International Journal of Modern Mathematical Sciences, 2014, 10(3): 239-246

Modern Scientific Press

International Journal of Modern Mathematical Sciences

ISSN:2166-286X

Journal homepage:www.ModernScientificPress.com/Journals/ijmms.aspx

Florida, USA

Article

On the Effect of Autocorrelation in Regression Model due to Specification Error

Babatunde O. S[1,*], Ikughur A. J[2], Ogunmola A.O[2], Oguntunde P. E[3]

[1]*Mathematics, Statistics and Computer Science Department, Kwararafa University Wukari, P.M.B 1019, Wukari, Taraba State, Nigeria*

[2]*Mathematics and Statistics Department, Federal University Wukari, P.M.B 1020, Wukari, Taraba State, Nigeria*

[3]*Department of Mathematics, Covenant University, Ota, Ogun State, Nigeria*

* Author to whom correspondence should be addressed; Email: olamee2323@gmail.com

Article history: Received 19 February 2014, Received in revised form 19 June 2014, Accepted 26 June 2014, Published 29 June 2014.

Abstract: The main cause of autocorrelation is omitted variables from the model. When an important independent variable is omitted from a model, its effect on the dependent variable becomes part of the error term. Hence, if the omitted variable has a positive or negative correlation with the dependent variable, it is likely to cause error terms that are positively or negative correlated. There are number of tests for specification error in detecting the errors of omitted variables from a regression analysis, one rarely knows the best test to use. This research uses bootstrapping experiment and some properties which estimators should possess if they are to be accepted as good and satisfactory estimates of the population parameters, the models investigated in the bootstrapping experiment consist of two autocorrelation models with autocorrelation level $\rho = 0.5$ and 0.9. A bootstrap simulation approach was used to generate data for each of the models at different sample sizes (n) 20, 30, 50, and 80 respectively each with 100 replications(r). For the models considered, the experiment reveals that the estimated β's were seriously affected by autocorrelation which may be due to omitted variables as the autocorrelation level varies in the different models (i.e. it produces a bias and inefficient estimator).

Keywords: specification error, autocorrelation, omitted variables, bootstrapping, estimators.

2000 Mathematical Subject Classification: 62J05

---

1. Introduction

Specification error is the error associated with the specification of the model, which can be take many forms such as omission of relevant variable, inclusion of unnecessary variables, errors of measurement etc.

When a relevant variable in the model is excluded, the specification will affect the properties of OLS estimator, in the presence of such error, OLS estimators will be bias. Some important questions that arise in the specification of the model include what variables should be included in the model, what are the probabilistic assumptions made about the y_t(dependent variable), x_t(independent variable) and u_t. (unobserved error).

Autocorrelation denotes the correlation between a time series y_t and its own lagged values y_{t-s} with s = - ∞ ….∞.

$$y_t = \rho y_{t-1} + u_t$$

Or consider the error term in a linear

$$y_t = \beta X_t + \varepsilon_t$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t$$

An assumption in the linear regression model is that of zero value of the disturbance term covariance at all possible pairs of observation point. This is referred to as absence of autocorrelation of the error terms and when the disturbance term at any particular period is correlated with any other value of the disturbance term in the series, then we have autocorrelation. Some sources of autocorrelation include omitting explanatory variables, misspecification of the true random error. In many studies, the problem of autocorrelation has been subjected to both the theoretical and empirical investigation in the areas of investigating its consequences, detecting its presence and finding appropriate estimation methods. Test for detecting the presence of autocorrelation and alternative consistence methods of estimating linear methods with autocorrelated disturbance terms include the use of OLS and there are several approaches to resolving problems of autocorrelation, this include Lagging dependent variable, differencing the Dependent variable, GLS and ARIMA.

One common view of autocorrelation is that it is a technical violation of an OLS assumption that leads to incorrect estimates of the standard errors. From this perspective, analysts view autocorrelation as a potential sign of improper theoretical specification rather than a technical violation of an estimator assumption (Beck, 1985; Hendry and Mizon, 1978; Mizon, 1995). Regardless of which perspective on autocorrelation is adopted, lagged dependent variables have been proposed and utilized (for good or ill) as a solution. The effort of many researchers have been very rewarding like Cochrane and Orcutt (1949) brought to focus the attention of economist on the fact that the presence of autocorrelated errors terms require some modification of the usual least squares method of estimation. Bootstrap is a particular resampling scheme with replacement. In statistics and econometrics, bootstrapping has come to mean to resample repeatedly and randomly form an original initial sample using each bootstrapped sample to compute a statistic. The resulting empirical distribution of the statistic is then examined and interpreted as an approximation to the true sampling distribution.

2. Methods

Considering the classical statistical linear model

$$Y = XB + U_t$$

$$U_t = \rho U_{t-1} + \varepsilon_t, \qquad var(\varepsilon_t) = \sigma^2$$

$$|\rho| < 1$$

$$\varepsilon_t \sim NID(0, \sigma^2), \qquad U_0 \sim N(0, {\sigma^2}_{\varepsilon_t}) \qquad ,$$

$$where\ t = 1,2, \ldots \ldots . p$$

$Y = px1$ vector observations on a sample space

$X = pxk$ nonstochastic design matrix of rank k

$\beta$is a k $-$dimensional fixed vector of unknown parameters.

$U = px1$ vector of unobservable random variable with$(0,1)$ and a finite covariance matrix.

Consider a two model of the form

Model	Specification	Problem
I.	True: $y_t = 10.0 + 5.0x_{1t} - 2.0x_{2t} + u_t$	
	Null: $y_t = \ \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + 0.5u_{t-1} + v_t$	Autocorrelation
II.	True: $y_t = 10.0 + 5.0x_{1t} - 2.0x_{2t} + u_t$	
	Null: $y_t = \ \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + 0.9u_{t-1} + v_t$	Autocorrelation

Where the true model is specifying a correct without autocorrelation problem and the null model is specifying the model that has been affected by autocorrelation.

Observations on the dependent variables are generated according to one of the specification labeled true.

Generation of data

For the autocorrelation data, first obtain the autoregressive error term $\rho u_{t-1} + v_t (\rho = 0.5\ and\ 0.9)$ by generating 21 normal random deviates using Rand () command in Microsoft Excel, I standardized the series obtained. Calculate starting value v_0 by drawing a random value U from $N(0,1)$and divide by $\sqrt{1} - \rho^2$. Use successive value of v_t drawn and initial value of v_0 to calculate $\rho u_{t-1} + v_t$ ignoring the first value in the series to avoid problem of initial value. The process above is repeated in as many times to obtain 100 replicationseach of series $20,30,50, and$ 80 respectively. After generating the autoregressive error term, the experiment is then repeated for each of the sample sizes $20,30,50, and$ 80 for the generation of the dependent variable.

For the bootstrap experiment, we consider the specification labeled True model:

$y_t = 10.0 + 5.0x_{1t} - 2.0x_{2t} + u_t$, assign a numerical values to all the parameter ($\beta_0 = 10, \beta_1 = 5.0, \beta_2 = -2.0$ in the model, the variance σ^2 is also assigned a numerical value on the basis of assumedσ^2, and then the disturbance term U is generated. The U generated was standardized. A random sample of $X$was then selected from a pool of random numbers and numerical values of $y_t = 10.0 + 5.0x_{1t} - 2.0x_{2t} + u_t$ was computed for each sample size using Microsoft Excel software. The $X's$ and $Y's$ generated were then copied from Microsoft Excel into STATA and then bootstrapped and replicated 100times using a STATA command, each replication produces a bootstrap sample which give distinct values of $Y's$ which leads to have different estimate of$\beta's$for each bootstrap sample regression Yon fixed $X's$. The procedure above is then repeated for different sample sizes and was also performed on each of our two models.

Criteria for evaluating the performance of the estimators

In this study, the following criteria were used for comparing evaluation of the performance of our estimators;

Average or mean of estimators in comparison with the true parameter, let $\hat{\beta}$ be the average estimates of the parameter β obtained in the i^{th}bootstrap replication, we compute

$\hat{\beta} = \frac{\sum_i^r \beta}{r}$, where r = number of replications

- Bias$(\hat{\beta}) = \hat{\beta} - \beta$
- Variance$(\hat{\beta}) = \frac{1}{r}\sum_{i=1}^r (\hat{\beta} - \beta)^2$

MSE $(\hat{\beta}) = \frac{1}{r} E(\hat{\beta} - \beta)^2$

3. Results and Discussion

In this study, the principal calculations for each model, estimation procedure, degree of autocorrelation of the error term and each sample size are presented below.

From Table 1, as the autocorrelation level varies the value of $\beta's$ becomes unstable as the sample size varies. For the models considered, the $\beta's$ exhibit damped oscillation in nature with $\beta_1$and β_2 been negative at n= 20 and 30 respectively.

Table 1. Comparison of the models with the sample sizes

	β_0	β_1	β_2
Actual	10.0	5.0	- 2.0
Model 1 n=20	9.9916	4.2268	-2.1552
n=30	10.3243	-0.6252	0.7172
n=50	9.6816	4.0752	-0.6024
n=80	10.5673	0.1606	1.5598
Model 2 n=20	9.7138	3.8262	-1.8792
n=30	10.2175	-0.3775	0.7548
n=50	9.5298	3.0180	0.6547
n=80	9.9406	0.6641	1.9015

From table 2, when the sample size is 20, the β_0 and β_1 exhibit a steady downwards positive bias at which model 1 is best, for β_2,it rose steadily with model 1 as best. When the sample size is 30, β_0 performed well in the two models with model 5 giving a least bias, in β_1, there is a fluctuation with all values been negative, and at β_2, it rose steadily with model 1 as best. At n=50, β_0 and β_1 decreased steadily with model one as best and β_2 fluctuate as model 1 as best. At n=80, β_0 decreased steadily with model 1 as best, $\beta_1$andβ_2 produces a steady increase with all values been positive.

Table 2. Comparison of coefficients of the models based on the same sample size

	$\boldsymbol{\beta_0}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$
Actual	10.0	5.0	- 2.0
Model 1 n=20	9.9916	4.2268	-2.1552
Model 2 n=20	9.7138	3.8262	-1.8792
Model 1 n=30	10.3243	-0.6252	0.7172
Model 2 n=30	10.2175	-0.3775	0.7548
Model 1 n=50	9.6816	4.0752	-0.6024
Model 2 n=50	9.5298	3.0180	0.6547
Model 1 n=80	10.5673	0.1606	1.5598
Model 2 n=80	9.9406	0.6641	1.9015

From table 3, most of the biases produced are negative for both β_0 and $\beta_1$while for β_2, the bias increases as the sample sizes increase for each model.

From table 4, model 1 exhibit damped oscillation with n=20 as best while at model 2 there was a sudden decrease after n=30 with n=20 as best.

Table 3. Bias table based on 3.1

	$\boldsymbol{\beta_0}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$
Actual	10.0	5.0	- 2.0
Model n=20	-0.0084	-0.7732	-0.1552
n=30	0.3243	-5.6252	2.7172
n=50	-0.3184	-0.0752	1.3976
n=80	0.5673	4.8394	3.5598
Model 2 n=20	-0.2862	-1.1738	0.1208
n=30	0.2175	4.6225	2.7548
n=50	-0.4702	-1.9820	2.6547
n=80	-0.0594	-4.3359	3.9015

Table 4. The MSE and its ranking in ascending order

	β_0	β_1	β_2	Mean
Actual	**10.0**	**5.0**	**- 2.0**	
Model 1 n=20	1	2	1	1.3
n=30	3	4	3	3.3
n=50	2	1	2	1.7
n=80	4	3	4	3
Model 2 n=20	3	1	1	1.7
n=30	2	4	4	3.3
n=50	4	2	2	2.7
n=80	1	3	3	2.3

4. Conclusions

In this study, we have examined the performances of the estimators in estimating effect of specification error due to autocorrelation in regression model. Criteria considered are bias, variance and mean square error. Based on the criteria considered, the estimate has been seriously affected by the autocorrelation as the effects on the models considered are not stable with model 2preferable based on the minimum bias and the MSE at large. The effect of increasing autocorrelation of the error terms on the model is significant i.e as the autocorrelation level increases the model get better. As the sample size increases and the coefficient changes, the effect becomes unstable. It produces an unreliable and less precise estimates i.e. bias and inefficient estimates and the change in the coefficient of the autocorrelation level from 0.5 *and* 0.9 has seriously affected the instability of the results obtained.

Also, our simulation results suggest that the performance of the estimator (OLS) depends on the number of replications. In this research study, we observed a number of unexpected results since some of our findings do not follow a conclusive pattern which reveal that the search for best estimators of models plagues by autocorrelation disturbances could be hazardous.

Further study is being carried out to increase the number of models considered, bootstrap replication, sample size to see if the effect of the autocorrelation could be more noticeable.

References

[1] Achen, Christopher H., Why Lagged Dependent Variables Can Supress the Explanatory Power of Other Independent Variables, *Annual Meetings of the Political Methodolody Section of the American Political Science Association, UCLA*, July 20-22, **2000**.

[2] Anya McGuirk & ArisSpanos. The Linear Regression Model with Autocorrelated Errors: Just Say No to Error Autocorrelation. *Annual Meeting of the American Agricultural Economics Association*, July 28-31, **2002**.

[3] Beck, Nathaniel, Comparing Dynamic Specifications: The Case of Presidential Approval. "*Political Analysis*, 3(**1991**):51-87

[4] Cochrane, D. and Orcutt, G. H., Application of Least Squares Regression to Relationships Containing Auto-Correlated Error Terms. *Journal of the American Statistical Association*,44 (245)(**1949**): 32-61.doi:10.1080/01621459.1949.10483290

[5] Effron, B. and Tibshirani R.J., *An Introduction to the Bootstrap*, Chapman & Hall/CRC. New York, **1993**.

[6] Kayode Ayinde, Emmanuel O. Apata, Oluwayemisi O. Alaba. Estimators of Linear Regression Model and Prediction under Some Assumption Violation. *Open Journal of Statistics*, 2(**2012**): 534-546.

[7] Hesterberg, T., Simulation and Bootstrapping for Teaching Statistics. *American Statistical Association: Proceedings of the Section on Statistical Education*, (**1998**): 44-52.

[8] Kayode Ayinde and B.A Oyejola, A Comparative Study of the Performance of the OLS and Some GLS Estimators When Stochastic Regressors are Correlated with the Error Terms. *Research Journal of Applied Sciences*, 2(**2007**): 215-220.

[9] Mizon, Grayham E., A simple message for autocorrelation correctors: Don't, *Journal of Econometrics*, 69(1)(**1995**): 267-288

[10] Thursby, J. G., Alternative Specification Error Tests: A Comparative Study. *Journal of the American Statistical Association*. 74(365)(**1979**): 222-225. DOI: 10.1080/01621459.1979.10481641.

[11] Thursby, J. G and Schmidt, P., Some Properties of Tests for Specification Error in a Linear Regression Model. *Journal of the American Statistical Association*, 72(**1977**): 359

[12] Vaduva, I., Bootstrap Method in Density Estimation, *St. Cerc. Mat., Tom*, 46(3)(**1994**): 397-406