Multiple Regression Analysis: Heteroskedasticity

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u$$

BS1802 Statistics and Econometrics

Jiahua Wu

382 Business School
j.wu@imperial.ac.uk

Imperial College Business School

Imperial means Intelligent Business

Roadmap

- Regression analysis with cross-sectional data
  - The multiple regression analysis
    - Basics: estimation, inference, analysis with dummy variables
    - More technically involved: asymptotics, heteroskedasticity, specification and data issues
- Advanced topics
  - Limited dependent variable models
  - Panel data analysis
  - Regression analysis with time series data

Outline (Wooldridge, Ch. 8.1 - 8.3)

- Consequences of heteroskedasticity
- Testing for heteroskedasticity
- Heteroskedasticity-robust inference

Outline

- Consequences of heteroskedasticity
- Testing for heteroskedasticity
- Heteroskedasticity-robust inference

Homoskedasticity and Heteroskedasticity

- Recall that the variance of OLS estimator is given by

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}, \qquad j = 1, \dots, k,$$

where $SST_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$, $\bar{x}_j = n^{-1}\sum_{i=1}^{n} x_{ij}$, and R_j^2 is the R-squared from regressing x_j on all other independent variables.

- For the variance formula to be valid, we need

Assumption (homoskedasticity)

$Var(u_i|x_{i1}, \dots, x_{ik}) = \sigma^2$ for $i = 1, 2, \dots, n$. (It implies $Var(u_i) = \sigma^2$)

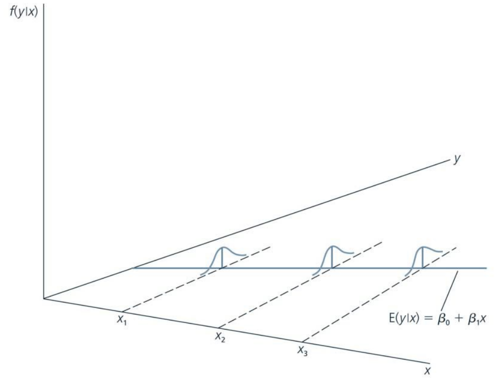
Homoskedasticity and Heteroskedasticity

Assumption (homoskedasticity)

$Var(u_i|x_{i1},\ldots,x_{ik}) = \sigma^2$ for $i = 1, 2, \ldots, n$. (It implies $Var(u_i) = \sigma^2$)

FIGURE 2.8

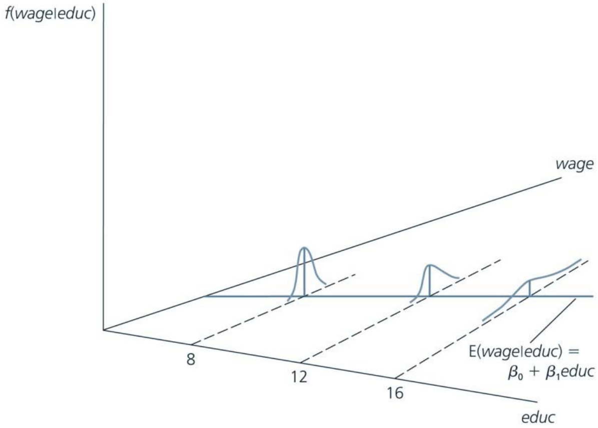
The simple regression model under homoskedasticity.



Homoskedasticity and Heteroskedasticity

FIGURE 2.9

Var(*wage*|*educ*) increasing with *educ*.



Consequences of Heteroskedasticity

- The OLS estimators are unbiased, even if we do not assume homoskedasticity
- Homoskedasticity is required for using the formula of the variance of the OLS estimator, which is important for inference
- The standard errors of the estimates are biased if we have heteroskedasticity
- If the standard errors are biased, we can not use the usual t statistics or F statistics for drawing inferences

Outline

- Consequences of heteroskedasticity
- Testing for heteroskedasticity
- Heteroskedasticity-robust inference

Heteroskedasticity Tests

- Essentially want to test $H_0 : Var(u|x_1, x_2, \ldots, x_k) = \sigma^2$, which is equivalent to

$$H_0 : E(u^2|x_1, x_2, \ldots, x_k) = E(u^2) = \sigma^2$$

Heteroskedasticity Tests: The Breusch-Pagan Test

- If assume a linear relationship between u^2 and x_j, i.e.,

$$u^2 = \delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k + v,$$

  the null hypothesis of homoskedasticity is equivalent to $H_0 : \delta_1 = \delta_2 = \cdots = \delta_k = 0.$

- The Breusch-Pagan test
  - OLS $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$ and save the squared residuals $\hat{u}^2$
  - OLS $\hat{u}^2 = \delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k + error$ and save the R-squared $R^2_{\hat{u}^2}$
  - The test statistic

$$F = \frac{R^2_{\hat{u}^2}/k}{(1 - R^2_{\hat{u}^2})/(n-k-1)} \sim F_{k,n-k-1} \quad \text{under the null}$$

  - Reject the null if F is too large (or has a too-small p-value)

Heteroskedasticity Tests: The White Test

- The Breusch-Pagan test will detect any linear forms of heteroskedasticity
- The White test allows for nonlinearities by using squares and crossproducts of all the x's
- The White test
  - OLS $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$ and save the residuals and the fitted values, $\hat{u}$ and $\hat{y}$
  - OLS $\hat{u}^2 = \delta_0 + \delta_1 \hat{y} + \delta_2 \hat{y}^2 + error$ and save the R-squared $R^2_{\hat{u}^2}$
  - The test statistic

$$F = \frac{R^2_{\hat{u}^2}/2}{(1 - R^2_{\hat{u}^2})/(n-3)} \sim F_{2,n-3} \quad \text{under the null}$$

  - Reject the null if F is too large (or has a too-small p-value)

Heteroskedasticity Tests: An Example

- Eg. Log wage model (wage1.RData)

$$\widehat{\log(wage)} = \underset{(.104)}{.284} + \underset{(.0073)}{.0920}\, educ + \underset{(.0017)}{.0041}\, exper + \underset{(.003)}{.022}\, tenure$$

$n = 526, R^2 = .316$

  - Breusch-Pagan test: p-value $= .013$
  - White test: p-value $= .035$

Outline

- Consequences of heteroskedasticity
- Testing for heteroskedasticity
- Heteroskedasticity-robust inference

Heteroskedasticity-Robust Inference

- It is possible to adjust the OLS standard errors to make the t stat (or F stat) valid in the presence of heteroskedasticity of unknown form
- The adjustment is called heteroskedasticity-robust procedure
- The procedure is "robust" because the adjusted t stat (or F stat) is valid regardless of the type of heteroskedasticity in the population (even if there is no heteroskedasticity)

Robust Standard Errors

- Denote $r.se(\hat{\beta}_j)$ as robust standard error
- The robust t stat is

$$t \text{ stat} = \frac{\hat{\beta}_j - a_j}{r.se(\hat{\beta}_j)}$$

- These robust standard errors only have asymptotic justification
  - With small sample sizes, robust t stat will not have a distribution close to t, and inferences will not be correct
- The robust F stat must be computed using a formula different from the original one.
  - But the robust F stat is easily obtained in R, using `linearHypothesis`.

Robust Standard Errors: An Example

- Eg. Log wage model (wage1.RData)

$$\widehat{\log(wage)} = \underset{(.104)}{.284} + \underset{(.0073)}{.0920}\,educ + \underset{(.0017)}{.0041}\,exper + \underset{(.003)}{.022}\,tenure$$

  The model with robust standard errors is

$$\widehat{\log(wage)} = \underset{[.112]}{.284} + \underset{[.0079]}{.0920}\,educ + \underset{[.0017]}{.0041}\,exper + \underset{[.004]}{.022}\,tenure$$

  $n = 526, R^2 = .316$ (the robust results are in [ ])

- Hypotheses: $H_0 : \beta_{educ} - \beta_{exper} = 0$ vs $H_1 : \beta_{educ} - \beta_{exper} \neq 0$
  - F stat: (158.46), $[143.46]$

When to Use Robust Standard Errors in Practice?

- Always report robust standard errors for linear probability model and panel data model
- For other models,
  - Test for heteroskedasticity
  - Report robust standard errors only if there is evidence of heteroskedasticity, as robust standard errors only have asymptotic justification