



ILJS-15-049

Developing a General Consistent Standard Error Estimator under Varying Strengths of Heteroscedasticity

Job*, O. and Oyejola, B. A.

Department of Statistics, University of Ilorin, Ilorin, Nigeria.

Abstract

In econometric studies involving cross-sectional data, the assumption of a constant variance for the disturbance term is a fluke. In consumer budget studies (micro consumption function), the residual variance about the population regression function is very likely to increase with income. Also, in cross-sectional studies of firms the residual variance probably increases with the size of the firm. In a simple linear regression model, the dependent variable Y is explained by Z . Thus we assume $y = f(z) + e$ and postulate that $\text{var}(y) = \text{var}(e) = \sigma^2 Z$.

Let $Z_i = X_i$. By implication we formulate the assumption about $\text{var}(e_i) = \sigma^2 Z_i$ in a rational and fairly general manner. In general, to validate this assumption, it is convenient and quite plausible to specify the form of association $\text{var}(e_i) = \sigma^2 Z_i^g$, where g is the strength of heteroscedasticity and the lower the strength (magnitude) of g , the smaller the difference between the individual variances. Except when $g = 0$, the model is homoscedastic otherwise $|g| \leq 2$ generally. This paper developed a general heteroscedasticity consistent standard error (HCSE) estimator using weight related to regressors that characterizes the random error term denoted by HC5. Comparative studies of the developed estimator with the existing HCSE estimators using various strengths of heteroscedasticity on a continuum scale at sample sizes 25, 30, 35, 40, 45, and 50 were implemented. The OLS estimator remains unbiased and the results showed that the developed estimator is indeed a generalization of all the existing HCSE estimators and proved to be consistent and asymptotically efficient.

Keywords: Linear, Regression Model, Heteroscedasticity Consistent Standard Error Estimator, Monte Carlo Simulation, Error terms, magnitude, weighting factor, Generalization.

1. Introduction

The usual OLS estimator for the variance of unknown model parameters in simple linear regression model minimizes the error sum of squares. Usually heteroscedasticity does not destroy the unbiasedness of the OLS estimators except that it becomes less efficient.

Assuming a simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + e_i, \text{ for } i=1 \dots n, \quad (1)$$

*Corresponding Author: Job, O.
Email: obalowu@unilorin.edu.ng

where $e_i \sim N(0, \sigma^2)$, i.e. homoscedasticity assumption holds. The variance of unknown model parameters is given by:

$$\text{var}(\beta_j) = \frac{\sum \hat{e}_i^2}{n-k} (X'X)^{-1}. \quad (2)$$

This is referred to as ordinary least squares covariance matrix estimator (OLSCME).

Specifically: $\text{var}(b_1) = \frac{\sum \hat{e}_i^2}{(n-k) \sum x_i^2}$, such that $x_i = X_i - \bar{X}$ for a simple linear regression model.

Building on the works of Eicker (1963, 1967) and Huber (1967) using the OLS residuals as estimators of the errors. The sandwich estimator HC0 was derived.

$$HC0 = (X'X)^{-1} X' \Phi X (X'X)^{-1}. \quad (3)$$

The symbol Φ is a diagonal matrix with the squared OLS residuals on the main diagonal and zero elsewhere. Hence equation (3) becomes

$$HC0 = (X'X)^{-1} X' \text{diag}(e_i^2) X (X'X)^{-1}. \quad (4)$$

When the assumption of homoscedasticity is violated, $\Phi \neq \sigma^2 I$ then the variance of the model parameters is no longer given by equation (2). Hinkley (1977) modified the OLSCM when he introduced a degree of freedom correction that inflates each residual by a weighting factor,

$\sqrt{\frac{n}{n-k}}$ which resulted in

$$HC1 = \frac{n}{n-k} (X'X)^{-1} X' \text{diag}(e_i^2) X (X'X)^{-1} = \frac{n}{n-k} HC0, \quad (5)$$

where $k = p + 1$ and p is the number of predictor variables in the regression model. Mackinnon and White (1985) developed another variation of OLS estimator by scaling the squares of residuals with $(1 - h_{ii})^{-1}$ where $h_{ii} = x_i (X'X)^{-1} x_i$, then

$$\text{Var}(e_i) = \sigma^2 (1 - h_{ii}) \neq \sigma^2,$$

that is

$$HC2 = (X'X)^{-1} X' \text{diag}\left[\frac{e_i^2}{(1 - h_{ii})}\right] X (X'X)^{-1}. \quad (6)$$

In such a way that, $(1 - h_{ii})^{-1}$ is a weighting factor.

A third variation by Davidson and Mackinnon (1993) gave an approximation to a more complicated jackknife estimator than the earlier one presented by Mackinnon and White (1985) as:

$$HC3 = (X'X)^{-1} X' \text{diag}\left[\frac{e_i^2}{(1 - h_{ii})^2}\right] X (X'X)^{-1}. \quad (7)$$

Such that $0 \leq h_{ii} \leq 1$, with $(1 - h_{ii})^{-2}$ as a weighting factor. This further inflates the squares of residuals. A fourth variation was derived by Cribari-Neto (2004) as:

$$HC4 = (X'X)^{-1}X' \text{diag}[e_i^2 / (1 - h_{ii})^{\delta_i}]X(X'X)^{-1}, \quad (8)$$

$$\text{where } \delta_i = \min(4, \frac{nh_{ii}}{k}), k = p + 1, \quad (9)$$

with $(1 - h_{ii})^{-\delta_i}$ as a weighting factor and p is the number of predictors in the model. In the previous studies the existing HSCE were subjected to analysis assuming heteroscedasticity of unknown form. The choice of weighting factors focuses on making OLS squares residuals less bias.

2. Methodology

Model Specification

We shall consider a classical simple linear regression model (1) above, where β_0 and β_1 are the unknown true parameters $\beta_0 = 2$, $\beta_1 = 3$ and $e_i \sim N(0, x_i^g)$. The model shall be studied under the strength of heteroscedasticity (g) ranging from -2 to +2. The developed estimator is

$$HC5 = (X'X)^{-1}X' \text{diag}[e_i^2 / (1 - h_{ii})^g]X(X'X)^{-1}, \quad (10)$$

such that $0 \leq h_{ii} \leq 1$ and $(1 - h_{ii})^{-g}$ as a weighting factor for the presence of heteroscedasticity of the form X_{ij}^g where “ g ” is the strength of heteroscedasticity that characterizes the error terms that relate to the predictor in the regression model and is estimable from the bivariate data (x_i, y_i) . In this study, strength of heteroscedasticity “ g ” ranges from -2 to +2.

The Design of Monte Carlo Simulation Studies Used for Data Generation is as follows:

- (I) Formulate the linear regression model for the study.
- (II) Specify the distribution of the normal error term to obtain values for the unobserved error terms (e_i).
- (III) Specify the distribution of the independent variable (X).
- (IV) Specify the values of the unknown true parameters of regression model β_0, β_1 .

- (V) Generate the values of the dependent variable (Y) from equation which involves using the true values of the model parameters, generated values of independent variable (X) and values of the unobserved error terms.
- (VI) Run the OLS regression of Y on X in order to obtain estimates of the residuals ($\hat{e}_i$).
- (VII) Generate the values of the heteroscedastic residuals using the estimated residuals above, that is $\hat{e}'_i = \hat{e}_i * X_i^{\frac{g}{2}}$, $g = -2, -1, -0.5, 0, 0.5, 1, 2$, respectively.
- (VIII) Generate values of the response variable (Y) using the pre-specified values of unknown parameters, independent variable (X) and the heteroscedastic residuals.
- (IX) Using bivariate data set (x_i, y_i) , test for the presence of heteroscedasticity using Goldfeld-Quandt test.
- (X) If heteroscedasticity is present in (IX) go to step (XI) otherwise go to step (V).
- (XI) Regress Y on X using the data set and obtain the squares of residuals ($\hat{e}_i^2$).
- (XII) Regress $\log(\hat{e}_i^2)$ on the $\log(X_i)$, carry out the appropriate test of significance in order to obtain the estimates of the strength of heteroscedasticity ($\hat{g}$) in the model.
- (XIII) If the estimated value of slope ($\hat{g}$) in (XII) is approximately equal to g introduced in the (VII) above and the test of significance supports the estimates in (XII), then subject the data set to the existing HCSE and HC5 estimators. Otherwise, restart the processes above.

Using the data generated with different strengths (g) of heteroscedasticity, estimates were therefore obtained for these estimators with varying sample sizes and 1000 replications.

Illustrative Example

Generation of sample values when sample size (n) =25, parameters $\beta_0=2, \beta_1=3$ and $g=-2$ Using R-Package. The generated error terms distributed mean zero and variance, $\sigma^2 = 1$ follow: `e<-rnorm(25,0,2)`.

`e<-c(0.79259937,0.15159755,0.22236327,4.07372642,-0.66352880,2.73412938, 13339313, 1.06999387, 1.41852392,-0.12010026,-2.39275398,-2.49368351,-0.45371828,0.03439936, -1.97778767, 1.85725137, 2.70763668,-0.25774271, -2.29936163, 3.75785389, 1.67069689, 0.91467720, 1.35848385, 2.66101714,-2.31255552).`

The values for the independent variable(X) simulated follow:

`X<-rnorm(25,5,2)`.

`X<-c( 7.227193, 4.850246, 4.624653, 7.142836, 6.245022, 5.349890, 3.538438, 5.432725, 4.447327, 4.657059, 7.345125, 5.589621, 4.050712, 6.692459, 2.492058, 3.683330, 335663, 8.707044, 6.726798, 1.482140, 2.764793, 7.475252, 5.348404, 7.265780, 8.207077).`

The generated values of the response variable (Y) follow:

$$y=2+3x+e \quad (11)$$

`y<-c( 23.779468, 16.447947, 15.769862, 23.984105, 20.577797, 18.461407, 2.381149, 18.400940, 15.493217, 15.795923, 23.701839, 18.237863, 13.832704, 22.050801, .1967150, 13.302403, 18.414223, 28.122153, 21.808233, 8.0030190, 10.484203, 24.544614, 8.199790, 24.153124, 26.357462).`

Obtain the estimates of residuals by regressing Y on X, and then generate values for heteroscedastic residuals as follows:

$$\hat{e}_i^* < -\hat{e}_i * x_i^{g/2}. \quad (12)$$

`$\hat{e}_i^* < -c(0.09788777, -0.10278944 -0.10409695, 0.55559569, -0.15726749, 0.41173639, -0.23416495, 0.10276592, 0.15123638, -0.17525494,-0.33353644, -0.53099915, -0.31943055, -0.02657525, -1.27945905, 0.25241322, 0.40723524, 0.00101939,-.37216107, 1.55659729, 0.18982553, 0.11885831, 0.15457802, 0.35578338, -0.26376994).$`

The values for the heteroscedastic dependent variable (Y) become:

$$y_i = 2 + 3x + \hat{e}_i^*. \quad (13)$$

y <-c(23.646433, 16.587695, 15.917761, 23.395921, 20.729712, 18.071472, 12.692068, 18.317463, 15.391162, 16.013998, 23.996650, 18.783392, 14.213348, 22.058451, 9.584670, 13.122348, 18.029221, 28.041093, 22.160426, 6.585554, 10.394600, 24.383083, 18.067059, 23.761022, 26.556359).

Result of test for the presence of heteroscedasticity using Goldfeld-Quandt Test:

Test statistic, GQ = 4.2001, with df1 = 11, df2 = 10, P-value = 0.01574.

Conclusion: Since the P-value is less than the pre-selected alpha=0.05, this indicates heteroscedasticity is likely present. Hence we can now estimate the strength of heteroscedasticity from the simulated data above.

To be followed by the next line:

The estimation of the strength of heteroscedasticity(g) in order to determine its form can be obtained by regressing $\log(e_i^2)$ on $\log(x_i)$. The estimates of intercept and g are in table below:

Variable	Estimate	s.e	t-value	P-value
Intercept	0.176	1.6173	0.109	0.9140
G	-2.0110	0.9637	-2.087	0.0482

From table above the strength of heteroscedasticity, g=-2.0110 and the form is given as

$$\text{var}(e_i) = \sigma^2 Z_i = \sigma^2 x_i^g = \sigma^2 x_i^{-2} \quad (14)$$

From Monte simulation process, the estimates of model parameters based on 1000 replications follows:

AVERAGE VALUE OF $\hat{\beta}_0 = 2.006492$, BIAS ($\hat{\beta}_0$) = 0.0065, ABSOLUTE BIAS ($\hat{\beta}_0$) = 0.0065

AVERAGE VALUE OF $\hat{\beta}_1 = 2.099893$, BIAS ($\hat{\beta}_1$) = -0.00344, ABSOLUTE BIAS ($\hat{\beta}_1$) = 0.00344.

Subjecting the existing HCSE and HC5 estimators to the bivariate data(x_i, y_i). The estimates of covariances, when g=-2, sample size (n) =25 are given as below:

$$\text{HC0} = \begin{pmatrix} 0.30949320 & -0.048650450 \\ & 0.007751424 \end{pmatrix} \quad \text{HC1} = \begin{pmatrix} 0.33640565 & -0.052880923 \\ & 0.008425461 \end{pmatrix}$$

$$\text{HC2} = \begin{pmatrix} 0.38641945 & -0.060778546 \\ & 0.009671662 \end{pmatrix} \quad \text{HC3} = \begin{pmatrix} 0.48403331 & -0.07617608 \\ & 0.01210936 \end{pmatrix}$$

$$\text{HC4} = \begin{pmatrix} 0.56679233 & -0.08933222 \\ & 0.01419241 \end{pmatrix} \quad \text{HC5} = \begin{pmatrix} 0.2005973 & -0.031494896 \\ & 0.005034438 \end{pmatrix}$$

The Standard error estimates of the model parameters shall be obtained by taking the square root of the entries on the main diagonals of the co variances matrices. See Appendix “A1” and Appendix “A2” respectively.

3. Results and Discussion

From Table1 it can be verified that OLS estimator remains unbiased even when regression homoscedasticity assumption is violated (see columns 3 and 4). The estimates of bias and absolute bias are very close to zero and particularly at $g=0$ the bias and absolute bias equal zero exactly. The standard error estimates is a sufficient criterion to judge the performance of the heteroscedasticity consistent covariance matrix estimators (HCCME), since

$$MSE(\beta_j) = Var(\beta_j) + [Bias(\beta_j)]^2, \text{ if } Bias(\beta_j) = 0 \text{ then } MSE(\beta_j) = Var(\beta_j)$$

Table1: Average Estimate of $\hat{\beta}_0$ and $\hat{\beta}_1$, Bias and Absolute Bias of OLS Estimator for Different Strengths(g) of Heteroscedasticity and Varying Samples Sizes(n)

Sample size (n)	Strength (g)	$\hat{\beta}_0$	$\hat{\beta}_1$	Bias($\hat{\beta}_0$)	Bias($\hat{\beta}_1$)	Absolute Bias($\hat{\beta}_0$)	Absolute Bias($\hat{\beta}_1$)
25	-2	2.006492	2.99893	0.006492	-0.001070	0.006492	0.001070
	-1	2.003437	2.999502	-0.003437	-0.000498	0.003437	0.000498
	-½	2.011658	2.997627	-0.011658	0.003730	0.011658	0.003730
	0	2.00000	3.00000	0.000000	0.000000	0.000000	0.000000
	½	2.003248	2.999622	-0.003248	0.000378	0.003248	0.000378
	1	2.020864	2.996797	-0.020864	0.003203	0.020864	0.003203
	2	1.994345	3.001383	0.005655	-0.001383	0.005655	0.001383
30	-2	2.014509	2.996844	0.014509	0.003156	0.014509	0.003156
	-1	1.996752	3.00061	0.003248	-0.000610	0.003248	0.000610
	-½	2.002684	2.999539	-0.002684	0.000461	0.002684	0.000461
	0	2.00000	3.00000	0.000000	0.000000	0.000000	0.000000
	½	2.00394	2.999241	-0.00394	0.000759	0.00394	0.000759
	1	1.978417	3.003813	0.021583	-0.003813	0.021583	0.003813
	2	2.00686	2.998429	-0.006860	0.001571	0.006860	0.001571
35	-2	2.019945	2.999136	0.019945	0.000864	0.019945	0.000864
	-1	1.984136	3.003443	0.015864	-0.003443	0.015864	0.003443
	-½	2.002723	2.999497	-0.002723	0.000503	0.002723	0.000503
	0	2.00000	3.00000	0.000000	0.000000	0.000000	0.000000
	½	1.998292	3.00281	0.001708	-0.002810	0.001708	0.002810
	1	1.993081	3.001168	0.006919	-0.001168	0.006919	0.001168
	2	1.991589	3.002033	0.008411	-0.002033	0.008411	0.002033
40	-2	2.005019	2.999136	0.005019	0.000864	0.005019	0.000864
	-1	2.00447	2.999065	-0.00447	0.000935	0.004470	0.000935
	-½	1.99870	3.000302	0.001300	-0.000302	0.001300	0.000302
	0	2.00000	3.00000	0.000000	0.000000	0.000000	0.000000
	½	1.996941	3.000495	0.003059	-0.000495	0.003059	0.000495
	1	1.991806	3.001398	0.008194	-0.001398	0.008194	0.001398
	2	1.989425	3.003047	0.010575	-0.003047	0.010575	0.003047
45	-2	1.998827	3.000183	-0.001173	-0.000183	0.001173	0.000183
	-1	1.984091	3.003301	0.015909	-0.003301	0.015909	0.003301
	-½	2.000513	2.999901	-0.000513	0.000099	0.000513	0.000099
	0	2.00000	3.00000	0.000000	0.000000	0.000000	0.000000
	½	1.990715	3.001506	0.009285	-0.001506	0.009285	0.001506
	1	1.986978	3.002143	0.013022	-0.002143	0.013022	0.002143
	2	2.001857	2.999581	-0.001857	0.000419	0.001857	0.000419
50	-2	1.992516	3.001224	-0.007484	-0.001224	0.007484	0.001224
	-1	2.010046	2.998024	-0.010046	0.001976	0.010046	0.001976
	-½	1.98750	3.002355	0.012500	-0.002355	0.012500	0.002355
	0	2.00000	3.00000	0.000000	0.000000	0.000000	0.000000
	½	1.998667	3.000217	0.001333	-0.000217	0.001333	0.000217
	1	1.998421	3.000265	0.001579	-0.000265	0.001579	0.000265
	2	2.061664	2.985942	-0.061664	0.014058	0.061664	0.014058

Table 2: Shows the estimates of $s.e(\hat{\beta}_1)$ for HCSE estimators at sample size (n) =25 and varying strengths of heteroscedasticity (g).

		When Sample size(n)=25					
	Estimator	HC0	HC1	HC2	HC3	HC4	HC5
Strength of heteroscedasticity(g)	$\hat{\beta}_1$	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)
-2	2.9697	0.0880	0.0918	0.0983	0.1100	0.1191	0.07095
-1	2.9805	0.1207	0.1258	0.1299	0.1398	0.1378	0.1122
-1/2	3.0600	0.1900	0.1981	0.2106	0.2339	0.2492	0.1806
0	3.0000	0.1914	0.1996	0.2012	0.2119	0.2081	0.1914
1/2	3.0342	0.1727	0.1800	0.1914	0.2145	0.2627	0.1816
1	3.1301	0.3179	0.3315	0.3865	0.4766	0.7321	0.3865
2	3.0469	0.3850	0.4014	0.4044	0.4250	0.4100	0.4250

From Table 2, when $g=-2$, -1 , and $-1/2$ at sample size $n=25$. HC5 gives the minimum estimates of $s.e(\hat{\beta}_1)$, While HC0 and HC5 offered the same estimate of $s.e(\hat{\beta}_1)$ at $g=0$. When $g=1/2$ HC0 gives the minimum $s.e(\hat{\beta}_1)$ followed by HC1, HC5, HC3, and HC4 in that order. When values of $g= -2, -1, -1/2, 0$, and $1/2$, $s.e(\hat{\beta}_1)$ are relatively small compared to when $g=1$ and 2 at sample size $n=25$. When $g=1$, HC2 and HC5 offered the same estimate of $s.e(\hat{\beta}_1)$. When $g=2$, HC3 and HC45 offered the same estimate of $s.e(\hat{\beta}_1)$.

Table 3: Shows the estimates of $s.e(\hat{\beta}_1)$ for HCSE estimators at sample size (n) =30 and varying strengths of heteroscedasticity (g).

		When Sample size(n)=30					
	Estimator	HC0	HC1	HC2	HC3	HC4	HC5
Strength of heteroscedasticity(g)	$\hat{\beta}_1$	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)
-2	2.9280	0.0835	0.0864	0.0885	0.0939	0.0924	0.0742
-1	3.0119	0.0659	0.0682	0.0694	0.0733	0.0729	0.0626
-1/2	2.9964	0.1254	0.1298	0.1327	0.1406	0.1402	0.1219
0	3.0000	0.2246	0.2325	0.2380	0.2526	0.2520	0.2246
1/2	3.0489	0.3630	0.3757	0.3823	0.4027	0.3934	0.3725
1	2.9444	0.1734	0.1795	0.1829	0.1931	0.1902	0.1829
2	3.1386	0.4378	0.4532	0.4722	0.5098	0.5222	0.5098

From Table 3, when $g=-2$, -1 , and $-1/2$ at sample size $n=30$. HC5 gives the minimum estimates of $s.e(\hat{\beta}_1)$, While HC0 and HC5 offered the same estimate of $s.e(\hat{\beta}_1)$ at $g=0$. When $g=1/2$ HC0 gives the minimum $s.e(\hat{\beta}_1)$ followed by HC5, HC1, HC4, and HC3 in that order. When $g=1$, HC2 and HC5 offered the same estimate of $s.e(\hat{\beta}_1)$. When $g=2$, HC3 and HC45 offered the same estimate of $s.e(\hat{\beta}_1)$.

Table 4: Shows the estimates of $s.e(\hat{\beta}_1)$ for HCSE estimators at sample size (n) =35 and varying strengths of heteroscedasticity(g).

		When sample size(n)=35					
	Estimator	HC0	HC1	HC2	HC3	HC4	HC5
Strength of heteroscedasticity(g)	$\hat{\beta}_1$	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)
	2.7961	0.1942	0.2000	0.2191	0.2475	0.3141	0.1533
-1	2.9191	0.2643	0.2722	0.2846	0.3066	0.3191	0.2456
-½	2.9857	0.0912	0.0939	0.0946	0.0980	0.0956	0.0896
0	3.0000	0.2077	0.2139	0.2158	0.2243	0.2194	0.2077
½	2.9764	0.1062	0.1093	0.1114	0.1170	0.1174	0.1087
1	3.0103	0.1080	0.1112	0.1121	0.1166	0.1148	0.1121
2	3.1334	0.2514	0.2589	0.2633	0.2760	0.2742	0.2760

From Table 4, when g=-2, -1, and -½ at sample size n=35. HC5 gives the minimum estimates of $s.e(\hat{\beta}_1)$, While HC0 and HC5 offered the same estimate of $s.e(\hat{\beta}_1)$ at g=0. When g=½ HC0 gives the minimum $s.e(\hat{\beta}_1)$ followed by HC5, HC1, HC3, and HC4 in that order. When g=1, HC2 and HC5 offered the same estimate of $s.e(\hat{\beta}_1)$. When g=2, HC3 and HC45 offered the same estimate of $s.e(\hat{\beta}_1)$.

Tab. 5: Shows the estimates of $s.e(\hat{\beta}_1)$ for HCSE estimators at sample size (n) =40 and varying strengths of heteroscedasticity(g).

		When sample size(n)=40					
	Estimator	HC0	HC1	HC2	HC3	HC4	HC5
Strength of heteroscedasticity(g)	$\hat{\beta}_1$	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)
-2	3.0146	0.0473	0.0486	0.0497	0.0522	0.0522	0.0430
-1	3.0180	0.0594	0.0610	0.0619	0.0644	0.0639	0.0571
-½	3.0126	0.1647	0.1690	0.1734	0.1827	0.1840	0.1606
0	3.0000	0.2173	0.2230	0.2301	0.2438	0.2505	0.2173
½	3.0138	0.0866	0.0889	0.0898	0.0932	0.0934	0.0882
1	3.0055	0.2008	0.2060	0.2103	0.2206	0.2230	0.2103
2	3.1458	0.3926	0.4028	0.4111	0.4308	0.4303	0.4308

From Table 5, when g=-2, -1, and -½ at sample size n=40. HC5 gives the minimum estimates of $s.e(\hat{\beta}_1)$, While HC0 and HC5 offered the same estimate of $s.e(\hat{\beta}_1)$ at g=0. When g=½ HC0 gives the minimum $s.e(\hat{\beta}_1)$ followed by HC5, HC1, HC4, and HC3 in that order. When g=1, HC2 and HC5 offered the same estimate of $s.e(\hat{\beta}_1)$. When g=2, HC3 and HC45 offered the same estimate of $s.e(\hat{\beta}_1)$.

Table 6: Shows the estimates of $s.e(\beta_1)$ for HCSE estimators at sample size (n) =45 and varying strengths of heteroscedasticity (g).

		When Sample size(n)=45					
	Estimator	HC0	HC1	HC2		HC4	HC5
Strength of heteroscedasticity (g)	$\hat{\beta}_1$	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)
-2	2.9856	0.0361	0.0369	0.0377	0.0395	0.0399	0.0331
-1	3.2079	0.2219	0.2270	0.2351	0.2494	0.2602	0.2096
$-\frac{1}{2}$	3.0210	0.1049	0.1073	0.1092	0.1138	0.1135	0.1028
0	3.0000	0.1623	0.1660	0.1680		0.1766	0.1623
$\frac{1}{2}$	2.9873	0.1116	0.1141	0.1174	0.1236		0.1144
1	2.9911	0.1673	0.1711	0.1748	0.1831	0.1942	0.1748
2	3.0108	0.3479	0.3559	0.3622	0.3773	0.3789	0.3773

From Table 6, when $g=-2$, -1 , and $-\frac{1}{2}$ at sample size $n=45$. HC5 gives the minimum estimates of $s.e(\beta_1)$, While HC0 and HC5 offered the same estimate of $s.e(\beta_1)$ at $g=0$. When $g=\frac{1}{2}$ HC0 gives the minimum $s.e(\beta_1)$ followed by HC1, HC5, HC2, HC3 and HC4 in that order. When $g=1$, HC2 and HC5 offered the same estimate of $s.e(\beta_1)$. When $g=2$, HC3 and HC45 offered the same estimate of $s.e(\beta_1)$.

Tab.7: Shows the estimates of $s.e(\beta_1)$ for HCSE estimators at sample size (n) =50 and varying strengths of heteroscedasticity(g)

		When Sample size(n)=50					
	Estimator	HC0	HC1	HC2	HC3	HC4	HC5
Strength of heteroscedasticity(g)	$\hat{\beta}_1$		S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)	S.E($\hat{\beta}_1$)
-2	2.9906	0.0318	0.0324	0.0327		0.0335	0.0299
-1	3.0505			0.1392	0.1455	0.1473	0.1276
$-\frac{1}{2}$	2.9213	0.1278	0.1305	0.1361	0.1452	0.1618	0.1239
0	3.0000	0.1503	0.1534	0.1599	0.1706	0.1893	
$\frac{1}{2}$	3.0065	0.0986	0.0950	0.0963	0.0997	0.0996	0.0947
1	2.9667	0.1551	0.1583	0.1598	0.1648	0.1635	0.1598
2	2.9341	1.0445	1.0661	1.1007	1.1617	1.2447	1.1617

From Table 7, when $g=-2$, -1 , and $-\frac{1}{2}$ at sample size $n=45$. HC5 gives the minimum estimates of $s.e(\beta_1)$, While HC0 and HC5 offered the same estimate of $s.e(\beta_1)$ at $g=0$. When $g=\frac{1}{2}$ HC5 gives the minimum $s.e(\beta_1)$ followed by HC1, HC2, HC0, HC4 and HC3 in that order. When $g=1$, HC2 and HC5 offered the same estimate of $s.e(\beta_1)$. When $g=2$, HC3 and HC45 offered the same estimate of $s.e(\beta_1)$.

4. Conclusion

Generally, the OLS estimators of model parameters remain unbiased when bivariate data set is fraught with heteroscedasticity problem. Also for SLRM, the estimated $se(\beta_1)$ for different strengths of heteroscedasticity(g) and varying sample sizes using the HCSE estimators are strictly less than the $se(\beta_0)$. The coefficient of determination (R^2) is usually unaffected by heteroscedasticity and for pedagogical purposes the observed R^2 ranges between 0.6032 to 0.9954. In addition the graphs of $se(\beta_j)$ showed consistent patterns in the estimates offered by the estimators (See Appendices A3 and A4). It is worthwhile to note that the following remarks are very pertinent from the study:

For negative strengths of heteroscedasticity HC5 is preferred as it has the minimum variances at $g=-2, -1, -0.5$ and 0 respectively.

At $g=0$, HC5 performs equally as HC0 (homoscedasticity assumption holds).

At $g=1$, HC5 performs equally as HC2.

At $g=2$, HC5 performs equally as HC3.

At $g=0.5$, HC5 performs almost equally as HC0 and HC1.

Furthermore, It is expected that HC5 will offer similar results to HC4 when $g=4$.

HC5 can be obtained for any strength “ g ” of heteroscedasticity that may be present in Simple Linear Regression model. HC5 is applicable to SLRM that is plagued with heteroscedastic error variance pinned down to the predictor the model. It may also be extended to multiple linear regression models having similar problem of heteroscedasticity. Finally HC5 is indeed a generalization of the existing family of HCSE estimators.

Acknowledgement

The authors acknowledge the comments of the reviewers for their suggestions and comments in improving the quality of the manuscripts.

References

- Ayinde, K. (2007): Equations to generate normal variates with desired inter correlations matrix. *Int. J. Stat. and syst.* **2**(2), 99-111.
- Behnken, D. W. and Draper, N. R. (1972): Residuals and their variance patterns. *Technometrics*. **14**, 102-111.
- Berry, W.D. and Feldman, S. (1985): *Multiple regressions in Practice*. Sage University paper Series on Quantitative Applications in the Social Science, Series no.07-050. Newbury Park, CA: Sage.
- Box, G. E. P. and Cox, D. R. (1964): An Analysis of transformations (with discussion). *J. R. Statist. Soc. ser. B*. **26**, 211-246.
- Breusch, T. S. and Pagan, A. R. (1979): A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*. **47**, 1287-1294.
- Cook, R. D. and Weisberg, S. (1983): Diagnostics for heteroskedasticity in regression. *Biometrika*. **70**, 1-10.
- Cook, R. D. and Weisberg, S. (1982): *Residuals and influence in Regression*. Chapman and Hall, New York, 123-153.
- Cook, R.D. (1977): Detection of Influential Observations in Linear Regression. *Technometrics*. **19** (1), 15-18.
- Cribari-Neto, F. (2004): Asymptotic inference under heteroskedasticity of unknown form. *Comp. Stat. & Data Analysis*. **45**, 215-233.
- David, R. and Mackinnon, J. G. (1993): *Estimation and Inference in Econometrics*. Oxford University Press, New York, 549-550.
- Davies, R. B. and Hutton, B. (1975): The Effects of errors in Independent Variables in Linear Regression, *Biometrika*. **62**, 383-391.
- Draper, N. R. and Smith, H. (1981): *Applied regression analysis*. (2nd ed.). Wiley & Sons, Inc., New York, 709.
- Eicker, F. (1967): Limit theorems for regression with unequal and dependent errors. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. **1**(1), 59-82. Eds.: L. M. Le Cam & J. Neyman. Berkeley, CA: University of California Press.
- Geary, R. C. (1966): A note on residuals heterocovariance and estimation of efficiency in regression. *American Statistician*. **20**, 30-31.
- Glejser, H. (1969): A new test of heterosecdasticity. *J. Am. Stat. Ass.* **64**, 314-323.

- Godfrey, C. (1978): Testing for Multiplicative Heteroscedasticity. *J. Econometrics*. **8**, 227-236.
- Goldfeld, S. M. and Quandt, R. E. (1965): Some tests for heteroskedasticity. *J. Am. Stat. Ass.* **60**, 539-547.
- Gujarati, D. N., (2004): *Basic Econometrics*. fourth edition. McGraw-Hill, New York, 387-427.
- Hinkley, D. V. (1997): Jackknifing in unbalanced situations. *Technometrics*. **19** (3), 285-292.
- Holt, D., Smith, M. F. and Winter, P. D. (1980): Analysis of Data from complex surveys. *J. Royal Stat. Soc. A*. **143** (4), 474-487.
- Huber, P. J. (1967): The behavior of maximum likelihood estimation under nonstandard conditions. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. **1** (Statistics), 221-233. Eds.: L. M. Le Cam & J. Neyman. Berkeley, CA: University of California Press.
- Huber, P. J. (1975): *Robustness and Design. A survey of Statistical Design and Linear Models*. North-Holland, Amsterdam. 287-303.
- Johnston, J. (1984): *Econometrics methods*. Second edition. McGraw-Hill Book Company, New York, 201-215.
- Karona, G. et al. (1954): *Contributions of Survey Methods to Econometrics*. Columbia: University Press, 203.
- Lancaster, T. (1968): Grouping Estimators on Heteroscedastic Data. *J. Am. Statist. Ass.* **63**, 191-198.
- Long, J. S. and Ervin, L. H. (2000): Using heteroskedasticity consistent standard errors in the linear regression model. *American Statistician* **54**, 217-224.
- Mackinnon, J. G. and White, G. P. (1985): Some heteroscedasticity Consistent Covariance Matrix Estimator with Improved finite sample properties. *J. Econometrics*. **29**, 305-325.
- Maddala, G. S. (2001): *Introduction to Econometrics*. Third edition. John Wiley, England. 199-216.
- Malinvaud, E. (1970): The Consistency of non-linear regression. *Annals of Mathematical Statistics*, **41**, 956-969.
- Park, R. E. (1966): Estimation with Heteroscedastic Error Term. *Econometrica*. **34** (4), 888.

The Park Test is a special case of the general test proposed by A. C. Harray (1976) in Estimating Regression models with Multiplicative Heteroscedasticity. *Econometrica*. **44** (3), 461-465.

Oduntan, E. A. (2004): *A Monte Carlo Study of the Problem of Multicollinearity in Simultaneous Equation Model*. (Ph.D. Thesis, Department of Statistics, University of Ibadan).

Olubusoye, O. E. (2001): *Consequences of the violation of the Assumption of zero correlation between pairs of Random Stochastic Terms used in Monte Carlo Experiment*. (Ph.D. Thesis, Department of Statistics, University of Ibadan).

White, H. (1980): A heteroskedasticity-consistent covariance estimator and a direct test for heteroskedasticity. *Econometrica*. **48**, 817-838.

APPENDIX “A1”

Table 8: Showing $Var(\hat{\beta}_0)$ for different strengths of heteroscedasticity and sample sizes using the existing HCSE and HC5 estimators.

Sample size(n)	HCSE Estimator	Strength of Heteroscedasticity(g)						
		-2	-1	-½	0	½	1	2
25	HC0	0.5563211	1.035072	0.918468	0.821779	0.591899	2.02742	0.985563
	HC1	0.5800049	1.079137	0.957569	0.856764	0.617097	2.113732	1.027521
	HC2	0.6216265	1.113811	1.009143	0.862423	0.646476	2.435896	1.031884
	HC3	0.695725	1.19929	1.112407	0.907213	0.713384	2.969556	1.081199
	HC4	0.7528561	1.18237	1.178001	0.897411	0.840795	4.437786	1.046338
	HC5	0.4478809	0.962537	0.877364	0.821779	0.617813	2.435896	1.081199
30	HC0	0.3980349	0.399925	0.808479	1.117747	1.592767	0.765355	1.482506
	HC1	0.412049	0.413962	0.836856	1.156978	1.64867	0.792218	1.534539
	HC2	0.4218287	0.418871	0.853973	1.177594	1.67193	0.810058	1.590764
	HC3	0.4471221	0.439054	0.903011	1.242048	1.755833	0.85822	1.709419
	HC4	0.4391035	0.431005	0.898114	1.231055	1.715428	0.85314	1.744768
	HC5	0.3545886	0.382114	0.786951	1.117747	1.631774	0.810058	1.709419
35	HC0	1.0312684	1.240713	0.572673	1.014122	0.518898	0.688361	0.890327
	HC1	1.0620594	1.277758	0.589771	1.044401	0.534391	0.708913	0.916909
	HC2	1.162775	1.331594	0.592426	1.047607	0.541842	0.709732	0.923821
	HC3	1.3127832	1.429958	0.612981	1.082529	0.566577	0.732105	0.959496
	HC4	1.6643906	1.477448	0.597427	1.056739	0.566996	0.716518	0.948983
	HC5	0.815022	1.156745	0.563086	1.014122	0.530156	0.709732	0.959496
40	HC0	0.2968187	0.316718	0.829541	1.141438	0.463849	1.083649	1.072228
	HC1	0.3045295	0.324946	0.851091	1.171091	0.475899	1.1118	1.100082
	HC2	0.311022	0.328728	0.873799	1.20906	0.482102	1.132815	1.117641
	HC3	0.3260455	0.341388	0.9207586	1.281772	0.501762	1.185506	1.165778
	HC4	0.3255925	0.338015	0.9267027	1.31811	0.508691	1.196716	1.162757
	HC5	0.270693	0.30532	0.8083742	1.141438	0.47281	1.132815	1.165778
45	HC0	0.247479	1.099726	0.570311	0.641874	0.60727	0.891327	1.266627
	HC1	0.2531689	1.12501	0.583424	0.656632	0.621232	0.91182	1.295749
	HC2	0.2582951	1.160713	0.592063	0.661466	0.638232	0.930093	1.317767
	HC3	0.2697435	1.225834	0.614939	0.682256	0.671297	0.972726	1.371859
	HC4	0.2716967	1.261305	0.612799	0.683161	0.688452	1.02579	1.377201
	HC5	0.2275922	1.042554	0.559839	0.641874	0.622497	0.930093	1.371859
50	HC0	0.2139438	0.692306	0.629736	0.762294	0.5807763	0.946478	3.9279155
	HC1	0.2183555	0.706582	0.642722	0.778013	0.5927524	0.965995	4.0089126
	HC2	0.2203181	0.721148	0.668647	0.803041	0.601495	0.977394	4.1334864
	HC3	0.2269386	0.751454	0.711214	0.847922	0.6232285	1.009593	4.3570139
	HC4	0.2245636	0.757036	0.786666	0.914542	0.6235912	1.002021	4.6650113
	HC5	0.2018907	0.664854	0.611554	0.762294	0.5910121	0.977394	4.3570139

APPENDIX “A2”

Table 9: Showing $Var(\hat{\beta}_1)$ for different strengths of heteroscedasticity and sample sizes using the existing HCSE and HC5 estimators.

Sample size(n)	HCSE Estimator	Strength of Heteroscedasticity(g)						
		-2	-1	$-\frac{1}{2}$	0	$\frac{1}{2}$	1	2
25	HC0	0.0880422	0.120685	0.189992	0.191438	0.172678	0.31792	0.384985
	HC1	0.0917903	0.125822	0.19808	0.199587	0.180029	0.331455	0.401375
	HC2	0.0983446	1	0.12988	0.210592	0.191432	0.386526	0.404424
	HC3	0.1100425	0.139857	0.233942	0.211869	0.214537	0.476623	0.425018
	HC4	0.1191319	0.137824	0.249211	0.208093	0.262712	0.732112	0.410023
	HC5	0.0709538	0.112207	0.180621	0.191438	0.181566	0.386526	0.425018
30	HC0	0.0834623	0.065911	0.125399	0.224585	0.363001	0.173395	0.437835
	HC1	0.0863917	0.068224	0.1298	0.232467	0.375742	0.179481	0.453202
	HC2	0.0885393	0.069444	0.132729	0.238036	0.382309	0.182921	0.472175
	HC3	0.0939435	0.073256	0.140637	0.252571	0.402745	0.193098	0.509755
	HC4	0.0924100	0.072852	0.140196	0.251993	0.393385	0.190222	0.522233
	HC5	0.0742088	0.062628	0.121932	0.224585	0.372518	0.182921	0.509755
35	HC0	0.1942126	0.264326	0.091224	0.207746	0.106169	0.107974	0.251405
	HC1	0.2000113	0.272218	0.093948	0.213949	0.109339	0.111198	0.258911
	HC2	0.2190869	0.284602	0.094553	0.215829	0.111408	0.112147	0.263312
	HC3	0.2474589	0.306582	0.098021	0.224297	0.117046	0.116559	0.276004
	HC4	0.3140991	0.319054	0.095563	0.219428	0.11735	0.114756	0.274242
	HC5	0.1533048	0.245616	0.08961	0.207746	0.108741	0.112147	0.276004
40	HC0	0.0473215	0.059425	0.164721	0.217342	0.086603	0.200768	0.392581
	HC1	0.0485509	0.060968	0.169000	0.222988	0.088852	0.205983	0.40278
	HC2	0.0496761	0.061852	0.173440	0.230111	0.089816	0.210338	0.411137
	HC3	0.0521664	0.064413	0.182698	0.243835	0.093236	0.220584	0.430764
	HC4	0.0521792	0.063909	0.183976	0.250514	0.093372	0.22299	0.430323
	HC5	0.0429512	0.057123	0.160552	0.217342	0.088185	0.210338	0.430764
45	HC0	0.0361061	0.221871	0.104899	0.162276	0.111582	0.167262	0.347905
	HC1	0.0369362	0.226972	0.107311	0.166007	0.114148	0.171107	0.355904
	HC2	0.0377308	0.235105	0.10923	0.167996	0.117381	0.1747797	0.362205
	HC3	0.0394531	0.24935	0.113778	0.17413	0.123582	0.183069	0.377324
	HC4	0.0398531	0.260214	0.113486	0.176567	0.127133	0.194167	0.378854
	HC5	0.0331252	0.209554	0.102812	0.162276	0.114433	0.1747787	0.377324
50	HC0	0.0317609	0.13325	0.127821	0.150327	0.09856	0.155084	1.044531
	HC1	0.0324158	0.135998	0.130457	0.153427	0.0949714	0.158282	1.066069
	HC2	0.0327446	0.139193	0.136137	0.159928	0.0962898	0.159837	1.100712
	HC3	0.0337681	0.14545	0.145227	0.170556	0.0996829	0.16479	1.161691
	HC4	0.0334787	0.147303	0.161776	0.189279	0.0996062	0.163468	1.244681
	HC5	0.0299055	0.127604	0.123933	0.150327	0.0946522	0.159837	1.161691

APPENDIX "A3"

Graphs of Standard Error of $Var(\hat{\beta}_0)$ Against Strength of Heteroscedasticity Figures 1.0-1.5

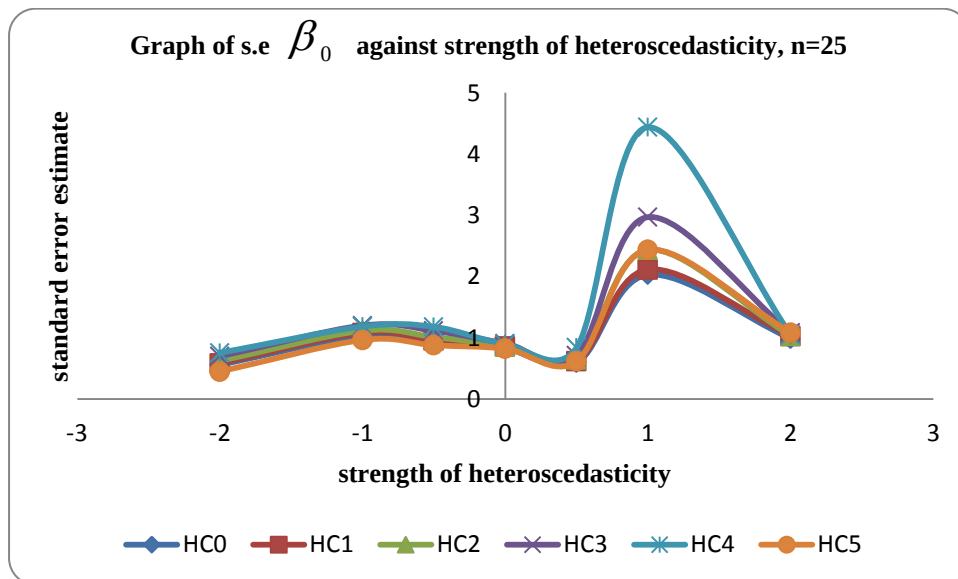


Figure 1.0

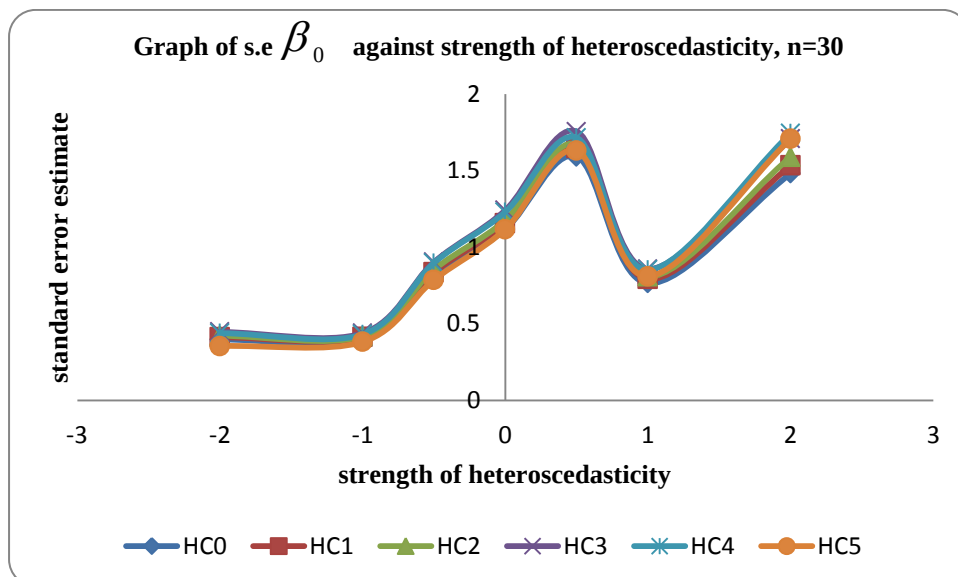


Figure1.1

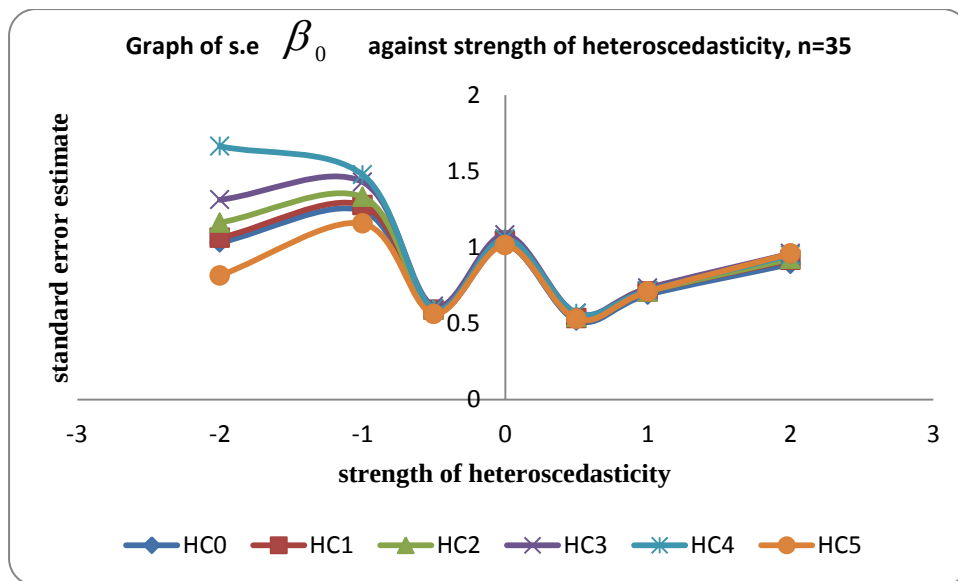


Figure1.2

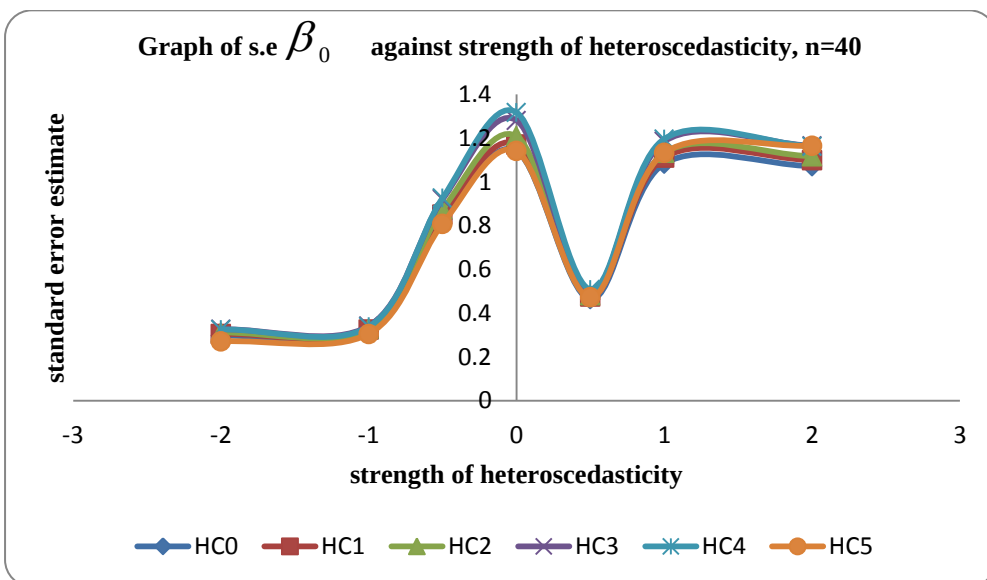


Figure 1.3

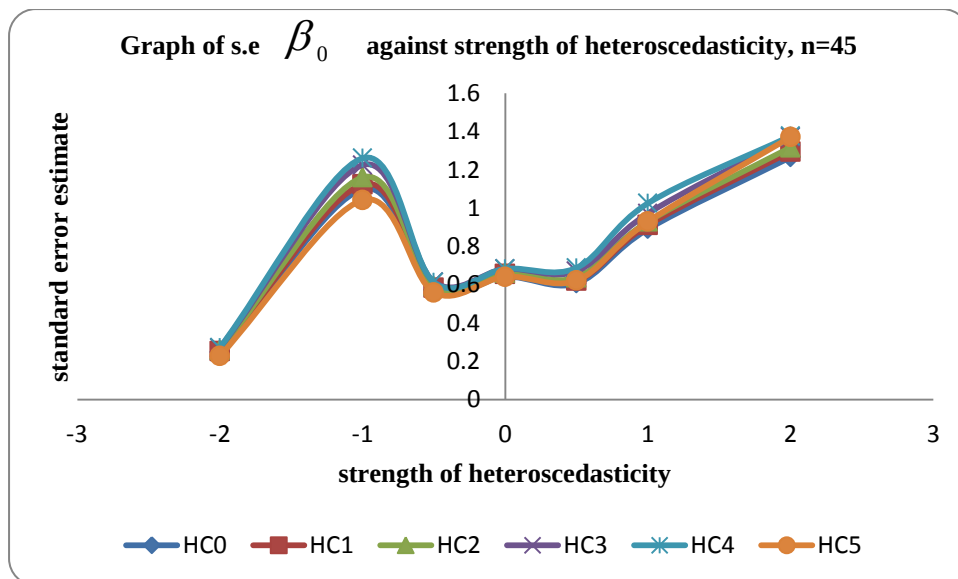


Figure 1.4

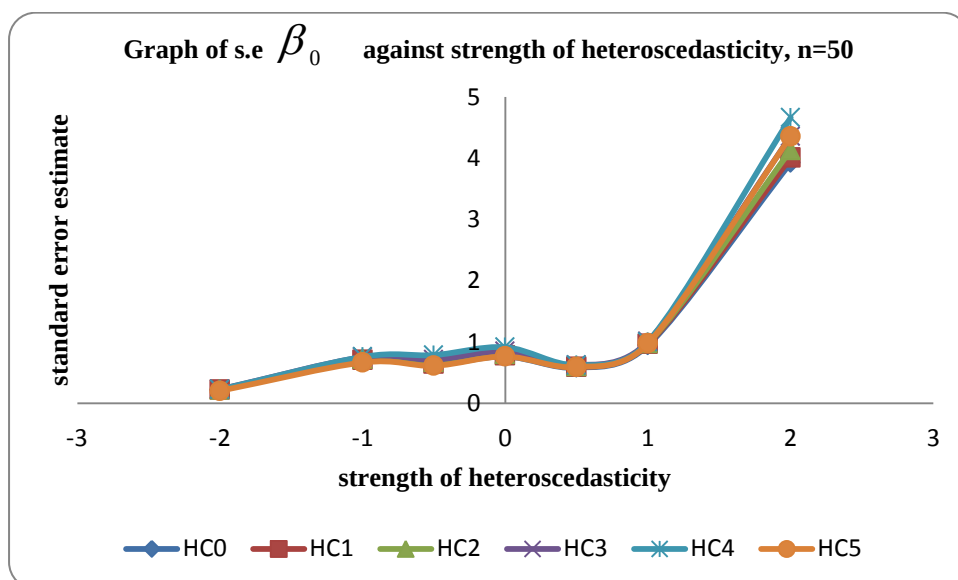


Figure 1.5

APPENDIX "A4"

Graphs of Standard Error of $Var(\hat{\beta}_1)$ Against Strength of Heteroscedasticity Figures 2.0-2.5

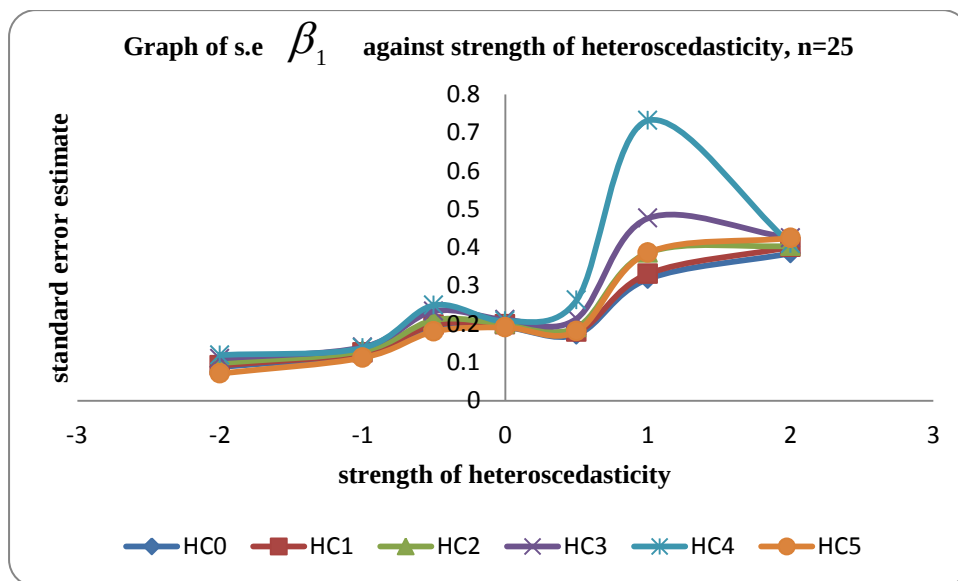


Figure 2.0

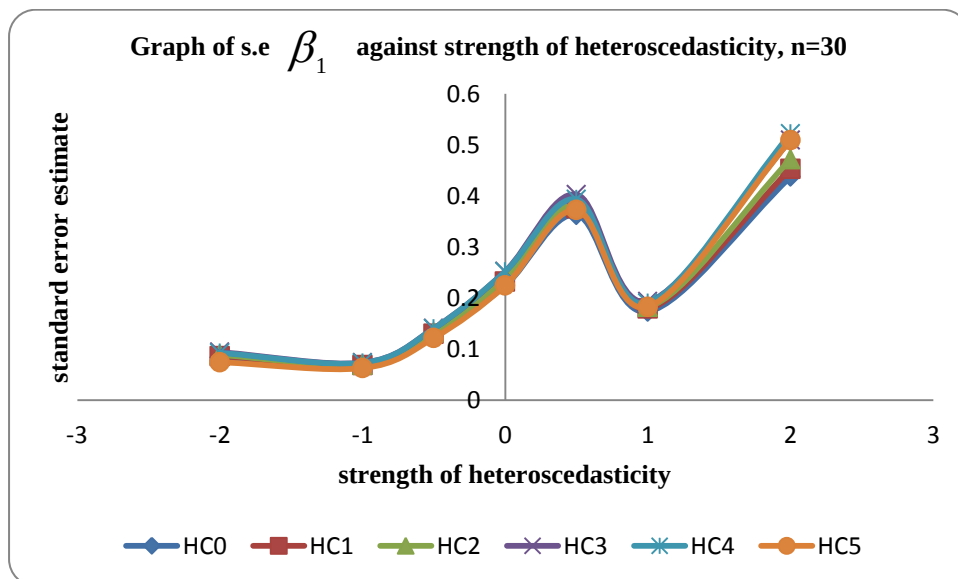


Figure 2.1

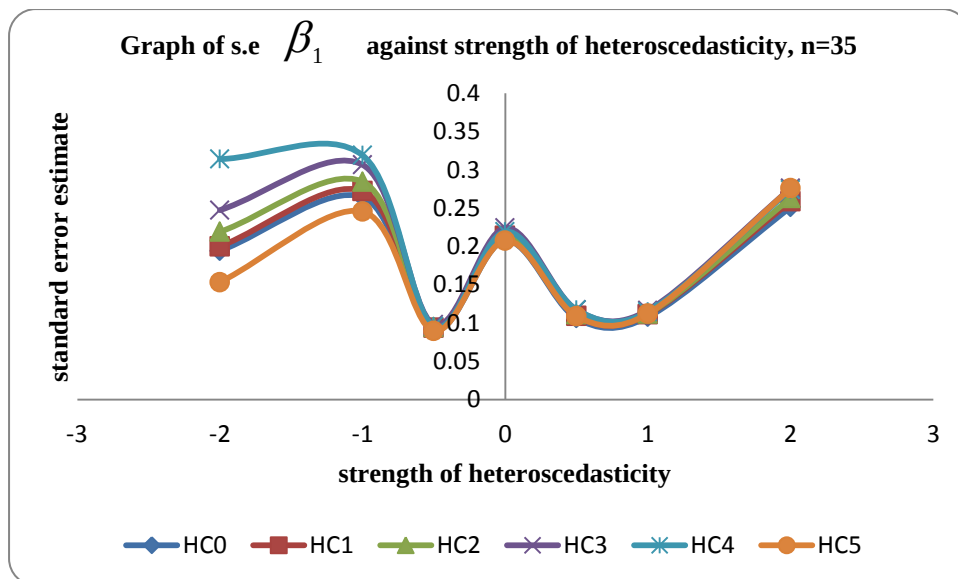


Figure 2.2

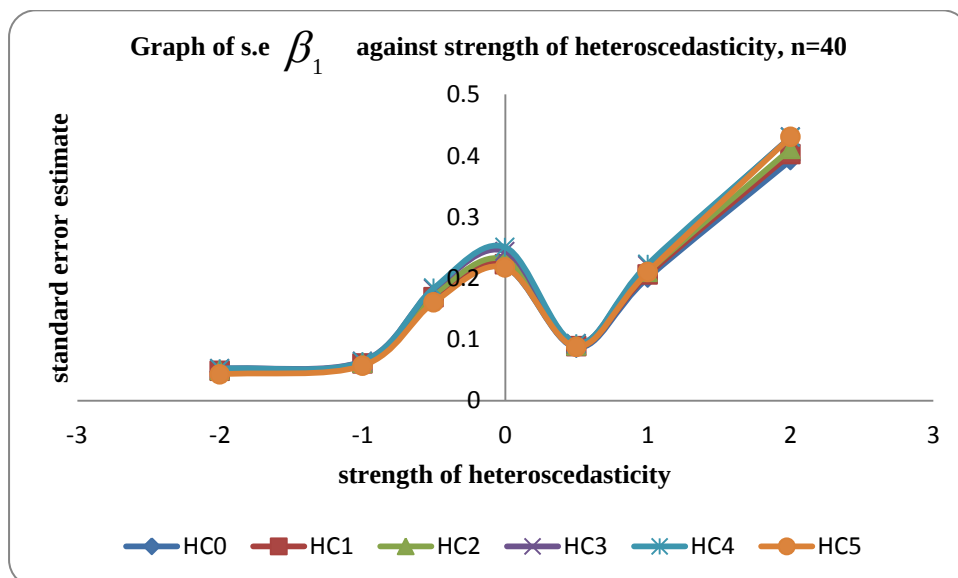


Figure 2.3

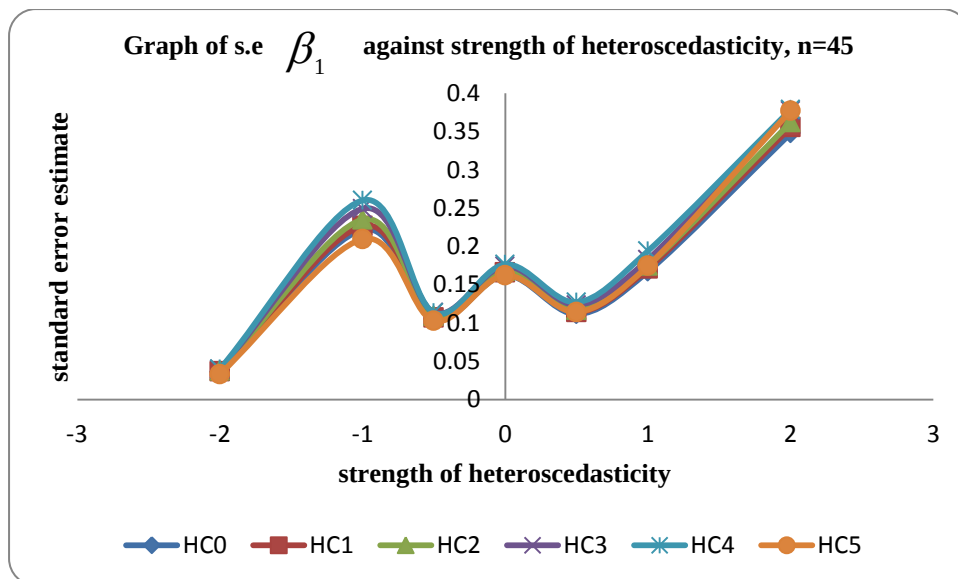


Figure 2.4

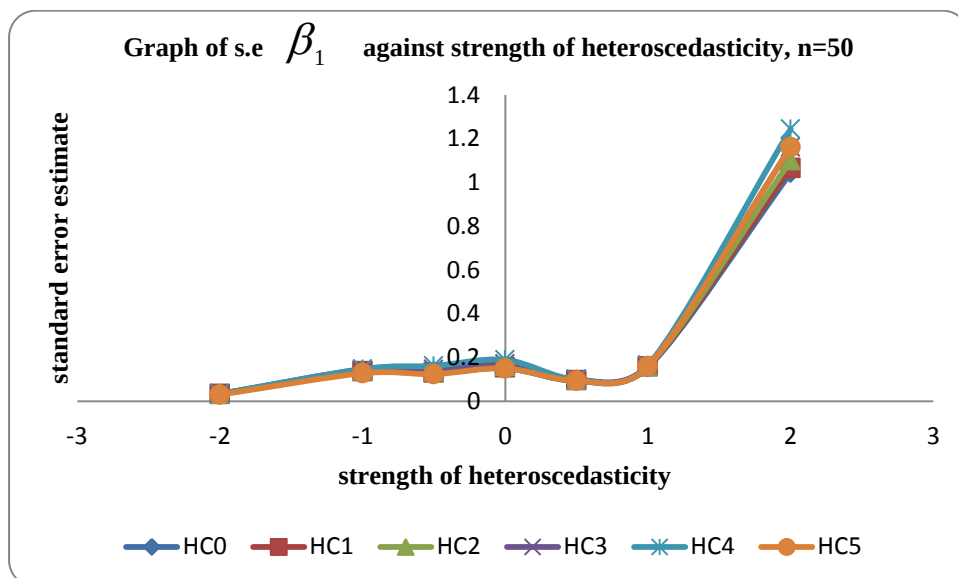


Figure 2.5