Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.2, No.11, 2012

Estimation under Heteroscedasticity: A Comparative Approach Using Cross-Sectional Data

Adebayo Agunbiade[1] Olawale Adeboye[2*]

1. Department of Mathematical Sciences, Faculty of Science, Olabisi Onabanjo University, Ago-Iwoye, Nigeria
2. Department of Mathematics & Statistics, School of Applied Science, Federal Polytechnic, Ilaro, Nigeria. P. M.B 50.

[*]E-mail of the corresponding authour: adeboye_olawale@yahoo.com

Abstract

A comparative investigation was done analytically for 4 different Estimation Techniques of a newly-designed Audit Fees model with four exogenous variables. The aim is to explore in depth the effects of the problem of heteroscedasticity in a CLRM of cross-sectional data and to determine an appropriate estimation technique(s) in the presence of such heteroscedasticity. Findings revealed that the estimates are virtually identical for three estimators: OLS, WH and NW, while the performance of the fourth estimator, GLS was found to be outstanding, as it completely eliminates the effect of heteroscedasticity by producing a “BLUE” result.

Key Words: Heteroscedasticity, Audit Fees, Exogenous variables, New-design and BLUE.

1. Introduction

One of the most frequently arisen problems in a Classical Linear Regression Model (CLRM) of Cross-Sectional data is Heteroscedasticity. By heteroscedasticity, we meant the existence of some non-constant variance function in a CLRM, Gujarati and Porter (2009). This paper is primarily concerns with heteroscedasticity which is one of the violations of the assumptions made regarding the regression model. The phenomenon, according to Maddala (2005), particularly deals with the problem of unequal error variance in the multiple regression model.

Heteroscedasticity is potentially a serious problem and a researcher needs to know whether it is present in a given situation and then take corrective action. Fox (1997) was of the opinion that the impact of non-constant error variance on the efficiency of Ordinary Least-Squares estimator and on the validity of least-squares inference depends on several factors, which include the sample size, the degree of variation in the σ_i^2 , the configuration of the regressor values and the relationship between the error variance and the X’s. It is therefore not possible to develop wholly general conclusions concerning the harm produced by heteroscedasticity. Thus, this paper therefore investigates the comparative estimation of heteroscedasticity effects using four different estimation techniques namely: Ordinary Least Squares (OLS), White Heteroscedasticity-Consistent Standard Errors and Covariance (WH), Newey-West HAC Standard Errors and Covariance (NW) and Weighted Least Squares (WLS). The performances of these estimators are evaluated based on relevant statistic such as R^2, $\bar{R}^2$, **F**, AIC and SWC. The aim is to explore in depth the phenomena effects of heteroscedasticity presence in cross-sectional data and examine which of the estimators is capable of producing a Best Linear Unbiased Estimates (BLUE) results.

Long and Ervin (2000) confirmed that in the presence of heteroscedasticity, OLS estimates are unbiased, but the usual tests of significance are generally inappropriate and their use can lead to incorrect inferences. Among other things, they suggested that data analysts should correct for heteroscedasticity using Heteroscedasticity Consistent Covariance Matrix (HCCM) whenever there is reason to suspect its presence.

Xavier, Bernadino and Juan (2012) were also of the opinion that with regard to day-to-day imprecision, the phenomenon called heteroscedasticity should be taken into account.

In this paper, we broaden the scope of heteroscedasticity by considering a K-1 variable classical linear regression model where the relation between a response variable Y and predictors is given by

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \text{-------} + \beta_k X_{kt} + \varepsilon_k \quad (1)$$

Where t = 1 , 2 , ---------, n

This model identifies K-1 explanatory variables (regressors) namely X_1, X_2,------- ,X_K and a constant term ε that assumed to influence the dependent variable (regressand). In the literature, model (1) has been thoroughly investigated for heteroscedasticity. **It is well known that when the assumptions of the linear regression model are correct, Ordinary Least Squares (OLS) provides efficient and unbiased estimates of the parameters. When the errors are heteroscedastic, the OLS estimates remains unbiased, but becomes inefficient. More importantly the usual procedures for hypothesis testing are no longer appropriate and their use can lead to incorrect inferences. According to Phoebus,J.D (1978), this means that confidence intervals based on OLS will be unnecessarily larger and as a result, the t and F tests are likely to give us inaccurate results.** Given that heteroscedasticity is common in cross-sectional data, methods that take care of heteroscedasticity are essential for prudent data analysis.

For the purpose of this paper, we shall assume four (4) predictors namely Total Assets (TA), Total Equity (TE), Customers Deposit (CD) and Profit Before Tax (PBT) with Audit Fees (AF) as the regressand variable. All these information are obtained from a cross-sectional data of eleven (11) commercial banks in Nigeria. Figures related to them were extracted from the year 2008, 2009 and 2010 audited financial statements as published by all the eleven banks.

Thus, an Audit Fees model is designed as

$$AF = f(TA, TE, CD, PBT) + \varepsilon \quad (2)$$

Explicitly, we have

$$AF_i = \beta_0 + \beta_1 TA_i + \beta_2 TE_i + \beta_3 CD_i + \beta_4 PBT_i + \varepsilon \quad (3)$$

In the course of this research, we shall demonstrate with great dexterity that the conditional variance of AF_i increases as each of TA_i, TE_i, CD_i and PBT_i increases. That is, the variance of AF is not the same for each of the banks. Hence, there is presence of heteroscedasticity. i.e

$$E(u_i^2) = \sigma_i^2 \quad (4)$$

Ole-Kristian et al (2007) examined the relation between excess audit fees and the implied required rate of return on equity capital in global markets, and they conjecture that when audit fees is excessively large, investors may perceive the auditor to be economically bonded to the client, leading to lack of independence. Meanwhile, they failed to establish a scientific procedure for the appropriate fixing or review of these audit fees, despite all the negative effects of its excess emphasized in their publication. To fill this gap, this paper intends to examine 4 different methods of estimating the parameters of Equation 3, to enable us know the most suitable technique for fitting an appropriate model for the review of Audit Fees in Nigeria banking industry.

Equation (2) also derives its justification from the board room negotiations that usually accompanied the review of Audit fees.

2. Materials And Methods

The estimation of heteroscedasticity effect in classical linear regression model (CLRM) enables us to know which of the following techniques a better estimator is in cross-sectional data:

1. Ordinary Least Squares (OLS) estimation in the presence of heteroscedasticity.
2. The method of Generalised Least Squares (GLS).
3. White heteroscedasticity-Consistent standard errors and covariance.
4. Newey-West HAC standard error and covariance.

2.1 OLS Estimation In The Presence Of Heteroscedasticity

If we introduce heteroscedasticity by letting $E(u_i^2) = \sigma_i^2$ but retain all other assumptions of the classical model, the OLS estimator $\hat{\beta}$ is the same with that of the situation under the assumption of homoscedasticity but its variance is obviously different from the usual variance obtained under the assumption of homoscedasticity.

We consider a two-variable model given as:

$Y_i = \beta_1 + \beta_2 X_i + u_i$

By minimizing the sum of square of error, the OLS estimator of β_2 becomes

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum x_i^2 - (\sum X_i)^2} \quad (5)$$

but its variance is now given by the following expression:

$\text{Var}(\hat{B}_2) = \text{E}[(\hat{\beta}_2 - \beta)^2] = \text{E}[(\sum k_i u_i)^2]$

$\text{Var}(\hat{\beta}_2) = \text{E}(k_1^2 u_1^2 + k_2^2 u_2^2 + \ldots\ldots + k_n^2 u_n^2 + 2 \text{ cross product terms})$

$= \text{E}(k_1^2 u_1^2 + k_2^2 u_2^2 + \ldots\ldots + k_n^2 u_n^2)$

Since the expectation of the cross product terms are zero because of the assumption of no serial correlation.

$\text{Var}(\hat{\beta}) = k_1^2 \text{E}(u_1^2) + k_2^2 E(u_2^2) + \ldots\ldots + k_n^2 E(u_n^2)$

Since, $\text{E}(u_i^2) = \sigma_i^2$, we have,

$\text{Var}(\hat{\beta}) = k_1^2 \sigma_1^2 + k_2^2 \sigma_2^2 + \ldots\ldots + k_n^2 \sigma_n^2$

$= \sum k_i^2 \sigma_i^2$

Since, $k_i = \frac{x_i}{\sum x_i^2}$ (from the linearity property of Gauss – Markov Theorem)

Therefore, $\text{Var}(\hat{\beta}) = \sum[(\frac{x_i}{\sum x_i^2})^2 \sigma_i^2]$

$$= \frac{\sum x_i^2 \sigma_i^2}{(\sum x_i^2)^2} \quad (6)$$

Equation (6) is obviously different from the usual variance formula obtained under the assumption of homoscedasticity, which is given as:

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma_i^2}{\sum x_i^2} \quad (7)$$

If $\sigma_i^2 = \sigma^2$ for each i, the two variance formulas will be identical. This is because $\hat{\beta}_2$ is still linear and unbiased under heteroscedasticity assumption when all other assumptions of CLRM hold. Since the variance of u_i, homoscedastic or heteroscedastic plays no part in the determination of the unbiasedness property.

Also, $\hat{\beta}_2$ is a consistent estimator under the assumption of the CLRM despite heteroscedasticity; that is, as the sample size increases indefinitely (i.e. becomes asymptotically large) the estimated β_2 converges to its true value. Furthermore, it can be shown that under regularity conditions, $\hat{\beta}_2$ is asymptotically normally distributed.

Granted that $\hat{\beta}_2$ is still linear, unbiased and consistent, it is pertinent to note that $\hat{\beta}_2$ is not efficient or best. That is, it does not have minimum variance in the class of unbiased estimators.

Thus, we can easily conclude that $\hat{\beta}_2$ is not BLUE in the presence of heteroscedasticity.

2.2 GLS Estimator

This is the procedure of transforming the original variables in such a way that the transformed variables satisfy the assumptions of classical model and then applying OLS to them. In short, GLS is OLS in the transformed variables that satisfy the standard least squares assumptions. The estimators thus obtained are known as GLS estimators and it is these estimators that are Best, Linear and Unbiased (BLUE).

Unlike the usual OLS method which does not make use of the information available in the unequal variability of the dependent variable Y, i.e. it assigns equal weight or importance to each observation. GLS takes such information into account explicitly and is therefore capable of producing estimators that are BLUE.

To illustrate this, we recall equation (3.1.10):

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

Which for ease of algebraic manipulation, we write as:

$$Y_i = \beta_1 X_{oi} + \beta_2 X_i + u_i \quad (8)$$

Where $X_{oi} = 1$ for each i.

By assuming that the heteroscedastic variances σ_i^2 are known, and divide equation (8) through by σ_i to obtain:

$$\frac{Y_i}{\sigma_i} = \beta_1\left(\frac{X_{oi}}{\sigma_i}\right) + \beta_2\left(\frac{X_i}{\sigma_i}\right) + \left(\frac{u_i}{\sigma_i}\right) \quad (9)$$

Which for ease of operation, we write as:

$$Y_i^* = \beta_1^* X_{oi}^* + \beta_2^* X_i^* + u_i^* \quad (10)$$

We used β_1^* and β_2^* the parameters of the transformed model, to distinguish them from the usual OLS parameters β_1 and β_2.

Hence, $\text{Var}(u_i^*) = E(u_i^*)^2 = E\left(\frac{u_i}{\sigma_i}\right)^2 \quad (11)$

Since, $E(u_i^*) = 0$

$\text{Var}(u_i^*) = \frac{1}{\sigma_i^2} E(u_i^2)$, since σ_i^2 is known

$= \frac{1}{\sigma_i^2}(\sigma_i^2)$, since $E(u_i^2) = \sigma_i^2$

$= 1$, which is a constant.

That is, the variance of the transformed disturbance terms u_i^* is now homoscedastic. Since we are still retaining the other assumptions of the classical model, the finding that it is u^* that is homoscedastic suggest that if we apply OLS to the transformed model (10), it will produce estimators that are BLUE.

In short, the estimated β_1^* and β_2^* are now BLUE and not the OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_2$.

2.3 White Heteroscedasticity-Consistent Standard Errors and Covariance

White, H. (1980) has derived a heteroscedasticity consistent covariance matrix estimator which provides estimates of the coefficient covariances in the presence of heteroscedasticity of unknown form. The white covariance matrix is given by:

$$\hat{\Sigma}_w = \frac{T}{T-\square}(X'X)^{-1}\left(\sum_{\square=1}^{\square} \square_{\square}^2 \square_{\square} \square_{\square}'\right)(X'X)^{-1} \quad (12)$$

Where T is the number of observations, K is the number of regressors and $\square_{\square}$ is the least squares residual.

2.4 Newey-West HAC Standard Errors and Covariance

The white covariance matrix described above assumes that the residuals of the estimated equation are serially uncorrelated. Newey and West (1987) have proposed a more general covariance estimator that is consistent in the presence of both heteroscedasticity and autocorrelation of unknown form.

The Newey-West estimator is given by:

$$\hat{\Sigma}_{NW} = \frac{T}{T-k}(X'X)^{-1}\hat{\Omega}(X'X)^{-1} \qquad (13)$$

Where

$$\hat{\Omega} = \frac{T}{T-k}\{\sum_{t=1}^{T} u_t^2 x_t x_t' + \sum_{v=1}^{q}[(1-\frac{v}{q+1})\sum_{t=v+1}^{T}(x_t u_t u_{t-v} x\_(t-v) + x\_(t-v)\ u\_t\ x\_t\text{^}')]\} \qquad (14)$$

Where q, the truncation lag, is a parameter representing the number of autocorrelations used in evaluating the dynamics of the OLS residuals u_t. Following the suggestion of Newey and West, E-views sets q to:

$$q = \text{floor}\left[4\left(T/100\right)^{2/9}\right] \qquad (15)$$

It is pertinent to note that using the white heteroscedasticity or Newey-West does not change the point estimates of the parameters; only the estimated standard errors are different.

3. Results And Discussion

The data collected is on the operational activities of eleven (11) commercial banks in Nigeria namely First Bank, United Bank for Africa, Zenith Bank, Stanbic ibtc, Skye bank, Union Bank, Access Bank, FCMB, Ecobank, GTBank and Diamond Bank for periods of year 2008, 2009 and 2010.

3.1 Summary of Measures of Model Validity

Here, we attempt to summarize the results of R^2, adjusted R^2, F-statistic, Akaike-info criterion (AIC) and Schwarz criterion (SWC) for the Ordinary Least Square (OLS), Generalized Least Squares (GLS), Newey-West HAC (NW) and White Heteroscedasticity (WH) procedures. This is to facilitate a better synopsis in comparative investigation of the phenomenon and to assist in the determination of which Estimation techniques perform best..

Table 1: Results of Estimation Techniques

YEAR		2008	2009	2010
OLS	R^2	0.4962	0.3538	0.5350
	$\bar{R}^2$	0.1603	-0.0770	0.2250
	F	1.4772 (with P value of 0.31826)	0.8213 (with P value of 0.556)	1.7258 (with P value of 0.2619)
	AIC	36.7444	37.5447	37.9998
	SWC	36.9249	37.7255	38.1806

GLS	R^2	0.9511	0.8702	0.9075
	$\bar{R}^2$	0.9187	07837	0.8459
	F	29.23 (with P-value of 0.000448)	10.0575 (with P-value of 0.0079)	14.7223 (with P-value of 0.002943)
	AIC	35.2123	37.7103	38.1665
	SWC	35.3932	37.8912	38.3473
NW	R^2	0.4962	0.3538	0.5350
	$\bar{R}^2$	0.1603	-0.0770	0.2250
	F	1.4772 (with P-value 0.31826	0.8213 (with P value of 0.556)	1.7258 (with P value of 0.2619)
	AIC	36.7444	37.5447	37.9998
	SWC	36.9249	37.7255	38.1806
WH	R^2	0.4962	0.3538	0.5350
	$\bar{R}^2$	0.1603	-0.0770	0.2250
	F	1.4772 (with P-value 0.31826	0.8213 (with P value of 0.556)	1.7258 (with P value of 0.2619)
	AIC	36.7444	37.5447	37.9998
	SWC	36.9249	37.7255	38.1806

The consistency of opinion or inference exhibited by the summarized results is a thing to be hold in a research of this magnitude. In fact, all the literature already reviewed about heteroscedasticity effect in a CLRM of cross-sectional data have been confirmed or established by these results. That is, only the results produced by WLS are worthy of any meaningful inference in the presence of heteroscedasticity.

According to year 2008 OLS results, the coefficient of determination (R^2) implies that only 49.6% of the variation in auditor's remuneration is explained by all the explanatory variables under consideration. The adjusted R^2 (0.160), akaike info criterion (36.74) and Schwarz criterion (36.92) further confirmed the position of our R^2, which adjudged the model as not a "best goodness of fit".

For the Year 2009, the coefficient of determination (R^2) implies that only 35.4% of the variation in auditor's remuneration is accounted for by all the regressors under consideration. The adjusted R^2 (-7.7%) in fact shows that the model is a poorly fitted one. The duo of akaike info (37.54) and Schwarz (37.72) criterion further confirmed this position, and that of year 2010 result also revealed that the Coefficient of determination (R^2) implies only 53.5% of the variation in auditor's remuneration is explained by the four explanatory variables. The adjusted R^2 (0.22), akaike info criterion (37.99) and Schwarz criterion (38.18) further confirmed the position of our R^2, which adjudged the model to have been poorly fitted.

The result of F statistic for the three years shows that the regression coefficients are not statistically significant, since its P-value is above 0.05.

Meanwhile, the above results are expected for OLS, since we used cross-sectional data which speaks volume of the presence of heteroscedasticity.

The GLS results presented for Year 2008 have taken care of all the deficiencies of OLS results presented/explained for the year 2008. Hence, it is adjudged to be a better model in the presence of heteroscedasticity. In fact, 95.11% variation in auditor's remuneration accounted for by the explanatory variables is one of the best situations for the measure of goodness of fit.

Accordingly, year 2009 results are also adjudged to be better than that of year 2009 OLS results. The coefficient of determination (R^2) implies that 87.02% of the variation in auditor's remunerations is explained by all the explanatory variables as against 35.38% presented by the OLS results.

Year 2010 equally gives a better result than that of the OLS which has a coefficient of determination (R^2) of 0.535. A 90.75% variation in auditor's remuneration, as explained by the explanatory variables makes the GLS model to be valid enough for reasonable inference.

The adjusted R^2, akaike info and Schwarz criterion results also pointed to the fact the GLS models are reasonably valid.

The result of F statistic shows that all the regression coefficients are statistically significant at both 5% and 1% levels of significance for the three years under consideration.

It is pertinent to note that the results of both the White and Newey heteroscedasticity test for the three years do not change the point estimates of the parameters from the ones obtained in the OLS analysis, which confirmed the presence of heteroscedasticity. Only the estimated standard errors of both test differed from that of the OLS estimates. Thus, the model arrived at by the two methods clearly show lack of goodness of fit as observed in OLS.

4. Conclusion

This paper has critically and analytically examined a comparative investigation from different estimation techniques for a newly designed Audit Fees model of four explanatory variables. It was found that, in the presence of heteroscedasticity; OLS, NW,WH produced virtually identical results while GLS results were outstanding.

Based on the results obtained by the empirical analysis of data collected, the following conclusions are therefore arrived at.

(i) That ordinary least squares (OLS), Newey-west (NW) and White Heteroscedasticity (WH) are not appropriate if heteroscedasticity is present in research data.
(ii) That generalized least squares (GLS) or weighted least squares (WLS) is the most appropriate method for estimation, in the presence of heteroscedasticity.
(iii) That cross-sectional data are usually heteroscedastic in nature.
(iv) That the best model for proper fixing and review of Audit Fees in banking industry could only be achieved with the use of GLS technique.

References

Agunbiade,D.A and J.O Iyaniwura (2010). Estimation under Multicollinearity: A Comparative Approach Using Monte Carlo Methods. Science Publications. Journal of Mathematics and Statistics 6(2): 183-192. ISSN 1549-3644.

Fox, J (1997): Applied Regression Analysis, Linear Models, and Related Methods, Sage Publications, California. Pp 306

Gujarati,D.N and D.C. Porter (2009). Basic Econometrics. 5th ed. Mc Graw-Hill, New York. Pp 922

Long,J.S. and L.H Ervin (2000). Using Heteroscedasticity Consistent Standard Errors in The Linear Regression Model. The American Statistician, Vol 54, no 3 pp 217-224. Retrieved April 5th, 2012 from http://www.jstor.org/stable/2685594

Maddala, G.S. (2005). Introduction to econometrics._ Wiley & Sons LTD, New York. Pp 836

Ole-Kristian, H. K. Tony, T. Wayne and K. Young (2007). Impact of Excess Auditor

Remuneration on Cost of Equity Capital Around the World. Retrieved April 4th,2012 from http://leeds-faculty.Colorado.edu/gunny/Workshop series/Papers 0708/Thomas.

Olutola,O. (2003). Principles and Practice of Auditing and Investigations. 1st Edition, Agbo Comm. Press, Lagos.

Phoebus, J.D. (1978). Introductory Econometrics. Springer-Verlag, New York Pp 110-111

White, H. (1980). A Heteroscedasticity Consistent Covariance Matrix Estimator and a Direct Test of Heteroscedasticity" Econometrica, vol 48, Pp 1180-1200

Xavier, F.A., G. Bernadino, and C.M Juan (2012). Uncertainty of Measurement and Heteroscedasticity. Journal of the International Federation of Clinical Chemistry And Laboraory Medicine, ejIFC Vol 14 no 1. Retrieved April 4th, 2012 from

http://www.ifcc.org/ejifcc/vol 14 no1/140103200306n.htm