Heteroskedasticity

⟹ 不变: ① OLS still unbiased and consistent.

② interpretation of R^2 is not changed.

变 ① invalidates variance formulas for OLS estimators. $Var(\beta_i)$ ✗

② usual F-test, t-test are not valid.

③ ~~Gauss-Markov Theorem~~ ⟹ OLS is no longer the best linear unbiased estimator (BLUE)

- **Heteroskedasticity-robust inference after OLS estimation**
  - Formulas for OLS standard errors and related statistics have been developed that are robust to heteroskedasticity of unknown form.
  - All formulas are only valid in large samples.
  - Formula for heteroskedasticity-robust OLS standard error.

$$\widehat{Var}(\hat{\beta}_j) = \frac{\sum_{i=1}^{n} \hat{r}_{ij}^2 \hat{u}_i^2}{SSR_j^2}$$

Also called White/Huber/Eicker standard errors. They involve the squared residuals from the regression and from a regression of x_j on all other explanatory variables.

t: 以前的能用

  - Using these formulas, the usual t test is valid asymptotically.
  - The usual F statistic does not work under heteroskedasticity, but heteroskedasticity robust versions are available in most software.

F: 以前不能用，有新版.

Example: $y = \beta_0 + \beta_1 x + u$.

$Var(u_i | x_i) = \sigma_i^2$ (heteroskedasticity).

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$= \beta_1 + \frac{\sum (x_i - \bar{x}) u_i}{\sum (x_i - \bar{x})^2}$$

$$\mathrm{Var}(\hat{\beta}_1) = \mathrm{Var}\left(\frac{\sum (x_i - \bar{x}) u_i}{\underbrace{\sum (x_i - \bar{x})^2}_{SST_x}} \right)$$

$$= \frac{1}{SST_x^2} \mathrm{Var}((x_1 - \bar{x}) u_1 + (x_2 - \bar{x}) u_2 \cdots)$$

$$= \frac{1}{SST_x^2} \sum_{i=1}^{n} (x_i - \bar{x})^2 \sigma_i^2$$

(heteroskedasticity)

$$\sigma_i^2 = \mathrm{Var}(u_i \mid x_i) = E(u_i^2 \mid x_i)$$

$$\Rightarrow \hat{\sigma}_i^2 = E(\hat{u}_i^2 \mid x_i) = \hat{u}_i^2$$

$$\Rightarrow \widehat{\mathrm{Var}(\hat{\beta}_1)} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 \hat{u}_i^2}{SST_x^2}$$

$$se(\hat{\beta}_1) = \sqrt{\widehat{\mathrm{Var}(\hat{\beta}_1)}}$$

heteroskedasticity-robust standard errors.

$$H_0: \hat{\beta}_1 = a$$

$$t = \frac{\hat{\beta}_1 - a}{se(\hat{\beta}_1)}$$

(n is large).

Multiple variables :

$$\widehat{Var}(\hat{\beta}_j) = \frac{\sum_{i=1}^{n} \hat{r}_{ij}^2 \hat{u}_i^2}{SSR_j^2}$$

Also called White/Huber/Eicker standard errors. They involve the squared residuals from the regression and from a regression of x_j on all other explanatory variables.