ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 13, July 2017

Nature and Some Tests of Heteroskedastacity

Siddamsetty Upendra [1], Prof. R. Abbaiah[2], B. Venkatesh[3]

Research Scholar, Department of Statistics, SVU, Tirupati, India[1]

Professor, Research Supervisor, Department of Statistics & Principal of Colleges of Sciences, SVU, Tirupati, India[2]

Research Scholar, Department of Statistics, SVU, Tirupati, India[3]

ABSTRACT: This paper proposes the meaning and nature of heteroskedastacity in econometric theory. This accommodates the sources to arise heteroskedastacity, consequences of heteroskedastacity and some tests like Goldfeld-Quandt test, Bartlett test, Breussch Pagan test, Glesjer test, Spearrman's Rank Correlation test, White test and Lagrange Multiplier test of heteroskedastacity. This paper also briefly discuss with Maximum Likelihood Estimation under Linear Probabilistic model.

KEYWORDS: Homoscedasticity, heteroskedastacity, sources, consequences, tests of heteroskedastacity

I. INTRODUCTION

Consider the general linear model $Y = X\beta + \varepsilon$ it is assumed that

$(i)E(\varepsilon) = 0$ $(ii)E(\varepsilon\varepsilon') = \sigma^2 I$ $(iii)\rho(X) = k, k < n$ (iv) X is a matrix of fixed known coefficients

(V) There is no error involved in the variable Y and X's

The assumption $E(\varepsilon\varepsilon') = \sigma^2 I = \begin{bmatrix} \sigma^2 & 0 & ..... & 0 \\ 0 & \sigma^2 & ..... & 0 \\ : & : & : & : \\ 0 & 0 & ..... & \sigma^2 \end{bmatrix}$

$$\Rightarrow \text{cov}(\varepsilon_i \varepsilon_j) = 0; \forall i \neq j = 1,2,----,n$$
$$= \sigma^2; \forall i = j = 1,2,----,n$$

The disturbances ε_i 's are uncorrelated with each other and have same unknown variance σ^2 .This assumption is known as the assumption of **Homoscedasticity** of disturbances.

The problem of Heteroscedasticity exists if the diagonal elements of variance co-variance matrix of ε are not identical, i.e., $\sigma^2 I = \left[\sigma_1^2 \quad \sigma_2^2 \quad ..... \quad \sigma_n^2\right]$ i.e., if the disturbances ε_i 's not having constant variance but having different variances and are uncorrelated then we say that there is a problem of **Heteroscedasticity**. In this case the disturbances are called Heteroscedastic disturbances.

II. SOURCE OF HETEROSCEDASTICITY

There are various reasons due to which the heteroscedasticity is introduced in the data. Some of them are as follows:

1. When the observations are based on the average data
2. In a number of random coefficients regression model
3. When the error variance changes with explanatory variables
4. The incorrect data transformations and incorrect functional form of the model can also give rise to the heteroscedastacity problem.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 13, July 2017

Consequences of the violation of the assumptions of Homoscedasticity
(Or)
Consequences of Heteroscedastic disturbances

Consider the general linear model $Y = X\beta + \varepsilon$,It is assumed that $(i)E(\varepsilon) = 0$ $(ii)E(\varepsilon\varepsilon') = \sigma^2 I$ $(iii)\rho(X) = k, k < n$ (iv) X is a matrix of fixed known coefficients(v)There is no error involved in the variable Y and X's

If $$E(\varepsilon\varepsilon') = \sigma^2\Omega = \sigma^2\begin{bmatrix} \frac{1}{\lambda_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\lambda_n} \end{bmatrix}_{nXn}$$

Where the λ 's are assumed to be known positive numbers but is unknown, i.e., ε 's heteroskedastic disturbances. If we apply Ordinary Least Squares (OLS) estimation to this model, the following consequences may be observed.

(i) OLS estimators are linearly unbiased, Consider the OLS estimator

$\hat{\beta} = (X'X)^{-1}(X'Y) = (X'X)^{-1}X'(X\beta + \varepsilon) = (X'X)^{-1}(X'X\beta) + (X'X)^{-1}X'\varepsilon = \beta + (X'X)^{-1}X'\varepsilon$ Hence, OLS estimators are linearly unbiased, i.e., $E(\hat{\beta}) = \beta \qquad \because E(\varepsilon) = 0$

(ii) OLS estimators are not having minimum variance when compared with Generalized Least Squares (GLS) Estimators,i.e., OLS estimators are not efficient under the condition of heteroskedastacity

Consider

$$\hat{\beta} = (X'X)^{-1}(X'Y) \Rightarrow \hat{\beta} - \beta = (X'X)^{-1}X'\varepsilon$$
$$\therefore V(\hat{\beta}) = E\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)\right]$$
$$= \sigma^2(X'X)^{-1}(X'\Omega X)(X'X)^{-1}$$
$$\because E(\varepsilon\varepsilon') = \sigma^2\Omega$$

The Best Linear Unbiased Estimator (BLUE) of β in the case of heteroskedastic disturbances is the GLS estimator $\tilde{\beta}$ which is given by $\tilde{\beta} = \left(X'\Omega^{-1}X\right)^{-1}\left(X'\Omega^{-1}Y\right)$ with $V\left(\tilde{\beta}\right) = \sigma^2\left(X'\Omega^{-1}X\right)^{-1}$

Let us consider a simple two variable linear model in deviation form as

$$y_i = x_i\beta + \varepsilon_i \text{ ; i=1,2,-----, n}$$
$$E(\varepsilon\varepsilon') = \sigma^2\Omega$$

If we define a non-singular matrix

$$\rho^{-1} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_n} \end{bmatrix}_{nxn} \text{ Then } \rho^{-1'}\rho^{-1} = \Omega^{-1} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}_{nxn} \text{ Also } X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \text{ and } Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\therefore \left(X'\Omega^{-1}X\right) = \sum_{i=1}^{n}\lambda_i x_i^2$$

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 13, July 2017

The GLS estimator of β is given by $\tilde{\beta} = \left(X'\Omega^{-1}X\right)^{-1}\left(X'\Omega^{-1}Y\right) = \dfrac{\sum_{i=1}^{n}\lambda_i x_i y_i}{\sum_{i=1}^{n}\lambda_i x_i^2}$

And $V(\tilde{\beta}) = \sigma^2\left(X'\Omega^{-1}X\right)^{-1} = \dfrac{\sigma^2}{\sum_{i=1}^{n}\lambda_i x_i^2}$ we have OLS estimator of β as $\hat{\beta} = \dfrac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$ and its variance is given by

$$V(\hat{\beta}) = \sigma^2\left(X'X\right)^{-1}\left(X'\Omega X\right)\left(X'X\right)^{-1} = \frac{\sigma^2}{\left[\sum_{i=1}^{n} x_i^2\right]^2}\left\{\sum_{i=1}^{n}\frac{x_i^2}{\lambda_i}\right\}$$

We can show that $V(\hat{\beta})$ is greater than $V(\tilde{\beta})$ under heteroskedastic disturbances.

Hence OLS estimators are not efficient under the heteroskedastacity.

(iii) OLS estimators are consistent

(iv) OLS estimators are asymptotically inefficient

(v) The OLS estimator of σ^2 is biased in the case of heteroskedastacity.

i.e., $\hat{\sigma}^2 = \dfrac{e'e}{n-k}$ is not unbiased estimator of σ^2 but $\tilde{\sigma}^2 = \dfrac{e'\Omega^{-1}e}{n-k}$ is unbiased estimator of σ^2.

Hence, we cannot apply the formula of variance of regression coefficients to conduct tests of significance and construct confidence intervals.

(vi) The prediction of Y for given X based on OLS estimators $\hat{\beta}$ will have high variance and hence the prediction will be inefficient.

III. TESTS OF HETEROSKEDASTACITY

Goldfeld- Quandt Test:

We can setup the null-hypothesis as H_0: There is no heteroskedastacity in the data.To test the above null-hypothesis, the steps are as follows:

Step-1: Order the observations on Y according to the magnitude of the observations on X_j and give ranks to the observations of X_j.

Step-2: Omit C central observations from the analysis.

Step-3: Divide the remaining (n-c) observations into two groups of equal size$\left(\frac{n-c}{2}\right)$ provided $\left(\frac{n-c}{2}\right)$>k.

Step-4: Fit separate regressions by OLS method to these two groups. Find residual sum of squares for each of these groups as $S_1=\sum e_1^2$ and $S_2=\sum e_2^2$ respectively. Then $F = \dfrac{S_1}{S_2}$ follows $F\left(\dfrac{n-c-2k}{2}, \dfrac{n-c-2k}{2}\right)$, here k is the number of parameters. Compare F-calculated value with F-table value and drawn the inference accordingly.

Bartlett Test:

First split Y data into m-classes according to the size of Y. We state the null-hypothesis asH_0: all the m-classes have same variance, i.e., $\sigma_1^2 = \sigma_2^2 = - - - - = \sigma_m^2$.

To test the above null-hypothesis, we use the likelihood ratio test statistic as follows

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 13, July 2017

$\lambda = \sum_{i=1}^{m}\left(\frac{S_1^2}{S_2^2}\right)^{\frac{n_i}{2}}$, Where S_i^2 is the variance of i^{th} sample of size of n_i observations. Also $nS^2 = \sum_{i=1}^{m} n_i S_i^2 i^{th}$, $n = \sum_{i=1}^{m} n_i$. Then $-2\log_e \lambda \sim \chi^2_{m-1}$ and the conclusions can be drawn accordingly.

Breusch Pagan Test:

This test can be applied when the replicated data is not available but only single observations are available. The null-hypothesis stated as H_0: $\sigma_1^{\ 2} = \sigma_2^{\ 2} = ---- = \sigma_n^{\ 2}$. The test procedure is as follows:

Step-1: Ignoring heteroskedastacity, apply OLS to $Y_i = \beta_1 + \beta_2 X_{i1} + ----+\beta_k X_{ik} + \varepsilon_i$ and obtain residual e=Y-X b, then $b = (X'X)^{-1}X'Y$.

Step-2: Construct the variables $g_i = e_i^2 \Big/ \left(\sum_{i=1}^{n} e_i^2 / n\right) = \frac{ne_i^2}{SS_{res}}$, where SS_{res} is the residual sum of squares based on e_i's.

Step-3: Run regression of 'g' on $Z_1, Z_2, ---, Z_n$ and get residual sum of squares $SS_{res}^{\ *}$.④. For testing, calculate $Q = \frac{1}{2}\left(\sum_{i=1}^{n} g_i^2 - SS_{res}^{\ *}\right)$ which is asymptotically distributed as $\chi^2_{(1-p)}$ and conclusions can be drawn accordingly.

Note: where $Z_i' = (1, Z_i^*)(1, Z_{i2}, Z_{i3}, ---, Z_{ip})$, $\sigma_i^{\ 2} = \gamma_1 + \gamma_2 Z_{i2} + ----+\gamma_p Z_{ip}$ and $\gamma' = (\gamma_1, \gamma_i^*) = (\gamma_1, \gamma_2, ---, \gamma_p)$.

Glesjer Test:

This test is based on the assumption that $\sigma_i^{\ 2}$ is influenced by one variable Z, i.e., there is only one variable which is influencing the heteroskrdastacity. This variable could be either one of the explanatory variable or it can be chosen from extraneous source also.

The test procedure is as follows:

Step-1: Calculate the residual 'e' by using OLS on the basis of available study and explanatory variables.

Step-2: Chose Z and apply OLS to $|e_i| = \delta_0 + \delta_1 Z_i^h + \varepsilon_i$ where ε_i be the disturbance term.

Step-3: Test H_0: δ_1 =0 using t-ratio test statistic.

Step-4: Conduct test for h=± 1 or h=$\pm \frac{1}{2}$. So the test procedure is repeated four times. In practice, one can chose any value of h, we chose h=1.

Spearman's Rank Correlation Test:

This test is based on the spearman's rank correlation coefficient between the absolute values of the residuals and X variable with which $E[\varepsilon_i^2] = \sigma_i^{\ 2}$ might be associated.By this test, we first regress Y on X as $y = \beta_0 + \beta_1 x_1 + \varepsilon$ and then find absolute residuals e's as $e = Y - \hat{Y}$. Compute the Spearman's Rank Correlation coefficient between residuals and X values and let it be ρ . A high value of ρ gives the presence of heteroskedastacity.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 13, July 2017

White test:

White test is a large sample Lagrange Multiplier test with a particular choice for the Z's, but it does not depends on the normality assumption.

The null-hypothesisH_0: $\sigma_i^2 = \sigma^2$, $\forall i$ and the alternative is that the null-hypothesis is false.

The motivation for the test is that if the null-hypothesis is true $S^2(X'X)^{-1}$ and $S^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$ both consistent estimators of $\text{var}(\beta)$, while if the null-hypothesis is false, the two estimators will diverge. *The procedure is:*

Regress e_i^2 on all the crosses and squares of X. The test statistic is $W = nR^2 \sim \chi^2_{(p-1)}$, where p is the total number of regressors, including the constant.

The advantage of White's test for heteroskedastacity is that you do not need to know the specific form of Ω.

Lagrange Multiplier(LM) Test:

$y_i = \beta_1 + \beta_2 X_{i2} + ---+ \beta_k X_{ik} + \varepsilon_i$, we are interested to test if assumption of heteroskedastacity is true; therefore we are to test the null-hypothesis $H_0 : \text{var}\left(\varepsilon_i / X_i\right) = \sigma^2$.

A simple approach is to assume a linear function for σ_i^2 is $\sigma_i^2 = \alpha_1 + \alpha_2 Z_{i2} + ----+\alpha_p Z_{ip}$, where the error variance σ_i^2 is related to a number of variables $Z_2, ---, Z_p$.

Then the null-hypothesis can be written as $H_0 : \alpha_2 = --- = \alpha_p = 0$.

The F-test is given by

Regress y against a constant term, $x_1, x_2, ---, x_k$ and obtain the estimated residual e_i for i=1, 2, ---, n. Regress e_i against a constant term, $Z_2, ----, Z_p$ and obtain OLS estimators $(\hat{\alpha}_1, \hat{\alpha}_2, ----, \hat{\alpha}_p)$. Denote the corresponding R-squared as $R^2_{e^2}$.

$$F = \frac{R^2_{e^2} / (p-1)}{1 - R^2_{e^2} / (n-k)} \sim F_{(p-1, n-k)}.$$

LM-test statistic for testing $\alpha_2 = --- = \alpha_p = 0$ is given by $LM = nR^2_{e^2} \sim \chi^2_{(p-1)}$.

Linear Probability Model (LPM)

$$y = \beta_1 + \beta_2 X_2 + ---+ \beta_k X_k + \varepsilon_i$$

$$E(y/X) \equiv E[y/X_2, ---, X_k] = \beta_1 + \beta_2 X_2 + ---+ \beta_k X_k$$

Where 'y' is a binary variable.

It results $P[y = 1/X_2, ---, X_k] = \beta_1 + \beta_2 X_2 + ---+ \beta_k X_k$.

More important, the linear probability model does violate the assumption of homoskedastacity.

When y is a binary variable, we have $\text{var}(y/X) = P(X)[1 - P(X)]$, where P(X) denotes the probability of success:

$$P[X] = \beta_1 + \beta_2 X_2 + ---+ \beta_k X_k.$$

This indicates that there exists heteroskedastacity in LPM. It implies the OLS estimators are inefficient in LPM.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 13, July 2017

Maximum Likelihood Estimation (MLE) in LPM:

Use MLE method to estimate β , the collection of $\beta_0, \beta_1, - - -, \beta_k, \sigma^2$ and $\Omega(\theta)$, the variance matrix of parameters, at same time.

Let Γ estimate Ω^{-1}. Then, the log likelihood can be written as

$$\log L = -\frac{1}{2}\left[\log(2\pi) + \log\sigma^2\right] - \frac{1}{2\sigma^2}\varepsilon'\Gamma\varepsilon + \frac{1}{2}\log|\Gamma|$$

First order conditions are given by

$$\frac{\partial \log L}{\partial \beta} = \frac{1}{\sigma^2}X'\Gamma(Y - X\beta)$$

$$\frac{\partial \log L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^2}(Y - X\beta)'\Gamma(Y - X\beta)$$

$$= \frac{1}{2}\left[\Gamma^{-1} - (\frac{1}{\sigma^2})\varepsilon\varepsilon'\right] = \frac{1}{2\sigma^2}\left(\sigma^2\Omega - \varepsilon\varepsilon'\right)$$

REFERENCES

1. Goldfield, S.M. and Quqndt, R.E., (1965), Some tests for homoskedastacity, Journal of American statistical Association, Vol. 60, No. 310, 540-541.
2. Amemiya, T. (1977), A Note on a Heteroscedastastic Model, Journal 0f Econometrica, 6, 365-370.
3. White, H. (1980), A heteroseadastacity-consistent covariance matrix estimator and direct test for heteroskedastacity, Econometrica, Vol. 48, No. 4, 821-827.
4. Geweke, K. (1981), The Approximate Slopes of Econometric Tests, Econometrica, 49, 1427-1442.
5. Glejser, H. (1969), A new test for heteroskedastacity, Journal of American Statistical Association, Vol. 64, 325, 361.
6. Breusch, T.S. and Pagan, A.R. (1979), A simple test for heteroskedastacity and random coefficient variation, Econometrica, Vol. 47, No.5, 1287-1294.
7. Theil H., Economic Forecasts and Policy, Second Revised Edition, (Amsterdam: North Holland Publishing Company, 1961).
8. White, H., “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity,” Econometrica 48 (1980), 817-838.