Department of Economics, Patna University, Patna

Name of the Programme: M A Economics (Sem. IV)

Name of the Course: EC- 1 Group C: Basic Econometrics

Module 4: Problems of Single Equation Model

Name of the Topic: Heteroscedasticity

By

Dr. Benoy Kumar Lal

Associate Professor

Department of Economics, Patna University, Patna

Email ID: benoykumarlal@yahoo.in

Mobile No. / WhatsApp No. 9608570960/ 9852059001

Heteroscedasticity

As we know that the Regression Model is based on some assumptions about the disturbance term. These are:

1. **Normality**: The first assumption of the Regression Model is that the disturbance term is normally distributed.
2. **Zero Mean:** The second assumption of the Regression Model is that the disturbance term/ error term has zero mean. i.e.

   $$E(u_i^2) = 0 \qquad \text{where, } i = 1, 2, 3, \ldots\ldots\ldots\ldots\, n.$$

3. **Homoscedasticity: It** assumes that the errors are distributed independently with Zero Mean and constant variance σ_u^2, **that is**

   $$E(\sigma_i^2) = \sigma_u^2 \qquad \text{where, } i = 1, 2, 3, \ldots\ldots\ldots\ldots\, n.$$

4. **Non- autoregression: Assumption of non-autoregression states that various disturbance terms are uncorrelated,** i.e.

   $$E(u_i, u_j) = 0 \qquad \text{where, } i \neq j \text{ and } i \,\&\, j = 1, 2, 3, \ldots\ldots\ldots\ldots\, n.$$

5. **Non-stochastic X: It** assumes that X_i is a non-stochastic variable with fixed values in repeated samples and such that for any size

   $$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

   is a finite value different from zero.

PTO

(b) **Definition:** In many econometric studies, the assumption of constant variance (Homoscedasticity) may not always be valid. It may happen that the errors are mutually uncorrelated but at the same time have different variance. This situation. where variance changes with independent variable is known as Problem of Heteroscedasticity. i.e.

If σ_u^2 is not constant; disturbance will be heteroscedastic

In general, heteroscedasticity occurs in observations, in which the size of economic units being observed over a wide range.

(c) Consequences or Effects of Heteroscedasticity:

If $E(u_i^2) \neq \sigma_u^2$ i.e. problem of heteroscedasticity is present in the model, following shall be the Consequences:

(i) Let the model is $Y_i = \alpha + \beta X_i + u_i$

then $\hat{\beta} = \frac{\Sigma x_i y_i}{\Sigma x_i^2} = \frac{\Sigma x_i (\beta x_i + u_i)}{\Sigma x_i^2}$

[Since $y_i = \beta x_i + u_i$]

or, $\hat{\beta} = \beta \frac{\Sigma x_i^2}{\Sigma x_i^2} + \frac{\Sigma x_i u_i}{\Sigma x_i^2} = \beta + \frac{\Sigma x_i u_i}{\Sigma x_i^2}$

taking expectation on both sides - we have

$$E(\hat{\beta}) = \beta + E\left[\frac{\Sigma x_i u_i}{\Sigma x_i^2}\right] \quad \text{-------} \quad (a)$$

$$= \beta + \frac{\Sigma x_i E(u_i)}{\Sigma x_i^2} = \beta + \frac{\Sigma x_i \times 0}{\Sigma x_i^2}$$

[Since $E(u_i) = 0$]

or, $E(\hat{\beta}) = \beta$

Thus it is clear that least square estimators are still unbiased (in the presence of heteroscedasticity).

(ii) Now, we shall test the effect of heteroscedasticity on efficiency of the estimators.

P.T.O.

Since, $\text{Var.}(\hat{\beta}) = E[\beta - E(\hat{\beta})]^2$

$$= E\left[\frac{\Sigma x_i u_i}{\Sigma x_i^2}\right]^2 \qquad \text{[from eq}^n \text{ (a)]}$$

$$= \frac{1}{(\Sigma x_i^2)^2} E\left[x_1^2 u_1^2 + x_2^2 u_2^2 + \cdots + 2x_1 x_2 u_1 u_2 + 2x_1 x_3 u_1 u_3 + \cdots\right]$$

$$= \frac{1}{(\Sigma x_i^2)^2}\left[x_1^2 E(u_1^2) + x_2^2 E(u_2^2) + \cdots + 2x_1 x_2 E(u_1 u_2) + 2x_1 x_3 E(u_1 u_3) + \cdots\right]$$

Let us suppose,

$$E(u_i^2) = k_i \sigma_u^2$$

then $\text{var.}(\hat{\beta}) = \frac{1}{(\Sigma x_i^2)^2}\left[x_1^2 k_1 \sigma_u^2 + x_2^2 k_2 \sigma_u^2 + \cdots\right]$

[Since $E(u_i u_j) = 0,\ i \neq j$]

$$= \frac{\sigma_u^2}{(\Sigma x_i^2)^2} \Sigma k_i x_i^2 = \frac{\sigma_u^2}{\Sigma x_i^2} \cdot \frac{\Sigma k_i x_i^2}{\Sigma x_i^2} \quad \text{—— (b)}$$

Thus, the variance is different. If x_i^2 and k_i are positively correlated then $\text{var}(\hat{\beta})$ under heteroscedasticity will be larger than variance under homoscedasticity and usual OLS (Ordinary Least Squares) formula will under estimate the true variance. Therefore, if disturbance term (u_i) is heteroscedastic, the variance of the estimates are inefficient.

(d) Tests for Heteroscedasticity:

Various methods/tests are used to detect the presence of heteroscedasticity in the model. These are (i) Spearman's Rank-correlation test (ii) Goldfeld-Quandt Test (iii) Glesjer test and (iv) Park test.

(I) <u>The Spearman's Rank Correlation Test</u>: This is applicable for both small and large samples.

<u>Practical Steps</u>.

* Regress Y on X and obtain the residuals e_i where $|e_i| = |Y - \hat{Y}|$

* Calculate Rank correlation co-efficient between X and $|e_i|$ from the given formula $\gamma_{|e|x} = 1 - \frac{6 \Sigma D^2}{n(n^2-1)}$

Where D is the rank difference.

* Use t-statistic to detect the presence of heteroscedasticity using the following formula

$$t = \frac{\gamma \sqrt{n-1}}{\sqrt{1-\gamma^2}}$$

[<u>Reference</u>: Page-152 of 'Econometrics' by Shyamala, Kaur and Pragasam for Numerical Example]

(II) <u>Goldfeld-Quandt Test</u>: This method is used for large samples.

<u>Practical Steps</u>.

1. Order the observations according to the magnitude of X (independent variable)

2. Select a certain number (one-third approx.) of central observations (c) which we omit from the analysis. The remaining $n-c$ observations are divided into two sub-groups of equal size i.e. $\frac{n-c}{2}$ where one subgroup includes small values of x and the other sub-group includes the large values of X.

(3) Now fit the separate regression to each of the sub-group, and obtain the sum of squared residuals from each of them (Σe_1^2 & Σe_2^2)

(4) Find value of F-statistic from the formula

$$F = \frac{\Sigma e_2^2}{\Sigma e_1^2}$$

[Where, Σe_1^2 denote the sum of squared residuals from the sample of small values of X_i and Σe_2^2 denotes the same for large values of X_i.

(5) Finally use the F-table at $\left(\frac{n-c}{2} - k\right)$ d.f.

(Reference: Page-154 of 'Econometrics' by Shyamala, Kaur and Pragasam. for Numerical Examples.]

(III) <u>The Park Test</u>: Prof. R.E. Park developed a specific functional form between the $\sigma_{u_i}^2$ and the explanatory variable to detect heteroscedasticity.

$$\sigma_{u_i}^2 = f(X_i) = \sigma^2 X^{\beta} e^{V_i}$$

taking log on both sides

$$\log \sigma_{u_i}^2 = \log \sigma^2 + \beta \log x + V_i \log_e e$$

$$\log \sigma_{u_i}^2 = \log \sigma^2 + \beta \log x + V_i$$

Since $\sigma_{u_i}^2$ is unobservable, Prof. Park suggested e_i^2 as a proxy for $\sigma_{u_i}^2$

$$\therefore \boxed{\log e_i^2 = \log \sigma_u^2 + \beta \log X_i}$$

<u>Steps</u>: * Run OLS and obtain e_i for the original data

* Run the log-linear regression between e_i^2 and X_i and examine whether β is significant

(IV): Glejser Test: It is similar to Park Test but suggested many functional forms instead of one as suggested by Park.

Steps:

* Regress Y on all explanatory variables and compute the residuals e_i's
* Regress the absolute values of e_i's ($|e_i|$) on explanatory variable with which σ_{ui}^2 is associated.

Since the actual form of this regression is not known, Glejser used the following functional forms to detect the presence of heteroscedasticity.

$$|e_i| = \beta_0 + \beta_1 X_i^2 + V_i$$

or, $$|e_i| = \beta_0 + \beta_1 X_i^{-1} + V_i$$

or, $$|e_i| = \beta_0 + \beta_1 \sqrt{x_i} + V_i$$

or, $$|e_i| = \beta_0 + \beta_1 X_i + V_i$$ .. and so on.

From the above models, we select that regression model which gives the best fit in the light of correlation coefficient and standard errors of coefficients β_0 and β_1. If β in the selected model turns out to be significant, it would suggest presence of heteroscedasticity.

If $\beta_0 = 0$ while $\beta_1 \neq 0 \rightarrow$ pure heteroscedasticity

If both β_0 & $\beta_1 \neq 0 \rightarrow$ mixed heteroscedasticity.

Questions - Short Answer type:

(i) Define heteroscedasticity.

(ii) Write short note on Goldfeld-Quandt test.

P.T.O

Long Answer type Questions:

(i) Describe the consequences of heteroscedasticity on OLS estimator.

(ii) Explain the tests used to detect the presence of heteroscedasticity.

Suggested Readings

1. Introduction to Econometrics by GMK Madnani
2. Econometrics: Theory and Applications. by Shyamala, Kaur and Prasasam
3. Econometrics & Mathematical Economics by Singh, Parashar & Singh
4. Econometrics by - K. Dhanasekaran.

by.

Dr. Benoy Kumar Lal
Associate Professor
Department of Economics
Patna University, Patna

Mob. - 9608570960, 9852059001
E Mail - bkl165@gmail.com
benoykumarlal@yahoo.in.