	Homoskedasticity	Heteroskedasticity
OLS	unbiased	unbiased
	normal	normal
	efficient	inefficient
	$\widehat{Var}(b)$ reliable	$\widehat{Var}(b)$ needs "optional" command

---

$Var(\varepsilon_i) = \sigma_i^2$ - heteroskedasticity

Test for heteroskedasticity
- If homoskedasticity
  - use vanilla output → be happy
- If heteroskedasticity
  - use "robust" SE and be cautious

$$Y_i = \alpha + \beta X_i + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma_i^2)$$

$$b_{OLS} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \qquad x_i = X_i - \bar{X} \quad y_i = Y_i - \bar{Y}$$

these are nonrandom

$$b_{OLS} = \beta + \frac{\sum_{i=1}^{n} x_i \varepsilon_i}{\sum_{i=1}^{n} x_i^2} = \beta + \sum_{j=1}^{n} w_j \varepsilon_j \qquad \text{where } w_j = \frac{x_j}{\sum_{i=1}^{n} x_i^2}$$

$$w_i \varepsilon_i \sim N(0, w_i^2 \sigma_i^2)$$

$$\Rightarrow \sum_{i=1}^{n} w_i \varepsilon_i \sim N\left(0, \sum_{i=1}^{n} w_i^2 \sigma_i^2\right)$$

$$\Rightarrow b_{OLS} \sim N\left(\beta, \sum_{i=1}^{n} w_i^2 \sigma_i^2\right)$$

If by luck of the draw, $\sigma_i^2 = \sigma^2 \ \forall i \Rightarrow \sum_{i=1}^{n} w_i^2 \sigma_i^2 = \sigma^2 \sum_{i=1}^{n} w_i^2 = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}$

$\Rightarrow$ Estimate σ^2 by s^2 $\Rightarrow$ Variance of b is estimated by $s^2 \sum_{i=1}^{n} w_i^2$

$e_i = Y_i - (a + bX_i)$

$e_i \approx \varepsilon_i$
$e_i^2 \approx \varepsilon_i^2$
$\sigma_i^2 = Var(\varepsilon_i) = E[\varepsilon_i^2]$
} intuition (incorrect, but accept it for now)

$\Rightarrow \sum_{i=1}^{n} w_i^2 e_i^2$ as an estimator for b_{OLS}

If n is "large" then $\sum_{i=1}^{n} w_i^2 e_i^2$ is an ok estimator for $Var(b_{OLS})$ under heteroskedasticity,

- Robust standard error.
- White's heteroskedasticity corrected standard error
- Huber's Standard error

Inefficiency?

$Y_i = \alpha + \beta X_i + \varepsilon_i$ $\quad \varepsilon_i \sim N(0, \sigma^2)$

OLS is BLUE

$Y_i = \alpha + \beta X_i + \varepsilon_i$ $\quad \varepsilon_i \sim N(0, \sigma_i^2)$

$$\frac{Y_i}{\sigma_i} = \alpha \cdot \frac{1}{\sigma_i} + \beta \frac{X_i}{\sigma_i} + \frac{\varepsilon_i}{\sigma_i}$$

$Y_i^* = \alpha X_{1i}^* + \beta X_{2i}^* + \varepsilon_i^*$ $\quad \frac{\varepsilon_i}{\sigma_i} \sim N(0,1)$

Result: Regress Y_i^* on X_{1i}^* and X_{2i}^*

$$\min_{a,b} \sum_{i=1}^{n} \left[Y_i^* - aX_{1i}^* - bX_{2i}^*\right]^2$$

$$= \min_{a,b} \sum_{i=1}^{n} \left[\frac{Y_i}{\sigma_i} - a\frac{1}{\sigma_i} - b\frac{X_i}{\sigma_i}\right]^2$$

$$= \underbrace{\min_{a,b} \sum_{i=1}^{n} \frac{1}{\sigma_i^2}(Y_i - a - bX_i)^2}_{\text{Weighted (generalized) least squares}}$$

Don't use $\min_{a,b} \sum_{i=1}^{n} \frac{1}{e_i^2}(Y_i - a - bX_i)^2$

Weighted least squares is BLUE

Estimate $\hat{\sigma}_i^2$ by some reasonably precise method

Testing for heteroskedasticity

(1) Goldfeld - Quandt Test

(2) Breusch - Pagan Test

(3) White's Test ← very sophisticated

(1) Goldfeld - Quandt Test

$$\sigma_i^2 = \begin{cases} \sigma^2_{small} \\ \sigma^2_{large} \end{cases}$$ There is heteroskedasticity, but it takes this form

Ex ante, order the data into two groups: (using intuition)

1st group: n_1 observations $n_1 + n_2 = n$

2nd group: n_2 observations

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i, \quad i = 1, \dots, n_1 \quad \varepsilon_i \sim N(0, \sigma^2_{small})$$

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i, \quad i = n_1+1, \dots, n_1+n_2 \quad \varepsilon_i \sim N(0, \sigma^2_{large})$$

* Estimate σ^2_{small} by $\frac{ESS_1}{n_1 - k}$

* Estimate σ^2_{large} by $\frac{ESS_2}{n_2 - k}$

$$\frac{\sigma^2_{large}}{\sigma^2_{small}} \sim \frac{ESS_2 / (n_2 - k)}{ESS_1 / (n_1 - k)} \equiv GQ \sim F(n_2 - k, n_1 - k)$$

$H_0: \sigma^2_{large} = \sigma^2_{small}$

$H_A: \sigma^2_{Large} > \sigma^2_{small}$

Reject H_0 if GQ sufficiently large.

(2) Breusch-Pagan Test (BP)

$\sigma_i^2 = f(\alpha_1 + \alpha_2 z_{2i} + \dots + \alpha_p z_{pi})$

Remark: (3) White's Test (more general than Breusch-Pagan)

$\sigma_i^2 = f(z_{2i}, z_{3i}, \dots, z_{pi})$

Back to BP:

(i) Estimate β's by OLS

(ii) Compute e_i^2

(iii) Let $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2$

(iv) Regress $\frac{e_i^2}{\hat{\sigma}^2}$ on $z_{2i}, \dots, z_{pi}$

(v) Obtain RSS under the model in (iv)

(vi) $BP = \frac{RSS}{2}$

(vii) Reject H_0: Homoskedasticity if BP is "large"

accept if "small"

Use $\chi^2(p)$ distribution