Econometrics 1

Lecture 8

Heteroscedasticity

Causes of Heteroskedasticity

- Learning: reduces errors; driving practice, driving errors and accidents
  typing practice and typing errors, defects in productions;
improved machines
- Growth: saving and variance of saving increases with income
- Improved data collection: better formulas and goods software
- Outliers affect the value of estimates
- Specification Errors and omitted variables:- in a demand model if you regress demand of a product to only its own price, there is a danger variables such as the prices of complements and income may appear in the error term.

More heteroscedasticity exists in cross section than in time series data.

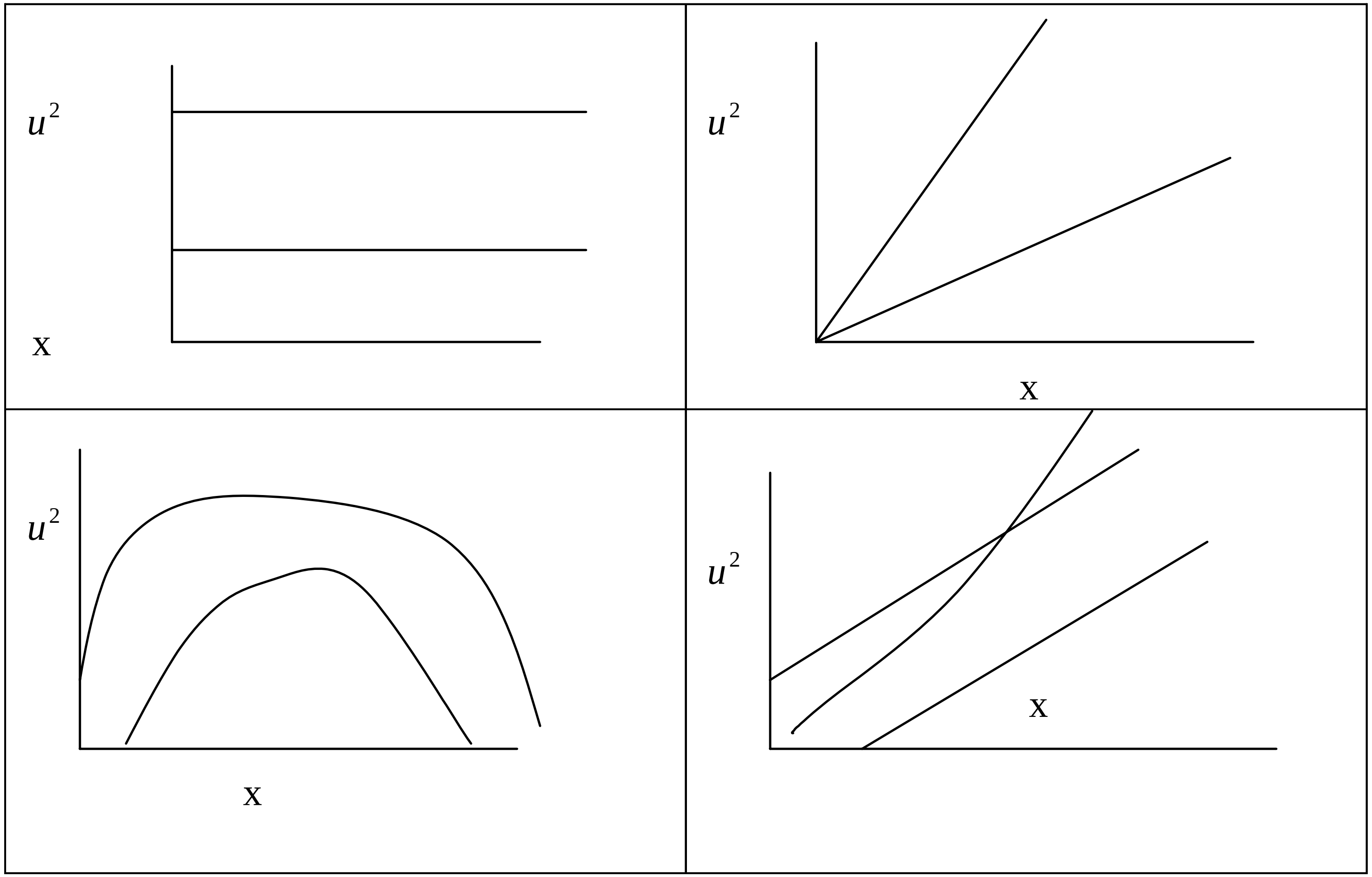
Nature and Causes

- LS assumption: variance of e_i is constant $\text{var}\left[e_i\right]=\sigma^2$ for every ith observation, $\text{var}\left(\hat{\beta}_2\right)=\hat{\sigma}^2\left[\frac{1}{\sum_i\left(x_i-\overline{x}\right)^2}\right]$ but it is possible that

$$\sigma_i^2=\sigma^2 x_i$$

Causes: Learning, growth, improved data collection, outliers, omitted variables;

Detection of Heteroscedasticity: Informal (Graphical) method



Consequence of Heteroscedasticity

OLS estimators give **unbaised** and **linear** estimates but not best because they variance large variance with the heteroscedasticity.

Assume a simple model: $Y_i = \beta_1 + \beta_2 x_i + e_i$

- OLS estimators are still unbiased $E\left(\hat{\beta}_2\right) = \beta_2$

Proof:

$$\hat{\beta}_2 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2} = \sum w_i y_i = \sum w_i (\beta_1 + \beta_2 x_i + e_i)$$

where $w_i = \frac{(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$

$$E\left(\hat{\beta}_2\right) = E\left[\sum w_i y_i\right] = E\left[\sum w_i (\beta_1 + \beta_2 x_i + e_i)\right] = E\left[\sum w_i \beta_1 + \beta_2 \sum w_i x_i + \sum w_i e_i\right] = \beta_2$$

Inefficiency of the OLS Estimator Due to Heteroscedasticity

- Variance of estimated parameters and the dependent variable
- $$\operatorname{var}\left(\hat{\beta}_2\right) = \hat{\sigma}^2 \left[\frac{1}{\sum_i \left(x_i - \bar{x}\right)^2} \right]$$

It was proved above that

$$E\left(\hat{\beta}_2\right) = E\left[\sum w_i y_i\right] = E\left[\sum w_i \left(\beta_1 + \beta_2 x_i + e_i\right)\right] = E\left[\sum w_i \beta_1 + \beta_2 \sum w_i x_i + \sum w_i e_i\right] = \beta_2$$

$$\operatorname{var}\left(\hat{\beta}_2\right) = E\left[E\left(\hat{\beta}_2\right) - \beta_2\right]^2 =$$

$$E\left[\sum w_i e_i\right]^2 = \left[\sum w_i^2 \operatorname{var}\left(e_i\right)^2 + \sum_{i \neq j} \sum w_i w_j \operatorname{cov}\left(e_i e_j\right)\right] = \sum w_i^2 \sigma_i^2 = \frac{\sum_i \left(x_i - \bar{x}\right)^2 \sigma_i^2}{\left[\sum_i \left(x_i - \bar{x}\right)^2\right]^2}$$

$$\operatorname{var}\left(\hat{\beta}_2\right) = \frac{\sum_i \left(x_i - \bar{x}\right)^2 \sigma_i^2}{\left[\sum_i \left(x_i - \bar{x}\right)^2\right]^2}$$

, thus variance of parameter is no longer constant. It rises with observations. When variances are larger, the standard errors are large and calculated t becomes smaller and coefficients become insignificant, though we may have correct variables in the model.

Tests for Heteroscedasticity

There are a series of formal methods developed in the econometrics literature to detect the existence of Heteroscedasticity in a given regression model.

Park test

Model $Y_i = \beta_1 + \beta_2 x_i + e_i$ (1)

Error square: $\sigma_i^2 = \sigma^2 x_i^{\beta} e^{v_i}$ (2)

Or taking log

$\ln \sigma_i^2 = \ln \sigma^2 + \beta \ln x_i + v_i$ (2')

steps : run the OLS regression for (1) and get the estimates of error terms e_i .

Square e_i , and then run a regression of $\ln e_i^2$ with x variable. Do t-test H_0: $\beta = 0$ against H_A: $\beta \neq 0$. If β is significant then that is the evidence of heteroscedasticity.

Spearman's rank correlation test

$$r_s = 1 - 6\left[\frac{\sum_i d_i^2}{n(n^2-1)}\right]$$

steps:

1. run OLS of y on x.
2. obtain errors e
3. rank e and y or x
4. find the difference of the rank
5. use t-statistics if ranks are significantly different assuming n>8 and rank correlation coefficient ρ =0.

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

If $t_{cal} > t_{crit}$ there is heteroscedasticity.

Tests for Heteroscedasticity

Glejser test	**Goldfeld-Quandt test**
$Y_i = \beta_1 + \beta_2 x_i + e_i$ There are several tests $\lvert e_i \rvert = \beta_1 + \beta_2 X_i + v_i$ $\lvert e_i \rvert = \beta_1 + \beta_2 \sqrt{X_i} + v_i$ $\lvert e_i \rvert = \beta_1 + \beta_2 \frac{1}{X_i} + v_i$ $\lvert e_i \rvert = \beta_1 + \beta_2 \frac{1}{\sqrt{X_i}} + v_i$ $\lvert e_i \rvert = \sqrt{\beta_1 + \beta_2 X_i} + v_i$ $\lvert e_i \rvert = \sqrt{\beta_1 + \beta_2 X_{i_i}^2} + v_i$ In each case do t-test H_0: $\beta = 0$ against H_A: $\beta \neq 0$. If β is significant then that is the evidence of heteroscedasticity.	Model $Y_i = \beta_1 + \beta_2 x_i + e_i$ (1) Steps: 1. Rank observations in ascending order of one of the x variable 2. Omit c numbers of central observations leaving two groups with $\frac{n-c}{2}$ number of osbervations 3. Fit OLS to the first $\frac{n-c}{2}$ and the last $\frac{n-c}{2}$ observations and find sum of the squared errors from both of them. 4. Set hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ against $H_A : \sigma_1^2 \neq \sigma_2^2$. 5. compute $\lambda = \frac{ERSS_2 / df2}{ERSS_1 / df1}$ it follows F distribution.

Tests for Heteroscedasticity

Breusch-Pagan,Godfrey test	White Test
$Y_i = \beta_1 + \beta_2 x_{2,i} + .. + \beta_k x_{k,i} + e_i$ 1.run OLS and obtain error squares 2. Obtain average error square $\tilde{\sigma}^2 = \frac{\hat{e}_i^2}{n}$ and $p_i = \frac{\hat{e}_i^2}{\tilde{\sigma}^2}$ 3. regress p_i on a set of explanatory variables $p_i = \alpha_1 + \alpha_2 x_{2,i} + .. + \alpha_k x_{k,i} + e_i$ 4. obtain squares of explained sum (EXSS) 5. $\theta = \frac{1}{2}(EXSS)$ 6. $\theta = \frac{1}{m-1}(EXSS) \approx \chi^2_{m-1}$ $H_0 : \alpha_2 = \alpha_2 = .. = \alpha_k = 0$ No heteroscedasticity and $\sigma_i^2 = \alpha_1$ a constant. If calculated χ^2_{m-1} is greater than table value there is an evidence of heteroscedasticity.	This is a more general test Model $Y_i = \beta_1 + \beta_2 x_{2,i} + \beta_3 x_{3,i} + e_i$ Run OLS to this and get $\hat{e}_i$ $\hat{e}_i^2 = \alpha_1 + \alpha_2 x_{2,i} + \alpha_3 x_{3,i} + \alpha_4 x_{2,i}^2 + \alpha_5 x_{3,i}^2$ $\alpha_6 x_{2,i} x_{3,i} + v_i$ Compute the test statistics $n.R^2 \sim \chi^2_{df}$ Again if the calculated χ^2_{df} is greater than table value there is an evidence of heteroscedasticity.

ARCH-GARCH Tests

- Autorregressive conditional Heteroscedasticity (ARCH)
- Generalised Autorregressive conditional Heteroscedasticity (ARCH)
- Moving average Autorregressive conditional Heteroscedasticity (ARCH)
- Autorregressive conditional Heteroscedasticity in Mean (ARCHM)

Variance and Covariance of Errors in Case of Heteroscedasticity Compared to the Variance of a Normal Error Term

$$E[ee'] = \begin{bmatrix} \text{var}(e_1) & \text{cov}(e_1 e_2) & \text{cov}(e_1 e_3) & \text{cov}(e_1 e_4) & \text{cov}(e_1 e_5) \\ \text{cov}(e_2 e_1) & \text{var}(e_2) & \text{cov}(e_2 e_3) & \text{cov}(e_2 e_4) & \text{cov}(e_2 e_5) \\ \text{cov}(e_3 e_1) & \text{cov}(e_3 e_2) & \text{var}(e_3) & \text{cov}(e_3 e_4) & \text{cov}(e_3 e_5) \\ \text{cov}(e_4 e_1) & \text{cov}(e_4 e_2) & \text{cov}(e_4 e_3) & \text{var}(e_4) & \text{cov}(e_4 e_5) \\ \text{cov}(e_5 e_1) & \text{cov}(e_5 e_2) & \text{cov}(e_5 e_3) & \text{cov}(e_5 e_4) & \text{var}(e_5) \end{bmatrix}$$

$$E[ee'] = \begin{bmatrix} \sigma_{11} & \sigma_{11} & \sigma_{11} & \sigma_{11} & \sigma_{11} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} & \sigma_{35} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} & \sigma_{45} \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_{55} \end{bmatrix} \quad \longleftarrow$$

$$E[ee'] = \sigma^2 I$$ where I is 5×5 identity matrix.

Variance and Covariance of Errors in Case of Heteroscedasticity Compared to the Variance of a Normal Error Term

If error term has a normal distribution as given by $e \sim N\left(0, \sigma^2\right)$

and if $\operatorname{cov}\left(e_i e_j\right) = 0 \; i \neq j$ then

$$E[ee'] = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 \end{bmatrix} \text{or}$$

←

$$E[ee'] = \sigma^2 \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \text{or}$$

$E[ee'] = \sigma^2 I$ where I is 5×5 identity matrix.

Variance and Covariance of Errors in Case of Heteroscedasticity Compared to the Variance of a Normal Error Term

$$E[ee'] = \begin{bmatrix} \operatorname{var}(e_1) & \operatorname{cov}(e_1 e_2) & \operatorname{cov}(e_1 e_3) & \operatorname{cov}(e_1 e_4) & \operatorname{cov}(e_1 e_5) \\ \operatorname{cov}(e_2 e_1) & \operatorname{var}(e_2) & \operatorname{cov}(e_2 e_3) & \operatorname{cov}(e_2 e_4) & \operatorname{cov}(e_2 e_5) \\ \operatorname{cov}(e_3 e_1) & \operatorname{cov}(e_3 e_2) & \operatorname{var}(e_3) & \operatorname{cov}(e_3 e_4) & \operatorname{cov}(e_3 e_5) \\ \operatorname{cov}(e_4 e_1) & \operatorname{cov}(e_4 e_2) & \operatorname{cov}(e_4 e_3) & \operatorname{var}(e_4) & \operatorname{cov}(e_4 e_5) \\ \operatorname{cov}(e_5 e_1) & \operatorname{cov}(e_5 e_2) & \operatorname{cov}(e_5 e_3) & \operatorname{cov}(e_5 e_4) & \operatorname{var}(e_5) \end{bmatrix}$$

$$E[ee'] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} & \sigma_{35} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} & \sigma_{45} \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_{55} \end{bmatrix}$$

← Hetroscedasti
Errors

$$E[ee'] = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_5^2 \end{bmatrix} = \quad E[ee'] = \sigma_t^2 I$$

Remedial Measures

Weighted Least Square and GLS when σ_i^2 known, divide the whole equation by σ_i

Apply OLS to transformed variables.

$$\frac{Y_i}{\sigma_i} = \frac{\beta_0}{\sigma_i} + \beta_1 \frac{X_1}{\sigma_i} + \beta_2 \frac{X_2}{\sigma_i} + \beta_3 \frac{X_3}{\sigma_i} + \beta_4 \frac{X_4}{\sigma_i} + \ldots\ldots + \beta_k \frac{X_k}{\sigma_i} + \frac{e_i}{\sigma_i}$$

Variance of this transformed model equals 1.

Other examples: $Y_i = \beta_1 + \beta_2 x_i + e_i$ and assume $e_i^2 = \sigma^2 x_i^2$

$$\frac{Y_i}{x_i} = \frac{\beta_1}{x_i} + \beta_2 + \frac{e_i}{x_i}; \quad E\left(\frac{e_i}{x_i}\right)^2 = \frac{\sigma^2 x_i^2}{x_i^2} = \sigma^2$$

In Matrix notation: $PY = PX\beta + Pe$; $P'P = \Omega^{-1}$ and $Y^* = X^*\beta + e^*$

$$\hat{\beta}_{GLS} = \left(X^{*}{}'X^{*}\right)^{-1} X^{*}{}'Y^{*} = aY$$

$\hat{\beta}_{GLS} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y = aY$ where Ω^{-1} is a variance-covariance matrix which makes the residual constant throughout all observations. when σ^2 unknown estimate σ^2 using the sample information and do the above procedures (**Gujarati** is a good text for Heteroscedasticity).

Some Heteroscedasticity Tests in Shazam

```
*GLS estimation
*Weighted Least Square estimation
*Glejser-Harvey-Pagan and ARCH tests
ols y x1 x2 /
diagnos/het
*Harvey and Phillips tests
ols y x1 x2 /
diagnos/recur
*Chow and Godfrey and Quandt test
ols y x1 x2 /
diagnos/chowone=5
*Jackknife test
ols y x1 x2 /cov=b anova
diagnos/jackknife
*Ramsey reset specification test
ols y x1 x2 /cov=b anova
diagnos/reset
```