LECTURE 4.1

Heteroskedasticity

ECON30130 Econometrics I

Sam Deegan
University College Dublin

Autumn 2027

Where We Are

Block 4 Breaks the Assumptions of Lecture 2.3

Block 3 added regressors and kept all five of Lecture 2.3's assumptions. Today's lecture drops assumption five, constant error variance, then lets errors correlate. Lecture 4.2 drops assumption four, the zero conditional mean.

Last Week's Lecture

Lecture 3.4 brought qualitative information into the regression:

1. **A Single Dummy Variable.**
   A 0/1 column shifts the intercept and leaves the two lines parallel.
2. **Multiple Categories.**
   Five qualifications need four dummies, each a gap from the base you chose.
3. **Interactions.**
   A product term lets the slopes differ, and the schooling coefficient describes the base group.
4. **Testing a Difference in Slopes.**
   Equal slopes is a test on one coefficient, and 19 graduates cannot reject it.

What Every Standard Error So Far Has Assumed

Every standard error so far assumed one error variance for everybody. So did the F-statistic of Lecture 3.2.

> **Question**
> Multicollinearity widened a standard error in Lecture 3.1 and left the estimate unbiased. Which other faults misstate a standard error, and which bias the estimate itself?

Lecture Plan

1. **Heteroskedasticity.**
   The error variance changes with schooling, and the estimate stays centred on 1.30.

2. **Robust Standard Errors.**
   Squared residuals replace the unseen variances, with HC2 or HC3 in small samples.

3. **Serial Correlation.**
   Correlated errors leave the estimate centred too, and a missing variable is sought first.

> **Wooldridge on Heteroskedasticity and Serial Correlation**
> The reading for this lecture is Wooldridge (2020), sections 8-1 and 8-2 on heteroskedasticity, then 12-1 on serial correlation.

SECTION 1

Heteroskedasticity

The Error Variance Changes with Schooling

Under heteroskedasticity, the error variance (σ_i^2) changes with a person's schooling:

$$\mathrm{Var}(u_i \mid x_i) = \sigma_i^2 \tag{1}$$

Heteroskedasticity Breaks Assumption Five

Assumption five, constant error variance, sets every σ_i^2 to one σ^2.

The Error Fans Out with Schooling

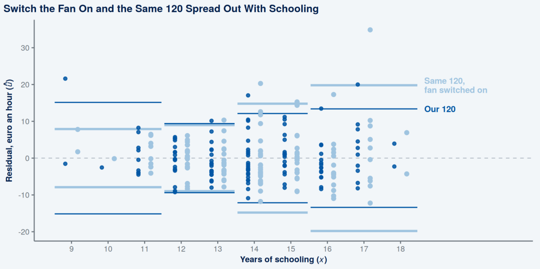

Figure 1: Our 120 (blue) and a counterfactual ghost: the same 120 with the fan on (pale).

> **Spread Widens as Schooling Rises**
> With the fan on, the spread around the line widens with schooling. Across samples the slope still centres on 1.60. That widening spread is heteroskedasticity.

Heteroskedasticity Keeps Only Unbiasedness

Of Lecture 2.3's three promises about the slope, heteroskedasticity keeps only unbiasedness:

- **Unbiased:** The estimate stays centred on the true return of 1.30.
- **Best:** A non-constant error variance costs efficiency.
- **Inference:** The printed standard error stops estimating the slope's spread across samples.

> **Intervals and t-Statistics Use That Standard Error**
> Every interval and *t*-statistic is built from the printed standard error. Those intervals need not cover the true slope 95 times in 100.

Each Fault Against the Promise of Lecture 2.3

Heteroskedasticity and serial correlation leave the estimate unbiased; a wrong shape biases it:

Table 1: Each fault against the promise of Lecture 2.3.

Fault	Still unbiased?	Still best?	Usual SE right?
Heteroskedasticity	Yes	No	No
Serial correlation	Yes	No	No
Wrong shape	**No**	**No**	**No**

Only Wrong Shape Loses Unbiasedness

Unbiasedness needs the error to average zero at every schooling level. Only a wrong shape shifts that average away from zero.

The Variance of the Slope Under Heteroskedasticity

With assumption five broken, each person's own error variance stays inside the sum:

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sum_i (x_i - \bar{x})^2 \, \sigma_i^2}{\underbrace{\left[\sum_i (x_i - \bar{x})^2 \right]^2}_{\text{schooling's squared-deviation sum}}} \tag{2}$$

Each Variance Is Weighted by Distance from the Mean
Each σ_i^2 enters with the weight $(x_i - \bar{x})^2$. People far from 13.45 years count for most.

A Common Error Variance Cancels One Sum

With every σ_i^2 equal to σ^2, the braced sum (SST_x) cancels once:

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2 \sum_i (x_i - \bar{x})^2}{\left[\sum_i (x_i - \bar{x})^2\right]^2} = \frac{\sigma^2}{\underbrace{\sum_i (x_i - \bar{x})^2}_{\mathrm{SST}_x}} \tag{3}$$

Our 120 give $\hat{\sigma}$ of 4.777 and SST_x of 633.9, so $\hat{\sigma}/\sqrt{\mathrm{SST}_x}$ gives Lecture 2.3's 0.1897.

What Heteroskedasticity Breaks: the Slope or Its Standard Error

Our sample's error now fans out with schooling, as in the fan plot. We fit the usual line and print its classical standard error.

> **Question**
> Name what the fan breaks, then name what the fan leaves alone. Does the printed standard error overstate or understate the slope's spread across samples?

The Printed Standard Error Can Miss in Either Direction

The fan leaves the slope centred on 1.30. Where the large variances sit decides which way the printed standard error misses.

> **Understated When Large Variances Sit Far From 13.45**
> Large variances far from 13.45 years make the printed number understate the spread. Large variances near 13.45 make it overstate the spread.

Nobody knows the direction in advance (Wooldridge, 2020, p. 266).

SECTION 2

Robust Standard Errors

Estimating the Slope Variance with Squared Residuals

Nobody sees any person's error variance. The robust estimator replaces each σ_i^2 with the squared residual:

$$\widehat{\text{Var}}(\hat{\beta}_1) = \frac{\sum_i (x_i - \bar{x})^2 \, \hat{u}_i^2}{\left[\sum_i (x_i - \bar{x})^2 \right]^2} \tag{4}$$

The Error and the Residual

The error u_i belongs to the population, and nobody ever sees it. The residual $\hat{u}_i$ belongs to the fit, and we compute one per person. The hat marks a quantity computed from our 120, not the population.

Noise in Each Squared Residual Averages Out Across 120 People

One squared residual carries enormous sampling noise as an estimate of one variance.

Only the Weighted Total Enters the Formula

The robust formula uses the squared residuals only through one weighted total. Errors in the individual terms average out across the 120.

Robust Standard Errors Need No Prior Test

The robust standard error is the square root of that estimate (White, 1980). White's standard error is valid whether or not the error variance is constant.

> **Correcting After a Test Hides the Test**
> A standard error chosen after a test depends on a test nobody reported. A robust standard error chosen from the start depends on no test.

Eicker (1967) and Huber (1967) had pointed to such standard errors in 1967.

Classical and Robust Standard Errors as the Fan Steepens

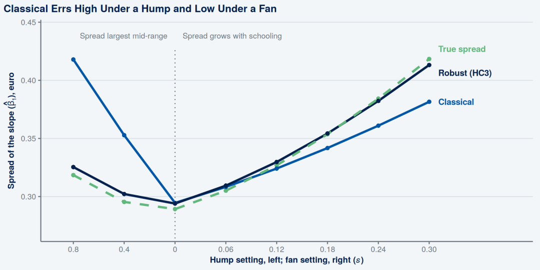

Figure 2: Averages over 1,000 repeated samples per setting: hump on the left, fan on the right.

> **The Classical Standard Error Misses Both Ways**
> The slope centres on 1.61 at every setting. Under a fan of 0.30, classical understates: 0.382 against a true 0.418. Under a hump of 0.8, it overstates: 0.418 against 0.318. HC3 stays within 0.007 of the truth.

Angrist and Pischke Make Robust Standard Errors the Default

Angrist & Pischke (2017) argue that most of introductory inference reduces to one instruction:

Report Robust Standard Errors

That instruction is to report robust standard errors. The paper would give assumption tests and generalised least squares no class time.

Which Robust Variant to Use by Default

The usual robust estimator carries its own bias in small samples.

> **Small Samples Need HC2 or HC3**
> Long & Ervin (2000) and Imbens & Kolesár (2016) both recommend the corrected variants, HC2 and HC3. Both were derived by MacKinnon & White (1985) from each observation's leverage. HC2 and HC3 scale up the squared residuals of high-leverage observations.

Robust standard errors belong in the default, and so does the corrected variant.

What a Steeper Fan Does to the Classical and Robust Standard Errors

Fit one sample with the fan off, then refit with the fan on.

> **Question**
> The error fans out more steeply. Which changes: the estimate, the classical standard error or the robust one?

A Steeper Fan Leaves the Slope Estimate Centred

A steeper fan leaves the estimate centred on 1.30. The gap between the robust and classical standard errors grows with the fan.

> **Curvature Shifts the Estimate Away From 1.30**
> The lab for this lecture turns on curvature instead. Both standard errors stay as narrow as they were with the fan off. The estimate itself stops being centred on 1.30.

SECTION 3

Serial Correlation

Correlated Errors Have the Same Consequences as Heteroskedasticity

Correlated errors cost efficiency and a correct standard error, exactly as heteroskedasticity does:

- **Unbiased:** The estimates stay centred on the true return of 1.30.
- **Standard Error:** The printed standard error stops estimating the standard deviation of $\hat{\beta}_1$ across samples.

Serial Correlation Arises Between Observations That Share a Unit

Serial correlation is dependence between observations that share a unit of clustering:

- **Time:** adjacent quarters share a time.
- **Firm:** one firm's workers share a firm.
- **Place:** one town's people share a place.

> **Random Sampling Rules Out Correlation Between Independent Draws**
> Random sampling already leaves different people's errors uncorrelated. Lecture 2.3's variance derivation drops the cross terms for that reason.

A Shared Town Shock Leaves Runs of Same-Signed Residuals

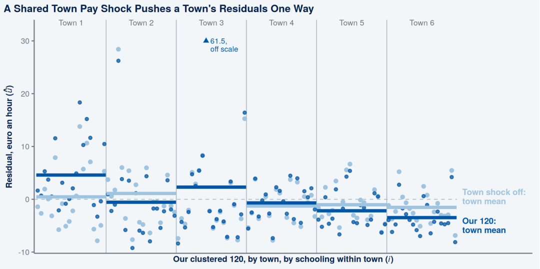

Figure 3: Clustered 120 (blue); counterfactual ghost, same people, town shock off (pale).

> **People in One Town Share a Pay Shock**
> Everyone in a town shares its earnings shock, so their errors are correlated. Town 1's residuals average +4.59 and town 6's −3.47. In repeated clustered samples, such runs appear 32 times in 100. With the shock off, they appear 6 times.

Why Correcting the Standard Error Comes Last

Newey & West (1987) give a standard error valid under both heteroskedasticity and autocorrelation. Applied first, that correction widens the interval around an unjustified coefficient.

> **Correlated Residuals Usually Mean a Missing Variable or Shape**
> Correlated residuals usually come from an omitted variable or a wrong functional form. Naming the variable or the form comes before any corrected standard error.

Two Unrelated Series Gave a Significant t-Statistic

Granger & Newbold (1974) regressed one time series on a second, unrelated series:

- **The Test:** the t-statistic rejected a true null far more often than 5 times in 100.
- **The Residuals:** the residuals were correlated from one period to the next.

Those correlated residuals were the only sign that the regression was spurious.

What Would Cause a Run of Same-Signed Residuals

The runs plot put one town's residuals above zero and another's below.

> **Question**
> Name two different faults that would produce that plot. Then name the evidence that would tell those two faults apart.

Three Faults Produce the Same Runs

Three faults produce the same runs, and the plot alone cannot separate them:

- **Correlated Errors:** errors within a shared unit move together.
- **Omitted Variable:** a missing variable runs in a pattern across schooling.
- **Wrong Shape:** the runs plot fits a straight line to a curve.

Evidence That Separates the Faults

Runs that vanish once a squared term is added point to the shape. Residuals that track a candidate variable point to that variable. Correlation inside a shared unit points to correlated errors.

What Would Have to Be True?

A robust standard error now sits beside our estimate of 1.5227. About 95 in every 100 intervals built this way contain the true slope.

> **Question**
> What would have to be true for this number to mean what we just said it means? And what is the plausible story where it isn’t?

The Interval Still Needs Assumption Four

The 95 in 100 still needs assumption four, the zero conditional mean.

> **Omitted Ability Leaves Residuals with No Pattern**
> Suppose a missing variable rises with schooling and raises earnings. The residuals then show no pattern. Least squares puts that variable's schooling-related part into the slope.

A robust standard error says nothing about whether schooling raised those earnings.

Summary: Heteroskedasticity, Robust Standard Errors and Serial Correlation

Today sorted two faults and one repair against Lecture 2.3's promise:

1. **Heteroskedasticity.**
   The estimates stay centred on 1.30, and the printed standard error stops estimating their spread across samples.

2. **Robust Standard Errors.**
   White's estimator holds with or without a constant variance, and HC2 or HC3 suits small samples.

3. **Serial Correlation.**
   Correlated errors also leave the estimate centred, and usually mean an omitted variable.

Before Lecture 4.2

1. **Play:**
   The app. Dial the fan up and watch the classical and robust standard errors separate.

2. **Apply:**
   Problem set 4, covering 4.1, with one question back from problem set 3. Then the appendix derivation.

3. **Review:**
   Wooldridge (2020), sections 8-1, 8-2 and 12-1, then chapter 9 before Lecture 4.2.

> **Problem Set 4**
> [Out this week, due week n.]

Next Lecture

Lecture 4.2 turns from the standard error to faults that bias the estimate:

1. **Functional Form and Outliers.**
   A wrong shape costs unbiasedness, and the residual plot shows it.
2. **Ability, Selection and Measurement Error.**
   Ability raises the estimate above 1.30; reporting noise pulls it toward zero.
3. **Reverse Causality.**
   Regression is symmetric, causation is not; the output never records which way you meant.
4. **The Research Designs of Part II.**
   Instruments, differencing and cutoffs each answer a fault no diagnostic detects.

Thank you

sam.deegan@ucdconnect.ie
sam-deegan.com

BACKUP

Appendix

Notation Used in This Lecture

Symbol	**What it is**	**First met**
i	the index running over the people in the sample	Lecture 1.1
x_i	years of schooling of person i	Lecture 1.1
$\bar{x}$	the mean years of schooling of our 120, 13.45	Lecture 1.1
β_1	the population slope, the true return to a year of schooling, 1.30	Lecture 2.1
$\hat{\beta}_1$	the least-squares slope from our 120 people, 1.5227	Lecture 2.1
u_i	person i's error: the part of earnings the population line misses	Lecture 2.1
$\hat{u}_i$	person i's residual: earnings minus the fitted value	Lecture 2.1
R^2	the share of the variation in earnings the fitted line accounts for	Lecture 2.2
w_i	the least-squares weight person i carries	Lecture 2.3
σ^2	the variance of the error, the same at every level of schooling	Lecture 2.3
σ	the error standard deviation	Lecture 2.3
$\hat{\sigma}$	the estimated error standard deviation, 4.777 on our 120	Lecture 2.3
SST_x	the sum of squared deviations of schooling, 633.9 in our sample	Lecture 2.3
t	an estimate minus its null value, over its standard error	Lecture 3.2
F	the statistic that tests several restrictions at once	Lecture 3.2

Notation Used in This Lecture

Symbol	What it is	First met
$\mathrm{Var}(\cdot)$	the variance of whatever is inside	Lecture 1.2
$\mathrm{Var}(u_i \mid x_i)$	the variance of the error given schooling	Lecture 2.3
σ_i^2	the error variance for person i	The Error Variance Changes With Schooling
$\widehat{\mathrm{Var}}(\hat{\beta}_1)$	the robust estimate of the slope's variance	Estimating the Slope Variance With Squared Residuals

Derivation: Why Heteroskedasticity Breaks the Variance Formula

Lecture 2.3 wrote the slope as the truth plus a weighted sum of errors. This derivation takes that sum's variance without assuming the errors share a common variance. That refusal gives the heteroskedastic variance formula. Putting $\hat{u}_i^2$ where σ_i^2 stands gives the robust estimator.

> **Lecture 2.3 Pulled a Common Variance Out in One Line**
> The derivation runs in three steps. The slope is β_1 plus a sum of weights w_i times errors. Its variance is the sum of squared weights times the individual error variances. The derivation stops there, because the errors share no common σ^2 to pull out. Lecture 2.3 pulled one out in a single line without naming the step.

[Full derivation. Set as homework; worked solution published with the set.]

Notes and Tools

The residual plot and each robust variant take one call in R or Stata:

- `plot(model, which = 1)`: residuals against fitted values, which is the first plot to run.
- `sandwich::vcovHC()` with `lmtest::coeftest()`, documented in
  1. Its default is HC3.
- `estimatr::lm_robust()`, which reports HC2 with adjusted degrees of freedom by default.
- `fixest::feols()` with `vcov = "hetero"`, if you are already using it.
- Stata's `vce(robust)` is HC1, not the HC3 that `sandwich` uses by default. Say which variant you used.

References

Angrist, J. D., & Pischke, J.-S. (2017). Undergraduate econometrics instruction: Through our classes, darkly. *Journal of Economic Perspectives*, *31*(2), 125–144. https://doi.org/10.1257/jep.31.2.125

Eicker, F. (1967). Limit theorems for regressions with unequal and dependent errors. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, volume 1: statistics* (pp. 59–82). University of California Press.

Granger, C. W. J., & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, *2*(2), 111–120. https://doi.org/10.1016/0304-4076(74)90034-7

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, volume 1: statistics* (pp. 221–233). University of California Press.

Imbens, G. W., & Kolesár, M. (2016). Robust standard errors in small samples: Some practical advice. *The Review of Economics and Statistics*, *98*(4), 701–712. https://doi.org/10.1162/REST_a_00552

Long, J. S., & Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*, *54*(3), 217–224. https://doi.org/10.1080/00031305.2000.10474549

References

MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, *29*(3), 305–325. https://doi.org/10.1016/0304-4076(85)90158-7

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703–708. https://doi.org/10.2307/1913610

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*(4), 817–838. https://doi.org/10.2307/1912934

Wooldridge, J. M. (2020). *Introductory econometrics: A modern approach* (7th ed.). Cengage Learning.