LECTURE 3.1

Multiple Regression: Estimation

ECON30130 Econometrics I

Sam Deegan
University College Dublin

Autumn 2027

Where We Are

> **The First Lecture with Two Regressors**
> Block 2 fitted one line and measured how far its slope sits from the truth. Today puts a second variable into that line. Lecture 3.2 then tests its coefficients.

Last Week's Lecture

Lecture 2.3 asked how far the fitted slope lands from the truth:

1. **The Gauss-Markov Assumptions.**
   Five claims about the population, and the weight each person carries in the slope.
2. **Unbiasedness.**
   Why assumption four makes the fitted slope centre on the truth.
3. **Precision.**
   How far the slope varies across samples, and the three things that narrow it.
4. **Consistency and Asymptotic Efficiency.**
   The slope concentrates on the truth, and least squares stays best in a wider class in the limit.

Ability Raises Pay and Is Missing from the Data

Block 2's 1.5227 assumed the error averages to zero at each schooling level. Today ability raises pay, rises with schooling, and is missing:

Question

Your estimate leaves out ability, which raises pay and raises schooling. Is the return you report too large or too small? Answer before the arithmetic.

Lecture Plan

1. **Partialling Out.**
   A coefficient is the slope on the part of its variable the others leave over.

2. **Reading a Partial Coefficient.**
   What controlling for a confounder, a mediator or a collider does.

3. **Omitted Variable Bias.**
   Leaving a variable out shifts the slope by a product of two slopes.

4. **Multicollinearity.**
   Correlated regressors widen standard errors and leave the estimates unbiased.

SECTION 1

Partialling Out

The Two-Regressor Model for Earnings and Schooling

A second slope, on ability, lets the model compare people of equal ability:

$$y_i = \beta_0 + \beta_1(x_i - x_0) + \gamma\, a_i + u_i \tag{1}$$

- x_0: the reference level of schooling, thirteen years in our population
- a_i: ability of person i, on the generator's own scale
- γ: the return to a unit of ability at fixed schooling

β_1 Now Holds Ability Fixed

β_1 is the return to schooling among people of equal ability. Lecture 2.3's β_1 held nothing fixed. The name is the same; the quantity and its assumptions differ.

Two Assumptions Change When a Regressor Is Added

The five assumptions carry over as MLR.1 to MLR.5, and two of them change (Wooldridge, 2020):

- **MLR.3:** neither regressor is constant, and neither is an exact linear function of the other.
- **MLR.4:** the error averages to zero given schooling and ability together, $E(u \mid x, a) = 0$.

> **MLR.3 Allows Correlated Regressors**
> MLR.3 rules out perfect collinearity only. Schooling and ability may be correlated, and in today's population they are.

Ability Is Available Only Because We Wrote the Population

Our ability column exists only because the generator wrote it:

- **Real Data:** no survey of Irish earnings carries a usable measure of ability.
- **Our 120:** we can regress earnings on schooling with and without ability.

> **A Simulation Is Not an Argument About Ireland**
> We wrote this population rather than observed it, so every number today is checkable. A written population verifies a formula and nothing more. A simulation is no evidence about the Irish return to schooling.

What Makes a Regressor a Control

Ability enters today's regression for the sake of the schooling coefficient:

> **A Control Serves Another Coefficient**
> A control is a regressor included for the sake of another coefficient. The control makes that coefficient a comparison at fixed values of itself.

No arithmetic marks a control, so you have to argue that claim.

The Part of Schooling That Ability Cannot Account For

A regression gives schooling only the credit ability has not already taken. The first step regresses schooling on ability:

$$x_i = \overbrace{\hat{\pi}_0 + \hat{\pi}_1 a_i}^{\text{the fitted value}} + r_i \tag{2}$$

- r_i: the part of person i's schooling that ability leaves over
- $\hat{\pi}_1$: the slope from regressing schooling on ability

> **Regression Anatomy Splits a Regressor in Two**
> Regression anatomy is this split of a regressor into fitted value and residual. The name and the phrase "other things equal" are Angrist & Pischke (2015)'s.

The Leftover Schooling Is Uncorrelated with Ability

The auxiliary fit sets two sums to zero, as least squares did in Lecture 2.1:

$$\sum_{i=1}^{n} r_i = 0 \qquad \sum_{i=1}^{n} a_i r_i = 0 \tag{3}$$

> **Centred on Zero, Uncorrelated with Ability**
> The first sum puts the mean of r_i at zero. With that mean at zero, the second leaves r_i uncorrelated with ability (Wooldridge, 2020).

The Two-Step Residual Recipe on Our 120 People

The coefficient on schooling is Lecture 2.1's slope formula with r_i in place of schooling:

$$\hat{\beta}_1 = \frac{\overbrace{\sum_{i=1}^{n} r_i y_i}^{\text{the cross-product}}}{\sum_{i=1}^{n} r_i^2} \tag{4}$$

The Recipe and the Full Fit Both Give 1.1022

Earnings on schooling alone gives 1.5201. Earnings on schooling and ability gives 1.1022. The two-step recipe gives 1.1022 as well.

Each Demonstration Today Is Its Own Simulated World

Each demonstration today runs the generator again with one switch on:

- **Its Own Population:** every run draws its own 200,000 people and its own 120.
- **No Before and After:** numbers from two runs are never one regression under different controls.

Why the Two Steps Return the Same Coefficient

The long fit writes each person's earnings exactly, residual included:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1(x_i - x_0) + \hat{\gamma}\, a_i + \hat{u}_i \tag{5}$$

Multiplying by r_i and summing over the 120 splits the recipe's numerator in four:

$$\sum_i r_i\, y_i = \hat{\beta}_0 \sum_i r_i + \hat{\beta}_1 \sum_i r_i(x_i - x_0) + \hat{\gamma} \sum_i a_i r_i + \sum_i r_i \hat{u}_i \tag{6}$$

Three of the Four Sums Are Zero

The auxiliary fit's two conditions and the long fit's own conditions remove three sums:

$$\sum_i r_i y_i = \hat{\beta}_0 \underbrace{\sum_i r_i}_{=0} + \hat{\beta}_1 \sum_i r_i (x_i - x_0) + \hat{\gamma} \underbrace{\sum_i a_i r_i}_{=0} + \underbrace{\sum_i r_i \hat{u}_i}_{=0} \tag{7}$$

Least Squares Had Already Set the Last Sum to Zero

The long fit's residuals are orthogonal to every regressor. The residual r_i is built from schooling, ability and a constant. Orthogonality is a first-order condition, not an assumption about the world (Wooldridge, 2020).

Only the Leftover Variation Survives

The fitted part of schooling is uncorrelated with r_i, so the surviving sum is $\sum r_i^2$:

$$\sum_i r_i y_i = \hat{\beta}_1 \underbrace{\sum_i r_i (x_i - x_0)}_{=\sum_i r_i^2} \quad \Longrightarrow \quad \hat{\beta}_1 = \frac{\sum_i r_i y_i}{\sum_i r_i^2} \tag{8}$$

The Recipe Holds in Every Sample

No step used the values of our 120 people. Any sample returns the same coefficient both ways, to every decimal.

Raw and Residualised Scatterplots Side by Side

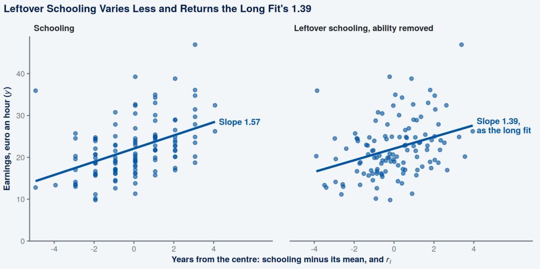

Figure 1: Our 120: earnings on schooling (slope 1.57) and on leftover schooling (1.39).

> **Leftover Schooling Varies Less Than Schooling**
> Ability has already used some of schooling's variation, widening the standard error on schooling.

SECTION 2

Reading a Partial Coefficient

What Holding Other Things Fixed Actually Covers

Multiple regression mimics a comparison at equal ability, though nobody sampled our 120 that way (Wooldridge, 2020):

$$\Delta\hat{y} = \hat{\beta}_1\,\Delta x \qquad \text{when} \qquad \Delta a = 0 \tag{9}$$

- $\Delta\hat{y}$: the difference in fitted earnings between two people
- Δx: the difference in their schooling

Holding Fixed Covers Only the Variables You Measured
Variables you did not measure are free to differ between two such people. The comparison is arithmetic, not evidence about what would happen if schooling rose.

Each Coefficient Holds a Different Variable Fixed

Each slope in a two-regressor table holds the other regressor fixed:

- $\hat{\beta}_1$ compares people who differ in **schooling** and agree on ability.
- $\hat{\gamma}$ compares people who differ in **ability** and agree on schooling.

> **Two Rows Answer Two Different Questions**
> Slopes that hold different variables fixed are not comparable with each other. Reading every row as its own effect is the commonest misreading of a regression table.

Why Only One Coefficient in Your Table Is Interpretable

Only the schooling coefficient was designed:

- **Schooling:** you put ability in to make assumption four credible for schooling.
- **Ability:** nothing in the model was chosen to make assumption four credible for ability.
- **A Route Left Out:** able people stay in school longer, and $\hat{\gamma}$ holds schooling fixed.

Read One Coefficient Causally

Report the rest as the adjustments that made it credible. A control's own coefficient rarely answers a question you asked, and is printed anyway.

Confounder, Mediator and Collider in One Causal Diagram

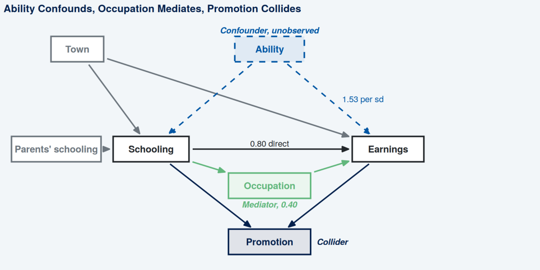

Figure 2: This population: ability causes both, occupation sits between, both cause promotion.

> **One Command, Three Different Consequences**
> The same `lm()` call adds all three, and the regression cannot tell them apart. Their places in the diagram are claims about the world.

Confounders: Controlling Removes a Bias

Ability raises schooling, and raises earnings for reasons unconnected to school:

- **Schooling Alone:** the regression credits ability's share of earnings to schooling, 1.5201.
- **With Ability:** the confounder's share comes back out, and the coefficient is 1.1022.

What Makes Ability a Confounder

Ability causes both schooling and earnings. Neither schooling nor earnings causes ability — genetics, upbringing and luck do.

Mediators: Controlling Removes Part of the Effect You Wanted

On a run with the occupation switch on, schooling raises pay partly through occupation:

With occupation left out the coefficient is 2.5041. Control for occupation and the coefficient falls to 0.8936. Roughly two thirds of it is gone.

> **Occupation Is Part of the Return**
> Occupation is not a nuisance to be removed. The missing two thirds is the part of the return that runs through occupation.

The Direct Effect Is Not the Total Return to Schooling

0.8936 is the return to schooling among people in the same occupation:

> **A Direct Effect on One Condition**
> 0.8936 earns the name direct effect only if nothing unmeasured causes both occupation and pay.
> Otherwise holding occupation fixed also creates a collider bias.

0.8936 is a real quantity, and it is not the total return to schooling.

Colliders: Controlling Creates a Bias That Was Not There

On a third run, schooling and earnings both raise promotion, which raises neither:

With promotion left out, the coefficient on schooling is 1.1020. Control for promotion and the coefficient is -0.9125. The sign reverses.

> **Controlling for Everything You Have Is Not a Defence**
> Adding promotion reversed the sign of a return whose true value is 1.30. A defence names each control and says which of the three it is. No dataset answers that question.

Holding Promotion Fixed Makes Schooling and Earnings Trade Off

Promotion here is an index built from schooling and earnings, so either can raise it:

- **At One Promotion Value:** people with less schooling have more earnings.
- **In the Regression:** holding promotion fixed reads that trade-off as a negative return.

Why the Collider Bias Is Not Sampling Noise

Both regressions ran on the same rows, and the promotion control is their only difference:

- **Nothing Omitted:** nothing was left out of the second regression or mismeasured in either.
- **Same Rows:** the gap between 1.1020 and -0.9125 is no sampling difference.

Holding Promotion Fixed Creates a Link Nothing Caused

No arrow in the diagram runs from schooling to earnings through promotion. The association appears only inside the regression that holds promotion fixed. That association would appear in a sample of any size (Elwert & Winship, 2014).

What Controlling Does in Each of the Three Cases

All three correlate with schooling; only the confounder does what “adding controls” promises:

- **Confounder:** controlling **removes a bias**, because the common cause’s share comes out.
- **Mediator:** controlling **removes part of the effect you wanted**, the mediated route.
- **Collider:** controlling **creates a bias**, because a common consequence induces an association.

Causal Order Is Argued, Not Fitted

All three outputs look identical: a coefficient, a standard error and a star. Causal order settles which case you are in. You argue it before you fit, and cannot recover it afterwards.

Whether Promotion Belongs in the Regression

People high on the promotion index both earn more and have more schooling.

> **Question**
> Should the promotion index go into the regression? What fact about the world, rather than about the data, would settle it?

Two Arrows Decide Whether Promotion Belongs

On this generator promotion stays out, because of where its arrows point:

- **Mediator:** if schooling raises promotion and promotion raises pay, controlling removes part of the return.
- **Collider:** if schooling and earnings both raise promotion, controlling creates a bias.

> **No Test on These 120 Rows Separates Them**
> Both stories fit the same correlations. A rise in R-squared is a fact about the data. So is a changed schooling coefficient.

SECTION 3

Omitted Variable Bias

The Regression We Ran and the Regression We Wanted

Real data lack the ability column, and leaving ability out breaks assumption four (Wooldridge, 2020):

$$\text{long:} \quad y_i = \hat{\beta}_0 + \hat{\beta}_1(x_i - x_0) + \hat{\gamma}\, a_i + \hat{u}_i \qquad \text{short:} \quad y_i = \tilde{\beta}_0 + \tilde{\beta}_1(x_i - x_0) + \tilde{u}_i \tag{10}$$

- $\tilde{\beta}_1$: the slope on schooling from the regression that left ability out

> **A Tilde Marks an Estimate with a Variable Left Out**
> A hat marks an estimate, as it has since Lecture 1.3. A tilde marks an estimate from a regression that left a variable out. Only the hatted slope estimates the quantity we said we wanted.

Deriving the Bias in Sample Algebra

Both regressions run on the same 120 rows, so the gap between them is arithmetic:

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})\, y_i}{\underbrace{\sum_{i=1}^{n} (x_i - \bar{x})^2}_{\mathrm{SST}_x}} \tag{11}$$

Each person's earnings equal the long fit plus its residual, so that fit replaces y_i:

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) \left(\hat{\beta}_0 + \hat{\beta}_1 (x_i - x_0) + \hat{\gamma}\, a_i + \hat{u}_i\right)}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{12}$$

Each Sum Reduces to Zero, One or a Slope

Two facts from Lecture 2.3 and the long fit's conditions settle three sums:

$$\tilde{\beta}_1 = \hat{\beta}_0 \underbrace{\frac{\sum_i (x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2}}_{=0} + \hat{\beta}_1 \underbrace{\frac{\sum_i (x_i - \bar{x})(x_i - x_0)}{\sum_i (x_i - \bar{x})^2}}_{=1} + \hat{\gamma} \underbrace{\frac{\sum_i (x_i - \bar{x})\, a_i}{\sum_i (x_i - \bar{x})^2}}_{=\hat{\delta}} + \underbrace{\frac{\sum_i (x_i - \bar{x})\, \hat{u}_i}{\sum_i (x_i - \bar{x})^2}}_{=0} \qquad (13)$$

- $\hat{\delta}$: the slope from regressing ability on schooling

> **The Auxiliary Slope Points the Other Way**
> $\hat{\delta}$ regresses ability on schooling. Section 1's $\hat{\pi}_1$ regressed schooling on ability. The two point opposite ways and take different values.

The Bias Is a Product of Exactly Two Slopes

Only $\hat{\beta}_1$ and one product survive the four sums:

$$\tilde{\beta}_1 = \hat{\beta}_1 + \underbrace{\hat{\gamma}\,\hat{\delta}}_{\text{the omitted variable bias}} \tag{14}$$

The Bias Runs Through Two Slopes
$\hat{\delta}$ converts schooling into ability, and $\hat{\gamma}$ converts ability into pay. The bias runs through both links, so it is their product (Wooldridge, 2020).

Worked example: our 120 people

Predicting the Short Estimate Before Moving the Sliders

The app's two sliders set how far the omitted variable raises schooling and pay.

> **Question**
> Both sliders sit at their midpoint. Is the short estimate above the long one, below it, or the same? Answer again for either slider set to zero.

The Signs of the Two Slopes Set the Sign of the Bias

The sign of the product sets which way the short slope misses (Wooldridge, 2020):

- **Same Signs:** both positive or both negative, and the short slope is too large.
- **Opposite Signs:** one of each, and the short slope is too small.
- **A Zero Slope:** either slider at zero closes the gap completely, not by half.

> **Leaving Ability Out Makes the Return Too Large**
> Ability raises pay, and able people stay in school longer. Both slopes are positive, so the short slope sits above the long one.

Block 2 and Today's Ability Run Are Different Populations

Block 2's sample and today's ability run come from different populations:

Sample	Slope on schooling	Standard error	R-squared
Block 2, every switch off	1.5227	0.1897	0.3531
Ability run, ability left out	1.6700	0.1652	0.4642

> **These Two Rows Are Not Before and After**
> Block 2's true return was 1.30, in a population of its own. Section 1's 1.5201 and the mediator and collider runs carry the same warning.

Short and Long Slopes Across Returns to Ability

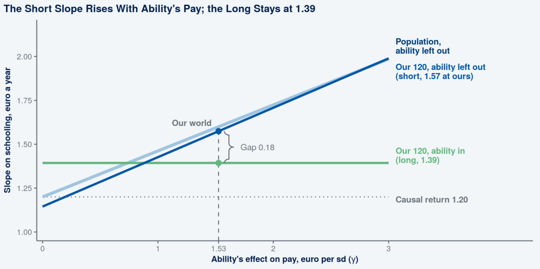

Figure 3: Counterfactual sweep of ability's pay on our 120: short slope rises, long stays.

> **Only the Short Regression's Line Rises**
> The long regression holds ability fixed, so ability's return cannot reach it. The short regression has nowhere to put ability but the schooling coefficient.

More Data Shrinks the Standard Error, Not the Bias

A hundredfold sample divides the short regression's standard error of 0.1652 by about ten:

- **The Spread:** narrows around 1.6700, not around the long regression's 1.2538.
- **The Gap:** 0.4162 in our 120, and it does not shrink as the sample grows.

No Sample Size Removes an Omitted Variable

A bias that shrank with more people would be a nuisance. Omitted variable bias does not shrink, so only the missing column removes it.

Only the Hats Come Off as the Sample Grows

Replacing each hat with its population slope gives the short slope's large-sample limit:

$$\text{in our 120:} \quad \tilde{\beta}_1 = \hat{\beta}_1 + \hat{\gamma}\,\hat{\delta} \qquad\qquad \text{as the sample grows:} \quad \operatorname{plim} \tilde{\beta}_1 = \beta_1 + \gamma\,\delta \tag{15}$$

- δ: the population slope of ability on schooling

> **The Short Slope Is Inconsistent**
> Lecture 2.3 called a rule consistent when it concentrates on its target. The short slope concentrates on $\beta_1 + \gamma\delta$ instead (Wooldridge, 2020).

An Objector Owes Us Two Slopes Whose Product Is 0.6700

Suppose someone says our 1.6700 should be one euro, so the bias would be 0.6700:

- **Many Pairs:** many pairs of slopes multiply to 0.6700.
- **Two Claims:** each slope is arguable on its own, and sometimes checkable.

The Two Ingredients Are Arguable Separately

Does the missing variable raise pay? Do the people who have it stay in school longer? Evidence can answer those two questions. Arguing about whether a study is credible cannot.

Different Slope Pairs Giving the Same Bias

Holding either of our two slopes fixed shows what the other would need to be:

- **Keep $\hat{\gamma}$ = 1.8867:** the auxiliary slope must rise from 0.2206 to 0.3551.
- **Keep $\hat{\delta}$ = 0.2206:** the return to ability must rise from 1.8867 to 3.0372.

> **The Return to Ability Is the Harder Claim**
> Almost everyone agrees that ability is worth something in pay. Three euro an hour per unit is a different matter. Nobody has measured that scale.

SECTION 4

Multicollinearity

Two Regressors Carrying Nearly the Same Information

Keeping ability in avoids that bias but widens the standard error on schooling:

> **Multicollinearity Is High Correlation Among Regressors**
> High but not perfect correlation between regressors is multicollinearity. Multicollinearity violates none of MLR.1 to MLR.5 (Wooldridge, 2020).

Little leftover schooling means a large standard error on schooling.

The Variance of the Schooling Coefficient with a Control

The slope's variance puts the population error variance (σ^2) over our 120's leftover schooling:

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n} r_i^2} = \frac{\sigma^2}{\underbrace{\sum_{i=1}^{n}(x_i - \bar{x})^2}_{\mathrm{SST}_x}} \times \underbrace{\frac{1}{1 - R_1^2}}_{\text{the variance inflation factor}} \tag{16}$$

- R_1^2: the R-squared from regressing schooling on the other regressors

Any Control That Predicts Schooling Widens the Standard Error

The inflation factor equals one only when nothing else predicts schooling:

- $R_1^2 = 0$**:** the factor is one, and the variance is Lecture 2.3's σ^2/SST_x.
- $R_1^2 > 0$**:** the factor rises above one, so the standard error on schooling gets larger.

> **Ability Removes Bias and Adds Variance**
> Drop ability again and the standard error narrows, but section 3's bias returns (Wooldridge, 2020).

A Control That Almost Perfectly Predicts Schooling

Suppose ability predicts schooling almost exactly in our 120, so every r_i sits near zero. Nothing about earnings has changed.

> **Question**
> What happens to the standard error on the coefficient for schooling? Name the part of the formula you reasoned about to get there.

The Standard Error Grows Without Limit Near Perfect Prediction

Every r_i near zero pushes R_1^2 towards one, and the inflation factor grows without limit:

$$R_1^2 \to 1 \quad \Longrightarrow \quad \frac{1}{1 - R_1^2} \to \infty \tag{17}$$

> **A Better Fit to Schooling Widens the Standard Error**
> The more of schooling ability accounts for, the wider the standard error on schooling. At $R_1^2 = 1$ exactly, $\hat{\beta}_1$ cannot be computed, and MLR.3 rules that case out (Wooldridge, 2020).

Standard Error and Estimate as the Correlation Rises

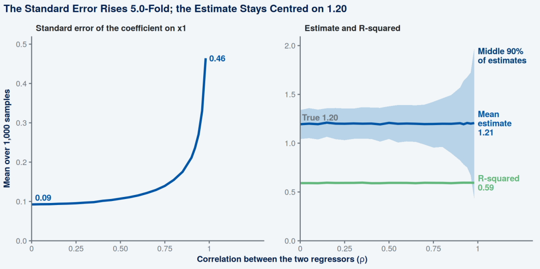

Figure 4: Constructed, not our sample: correlation 0 to 0.98, true coefficient fixed at 1.20.

> **The Standard Error Climbs Steeply Near One**
> The estimate stays centred on the true coefficient, and R^2 hardly changes.

Collinearity Widens Standard Errors Without Breaking Assumptions

Goldberger (1991) coined *micronumerosity*, the problem of small sample size, to mock the attention collinearity gets (Wooldridge, 2020):

- **Collinearity:** a large R_1^2 raises the variance of $\hat{\beta}_1$.
- **Micronumerosity:** a small SST_x, from few people, raises it too.

> **Nothing Is Broken and the Standard Errors Are Wide**
> Collinearity breaks no assumption, and Gauss-Markov still holds. The estimates are unbiased, and the fitted values and R^2 are hardly touched. The standard errors still tell you how far each estimate typically sits from the truth. That distance is large.

Why Dropping a Collinear Regressor Introduces Bias

The usual instinct is to drop one, which narrows the survivor's standard error:

- **Common Practice:** Lindner et al. (2020) trace how routinely published work does exactly that.
- **Other Remedies:** Vanhove (2021) works through the rest of the standard remedies.

> **Dropping One Regressor Brings Back the Bias**
> The two regressors move together, so the survivor absorbs the dropped one's effect. The survivor then carries section 3's omitted variable bias. Its narrower spread centres on the partial effect plus that bias.

Variance Inflation Factors and Their Conventional Thresholds

Some applied work calls collinearity a problem once the inflation factor passes ten (Wooldridge, 2020):

- **Variance Inflation Factor:** how many times larger the variance is than with uncorrelated regressors.
- **Five and Ten:** conventions rather than results (O'Brien, 2007).
- **What Narrows the Standard Error:** Lecture 2.3's answer, more people or more spread in schooling.

Collinearity Is in Your Data, Not in Your Model

Wide standard errors report what the data do not know. An inflated standard error matters only against the question you asked.

What Would Have to Be True?

Our claim: 1.2538 is the return to schooling among people of equal ability.

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?

Three Stories That Would Break 1.2538

Assumption four needs no other omitted variable that both raises pay and predicts schooling:

- **Family Income:** could pay for longer schooling and for the contacts that raise pay.
- **School Quality:** could raise both schooling and pay.
- **Persistence:** whatever makes someone finish what they start could raise both.

> **Each Story Needs Both Slopes**
> Each candidate needs its own $\hat{\gamma}$ and $\hat{\delta}$. A story with only one of the two adds no bias.

Summary: Partialling Out, Controls, Bias and Collinearity

1. **Partialling Out.**
   A coefficient is the slope on its variable's leftover part, 1.1022 either way.

2. **Reading a Partial Coefficient.**
   Controlling removes a bias, removes part of the effect, or reverses the sign.

3. **Omitted Variable Bias.**
   Short minus long is $\hat{\gamma}\hat{\delta}$, here $0.2206 \times 1.8867 = 0.4162$.

4. **Multicollinearity.**
   A large R_1^2 widens the standard error and biases nothing.

Before Lecture 3.2

1. **Play:**
   The app. Set the return to ability to zero, then try to restore the bias.

2. **Apply:**
   Problem set 3, covering 3.1 and 3.2. The set asks you to defend one control and reject another.

3. **Review:**
   Wooldridge (2020), sections 3-2 to 3-4, on partialling out, the bias and multicollinearity. Then Cunningham (2021), the chapter on directed acyclic graphs from p.96, for the collider. Wooldridge never uses the word.

Next Lecture

Lecture 3.2 asks what our 120 people can settle about a coefficient:

1. **Testing a Coefficient.**
   Comparing an estimate with its own standard error, and what that settles.
2. **Confidence Intervals and the Regression Table.**
   A range built from the same two numbers, and what a published table prints beside it.
3. **Why the t Survives Non-Normal Errors.**
   Our earnings are skewed, and the slope is an average of 120 small errors, so its spread across samples is close to normal.
4. **Testing Restrictions.**
   One F-statistic compares the fit with and without a block of coefficients.

Thank you

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

BACKUP

Appendix

Notation Used in This Lecture

Symbol	What it is	First met
i	which person, running from 1 to n	Lecture 1.1
n	how many people are in the sample, here 120	Lecture 1.1
y_i	hourly earnings of person i, in euro	Lecture 1.1
x_i	years of schooling of person i	Lecture 1.1
$\bar{x}$	the mean years of schooling of our 120	Lecture 1.1
$\mathrm{Var}(\cdot)$	the variance of what is inside it	Lecture 1.2
x_0	the reference level of schooling; ours is thirteen years	Lecture 2.1
β_0	the population intercept: expected earnings at x_0 years and zero ability	Lecture 2.1
u_i	everything else that raises or lowers person i's earnings	Lecture 2.1
$\hat{\beta}_0$	the least-squares intercept from our 120	Lecture 2.1
$\hat{u}_i$	person i's residual: earnings minus the fitted value	Lecture 2.2
R^2	the share of the sample variation in earnings the fit accounts for	Lecture 2.2
SST_x	the sum of squared deviations of schooling	Lecture 2.3

Notation Used in This Lecture

Symbol	What it is	First met
σ^2	the variance of the error	Lecture 2.3
$\hat{\sigma}^2$	estimated error variance, divisor $n-3$	Lecture 2.3
plim	value an estimate concentrates on	Lecture 2.3
a_i	ability of person i, generator's scale	The Two-Regressor Model for Earnings and Schooling
β_1	return to schooling at equal ability	The Two-Regressor Model for Earnings and Schooling
γ	return to ability at fixed schooling	The Two-Regressor Model for Earnings and Schooling
$E(u \mid x, a)$	mean error given schooling and ability	Two Assumptions Change When a Regressor Is Added
$\hat{\pi}_0$	intercept, schooling on ability	The Part of Schooling That Ability Cannot Account For
$\hat{\pi}_1$	slope, schooling on ability	The Part of Schooling That Ability Cannot Account For
r_i	schooling that ability leaves over	The Part of Schooling That Ability Cannot Account For
$\hat{\beta}_1$	fitted schooling return, ability fixed	The Two-Step Residual Recipe on Our 120 People

Notation Used in This Lecture

Symbol	What it is	First met
$\hat{\gamma}$	fitted ability return, schooling fixed	Why the Two Steps Return the Same Coefficient
$\Delta\hat{y}$	gap in fitted earnings, two people	What Holding Other Things Fixed Actually Covers
Δx	gap in their schooling	What Holding Other Things Fixed Actually Covers
Δa	gap in their ability, set to zero	What Holding Other Things Fixed Actually Covers
$\tilde{\beta}_0$	intercept, ability left out	The Regression We Ran and the Regression We Wanted
$\tilde{\beta}_1$	slope on schooling, ability left out	The Regression We Ran and the Regression We Wanted
$\tilde{u}_i$	residual, ability left out	The Regression We Ran and the Regression We Wanted
$\hat{\delta}$	slope, ability on schooling	Each Sum Reduces to Zero, One or a Slope
δ	population slope, ability on schooling	Only the Hats Come Off as the Sample Grows
R_1^2	R-squared, schooling on the others	The Variance of the Schooling Coefficient With a Control
k	number of slopes	Derivation: The First-Order Conditions for Two Regressors

Worked Example: the Formula Closing on Our Own 120 People

Both fits use the same 120 rows, so the formula holds exactly, not on average:

$$\underbrace{1.6700}_{\tilde{\beta}_1} - \underbrace{1.2538}_{\hat{\beta}_1} = 0.4162 = \underbrace{0.2206}_{\hat{\delta}} \times \underbrace{1.8867}_{\hat{\gamma}} \tag{18}$$

A year more schooling goes with 0.2206 more ability. A unit more ability is worth 1.8867 euro an hour. Their product is 0.4162. The two regressions sit 0.4162 apart, to four decimals. Nothing was fitted to make that work.

> **The Formula Is an Identity, Not an Approximation**
> The identity holds in every sample, for any omitted variable. Real work has the formula but lacks $\hat{\gamma}$ and $\hat{\delta}$. Both need the column we do not have (Wooldridge, 2020).

Back to the slide

Derivation: the First-Order Conditions for Two Regressors

The anatomy proof used three conditions on the long fit's residuals, and each comes from setting a derivative of the residual sum to zero:

$$\sum_{i=1}^{n} \hat{u}_i = 0 \qquad \sum_{i=1}^{n} (x_i - x_0)\, \hat{u}_i = 0 \qquad \sum_{i=1}^{n} a_i\, \hat{u}_i = 0 \tag{19}$$

One Condition per Coefficient

Three coefficients give three conditions. So the divisor for $\hat{\sigma}^2$ is $n - 3$ here, where Lecture 2.3 used $n - 2$. Wooldridge writes the general set for k regressors as $k + 1$ equations (Wooldridge, 2020).

Derivation: Why the Centring Constant Cancels

Lecture 2.3's appendix proved the two facts both body derivations used:

$$\sum_{i=1}^{n}(x_i - \bar{x})(x_i - x_0) = \sum_{i=1}^{n}(x_i - \bar{x})^2 + (\bar{x} - x_0)\sum_{i=1}^{n}(x_i - \bar{x}) = \mathrm{SST}_x \tag{20}$$

Centring at the Leaving Certificate Costs Nothing

The second sum is zero, so x_0 contributes nothing. Both derivations run unchanged on raw schooling. Centring sets what the intercept means and leaves every slope where it was.

Notes and Tools

- **The App:** two sliders, ability to schooling and ability to pay, with both coefficients printed.
- **The Lecture in One Line:** `coef(lm(y ~ x + a))[2]` beside `coef(lm(y ~ x))[2]`.
- **This Lecture's Lab:** `resid()` and `lm()`, checked against the two-variable fit and against $\hat{\gamma}\hat{\delta}$.
- **Variance Inflation Factors:** `car::vif()`, read alongside O'Brien (2007).
- **Where This Module Sits:** Angrist & Pischke (2017), on what a first econometrics course should teach.

References

Angrist, J. D., & Pischke, J.-S. (2015). *Mastering 'metrics: The path from cause to effect*. Princeton University Press.

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

Cunningham, S. (2021). *Causal inference: The mixtape*. Yale University Press.

Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology*, *40*, 31–53. https://doi.org/10.1146/annurev-soc-071913-043455

Goldberger, A. S. (1991). *A course in econometrics*. Harvard University Press.

Lindner, T., Puck, J., & Verbeke, A. (2020). Misconceptions about multicollinearity in international business research: Identification, consequences, and remedies. *Journal of International Business Studies*, *51*(3), 283–298. https://doi.org/10.1057/s41267-019-00257-1

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, *41*(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6

References

Vanhove, J. (2021). Collinearity isn't a disease that needs curing. *Meta-Psychology*, *5*, MP.2021.2548. https://doi.org/10.15626/MP.2021.2548

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