeneous variables as well as random normal deviates. The values so generated for

2010 Mathematics Subject Classification: 62Jxx.

Keywords and phrases: Monte Carlo study, multicollinearity, simultaneous equation, instrumental variable estimators, exogeneous variables.

Received December 25, 2011

the exogeneous variables and random normal deviates are invariably not examined for orthogonality before they are used in such studies. By so doing, many researchers make the unverified and potentially wrong assumption that the randomly generated variables as well as the random normal deviates are independent and uncorrelated. Results based on variables so generated and based, as they are, on the assumption that such random variables and error terms are inherently orthogonal, may be misleading. In view of the possibility of multicollinearity problem affecting on such results, inferences drawn from them may be impaired. Consequently, conclusion based on such results may be misleading. As a result of the problem stated above, there is need for the presence of multicollinearity between the exogeneous variables especially those, which appear in the same equation to be explored.

This study is therefore designed to answer the following question: Should pairs of random variables and deviates be taken on their face values as being independent? Or should a researcher look for such random normal deviates and variables, which are independent before using them in Monte Carlo studies? Many Monte Carlo studies have really not taken these into consideration. This paper therefore sought to examine both the intra and inter equation effects of multicollinearity on the exogeneous variables of a simultaneous equation model and also investigate the effect of the unsuspected dependence between random normal deviates used for generating the stochastic components of a simultaneous equation model.

2. Theoretical Background and Literature Review

In reviewing the existing literature on the estimation of simultaneous equation models, a host of authors believed that the ordinary least squares (OLS) method is known to lead to biased and inconsistent estimate, when used in estimating the parameters of each equation in a simultaneous equation model (Mishra [12], Johnston [9], Koutsoyiannis [10], Agunbiade and Iyaniwura [3]) except in the case of a recursive model (Montgomery et al. [13], Hassen [7]).

In the typical simultaneous case, where the special assumptions of a recursive system are not fulfilled, the main estimating techniques are indirect least squares (ILS), two-stage least squares (2SLS), both of which may be interpreted as instrumental variable estimators, limited-information maximum likelihood (LIML), three-stage least squares (3SLS), and full-information maximum likelihood (FIML). ILS, 2SLS, and LIML are essentially single-equation methods, in which attention is focused on one equation at a time without using all the information contained in the detailed specification of the rest of the model. 3SLS and FIML are systems methods, where all the equations of the model are estimated simultaneously. In principle, information on the complete structure, if correct, will yield estimators with greater asymptotic efficiency than that attainable by limited-information methods (Hassen [7] and Montgomery et al. [13]) supported this. However, in practice, ILS is not a widely used technique because of the stringent requirement for an equation to be exactly identified. 2SLS is perhaps the most important and widely used procedure because, it is applicable to equations that are over-identified or exactly identified. Moreover, in the case of an exactly identified equation, the 2SLS estimates are identical with the ILS estimates. The LIML estimators have the same asymptotic variance-covariance matrix as 2SLS. The estimates of the asymptotic variances, however, will differ since s^2 the sample variance is computed from the estimated structural coefficients, which will be different in the two cases.

A necessary condition for the greater efficiency of the full-information or systems method of estimation over a limited-information method is that the specification of the complete model should be correct. In many systems, this is a formidable requirement, and the larger and more detailed the system, the more difficult it is to specify it completely. Even granted a correct full-system specification, there are two conditions under which 2SLS and 3SLS will give identical point estimates with identical asymptotic sampling variances. The first is

$$\sigma_{ij} = 0 \text{ for all } i \neq j,$$

that is the contemporaneous correlations between the disturbances in different structural equations are all zero. The other condition for the equivalence of 2SLS and 3SLS estimators is that of every equation must be exactly identified. It must be noted also that before applying 3SLS in practice, one must omit all unidentified equations and also all identities, since the latter have zero disturbance, which would render the $\sum$ matrix singular. As with 3SLS, the full-information maximum likelihood method is a complete system method of estimation. It is computationally more tedious than the 3SLS because it often involves the solution of nonlinear equations. The asymptotic variance matrix of the FIML estimator is identical with that of 3SLS thus indicating the asymptotic efficiency of the latter method. This feature, combined with its less severe computational problems leads some authors to recommend 3SLS over FIML (see Johnston [9] for more details). It is upon this background that we seek to make a comparative analysis of the various estimation techniques in producing a robust estimator when applying to the simultaneous equation model. In this paper, we employ the above mentioned estimation methods on our multicollinear generated Monte Carlo data to examine the intra and inter equation effects of multicollinearity on the exogeneous variables of our two equation simultaneous model having five structural parameters.

3. The Concepts of Multicollinearity

The OLS estimators can be stated as

$$\mathbf{b} = (\mathbf{X}^1\mathbf{X})^{-1}\mathbf{X}^{/}\mathbf{Y}, \tag{1}$$

with the variance matrix

$$\mathbf{var}(\mathbf{b}) = \sigma^2(\mathbf{X}^1\mathbf{X})^{-1}.$$

Thus the sampling variances depend not only on the disturbance variance σ^2, but also on the sample values of the explanatory variables. Where the explanatory variables are non-orthogonal, the correlation between the explanatory variables will be high. In such a case, the numerical value of the off-diagonal (co-variation) term will be high. When two explanatory variables are orthogonal, the coefficient of the X's in the multiple regression equation would be the same as those given by the simple regression of Y on each X in turn. Olayemi and Olayide ([15]) believed that orthogonal variable may be set up in experiment designs, but they are the exception, not the rule, in economic data. Increasing correlation between the two explanatory variables is evidenced by the increasing numerical value for the off-diagonal (co-variation) term. This is also reflected in the dramatic fall in the value of the determinant. This is described as a situation of collinearity (or multicollinearity) between the explanatory variables.

4. Methodology

The principle of asymptotic theory is the utilization of distributional properties in large samples to finding estimates of unknown population parameters and drawn inference about them. Some of these properties include consistency, asymptotic efficiency, and asymptotic normality. In econometrics, attention is often focused on evaluating the behaviour of estimators in small samples and asymptotic theory fails most of the time to furnish enough useful information about these estimators. This inevitably creates a problem in trying to choose from among competing estimators. One way of studying the small-sample properties of estimators is to utilize the Monte-Carlo technique.

The Monte Carlo approach has been applied not only to the choice of alternative estimators, but also in determining the impact of sample size, serial correlation, multicollinearity, and other factors on the different possible estimators in a given study (Intriligator et al. [8]). Maddala ([11]) is of the opinion that this approach creates a "laboratory environment", where "control experiments" on estimators are performed. Another reason for the desirably of the Monte Carlo technique is its wide application in providing a way out of mathematical problems, which are analytically intractable.

Despite its widespread use, the Monte Carlo technique suffers from two main criticisms, namely, specifically and imprecision, although the impact of these deficiencies can be reduced to a reasonable extent. For instance, response surface methodology may be used to reduce specificity, while control variate technique may be used to reduce the problem of imprecision. By specifically, we mean that the results obtained are largely dependent on the specification of the model(s) used in a given study, imprecision simply means that the results are not precise.

The implication of these deficiencies is two-fold. One is that simulation results are most of the time suggestive rather than conclusive, since the extent of generalization of sampling experiments compares to their analytical counterparts is smaller. Secondly, Monte Carlo results are not end-product in econometric theory in general or in the investigation of finite sample distributions in particular, but stand somewhere in between. This seems to be the view of Intriligator et al. ([8]) in his assertion:

"Monte Carlo experimentation can efficiently complement analysis to establish numerical analytic formulae, which jointly summarize the experimental findings and known analytical results in order to help interpret empirical evidence and to compute outcomes at other points within the relevant parameter space. The accuracy obtainable depends on the given budget constraint and the relevant price of capital of labour, the later determining the point at which simulation is substituted for analysis and the former the overall precision of the exercise".

Up to this day, the small sample properties of the various econometric techniques have been studied from simulated data in what are known as Monte Carlo studies, and not with direct application of the techniques to actual deviations. This approach is due to the fact the actual observations on economic variables usually involve multicollinearity, autocorrelation, errors of measurement, and most other econometric 'diseases' simultaneously. Studies involving small sample properties of estimators are usually based on the assumption of the simultaneous occurrence of all these problems. By the Monte Carlo approach, the econometrician can generate data sets and stochastic terms, which are free to all, but containing one of the problems listed above and therefore generate data resembling those obtained from controlled experiments.

The Monte Carlo approach used is, therefore, the process of drawing conclusions from simulated data. In the theoretical or analytical approach, conclusions are deducted from postulates.

Basically, the difference is between induction and deduction. This methodology, in the past two decades, has found extensive use in the field of operational research and nuclear physics, where there is a variety and complexity of problems beyond the available resources of the theoretician. They have been employed sporadically in numerous other fields of sciences and in the social sciences.

Parker ([16]) employed the Monte Carlo method for solving large-scale problems in neutronics. Also, Wagnar ([17]) employed Monte Carlo sampling technique to compare small sample properties of limited information-single equation (L.I.S.E.), least squares, and instrumental variable estimate. Two versions of an essentially two-equation model were used to generate 100 sets of observations over 20 time periods. With these observations and various statistical techniques, estimates of the parameters of an over-identified equation were obtained and compared.

The analysis of the results of the 100 sets of data produced for each of the model indicates that, the least squares generally gives more biased but less variable estimates than the L.I.S.E. method. Furthermore, the sample estimates of the variance of an L.I.S.E. statistics is reliable on the average, and that the t distribution may be used in constructing confidence intervals with L.I.S.E. estimates of the parameters and the sampling variance of the estimates.

Frees ([5]) considered four alternative forms of two-parameter normal and non-normal error distributions. He also reported on a Monte Carlo study of the small-sample properties of least squares, two-stage least squares, three-stage least squares, and full-information maximum likelihood estimators.

On the basis of 1,000 replications of sample size 20 in two experiments on an over-identified model, he found that the small-sample ranking of econometric estimators of both structural coefficient and forecasts of endogeneous variables, according to parametric and non-parametric measures of bias, dispersion and dispersion including bias, do not change for any of the four error distributions considered. He found out further that least squares is the most biased and maintains the Gauss-Markov property of minimum variance. According to his result, the large bias of least squares is more than offsets the small variance, so that least square exhibits the largest mean squares of the four estimators.

5. Data Generation

Our model can be rewritten as

$$\beta Y_t + \Gamma X_t = \varepsilon_t, \tag{2}$$

where

$$\beta = \begin{pmatrix} 1 & \beta_{12} \\ -\beta_{21} & 1 \end{pmatrix}, \quad \Gamma = \begin{pmatrix} -\gamma_{11} & -\gamma_{12} & 0 \\ 0 & -\gamma_{21} & -\gamma_{23} \end{pmatrix},$$

$$Y_t = \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix}, \quad X_t = \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix}, \quad \varepsilon_t = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}.$$

From Equation (2), we can obtain the reduced form of our model by simplify for Y_t as follows:

$$Y_t = -\beta^{-1}\Gamma X_t + \beta^{-1}\varepsilon_t. \tag{3}$$

This can be further simplified thus

$$Y_t = 1/(1-\beta_{12}\beta_{21})\begin{pmatrix} 1 & \beta_{12} \\ \beta_{21} & 1 \end{pmatrix}\begin{pmatrix} \gamma_{11} & -\gamma_{12} & 0 \\ 0 & -\gamma_{21} & -\gamma_{23} \end{pmatrix}\begin{pmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{pmatrix}$$

$$+ 1/(1-\beta_{12}\beta_{21})\begin{pmatrix} 1 & \beta_{21} \\ \beta_{12} & 1 \end{pmatrix}\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},$$

$$y_{1t} = \left(\frac{\gamma_{11}}{1-\beta_{12}\beta_{21}}\right) x_{1t} + \left(\frac{\gamma_{12}+\beta_{21}\gamma_{21}}{1-\beta_{12}\beta_{21}}\right) x_{2t} + \left(\frac{\beta_{12}\gamma_{23}}{1-\beta_{12}\beta_{21}}\right) x_{3t} + \left(\frac{\varepsilon_{1t}+\beta_{21}\varepsilon_{2t}}{1-\beta_{12}\beta_{21}}\right),$$

$$y_{2t} = \left(\frac{\gamma_{11}\beta_{12}}{1-\beta_{12}\beta_{21}}\right) x_{1t} + \left(\frac{\gamma_{11}+\beta_{12}+\gamma_{21}}{1-\beta_{12}\beta_{21}}\right) x_{2t} + \left(\frac{\gamma_{23}}{1-\beta_{12}\beta_{21}}\right) x_{3t} + \left(\frac{\beta_{12}\varepsilon_{1t}+\varepsilon_{2t}}{1-\beta_{12}\beta_{21}}\right). \quad (4)$$

For the purpose of this research work, we assumed numerical values for the structural parameters. These are shown below:

$$\beta_{12} = 1.8, \quad \gamma_{11} = 1.2, \quad \gamma_{21} = 0.6,$$

$$\beta_{21} = 0.4, \quad \gamma_{12} = 10.5, \quad \gamma_{23} = 1.4. \quad (5)$$

Solving for y_{1t} and y_{2t} in Equation (4), we have

$$y_{1t} = 4.2857\ x_{1t} + 2.6429\ x_{2t} + 9.0\ x_{3t} + (\varepsilon_{1t} + 0.4\ \varepsilon_{2t})/0.28,$$

$$y_{2t} = 7.7143\ x_{1t} + 5.3571\ x_{2t} + 5.0\ x_{3t} + (1.8\ \varepsilon_{1t} + \varepsilon_{2t})/0.28,$$

model (3) is the reduced form of our original model in (1).

The X's and ε's earlier generated were passed through Equation (4) to obtain numerical values for the Y's in five replications.

6. Estimation of the Structural Parameters

At this juncture, we are in the possession of numerical values of the X's-exogeneous variables, the Y's-endogeneous variables, and ε's-the error terms. Armed with this information, efforts were made to estimate the structural parameters using various estimation methods discussed above. The time series statistical analysis program (TSP) was employed for the purpose of this estimation. The estimates obtained are then subjected to the least squares estimation.

7. Results Analysis and Discussion

Based on the results generated from our estimation of the structural parameters, we did the following:

(1) Computed biases of the estimates of the parameters of the model.

(2) The biases were ranked on the basis of the various correlation explored. The smallest bias was ranked 1, while the largest bias was ranked 3.

(3) The R-squared of the estimates were also ranked on the basis of the various levels of correlation explored. Here the highest R-squared was ranked 1, while the lowest was ranked 3.

Although, the biases obtained from the estimation of the structural parameters were noted to be large, the R-squared were quite high. Also, a few outliners were noted among the structural parameter estimates. The estimates turned out by the reduced form estimation process have smaller biases. However, the R-squared values of the estimates were very low. As expected, the estimates generated by the two stage least squares method are identical with those generated by the limited information maximum likelihood method. However, as would be expected, our results did not yield identical estimates under 2SLS and 3SLS methods. By and large, our Monte Carlo experiment has not been able to produce estimates that are identically the same as the original parameters we started with. This situation may be attributable to the presence of multicollinearity amongst other reasons. Furthermore, the ranking effected on our biases has not been able to clearly bring out the effect of multicollinearity on our experiment.

8. Conclusion

The Monte Carlo study set up to determine the effect of multicollinearity phenomenon in a simultaneous equation econometric model by using a two-equation having five structural parameters has essentially shown that results of the 2SLS and LIML estimation techniques produced essentially virtual identical results while variations exists with other estimation techniques. Again, some outliers were noted among the structural parameters as the biases were noted to be high with very high R-squared values. Since Monte Carlo study serves as alternative to the real-life experiment; the results obtained from it, though, has not produced those that are identically the same with the original data/parameter values we started with, the situation may be attributable to the presence of multicollinearity among other reasons. We therefore call on researchers to be weary on the adverse effect of multicollinearity phenomenon and thus should screen data before embarking on analysis.

References

[1] D. A. Agunbiade and J. A. Osilagun, Identifiability criteria in a simultaneous equation model, African Journal of Pure Applied Sciences 1(1) (2008), 34-38.

[2] D. A. Agunbiade, A Monte-Carlo Approach to the Estimation of a Just Identified Simultaneous Three-Equations Model with Three Multicollinear Exogeneous Variables, Unpublished PhD. Thesis, University of Ibadan, Nigeria, 2008.

[3] D. A. Agunbiade and J. O. Iyaniwura, Estimation under multicollinearity: A comparative approach using Monte Carlo methods, Journal of Mathematics and Statistics 6(2) (2010), 183-192.

www.scipub.org

[4] B. R. Casteneira and L. C. Nunes, Testing endogeneity in a regression model: An application of instrumental variable estimation, Investigation Operative 8(1-3) (1999).

[5] E. D. Frees, Regression Modelling with Actuarial and Financial Applications, First Published by Cambridge University Press, ISBN 978-052-113596-2, Pg. 565, 2010.

http://research.bus.wisc.edu/Reg/Acturies

[6] D. N. Gujarati and Sangeetha, Basic Econometrics, Fourth Edition, Tata McGraw-Hill Publishing Company Ltd., India, ISBN 978-0-07-066005-2, Pg. 1036, 2007.

www.tatamcgrawhill.com

[7] B. E. Hansen, Econometrics, Revised Edition, Page 291, 2011.

www.ssc.wisc.edu/~bhansen

[8] M. Intriligator, R. Bodkin and C. Hsiao, Econometric Models, Techniques and Applications, International Edition, Prentice Hall, 1996.

[9] J. Johnston, Econometrics Methods, Fourth Edition, McGraw- Hill Book Company, New York, 2007.

[10] A. Koutsoyiannis, Theory of Econometrics: An Introductory Exposition of Econometric Methods, Second Edition, Palgrave, New York, 2003.

[11] G. S. Maddala, Introduction to Econometrics, John Wiley and Sons Ltd., Third Edition, ISBN 9971-51-383-8, Pg. 636, 2005.

[12] S. K. Mishra, Estimation under Multicollinearity: Application of Restricted Liu and Maximum Entropy Estimators to the Portland Cement Dataset, 2004.

http://mpra.ub.uni-muenchen.de/1809/

[13] D. C. Montgomery, E. A. Peck and G. G. Vining, Introduction to Linear Regression Analysis, Reprint Third Edition, John Wiley Series in Probability and Statistics, ISBN 81-265-1047-1, Pg. 641, 2007.

[14] A. L. Nagar, A Monte Carlo study of alternative simultaneous equation estimators, Econometrica 28(3) (1969), 573-590.

[15] J. K. Olayemi and S. O. Olayide, Elements of Applied Econometrics, 1998.

[16] J. B. Parker, The use of the Monte Carlo methods for solving large scale problems in neutronics, Journal of Royal Statistical Society of America 135(1) (1972), 16-37.

[17] H. M. Wagnar, A Monte Carlo study of estimates of simultaneous linear structural equations, Econometrica 26 (1958), 117-133.

■