

Endogeneity Effect on AR (1) Models in Small Samples

Yakubu Dekongmene Kanyir¹, John O. Olaomi², Albert Luguterah³

¹University for Development Studies, Navrongo Campus, Tamale, Ghana

²University of South Africa (UNISA), Science Campus, Florida, South Africa

³C.K. Tedam University of Technology and Applied Sciences, Navrongo, Ghana

Email: kd_yak@yahoo.com, Olaomjo@unisa.ac.za, adlugu@yahoo.com

How to cite this paper: Kanyir, Y. D., Olaomi, J. O., & Luguterah, A. (2022). Endogeneity Effect on AR (1) Models in Small Samples. *Modern Economy*, 13, 1194-1205. <https://doi.org/10.4236/me.2022.139063>

Received: June 13, 2022

Accepted: September 18, 2022

Published: September 21, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This study examines the endogeneity effect on autoregressive linear models of AR (1) in small samples, making use of the Ordinary Least Square (OLS) estimator, Two-Stage Least Squares (2SLS) estimator, and Generalized Method of Moment (GMM) estimator, based on the sensitivity analysis of sample size and specification errors in estimator determination in linear regression model through the use of Monte Carlo simulation and application to real-life data. The simulation indicates that 2SLS and GMM estimators show the smallest biases when the sample size is varied from $n = 10, 25, 50$ to 100. The estimator that performs best when sample size $n = 10$ across autocorrelation (ρ) and significant correlation (α) at all levels of replication of 10,000 is GMM. In the real-life data, OLS and 2SLS exhibit higher endogeneity characteristics from the dataset used. The empirical analysis base on MSE criteria GMM is the best estimator for dealing with external shock factors to inflations embedded with endogeneity in the linear model. When endogeneity and autocorrelation are bedeviled in a linear AR (1) model, in small samples, using the GMM estimator will provide the best results in small samples than using 2SLS and OLS.

Keywords

Autoregressive Linear Model, Endogeneity, OLS, 2SL, GMM

1. Introduction

Much research has indicated that there is enough evidence for the large samples theory in assessing frameworks in models, what is still lacking is how to get a complete set of the sample in the regression model, which is still not been ap-

appropriately dealt with (Kramer, 1998). The dependence on the theory of asymptotic lead to the most problem of bias and sometimes the inferential accuracy level when the sample size, are small (Philips, 1982; Olaomi & Sangodoyin, 2010). Many statisticians are often concerned about, the less subject to sampling fluctuation statistic presumptions, when they seem to be failing in the model. Wooldridge (2002) did mention the things that bring about endogeneity bias in models such as; error in measurement, variables occurring at the same time, and omission of some important variables.

Cochrane and Orcutt (1949) in their findings reveal that there were highly positive autocorrelated error terms in most economic relations. The findings in Rao and Griliches (1969) indicated in their study that there is more to benefit from when one tries to deal with the presumptions in regressions regarding the predictor variables and disturbance errors in the linear model than its original form. Endogeneity is the variable or a change that sets in from within a system or a model. A variation in customer choice of food with regards to high cholesterol to low cholesterol is an endogenous change that may affect any meaningful marching model (Bedri et al., 2010).

Kennedy (2008) stated that four different issues may significantly introduce endogeneity in OLS regression models such as; errors in variables (measurement error), autoregression, omitted variables, and simultaneous causality. In all of these scenarios, OLS regression many times report biased coefficients instead of estimating the true relationship between the independent variable and the dependent variable, OLS regression mistakenly includes the correlation between the independent variable and the error term in the estimation of the independent variable's coefficient.

Infractions in predictor disturbance term presumptions contain certain vital components for the OLS model. For instance, the predictor outcomes may be wrong when testing it for significance with the parameters. The coefficients are always not as strong as they would have been when you consider their autocorrelations in the estimations of the parameters in the regression model. Lastly, because of the nature of the predictor variables, many at times carry ideas that may be made use of during the process of prediction of values in the future in the linear model.

Clougherty et al. (2016) say endogeneity bias renders coefficient estimates from standard regressions practically difficult to explain as the estimates will be inconsistent in the manner that they do not converge to the exact coefficient values. Some studies have been done by many researchers in estimators and estimation of linear models any time least squares assumptions of error terms of independence and zero correlation within regressors and their error terms are violated making use of Monte Carlo design (Olaomi & Shangodoyin, 2010). Other studies emphasize that no matter how small the presence of endogeneity is, it can lead to biased and inconsistent results which will lead to causal inference (Semadeni et al., 2014), "Little experience is sufficient to show that the traditional machinery of statistical processes is wholly unsuited to the needs of

practical research. Not only does it take cannon to shoot a sparrow, but it misses the sparrow. The elaborate mechanism built on the theory of infinitely large samples is not accurate enough for simple laboratory data. Only by systematically tackling small sample problems on their merits does it seem possible to apply the accurate test to practical data” (Fisher, 1925). It is known that in an autocorrelated but none endogenized model, the Feasible Generalized Least Square (FGLS) estimator is better than the OLS estimator when it comes to efficiency in their estimates. Two-Stage Least Squares (2SLS) estimator similarly performs better than other estimators with the presence of endogeneity in the model and absence of autocorelation (Olaomi & Iyaniwura, 2006; Olaomi, 2008). Infrac-tions in predictor disturbance term presumptions contain certain vital components of the OLS model. For instance, the predictor outcomes may be wrong when testing it for significance with the parameters. The coefficients are always not as strong as they would have been when considering their autocorrelations in the estimations of the parameters in the regression model, lastly, because of the nature of the predictor variables, many at times carry ideas that may be made use of during the process of prediction of values in the future in linear models.

Violation of the presumptions underlining the independence of regressors and disturbance terms in most linear models has brought about the problems of autocorrelation and multicollinearity. All of these have an effect on the estimates, which also affect predictions (Kayode et al., 2012).

Reeb et al. (2012) were of the view that there is still much work that needs to be done to increase knowledge on endogeneity in models and how researchers can provide methods of resolving this crucial methodological problem. It has been proven that large sample properties of estimators can be established, while that of small sample properties typically remains a problem (Adedayo, 2008). One of the estimation procedures in some situations may be preferred due to its ability to give better parameter estimates precisions over the others (Kayode, 2007).

Blundell and Bond (1998) in their studies proposed another method in dealing with endogeneity estimation in a linear model with the technique of Generalized Method of Moments (GMM) aims at exploiting all the conditions between the dependent variables and the disturbance term.

Nicola and Mathias (2017) did extensive work on whether the preference is affected by the support for democracy for a certain number of years regarding the endogeneity of political preferences. What they did was to find out inside countries changes in the individual interest for democracy on the preference for it.

Some methods of estimation in models were developed by (Fair, 1984, 1973) what was left was violations in their least-squares in the model which has the potential to render them not responsive, therefore needs to be given further studies with regards to its sample size, specification error, effects, degree of level of significance and to do that by comparing our results to the other estimators in

literature.

Accordingly, a well-designed study must be clear about how and why variables influence one another and the logic and direction of the relationship must be specified (Larcker & Rusticus, 2007). Therefore, this paper presents results of the endogeneity effect on AR (1) models, in small samples, making use of existing estimators of OLS, 2SLS, and GMM, based on the sensitivity analysis of sample size and specification errors in estimator determination in linear regression model through the use of Monte Carlo simulation when the least square assumption of lack of autocorrelation and zero correlation between regressor and error terms are violated.

This study thus made use of these existing estimators which have been established asymptotically in nature (for large samples) but seek to establish their behaviors in a small samples environment and to find out estimator's in dealing with it when the least square assumption of lack of autocorrelation and zero correlation between regressor and error terms are violated and when there is endogeneity and autocorrelation in the model is present. Also of interest are the characteristics such as rho (correlation between regressor and error term), significant level, and autocorrelation increase in the model. Hence this study included large samples in the design and confirmed their known asymptotic nature in literature during the simulation process.

2. Model

We assume a simple linear regression and nonlinear model in our study as:

$$Y_t = \alpha + \beta X_t + U_t \quad (1)$$

$$Y_t = \alpha + \beta X_t + \gamma X_{t-1} + U_t \quad (2)$$

$$Y_t = \alpha \ell^{x\sqrt{\beta} + U_t} \quad (3)$$

$$Y_t = \alpha \ell^{\beta x_1} + \gamma \ell^{x_2 + U_t} \quad (4)$$

$$U_t = \rho U_{t-1} + \varepsilon_t, \quad X_t = \lambda X_{t-1} + v_t, \quad \varepsilon_t \approx N(0, \sigma^2), \quad U_t \rightarrow AR(1)$$

$$E(X_t, U) \neq 0, \quad E(U_t, U_j) \neq 0, \quad E(\varepsilon_1, \varepsilon_2) \neq 0,$$

$$|\rho, \lambda| < 1, \quad (\alpha, \beta) = (1, 1), \quad r = Cor(U_t, X_t)$$

$$U_t \approx N\left(0, \frac{\sigma^2}{1 - \rho^2}\right), \quad X_t \approx N\left(0, \frac{\sigma^2}{1 - \lambda^2}\right),$$

Y_t endogenous variable, U_t and X_t represents first-order autoregressive variables, ε_t white noise processes, ρ and λ for stationary parameters, α, β are usually assumed to be unity or fixed and significant at α when $E(\varepsilon_1, \varepsilon_2) \neq 0$ and autocorrelation level (ρ).

2.1. Experimental Structure

The study investigated estimators to ascertain autocorrelation levels (ρ), their efficiencies, significance levels (α) of correlation within, X_t and U_t , the ef-

fects they have on the endogenous variable Y_t employing Means Square Error (MSE) and Bias criteria simultaneously. We performed serious sensitivity analysis on GMM, 2SLS, and OLS on the estimation of the stationary parameters α and β when, $E(x_i, u_j) \neq 0$, $E(u_i, u_j) \neq 0$ and $E(\varepsilon_i, \varepsilon_j) \neq 0$ simultaneously as their assumptions are violated and therefore we perform a Monte Carlo experiment on them as well.

Employing the Model (3) above, a value U_0 (for a certain sample size-specific) was generated and drawing a value at random ε_0 coming from this $N(0, 1)$ which was then divided by $\sqrt{1-\rho^2}$. From $N(0, 1)$ t values taking successively and those values were used to calculate the autoregressive U_t , X_t and Z_t which similarly were generated to be AR (1). In all these processes, Monte Carlo experiments involving endogeneity Z_t was drawn once and then held constant throughout the replication process (Nelson & Startz, 1990).

The study used simulation approach as this $Cov(X_t, U_t) \neq 0$, hence the closer to intractability by the procedure of analysis in our sensitivity approach using small sample method during the investigation in Monte Carlo design.

In this sensitivity investigation, the degree of autocorrelation was varying (ρ) 0.4, 0.8, and 0.9. The effect of the sample size was also changing from 10, 15, 25, 50, and 100 during each replication procedure in total 10,000 times in the experimental set. The effectiveness of our estimators was examined by making use of accuracy test criteria of Bias and MSE. We involved a design set of 27 which was spread across the sample size as mentioned earlier to help in the data generation process.

2.2. Data Generation

We used the following to enable us to generate the needed data and they are, U_t , Z_t and X_t . Data were generated to be AR (1) and in the replication process, Z_t is drawn once and equally held constant $Cor(Z, X) > 0.8$. It was also held constant to make sure that estimators are not being driven by frail different variables. With the model above, each of these Z_t , X_t and U_t was generated in AR (1).

The values of $Cor(U_t, X_t)$ and $Cor(U_t, Z_t)$ were computed their values in absolute terms tested against the following significance levels 1%, 2%, and 5% respectively. In the process of selection, after the simulation, anytime $Cor(U_t, X_t)$ is significant and this $Cor(U_t, Z_t)$ is not significant, the series U_t then was selected on the other hand, if $Cor(U_t, Z_t)$ is insignificant then we disregard it in our selection process. This procedure of selections were replicated for each ρ , α and N in 10,000 times. After all the selection procedure Y_t was computed as our endogeneity variable for each selected U_t and X_t to form our model.

3. Best Estimator of the Model Results

Model 3 Monte—Carlo Simulation results (Table 1 & Table 2)

$$y_t = \alpha \ell^{x\sqrt{\beta} + U_t}$$

Table 1. Average bias (estimations).

N	10			15			25			50			100		
ρ	0.4	0.8	0.9	0.4	0.8	0.9	0.4	0.8	0.9	0.4	0.8	0.9	0.4	0.8	0.9
0.01															
OLS	0.5509	0.5425	0.5421	0.5009	0.5300	0.5366	0.5600	0.5548	0.5326	0.5701	0.6015	0.6148	0.4599	0.4713	0.5012
2SLS	0.5914	0.3865	0.6881	0.5899	0.4228	0.3302	0.4210	0.4032	0.4125	0.5953	0.5815	0.4925	0.4259	0.4013	0.4259
GMM	-0.3255	-0.3995	-0.3181	-0.3302	-0.3452	-0.3502	-0.1125	-0.1225	-0.1325	-0.1802	-0.2033	-0.3863	-0.4325	-0.4222	-0.4312
0.02															
OLS	0.6021	0.7025	0.7199	0.6105	0.5925	0.6054	0.6012	0.6201	0.6302	0.6352	0.6023	0.5856	0.4268	0.4329	0.4316
2SLS	0.5548	0.5602	0.5599	0.5662	0.4089	0.5006	0.4896	0.4712	0.4023	0.4412	0.4124	0.4123	0.4132	0.4099	0.4012
GMM	-0.132	-0.1502	-0.1635	-0.3425	-0.2329	-0.3254	-0.4526	-0.4625	0.4025	-0.3206	-0.4025	-0.3123	-0.432	-0.4513	-0.4368
0.05															
OLS	0.3588	0.3682	0.3005	0.4502	0.4402	0.4369	0.3725	0.4012	0.4201	0.4399	0.4612	0.4512	0.4802	0.4995	0.4756
2SLS	0.3512	0.3316	0.3528	0.3856	0.3995	0.4025	0.3528	0.3836	0.4012	0.4113	0.4015	0.3956	0.3866	0.3715	0.3815
GMM	-0.1805	-0.1785	-0.1528	-0.3915	-0.3952	-0.3159	0.2523	0.1206	-0.2254	-0.2235	-0.2299	-0.3355	-0.3402	-0.3271	-0.31526

Table 2. Average MSE = Variance + (Bias)².

N	10			15			25			50			100		
ρ	0.4	0.8	0.9	0.4	0.8	0.9	0.4	0.8	0.9	0.4	0.8	0.9	0.4	0.8	0.9
0.01															
OLS	1.6579	1.9025	1.6470	1.9059	1.9776	2.0098	2.0187	2.4161	2.4188	2.9825	2.9221	3.0020	3.1702	3.3567	3.1140
2SLS	0.3690	0.1682	0.4918	0.3966	0.2366	0.2479	0.2227	0.2303	0.2406	0.3240	0.4417	0.4831	0.3314	0.3253	0.3349
GMM	0.0504	0.0428	0.0446	0.0699	0.0712	0.068	0.0768	0.0987	0.2273	0.1809	0.1067	0.1025	0.3877	0.3196	0.3419
0.02															
OLS	1.5042	1.7090	1.7734	1.4961	1.6134	1.6681	1.9249	1.9868	2.0353	2.8131	2.6152	2.7216	2.3445	2.312	2.3775
2SLS	0.3328	0.3041	0.1320	0.3469	0.3641	0.3398	0.3773	0.3607	0.4258	0.4021	0.3588	0.3922	0.3263	0.3373	0.3466
GMM	0.0537	0.0631	0.0723	0.0645	0.0680	0.0640	0.0602	0.0496	0.0393	0.0676	0.0840	0.0966	0.1081	0.1615	0.1655
0.05															
OLS	1.2155	1.2667	1.4062	1.2258	1.2399	1.5181	2.2648	1.7733	1.8932	2.6274	2.1481	2.1705	2.5419	2.5595	2.5677
2SLS	0.1431	0.1330	0.1423	0.1719	0.1007	0.1045	0.2539	0.2378	0.2552	0.2536	0.0404	0.3590	0.0612	0.3596	0.3845
GMM	0.0726	0.0808	0.0788	0.0849	0.1017	0.1051	0.0862	0.0645	0.0717	0.1413	0.0897	0.0843	0.1470	0.1547	0.1739

3.1. Best Estimator of the Model Results

Table 3 contains the summaries of the best scenario in terms of the estimates in accordance to the criteria of bias and MSE in Small samples concerning our Model (3) which is an intrinsically nonlinear model with one variable in the simulation has also been accomplished. The sensitivity analysis reveals that GMM and 2SLS estimators from the simulation indicate the one with minor biases and

Table 3. Best estimator of Model (3) for each scenario.

n	10			15			25			50			100			
ρ	0.4	0.8	0.9	0.4	0.8	0.9	0.4	0.8	0.9	0.4	0.8	0.9	0.4	0.8	0.9	
Criteria	$\alpha \downarrow$															
	0.01	GMM	2SLS	GMM	GMM	GMM	2SLS	GMM	GMM	GMM	GMM	GMM	GMM	GMM	2SLS	2SLS
BIAS	0.02	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM	2SLS	GMM	GMM	GMM	2SLS	2SLS	2SLS
	0.05	GMM	GMM	GMM	2SLS	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM
	0.01	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM	2SLS	GMM	2SLS
MSE	0.02	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM
	0.05	GMM	GMM	GMM	GMM	2SLS	2SLS	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM

that of OLS the one with the substantial was bias in the findings. In the analysis following the MSE criterion, GMM and 2SLS have the minimum and OLS however has the complete worst performance. The estimator that did well in terms of dealing with sample size $n = 10$ across autocorrelation (ρ) and significant (α) at all levels is GMM, meanwhile, 2SLS did somehow better when the sample size was $n = 100$. When sample size ($n = 10, 25, \text{ and } 50$) the estimator that produces the best outcomes is GMM at all ρ and α levels. 2SLS performs somewhat better when levels of ρ and α respectively. OLS estimator from our discuss researchers indicates that it is a biased estimator and is consistent with models of similar characters in any sensitivity econometric analysis. GMM and 2SLS are unbiased from the analysis same way in makes that they manifest that and indeed they are consistent estimators according to (Koutsoyannis, 2003; Fair, 1984) and should be dependent on when researchers are to deal with a smaller sample size in line with conducting any studies making use of our intrinsically nonlinear model of such nature as we have it in the simulation. The simulation results make use of the bias criterion, and the estimators can be ranked as, GMM, 2SLS, and OLS. In the case of criterion for MSE same estimators can equally be ranked as GMM, 2SLS, and OLS. When sample sizes ($n = 10, 15, 25, \text{ and } 50$) using the criterion of bias, GMM put on the best outcomes at all levels of ρ and α . 2SLS discharge well when sample size ($n = 100, \rho = 0.9$). The findings also convey that GMM performs best at all levels of ρ and α when the sample size is ranked from ($n = 10, n = 15, 25, \text{ and } n = 50$) from the MSE point of view.

3.2. Real-Life Data Application

In this analysis, three different datasets from (World Bank, Bank of Ghana, Ministry of Finance, and Ghana Statistical Service) were applied. Each of the datasets has a small sample of 20 yearly observations from 1998 to 2017. The dataset comprises of Exchange Rate (Monthly Average GHC/USD) from (the Bank of Ghana and Ministry of Finance), International Oil Price (in \$) from (World Bank), Inflation from (Ghana Statistical Service), and Trade Openness (World

Bank). Here in this dataset, factors that contribute to changes in inflation in time regimes from s 1998 through 2017 are assessed for the presence of autocorrelation and endogeneity, and other econometric factors.

3.3. Ordinary Least Square Parameter Estimations

The dataset is applied to Model (4) to investigate the correlational characteristics of external shock factors (exchange rate, oil price, and trade openness) on Ghana's inflation. The model is

$$\text{Inf} = \varpi_0 + \varpi_1(\text{exch}) + \varpi_2(\text{oil}) + \varpi_3(\text{To}). \quad (4)$$

3.4. Endogeneity Test

Wu Hausman test in **Table 4** was conducted to check the existence of correlation between independent variables and the error term in the model. The test revealed *p*-value greater than 0.05, indicating the presence of endogeneity in the model.

3.5. The 2SLS Estimates

The 2SLS estimation technique was used on the dataset based on the instrumental variable model

$$\text{Inf} = \varpi_0 + \varpi_1(\text{exch}) + \varpi_2(\text{oil}) + \varpi_3(\text{To}) | \varpi_1(\text{exch}) + \varpi_2(\text{oil}) + \varpi_4(\text{Exp}) \quad (5)$$

$\varpi_v = (\varpi_0, \varpi_1, \varpi_2, \varpi_3, \varpi_4)^T$ is a vector of parameters.

Table 5 shows OLS, the 2SLS suggests that Ghana's inflation increases (75.306) when there are decreasing exchange rates (-2.518) oil prices (-0.234), and openness to trade (-51.169). The results on the model are not satisfactory for 2SLS as all the parameters were not significant. This is because the 2SLS result has no power to control endogeneity in the model.

3.6. The Generalised Method of Moment Estimator

The GMM model controls for endogeneity by internally transforming the data

Table 4. Endogeneity test results.

Test	df1	df2	Statistics	<i>p</i> -value
Weak instrument	1	16	0.107	0.748
Wu-Hausman	1	15	0.286	0.601

Table 5. 2SLS estimates of Model (5).

Parameter	Estimate	Std. Error	<i>t</i> -value	<i>p</i> -value
ϖ_0 (constants)	75.306	248.069	0.304	0.765
ϖ_1 (exch)	-2.518	11.351	-0.222	0.827
ϖ_2 (oil)	-0.234	0.491	-0.476	0.640
ϖ_3 (To)	-51.169	245.542	-0.208	0.838

and by including lagged values of the dependent variable. In this, the GMM model provides a better estimation method compared to the OLS model and the 2SLS. Results in **Table 6** show all the factors are significant, an exchange rate (0.019), oil price (0.002), and openness to trade (0.005). Only prices of oil are significant (p -value = 0.002) and negatively (t -value = -3.043) correlated to Inflation, the other factors like exchange rate, and openness to trade showed a positive correlation to inflation in the GMM estimation.

3.7. Selection of Best Estimator for the Dataset

All three estimators have demonstrated differences in their capacities for econometric properties in small samples when apply to the dataset. These three estimators (OLS, 2SLS, and GMM) were then computed. The estimators after the analysis were compared based on mean square error (MSE) criteria. The results in **Table 7** show that GMM is a better estimator to consider when having a small sample with the presence of endogeneity and autocorrelation in the dataset than OLS and 2SLS in computing model parameters of inflation dependent on shock factors such as exchange rate, oil prices and openness to trade. For MSE the smallest the value the better and the closer it is to find the best line fit on the dataset and again the more perfect the estimator is, hence GMM estimator is more suitable.

Table 6. GMM estimates of Model (4).

Parameter	Estimate	Std. Error	t -value	p -value
ϖ_0 (constants)	5.969	10.827	0.551	0.581
ϖ_1 (exch)	0.5957	0.613	0.971	0.019
ϖ_2 (oil)	-0.098	0.0325	-3.043	0.002
ϖ_3 (To)	17.480	12.907	1.354	0.005

Table 7. Selection of best estimator for the dataset.

Estimator	MAE	MSE	RMSE
GMM	4.020	27.858	5.278
2SLS	7.101	105.889	10.290
OLS	69.999	5005.841	70.751

4. Conclusion

In the sensitivity analysis of the endogeneity effect on an autoregressive linear model of Order (1) in small samples in this intrinsically nonlinear regression model with one variable to determine the best estimator using OLS, 2SLS, and GMM respectively, we were able to attain its expected results in the simulation. When endogeneity and autocorrelation are assailed in a nonlinear autoregressive

model of Order (1) then using GMM and 2SLS estimators stand the chance to produce the best results in small samples than using OLS estimator. Furthermore, GMM estimators also represent more perfect results than 2SLS and OLS when the sample size is ($n = 100$) across all specifications. When there is an increase in autocorrelation and the sample size is small, efficiency reduces in 2SLS and OLS accept GMM at all levels of ρ and α respectively. Sample size issue has been a worrying case to a lot of empirical applied studies in literature; from the simulation, such workers can have a breath since they can make use of GMM and 2SLS estimators as a solution such as in Model (3) with all underlining conditions intrinsically nonlinear regression model therein. The effect of the error term, the extent of correlation in Model (3), and specification error when dealing with the endogeneity effect with a minimum bias has been accomplished and the GMM estimate causes the best outcome across all levels. The best estimator ranking from the analysis is GMM, 2SLS, and OLS. The OLS and 2SLS exhibit higher characteristics of endogeneity from the dataset analyzed used. The empirical analysis puts GMM as the best estimator for handling and controlling endogeneity on external shock factors to inflations and by extension in small samples when there is endogeneity presence in the model.

The limitations of the study are it is difficult to consider which sample was the smallest as a researcher in the process and also 2SLS estimator has no in-build mechanisms to internally transform the dataset when there is the detection of endogeneity present in the dataset.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- Adedayo, A. A. (2008). Comparative Performance of the Limited Information Techniques in a Two-Equation Structural Model. *European Journal of Scientific Research*, 20, 197-205.
- Bedri, K., Onur, T., & Selahattin, T. (2010). A Direct Test of the Endogeneity of Money. *Turkish Economic Association Discussion*, 29, 577-585.
- Blundell, R., & Bond, S. (1998). Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics*, 87, 115-143. [https://doi.org/10.1016/S0304-4076\(98\)00009-8](https://doi.org/10.1016/S0304-4076(98)00009-8)
- Clougherty, J.A., Duso, T., & Muck, J. (2016). Correcting for Self-selecting Based Endogeneity in Management Research: Review, Recommendations and Simulations. *Organizational Research Methods*, 19, 286-347.
- Cochrane, D., & Orcutt, G. H. (1949). Application of Least Square to Relationship Containing Autocorrelated Error Terms. *Journal of the American Statistical Association*, 44, 32-61. <https://doi.org/10.1080/01621459.1949.10483290>
- Fair, R. C. (1973). A Comparison of Alternative Estimators of Macroeconomic Models. *International Economic Review*, 14, 261-277. <https://doi.org/10.2307/2525920>
- Fair, R. C. (1984). *Specification, Estimation, and Analysis of Macro Econometric Models* (pp. 210-214). Harvard University Press. <http://books.google.com>

- Fisher, R. (1925). Theory of Statistical Estimation. *Proceedings of the Cambridge Philosophical Society*, 22, 700-725. <https://doi.org/10.1017/S0305004100009580>
- Kayode, A., (2007). Performance of Some Estimators of a Linear Model with Autocorrelated Error Terms in the Presence of Multicollinearity. *Research Journal of Applied Science*, 2, 536-543.
- Kayode, A., Alao, R. F., & Femi J. A. (2012). Effect of Multicollinearity and Autocorrelation on Predictive Ability of Some Estimators of Linear Regression Model. *Mathematical Theory and Modeling*, 2, 41-52.
- Kennedy, P. (2008). *A Guide to Econometrics* (2nd ed.). Blackwell Publishing Ltd.
- Koutsoyannis, A. (2003). *Theory of Econometrics* (2nd ed., pp. 200-223). Palgrave Publishers.
- Kramer, W. (1998). Asymptotic Equivalence of Ordinary Least Squares and Generalized Least Squares with Trending Regressors and Stationary Autoregressive Disturbances. In R. Galata, & H. Küchenhoff (Eds.), *Econometrics in Theory and Practice* (pp. 137-142). Physica-Verlag HD. https://doi.org/10.1007/978-3-642-47027-1_13
- Larcker, D., & Rusticus, T. (2007). Endogeneity and Empirical Accounting Research. *European Accounting Review*, 16, 207-215. <https://doi.org/10.1080/09638180701269905>
- Nelson, C. R., & Startz, R. (1990). The Distributions of Instrumental Variables Estimator and Its T-Ratio When the Instrument Is a Poor One. *The Journal of Business*, 63, 125-140. <https://doi.org/10.1086/296497>
- Nicola, F.S., & Matthias, S. (2017). On the Endogeneity of Political Preferences. *Political Economy Research Reports*, 347, 1145-1148.
- Olaomi, J. O. (2008). Performance of the Estimators of Linear Regression Model with Autocorrelated Error Terms Which Are Also Correlated with the Geometric Trended Regressor. *European Journal of Scientific Research*, 20, 187-196.
- Olaomi, J. O., & Iyaniwura, J. O. (2006). The Efficiency of GLS Estimators in a Linear Regression Model with Autocorrelated Error Terms Which Are Also Correlated with the Regressor. *International Journal of Biological and Physical Sciences*, No. 11, 129-133.
- Olaomi, J. O., & Shangodoyin D.K. (2010). Comparative Study of Estimators in Autocorrelated-Endogenized Linear Model. *Interstat*, No. 6, 1-10. <http://interstat.statjournals.net>
- Philips, P. C. B. (1982). *Small Sample Distribution Theory in Econometric Models of Simultaneous Equations*. Discussion Paper, Cowles Foundation.
- Reeb, D., Sakakibara, M., & Mahmood, I. (2012). From the Editors: Endogeneity in International Business Research. *Journal of International Business Studies*, 43, 211-218. <https://doi.org/10.1057/jibs.2011.60>
- Semadeni, M., Withers, M. C., & Certo, S. T. (2014). The Perils of Endogeneity and Instrumental Variables in Strategy Research: Understanding through Simulations. *Strategic Management Journal*, 35, 1070-1079. <https://doi.org/10.1002/smj.2136>
- Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.
- Rao, P., & Griliches, Z. (1969). Small Sample Properties of Several Two-Stage Regression Methods in the Context of Autocorrelated Errors. *Journal of the American Statistical Association*, 64, 251-272. <https://doi.org/10.1080/01621459.1969.10500968>

Appendix

Datasets Source

s/n	Variable data	Description	Period	Source
1	GDP	The data is published annual I have adjusted it for monthly periods	1998-2017	Ghana Statistical Service
2	Interest rate	data are presented as the same as from source	1998-2017	Bank of Ghana
3	inflation rate	data are presented as the same as from the source	1998-2017	Ghana Statistical Service
4	exchange rate	data are presented as the same as from the source	1998-2017	Bank of Ghana, Ghana Ministry of finance
5	international oil price	adjusted	1998-2017	World Bank
6	FDI	data is in the percentage of Ghana's GDP	1998-2017	Ghana Investment and promotion cancel
7	Trade Openness	this is calculated as (Export + Import)/GPD	1998-2017	World Bank