L8: Heteroscedasticity

Feng Li
feng.li@cufe.edu.cn

School of Statistics and Mathematics
Central University of Finance and Economics

What is so-called heteroscedasticity

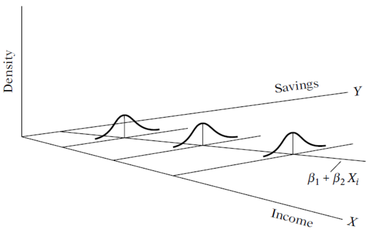
- In a linear regression model, we assume the error term has a normal distribution with mean zero and variance of σ^2, i.e.

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

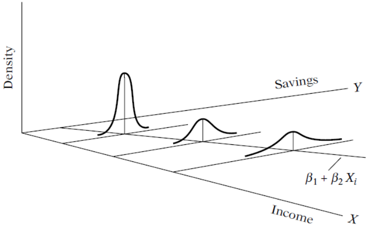
  which is called **homoscedasticity**.
- But when the error term does not have constant variance, i.e.

$$Var(u_i) = \sigma_i^2$$

  we call it **heteroscedasticity**.
- See the differences between the two pictures for the model $Saving = \alpha + \beta Income + u_i$



Homoscedastic disturbances.



Heteroscedastic disturbances.

An OLS example

- Recall the model $Y_i = \alpha_1 + \alpha_2 X_i + u_i$.
- If the error term u_i is homoscedastic with variance σ^2, we know we have BLUE estimators and

$$\hat{\alpha}_2 = \frac{\sum x_i y_i}{\sum x_i^2},\ Var(\hat{\alpha}_2) = \frac{\sigma^2}{\sum x_i^2}.$$

- If the error term u_i is homoscedastic with variance σ_i^2, we have

$$\hat{\alpha}_2 = \frac{\sum x_i y_i}{\sum x_i^2},\ Var(\hat{\alpha}_2) = \frac{\sum x_i^2 \sigma_i^2}{(\sum x_i^2)^2} \neq \frac{\sigma^2}{\sum x_i^2}.$$

  why? see Appendix 11A.1.
  - $\hat{\alpha}_2$ is still linear and unbiased, **why?**
  - But it is not "best" anymore, i.e. will not grant the minimum variance.

Use GLS to take heteroscedasticity into account I

- The OLS method treats every observation equally and does not take heteroscedasticity into account.
- The **generalized least squares** (GLS) will.
  - Consider the heteroscedasticity model

$$Y_i = \beta_1 + \beta_2 X_{1i} + u_i, \text{ where } Var(u_i) = \sigma_i^2$$

  - If we transform the model by dividing $1/\sqrt{w_i}$ where $w_i = 1/\sigma_i^2$ at both sides (assume σ_i is known),

$$\frac{Y_i}{\sigma_i} = \beta_1 \frac{1}{\sigma_i} + \beta_2 \frac{X_i}{\sigma_i} + \frac{u_i}{\sigma_i}$$

    which can be rewritten as

$$Y_i^* = \beta_1 X_{0i}^* + \beta_2 X_{1i}^* + u_i^*,$$

    and $u_i^* = u_i/\sigma_i$ is the new error term for the new model.
  - $Var(u_i/\sigma_i) = 1$ is now a constant. **why?**
  - We call $\hat{\beta}_1^*$ $\hat{\beta}_2^*$ as GLS estimators

$$\hat{Y}_i^* = \hat{\beta}_1^* X_{0i}^* + \hat{\beta}_2^* X_{1i}^*,$$

Use GLS to take heteroscedasticity into account II

- To obtain GLS estimators, we minimize

$$\sum(\hat{u}_i^*)^2 = \sum(Y_i^* - \hat{\beta}_1^* X_{0i}^* - \hat{\beta}_2^* X_{1i}^*)^2 = \sum w_i (Y_i - \hat{\beta}_1^* X_{0i} - \hat{\beta}_2^* X_{1i})^2$$

  which is done by the usual way we have done in OLS.

- The GLS estimator of β_2^* is

$$\hat{\beta}_2^* = \frac{\sum w_i \sum w_i X_i Y_i - \sum w_i X_i \sum w_i Y_i}{\sum w_i \sum w_i X_i^2 - \left(\sum w_i X_i\right)^2}$$

  and the variance is

$$Var\left(\hat{\beta}_2^*\right) = \frac{\sum w_i}{\sum w_i \sum w_i X_i^2 - \left(\sum w_i X_i\right)^2}$$

  where $w_i = 1/\sigma_i^2$.

- When $w_i = w = 1/\sigma^2$, the GLS estimator will reduce to the OLS estimator. **Verify this!**
- $\hat{\beta}_1^* = \bar{Y}^* - \hat{\beta}_2^* \bar{X}^*$ where $\bar{Y}^* = (\sum w_i Y_i)/\sum(w_i)$, $\bar{X}^* = (\sum w_i X_i)/\sum(w_i)$.

Use GLS to take heteroscedasticity into account III

- At this particular setting $w_i = 1/\sigma_i^2$, we call this is **weighted least squares** (WLS) which is a special case of GLS.
- $\hat{\beta}_2^*$ is unbiased and $Var(\hat{\beta}_2^*) < Var(\hat{\beta}_2)$.
- It can be shown that

$$\hat{\beta}_2^* = \frac{\sum w_i x_i^* y_i^*}{\sum w_i \left(x_i^*\right)^2}$$

  and

$$Var\left(\hat{\beta}_2^*\right) = \frac{1}{\sum w_i \left(x_i^*\right)^2}$$

  where

$$x_i^* = X_i - \bar{X}^*, \quad y_i^* = Y_i - \bar{Y}^*$$

- See **Exercise 11.5**.

GLS in matrix form I

- Let Y given X is a linear function of X, whereas the conditional variance of the error term given X is a **known** matrix Ω

$$Y = X\beta + \varepsilon, \qquad E[\varepsilon|X] = 0,\ \text{Var}[\varepsilon|X] = \Omega.$$

- Generalized least squares method estimates β by minimizing the squared Mahalanobis length of this residual vector:

$$\hat{\beta} = \underset{\beta}{\arg\min}\,(Y - X\beta)'\,\Omega^{-1}(Y - X\beta),$$

- The estimator has an explicit formula:

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y.$$

- The GLS estimator is unbiased, consistent, efficient, and asymptotically normal:

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}\big(0,\ (X'\,\Omega^{-1}X)^{-1}\big).$$

GLS in matrix form II

- If the covariance of the errors Ω is **unknown**, one can get a consistent estimate of Ω , say $\widehat{\Omega}$, we proceed in two stages:
- The model is estimated by OLS or another consistent (but inefficient) estimator, and the residuals are used to build a consistent estimator of the errors covariance matrix;

$$\begin{aligned}
\widehat{\beta}_{OLS} &= (X'X)^{-1}X'y \\
\hat{u}_j &= (Y - Xb)_j \\
\widehat{\Omega}_{OLS} &= \text{diag}(\widehat{\sigma}_1^2, \widehat{\sigma}_2^2, \ldots, \widehat{\sigma}_n^2).
\end{aligned}$$

- Using the consistent estimator of the covariance matrix of the errors, we implement GLS ideas.

$$\widehat{\beta}_{GLS} = (X'\widehat{\Omega}_{OLS}^{-1}X)^{-1}X'\widehat{\Omega}_{OLS}^{-1}y$$

GLS in matrix form III

- The procedure can be iterated and this estimation of $\widehat{\Omega}$ can be iterated to convergence.

$$
\begin{aligned}
\widehat{u}_{GLS} &= Y - X\widehat{\beta}_{GLS} \\
\widehat{\Omega}_{GLS} &= \text{diag}(\widehat{\sigma}^2_{GLS,1}, \widehat{\sigma}^2_{GLS,2}, \ldots, \widehat{\sigma}^2_{GLS,n}) \\
\widehat{\beta}_{GLS} &= (X'\widehat{\Omega}^{-1}_{GLS}X)^{-1}X'\widehat{\Omega}^{-1}_{GLS}y
\end{aligned}
$$

- **Question**: How do you calculate the degrees of freedom of the model? [Hint: think about the hat matrix]

WLS example – Example 11.7 I

- Assume we want to make WLS regression with the give data.

Y	X	σ_i	Y_i/σ_i	X_i/σ_i
3396	1	743.7	4.5664	0.0013
3787	2	851.4	4.4480	0.0023
4013	3	727.8	5.5139	0.0041
4104	4	805.06	5.0978	0.0050
4146	5	929.9	4.4585	0.0054
4241	6	1080.6	3.9247	0.0055
4387	7	1243.2	3.5288	0.0056
4538	8	1307.7	3.4702	0.0061
4843	9	1112.5	4.3532	0.0081

- What can you do then?
  - **Option 1:** Apply the general GLS formula in p.5 to obtain the estimators.
  - **Option 2:** Use OLS to regress Y/σ_i with $1/\sigma_i$ and X_i/σ_i *without intercept*.
- Can you obtain the same results?
- Compare the WLS results with OLS results.

WLS example – Example 11.7 II

The regression results of WLS are as follows:

$$\widehat{(Y_i/\sigma_i)} = 3406.639(1/\sigma_i) + 154.153(X_i/\sigma_i)$$

$$(80.983) \qquad (16.959)$$

$$t = \quad (42.066) \qquad (9.090)$$

$$R^2 = 0.9993^{31}$$

OLS regression results:

$$\hat{Y}_i = 3417.833 + 148.767 X_i$$

$$(81.136) \qquad (14.418)$$

$$t = \quad (42.125) \qquad (10.318) \qquad R^2 = 0.9383$$

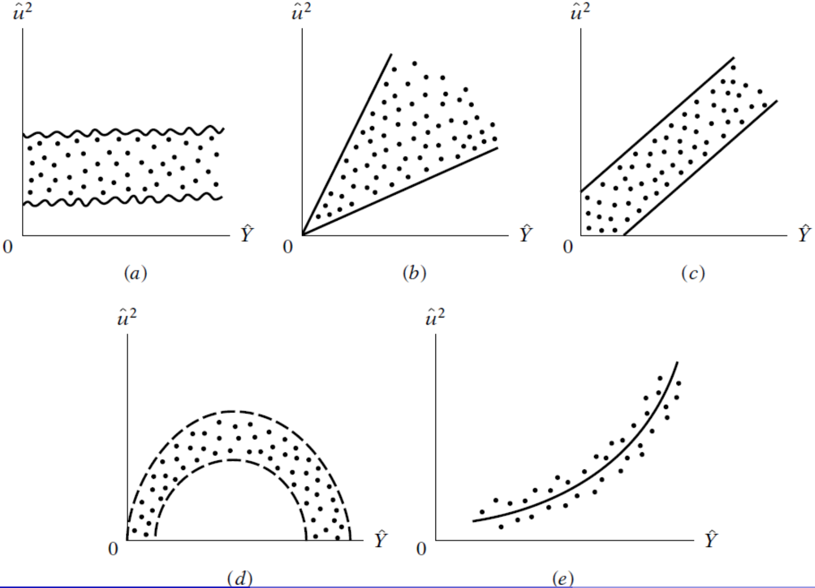
- How do the standard errors and t statistics change?

Consequences of using OLS when heteroscedastic

- Suppose there are heteroscedastic but we insist using OLS. What will go wrong? – **whatever conclusions we draw may be misleading**.
- We could not establish confidence intervals and test hypotheses with usual *t*, *F* tests.
- The usual tests are likely to give larger variance than the true variance.
- The variance estimator of $\hat{\beta}$ by OLS is a **biased** estimator of the true variance.
- The usual estimator of σ^2 which was $\sum \hat{u}_i^2/(n-2)$ is **biased**.

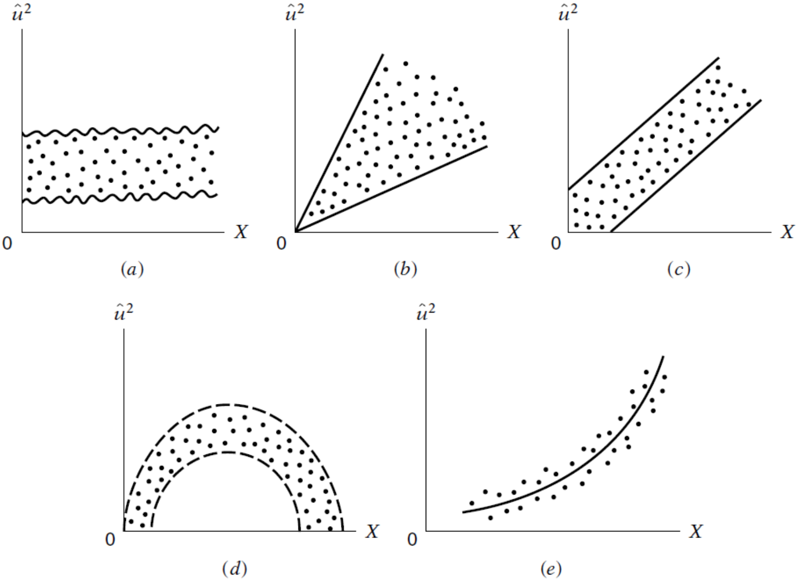
Detecting heteroscedasticity (1)

↪ Plot $\hat{u}_i^2$ against $\hat{Y}_i$



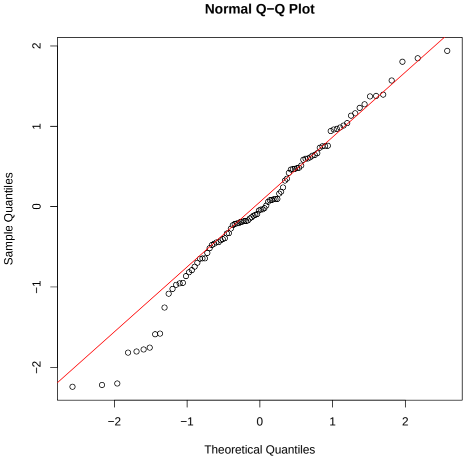
Detecting heteroscedasticity (2)

↛ Plot $\hat{u}_i^2$ against X_i



Detecting heteroscedasticity (3)
↛ QQ plot

- If the residual is normally distributed, plot sample quantile for the residual against the theoretical quantile from standard normal distribution should on the 45 degree line.



Detecting heteroscedasticity (4)
$\hookrightarrow$ White's general heteroscedasticity test

- H_0: No heteroscedasticity.
- Consider $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$, (other models are the same)
- **step 1:** Do the OLS to obtain the residuals $\hat{u}_i$.
- **step 2:** Run the following model with the covariates and their crossproducts

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} + \nu_i$$

  and obtain R^2.
- **step 3:** $nR^2 \sim \chi^2(k-1)$ where k is no. of unknown parameters in **step 2**.
- **step 4:** If $\chi^2_{obs}(k-1) > \chi^2_{crit}(k-1)$, reject H_0.
- Question: How do you carry out White's test with the model $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$?

Example of White's test

- Consider the following regression model with 41 observations,

$$\ln Y_i = \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i$$

where Y = ratio of trade taxes (import and export taxes) to total government revenue, X_2 = ratio of the sum of exports plus imports to GNP, and X_3 = GNP per capita.

- By applying White's heteroscedasticity test. We first obtain the residuals from regression.
- Then we do the following auxiliary regression

$$\hat{u}_i^2 = -5.8 + 2.5 \ln X_{2i} + 0.69 \ln X_{3i} - 0.4 (\ln X_{2i})^2 - 0.04 (\ln X_{3i})^2 + 0.002 \ln X_{2i} \ln X_{3}$$

and $R^2 = 0.1148$.

- Can you compute the white's test statistic?
- What is your conclusion of heteroscedasticity? (The 5% critical value of $\chi^2_{df=5} = 11.07$, and the 10% critical value of $\chi^2_{df=5} = 9.24$)

Detecting heteroscedasticity (5)
↬ Goldfeld-Quandt test

- It is popular to assume σ_i^2 is positively related to one of the covariates e.g. $\sigma_i^2 = \sigma^2 X_{2i}^2$ in a three covariates model.
- The bigger X_i we have, the bigger σ_i^2 is.
- H_0 : Homoscedasticity
- **step 1:** Sort covariates with the order of X_{2i}
- **step 2:** Delete the c central observations and dived the remaining parts into two groups.
- **step 3:** Fit the two groups separately with OLS and obtain RSS_1 (for the small values group) and RSS_2 (for the large values group) with both $(n-c)/2-k$ degrees of freedom. **Why?**
- **step 4:** Compute the ratio

$$\lambda = \frac{\mathrm{RSS}_2/[(n-c)/2-k]}{\mathrm{RSS}_1/[(n-c)/2-k]} \sim F(((n-c)/2-k),((n-c)/2-k))$$

- Reject H_0 if $\lambda > F_{crit}(((n-c)/2-k),((n-c)/2-k))$.

Detecting heteroscedasticity (6)
↪ Breusch-Pagan-Godfrey test

- Consider the model $Y_i = \beta_1 + \beta_2 X_{2i} + ... + +\beta_k X_{ki} + u_i$
- Assume that $\sigma_i^2 = \alpha_1 + \alpha_2 Z_{2i} + ... + \alpha_m Z_{mi}$ where Z_i are known variables which can be X_i.
- If there no heteroscedasticity, then $\alpha_2 = ... = \alpha_m = 0$ and $\sigma_i^2 = \alpha_1$.
- **step 1:** Obtain $\hat{u}_1, ..., \hat{u}_n$ by the model.
- **step 2:** Obtain $\tilde{\sigma}^2 = \sum \hat{u}_i^2 / n$.
- **step 3:** Construct variable $p_i = \hat{u}_i^2 / \tilde{\sigma}^2$
- **step 4:** Regress $p_i = \alpha_1 + \alpha_2 Z_{2i} + ... + \alpha_m Z_{mi} + v_i$
- **step 5:** Obtain

$$ESS/2 \sim \chi^2(m-1)$$

- Evidence of heteroscedasticity when $ESS/2 > \chi^2_{crit}(m-1)$.

Detecting heteroscedasticity (7)

↬ Spearman's rank correlation test

1. The null hypothesis: Heteroscedasticity
2. Obtain the residuals $\hat{u}_i$ from the regression.
3. Rank $|\hat{u}_i|$ and X_i (or Y_i)
4. Compute the Spearman's rank correlation coefficients

$$r_s = 1 - 6\frac{\sum d_i^2}{n(n^2-1)}$$

   where d_i are the differences of $|u_i|$ and X_i in the ranked order and n is number of individuals.

5. The significance of the sample r_s can be tested by the t test as

$$t_{obs} = \frac{r_s\sqrt{n-2}}{1-r_s^2}$$

6. Decision rule: if $t_{obs} > t_{critical}$, accept H_0. Otherwise, there is no heteroscedasticity. Multiple regressors should repeat this procedure multiple times.

How to obtain estimators

$\looparrowright$ with $Y_i = \beta_1 + \beta_2 X_i + u_i$ when $E(u_i) = 0$ and $Var(u_i) = \sigma_i^2$

- **When σ_i is known:** use WLS method to obtain BLUE estimators. pp. 4–5
- **When σ_i is not known:**
  - If $V(u_i) = \sigma^2 X_i^2$, do OLS with model

    $$\frac{Y_i}{X_i} = \beta_2 + \beta_1 \frac{1}{X_i} + \frac{u_i}{X_i}$$

    and $Var(u_i/X_i) = \sigma^2$. **Why?**
  - If $Var(u_i) = \sigma^2 X_i$ $(X_i > 0)$, do OLS with model

    $$\frac{Y_i}{\sqrt{X_i}} = \beta_2 + \beta_1 \frac{1}{\sqrt{X_i}} + \frac{u_i}{\sqrt{X_i}}$$

    and $Var(u_i/\sqrt{X_i}) = \sigma^2$. **Why?**
  - If $Var(u_i) = \sigma^2 [E(Y_i)]^2$ $(X_i > 0)$, do OLS with model

    $$\frac{Y_i}{\hat{Y}_i} = \beta_2 + \beta_1 \frac{1}{\hat{Y}_i} + \frac{u_i}{\hat{Y}_i}$$

    and $Var(u_i/\hat{Y}_i) \approx Var(u_i)/[E\hat{Y}_i]^2 = Var(u_i)/{Y_i}^2 = \sigma^2$.
  - Do OLS with log transformed data $\ln Y_i = \beta_1 + \beta_2 \ln X_i + \nu_i$ can also reduce heteroscedasticity.

Take home questions

- **11.1, 11.2, 11.6, 11.21**
- How do you perform the maximum likelihood estimation when there is heteroscedastic and Ω is known?