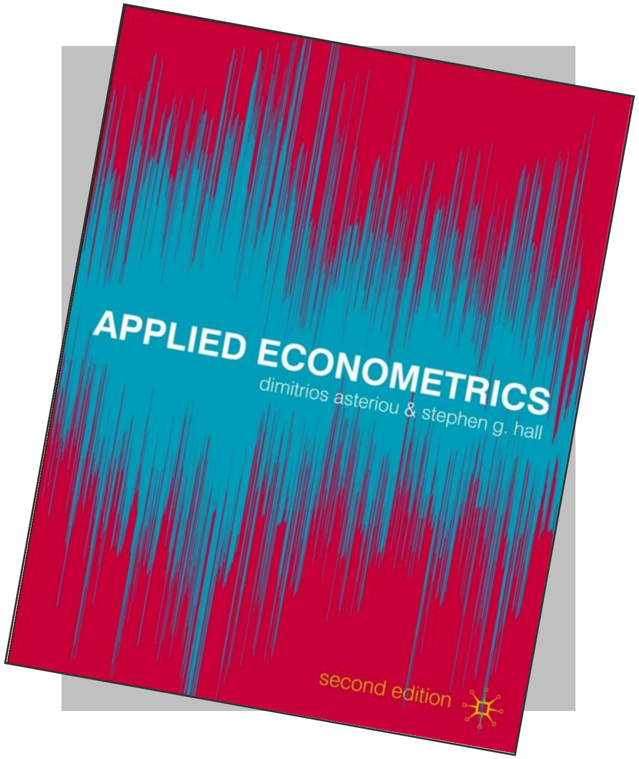
Applied Econometrics
Second edition

Dimitrios Asteriou and
Stephen G. Hall



palgrave
macmillan

HETEROSKEDASTICITY

1. What is Heteroskedasticity
2. Consequences of Heteroskedasticity
3. Detecting Heteroskedasticity
4. Resolving Heteroskedasticity

Learning Objectives

1. Understand the meaning of heteroskedasticity and homoskedasticity through examples.
2. Understand the consequences of heteroskedasticity on OLS estimates.
3. Detect heteroskedasticity through graph inspection.
4. Detect heteroskedasticity through formal econometric tests.
5. Distinguish among the wide range of available tests for detecting heteroskedasticity.
6. Perform heteroskedasticity tests using econometric software.
7. Resolve heteroskedasticity using econometric software.

What is Heteroskedasticity

Hetero (different or unequal) is the opposite of Homo (same or equal)…

Skedastic means spread or scatter…

Homoskedasticity = equal spread

Heteroskedasticity = unequal spread

What is Heteroskedasticity

Assumption 5 of the CLRM states that the disturbances should have a constant (equal) variance independent of *t*:

$$Var(u_t)=\sigma^2$$

Therefore, having an equal variance means that the disturbances are homoskedastic.

What is Heteroskedasticity

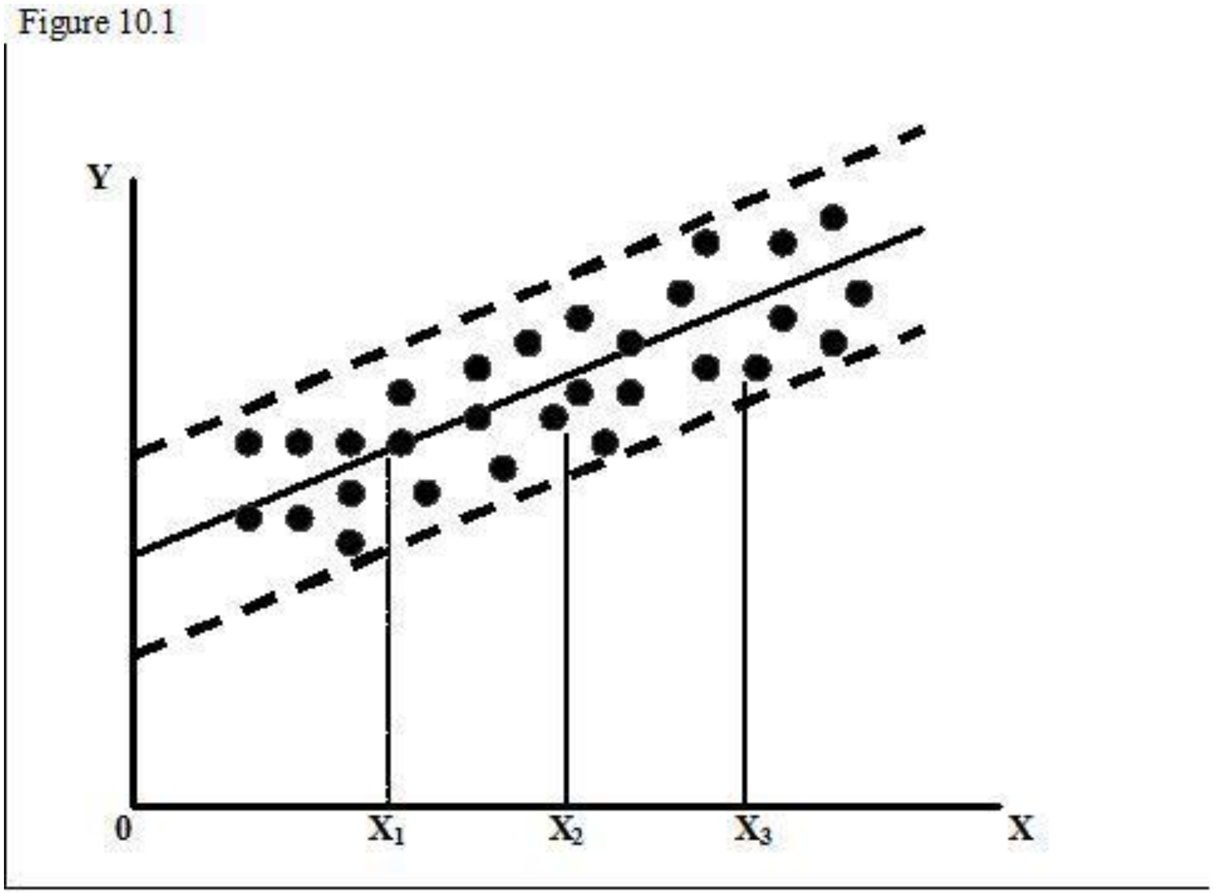
If the homoskedasticity assumption is violated then

$$Var(u_t)=\sigma_t^2$$

Where the only difference is the subscript *t*, attached to the σ_t^2, which means that the variance can change for every different observation in the sample $t=1, 2, 3, 4, \ldots, n$.

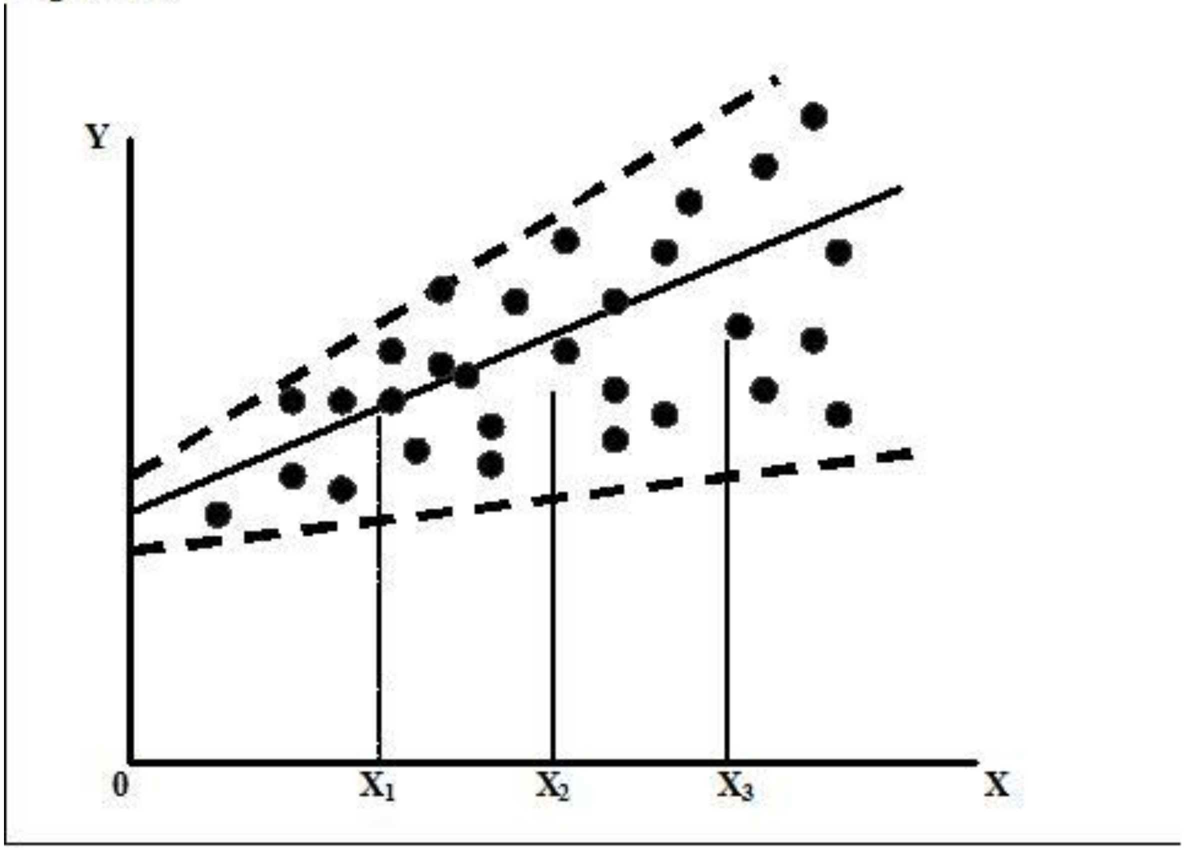
Look at the following graphs…

What is Heteroskedasticity

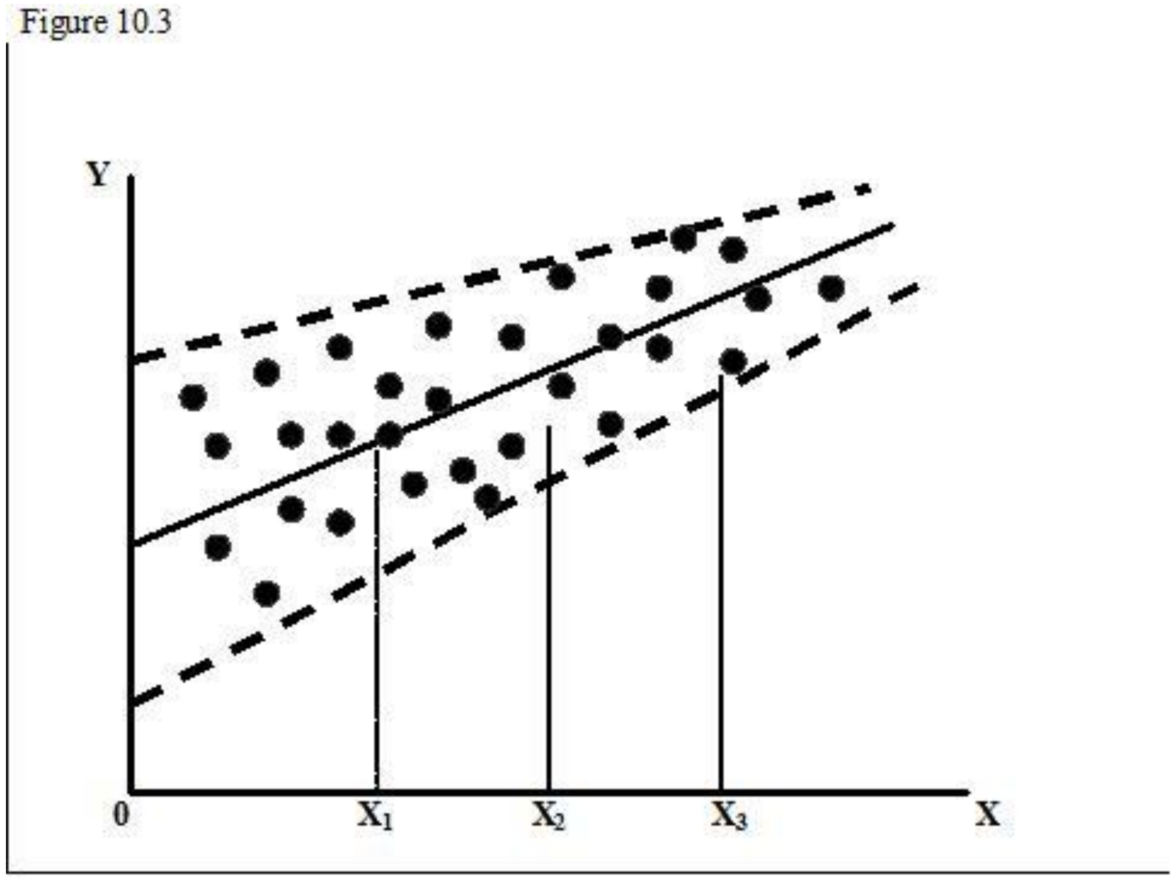


What is Heteroskedasticity

Figure 10.2



What is Heteroskedasticity



What is Heteroskedasticity

First graph: Homoskedastic residuals

Second graph: income-consumption patterns, for low levels of income not much choices, opposite for high levels.

Third graph: improvements in data collection techniques (large banks) or to error learning models (experience decreases the chance of making large errors).

Consequences of Heteroskedasticity

1. The OLS estimators are still unbiased and consistent. This is because none of the explanatory variables is correlated with the error term. So a correctly specified equation will give us values of estimated coefficient which are very close to the real parameters.

2. Affects the distribution of the estimated coefficients increasing the variances of the distributions and therefore making the OLS estimators inefficient.

3. Underestimates the variances of the estimators, leading to higher values of *t* and *F* statistics.

Detecting Heteroskedasticity

There are two ways in general.

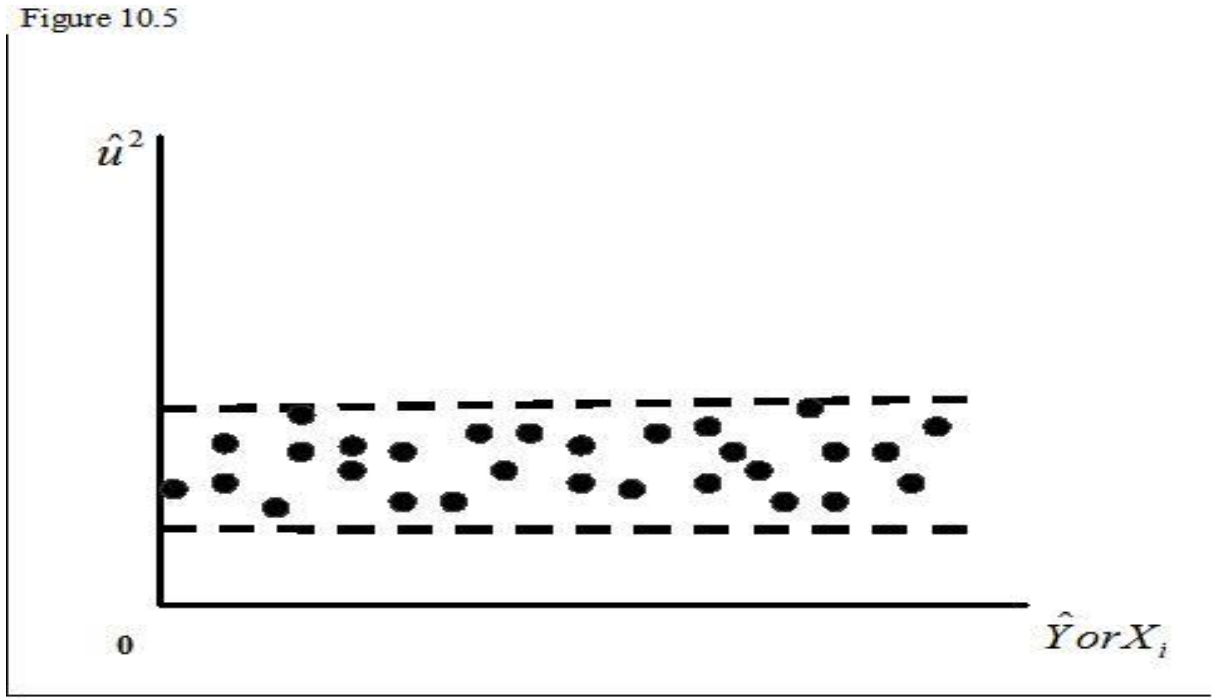
The first is the informal way which is done through graphs and therefore we call it the **graphical method**.

The second is through **formal tests** for heteroskedasticity, like the following ones:

1. The Breusch-Pagan LM Test
2. The Glesjer LM Test
3. The Harvey-Godfrey LM Test
4. The Park LM Test
5. The Goldfeld-Quandt Tets
6. White's Test

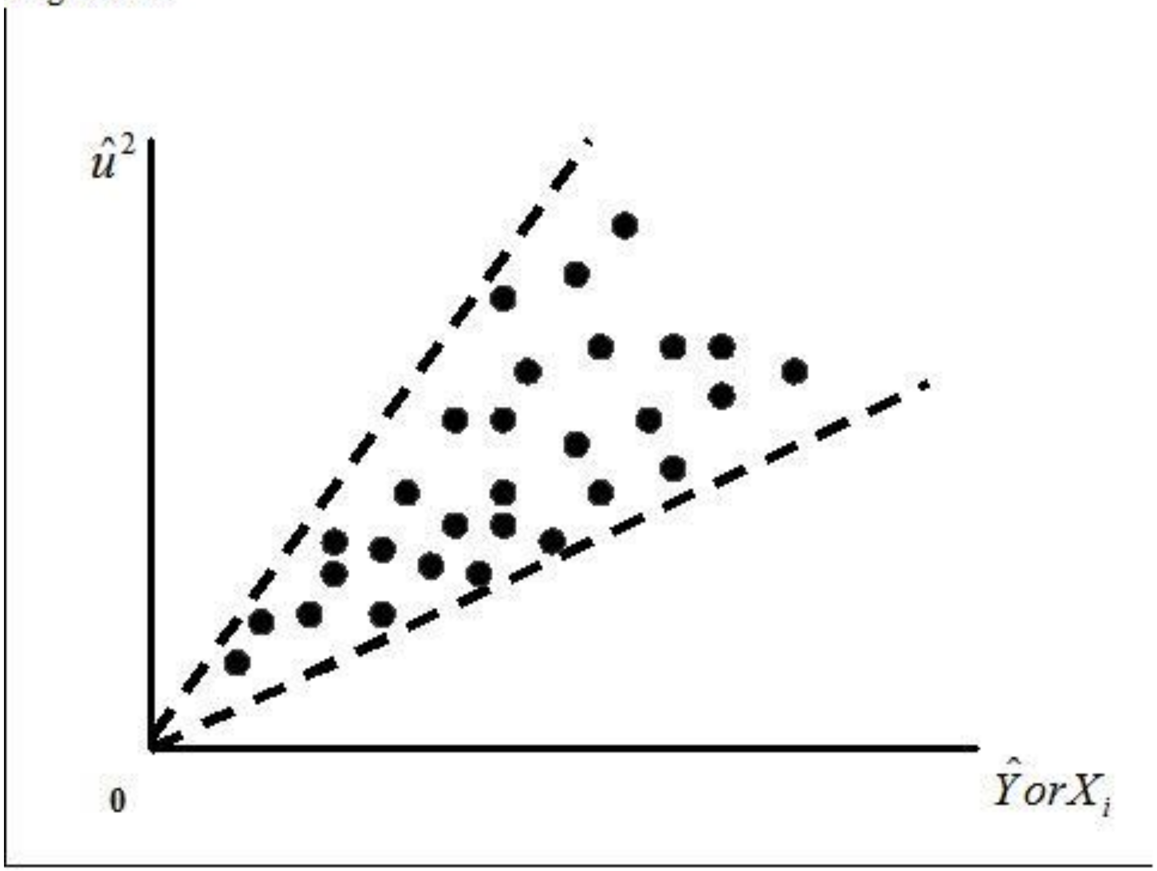
Detecting Heteroskedasticity

We plot the square of the obtained residuals against fitted Y and the X's and we see the patterns.



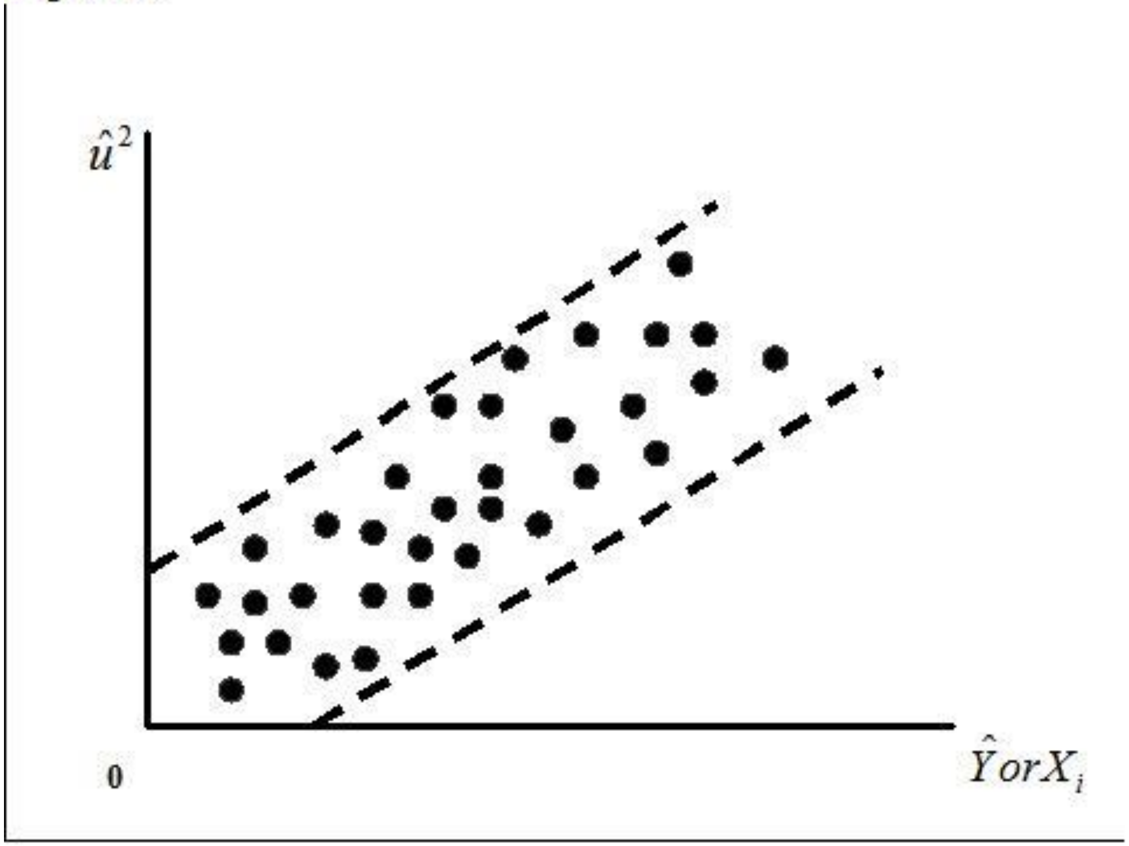
Detecting Heteroskedasticity

Figure 10.6



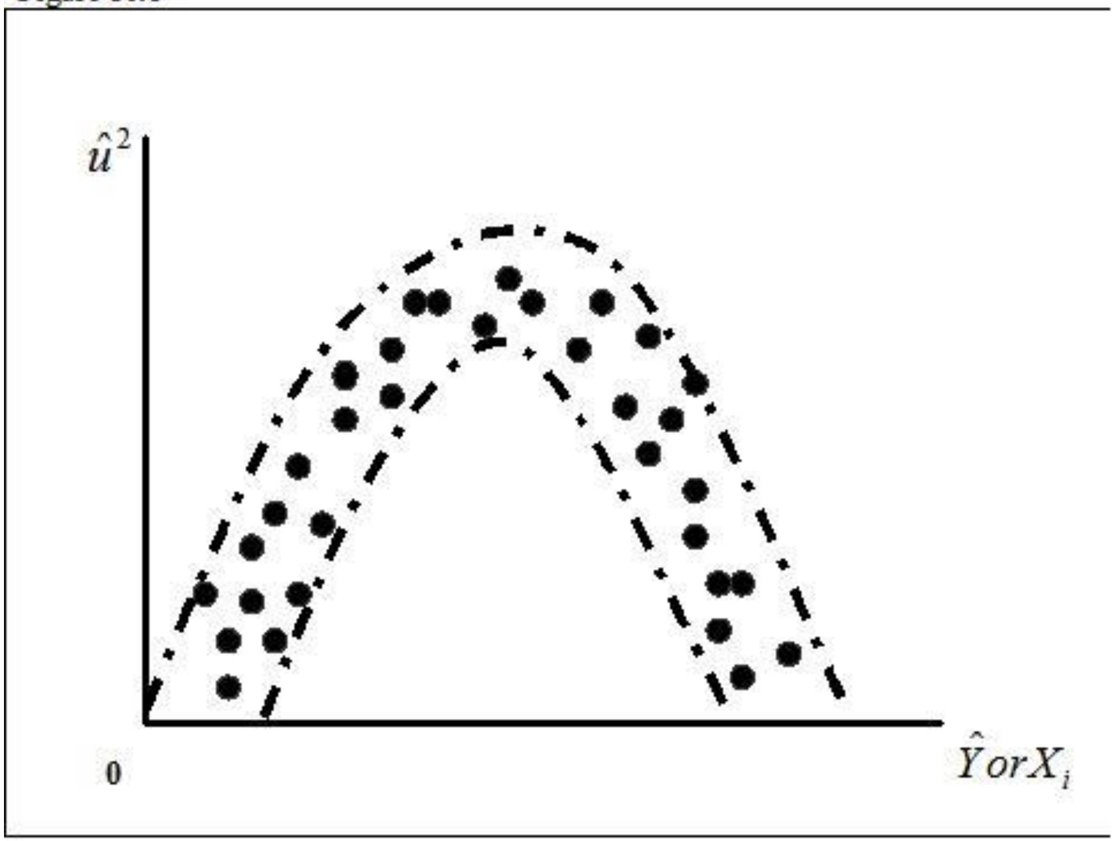
Detecting Heteroskedasticity

Figure 10.7

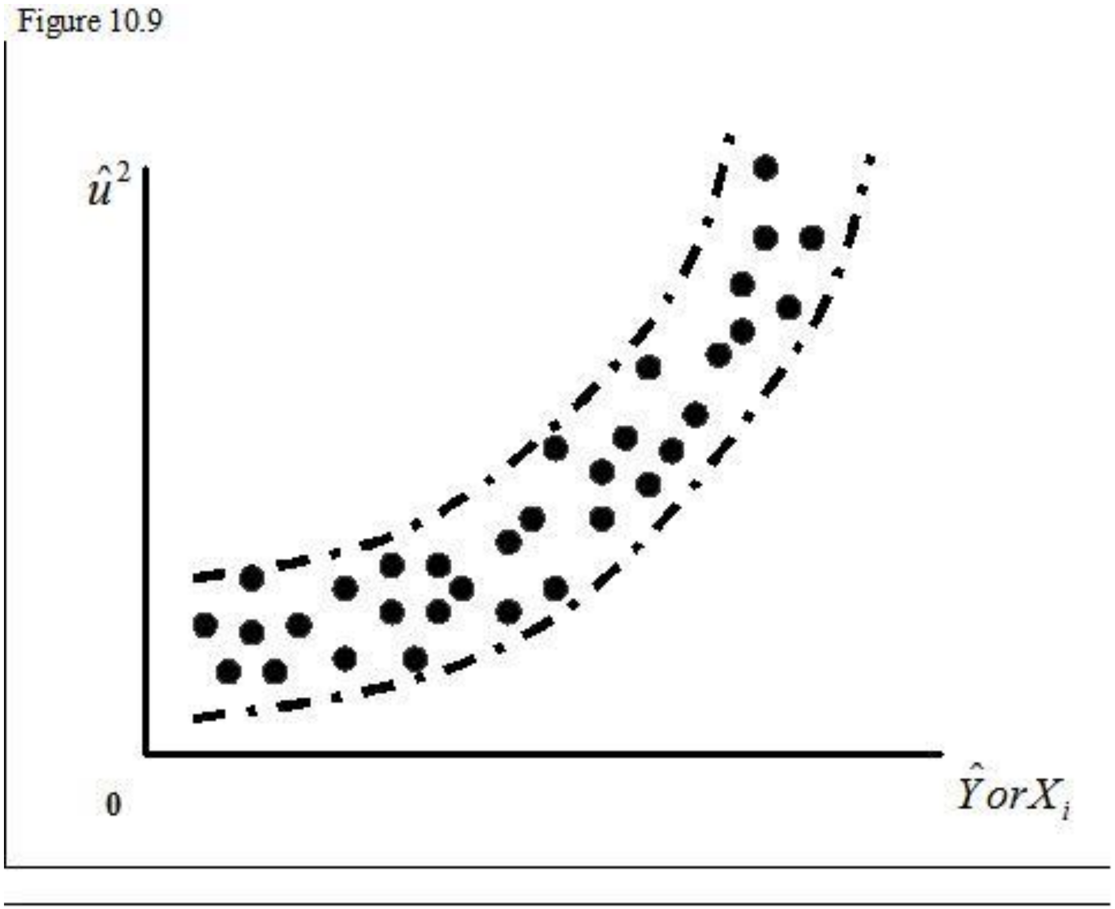


Detecting Heteroskedasticity

Figure 10.8



Detecting Heteroskedasticity



The Breusch-Pagan LM Test

Step 1: Estimate the model by OLS and obtain the residuals

Step 2: Run the following auxiliary regression:

$$\hat{u}_t^2 = a_1 + a_2 Z_{2t} + a_3 Z_{3t} + \ldots + a_p Z_{pt} + v_t$$

Step 3: Compute $LM = nR^2$, where n and R^2 are from the auxiliary regression.

Step 4: If LM-stat>χ^2_{p-1} critical reject the null and conclude that there is significant evidence of heteroskedasticity

The Glesjer LM Test

Step 1: Estimate the model by OLS and obtain the residuals

Step 2: Run the following auxiliary regression:

$$|\hat{u}_t| = a_1 + a_2 Z_{2t} + a_3 Z_{3t} + \ldots + a_p Z_{pt} + v_t$$

Step 3: Compute $LM=nR^2$, where n and R^2 are from the auxiliary regression.

Step 4: If LM-stat>χ^2_{p-1} critical reject the null and conclude that there is significant evidence of heteroskedasticity

The Harvey-Godfrey LM Test

Step 1: Estimate the model by OLS and obtain the residuals

Step 2: Run the following auxiliary regression:

$$\ln \hat{u}_t^2 = a_1 + a_2 Z_{2t} + a_3 Z_{3t} + \ldots + a_p Z_{pt} + v_t$$

Step 3: Compute $LM=nR^2$, where n and R^2 are from the auxiliary regression.

Step 4: If LM-stat>χ^2_{p-1} critical reject the null and conclude that there is significant evidence of heteroskedasticity

The Park LM Test

Step 1: Estimate the model by OLS and obtain the residuals

Step 2: Run the following auxiliary regression:

$$\ln \hat{u}_t^2 = a_1 + a_2 \ln Z_{2t} + a_3 \ln Z_{3t} + \ldots + a_p \ln Z_{pt} + v_t$$

Step 3: Compute $LM=nR^2$, where n and R^2 are from the auxiliary regression.

Step 4: If LM-stat>χ^2_{p-1} critical reject the null and conclude that there is significant evidence of heteroskedasticity

The Engle's ARCH Test

Engle introduced a new concept allowing for hetero-skedasticity to occur in the variance of the error terms, rather than in the error terms themselves.

The key idea is that the variance of u_t depends on the size of the squarred error term lagged one period u^2_{t-1} for the first order model or:

$$Var(u_t)=\gamma_1+\gamma_2 u^2_{t-1}$$

The model can be easily extended for higher orders:

$$Var(u_t)=\gamma_1+\gamma_2 u^2_{t-1}+\ldots+\gamma_p u^2_{t-p}$$

The Engle's ARCH Test

Step 1: Estimate the model by OLS and obtain the residuals

Step 2: Regress the squared residuals to a constant and lagged terms of squared residuals, the number of lags will be determined by the hypothesized order of ARCH effects.

Step 3: Compute the *LM statistic* $= (n\text{-}\rho)R^2$ from the LM model and compare it with the chi-square critical value.

Step 4: Conclude

The Goldfeld-Quandt Test

Step 1: Identify one variable that is closely related to the variance of the disturbances, and order (rank) the observations of this variable in descending order (starting with the highest and going to the lowest).

Step 2: Split the ordered sample into two equally sized sub-samples by omitting c central observations, so that the two samples will contain $\frac{1}{2}(n-c)$ observations.

The Goldfeld-Quandt Test

Step 3:Run and OLS regression of Y on the X variable that you have used in step 1 for each sub-sample and obtain the *RSS* for each equation.

Step 4: Caclulate the $F\text{-}stat = RSS_1/RSS_2$, where RSS_1 is the *RSS* with the largest value.

Step 5: If $F\text{-stat} > F\text{-crit}_{(1/2(n-c)-l,1/2(n-c)-k)}$ reject the null of homoskedasticity.

The White's Test

Step 1: Estimate the model by OLS and obtain the residuals

Step 2: Run the following auxiliary regression:

$$\hat{u}_t^2 = a_1 + a_2 X_{2t} + a_3 X_{3t} + a_4 X_{2t}^2 + a_5 X_{3t}^2 + a_6 X_{2t} X_{3t} + v_t$$

Step 3: Compute $LM = nR^2$, where n and R^2 are from the auxiliary regression.

Step 4: If LM-stat>χ^2_{p-1} critical reject the null and conclude that there is significant evidence of heteroskedasticity

Resolving Heteroskedasticity

We have three different cases:

(a) Generalized Least Squares

(b) Weighted Least Squares

(c) Heteroskedasticity-Consistent Estimation Methods

Generalized Least Squares

Consider the model

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \ldots + \beta_k X_{kt} + u_t$$

where

$$Var(u_t) = \sigma_t^2$$

Generalized Least Squares

If we divide each term by the standard deviation of the error term, σ_t we get:

$$Y_t=\beta_1 (1/\sigma_t) +\beta_2 X_{2t}/\sigma_t +\beta_3 X_{3t}/\sigma_t +\ldots+\beta_k X_{kt}/\sigma_t +u_t/\sigma_t$$

or

$$Y^*_t= \beta^*_1+ \beta^*_2 X^*_{2t}+ \beta^*_3 X^*_{3t}+\ldots+ \beta^*_k X^*_{kt}+u^*_t$$

Where we have now that:

$$Var(u^*_t)=Var(u_t/\sigma_t)=Var(u_t)/\sigma_t^2=1$$

Weighted Least Squares

The GLS procedure is the same as the WLS where we have weights, w_t, adjusting our variables.

Define $w_t=1/\sigma_t$, and rewrite the original model as:

$$w_tY_t=\beta_1 w_t+\beta_2 X_{2t}w_t+\beta_3 X_{3t}w_t+\ldots+\beta_k X_{kt}w_t+u_tw_t$$

Where if we define as $w_tY_{t-1}=Y^*_t$ and $X_{it}w_t=X^*_{it}$

we get

$$Y^*_t=\beta^*_1+\beta^*_2X^*_{2t}+\beta^*_3X^*_{3t}+\ldots+\beta^*_kX^*_{kt}+u^*_t$$