Journal of the Nigerian Association of Mathematical Physics
***Volume34,(March, 2016*), pp 209 – 212**
© J. of NAMP

Weighted Cochrane Two Stage Estimator for Handling Autocorrelation and Heteroscedasticity

[1]Adewale F. Lukman,[2]Oranye Ebele,[1]Okegbade Ibukunand[3]Arowolo Olatunji

[1]Department of Statistics, Ladoke Akintola University of Technology, P.M.B. 4000, Ogbomoso, Oyo State, Nigeria.
[2]Federal University Ndufu Alike, P.M.B. 1010, Ikwo, Ebonyi State.
[3]Department of Mathematics & Statistics, Lagos State Polytechnic, P.M.B. 1007, Ikorodu, Lagos.

Abstract

The Ordinary Least Squares (OLS) estimator is the most popularly used estimator to estimate the parameters in a linear regression model when certain basic assumptions are satisfied. When the problem of autocorrelation and heteroscedasticity jointly exists in a model, OLS estimators, though unbiased, are no longer minimum variance among all linear unbiased estimators. The estimator prposed in this work, based on the standard error and other criteria perform well when compared to OLS. A real data is used to investigate the performance of the proposed method.

Key words: Ordinary Least Square, Weighted Least Square, Weighted Cochrane Two Stage Estimator

1.0 Introduction

Consider the standard linear regression model:

$Y = X\beta + U$ (1)

where X is an $n \times p$ matrix with full rank, Y is a $n \times 1$ vector of dependent variable, β is a $p \times 1$ vector of unknown parameters, and U is the error term such that $E(U) = 0$ and $E(UU')=\sigma^2 I_n$.

The Ordinary Least Squares (OLS) estimator is the most popularly used estimator to estimate the parameters in a linear regression model [1]. The Estimator is defined as:

$\hat{\beta} = (X'X)^{-1}X'Y$ (2)

Under certain classical assumptions, this estimator has some very attractive properties which have made it one of the most powerful estimators.One of the assumptions is that the error term should not be correlated. A violation in this assumption is known as autocorrelation. It is well known that performance of OLS estimator is unsatisfactory in the presence of autocorrelation in that the regression coefficients possess large standard errors [2]. In literature, there are various methods existing to solve this problem. Among them is the Cochrane-Orcutt estimator [3]. Another important assumption of the linear regression model is the assumption of constant variance. A violation in this assumption is known as heteroscedasticity, that is,$E(UU')=\sigma^2 V$. When heteroscedasticity occurs, the OLS estimator is still linear, unbiased and consistent but is not best and no longer has minimum variance [4].Consequently, the statistical inference based on t and F test gives an inaccurate result leading to large variances and statistically insignificant regression coefficient.

A popular method of estimation in the presence of heteroscedasticity is the weighted least square which is a special type of the generalized least square. Inevitably, Heteroscedasticity and Autocorrelation can simultaneously exist in linear regression model estimation[5]. Harrison and McCabe [6] concluded that the power of the Durbin-Watson test is not appreciably affected by Heteroscedasticity. Epps and Epps [5]revealed that autocorrelation can seriously invalidate the use of Goldfeld-Quandt and Glejser tests.Epps and Epps [5] revealed that Durbin-Watson test for autocorrelation is robust to the presence of heteroscedasticity and suggest a correction using Cochrane-Orcutt transformation. In this article the objective is to harmonize Cochrane-Orcutt Estimator and weighted least square with the heteroscedasticity corrected estimates to jointly handle both problem.

Corresponding author: Adewale F. Lukman, E-mail:wale3005@yahoo.com, Tel.: +2347032328232

2.0 Weighted Least Square

Consider the model in (1), where X is an n$\times p$ matrix with full rank, Y is a n$\times$ 1 vector of dependent variable, β is a p$\times$ 1 vector of unknown parameters, and U is the error term such that E$(U) = 0$ and $E(UU')$=$\sigma^2 V$.$\sigma^2 V$is the variance covariance matrix of the error term, V is positive definite. Let P be an n×n symmetric matrix such that $P'P$ =V. Pre-multiply model (1) by P^{-1}, model (1) becomes:

$$P^{-1}Y = P^{-1}XB + P^{-1}u$$

Let $Y^* = P^{-1}Y,\ X^* = P^{-1}X,\ u^* = P^{-1}u$

Therefore the transformed model:

Y*=X*β*+u* (3)

Suchthat E(u*)=0 and Cov(u*)=$\sigma^2 I$. The OLS estimator for the model (4) is defined as:

$$\hat{\beta} = \left(X^{*\prime}X^*\right)^{-1} X^{*\prime}Y = \left(X'P^{-1}PX\right)^{-1} X'P^{-1}PY \qquad (4)$$

where $P^{-1}P = W$

Therefore, equation (4) becomes:

$$\hat{\beta}_W = (X'WX)^{-1}X'WY \qquad (5)$$

which is called weighted Least Squares (WLS) estimator of β.

3.0 Proposed Estimator

3.0.1 Weighted Cochrane-Orcutt Two Stage Estimator

Cochrane-Orcutt Two Stage Least Square [7]and weighted least square [8]are combined sequentially to form weighted Cochrane Two Stage Estimator (WCTSE). This method is adopted to deal with the problem of autocorrelated and heteroscedastic error. The procedures are highlighted as follows:

Compute the ordinary least square estimates of model in (1) as given in (2).Obtain the autocorrelation coefficient ρ and Pre-multiply model (1) by ρ, model (1) becomes:

$$\rho Y = \rho X\beta + \rho U \qquad (6)$$

This can be written as:

$$Y^* = X^*\beta + U^* \qquad (7)$$

where $U^* \sim N(0, \sigma^2 I)$, $Y^* = \rho Y, X^* = \rho X and U^* = \rho U.$

Therefore, OLS Estimator to the transformed model (7) is defined as:

$$\hat{\beta} = \left(X^{*\prime}X^*\right)^{-1}X^{*\prime}Y^* \qquad (8)$$

$$= (X'\rho'\rho X)^{-1}X'\rho'\rho Y \qquad (9)$$

$$=(X'\Omega X)^{-1}X'\Omega Y \qquad (10)$$

where$Y^* = \rho Y = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}$

$$X^* = \rho X = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p} \\ 1 & X_{31} & X_{32} & \cdots & X_{3p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix}$$

$$\Omega = \rho'\rho = \begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Therefore, equation (10) is referred to as Cochrane Two stage least squares.

In order to jointly handle the problem of autocorrelation and heteroscedasticity, the weighted least square Estimator (5) is combined with the Cochrane Two stage least squares Estimator (10).Use Cochrane-Orcutt two stage to find the estimates of the autocorrelation coefficient, ρ, use the residual that has been corrected of the problem of autocorrelation to generate the weight. A Feasible Generalized Least square procedure to correct for heteroscedasticty was adopted to generate the weight [8]. The procedures are as follows:

i. Run the regression of dependent variable (Y*) on the regressors (X*'s) of the transformed data and obtain the squared residuals.

$Y^*=\beta_0^* + \beta_1^* X_1^* + \cdots + \beta_p^* X_P^* + \varepsilon^*$ (11)

where Y* is the transformed dependent variable when autocorrelation has been removed.
X*'s are the transformed regressors when autocorrelation has been removed.

ii. Regress the log of the squared residuals ($\log\varepsilon^{2*}$) on the regressors.

$\log\varepsilon^{2*}=\beta_0^* + \beta_1^* X_1^* + \cdots + \beta_p^* X_P^* + v$ (12)

iii. Call the fitted values from this regression $\hat{y}_t$. Exponentiatethe fitted values from (12) as $\hat{h}=\exp(\hat{y}_t)$. Therefore, weighted least squares is performed defining weight as $W_I = \frac{1}{\hat{h}}$ (13)

That is, the reciprocal of the exponentiated fitted values. This procedure is a modification of the result in [5] that if autocorrelation and heteroscedasticity coexist, the autocorrelation tests are robust. He recommended that the dataset should be transformed using Cochrane–Orcutt iterative method to handle the problem of autocorrelation, then, test the non-autocorrelated error for heteroscedasticty. However, if heteroscedasticty is present suitable correction is applied to the transformed data set that is free of autocorrelation.

4.0 Application Example

Data were collected from Central bank bulletin in Nigeria from 1981 to 2011. The dataset consist of two explanatory variables Recurrent Expenditure (X_1) and Capital Expenditure (X_2) and a Nigerian Nominal Gross Domestic product as the dependent variable (Y).Linear regression model was fitted to the dataset and diagnosed for the following assumptions: (i) Constant Error Variance (ii) Multicollinearity (iii) Autocorrelation (iv) Normality of Error term

Table 1: OLS Regression Estimate

OLS estimator/Goodness Of Fit					**DIAGNOSTIC CHECK**		
Model	**B**	**Std .Error**	**t-ratio**	**Pvalue**	**Statistic**	**Statistic (P-value)**	**VIF**
Constant	-143118	362941	-0.39	0.6963	RHO	0.2944	
X_1	10.8087	0.7248	14.91	0.0000***	Jarque-Bera	2.9183 (0.2324)	5.981
X_2	1.5476	1.9341	0.80	0.4303	DW	1.3432 (0.0138)	5.981
R-square	0.9812				White test	17.8251 (0.0032)	
Adjusted R-SQ	0.9799				RESET	2.3544 (0.115)	
AIC	973.0712			SBIC	977.3732	HQC	974.4735

NOTES: Adjusted R-SQ = Adjusted Coefficient of Determination, AIC = Akaike Criterion, SBIC=Schwarz Criterion, HQC=Hannan-Quinn, RHO = Autocorrelation coefficient.VIF=Variance Inflation Factor, DW=Durbin Watson test for autocorrelation, * indicate stationarity at 10% level of significance.**

From Table 1, the fitted regression model based on is:$\hat{Y} = -143118 + 10.8087X_1 + 1.5476X_2$ (14)

The results confirmed a positive influence of recurrent and capital expenditure on gross domestic product. Also, the P-value confirmed that recurrent expenditure has a significant contribution but capital expenditure is not statistically significant.

From Table 1, the estimated white test statistic and P-value for the linear model are 17.8251 and 0.0032 respectively. This revealed that the constant variance assumption is not satisfied. The variance inflation factors are less than 10, therefore the model suffers from multicollinearity. The estimated Durbin-Watson value (1.3432) and P value (0.0138) revealed that the error terms are correlated.Jarque-Bera statistic (2.9183) and P-value (0.2324) showed that the error tem is normally distributed.The P-value of the regression coefficient showed that recurrent expenditure is significant to Nigerian economic growth.

However, the model suffers two major violations of classical linear regression model assumptions. These are identified as the problem of autocorrelation and heteroscedasticity.Therefore, the model is estimated using Cochrane Two stage estimator in Table 2.

Table 2: Cochrane Two Stage Estimate

Cochrane Two Stage Estimator/Goodness of fit				Diagnostic Checks		
Model	**B**	**Std .Error**	**P value**	**Statistic**	**Value (P-value)**	**VIF**
Constant	-60731.0	343976	0.8611	RHO	0.0390	
X_1^*	10.77	0.7794	0.0000***	Jarque-Bera	11.8775 (0.0026)	3.856
X_2^*	1.62	2.0928	0.4447	DW	1.6438 (0.1035)	3.856
R-square	0.9666			White test	17.5538 (0.0036)	
Adjusted R-SQ	0.9642			SBIC	975.0505	
HQC	972.1509			AIC	970.7485	

NOTES: Adjusted R-SQ = Adjusted Coefficient of Determination, AIC = Akaike Criterion,SBIC=Schwarz Criterion, HQC=Hannan-Quinn,RHO = Autocorrelation coefficient,VIF=Variance Inflation Factor, DW=Durbin Watson test for autocorrelation, * indicate stationarity at 10% level of significance.**

The rho value from Table 1 is estimated to be 0.2944 and this is used for the data transformation. The estimated model using Cochrane Two Stage estimator is:

$$\hat{Y} = -60731.0 + 10.77X_1^* + 1.62X_2^* \quad (15)$$

Durbin-Watson Results from Table 2 showed that the problem of autocorrelation has been handled. However, the assumption of constant variance is still not satisfied since white test p value is less than 5% level of significance.Therefore, weighted Cochrane Two stage estimator is adopted.

Table 3: Weighted Cochrane Two Stage Estimate

Model	B	Std .Error	t-ratio	P value	Statistic	Value (P-value)	VIF
Constant	-57468.7	56361.1	-1.02	0.3166	AIC	173.1426	
X_1^*	8.3184	0.6746	12.33	0.0000***	HQC	174.5450	3.856
X_2^*	5.6550	1.5441	3.662	0.0010***	SBIC	177.4446	3.856
R-square	0.9429						
Adjusted R-SQ	0.9388						

NOTES: Adjusted R-SQ = Adjusted Coefficient of Determination, AIC = Akaike Criterion, SBIC=Schwarz Criterion, HQC=Hannan-Quinn,VIF=Variance Inflation Factor, * indicate stationarity at 10% level of significance.**

The estimated model using Weighted Cochrane Two Stage estimator is:

$$\hat{Y} = -57468.7 + 8.3184X_1^* + 5.6550X_2^* \quad (16)$$

The standard errors are smaller than the ones obtained from OLS and Cochrane-Orcutt Estimators and therefore the t values are much smaller than those obtained by OLS. Therefore, the impact of recurrent and capital expenditure on the Nigerian economic growth is positive and significant at 5%.

5.0 Conclusion

In this article, the focus is to propose an alternative estimator to OLS when autocorrelation and heteroscedasticty co-exist in a linear model. The results show that the standard error of OLS estimates is affected by the presence of both problems and this in turn inflate the value of the estimates. However, the standard error estimate of weighted Cochrane two stage estimator is more efficient and as a result the t-test shows that both recurrent and capital expenditure has a positive and significant effect on the economic growth of Nigeria. Also, in terms of the model adequacy using the Akaike criterion (AIC), Schwarz (SBIC) and Hannan-Quinn (HQC) criterion, the weighted Cochrane provide more precise estimates than OLS. Consequently, the weighted Cochrane two stage estimator jointly handle the problem of autocorrelation and heteroscedasticity in a linear regression model.

6.0 References

[1] Dempster, A. P. (1973). Alternatives to Least Squares in Multiple Regression in D. Kabe and R.P. Gupta (eds.), Multivariate Statistical Inference. Amsterdam North-Holland Publishing Co. 25-40.

[2] Greene, W. (2003). Econometrics Analysis. New Jersey, Prentice Hall.

[3] Gujarati, N. D. (2003). Basic Econometrics (4th Ed.). New Delhi: Tata McGraw-Hill, 748, 807.

[4] Koutsoyiannis, A. (1977). Theory of Econometrics. Palgrave publishers, 2nd Edition.

[5] Epps, T.W. and Epps, M.L. (1977). The robustness of some standard tests for autocorrelation and Heteroscedasticity when both problems are present. Econometrica. 45, 745-752.

[6] Harrison, M.J. and McCabe, B.P.M. (1975). Autocorrelation with Heteroscedasticity: A note on the Robustness of the Durbin-Watson, Geary, and Henshaw tests. Biometrica, 62, 215-216.

[7] Hussein, Y. A and Abdalla, A. A (2012): Generalized Two stages Ridge Regression Estimator for Multicollinearity and Autocorrelated errors.Canadian Journal on Science andEngineering Mathematics, 3, 79-85.

[8] Wooldridge, J. M. (2002). Econometric Analysis of Cross Section and Panel Data. Cambridge, MA: MIT Press.

[9] Central Bank of Nigeria Statistical bulletin, 2011 Edition.