On Inference of the Linear Regression Model with Groupwise Heteroscedasticity

Abdul Majid
Pakistan Bureau of Statistics, Pakistan
amajid_1980@yahoo.com

Muhammad Aslam
Department of Statistics, Bahauddin Zakariya University, Pakistan
aslamasadi@bzu.edu.pk

Muhammad Amanullah
Department of Statistics, Bahauddin Zakariya University, Pakistan
aman_stat@yahoo.com

Saima Altaf
Department of Statistics, Bahauddin Zakariya University, Pakistan
drsaimaaltaf27@gmail.com

Abstract

The performance of heteroscedasticity consistent covariance matrix estimators (HCCMEs), namely, HC0, HC1, HC2, HC3 and HC4 have been evaluated by numerous researchers for the heteroscedastic linear regression models. This study focuses on examining the performance of these covariance estimators in case of groupwise heteroscedasticity. With the help of the Monte Carlo simulations, we evaluate the performance of these covariance estimators and the associated quasi-*t* tests. We consider the cases when data are divided into 10, 20 and 30 groups of different sizes and the regression is run on the mean values of the dependent variable and the regressor of these groups. The numerical results show that HCCMEs perform appealingly well in case of groupwise heteroscedasticity.

Keywords: Groupwise heteroscedasticity; HCCME; Size distortion; White's estimator.

1. Introduction

The linear regression models for cross-sectional data often exhibit the problem of heteroscedasticity i.e. the error variances are not constant for all observations. In this situation, the ordinary least square (OLS) estimators of the parameters remain unbiased and consistent but become inefficient. Since the form of heteroscedasticity is usually unknown, the practitioners usually use the OLS estimators even when the error variances are not constant. However, the usual OLS covariance matrix estimator becomes biased and does not remain consistent when the homoscedasticity assumption is violated. Since the OLS standard errors are based directly on these variances, so the inferences drawn on the basis of these estimators become misleading and erroneous. It thus, becomes necessary to build and use alternative covariance matrix estimators that are consistent under both homoscedasticity and heteroscedasticity of unknown form. Several consistent covariance matrix estimators are available in the literature.

In his econometrics text book, Wooldridge (2008, pp. 249) writes, "In the last two decades, econometricians have learned to adjust standard errors, *t, F* and *LM* statistics so

that they are valid in the presence of heteroscedasticity of unknown form. This is very convenient because it means we can report new statistics that work, regardless of the kind of heteroscedasticity present in the population".

The most commonly used heteroscedasticity consistent covariance matrix estimator (HCCME) was presented by White (1980). White's estimator is known as HC0 in the literature. MacKinnon and White (1985) and Davidson and MacKinnon (1993) presented three alternative improved versions of HCCMEs, commonly known as HC1, HC2 and HC3. Numerous researchers have evaluated the performance of HCCMEs for the heteroscedastic linear regression models (e.g., see Ahmed et al. 2011, Aslam et al. 2013, Cribari-Neto 2004, Cribari-Neto and Zarkos 1999, 2001, 2004, Flachaire 2005, Long and Ervin 2000, MacKinnon 2011 etc.).

In real life, there are many situations when the dependent and explanatory variables have to be aggregated as average values i.e. the means of dependent and explanatory variables are available and we have to run regression on these average values. As a result of such aggregation, the error tem in the resulting model becomes heteroscedastic, in spite of the fact that it is homoscedastic in the model of individual values. It is because the number of observations (say) n_g varies across the different groups. Such type of heteroscedasticity is called as groupwise heteroscedasticity (see Greene 2000, for more details). For groupwise heteroscedastic regression model, obviously, the usual OLS covariance matrix of estimates will be inconsistent. Thus, it becomes needful to use the HCCMEs. Now the question arises that what would be the performance of the HCCME in case of groupwise heteroscedasticity but the available literature is found to be silent in this situation. This thing motivated us to evaluate the performance of HCCMEs in case of groupwise heteroscedasticity. Using the Monte Carlo simulations, we evaluated the performance of HCCMEs for groupwise heteroscedasticity.

2. Heteroscedasticity-Consistent Covariance Matrix Estimator (HCCME)

Consider the regression model

$$y = X\beta + \varepsilon, \tag{1}$$

where y is an $n\times 1$ vector of observations on the dependent variable, X is an $n\times p$ known design matrix of rank p, β is an $p\times 1$vector of unknown parameters, and ε is an $n\times 1$ vector of random errors with zero mean and variance $\Omega = \sigma^2 I_n$, where I_n is an identity matrix of order n.

The ordinary least squares (OLS) estimator is

$$\hat{\beta}_{OLS} = \left(X'X\right)^{-1} X'y,$$

and its covariance matrix is

$$Cov\left(\hat{\beta}_{OLS}\right) = \left(X'X\right)^{-1} X'\Omega X \left(X'X\right)^{-1}, \tag{2}$$

where

$$E\left(\varepsilon\varepsilon'\right) = \Omega.$$

If the errors are homoscedastic, then $E(\varepsilon\varepsilon') = \sigma^2 I$.

So Eq. (2) becomes

$$Cov\left(\hat{\beta}_{OLS}\right) = \sigma^2 \left(X'X\right)^{-1}.$$

Defining the residuals $e_i = y - X\hat{\beta}_{OLS}$, we can estimate the OLS covariance matrix (OLSCM) as

$$OLSCM\left(\hat{\beta}_{OLS}\right) = \frac{\sum e_i^2}{n-p}\left(X'X\right)^{-1}, \tag{3}$$

where *n* is the number of observations and *p* is the number of unknown parameters.

But if the errors are heteroscedastic and the form of heteroscedasticity is unknown then the HCCME, presented by White (1980), commonly known as HC0, can be used. White's estimator is defined as

$$HC0 = \left(X'X\right)^{-1} X'diag\left[e_i^2\right] X \left(X'X\right)^{-1}.$$

The disadvantage of White's estimator is that it can be biased, usually downward, in small samples (see MacKinnon and White, 1985) and takes no account of the fact that OLS residuals tends to be too small. MacKinnon and White (1985) provided an improved estimator i.e., HC1, that was tailored after making the adjustment for the degree of freedom (d.f) as suggested by Hinkley (1977).

$$HC1 = \frac{n}{n-p}\left(X'X\right)^{-1} X'diag\left[e_i^2\right] X \left(X'X\right)^{-1}$$

$$= \frac{n}{n-p} HC0.$$

Another version of HCCME considered by MacKinnon and White (1985) is:

$$HC2 = \left(X'X\right)^{-1} X'diag\left[\frac{e_i^2}{1-h_i}\right] X \left(X'X\right)^{-1}.$$

The *i*th squared OLS residual is weighted by the reciprocal of (*1*- h_i), where h_i is the *i*th diagonal element of the hat matrix, $H = X\left(X'X\right)^{-1}X'$.

The third HCCME is the jackknife estimator of Efron (1979, as modified by Davidson and MacKinnon, 1993) and is defined as:

$$HC3 = \left(X'X\right)^{-1} X'diag\left[\frac{e_i^2}{\left(1-h_i\right)^2}\right] X \left(X'X\right)^{-1}.$$

All the three alternative estimators HC1, HC2 and HC3, are asymptotically equal to the HC0 but have superior small sample properties relative to HC0 (Long and Ervin, 2000). Long and Ervin examined the small sample properties of tests using the above stated four

versions of the HCCMEs in the linear regression model. Based on their simulation results, Long and Ervin showed that HC0 results in incorrect inferences when $n \leq 250$ while other three versions of HCCMEs work well for small sample size, especially HC3. They strongly recommended that HC3 should be used for samples less than 250 regardless of the presence or absence of heteroscedasticity.

Cribari-Neto and Zarkos (2001) showed that HCCMEs do not perform well if there are high leverage points in the design matrix. They showed that the quasi-*t* tests based on HCCMEs mentioned above are not reliable in the presence of high leverage points in the design matrix. Cribari-Neto (2004) presented the HCCME that performs quite well in the presence of high leverage points in the design matrix. This estimator was named as HC4.

$$HC4 = \left(X'X\right)^{-1} X'diag\left[\frac{e_i^2}{\left(1-h_i\right)^{\delta_i}}\right] X\left(X'X\right)^{-1},$$

where

$$\delta_i = \min\left\{4, \frac{h_i}{\bar{h}}\right\} = \min\left\{4, \frac{nh_i}{\sum_{i=1}^{n} h_i}\right\} = min\left\{4, \frac{nh_i}{p}\right\}.$$

The numerical results in Cribari-Neto (2004) showed that the asymptotic inferences in linear regression models were much affected by the presence of high leverage points in the design matrix. By the Monte Carlo simulations, Cribari-Neto (2004) showed that the quasi-*t* tests, based on the HC4 estimator, were reliable even in the presence of influential observations in the design matrix.

3. Groupwise Heteroscedasticity

The linear regression models are usually used in real life in which both the dependent and explanatory variables have individual values. There are many situations in real life, when the dependent and explanatory variable have aggregated or average values i.e. the means of dependent and explanatory variables are available and the regression is run for these average values. In spite of the fact that the error tem in the model of individual values is homoscedastic, when the observations are grouped and the regression is run on the means of dependent and explanatory variables, the error tem in the resulting model becomes heteroscedastic.

If the *n* observations are grouped into *G* groups each with n_g observations and the means $\bar{y}_g$ and $\bar{x}_g$ for $g = 1, 2, \ldots, G$ groups are observed then the regression model, given in Eq. (1), becomes

$$\bar{Y}_g = \bar{X}_g \beta + \bar{\varepsilon}_g \tag{4}$$

or

$$\bar{Y}_g = \bar{X}_g \beta + v_g.$$

The error term v_g in the above model is now heteroscedastic, irrespective of the fact that the error term in regression model (1) of individual values y_i and x_i, is homoscedastic because

$$\sigma^2{}_g = Var(v_g) = Var\left(\bar{\varepsilon}_g\right) = \sigma^2 / n_g ,$$

where n_g is the number of observations in group *G*.

4. HCCME under Groupwise Heteroscedasticity

In case of groupwise heteroscedasticity, instead of the individual values, the mean of dependent variable is regressed on the means of explanatory variables. Consider regression model (4) where both, the dependent and independent, represent the respective mean values. For this model, White's estimator can be written as

$$HC0 = \left(\bar{X}'\bar{X}\right)^{-1}\bar{X}'\,diag\left[\tilde{v}_i^2\right]\bar{X}\left(\bar{X}'\bar{X}\right)^{-1}, \tag{5}$$

where $\tilde{v}_i = \bar{y} - \bar{X}\tilde{\beta}_{OLS}$, are the OLS residuals for $\tilde{\beta}_{OLS} = (\bar{X}'\bar{X})^{-1}\bar{X}'\bar{y}$.

After applying the degree of freedom adjustment, we get the HC1 estimator as

$$\begin{aligned} HC1 &= \frac{G}{G-p}\left(\bar{X}'\bar{X}\right)^{-1}\bar{X}'\,diag\left[\tilde{v}_i^2\right]\bar{X}\left(\bar{X}'\bar{X}\right)^{-1}, \\ &= \frac{G}{G-p}HC0, \end{aligned} \tag{6}$$

where *G* is the number of groups which behaves like the sample size and *p* is the number of unknown parameters in groupwise heteroscedastic model.

The HC2 estimator can be written as

$$HC2 = \left(\bar{X}'\bar{X}\right)^{-1}\bar{X}'\,diag\left[\frac{\tilde{v}_i^2}{1-r_i}\right]\bar{X}\left(\bar{X}'\bar{X}\right)^{-1}, \tag{7}$$

where r_i is the *i*th diagonal element of the hat matrix, $R = \bar{X}\left(\bar{X}'\bar{X}\right)^{-1}\bar{X}'$.

The HC3 estimators becomes

$$HC3 = \left(\bar{X}'\bar{X}\right)^{-1}\bar{X}'\,diag\left[\frac{\tilde{v}_i^2}{\left(1-r_i\right)^2}\right]\bar{X}\left(\bar{X}'\bar{X}\right)^{-1}. \tag{8}$$

Finally, the HC4 estimator becomes

$$HC4 = \left(\bar{X}'\bar{X}\right)^{-1}\bar{X}'\,diag\left[\frac{\tilde{v}_i^2}{\left(1-r_i\right)^{\delta_i}}\right]\bar{X}\left(\bar{X}'\bar{X}\right)^{-1}, \tag{9}$$

where

$$\delta_i = \min\left\{4, \frac{r_i}{\bar{r}}\right\} = \min\left\{4, \frac{Gr_i}{\sum_{i=1}^{n} r_i}\right\} = min\left\{4, \frac{Gr_i}{p}\right\},$$

where $\bar{r}$ is the average of r_i.

These estimators are used to get the consistent covariance matrix estimator of coefficient estimates for the model with groupwise heteroscedasticity.

5. Numerical Evaluation

The numerical results, presented in this study, were obtained using the simple regression model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad ; \qquad i = 1, 2, \ldots, n. \tag{10}$$

We generated 275, 550 and 825 observations. For the case of 275 observations, the values of covariate *x* were obtained as random draws from the *U* (0, 1) distribution and then kept fixed. The errors ε_i's were taken from the normal distribution with mean 0 and variance 1 i.e. $\varepsilon_i \sim$ N (0, 1). Clearly, the error term is homoscedastic with ${\sigma_i}^2 = 1$. The dependent variable *y* was constructed using $\beta_0 = \beta_1 = 1$ under data generating process (DGP-I) and $\beta_0 = 10$, $\beta_1 = 0.5$ (DGP-II). These 275 observations were grouped into *G* = 10 groups each of size n_g where n_g = 5, 10, 15, …, 50 and the means $\bar{y}_g$ and $\bar{x}_g$ for all the 10 groups were obtained. It is obvious that $\sum_{j=1}^{G=10} n_g = 275$ and that is why we, initially generated 275 observations. Then the simple regression was run on these average values as:

$$\bar{y}_g = \beta_0 + \beta_1 \bar{x}_g + \bar{\varepsilon}_g \quad ; g = 1, 2, \ldots, 10. \tag{11}$$

The error term in the above model is now heteroscedastic irrespective of the fact that the error term in the regression model of individual values y_i and x_i is homoscedastic because ${\sigma^2}_g = Var\left(\bar{\varepsilon}_g\right) = \sigma^2 / n_g$ where n_g is the number of observations in group *G*. Then these 275 observations were replicated twice and three times to get 550 and 825 observations, respectively. For such replication, the degree of heteroscedasticity remained unchanged to make comparisons for different data sets.

Thus obtained 550 observations were divided into *G* = 20 groups of sizes n_g where n_g = 5, 10, 15, …, 50, 5, 10, 15, …, 50 and the means $\bar{y}_g$ and $\bar{x}_g$ for *g* = 1, 2, …, 20 were obtained. Thus, there were two groups of size 5, two of size 10 and so on. The data on 825 observations were divided into *G* = 30 groups of sizes n_g where n_g = 5, 10, 15, …, 50, 5, 10, 15, …, 50, 5, 10, 15, …, 50 and the means $\bar{y}_g$ and $\bar{x}_g$ for *G* = 1, 2, …, 30

groups were observed. The regression for model (11) was run on the means of 20 and 30 groups, separately. These experiments were replicated 5000 times each for G =10, 20 and 30. The estimates of the β's were computed by the OLS, and the estimates of $Cov\left(\hat{\beta}_{OLS}\right)$ were computed using Eq. (3) i.e. using the OLS and the HCCMEs, HC0, HC1, HC2, HC3 and HC4, given in Eqs. (5)-(9), respectively. All the computations were performed through programming routines in the econometric package EViews 5.0 (visit www.eviews.com).

Following Cribari-Neto and Galvão (2003) and Cribari-Neto (2004), we computed the total relative bias (TRB) of the OLS, the HC0, HC1, HC2, HC3 and HC4 covariance estimators. The total relative bias yields the sum of the absolute relative biases of the estimated variances of $\hat{\beta}_0$ and $\hat{\beta}_1$. The relative bias is defined as the difference between the mean value of variance estimator and the true value of variance divided by the true variance. Thus, the TRB is defined as

$$\text{TRB} = \frac{\left|E\left\{\hat{Var}\left(\hat{\beta}_0\right)\right\} - Var\left(\hat{\beta}_0\right)\right|}{Var\left(\hat{\beta}_0\right)} + \frac{\left|E\left\{\hat{Var}\left(\hat{\beta}_1\right)\right\} - Var\left(\hat{\beta}_1\right)\right|}{Var\left(\hat{\beta}_1\right)}.$$

Table 1 and 2 presents the TRB under DGP-I and -II respectively for the OLS, HC0, HC1, HC2, HC3 and HC4 variance estimators.

Table 1: Total Relative Bias under DGP-I

G	OLS	HC0	HC1	HC2	HC3	HC4
10	0.78	1.41	1.27	0.56	2.96	70.39
20	0.74	0.72	0.58	0.13	0.83	4.31
30	0.79	0.81	0.72	0.50	0.10	0.95

Table 2: Total Relative Bias under DGP-II

G	OLS	HC0	HC1	HC2	HC3	HC4
10	1.15	2.91	2.47	1.12	3.31	38.32
20	1.23	1.19	0.97	0.44	1.83	3.51
30	1.29	1.22	1.13	0.97	1.21	1.79

The figures in Table 1 and 2 convey that when G = 10, the HC4 estimator is much biased than the other estimators. On the other hand, the HC2 estimator yields lowest bias among all the estimators. When G = 20, we observe that the HC4 estimator is again much biased as compared to other estimators but the amount of bias reduces considerably when the number of groups (G) in which the data have been divided, increases. The HC2 estimator is again less biased than the other estimators. Cribari-Neto (2004) reported that the HC4 estimator was much biased among all the estimators and the same behavior has been observed in our case. We note that as the number of groups i.e. G increases, the TRB of

HC3 and HC4 estimators reduces, dramatically but the similar behavior is not observed for the OLS, HC0, HC1 and HC2 estimators.

Again following Cribari-Neto (2004), we computed the root mean square error (RMSE) of the OLS, HC0, HC1, HC2, HC3 and HC4 estimators. Table 2 presents the square root of the total mean squared error ($\times\sqrt{5000}$) for each variance estimator under consideration. It is defined as

$$RMSE = \sqrt{MSE\left\{\hat{Var}\left(\hat{\beta}_0\right)\right\} + MSE\left\{\hat{Var}\left(\hat{\beta}_1\right)\right\}} \times \sqrt{5000},$$

where

$$MSE\left(\hat{Var}\left(\hat{\beta}_0\right)\right) = Var\left\{\hat{Var}\left(\hat{\beta}_0\right)\right\} + \left\{Bias\left\{\hat{Var}\left(\hat{\beta}_0\right)\right\}\right\}^2,$$

and

$$MSE\left(\hat{Var}\left(\hat{\beta}_1\right)\right) = Var\left\{\hat{Var}\left(\hat{\beta}_1\right)\right\} + \left\{Bias\left\{\hat{Var}\left(\hat{\beta}_1\right)\right\}\right\}^2,$$

where $\hat{Var}\left(\hat{\beta}_0\right)$ and $\hat{Var}\left(\hat{\beta}_1\right)$ are the estimates of $Var\left(\hat{\beta}_0\right)$ and $Var\left(\hat{\beta}_1\right)$, respectively. The MSE thus comprises of not only bias but also the variance of different estimators.

Table 3: Total RMSE ($\times\sqrt{5000}$) under DGP-I

G	OLS	HC0	HC1	HC2	HC3	HC4
10	63.01	83.47	79.21	93.76	387.63	6751.08
20	25.96	36.35	37.37	48.75	81.62	226.84
30	18.08	23.69	23.83	26.20	32.69	58.65

Table 4: Total RMSE ($\times\sqrt{5000}$) under DGP-II

G	OLS	HC0	HC1	HC2	HC3	HC4
10	87.03	115.29	109.41	129.50	533.40	7324.62
20	34.86	50.21	49.61	66.33	112.73	213.37
30	23.97	32.72	32.91	36.19	45.15	81.01

Table 3 and 4 display the values of Total RMSE ($\times\sqrt{5000}$) under DGP-I and –II, respectively for each variance estimator under consideration. The numerical results presented in Table 3 and 4 show that the OLS estimator yields the smallest RMSE as compared to other estimators under consideration for all *G* (number of groups). The HC4 estimator yields the largest RMSE as compared to other estimators. Cribari-Neto (2004) also reported the similar results but for the ungrouped data.

Following Cribari-Neto (2004), we computed the estimated null rejection rate (NRR) of quasi-*t* tests, based on the OLS and HCCM estimators. Table 5, 6 and 7 presents the NRR

(in percentage) of quasi-*t* tests corresponding to the nominal level α = 1%, 5% and 10%, respectively. For convenience, we focus on the testing the null hypothesis $H_0 : \beta_1 = \beta_1^0$ against the alternative hypothesis $H_1 : \beta_1 \neq \beta_1^0$, where β_1^0 is the true value of β_1. The test statistic that has been used is

$$t = \frac{\hat{\beta}_1 - \beta_1^0}{\sqrt{\hat{\text{var}}(\hat{\beta}_1)}},$$

where $\hat{\text{var}}(\hat{\beta}_1)$ denotes the estimate of variance of $\hat{\beta}_1$ obtained from the OLS and HCCMEs. Under the null hypothesis, the test statistic has a limiting normal distribution with zero mean and unit variance. The absolute values of test statistics thus obtained are compared with the critical values from this limiting distribution corresponding to the nominal level α = 1%, 5% and 10%.

Table 5: Estimated null rejection rates of quasi *t*-tests (α = 1 %) under DGP-I

G	OLS	HC0	HC1	HC2	HC3	HC4
10	8.34	24.50	20.22	13.14	6.10	1.94
20	6.10	10.26	9.00	7.26	5.10	2.46
30	14.02	18.94	17.56	15.52	12.64	9.22

Table 5 presents the estimated NRR of the quasi *t*-tests corresponding to the nominal level α = 1% under DGP-I. The figures presented in Table 5 show that when *G* =10, the tests based on the HC0, HC1 and HC2 estimators result in larger size distortion. However, when *G* = 20, the performance of tests based on these estimators improves regarding size distortion. The tests that use the OLS, HC3 and HC4 estimators result in relatively less size distortion. We note that when *G* = 10 and 20, the tests based on the HC4 estimator, show satisfactory performance if the criterion is size distortion. For instance, when *G* = 10 and 20, the HC4-based test rejects the null hypothesis approximately 2% of the times that is close to the nominal level of the test. For *G* = 30, the tests based on all the covariance estimators, considered here, show large size distortion. However, the HC4 estimator results in less size distortion among rest of the estimators. Table 6 and 7 reveal the similar behavior for 5% and 10% level of significance, respectively under DGP-I. Table 8 presents the estimated NRR of the quasi *t*-tests corresponding to the nominal level α = 5% under DGP-II. Almost similar behavior of NRR is observed under both DGP-I and DGP-II

Table 6: Estimated null rejection rates of quasi *t*-tests (α = 5 %) under DGP-I

G	OLS	HC0	HC1	HC2	HC3	HC4
10	17.14	36.28	31.28	21.90	10.74	3.08
20	14.46	19.52	17.48	14.10	10.36	5.90
30	27.38	30.50	29.22	26.66	22.58	16.66

Table 7: Estimated null rejection rates of quasi *t*-tests (α = 10 %) under DGP-I

G	OLS	HC0	HC1	HC2	HC3	HC4
10	24.92	43.62	38.78	28.92	14.68	4.38
20	22.22	27.08	24.86	20.34	15.06	9.10
30	35.48	38.58	37.08	34.24	30.08	23.18

Table 8: Estimated null rejection rates of quasi *t*-tests (α = 5 %) under DGP-II

G	OLS	HC0	HC1	HC2	HC3	HC4
10	15.12	32.12	27.60	19.42	9.48	2.62
20	12.36	17.24	15.42	12.46	9.14	5.20
30	24.14	26.92	25.68	23.34	19.82	14.66

Following Cribari-Neto and Lima (2009), we evaluated the performance of the OLS and HCCMEs, under groupwise heteroscedasticity, for interval estimation. The confidence intervals based on the OLS and HCCM estimators are constructed for β_0 and β_1. The 100(1-α) % two sided confidence intervals for β_j (j = 0, 1) are constructed as:

$$\hat{\beta}_j \pm Z_{1-\frac{\alpha}{2}} \sqrt{\hat{Var}(\hat{\beta}_j)},$$

where $\hat{\beta}_j$ are the OLS estimator of β_j and $\hat{Var}(\hat{\beta}_j)$ the is the estimator of $Var(\hat{\beta}_j)$ obtained by the OLS and HCCM estimators, under consideration. The nominal coverage of all the confidence intervals, under consideration, is $1-\alpha = 0.95$. We compute the empirical coverage and average length of confidence intervals for the both coefficients β_0 and β_1 but for discussion here, we focus on β_1 only.

Table 9 shows the empirical coverage and average length of confidence intervals for β_1, under DGP-I, based on the OLS and HCCM estimators. The figures show that when G = 10, the empirical coverage of confidence intervals based on the HC4 estimator is good as compared to the other estimators (i.e. OLS, HC0, HC1, HC2 and HC3). This coverage is 96.92% which is close to the nominal level (95%). The HC0-intervals show the worst performance in the sense of coverage which is 63.72%. Additionally, we see that the average length of the HC4-intervals is largest. When the group size increases from 10 to 20, the empirical coverage of all the estimators tends to increase, except for the HC4-intervals. On the other hand, the average length of the HC4-intervals decreases considerably. When G = 30, the empirical coverage of confidence intervals based on all estimators reduces considerably. Here all the estimators show poor performance regarding the coverage of confidence intervals. However, the HC4 estimator performs relatively well as compared to the other estimators, giving about 83% coverage. The same measures have also been calculated under DGP-II and almost the similar behavior is observed. That is why, they have not been reported here.

Table 9: Confidence intervals for β_1: Coverage (%) and average length under DGP-I

G	OLS		HC0		HC1		HC2		HC3		HC4	
	Coverage	Length	Coverage	Length	Coverage	Length	Coverage	Length	Coverage	Length	Coverage	Length
10	82.86	3.61	63.72	2.48	68.72	2.77	78.10	3.68	89.26	6.41	96.92	22.94
20	85.54	2.66	80.48	2.57	82.52	2.71	85.90	3.05	89.64	3.68	94.10	5.29
30	72.62	2.13	69.50	2.04	70.78	2.11	73.34	2.26	77.42	2.52	83.34	3.05

6. Conclusion

It is common practice to use the HCCMEs to get the consistent estimates of variances of the OLS estimates for heteroscedastic linear regression models. In many practical situations, averages of the dependent and explanatory variables are used and regression is run on these average values. In this situation, the problem of groupwise heteroscedasticity is evident. The present study unfolds the performance of the conventional HCCMEs to encounter groupwise heteroscedasticity by providing consistent covariance matrix estimates. Five versions of HCCME i.e. HC0-HC4, have been computed for the linear regression model, facing groupwise heteroscedasticity. The Monte Carlo results reveal that the HC3 and HC4 estimators perform adequately well in the presence of groupwise heteroscedasticity as they do in the usual heteroscedastic linear regression model. However, the HC4 estimator outperforms the HC3 estimator, giving less size distortion and better coverage in interval estimation.

References

1. Ahmed, M., Aslam, M. and Pasha, G. R. (2011). Inference under heteroscedasticity of unknown form using some adaptive estimator. *Communications in Statistics-Theory & Methods,* 40, 4431-4457.

2. Aslam, M., Riaz, T. and Altaf, S. (2013). Efficient estimation and robust inference of linear regression models in the presence of heteroscedastic errors and high leverage points. *Communications in Statistics- Simulation and Computation,* 42, 2223-2238.

3. Cribari-Neto, F. (2004). Asymptotic inference under heteroskedasticity of unknown form. *Computational Statistics & Data Analysis,* 45, 215–233.

4. Cribari-Neto, F. and Galvão, N. M. S. (2003). A class of improved heteroskedasticity-consistent covariance matrix estimators. *Communications in Statistics-Theory & Methods,* 32, 1951–1980.

5. Cribari-Neto, F. and Lima, M. G. A. (2009). Heteroskedasticity-consistent interval estimators. *Journal of Statistical Computation and Simulation,* 79, 787-803.

6. Cribari-Neto, F. and Zarkos, S. G. (1999). Bootstrap methods for heteroskedastic regression models: evidence on estimation and testing. *Econometric Reviews,* 18, 211- 228.

7. Cribari-Neto, F. and Zarkos, S. G. (2001). Heteroskedasticity-consistent covariance matrix estimation: White's estimator and the bootstrap. *Journal of Statistical Computation and Simulation*, 68, 391- 411.

8. Cribari-Neto, F. and Zarkos, S. G. (2004). Leverage-adjusted heteroskedastic bootstrap methods. *Journal of Statistical Computation and Simulation*, 74, 215-232.

9. Davidson, R. and MacKinnon, J.G. (1993). *Estimation and Inference in Econometrics*. Oxford University Press, New York.

10. Efron, B. (1979). Bootstrap methods: Another look at the Jackknife. *Annals of Statistics*, 7, 1–26.

11. Flachaire, E. (2005). More efficient tests robust to heteroskedasticity of unknown form. *Econometric Reviews*, 24, 219-241.

12. Greene, W. H. (2000). Econometric Analysis. Prentice Hall, New Jersey.

13. Hinkley, D. V. (1977). Jackknifing in unbalanced situations. *Technometrics*, 19, 285–292.

14. Long, J. S., Ervin, L. H. (2000). Using heteroskedasticity consistent standard errors in the linear regression model. *The American Statistician*, 54, 217-224.

15. MacKinnon, J. G. (2011). Thirty years of heteroskedasticity-robust inference. Queen's Economics Department Working Paper No. 1268, Queen's University, Canada.

16. MacKinnon, J. G. and White, H. (1985). Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29, 305-25.

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

18. Wooldridge, J. M. (2008). Introductory Econometrics: A Modern Approach. South-Western Cengage Learning, Mason, USA.