Confidence Intervals and Heteroskedasticity

23 Sep 2026 | Chapter 5

ECO 351 – Introduction to Econometrics

Shi Qiu – Fall 2026

Recall: Coefficient uncertainty

From Lecture 9

The school slope estimate was -2.28 with estimated SE .52. Its approximate 95% interval was $[-3.30, -1.26]$.

$$CI_{95\%} = \widehat{\beta}_1 \pm 1.96 SE(\widehat{\beta}_1)$$

What determines the margin of error? If the SE doubled with the estimate unchanged, what would happen to the interval center and width?

Why the usual standard-error formula can fail

Answer

$$\text{margin of error} = 1.96 \times SE$$

Doubling the SE doubles the margin and total interval width. The unchanged coefficient estimate stays at the center.

What features of the regression errors determine whether a standard-error formula is appropriate?

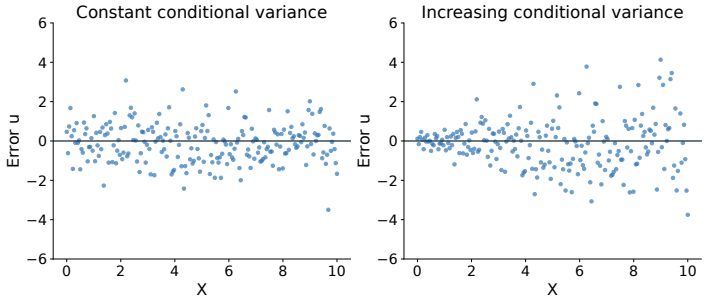
The line can be right while its uncertainty is wrong

A regression describes the conditional mean of scores. Districts may also have different amounts of score variation around that mean.

$$Score_i = \beta_0 + \beta_1 STR_i + u_i$$

We now study the spread of u_i at different STR values, while keeping the fitted-mean question separate.

The mean can stay at zero while the spread changes



Homoskedasticity

$\text{Var}(u_i \mid X_i) = \sigma^2$: the conditional variance is constant.

Heteroskedasticity

$\text{Var}(u_i \mid X_i)$ varies with X_i: the vertical spread changes.

Why does the variance pattern enter a slope's SE?

Conditional on the regressors, the OLS slope error is a weighted average of regression errors:

$$\widehat{\beta}_1 - \beta_1 = \frac{\sum_i (X_i - \bar{X}) u_i}{\sum_i (X_i - \bar{X})^2}.$$

Observations farther from the mean of X receive more weight. If those observations also have larger error variance, that matters for slope uncertainty.

Two variance formulas for the same OLS slope

Let $S_{XX} = \sum_i (X_i - \bar{X})^2$.

Constant conditional variance

$$\widehat{\mathrm{Var}}_{hom}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{S_{XX}}$$

One estimated error variance is applied to every observation.

Variance may differ

$$\widehat{\mathrm{Var}}_{HC1}(\hat{\beta}_1) = \frac{n}{n-2} \times \frac{\sum_i (X_i - \bar{X})^2 \hat{u}_i^2}{S_{XX}^2}$$

Observed squared residuals estimate the separate contributions.

The OLS coefficient is unchanged; only its estimated variance changes.

Stata: change only the standard-error calculation

Predict which entries will match exactly and which may differ.

```
do "setup.do" schools
regress score str
estimates store conventional
regress score str, vce(robust)
estimates store robust
estimates table conventional robust, b(%9.3f) se(%9.3f)
```

Read the result

Both columns have identical coefficients. The standard errors differ because they use different variance assumptions.

Stata output excerpt (1/2)

```
. regress score str
Number of obs   =       420

------------------------------------------------------------------------------
       score | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         str |  -2.279808   .4798256    -4.75   0.000     -3.22298   -1.336637
       _cons |    698.933   9.467491    73.82   0.000     680.3231    717.5428
------------------------------------------------------------------------------

. regress score str, vce(robust)
Number of obs     =         420

------------------------------------------------------------------------------
             |               Robust
       score | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         str |  -2.279808   .5194892    -4.39   0.000    -3.300945   -1.258671
       _cons |    698.933   10.36436    67.44   0.000     678.5602    719.3057
------------------------------------------------------------------------------
```

Stata output excerpt (2/2)

```
. estimates table conventional robust, b(%9.3f) se(%9.3f)
-------------------------------------
    Variable | conventional    robust
-------------+-----------------------
         str |     -2.280      -2.280
             |      0.480       0.519
       _cons |    698.933     698.933
             |      9.467      10.364
-------------------------------------
                          Legend: b/se
```

Robust inference leaves the OLS line unchanged

The OLS minimization problem is unchanged. Changing the variance estimator can change SEs, t statistics, p-values and confidence intervals.

What it does not repair

Omitted-variable bias, an inappropriate functional form, measurement error, and sample selection require separate arguments or methods.

Recalculate an interval under two SE estimates

An illustrative slope is -0.80. The conventional SE is 0.25 and the robust SE is 0.50. Calculate the two approximate 95% intervals and compare their zero-null decisions.

Solution: Recalculate an interval under two SE estimates

$$CI_{conv} = [-1.29, -0.31], \qquad CI_{robust} = [-1.78, 0.18]$$

The point estimate is the same; only the uncertainty calculation changed. The robust interval includes zero.

Stata: inspect what the residual plot can reveal

Look for curvature, unequal spread, or unusual residuals. Which of these concerns the conditional mean and which concerns variance?

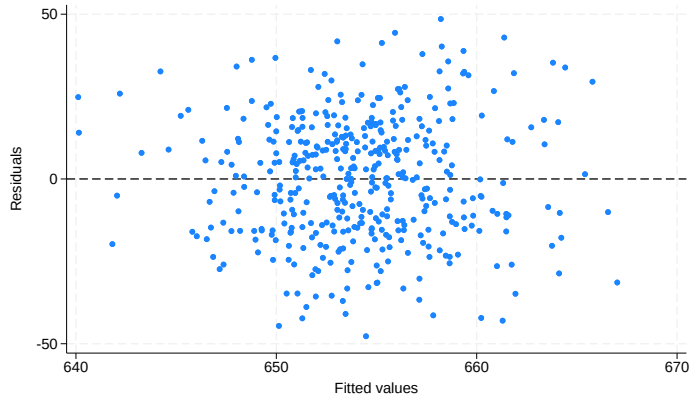
```
rvfplot, yline(0)
```

Read the result

A fan shape suggests unequal variance. Curvature suggests that the fitted mean may need a different functional form.

Stata graph: residuals versus fitted values

Direct export from the displayed Stata graph command



Separate two problems in one residual plot

A residual plot shows both a U-shaped average pattern and wider spread at high fitted values. Which problem does a robust SE address?

Solution: Separate two problems in one residual plot

It addresses the variance calculation for unequal spread. It does not correct the U-shaped conditional-mean pattern; the regression function itself needs attention.

Homoskedastic efficiency is a conditional result

Gauss–Markov

Under the additional homoskedasticity assumption, OLS is the most efficient linear unbiased estimator among a specified class. This is not a claim that OLS is best under every data-generating process.

Course practice

Because constant error variance is often implausible in cross sections, the empirical default is to report robust standard errors and explain the sampling design.

What robust inference changes

Conclusion

A slope's uncertainty depends on the distribution of errors. Robust SEs address unequal conditional variance while leaving the OLS coefficients unchanged.

Could a very small robust SE coexist with a biased class-size slope? Explain.

Quiz 2: finish class here

Time

The quiz occupies the final 15–20 minutes of class. I will announce the stopping time.

1. Open Quiz 2 in Blackboard and answer all 4 questions.
2. The entire quiz is worth 4 points.
3. Submit before time is called; we will discuss the answers next class.