HETEROSCEDASTICITY AND ITS DETECTION

By:
Dr. Akash Asthana
Assistant Professor,
Dept. of Statistics,
University of Lucknow, Lucknow

HETEOSCEDASTICITY

- In the linear regression model, it is assumed that the variance of error terms should be constant and also independent of each other.

  i.e. $V(\epsilon) = \sigma_\epsilon^2 I \text{ or } V(\epsilon_i) = \sigma_\epsilon^2 \ \& \ Cov(\epsilon_i, \epsilon_j) = 0$
- If this assumption is not fulfilled then heteroscedasticity is said to be present.
- In presence of heteroscedasticity the estimates of regression coefficients not remain BLUE.

DETECTION OF HETEROSCEDASTICITY

- There are several methods used for the detection of heteroscedasticity among which most commonly used methods are:

1. Graphical Method
2. Park’s test
3. Glejser Test
4. Speaeman’s rank correlation test
5. Goldfeld – Quandt Test
6. Breusch – Pagan – Godfrey Test

RESIDUALS

- Residuals (r_i) are nothing but the difference between the observed and predicted value of the dependent variable.
- It can be considered as estimate of the error term in the regression model.

$$r_i \text{ (or } \hat{U}_i) = Y_i - \hat{Y}_i$$

- Residuals are of following types:

1. Simple Residual
2. Standardized Residual
3. Studentized Residual

SIMPLE RESIDUAL

- The simple residual (r_i) is obtained by the difference between the observed and predicted value of the dependent variable.

STANDARDIZED RESIDUAL

- The simple residual has a drawback that its distribution is not Normal therefore standardized residuals are used these are given by: $$r_i' = \frac{r_i}{\hat{\sigma}_e}$$

STUDENTIZED RESIDUAL

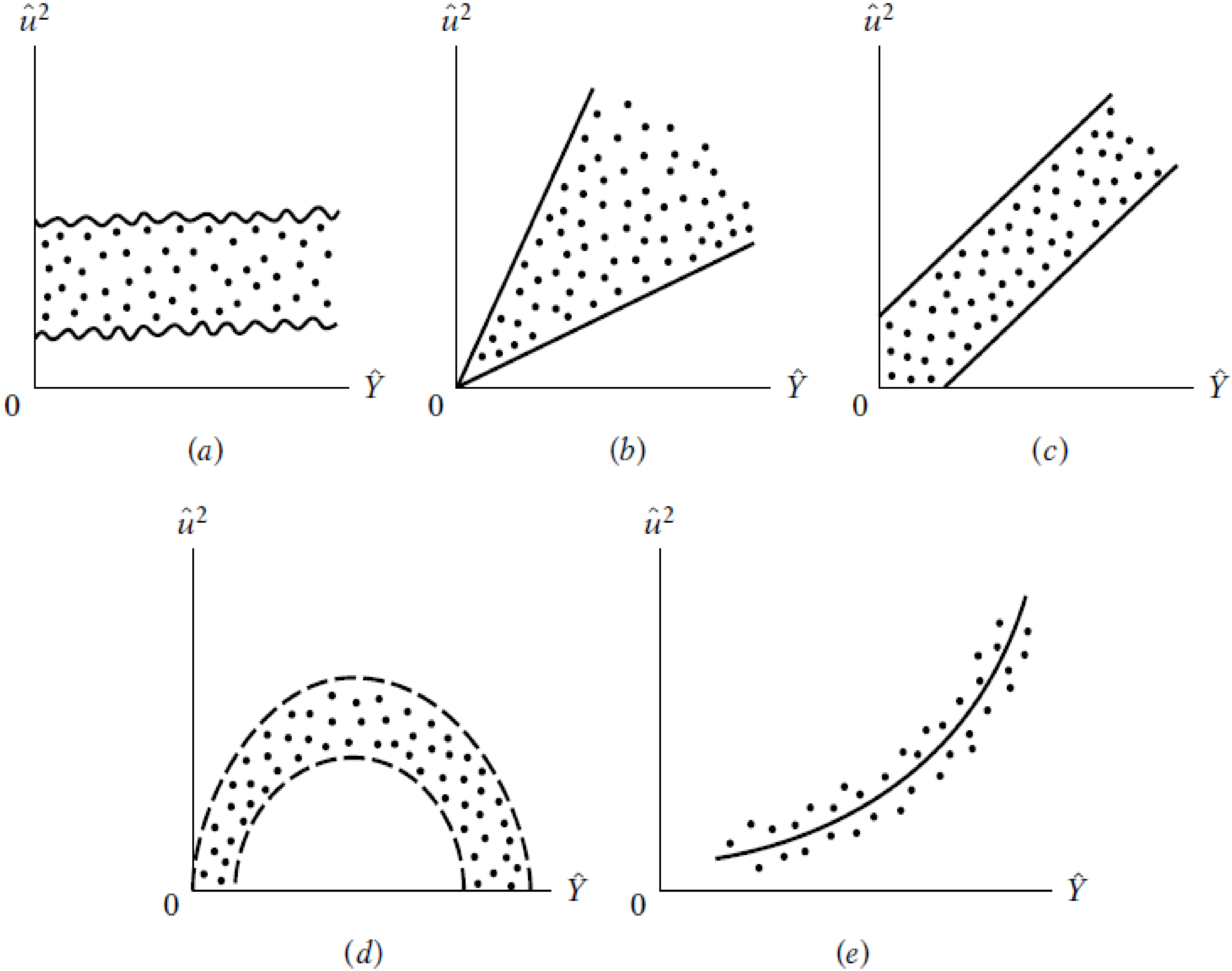
The standardized residual has a drawback that its values are unbounded. To overcome this difficulty the standardized residuals are used these are given by: $$r_i'' = \frac{r_i}{\hat{\sigma}_e\sqrt{1-h_{ii}}}$$

Where hii is the ith diagonal element of hat matrix (H) = $X(X'X)^{-1}X'$

It follows the student's t-distribution.

GRAPHICAL METHOD

- In this method Residual square (r_i^2 or $r_i^{2'}$)is plotted against the predicted value of the dependent variable. If the plot doesn't show any pattern then heteroscedasticity is said to be absent otherwise it is said to be absent.
- In the figures on next slide first figure (a) shows case of homoscedastic data whereas other figures (b, c, d & e) are the examples of heteroscedastic data [1].


$\hat{u}^2$
$\hat{Y}$
0
(a)
(b)
(c)
(d)
(e)

PARK'S TEST

Park had modeled the error variance as a function of explanatory variables defined as:

$$\sigma_i^2 = \sigma^2 X_i^{\beta} e^{\nu_i}$$

or,

$$\log_e \sigma_i^2 = \log_e \sigma^2 + \beta \log_e X_i + \nu_i$$

Where ν_i is homoscedastic error term.

However, since σ_i^2 is unknown Park had been suggested the use of r_i^2 in its place. If β comes out to be significant then heteroscedasticity is said to be present in the data.

GLEJSER TEST

- The Glejser test is similar to Park's test. Instead of one Glejser had used different functional forms to model error variance (or its estimate) over explanatory variables. If any of the model comes out to be significant then heteroscedasticity is said to be present.
- The functional forms used by Glejser are on next slide.

GLEJSER TEST

$$|r_i| = \beta_1 + \beta_2 X_i + v_i$$

$$|r_i| = \beta_1 + \beta_2 \sqrt{X_i} + v_i$$

$$|r_i| = \beta_1 + \beta_2 \frac{1}{X_i} + v_i$$

$$|r_i| = \beta_1 + \beta_2 \frac{1}{\sqrt{X_i}} + v_i$$

$$|r_i| = \sqrt{\beta_1 + \beta_2 X_i} + v_i$$

$$|r_i| = \sqrt{\beta_1 + \beta_2 X_i^2} + v_i$$

However this test has a drawback that in last two models parameters are nonlinear and therefore can’t be estimated with the help of OLS method. Also these two method considered the term νi as white noise, which is not necessary according to Goldfled and Quandt i.e. it may be heteroscedastic itself.

SPEARMAN'S RANK CORRELATION TEST

- It has following steps:

1. Fit the regression of Y on X and otain the residuals.
2. Compute the Spearman's rank correlation between absolute value of residuals and Xi (or $\hat{Y}_i$)
3. Test the null hypothesis that population correlation coefficient is zero using t-test. If the hypothesis is rejected then heteroscedasticity is said to be present. The t-statistics is given by:

$$t = \frac{r_s\sqrt{(n-2)}}{\sqrt{(1-r_s^2)}}$$

r_s = Spearman's rankcorrelation coefficient.

GOLDFLED – QUANDT TEST

- This method is applicable only if the heteroscedastic variance (σ_i^2) is positively related with one of the explanatory variable.
- In this method it is assumed that σ_i^2 is proportional to the square of the explanatory variable.
- Goldfled and Quandt had suggested a number of steps to detect the heteroscedasticity which are on next slide.

GOLDFLED – QUANDT TEST

1. Order (or arrange) the observations in the ascending order of values of X.
2. Omit c central values (c is a specified a priori) and divide the remaining c central values into two equal halves having (n-c)/2 observations.
3. Fit separate OLS regression for both the halves and compute residual sum of squares for each (say RSS_1 and RSS_2) having (n-c-2k)/2 degrees of freedom. (k is no. of parameters estimated.)
4. Compute the ratio $\lambda = RSS_1/RSS_2$.
5. Heteroscedasticity is said to present if :

$$\lambda > F(n-c-2k)/2,\ (n-c-2k)/2,\ \alpha$$

BREUSCH–PAGAN–GODFREY TEST

- The goldfled–Quandt test depend upon the correct selection of the value of c and the correct explanatory variable according to which observations are to be arranged.
- To overcome this difficulty Breusch–Pagan–Godfrey defined another test.
- This test has a number of steps which are on next slide.

BREUSCH–PAGAN–GODFREY TEST

1. Fit the regression model Y = Xβ + ϵ using OLS method and obtain the residuals $r_1, r_2, \ldots, r_n$.
2. Obtain the estimate of σ2 using: $\tilde{\sigma}^2 = \sum_{i=1}^{n} r_i^2 / n$
3. Construct the variable p_i using: $p_i = r_i^2 / \tilde{\sigma}^2$
4. Regress the p_i over the Z_j's. Some or all X_j's may serve as Z_j's.

   $$p_i = \alpha_0 + \alpha_1 Z_{1i} + \alpha_2 Z_{2i} + \cdots + \alpha_m Z_{mi} + \nu_i$$

   Where νi is the homoscedastic error term.
5. Obtain the Explained (Regression) Sum of Square (ESS) and define: Θ = ESS/2
6. If Θ exceeds the $\chi^2_{(m-1)}$ at given level of significance then heteroscedasticity is said to be present.

REFRENCES

1. Gujrati DN, Basic Econometrics, 4th edition (2004) , The McGraw-Hill Companies.
2. Draper NR & Smith H, Applied Regression Analysis, 3rd edition (1998), John Wiley & Sons Inc.
3. Johnston J & Dinardo J, Econometric Methods, 4th edition (1997), McGraw-Hill Companies.