Theoretically Proof of the Efficiency of the GLSE with Compare to the OLSE of the Parameters for the Specific From of Heteroscedasticity that is, σ ^ 2 _ i=σ ^ 2 X ^ 2 _ i with Σ ^ n _ i=1X ^ 2 _ i=n

Md. Sharif Hossain

https://doi.org/10.15017/3000219

出版情報：経済論究. 104, pp.63-79, 1999-07-30. 九州大学大学院経済学会
バージョン：
権利関係：

Theoretically Proof of the Efficiency of the GLSE with Compare to the OLSE of the Parameters for the Specific Form of Heteroscedasticity that is, $\sigma_i^2=\sigma^2X_i^2$ with $\sum_{i=1}^{n}X_i^2=n$

Md. Sharif Hossain[1)]

Abstract

The principal purpose of this research is to show theoretically that the generalised least squares estimators (GLSE) are more efficient than that of the odinary least squares estimators (OLSE) of the parameters for the specific form of heteroscedasticity that is, error variances are proportional to the square of the independent variable with the additional condition that sum of squares of the observations of the explanatong vaniable is equal to the number of observations that is, $\sum_{i=1}^{n}X_i^2=n$, where, "X" is the explanatory variable and "n" is the total number of observations. Moreover, the particular form of heteroscedasticity has been considered for a specific linear regression equation without the intercept term imposing the additional condition. In order to show theoretically the efficiency of the GLSE as compared to the OLSE of the parameters for the specific form of heteroscedasticity with the additional condition, we have to derive some important properties of the OLSE of parameters for the specific form of heteroscedasticity with the additional condition that, $\sum_{i=1}^{n}X_i^2=n$.

For empirical verification of the theoretical construct of this research, a set of real data on the total collections and collection of bound journals for 45 libraries in Bangladesh for the year 1995 has been used. The rationale behind the choice of the data set for the current purpose is also delineated.

1) The first and foremost words of thanks from me are for my reverend teacher Professor Chikayoshi Saeki who provided me for the higher study about Econometrics at the Faculty of Economics, Department of Economic Engineering of Kyushu University in Japan.

I express my indebtness to my honorable teachers Dr. Matiur Rahman, Associate Professor, Department of Statistics, University of Dhaka, Bangladesh, Dr. A.B.M.A. Sobhan Miah, Professsor, Department of Statistics, Jahangirnagar University, Savar, Dhaka, Bangladesh, for their vigilant guidance and encouragement during the entire phase of the work.

I also acknowledge my gratefulness to my honorable tutor Mr. Sadayuki Takii, who has completed the Ph. D. degree from the Faculty of Economics, Department of Economic Engineering of Kyushu University, Japan, for his hearty and sincere co-operations.

I like to express my gratefulness to my friend Noor-E-Shahreen Tuly for her valuable advice.

Table of Contents

1 Introduction

Econometric model building is based on some crucial assumptions. One such assumption is the assumption of homoscedasticity. However the assumption of homoscedasticity may not remain valid in many practical situations, particularly when one deals with cross-section data. Thus the problems associated with heteroscedasticity arise. It is well known that the violation of homoscedasticity in regression model results in impairing the properties of the parameter estimates. Now depending on the form of heteroscedasticity, properties of the parameter estimates, testing procedures and estimation techniques etc differ.

With respect to testing the presence of heteroscedasticity and estimation of the heteroscedastic model a lot of works have been done. Now, for the testing procedure, one can refer Bartlett(1937), Goldfeld and Quandt(1965, 1972), Glejser (1969), Spearman's Rank Correlation test (1972), Harvey (1974, 1976), Breush-Pagan (1979), White (1980), Harvey and Phillips (1981), Engle (1982) etc. And, also for estimation of the heteroscedastic model, one can refer to Thiel (1951, 1971), Prais (1953), Prais and Houthakker (1955), Park (1966), Kakwani and Gupta (1967), Hildreth and Houck (1968), C.R.Rao (1970, 1972), Kmenta(1971), Theil(1971), Goldfeld and Quandt (1972), Amemiya (1973), Froehlich (1973), Hartly and Jayatllake (1973), Harvey (1974), Oberhofer and Kmenta (1974), Raj (1975), Magnus (1975), Amemiya (1977), Brown (1978), Jobson and Fuller (1980), Battese and Bony Hady (1981). Harvey (1976) also developed the estimation technique for autoregressive conditional heteroscedastic (ARCH) model and estimation technique of heteroscedastic regression model with correlated disturbances have been developed by Dhrymes (1971), Oberhofer and Kmenta (1974), Magnus (1978). Recently in the field of Time Series Econometrics, generalized autoregressive conditional heteroscedastic (GARCH) model has been developed and many researchers have also paid their attentions. All the testing procedures and estimation techniques reffered to above differ among themselves depending on the form of heteroscedasticity. Likewise, properties of the parameter estimates also depend on the form of heteroscedasticity. In this paper, it is very much important and interesting to show theoretically that the GLSE are more efficient than that of the OLSE of the parameters for the assumed form of heteroscedasticity that is, $\sigma_i^2=\sigma^2X_i^2$ with the additional condition that, $\sum_{i=1}^{n}X^2{}_i=n$. To show theoretically of the efficiency of the GLSE with compare to the OLSE of the parameters for the particular form of heteroscedasticity with the additional condition, it is very much important to derive some important properties of the parameter estimates for the specific form of heteroscedasticity with the additional condition that, $\sum_{i=1}^{n}X^2{}_i=n$.

One of the main concerns of this paper is to review critically the works done so far regarding heteroscedasticity.

Finally an empirical investigation in this regard has also been provided in this paper for a given data set.

2 Meaning of Heteroscedasticity

In this part we will present a brief overview of the underlying concept of heteroscedasticity. Let us consider the general linear regression model of the form,

$$Y=X\beta+\epsilon \tag{1}$$

where,

$$Y=\begin{pmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_n \end{pmatrix};\ X=\begin{pmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{pmatrix};\ \beta=\begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_k \end{pmatrix};\ \epsilon=\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \epsilon_n \end{pmatrix}$$

where;

Y denotes the $(n\times1)$ vector of the observations of a dependent variable,

X denotes the $(n\times(k+1))$ matrix of the explanatory variables,

β denotes the $((k+1)\times1)$ vector of unknown constants,

ϵ denotes $(n\times1)$ vector of the stochactic disturbances.

The following are the usual underlying assumptions of the model.

1. ϵ_i is normally distributed for all i $(i=1, 2,...n.)$
2. $E(\epsilon_i)=\sigma^2$; for all i.
3. $\mathrm{Var}(\epsilon_i)=\sigma^2$; for all i.
4. X is a fixed matrix with no stochastic regression. X has a full column rank (no perfect multicollinearity).

5. $E(X_i, \epsilon_j)=0$;

6. $Cov(\epsilon_i, \epsilon_j)=0$; for $i \neq j$.

Accordingly we can write, $\epsilon_i \sim NID(0, \sigma^2)$.

We note that we shall refer the model given above as the general linear model untill and unless some specific form is mentioned. In OLS method the errors are assumed to be distributed independently with zero mean and constant variance. However, in many economic studies, specially in cross-section or micro-economic analysis, the assumption of constant variance may not valid. In this situation the problems of heteroscedasticity arise.

Now, heteroscedasticity may arise with correleted or uncorreleted errors, whose structures are shown below,

i) Heteroscedasticity with no autocorrelation;

$$Var(\epsilon)=\begin{pmatrix} \sigma_1^2 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \sigma_2^2 & 0 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & \sigma_n^2 \end{pmatrix};$$

ii) Heteroscedasticity with autocorrelation;

$$Var(\epsilon)=\begin{pmatrix} \sigma_1^2 & \rho^1\sigma_1\sigma_2 & \rho^2\sigma_1\sigma_3 & \cdot & \cdot & \cdot & \rho^{n-1}\sigma_1\sigma_n \\ \rho^1\sigma_2\sigma_1 & \sigma_2^2 & \rho^1\sigma_2\sigma_3 & \cdot & \cdot & \cdot & \rho^{n-2}\sigma_2\sigma_n \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho^{n-1}\sigma_n\sigma_1 & \rho^{n-2}\sigma_n\sigma_2 & \rho^{n-3}\sigma_n\sigma_3 & \cdot & \cdot & \cdot & \sigma_n^2 \end{pmatrix};$$

In either case, heteroscedasticity may occur principally in the following two forms;

1 Multiplicative form:

$\sigma_i^2=\sigma^2 Z_i^\delta$; (Lahari and Egy, 1981), where δ measures the importance of heteroscedasticity, σ^2 is the proportional constant, and Z's are variables which are responsible for heteroscedasticity.

A general representation of the multiplicative form of heteroscedasticity is as follows,

$$\log(\sigma^2{}_i)=\log\sigma^2+\delta_1\log(Z_{i1})+......+\delta_k\log(Z_{ik}) \tag{2}$$

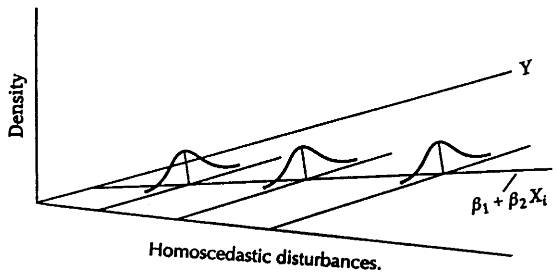
2 Additive form:

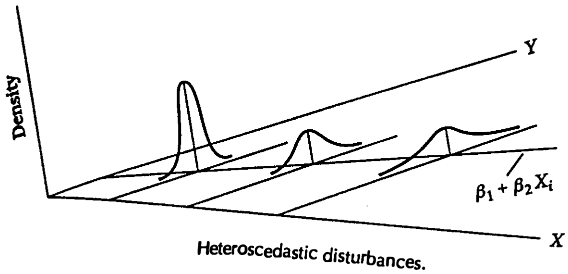
$\sigma_i^2=a+bX_i+cX_i^2$; (Goldfeld and Quandt, 1972), where a, b, and c are constant, b and c measure the importance of heteroscedasticity. X_i be the explanatory variable.

A general representation of the additive form of heteroscedasticity is as follows,

$$\sigma^2{}_i=a_0+a_1Z_{i1}+......+a_kZ_{ik};\ where;\ Z_{ik}=X^{ik} \tag{3}$$

Also the graphical representation of heteroscedasticity is shown below.





3 Plausible Sources of Heteroscedasticity

In general, heteroscedasticity may arise due to various reasons, among which the following are the most important ones.

Heteroscedasticity may be due to the omission of the explanatory variables and inclusions of the irrelevant variables in the model. Heteroscedasticity may also arise in a "number of random coefficient models". Heteroscedasticity may arise in a group data. Hateroscedasticity may arise in many econometric studies especially in cross-section or micro-economic and time series analysis.

Now, important issue arises about what are the adverse impacts of heteroscedasticity and we discuss;

4 Consequences of Heteroscedasticity

Now it is pertinent to focus on the consequences resulting from the presence of heteroscedasticity. Such consequences are delineated below.

If we apply the OLS method to the regression equation, we have $Var(\hat{\beta})=(X'X)^{-1}\sigma^2$. With heteroscedasticity the variance of ϵ would not be a finite constant figure rather would tend to change with an increasing or decreasing values of X. Thus heteroscedasticity renders the above formula inapplicable to conduct test of significance and to cnstruct confidence intervals. Moreover, for the presence of heteroscedasticity the unbiasedness of the parameters is not affected but they have no the minimum variance property. So we can say that in the case of the presence of heteroscedasticity, the OLS estimates do not satisfy the BLUE property. And also they are not asymptotically efficient. We can say that the prediction of a dependent variable Y for a given values of the independent variables X's based on estimators $\hat{\beta}$ from the data exhibiting heteroscedasticity would be inefficient, because the variance of the prediction includes the variances of ϵ and of the parameter estimates, which are not minimum due to the incidence of heteroscedasticity.

An important aspect of applied econometric research is testing the assumptions involved in econometric model. Thus, when dealing with a cross-section data set, test for heteroscedasticity is an important issue because of the assumption of the constant variance of the random disturbances across observations is likely to be violated. However, none of the tests for heteroscedasticity and none of the estimation techniques for the heteroscedastic model can be considered to be final. Depending on the structure of heteroscedasticity both testing procedures and estimation procedures will differ.

In this paper an attempt has been made to show theoretically of the efficiency of the GLSE with compare to the OLSE of the parameters for the specific form of heteroscedasticity that is, $\sigma_i^2=\sigma^2X_i^2$, with the additional condition that, $\sum_{i=1}^{n}X_i^2=n$. For this purpose at first an attempt has been made to derive some important properties of the ordinary least squares estimates when a particular heteroscedastic structure is considered with the additional condition that, $\sum_{i=1}^{n}X_i^2=n$. Hence, now we pass to next to derive some important properties of the OLSE of parameters in a linear regression model for a specific form of heteroscedasticity with the additional condition that, $\sum_{i=1}^{n}X_i^2=n$.

5 Derivation of Some Important Properties of the OLSE of the Parameters for a Specific Form of Heteroscedasticity that is, $\sigma_i^2=\sigma^2X_i^2$ with the Additional Condition that, $\sum_{i=1}^{n}X_i^2=n$.

5.1 Introduction

Econometric model may be linear or non-linear. If the model is non-linear, then in most cases by taking a suitable transformation, the non-linear model can be transformed into a linear model.

However, let us consider the general linear regression equation

$$Y=X\beta+\epsilon \tag{4}$$

where,

Y be the $(n\times1)$ vector of the observations of a dependent variable,

X be the $(n\times(k+1))$ non-stochastic designed matrix of the explanatory variable,

β be the $((k+1)\times1)$ vector of unknown parameters,

ϵ be the $(n\times1)$ vector of random error terms.

The above model is based on the following assumptions,

$E(\epsilon_i)=0,\ E(\epsilon_i^2)=\sigma^2,\ E(\epsilon_i\epsilon_j)=0$; *for* $i\neq j$.

$\epsilon_i \sim NIID(0, \sigma^2)$ for all i.

When, $E(\epsilon_i^2)\neq\sigma^2$, then the problem of heteroscedasticity arises.

Heteroscedasticity may arise in various forms, but our interest lies in a specific multiplicarive form of heteroscedasticity that is, $\sigma_i^2=\sigma^2X_{i1}^{\delta 1}\ X_{i2}^{\delta 2}...X_{ik}^{\delta k}$

Now, the specific form of heteroscedasticity, for which our interest lies in investigating some important properties of the OLSE of parameters in general linear model which is passing through the origin, is $\sigma_i^2=\sigma^2 X_i^2$, with the additional condition that, $\sum_{i=1}^{n}X_i^2=n$.

For this specific form of heteroscedasticity with the additional condition we derive some important properties of the OLSE of parameters.

In order to derive some important properties of the OLSE of parameters, our considerable regression equation is,

$$Y_i=\beta X_i+\epsilon_i,\ (i=1, 2, 3......n.) \tag{5}$$

where,

Y_i be the ith value of the dependent variable,

X_i be the ith value of the independent variable,

β be the unknown parameter,

ϵ_i be the ith random error term.

The underlying assumptions are as follows,

$E(\epsilon_i)=0,\ E(\epsilon_i^2)=\sigma^2X_i^2,\ E(\epsilon_i, \epsilon_j)=0$; *for* $i\neq j$, and $\sum_{i=1}^{n}X_i^2=n$.

$\epsilon_i \sim NID(0, \sigma_i^2)$

In matrix notation the equation (5) can be written as follows,

$$Y=X\beta+\epsilon \tag{6}$$

where,

$$Y=\begin{pmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_n \end{pmatrix};\ X=\begin{pmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ X_n \end{pmatrix};\ \epsilon=\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \epsilon_n \end{pmatrix}$$

If we apply the OLS method to the equation (6) we have,

$$\hat{\beta}=(X'X)^{-1}X'Y \tag{7}$$

Let us now examine the OLSE of β for the assumed form of heteroscedasticity with the additional condition that, $\sum_{i=1}^{n}X_i^2=n$.

5.2 Properties of $\hat{\beta}$

5.2.1: Unbiasedness

We have,

$$\begin{aligned}\hat{\beta}&=(X'X)^{-1}X'Y\\&=(X'X)^{-1}X'(X\beta+\epsilon)\\&=\beta+(X'X)^{-1}X'\epsilon\end{aligned} \tag{8}$$

Now, taking expectation on both sides of the equation (8) we have,

$$E(\hat{\beta})=\beta,\ [since,\ E(\epsilon)=0]$$

This shows, that unbiasedness of the OLSE of β is not affected, if the disturbances are heteroscedastic of the given form for the given model.

5.2.2: Consistency

We have,

$$\hat{\beta}=\beta+(X'X)^{-1}X'\epsilon \tag{9}$$

Now, taking probability limit of the equation (9) we can write that,

$$\mathrm{Plim}(\hat{\beta})=\beta+\lim_{n\to\infty}\left(\frac{X'X}{n}\right)^{-1}\mathrm{plim}\left(\frac{X'\epsilon}{n}\right) \tag{10}$$

But, the term $\frac{X'\epsilon}{n}$ has a zero mean, and variance-covariance matrix is $\frac{\sigma^2X'\Omega X}{n^2}$.

where,

$$\Omega=\begin{pmatrix}X_1^2 & 0 & 0 & \cdot & \cdot & 0\\ 0 & X_2^2 & 0 & \cdot & \cdot & 0\\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot\\ 0 & 0 & 0 & \cdot & \cdot & X_n^2\end{pmatrix}$$

If $\lim_{n\to\infty}\frac{X'\Omega X}{n}$ *is finite then* $\lim_{n\to\infty}\frac{X'\Omega X}{n^2}$ *will be zero*

Here, $\frac{X'\epsilon}{n}$ has a zero mean and its variance-covariance matrix vanishes asymptotically, which implies that,

$$\mathrm{plim}\frac{X'\epsilon}{n}=0$$

hence,

$$\mathrm{plim}(\hat{\beta})=\beta$$

This shows that if $\lim_{n\to\infty}\frac{X'\Omega X}{n}$ is finite, and then $\hat{\beta}$ is consistent, if the disturbances are of the given form.

5.2.3: Variance of $\hat{\beta}$ for the Assumed Form of Heteroscedasticity with the Additional Condition that, $\sum_{i=1}^{n}X_i^2=n$

We have,

$$\hat{\beta}=(X'X)^{-1}X'Y \tag{11}$$

Now,

$$X'X=(X_1\ X_2 \ldots Xn)\begin{pmatrix}X_1\\X_2\\\cdot\\\cdot\\X_n\end{pmatrix}$$

$$=\sum_{i=1}^{n}X_i^2$$

$$=n[since,\ \sum_{i=1}^{n}X_i^2=n] \tag{12}$$

And therefore we have,

$$(X'X)^{-1}=\frac{1}{n}$$

Again,

$$X'Y=(X_1\ X_2 \ldots Xn)\begin{pmatrix}Y_1\\Y_2\\\cdot\\\cdot\\Y_n\end{pmatrix}$$

$$=\sum_{i=1}^{n}X_iY_i$$

$$=\sum_{i=1}^{n}X_i(\beta X_i+\epsilon_i)$$

$$=\beta\sum_{i=1}^{n}X_i^2+\sum_{i=1}^{n}X_i\epsilon_i$$

$$=n\beta+\sum_{i=1}^{n}X_i\epsilon_i \tag{13}$$

Now, putting these values in equation (11) we can write that,

$$\hat{\beta}=\beta+\sum_{i=1}^{n}\frac{X_i\epsilon_i}{n}$$

$$\Longrightarrow\hat{\beta}-\beta=\sum_{i=1}^{n}\frac{X_i\epsilon_i}{n} \tag{14}$$

The variance of $\hat{\beta}$ is given by,

$$Var(\hat{\beta})=E(\hat{\beta}-\beta)^2$$

$$=E\left(\frac{\sum_{i=1}^{n}X_i\epsilon_i}{n}\right)^2[putting\ the\ value\ of\ (\hat{\beta}-\beta)]$$

$$=\frac{1}{n^2}\left[E\left(\sum_{i=1}^{n}X_i^2\epsilon_i^2+\sum_{i=1}^{n}\sum_{j=1, i\neq j}^{n}X_iX_j\epsilon_i\epsilon_j\right)\right]$$

$$=\frac{\sum_{i=1}^{n}X_i^2}{n^2}E(\epsilon_i^2)[since,\ E(\epsilon_i,\ \epsilon_j)=0\ for\ i\neq j]$$

$$=\frac{\sum_{i=1}^{n}X_i^2}{n^2}\sigma^2X_i^2$$

$$=\sigma^2\frac{\sum_{i=1}^{n}X_i^4}{n^2} \quad (15)$$

5.2.4: BLUE Property

We know that

$$Y_i=\beta X_i+\epsilon_i \quad (16)$$

Now, dividing the equation (16) by X_i, we can write that,

$$\frac{Y_i}{X_i}=\beta+\frac{\epsilon_i}{X_i}$$

$$Y_i^*=\beta+\epsilon_i^* \quad (17)$$

Here, $E(\epsilon_i^*)=0,\ Var(\epsilon^*_i)=\frac{1}{X_i^2}Var(\epsilon_i)=\sigma^2,\ E(\epsilon_i^*,\epsilon_j^*)=\frac{1}{X_iX_j}E(\epsilon_i,\epsilon_j)=0.$

Since ϵ_i^* is a function of a normal variate ϵ_i, hence ϵ_i^* is also a normal variate. Thus, the new random error term ϵ_i^* satisfies all the usual assumptions, hence we can apply the OLS method to the transformed equation (17), after applying the OLS method to the transformed equation (17) we have,

$$\hat{\beta}=\sum_{i=1}^{n}\frac{Y_i^*}{n}$$

$$=\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i}{X_i} \quad (18)$$

This formula for the best linear unbiased estimator of β is obviously different from that for the least squares estimator. Thus our conclusion is that, the least squares estimators are not BLUE when disturbances are heteroscedastic. From this conclusion we can say that the least squares estimators do not have the smallest variance among all unbiased estimators and therefore they are not efficient.

5.2.5: Asymptotically Efficiency

The likelihood function of the sample observations $(Y_1\ Y_2 \ldots\ldots Y_n)$ is given by

$$L=\left(\frac{1}{\sqrt{2\pi\sigma^2_i}}\right)^2 exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{Y_i-\beta X_i}{\sigma_i}\right)^2\right] \quad (19)$$

Taking logarithm of both sides of the equation (19) then we have,

$$\log(L)=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log(\sigma_i^2)-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{Y_i-\beta X_i}{\sigma_i}\right)^2 \tag{20}$$

Now, taking the partial derivatives of the equation (20) with respect to $\tilde{\beta}$ (MLE of β) and then equating to zero we can write that,

$$\frac{\delta \log L}{\delta\beta}=-2\sum_{i=1}^{n}\frac{1}{\sigma_i^2}(Y_i-X_i\tilde{\beta})(-X_i)=0$$

$$\Longrightarrow\sum_{i=1}^{n}\frac{Y_iX_i}{\sigma_i^2}-\tilde{\beta}\sum_{i=1}^{n}\frac{X_i^2}{\sigma_i^2}=0$$

$$\Longrightarrow\tilde{\beta}\sum_{i=1}^{n}\frac{X_i^2}{\sigma^2X_i^2}=\sum_{i=1}^{n}\frac{Y_iX_i}{\sigma^2X_i^2}$$

$$\Longrightarrow n\tilde{\beta}=\sum_{i=1}^{n}\frac{Y_i}{X_i}$$

$$\Longrightarrow\tilde{\beta}=\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i}{X_i} \tag{21}$$

Therefore, from the equation (21) we have,

$$Var(\tilde{\beta})=\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{X_i^2}Var(Y_i)\ \ [\textit{Covariance terms will be vanished}]$$

$$=\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{X_i^2}Var(\beta X_i+\epsilon_i)$$

$$=\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{X_i^2}Var(\epsilon_i)$$

$$=\frac{1}{n^2}\sum_{i=1}^{n}\frac{\sigma^2X_i^2}{X_i^2}$$

$$=\frac{\sigma^2}{n} \tag{22}$$

Clearly, the variance of the least squares estimator of β given in (15) is different from that of the variance of the maximum likelihood estimator given in (22), what ever may be the sample size.

Therefore, since the variance of the least squares estimator is not asymptotically equivalent to the variance of the maximum likelihood estimator, hence the least squares estimators, are not asymptotically efficient, when the disturbances are heteroscedastic of the given form.

5.2.6: Varaince of the GLSE for the Assumed Form of Heteroscedasticity with the Additional Condition that, $\sum_{i=1}^{n}X_i^2=n$

From equation (17) the GLSE of the parameter β is given by,

$$\hat{\beta}_g=\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i}{X_i} \tag{23}$$

From equation (23), variance of $(\hat{\beta}_g)$ is given by,

$$Var(\hat{\beta}_g)=\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{X_i^2}Var(Y_i)\ \ [\textit{Covariance terms will be vanished}]$$

$$=\frac{1}{n^2}\sum_{i=1}^{n}\frac{\sigma^2 X_i^2}{X_i^2}$$

$$=\frac{\sigma^2}{n} \tag{24}$$

Now, we will pass to show theoretically that the GLSE are more efficient than that of the OLSE for the assumed form of heteroscedasticity with the additional condition that, $\sum_{i=1}^{n} X_i^2 = n$.

6 Theoretically Proof of the Efficiency of the GLSE with Compare to the OLSE of the Parameters for the Assumed Form of Heteroscedasticity that is, $\sigma_i^2=\sigma^2 X_i^2$ with the Additional Condition that, $\sum_{i=1}^{n} X_i^2 = n$

We have also derived the varince of the OLSE and also the variance of the GLSE in the previous section which are given below,

Variance of the OLSE of the parameter is as follows,

$$Var(\hat{\beta})=\frac{\sigma^2}{n}\sum_{i=1}^{n}\frac{X_i^4}{n} \tag{25}$$

And variance of the GLSE of the parameter is as follows,

$$Var(\hat{\beta}_g)=\frac{\sigma^2}{n} \tag{26}$$

The ratio of the variances of OLSE and GLSE of the parameter is given by

$$\frac{Var(\hat{\beta})}{Var(\hat{\beta}_g)}=\frac{\sum_{i=1}^{n} X_i^4}{n} \tag{27}$$

Let us consider the two real number a_i and b_i. Now, using the Cauchy-Schwartz inequality we can write that,

$$\left(\sum_{i=1}^{n} a_i, b_i\right)^2 \leq \sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} b_i^2 \tag{28}$$

Let, us now putting $a_i=X_i^2$, and $b_i=1$, then from the equation (28) we can write that,

$$\left(\sum_{i=1}^{n} X_i^2\right)^2 \leq n\sum_{i=1}^{n} X_i^4$$

$$\Longrightarrow n^2 \leq n\sum_{i=1}^{n} X_i^4 \ \left[since,\ \sum_{i=1}^{n} X_i^2 = n\right]$$

$$\Longrightarrow 1 \leq \frac{\sum_{i=1}^{n} X_i^4}{n}$$

$$\Longrightarrow \frac{Var(\hat{\beta})}{Var(\hat{\beta}_g)} \geq 1 \quad \left[since,\ \frac{Var(\hat{\beta})}{Var(\hat{\beta}_g)}=\frac{\sum_{i=1}^{n} X_i^4}{n}\right]$$

$$\Longrightarrow Var(\hat{\beta}) \geq Var(\hat{\beta}_g) \tag{29}$$

Thus from the equation (29) we have to conclude that the GLSE of regression coefficients are more efficient than that of the OLSE for the assumed form of heteroscedasticity with the additional condition that, $\sum_{i=1}^{n}X_i^2=n$.

Now to show empirically the efficiency of the GLSE as compare to the OLSE, a real data set, includes the number of journals and total collections for 45 libraries in Bangladesh for the year 1995 has also been collected. The reasons, for choicing the data set for current purpose is also delineated below.

7 Data Collection

To demonstrate the appropriateness of the theoretical construct, when heterogeneity of variances are assumed to be of the particular form, *i.e.* $\sigma_i^2=\sigma^2X_i^2$, a real data set has been collected, which includes the number of bound journals and total collections for 45 libraries in Bangladesh for the year 1995.

The total collections including the collection of books represents the explanatory variable and the collection of bound journals represents the dependent variable.

The suitability of data set considered is highlighted below.

1. In the study, the specific form of regression equation considered is, $Y_i=\beta X_i+\epsilon_i$. The type of data considered is supposed to conform to the above model, because when the value of the total collections is zero then the value of the collection of bound journals must be zero. Here Y_i represent the collection of bound journals and the explanatory variable X_i represents the total collections corresponding to the ith library.

2. High possibility for the existence of the heteroscedastic problem in the data can be justified in the following way. We are dealing with a cross-section data on a number of micro-economic units which are libraries. It can be postulated that variability of bound journals in the libraries with small number of total collections is smaller than that of the libraries with large number of total collections.

Thus if X_i be the ith value of the X vector, the variance of Y_i around its expected value $X_i\beta$ can be assumed to be low for small number of total collections and high for large number of total collections. At a point of time for each library variation in the collection in different libraries are basically due to the following conceivable reasons.

i : The budget allocation for each library is different.

ii : Variation in demands for library materials of different libraries. In this location of the library plays an important role basically regarding the number of users and purpose to be surve by the library.

iii: The libraries which are established in most important regions have the larger collection of books than the libraries which are established in less important regions.

For example, the libraries which are established in Dhaka which is the capital city of Bangladesh has the larger collections than the libraries which are established in other district towns like as Chittagong, Commilla, Munshigonj, Rajshahi etc.

The varying nature of the factors omitted in the regression equation for explaying variation in the total number of bound journals may introduce disturbances with different variances. Thus, we have sufficient reasons to suspect the problem of heteroscedasticity of error terms in the given data set.

Now, on the basis of the real data set, we discuss the results for the efficiency of the GLSE as compare to the OLSE.

8 Discussion on Results on the Basis of the Given Data Set

8.1 Introduction

The efficiency of the GLSE relative to the OLSE, is illustrated with the help of the real data set. Now we pass to discuss the empirical results.

8.2 Empirical Results:

In this section we have principally presented the different results obtained by OLS and GLS method for the specific regression equation formulated for the analysis of the data. Since the data set, does not satisfy the assumed condition $\sum_{i=1}^{n} X_i^2 = n$. Hence the results are obtained by omitting this condition. The omission of the condition has no any bad impact on the results. For the given data set we have obtained that the estimated values of the OLSE and GLSE of β are respectively 2.11410E-001 and 9.22671E-002. Thus, for the given data set we can say that per unit increase in value of total collections, the collection of bound journals will be increased as 2.11410E-001.

In addition, the loss of efficiency under the assumed form of heteroscedasticity of the simple least squares estimates as a compared with the generalized least squares estimation can be illustrated as $\frac{Var(\hat{\beta})}{Var(\beta_g)}=4.994058$. Thus we can say that the variance of the least squares estimate of β is more than that 5 times as large as that of the GLSE of β. That implies that the GLSE is more efficient than that of the OLSE for the given data set.

We also presented the results of OLSE, GLSE and the ratio of the variances of OLSE and GLSE of the parameter obtained for the given data set with the following table.

Table
Results of OLSE, GLSE and the Ratio of the Variances of OLSE and GLSE

Name of Estimators	Value
OLSE	2.114104E-001
GLSE	9.226721E-002
Ratio of the Variances	4.9940580

Now, we pass on the next section to draw an overall conclusion

9 Discussion and Conclusion

The main concern of this thesis is to show theoretically of the efficiency of the GLSE as compare to the OLSE of the parameters for the specific heteroscedastic model without intercept term with the additional condition that, $\sum_{i=1}^{n}X_i^2=n$. In order to show the efficiency of the GLSE of parameters as compared to the OLSE of parameters for the specific heteroscedastic model with the additional condition, we have also derived some important properties of the OLSE of parameters for the specific heteroscedastic model with the additional condition. Now to make an appropriate start of the research we have to made a wide literature survey. We have also highlighted some statistical facts to provide some motivations for the particular form of heteroscedasticity.

It is noted that in many economic studies, specially in micro-economic analysis the heteroscedastic problem arises. The heteroscedastic problem may arise with correlated error terms or without correlated error terms. Thus, for the analysis of the heteroscedastic data set, it is very much important to know the particular form of heteroscedasticity, and in this paper we concentrated on a particular type of multiplicative form of heteroscedasticity.

As regards of the optimum properties of the OLSE of the parameters, it has been examined that the OLSE of the parameters are unbiased, consistent but not BLUE for the given heteroscedastic model. It has also been observed that OLS estimators are not asymptotically efficient for the given form.

We have also obtained the GLSE of the parameters for the given form of heteroscedasticity and we have examined theoretically that the GLSE of the parameter is BLUE and it is also efficient as compared to the OLSE for the given form of heteroscedastic model with the additional condition. In order to examine empirically how much efficient is the GLSE than that of OLSE of parameter we have calculated the variance of the GLSE and of the OLSE of the parameters for a given data set. In order to demonstrate the results, a real data set has been used which was collected from the 45 libraries in Bangladesh for the year 1995. From the estimated results we have concluded that the per unit increase in value of total collections the collection of bound journals will be increased as 2.11410E-001, which indicates that the collection of average number of bound journals of the libraries in Bangladesh is very poor. For the higher study, this collection of bound journals is insufficient. That is why, we can conclude that, it is very necessary to increase the collection of bound journals for the higher study in Bangladesh. From the another estimated results we have also concluded that the variance of the OLSE of the parameters is about 5 times as large as that of the GLSE of the parameters. Therefore we have to conclude that for the given data set the generalised least squares estimates are more efficient (about 5 times) than taht of the OLSE of the parameters. Finaly, we can conclude that the variability of the collection of bound journals from library to liabrary is very high in Bangladesh.

BIBLIOGRAPHY

1. Ali Mukhter, M. and Silver, J. Lew. (1985) "Test for Equality Between Sets of Cofficients in Two Linear Regression Models Under Heteroscedasticity," *Journal of the American Statistical Association*, 80, 730-735.
2. Amemiya, T. (1937) "Regression Analysis When the Variance of the Dependent variable is Proportional to the Square of its Expectation," *Journal of the American Statistical Association*, 68, 928-934.
3. Amemiya, T. (1977) "A Note on a Heteroscedastic Model," *Journal of Econometrics*, 6, 365-370.
4. Anscombe, F. J. and Tukey, J. W. (1965) "The Experiment and Analysis of Residuals," *Technometrics*, 5, 141-160.
5. Battese, G. E. and B. P. Bonyhard (1981) "Estimation of Household Expenditure Function; An Application of a Class of Heteroscedastic Regression Models," *The Economic Record*, 57, 80-85.
6. Box, G. E. P. and W. J. Hill (1974) "Correcting Inhomogeneity of Variances with Power Transformation Weighting," *Technometrics*, 16, 385-398.
7. Brusch, T. S. and A. R. Pagan (1979) "A Simple Test for Heteroscedasticity and Random Coefficient Variation," *Econometrica*, 47, 1287-1294.
8. Brown, K. G. (1978) "Estimation of Variance Components Using Residuals," *Journal of the American Statistical Association*, 73, 141-146.
9. Carroll, R. J. and D. Ruppert (1982) "A Comparison Between Maximum Likelihood and Generalized Least Squares in a Heteroscedastic Linear Model," *Journal of the American Statistical Association*, 77, 878-882.
10. Carroll, R. J (1982) "Adapting for Heteroscedasticity in Linear Models," *Annals of Statistics*, 10, 1223-1224.
11. Carroll, R. J. Wu, C. F. J. and Ruppert (1988) "The Effect of Estimating Weights in Weighted Least Squares," *Journal of the American Statistical Association*, 83, 1045-1054.
12. Chew, V. (1965) "Covariance Matrix Estimation in Linear Model," *Journal of the American Statistical Association*, 65, 171-181.
13. Chow, G. C. (1960) "Test of Equality Between Sets of Coefficients in Two Linear Regression Models," *Econometrica*, 28, 591-605.
14. Cragg, J. G. (1982) "Estimation and Testing in Time Series Regression Models with Heteroscedastic Disturbances," *Journal of Econometrics*, 20, 135-157.
15. Dyer, D. D. and J. P. Keating (1980) "On the Determination of Critical Values for Bartlett's Test," *Journal of the American Statistical Association*, 75, 313-319.
16. Egy, D. and K. Lahiri (1979) "On Maximum Likelihood Estimation of Functional Form and Heteroscedas-

ticity," *Economics Letters*, 2, 155-159.

17. Engle, R. F. (1982) "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987-1008.
18. Froehlich, B. R. (1973) "Some Estimation for a Random Coefficient Regression Model," *ournal of the American Statistical Association*, 68, 329-331.
19. Fuller, W. A. and Rao, J. N. K. (1978) "Estimation for a Linear Regression Model with Unknown Diagonal Covariance Matrix," *Annals of Statistics*, 6, 1149-1158.