LECTURE 2.3

Unbiasedness, Precision and Large Samples

ECON30130 Econometrics I

Sam Deegan
University College Dublin

Autumn 2027

Where We Are

1. Foundations	2. Simple Regression	3. Multiple Regression	4. Broken Assumptions
1.1 Data and Variation	2.1 Least Squares	3.1 Estimation	4.1 Heteroskedasticity
1.2 Conditional Means	2.2 Units and Fit	3.2 Inference	4.2 Specification
1.3 Sampling Variation	**2.3 Bias and Precision**	3.3 Functional Form	
		3.4 Dummies	

> **The Last of Three Lectures on One Regressor**
> Lecture 2.1 fitted the line and 2.2 measured how well it fits. Today asks how far the fitted slope sits from the true one.

Last Week's Lecture

Lecture 2.2 read the slope in euro and asked what R-squared cannot report:

1. **Reading the Slope.**
   What 1.57 euro an hour per year says, and how a few high earners change it.

2. **What R-Squared Will Not Tell You.**
   A share of variation reports neither the line's shape nor the slope's size.

3. **Fit on a Hold-Out Sample.**
   Why R-squared on the fitting rows usually exceeds R-squared on new rows.

What Lectures 1.2 and 2.2 Left Open

The sample mean was judged on three properties in Lecture 1.2: unbiasedness, consistency and efficiency. Lecture 2.2's R-squared says nothing about how far 1.5744 sits from the truth.

> **Question**
> If I drew a different 120 people from the same population and fitted again, would I get the same slope? Should you want me to?

A Different 120 Would Return a Different Slope

The slope is built from 120 unobserved errors. A new draw brings new ones.

> **A Rule That Never Varies Ignores the Data**
> Lecture 1.2's rule C reports 12.00 euro whatever the sample. The rule never varies, and differs from the truth by the same amount every time.

You want a rule centred on the truth, with a small spread across samples.

Lecture Plan

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.

SECTION 1

The Gauss-Markov Assumptions

Five Claims About the Population, Not the Sample

The arithmetic of 2.1 and 2.2 works on any 120 rows, including nonsense. Five assumptions make that arithmetic a claim about the world:

1. **Linear in Parameters:** the population line is $y = \beta_0 + \beta_1(x - x_0) + u$, with x_0 at thirteen years.
2. **Random Sampling:** our 120 people were drawn at random from that population.
3. **Variation in Schooling:** the 120 do not all have the same years of schooling.
4. **Zero Conditional Mean:** at every level of schooling, the error averages to zero.
5. **Constant Error Variance:** at every level of schooling, the error has the same variance.

> **Wooldridge Numbers These SLR.1 to SLR.5**
> SLR is simple linear regression; 3.1's multiple regression uses MLR.1 to MLR.5 (Wooldridge, 2020).

What Assumptions One to Three Are Doing

None of the first three assumptions is controversial, and each does one job:

- **Linear in Parameters:** β_0 and β_1 are only added and multiplied, which is what makes β_1 a single number.
- **Random Sampling:** one person's error says nothing about another's.
- **Variation in Schooling:** identical schooling for all 120 makes the slope's denominator zero.

> **Assumption Three on Our 120 People**
> Squaring each person's distance from mean schooling and adding gives 439.5. If all 120 had thirteen years of schooling, every distance would be zero. The total would then be zero, and no slope would exist.

What Each Assumption Is Needed For

The five assumptions do five different jobs:

Assumption	What it is needed for
Linear in Parameters	the assumption is about β_0 and β_1, not about the shape the line draws
Random Sampling	different people's errors are uncorrelated, which the slope's variance needs
Variation in Schooling	the spread of schooling is not zero, so a slope exists
Zero Conditional Mean	assumption four is the only one that delivers unbiasedness
Constant Error Variance	assumption five makes the slope's variance formula correct

Four and Five Fail in Different Ways

Assumption four is about the error's average, and five about its spread. Break four and the estimator centres on a value other than β_1. Break five and the centring survives. The printed standard error no longer tells you how far our slope typically sits from the true one.

Assumption Four: Zero Conditional Mean

The condition that made 2.1's line the conditional expectation also makes the slope unbiased:

$$E(u \mid x) = 0 \tag{1}$$

- u: everything other than schooling that raises or lowers earnings
- $E(u \mid x)$: the average of the error among people with a given schooling

> **Assumption Four Is a Claim About Ability**
> Assumption four says the average of everything else does not rise with schooling. In our population the errors around the line average zero at every level of schooling, so four holds for that line here. Outside a simulation it almost never holds: Irish earnings data would break it. Lecture 3.1 adds ability to the model.

Assumption Five: Constant Error Variance

Assumption five is about the spread of the error, not about its average:

$$\text{Var}(u \mid x) = \sigma^2 \tag{2}$$

- σ^2: the variance of the error

In this population σ is 5.93 euro, and close to that at every level of schooling our 120 reach.

Wooldridge Adds Assumption Five for the Variance

Assumption five plays no role in unbiasedness, and Wooldridge adds it for two reasons (Wooldridge, 2020):

- **A Short Formula:** the slope's variance collapses to one short expression.
- **The Gauss-Markov Theorem:** the theorem in the Precision section is false without it.

> **A Broken Five Leaves the Estimate Unbiased**
> Assumption five fails if earnings spread out more at higher schooling. The estimator stays unbiased. The printed standard error no longer tells you how far our slope typically sits from the true one.

One Sample Fitted Against the Line We Know Is True

Figure 1: Population pale; our 120 and their line blue; the population line green dashed.

> **Only the Fitted Line Is Ever Available**
> The two lines differ because our 120 people are one draw. We know the true line only because we wrote the population ourselves.

The Weight Each Person Carries

Least squares does not treat our 120 people alike. Each person carries a weight:

$$w_i = \frac{\overbrace{x_i - \bar{x}}^{\text{the deviation}}}{\underbrace{\sum_{j=1}^{n}(x_j - \bar{x})^2}_{\text{SST}_x}} \tag{3}$$

- w_i: the weight person i carries
- j: a second index over the same 120 people, used inside w_i
- SST_x: the sum of squared deviations of schooling, 439.5 in our sample

Worked example: our 120 people

A Weight Is Negative Below Mean Schooling

Our mean schooling is 13.93, so a person below it carries a negative weight:

$$\underbrace{\frac{x_i - \bar{x}}{\sum_{j=1}^{n}(x_j - \bar{x})^2}}_{\text{above the mean}} > 0 \qquad \underbrace{\frac{x_i - \bar{x}}{\sum_{j=1}^{n}(x_j - \bar{x})^2}}_{\text{below the mean}} < 0 \tag{4}$$

Neither weight depends on earnings.

Two Facts About the Weights

$\sum w_i = 0$, because the deviations of schooling sum to zero. And $\sum w_i(x_i - x_0) = 1$, because that numerator is SST_x itself. The appendix proves both.

Worked example: our 120 people

SECTION 2

Unbiasedness

Multiplying Out the Numerator of the Slope

The random earnings sit inside both sums of Lecture 2.1's covariance over variance:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\underbrace{\sum_{i=1}^{n}(x_i - \bar{x})^2}_{\text{SST}_x}} \tag{5}$$

Multiplying the deviations out splits the numerator into two sums:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\, y_i \;-\; \bar{y}\sum_{i=1}^{n}(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{6}$$

The Slope Is a Weighted Sum of the Earnings

The deviations of schooling sum to zero, so the $\bar{y}$ term drops out:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\, y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sum_{i=1}^{n} w_i\, y_i \tag{7}$$

The Same Slope, Rearranged

Only the arrangement has changed: 691.9 over 439.5 still gives our 1.57. Each person contributes a weight fixed by schooling, times one random earnings figure.

The Population Model Supplies Each Person's Earnings

The population model says what each person's earnings are made of:

$$\hat{\beta}_1 = \sum_{i=1}^{n} w_i \, y_i \qquad y_i = \beta_0 + \beta_1 (x_i - x_0) + u_i \tag{8}$$

- u_i: person i's error

Substituting the Model Gives Three Sums

The model replaces every y_i in the weighted sum:

$$\hat{\beta}_1 = \sum_{i=1}^{n} w_i \left(\beta_0 + \beta_1(x_i - x_0) + u_i\right) \tag{9}$$

Multiplying each weight through gives one sum for each term:

$$\hat{\beta}_1 = \beta_0 \sum_{i=1}^{n} w_i \; + \; \beta_1 \sum_{i=1}^{n} w_i(x_i - x_0) \; + \; \sum_{i=1}^{n} w_i u_i \tag{10}$$

The Slope Equals the True Slope Plus One Term

The first sum is zero, so the intercept disappears. The second sum is one, so the true slope comes through unchanged:

$$\hat{\beta}_1 = \beta_1 + \underbrace{\sum_{i=1}^{n} w_i u_i}_{\text{the weighted sum of errors}} \tag{11}$$

One Term Separates the Estimate from the Truth

$\sum w_i u_i$ is the entire distance between $\hat{\beta}_1$ and β_1, and nobody observes it. Unbiasedness is its average across samples, and precision is its variance.

Worked example: our 120 people

Why the Rule Centres on the True Slope

The weights come from schooling alone, and assumption four sets each error's average to zero:

$$E\left(\hat{\beta}_1 \mid x\right) = \beta_1 + \sum_{i=1}^{n} w_i \, E(u_i \mid x) = \beta_1 \tag{12}$$

Only Assumption Four Is Used at This Step

Assumptions one to three were used to write the decomposition down. Assumption four alone says what the error averages to. Breaking five therefore leaves unbiasedness intact.

Least Squares Is Unbiased Under Assumptions One to Four

Every set of schooling values gives an average of β_1, so averaging over all the sets still gives β_1 (Wooldridge, 2020):

$$E(\hat{\beta}_1) = \beta_1 \tag{13}$$

The Sampling Distribution of the Slope, Simulated

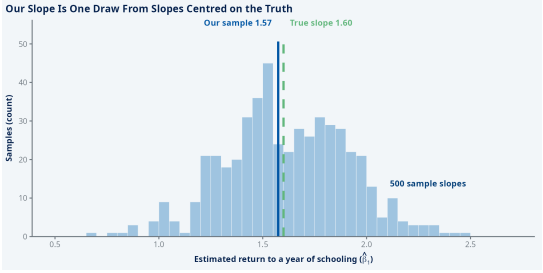

Figure 2: Slopes from 500 samples of 120 (repeated-sample ghost), our 1.57, the true 1.60.

> **Our 1.5744 Is One Bar in This Histogram**
> The five hundred slopes pile up over the true 1.60. Lecture 1.3 drew the same picture for a mean.

Unbiasedness Is a Property of the Rule, Not of Your Estimate

Unbiasedness says where the histogram sits, and nothing about any single bar (Wooldridge, 2020):

- **True:** “least squares is unbiased” describes the pile of five hundred slopes.
- **Meaningless:** “our estimate is unbiased”: one number has no sampling distribution.

> **An Unbiased Rule Still Misses in Any One Sample**
> Across samples the estimate falls either side of 1.60 and averages to it. Least squares is unbiased, and our 1.5744 still differs from the truth. A student who reads unbiased as “correct” misreads every standard error after today.

Dropping the Highest Earner Is Still an Estimator

A colleague deletes the highest-earning person before fitting every regression. Fixed before the data arrive, the rule is an estimator with its own histogram.

> **Question**
> Is that estimator unbiased for β_1? Which of the five assumptions does the deletion break?

Deleting the Highest Earner Breaks Assumption Four

The rule is biased, because whether a row survives depends on its own error:

- **Assumption Four:** the kept errors no longer average to zero at each schooling level.
- **Not Assumption Two:** the 120 arrived at random, and the damage comes afterwards.
- **Not Assumption Three:** the 119 people left still vary in schooling.

> **The Bias Tends to Run Downwards**
> The highest earner tends to have a large positive error and above-average schooling. Deleting that person tends to pull the fitted line's right-hand end down.

SECTION 3

Precision

Deriving the Variance of the Slope

Unbiasedness took the average of the leftover term. Precision takes its variance instead, with our 120 schooling values held fixed:

$$\text{Var}(\hat{\beta}_1) = \text{Var}\left(\sum_{i=1}^{n} w_i u_i\right) = \sum_{i=1}^{n} w_i^2 \, \text{Var}(u_i) \tag{14}$$

Random Sampling Removes the Cross Terms

Random sampling leaves different people's errors uncorrelated. So the variance of the sum contains no covariance between two people's errors (Wooldridge, 2020).

The Variance of the Slope Is σ^2 Over SST_x

Assumption five puts one σ^2 in place of every $\text{Var}(u_i)$. What is left on top is SST_x, so one factor cancels:

$$\text{Var}(\hat{\beta}_1) = \sigma^2 \sum_{i=1}^{n} w_i^2 = \sigma^2 \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\left(\sum_{j=1}^{n}(x_j - \bar{x})^2\right)^2} = \underbrace{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}_{\text{SST}_x} \tag{15}$$

Why the Divisor Is SST_x

A sample mean gives every person equal weight, so n sits underneath it. A slope weights people by their distance from mean schooling, so SST_x does.

Three Ingredients, Read Off the Formula

Splitting SST_x names the only three things that move the slope from sample to sample:

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{(n-1)\,s_x^2} \qquad \text{sd}(\hat{\beta}_1) = \frac{\sigma}{s_x\sqrt{n-1}} \tag{16}$$

- **The Error Variance:** σ^2 sits alone on top, so noisier earnings widen the histogram.
- **The Spread in Schooling:** s_x is below: bunched schooling widens the histogram.
- **The Sample Size:** $\sqrt{n-1}$ is underneath, so four times the data halves the spread.

You Choose Two of the Three Ingredients

No sampling design lowers σ. Who you sample sets s_x; how many sets n. Widening schooling lowers the standard error faster than adding people: s_x enters squared.

A Larger Sample Narrows the Histogram

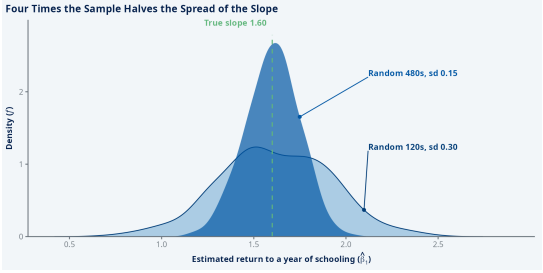

Figure 3: Slopes from 500 samples of 480 (blue) and of 120 (pale): repeated-sample ghosts.

> **Quadrupling the Sample Halves the Spread**
> Both densities centre near 1.60, because n appears nowhere in the unbiasedness proof. Four times the data halves the spread, 0.30 to 0.15, at the rate $\sqrt{n-1}$ sets.

More Spread in Schooling Narrows the Histogram

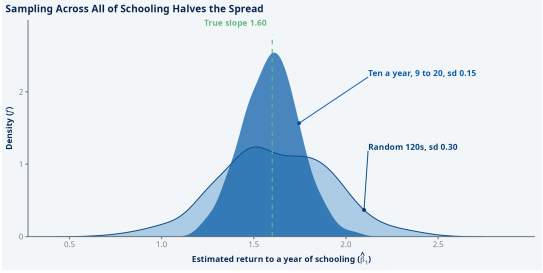

Figure 4: Same population, samples spread evenly across schooling (blue) against random 120s.

> **Data Cannot Replace Spread in Schooling**
> Ten people at each year from nine to twenty double s_x, from about 1.8 to 3.5. SST_x is larger, and the slope varies half as much across samples. With no spread at all SST_x is zero, which assumption three rules out.

Choosing Between More People and Wider Schooling

You may have only one of two samples from the same population:

- **Many People:** two hundred people have eleven to thirteen years of schooling.
- **Wide Schooling:** eighty people have nine to twenty years of schooling.

> **Question**
> Which sample gives the more precise slope? Which two ingredients have you traded against each other?

The Eighty Give the More Precise Slope

The people enter $\text{SST}_x = (n-1)s_x^2$ once, and the spread enters squared:

- **Sample Size:** 79 against 199 cuts SST_x by a factor of about 2.5.
- **Spread:** a range of eleven years against two raises it by a factor of about 30.
- **Error Variance:** both samples share one population, so σ does not change.

> **The Eighty's SST_x Is Roughly Twelve Times Larger**
> The eighty's standard error is therefore roughly 3.5 times smaller. The twelve treats spread as proportional to the range of schooling.

A Smaller Error Standard Deviation Narrows the Histogram

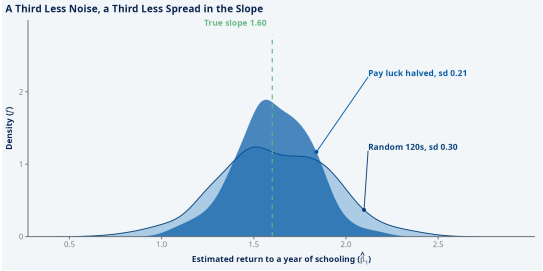

Figure 5: Pay luck halved (blue), our world (pale): repeated-sample ghosts of the same people.

> **Noise Is a Population Fact**
> Halving each person's own pay luck takes σ from 5.93 to 4.04, a third lower. σ sits alone on top, so the width falls by the same third, from 0.30 to 0.21. No sample size and no sampling design lowers σ itself.

Estimating the Error Standard Deviation from the Residuals

Nobody observes the errors, so our residuals give an estimate of σ ($\hat{\sigma}$):

$$\hat{\sigma} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2} \tag{17}$$

- $\hat{u}_i$: person i's residual; the sum of their squares is SSR

> **Our 5.993 Estimates the True 5.93**
> The error's true standard deviation is 5.93, and our residuals give 5.993. Two fitted coefficients have already used two residuals, so the divisor is $n - 2$.

Worked example: our 120 people

The Standard Error on Our 120 People

Putting $\hat{\sigma}$ in place of σ turns the standard deviation into a standard error:

$$\text{se}(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \tag{18}$$

The Spread of the Histogram, Not of Earnings
All three ingredients sit inside: 5.993 for the error, and 439.5 for the spread of schooling and the 120 people. $\text{se}(\hat{\beta}_1)$ estimates the spread of the histogram of slopes (Angrist & Pischke, 2015).

Worked example: our 120 people

Least Squares Against a Two-Point Estimator

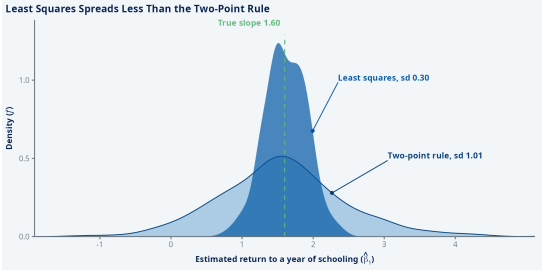

Figure 6: Least squares (blue) and the two-point rule (pale), same 500 samples: ghosts.

> **The Two-Point Rule Is Also Centred on 1.60**
> The two-point rule uses two of the 120 people and is still unbiased. Unbiasedness therefore cannot choose between it and least squares.

What the Gauss-Markov Theorem Claims

The two-point rule is one of a whole class of linear unbiased rules. Linear, unbiased and best each restrict what the theorem claims:

- **Linear:** the estimator is a weighted sum of the y_i, as least squares is.
- **Unbiased:** the estimator centres on β_1, as the two-point rule also does.
- **Best:** least squares has the smallest sampling variance, and best claims nothing more.

> **Best Linear Unbiased**
> Take every estimator that is a weighted sum of the earnings and unbiased for β_1. Least squares has the smallest sampling variance among them. The five assumptions are named after this theorem (Wooldridge, 2020).

What the Theorem Does Not Claim

Three limits bound the theorem, and the third, on bias, is most often overlooked:

- **A Class, Not Every Rule:** the theorem is silent about biased and non-linear rules.
- **All Five Assumptions:** without the fifth, least squares stops being best.
- **Precision, Not Centring:** the theorem ranks variances and says nothing about bias.

> **A Biased Rule Can Have a Smaller Variance**
> Lecture 1.2's trimmed mean was biased and had the smaller sampling variance. The theorem neither permits nor forbids such a rule.

Unbiased, Consistent and Efficient, Now for the Slope

The unbiasedness proof and the Gauss-Markov theorem settle two of Lecture 1.2's three properties:

- **Unbiased:** Over the samples we might have drawn, the rule returns β_1.
- **Consistent:** As n grows without bound, the rule concentrates on β_1.
- **Efficient:** Of two unbiased rules, the efficient one has the smaller sampling variance.

Neither result says what happens to the slope as the sample grows without bound. The variance already derived has the sample size in its denominator.

SECTION 4

Consistency and Asymptotic Efficiency

The Variance Shrinks, So the Slope Concentrates on the Truth

A distribution centred on β_1 whose spread goes to nothing collapses onto β_1:

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{(n-1)\, s_x^2} \longrightarrow 0 \qquad \Longrightarrow \qquad \text{plim}\, \hat{\beta}_1 = \beta_1 \tag{19}$$

Reading plim

plim is the value an estimator concentrates on as the sample grows without bound. Lecture 1.2 called that idea the law of large numbers for the sample mean (Wooldridge, 2020).

The Spread of the Slope at Four Sample Sizes

Our schooling variance is 439.5 over 119, or 3.69. Holding that and σ at 5.93 fixed, n alone sets the spread:

n	$\text{SST}_x = (n-1)\,s_x^2$	$\text{sd}(\hat{\beta}_1)$
30	107.1	0.5731
120	439.5	0.2829
480	1768.9	0.1410
2000	7382.3	0.0690

These Are Spreads, Not Standard Errors

Every number in the third column uses the true σ of 5.93. Only a simulation hands you that. Our 0.2859 is one sample's estimate of the 0.2829.

Four Sample Sizes Against One True Slope

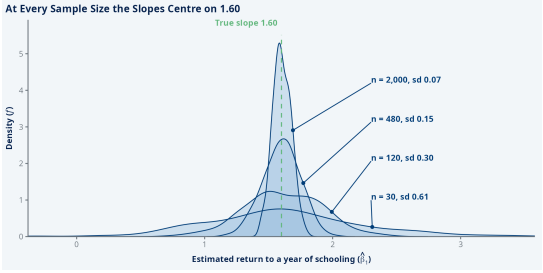

Figure 7: 500 slopes at n = 30, 120, 480 and 2,000, repeated-sample ghosts, against 1.60.

> **All Four Sample Sizes Centre on 1.60**
> All four pile up over 1.60, because *n* enters no step of the unbiasedness proof. Only the width changes: 0.61, 0.30, 0.15 and 0.07.

Consistent but Biased, Unbiased but Inconsistent

Two rules built from ours each have exactly one of the two properties:

Rule	Averages to	As n grows
our slope plus $1/n$	$\beta_1 + 1/n$	the bias shrinks: 0.0083 at 120, 0.0005 at 2000
1.2's first-ten rule, for a slope	β_1	the spread never falls

Unbiasedness Is the Claim About Our 120

Unbiasedness holds at the sample size you actually have. Consistency describes a sequence of samples, of which you drew one. We want both, because each of them can hold while the other fails.

Dividing Both Sums by n Gives Two Averages

The weighted sum of errors is one sum over another. Dividing both by n changes nothing:

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})\,u_i}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{20}$$

Each Average Concentrates on a Population Number
The top average concentrates on $\text{Cov}(x, u)$. The bottom concentrates on $\text{Var}(x)$, which s_x^2 estimates.

Consistency Needs Only That Schooling and the Error Are Uncorrelated

In the limit each average is replaced by the population number it concentrates on:

$$\operatorname{plim} \hat{\beta}_1 = \beta_1 + \underbrace{\frac{\operatorname{Cov}(x, u)}{\operatorname{Var}(x)}}_{\text{the inconsistency}} \tag{21}$$

What Consistency Asks for Instead

Assumption three keeps $\operatorname{Var}(x)$ above zero, so the inconsistency vanishes exactly when $\operatorname{Cov}(x, u) = 0$. Assumption four asks more: a zero average error at every level of schooling (Wooldridge, 2020).

Does a Bigger Sample Repair a Broken Assumption Four?

Every spread in the table fell as the sample grew. The inconsistency, $\mathrm{Cov}(x, u)/\mathrm{Var}(x)$, contains no sample size at all.

> **Question**
> If schooling and the error are correlated, does collecting more people bring the slope back to 1.60?

A Bigger Sample Does Not Repair Assumption Four

Both parts of the inconsistency are population numbers, so neither shrinks as the sample grows:

- **The Spread:** falls from 0.5731 at thirty people to 0.0690 at two thousand.
- **The Inconsistency:** stays the same size at every sample size.

> **More Data Concentrates the Slope on the Wrong Number**
> The histogram narrows around β_1 plus the inconsistency, not around β_1. The reported standard error shrinks all the same.

Each Function of Schooling Defines a Rule for the Slope

Each function of schooling (g), picked before the data arrive, gives its own slope ($\tilde{\beta}_1$):

$$\sum_{i=1}^{n} g(x_i)\left(y_i - \tilde{\beta}_0 - \tilde{\beta}_1(x_i - x_0)\right) = 0 \tag{22}$$

Only a g Correlated with Schooling Qualifies

Under the first four assumptions a choice of g gives a consistent rule when $g(x_i)$ is correlated with schooling. Squaring schooling qualifies. A g that returns the same number for everyone does not, because the denominator it builds is then zero.

No Choice of g Beats Least Squares in Large Samples

Least squares is the choice $g(x_i) = x_i - x_0$. Under all five assumptions no other choice has a smaller large-sample variance (Wooldridge, 2020).

What Each Efficiency Theorem Covers

Gauss-Markov ranks least squares among linear unbiased rules at every sample size. Asymptotic efficiency ranks it inside a wider family, and only as the sample grows. Neither of them ranks bias.

What Would Have to Be True?

Two claims now rest on our numbers. The rule behind 1.5744 centres on the truth, and 0.2859 measures its spread across samples.

> **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?

What 1.5744 and 0.2859 Each Rely On

The two claims rest on different assumptions:

- **The Centring on the Truth:** assumption four, and nothing substitutes for it.
- **The 0.2859:** assumption five, random sampling, and 5.993 a fair estimate of 5.93.

> **The Standard Error Gives No Warning of Bias**
> If people who stayed in school longer would have earned more anyway, assumption four fails. The histogram then sits in the wrong place by construction. The 0.2859 measures its width and says nothing about its centre.

Switching Off Assumption Four in the App

The app draws five hundred samples and builds the histogram of slopes. Each of the five assumptions can be switched off alone.

> **Question**
> If we break assumption four, does the histogram shift, widen, both, or neither?

Breaking Assumption Four Shifts the Histogram Off 1.60

Assumption four enters one step of the unbiasedness proof, so breaking it shifts the centre:

- **The Centre:** non-zero $E(u_i \mid x)$ terms survive, so the histogram sits off 1.60.
- **The Width:** assumption four is absent from σ^2/SST_x, so the width holds.

> **A Switch Can Change Two Things**
> A switch that also changes the error variance changes the width too. In real work you would see neither histogram, only one number and its standard error.

Summary: Assumptions, Precision and Large Samples

1. **The Gauss-Markov Assumptions.**
   Assumption four makes the estimator unbiased; assumption five makes the variance formula correct.

2. **Unbiasedness.**
   The fitted slope is β_1 plus a weighted sum of errors whose average is zero.

3. **Precision.**
   The same sum has variance σ^2/SST_x, which on our 120 people gives a standard error of 0.2859.

4. **Consistency and Asymptotic Efficiency.**
   That variance falls to nothing, so the slope concentrates on β_1, and no rule in a wider class beats it in the limit.

Before Lecture 3.1

1. **Play: Sample Size Against Spread.**
   Raise the sample size, then widen schooling, and see which narrows the histogram more.

2. **Apply: Problem Set 2.**
   Today's decomposition again, with $n = 6$ and numbers you can do by hand.

3. **Review: Three Wooldridge Sections.**
   2-5 on the assumptions and the slope's variance, then 5-1 and 5-3 (Wooldridge, 2020).

Next Lecture

Lecture 3.1 puts a second variable into the model:

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.

Thank you

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

BACKUP

Appendix

Notation Used in This Lecture, Part 1

Symbol	What it is	First met
$\sum$	a sum over the people in the sample	Lecture 1.1
i	which person, running from 1 to n	Lecture 1.1
j	a second index over the same people, inside a term indexed i	The Weight Each Person Carries
n	the number of people in the sample, 120 here	Lecture 1.1
y	hourly earnings in the population, in euro	Lecture 1.1
y_i	hourly earnings of person i, in euro	Lecture 1.1
$\bar{y}$	the mean hourly earnings of our 120	Lecture 1.1
x	years of schooling in the population	Lecture 1.1
x_i	years of schooling of person i	Lecture 1.1
$\bar{x}$	the mean years of schooling of our 120, 13.93	Lecture 1.1

Notation Used in This Lecture, Part 2

Symbol	What it is	First met
s_x^2	the sample variance of schooling, 3.69 on our 120	Lecture 1.1
s_x	the sample standard deviation of schooling	Lecture 1.1
β_0	the population intercept, expected earnings at thirteen years	Lecture 2.1
x_0	the reference level of schooling, thirteen years in our population	Lecture 2.1
β_1	the population slope, 1.60	Lecture 2.1
$\hat{\beta}_1$	the least-squares slope from our 120, 1.5744	Lecture 2.1
u	everything other than schooling that raises or lowers earnings	Lecture 2.1
u_i	person i's error	Lecture 2.1
$\hat{u}_i$	person i's residual	Lecture 2.1
SSR	the residual sum of squares, 4238.3 on our 120	Lecture 2.1

Notation Used in This Lecture, Part 3

Symbol	What it is	First met
σ^2	the variance of the error	Assumption Five: Constant Error Variance
σ	the error standard deviation, 5.93 in this population	Assumption Five: Constant Error Variance
w_i	the weight person i carries	The Weight Each Person Carries
SST_x	the sum of squared deviations of schooling, 439.5 in our sample	The Weight Each Person Carries
$\hat{\sigma}$	the estimated error standard deviation, 5.993 on our 120	Estimating the Error Standard Deviation From the Residuals
$\hat{\sigma}^2$	the estimated error variance	Derivation: Why the Divisor Is n Minus Two
$g(x_i)$	any function of person i's schooling	Each Function of Schooling Defines a Rule for the Slope
$\tilde{\beta}_0$	the intercept a competing rule returns	Each Function of Schooling Defines a Rule for the Slope
$\tilde{\beta}_1$	the slope a competing rule returns	Each Function of Schooling Defines a Rule for the Slope

Notation Used in This Lecture, Part 4

Symbol	What it is	First met
the hat, $\hat{\beta}_1$	an estimate of whatever sits under it	Lecture 1.2
the tilde, $\tilde{\beta}_1$	a competing rule's estimate, set against least squares	Each Function of Schooling Defines a Rule for the Slope
$E(\cdot)$	the population average of whatever is inside	Lecture 1.2
$E(u \mid x)$	the average of the error among people with a given schooling	Lecture 2.1
$\mathrm{Var}(\cdot)$	the variance of whatever is inside	Lecture 1.2
$\mathrm{Var}(u \mid x)$	the variance of the error among people with a given schooling	Assumption Five: Constant Error Variance
$\mathrm{Cov}(\cdot, \cdot)$	the covariance of the two quantities inside	Lecture 1.2
$\mathrm{sd}(\hat{\beta}_1)$	the standard deviation of the slope across repeated samples	Three Ingredients, Read Off the Formula
$\mathrm{se}(\hat{\beta}_1)$	the estimated standard deviation of the slope across samples, 0.2859	The Standard Error on Our 120 People
plim	the value an estimator concentrates on as the sample grows without bound	The Variance Shrinks, So the Slope Concentrates on the Truth

Worked Example: the Weights on Our 120

Each weight divides a person's distance from mean schooling by one total:

$$w_i = \frac{x_i - \bar{x}}{\underbrace{\sum_{j=1}^{n}(x_j - \bar{x})^2}_{\text{SST}_x}} \tag{23}$$

On our 120 people mean schooling is 13.93, and the squared deviations add to 439.5:

$$w_i = \frac{x_i - 13.93}{439.5} \qquad \text{SST}_x = 439.5 \tag{24}$$

Back to the slide

Worked Example: a Weight Above and a Weight Below the Mean

Take one person with eighteen years of schooling and one with ten:

$$w_i = \frac{x_i - \bar{x}}{\sum_{j=1}^{n}(x_j - \bar{x})^2} \tag{25}$$

Our mean schooling of 13.93 replaces $\bar{x}$, and our 439.5 replaces the sum, SST_x:

$$\underbrace{\frac{18 - 13.93}{439.5}}_{\text{eighteen years}} = 0.00926 \qquad \underbrace{\frac{10 - 13.93}{439.5}}_{\text{ten years}} = -0.00894 \tag{26}$$

Back to the slide

Worked Example: Our Draw's Weighted Sum of Errors

Our sample gave 1.5744 when the true slope is 1.60:

$$\hat{\beta}_1 = \beta_1 + \sum_{i=1}^{n} w_i u_i \qquad \Longrightarrow \qquad 1.5744 = 1.60 + (-0.0256) \tag{27}$$

> **Nobody Sees the Histogram in Real Work**
> In real work you draw once and never see the histogram. You see 1.5744, and neither the 1.60 nor the -0.0256.

Back to the slide

Worked Example: the Error Standard Deviation on Our 120

The estimate of σ divides the residual sum of squares by $n - 2$:

$$\hat{\sigma} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2} \tag{28}$$

Our 4238.3 replaces the sum, SSR, and 120 people less two leave 118:

$$\hat{\sigma} = \sqrt{\frac{4238.3}{118}} = 5.993 \tag{29}$$

Back to the slide

Worked Example: the Standard Error on Our 120

The standard error puts $\hat{\sigma}$ over the square root of SST_x:

$$\text{se}(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \tag{30}$$

Our 5.993 replaces $\hat{\sigma}$, and our 439.5 replaces the sum, SST_x:

$$\text{se}(\hat{\beta}_1) = \frac{5.993}{\sqrt{439.5}} = 0.2859 \tag{31}$$

Back to the slide

Derivation: Two Facts About the Weights

The body used $\sum w_i = 0$ and $\sum w_i(x_i - x_0) = 1$ without proof. Both are summation algebra from Lecture 1.1:

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0 \qquad \sum_{i=1}^{n}(x_i - \bar{x})(x_i - x_0) = \sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{32}$$

The Centring Constant Cancels Once Again

The second holds because the first is zero: replacing $x_i - x_0$ by $(x_i - \bar{x}) + (\bar{x} - x_0)$ adds a multiple of the first sum, which contributes nothing. Dividing both by that sum, SST_x, gives the two facts.

Derivation: Why the Divisor Is n Minus Two

The body used $\hat{\sigma}^2 = \text{SSR}/(n-2)$. Residuals obey two restrictions the errors do not, Lecture 2.1's two first-order conditions:

$$\sum_{i=1}^{n} \hat{u}_i = 0 \qquad \sum_{i=1}^{n} (x_i - x_0)\,\hat{u}_i = 0 \tag{33}$$

Two Conditions, Two Degrees of Freedom

Given 118 of the residuals, those two equations return the other two, so the 120 residuals carry only 118 free numbers. Dividing by 118 rather than 120 is what makes $\hat{\sigma}^2$ unbiased for σ^2 (Wooldridge, 2020).

Notes and Tools

Three tools go with this lecture:

- **The App:** the sampling app, with the five assumptions switchable one at a time.
- **The Lecture in One Line:** `summary(lm(y ~ I(x - 13)))$coefficients[2, 2]`.
- **This Lecture's Lab:** `replicate()` and `sd()` on 500 slopes, checked against `summary()`.

References

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

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