EKOIST Journal of Econometrics and Statistics
EKOIST 2024, 41 : 13–21
DOI: 10.26650/ekoist.2024.41.1251589

RESEARCH ARTICLE

A Monte Carlo Study of the Residuals of System Estimators in the Presence of Multicollinearity

Emmanuel Oduntan[1]

[1](Dr.), American University of Nigeria, Department of Mathematics and Statistics, Adamawa State, Nigeria

ABSTRACT
Residual analysis is often used to evaluate the precision of the parameter estimates of econometric models. Analysis of residuals from regression is an important way of assessing the performance of a regression model in achieving the goal of accounting for the independent variable under the underlying assumption. With the Monte Carlo Simulation (MCS) of a data set of sample size 40 over varied replications R = 20, 50, 100 and 150, we used residual analysis to study the relative performance of six estimators of a simultaneous equation model under varied multicollinearity conditions. We found that the two-stage least squares (2SLS), Limited Information Maximum Likelihood (LIML), and Three-Stage Least Squares (3SLS) estimators generated virtually similar estimates. This is in agreement with the theory. In addition, the results revealed that notwithstanding the level of multicollinearity, Ordinary Least Squares (OLS), followed by Indirect Least Squares (ILS), produced the lowest Sum of Squared Residuals (SSR) of parameter estimates, an indication of the robustness of OLS in the presence of multicollinearity. This result also showed that the single equation estimators (OLS and ILS) performed better than the system estimators under the condition of multicollinearity to which we subjected our model. Furthermore, the Sum of Squared Residuals (SSR) generated for cases of low multicollinearity are lower than those generated for cases of high multicollinearity.

Keywords: Residual Analysis, Simultaneous Equation Model, Monte Carlo Simulation, Estimators, Multicollinearity, Replications.

Introduction

Using the Monte Carlo Simulation framework proposed by Oduntan and Iyaniwura (2021), this paper applied a residual analysis based on the Sum of Squared Residuals (SSR) criteria to examine the performance of 6 estimators vis-à-vis, Full Information Maximum Likelihood [FIMF], Limited Information Maximum Likelihood [LIMF], Two-Stage Least Squares [2SLS], Three-Stage Least Squares [3SLS], Indirect Least Squares [ILS], and Ordinary Least Squares [OLS], in the presence of multicollinearity. The results show that OLS produced the lowest Sum of Squared Residuals (SSR), which is an indication of the robustness of OLS in the presence of multicollinearity.

Analysis of residuals from regression is an important way of assessing the performance of a regression model in achieving the goal of accounting for the independent variable under the underlying assumption. All residuals are expected to be small and unstructured. Structured residual or those that are non-random sheds a "bad light" on the regression. Most problems that are associated with the appointment of variables into the model as well as a choice of model estimator will turn up in the residuals. Model diagnostics is an integral part of model determination, and an important part of the model diagnostics is residual analysis (Farias and Branco, 2012). Analysis of the residuals plays an important role in validating the regression model. It is a powerful diagnostic tool, as it helps in assessing whether or not some underlying assumptions of regression have been violated. These violations may take a toll on the appropriateness of the model estimators. Residual analysis can be used as a powerful tool in model improvement (Zhuang, 2006)

Graphical and numerical analysis of residuals can be informative about model misspecification even when data are censored or grouped.(Chesher and Irish, 1987). In addition, many applied workers are strongly oriented to residual analysis for assessing model adequacy (Pagan and Hall, 1983). Using a normal linear model, Chesher and Irish (1987) developed procedures for calculating diagnostic statistics to detect model misspecification when grouped or censored data are analysed. Jalilian and Vahidi-Asl (2011),

Corresponding Author: Emmanuel Oduntan **E-mail:** emmanuel.oduntan@aun.edu.ng
Submitted: 08.10.2023 • **Revision Requested:** 24.09.2024 • **Last Revision Received:** 09.10.2024 • **Accepted:** 15.10.2024

in their study on the use of residual analysis for inhomogeneous Neyman-Scott models, stated that ‘residual analysis method has a good performance in assessing goodness-of-fit and revealing inadequacy of the fitted model’. Baddeley et al. (2005) defined residuals for point process models fitted to spatial point pattern data and proposed diagnostic plots based on them. They observed that ‘a plot of smoothed residuals against spatial location, or against a spatial covariate, is effective in diagnosing spatial trend or covariate effects’.

In the literature, many studies used residuals of parameter estimates from regression to assess the performance of the regression model in achieving the goal of accounting for the independent variable under the underlying assumption. Some of these studies are as follows: Albert and Chib (1995), Farias and Branco (2012), Ogata (1988), Pagan and Hall (1983), Schoenberg (2003), Clements, et al. (2011), Ardalani-Farsa, et al. (2010), Chaloner and Brant (1988) and Chaloner (1991).

Based on the expectation that the residuals should be small, in this paper, we conducted a numerical analysis of the residuals of our parameter estimates to examine the performance of our estimators in the presence of multicollinearity using the Monte Carlo Simulation (MCS). In section 2, we present the theoretical framework and empirical strategy of the study. Section 3 presents our results and discussions, while section 4 provides the conclusion of the study.

Materials and Methods

Theoretical Framework

Let

$$Y = F(X, \phi) + mu \tag{1}$$

where µ $\sim N(0, \sigma^2)$ satisfies the classical least squares assumptions.

Assign numerical values to ϕ,and σ^2. On the basis σ^2, select normal deviates to be used in generating µ. Select a random sample of size T for X and compute the numerical values of F(X,ϕ). Obtain vector Y by computing F(X,ϕ) + µ. Lastly, regress Y on X to generate $\hat{\phi}$

With the same sample size, repeat this process to facilitate the construction of the sampling distribution of ($\hat{\phi}$) and to investigate the stability of the results. We then evaluate the precision of $\hat{\phi}$ and make further evaluation of the efficiency of different estimators of ϕ, using the empirical distribution obtained.

Furthermore, as is well known, the OLS estimator $\hat{\phi}$ of ϕ is the minimizer of

$$Q = \sum_i (y_i - X\phi)^2$$

Hence, the residual from the model can be expressed as;

$$\hat{\varepsilon}_i = (y - X\hat{\phi})$$

Using the Sum of Squared Residuals (SSR) criteria for the comparative analysis of the performance of different estimators on the basis of the residuals generated by each estimator, the estimator with the least SSR is ranked best among the others.

The Empirical Strategy

Adopting the Monte Carlo simulation strategy of Oduntan and Iyaniwura (2021), we assumed the model

$$\begin{aligned} y_{1t} &= \beta_{12} y_{2t} + \gamma_{11} X_{1t} + \gamma_{12} X_{2t} + u_{1t} \\ y_{2t} &= \beta_{21} y_{1t} + \gamma_{22} X_{2t} + \gamma_{23} X_{3t} + u_{2t} \end{aligned} \tag{2}$$

where

y’s are endogenous variables
x’s are exogenous variables
u’s are disturbance terms

In the matrix form, equation 2 becomes

$$Y = X\phi + u$$

where

$$Y = [y_1 y_2], \quad X = [1\,1\,1\,X_1 X_2 X_3\,1\,1\,1] \quad \phi = [\beta_1\beta_2\gamma_1\gamma_2\gamma_3]$$
$$u = [u_1 u_2]$$

where u ~ $N(0,\sigma^2)$ satisfies the classical least squares assumptions.

We generated our data series as follows:

1. Set sample size = 40, for the purpose of this study. Arbitrarily assigned the following numerical values to the model's structural parameters.
2. Arbitrarily assigned the following numerical values to the model's structural parameters.
$$\beta_{12} = 1.8,\ \gamma_{11} = 1.2,\ \gamma_{12} = 0.6$$
$$\beta_{21} = 0.4,\ \gamma_{22} = 0.5,\ \gamma_{23} = 1.4 \tag{3}$$
3. Assign arbitrary values to the elements of the variance-covariance matrix of the disturbance terms at any given sample point.
$$\Omega = [4.5\,3.0\,3.0\,3.5] \tag{4}$$
4. Select the values of the predetermined variables X_{1t}, X_{2t} and X_{3t} from a pool of uniformly distributed random numbers with the correlation coefficients $r_{(x_1x_2)}$, $r_{(x_2x_3)}$ and $r_{(x_1x_3)}$ defined as; (a) Low multicollinearity - insignificant at the 5(b) High multicollinearity - significant at the 1This results in six sets of X' s defined as $r_{(x_1x_2)}$,$r_{(x_2x_3)}$ and $r_{(x_1x_3)}$: Low Multicollinearity $r_{(x_1x_2)}$,$r_{(x_2x_3)}$ and $r_{(x_1x_3)}$: High Multicollinearity
5. From a Normal (0,1) distribution, generate the values of U_{1t} and U_{2t} for each sample point using the following two-step procedure:
(a) Draw independent series ε_t of random normal deviates ~ N(0,1) from a pool of random normal deviates.
(b) Transform the generated series into a series of random disturbances to guarantee conformity with the variance-covariance matrix Ω using the method presented by Nagar (1969) as described below:
Oduntan and Iyaniwura (2021), further define a positive definite matrix Σ such that
$$\Sigma = PP^{/} \tag{5}$$
where P is an upper triangular matrix
Let
$$P = (S_{11} S_{21} 0 S_{22}) \tag{6}$$
Then
$$S_{22} = +\sqrt{\sigma_{22}}$$
$$S_{21} = \frac{\sigma_{12}}{S_{22}} \tag{7}$$
$$S_{11} = +\sqrt{(\sigma_{11} - S_{21}^2}$$
We generated the random disturbance series using
$$u = P\varepsilon_t = \left(\frac{u_{t1}}{u_{t2}}\right)$$
$$= P\left(\frac{\varepsilon_{t1}}{\varepsilon_{t2}}\right) \tag{8}$$
$$= (S_{11}\,S_{21}\,0S_{22})(\varepsilon_{t1}\,\varepsilon_{t2})$$

Hence,

$$u_{t1} = S_{11}\varepsilon_{t1} + S_{21}\varepsilon_{t2}$$
$$u_{t2} = S_{22}\varepsilon_{t2} \quad (9)$$

(vi) By reduced form, generate the endogenous variables from the values obtained for the X's and U's and the values assigned to the structural parameters.
Consider the model,

$$y_{1t} = \beta_{12}y_{2t} + \gamma_{11}X_{1t} + \gamma_{1t}X_{2t} + u_{1t}y_{2t} = \beta_2 1 y_{1t} + \gamma_{21}X_{2t} + \gamma_2 t X_{3t} + u_{2t}$$

Rearranging, we have,

$$y_{1t} - \beta_{12}y_{2t} - \gamma_{11}X_{1t} - \gamma_{2t}X_{2t} - 0X_{3t} = u_{1t} - \beta_{21}y_{1t} + y_{2t} - 0X_{1t} - \gamma_{21}X_{2t} - \gamma_{23}X_{3t} = u_{2t}$$

or

$$BY_t + \Gamma X_t = u \quad (10)$$

Where,

$$B = [1 - \beta_{12} - \beta_{21}1],\ \Gamma = [-\gamma_{11} - \gamma_{12}00 - \gamma_{21} - \gamma_{23}],$$
$$Y_t = [y_{1t}y_{2t}],\ X_t = [X_{1t}X_{2t}X_{3t}],\ u = [u_{1t}u_{2t}]$$

Rewriting Equation (10), we have

$$\begin{aligned} Y_t &= -B^{-1}\Gamma X_t + B^{-1}u \\ &= -\frac{1}{1-\beta_{12}\beta_{21}}[1\beta_{21}\beta_{12}1][-\gamma_{11} - \gamma_{12}00 - \gamma_{21} - \gamma_{23}][X_{1t}X_{2t}X_{3t}] \\ &+ \frac{1}{1-\beta_{12}\beta_{21}}[1\beta_{21}\beta_{12}1][u_{1t}u_{2t}] \\ \text{where } B^{-1} &= \frac{1}{(1-\beta_{12}\beta_{21}}[1\beta_{21}\beta_{12}1] \end{aligned} \quad (11)$$

Furthermore,

$$\begin{aligned} y_{1t} &= \left[\frac{\gamma_{11}}{1-\beta_{12}\beta_{21}}\right]X_{1t} + \left[\frac{\gamma_{12}+\beta_{21}\beta_{21}}{1-\beta_{12}\beta_{21}}\right]X_{2t} + \left[\frac{\beta_{12}\beta_{23}}{1-\beta_{12}\beta_{21}}\right]X_{3t} + \left[\frac{\varepsilon_{1t}+\beta_{21}u_{2t}}{1-\beta_{12}\beta_{21}}\right] \\ y_{2t} &= \left[\frac{\gamma_{11}\beta_{21}}{1-\beta_{12}\beta_{21}}\right]X_{1t} + \left[\frac{\gamma_{11}+\beta_{21}\gamma_{21}}{1-\beta_{12}\beta_{21}}\right]X_{2t} + \left[\frac{\gamma_{23}}{1-\beta_{12}\beta_{21}}X_{3t}\right] \\ &+ \left[\frac{\beta_{12}u_{1t}+u_{1t})}{1-\beta_{12}\beta_{21}}\right] \end{aligned} \quad (12)$$

We used equation 12 to produce the values of the dependent variables at each sample point.

6. The procedure described above is repeated over replications R = 20, 50, 100, 150, and
7. (viii) With the generated data sets for $y_{1t}, y_{2t}, y_{3t}, X_{1t}, X_{2t}$,and X_{3t} structural parameters were generated using FIML, 3SLS, LIMF, 2SLS, , ILS, and OLS estimators. (ix) Finally, we obtained the residuals of the estimates for further review and analysis.

The Estimators

If OLS is applied to an equation in a simultaneous model, there will usually be more than one current endogenous variable in the relation, and whichever variable is selected as the dependent variable, the remaining endogenous variables that are correlated with the disturbance term will appear in the equation as explanatory variables. Hence, the OLS will be biased and inconsistent. In the simultaneous equation models where the special assumptions of a recursive system are not valid, the valid estimating techniques are ILS, 2SLS, LIML, 3SLS and FIML.

Indirect Least Squares or Reduced-form Methods (ILS): This method of estimation involves the application of the ordinary least squares to each reduced-form equation of a model. Having obtained the reduced-form estimates in this way, the structural coefficients were obtained using the algebraic transformation of the relationships between the reduced-form and the structural coefficients of the model. ILS is a feasible estimation technique for an equation that is just identified.

Two-Stage Least Squares (2SLS): This is a single-equation method and is probably the most popular for estimating over-identified models. It is directly applied to the structural models. The objective of the two-stage least squares method is to reduce the correlation of the explanatory endogenous variables with error terms as much as possible so that the ordinary least squares method can be appropriately applied to each equation of the structural model. Where there exists exact identifiability, the 2SLS estimates are identical to the ILS estimates

Limited Information Maximum Likelihood (LIMF): This is a single-equation method that uses the principle of maximum likelihood. It is a "limited-information" method because it does not make full use of the information provided by the equations of the model other than those of the particular equation under consideration. The limited information it requires on the other equation of the model is the specification of the truly exogenous variables that are contained in those other equations. It is an appropriate method for estimating over-identified models. The LIML estimator has the same asymptotic variance-covariance matrix as 2SLS. However, the estimates of the asymptotic variances will differ.

Three-Stage Least Squares (3SLS): This method involves the application of the least squares method in three stages. The method is related to the two-stage least squares method in that the first two stages are similar to those of the two-stage method, while in the two-stage method, the two stages involve the structural-form equations of the model, and the first two stages of the three-stage least squares method involve the reduced-form equations of the model. In the third stage of the three-stage least squares method, the generalised least squares technique is applied to correct for any problem of heteroscedasticity that may arise in the model

Full Information Maximum Likelihood (FIMF): This is a system method based on the principle of maximum likelihood. It requires a full knowledge of the structure of the equations in the model. It is computationally more expensive than 3SLS as it involves the solution of non-linear equations. The most practical application of 3SLS and FIML occurs with fairly small models.

Results and Discussion

Our model is exactly identified (by Order and Rank conditions of identification). Hence, the unique estimates of the model are realisable. Two scenarios of the presence of multicollinearity were considered. Scenario 1 relates to cases where the correlation among the exogenous variables is significant at the 1% level designated as High Multicollinearity, while Scenario 2 relates to cases where the correlation among the exogenous variables is insignificant at the 5% level designated as Low Multicollinearity. We simulated a finite data set of sample size = 40 over replications R = 20, 50, 100 and 150 to evaluate the two cases of multicollinearity under consideration. The parameter estimates generated from our estimation are highlighted in tables 1 to 4.

Table 1. Average of Parameter Estimates for *Sample Size=40 over Replications R*=20

		Parameter Estimates Equation 1			Parameter Estimates Equation 2		
Estimator	Level of Multicollinearity	β_{12} (1.8)	γ_{11} (1.2)	γ_{12} (0.6)	β_{21} (0.4)	γ_{22} (0.5)	γ_{23} (1.4)
	Low	0.4880	2.0917	3.7904	1.9980	-2.0296	-12.0598
OLS	High	0.5207	3.3235	1.6274	1.9616	0.1875	-13.4809
	Low	8.3308	-39.5354	-99.5501	2.01144	-2.1363	-12.1603
LIML	High	1.0153	-1.8051	-4.4980	2.2571	-3.0877	-15.6055
	Low	8.3308	-39.5354	-99.5501	2.01144	-2.1363	-12.1603
2SLS	High	1.0153	-1.8051	-4.4980	2.2571	-3.0877	-15.6053
	Low	8.3308	-39.5354	-99.5501	2.01144	-2.1363	-12.1603
ILS	High	1.0153	-1.8051	-4.4980	2.2571	-3.0877	-15.6055
	Low	8.3308	-141.9700	-101.3660	2.0114	-2.1363	-12.16603
3SLS	High	1.0153	-4.4980	-1.7601	1.8784	0.3502	-12.3540
	Low	-0.5334	4.8014	23.9146	1.33489	2.2288	-2.6684
FIML	High	0.8351	1.2333	-1.7886	2.0405	0.0048	-14.6314

Table 1 presents the average of the parameter estimates when the sample size is 40 over 20 replications. A review of the estimates revealed that of the six estimators considered, LIML, 2SLS and ILS produced identical parameter estimates.

Table 2 presents the average of parameter estimates when the sample size is 40 over 50 replications. A review of the estimates revealed that of the six estimators considered, LIML and 2SLS produced identical parameter estimates.

Table 3 presents the average of parameter estimates when the sample size is 40 over 100 replications. Here, similar to the case of 50 replications, LIML and 2SLS produced identical parameter estimates.

Table 4 presents the average of the parameter estimates when the sample size is 40 over 150 replications. Under this scenario, of

Table 2. Average of Parameter Estimates for *Sample Size=40 over Replications R*=50

		Parameter- Estimates Equation 1			Parameter Estimates Equation 2		
Estimator	Level of Multicollinearity	β_{12} (1.8)	γ_{11} (1.2)	γ_{12} (0.6)	β_{21} (0.4)	γ_{22} (0.5)	γ_{23} (1.4)
	Low	0.5045	2.2038	3.3187	1.9785	-1.2544	-12.0612
OLS	High	0.5032	4.0824	1.3123	1.9463	0.2618	-13.2990
	Low	3.2581	-19.6020	-25.4824	1.8961	-0.9107	-11.5692
LIML	High	1.4652	10.9583	-24.8366	3.9684	-38.1473	-18.3811
	Low	3.2581	-19.6020	-25.4824	1.8961	-0.9107	-11.5692
2SLS	High	1.4652	10.9583	-24.8366	3.9684	-38.1473	-18.3811
	Low	2.0035	-0.1073	-2.6265	0.6312	-1.415	2.1547
ILS	High	4.0433	2.3715	-2.2671	0.6342	-39.129	7.3114
	Low	3.2582	-60.6614	-26.2084	2.0034	-1.4163	-12.801
3SLS	High	1.4652	-24.8340	11.0163	1.9563	0.1316	-13.289
	Low	-19.6766	-194.468	568.2325	45.6193	-195.3530	-748.395
FIML	High	0.5733	8.7574	-3.2427	1.9434	0.3035	-13.4245

Table 3. Average of Parameter Estimates for *Sample Size=40 over Replications R*=100

		Parameter Estimates Equation 1			Parameter Estimates Equation 2		
Estimator	Level of Multicollinearity	β_{12} (1.8)	γ_{11} (1.2)	γ_{12} (0.6)	β_{21} (0.4)	γ_{22} (0.5)	γ_{23} (1.4)
	Low	0.5121	2.3426	2.9612	1.9638	-0.7421	-12.4434
OLS	High	0.5073	3.9340	1.3713	1.9448	0.1395	-13.1436
	Low	1.6505	-3.4215	-11.9137	2.0076	-0.8488	-13.5513
LIML	High	1.0365	-0.6845	48.8241	3.7157	-23.8199	-24.1025
	Low	1.6505	-3.4215	-11.9137	2.0076	-0.8488	-13.5513
2SLS	High	1.0365	-0.6845	48.8241	3.7157	-23.8199	-24.1025
	Low	2.0178	0.2571	-0.7066	0.5208	-0.9126	1.4084
ILS	High	3.6042	0.1563	0.7701	0.6334	-22.1673	5.2065
	Low	1.6427	-23.9515	-12.2718	2.0593	-1.0705	-14.1672
3SLS	High	1.0366	-4.5398	0.6456	1.9414	0.0316	-12.9156
	Low	-9.4703	-95.3091	298.4402	23.8973	-97.9823	-382.582
FIML	High	0.6691	7.5479	-5.4429	1.9452	0.1703	-13.2548

the six estimators considered, only LIML and 2SLS generated identical parameter estimates. In summary, in tables 1 to 4, while LIML, 2SLS and ILS produced identical estimates over 20 replications, 50 replication appears to be the turning point for ILS as it dropped off in the production of similar estimates with LIML and 2SLS from 50 replications and above.

Furthermore, in tables 1 to 4, the values in parentheses are the suggested true values of the parameters. We designated the best estimators as those whose average estimates are closest to the true parameter value. For parameter β_{12} whose true value is 1.8, the best estimates are obtained from OLS for both low and high multicollinearity. Similarly, for parameter β_{21} whose true value is 0.4, the best estimates are obtained from OLS for both low and high multicollinearity. By theory, for an equation that is just identified, estimates of parameters obtained by 2SLS, LIML, ILS and 3SLS should be identical (Johnston 1991). From our results, 3SLS, 2SLS and LIML produced virtually similar estimates. For our residual analysis, the six estimators we considered were reclassified into 4 groups vis-à-vis: OLS, ILS, FIML and L23 (for LIML,2SLS and 3SLS).

Table 4. Average of Parameter Estimates for *Sample Size =40 over Replications R*=150

		Parameter Estimates Equation 1			Parameter Estimates Equation 2		
Estimator	Level of Multicollinearity	β_{12} (1.8)	γ_{11} (1.2)	γ_{12} (0.6)	β_{21} (0.4)	γ_{22} (0.5)	γ_{23} (1.4)
	Low	0.5155	2.3845	2.8425	2.0237	-0.5955	-12.5358
OLS	High	0.5111	3.7765	1.4361	1.9514	0.1124	-13.2334
	Low	1.6329	-1.3810	-8.9139	1.9109	-0.3902	-12.6113
LIML	High	4.0277	-51.3888	9.3297	3.1286	-15.9798	-20.4538
	Low	1.6329	-1.3810	-8.9139	1.9109	-0.3902	-12.6113
2SLS	High	4.0277	-51.3888	9.3297	3.1286	-15.9798	-20.4538
	Low	1.9186	0.5685	0.6988	0.5188	-0.6004	1.5265
ILS	High	2.8916	-0.6146	-0.0614	0.5298	-14.0596	4.0739
	Low	1.4335	-15.0677	-9.1526	1.9654	-0.4785	-13.8968
3SLS	High	4.0278	-53.9570	-23.6568	1.9056	0.0891	-12.4936
	Low	-6.0768	-62.1419	198.4648	16.5661	-65.2997	-259.0820
FIML	High	0.6543	7.7370	-4.6519	-0.9192	11.2880	28.1733

Table 5 highlights the further evaluation of results in tables 1 to 4 using the Sum of Squared Residuals (SSR) of the parameter estimates. Hence, the performance of the estimators in the presence of multicollinearity was assessed on the basis of the analysis conducted on the SSR.

Table 5. Sum of Squared Residuals of Parameter Estimates (SSR)

		Parameter Estimates Equation 1			Parameter Estimates Equation 2		
Estimator	Level of Multicollinearity	β_{12} (1.8)	γ_{11} (1.2)	γ_{12} (0.6)	β_{21} (0.4)	γ_{22} (0.5)	γ_{23} (1.4)
	Low	0.5155	2.3845	2.8425	2.0237	-0.5955	-12.5358
OLS	High	0.5111	3.7765	1.4361	1.9514	0.1124	-13.2334
	Low	1.6329	-1.3810	-8.9139	1.9109	-0.3902	-12.6113
LIML	High	4.0277	-51.3888	9.3297	3.1286	-15.9798	-20.4538
	Low	1.6329	-1.3810	-8.9139	1.9109	-0.3902	-12.6113
2SLS	High	4.0277	-51.3888	9.3297	3.1286	-15.9798	-20.4538
	Low	1.9186	0.5685	0.6988	0.5188	-0.6004	1.5265
ILS	High	2.8916	-0.6146	-0.0614	0.5298	-14.0596	4.0739
	Low	1.4335	-15.0677	-9.1526	1.9654	-0.4785	-13.8968
3SLS	High	4.0278	-53.9570	-23.6568	1.9056	0.0891	-12.4936
	Low	-6.0768	-62.1419	198.4648	16.5661	-65.2997	-259.0820
FIML	High	0.6543	7.7370	-4.6519	-0.9192	11.2880	28.1733

From the SSRs in table 5, at both levels of multicollinearity, for the two equations and over all replications, OLS generally rendered the lowest SSR followed by ILS compared to other estimators. Thus, on the basis of our analysis of the residual, OLS performed best in all cases compared with other estimators in both equations. However, the SSRs generated for cases of low multicollinearity are lower than those generated for high multicollinearity. Furthermore, for equation 1, at low multicollinearity, OLS generated SSR values of 151.55, 170.10, 176.10 and 175.90 for replications R = 20, 50, 100 and 150,, respectively. Hence, 100 replications appear to be the turning point in the performance of OLS at this instance. A similar turning point could not be established for OLS in the case of high multicollinearity in equation 1. For equation 2, at low multicollinearity, OLS generated SSR values of 9.95, 20.51, 30.08 and 30.20 for replications R = 20, 50, 100 and 150, respectively. Hence, for this case, the SSR progressively increased as the replications increased. Also for equation 2, at high multicollinearity, OLS generated SSR values of 27.64, 37.47, 39.90 and 37.32 for replications R = 20, 50, 100 and 150, respectively, with a turning point at replication R = 100. Table 6 further highlights the performance of the six estimators on the basis of the SSR of the parameter estimates and also in terms of the reclassification adopted above (OLS, ILS, FIML and L23 (for LIML,2SLS and 3SLS).

Table 6. Performance of Estimators using the Sum of Squared Residuals (SSR)

		Parameter Estimates Equation 1			Parameter Estimates Equation 2		
Estimator	Level of Multicollinearity	β_{12} (1.8)	γ_{11} (1.2)	γ_{12} (0.6)	β_{21} (0.4)	γ_{22} (0.5)	γ_{23} (1.4)
	Low	0.5155	2.3845	2.8425	2.0237	-0.5955	-12.5358
OLS	High	0.5111	3.7765	1.4361	1.9514	0.1124	-13.2334
	Low	1.6329	-1.3810	-8.9139	1.9109	-0.3902	-12.6113
LIML	High	4.0277	-51.3888	9.3297	3.1286	-15.9798	-20.4538
	Low	1.6329	-1.3810	-8.9139	1.9109	-0.3902	-12.6113
2SLS	High	4.0277	-51.3888	9.3297	3.1286	-15.9798	-20.4538
	Low	1.9186	0.5685	0.6988	0.5188	-0.6004	1.5265
ILS	High	2.8916	-0.6146	-0.0614	0.5298	-14.0596	4.0739
	Low	1.4335	-15.0677	-9.1526	1.9654	-0.4785	-13.8968
3SLS	High	4.0278	-53.9570	-23.6568	1.9056	0.0891	-12.4936
	Low	-6.0768	-62.1419	198.4648	16.5661	-65.2997	-259.0820
FIML	High	0.6543	7.7370	-4.6519	-0.9192	11.2880	28.1733

The estimator that generated the lowest SSR is classified as a Good Performer, while the estimator with the next lowest SSR is displayed in brackets. Estimators that generated large numerical SSRs are classified as Large SSRs. These results confirm the superiority of OLS at both levels of multicollinearity for both equations in the model and over the different replications considered. From table 6 it can be deduced that generally, OLS yielded the lowest SSR followed by ILS, while the system estimators generally produced a large SSR.

Conclusions

In a Monte Carlo Simulation (MCS) study, we analysed the residuals of our parameter estimates with a view to evaluating the relative performance of six system estimators (OLS, ILS, 2SLS, 3SLS, LIML and FIML) under varied levels of multicollinearity. Our simultaneous equation model was estimated with a Monte Carlo simulated data set of sample size 40 over different replications R = 20, 50, 100 and 150. We found that, in line with theory and as also confirmed by Johnston (1991) for the just identified equation, the 3SLS, LIML and 2SLS estimators produced virtually identical estimates. For the OLS in equation 1, replication R = 100 appears to be the turning point in the progressive rise of the SSR. The results revealed that irrespective of the magnitude of multicollinearity, OLS yielded the lowest SSR followed by ILS, while all the system estimators generated large SSRs. This result indicates the robustness of the OLS in the presence of multicollinearity. Also, the SSRs generated for cases of low multicollinearity are lower than those generated for cases of high multicollinearity. This is in agreement with the theory on the need to maintain low multicollinearity in econometric models for the optimal performance of the estimators.

Peer Review: Externally peer-reviewed.
Conflict of Interest: The author have no conflict of interest to declare.
Grant Support: The author declared that this study has received no financial support.

ORCID:

Emmanuel Oduntan 0000-0002-6115-480X

REFERENCES

Albert J. and Chib S. (1995), "Bayesian residual analysis for binary response regression models", *Biometrika*, Vol. 82, Issue 4, No. 1, pp. 747–769.

Ardalani-Farsa M. and Zolfaghari S. (2010), "Chaotic time series prediction with residual analysis method using hybrid Elman–NARX neural networks", *Neurocomputing*, Vol. 73, Issues 13–15, pp. 2540-2553.

Baddeley, A., Turner, R., Moller, J. and Hazelton, M. (2005), "Residual Analysis for Spatial Point Processes" *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* Volume 67, No. 5, pages 617-666.

Chaloner, K. and Brant, R. (1988), "A Bayesian approach to outlier detection and residual analysis", Biometrika, Volume 75, Issue 4, 1, pages 651–659.

Chaloner K. (1991), "Bayesian residual analysis in the presence of censoring", Biometrika, Vol. 78, Issue 3, No. 1, pp. 637–644.

Chesher A and Irish M (1987), "Residual analysis in the grouped and censored normal linear model", *Journal of Econometrics*, Vol. 34, Issues 1–2, pages 33-61.

Clements, R. A., Schoenberg, F. P. and Schorlemmer, D. (2011), "Residual Analysis Methods for Space-Time Point Processes with Applications to Earthquake Forecast Models in California". *The Annals of Applied Statistics* Vol. 5, No. 4, pages. 2549-2571.

Farias R. B. A. and Branco M. D. (2012), "Latent residual analysis in binary regression with skewed link", (Contributions to the 10th Bayesian Statistics Brazilian Meeting), *Brazilian Journal of Probability and Statistics*, Vol. 26, No. 4, pp. 344 357.

Johnston, J. (1991), Econometrics Methods, McGraw–Hill Book Company.

Schoenberg F. P. (2003), "Multidimensional Residual Analysis of Point Process Models for Earthquake Occurrences", *Journal of the American Statistical Association* Vol. 98, No. 464, pages 789-795.

Jalilian, A. H. and Vahidi-Asl, M. Q. (2011), "Residual Analysis for Inhomogeneous Neyman-Scott Processes", Scandinavian Journal of Statistics. Volume 38, No. 4, pages 617-630.

Nagar, A.L. (1960), "A Monte Carlo Study of Alternative Simultaneous Equation Estimators", *Econometrica*, Vol. 28, No. 3, pages 573-590.

Oduntan, Emmanuel A. and Iyaniwura, J. O. (2021) "A Monte Carlo Simulation Framework on the Relative Performance of System Estimators in the Presence of Multicollinearity" Cogent Social Sciences, Vol. 7, Issue 1.

Ogata Y. (1988), "Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes", Journal of the American Statistical Association, Vol. 83, Issue 401.

Pagan, A.R. and Hall, A.D. (1983), "Diagnostic tests as residual analysis", Econometric Reviews, Volume 2, Issue 2.

Zhuang J. (2006), "Second-Order Residual Analysis of Spatiotemporal Point Processes and Applications in Model Evaluation", *Journal of the Royal Statistical Society, (Statistical Methodology), Series B*, 68, No. 4.

How cite this article

Oduntan, E. (2024). A monte carlo study of the residuals of system estimators in the presence of multicollinearity. *EKOIST Journal of Econometrics and Statistics*, 41, 13-21. https://doi.org/10.26650/ekoist.2024.41.1251589