

OLS assumptions, heteroskedasticity and robust standard errors

The ideas in six lines

1. **Unbiased OLS** needs: linear in parameters, random sample, no perfect collinearity, $\mathbb{E}[u \mid x] = 0$.
2. The **usual standard error** also needs **homoskedasticity**: the same error variance for every x .
3. All five together: OLS is the best linear unbiased estimator (**Gauss–Markov**).
4. **Heteroskedasticity** leaves the slope unbiased but makes the usual standard errors wrong.
5. Look for it with a residual plot, or the **Breusch–Pagan** and **White** tests.
6. **Robust (White) standard errors** change the standard error only, not the estimate.

The formulas

$$\text{Usual: } \widehat{\text{se}}(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{\sum_i (x_i - \bar{x})^2}}$$

$$\text{Robust: } \widehat{\text{Var}}(\hat{\beta}_1) = \frac{\sum_i (x_i - \bar{x})^2 \hat{u}_i^2}{(\sum_i (x_i - \bar{x})^2)^2}$$

95% interval: $\hat{\beta}_1 \pm 1.96 \times \text{se}$; t-statistic for $\beta_1 = 0$: $\hat{\beta}_1 / \text{se}$.

Worked example

Simulated data from the chapter: 200 workers, the error spread grows with AI use ($h = 3$).

- Usual standard error 0.45; actual spread of the estimate 0.57; robust standard error 0.56.
- 95% half-width: usual $1.96 \times 0.45 \approx 0.88$; robust $1.96 \times 0.56 \approx 1.10$.
- The usual interval is too narrow and covers the true slope less than 95% of the time.

Four traps

- “Heteroskedasticity biases the slope.” No. It affects the standard errors.
- “Robust standard errors fix the slope.” No. The slope stays exactly the same.
- “Robust standard errors are always larger.” Not always. They can be smaller.
- “A tight interval means a precise estimate.” Only if the formula behind it fits the data.

Words to know

unbiased · zero conditional mean · homoskedasticity · heteroskedasticity · Gauss–Markov · robust (White) standard error · Breusch–Pagan test

Try these

1. Which assumption makes the OLS slope unbiased (together with linearity, random sampling and no perfect collinearity)?
2. $\hat{\beta}_1 = 1.2$ with a robust standard error of 0.3. What is the 95% interval? (Two decimals.)
3. With the numbers in question 2, what is the t-statistic for $\beta_1 = 0$?
4. Residuals fan out as x grows. What is this called, and what do you report?
5. True or false: robust standard errors are always larger than the usual ones.
6. Usual standard error 0.45, robust 0.56. By how many percent is the robust interval wider?

Answers

1. $\mathbb{E}[u \mid x] = 0$ (zero conditional mean).
2. $1.2 \pm 1.96 \times 0.3 = 1.2 \pm 0.59$, so about $[0.61, 1.79]$.
3. $1.2/0.3 = 4.0$.
4. Heteroskedasticity. Keep the OLS slope and report robust standard errors.
5. False. They can be smaller than the usual ones.
6. $0.56/0.45 - 1 \approx 24\%$.