The Multiple Regression Model

What Happens When One Regressor Isn't Enough

Jake Anderson

March 21, 2026

Outline

Outline

A Familiar Regression

From earlier lectures, you can estimate a simple wage equation:

$$\log(\text{wage}_i) = \beta_1 + \beta_2\,\text{educ}_i + e_i$$

You run OLS and get $b_2 = 0.076$: each additional year of education is associated with about a 7.6% increase in wages.

But is that really the return to education?

Think about what we left out:

- More experienced workers earn more
- Workers with more education tend to have *less* labor market experience (they spent those years in school)

$\Longrightarrow$ If experience affects wages *and* is correlated with education, our estimate of β_2 absorbs both effects.

The Thought Experiment

Imagine two workers with **the same experience** but different education levels.

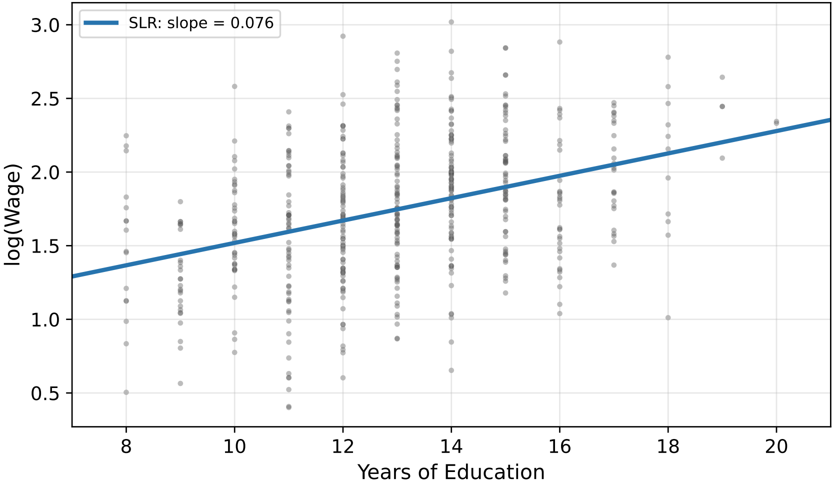
Worker A: 12 years of education, 20 years of experience
Worker B: 16 years of education, 20 years of experience

The wage gap between A and B is **purely due to education**, because experience is held constant.

That's what we want β_2 to measure: the effect of education *holding experience constant*.

But simple regression doesn't hold anything constant. It only has one regressor.
$\implies$ We need a model with **multiple regressors**.

What Happens in the Data?



Simple regression of log(wage) on educ: slope $= 0.076$.

Outline

From One Regressor to Many

Simple regression (Chapters 2–4): one explanatory variable

$$y_i = \beta_1 + \beta_2 x_i + e_i$$

Multiple regression: two or more explanatory variables

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i$$

The general model with K parameters ($K - 1$ regressors + intercept):

$$\boxed{y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_K x_{iK} + e_i}$$

Everything you learned in SLR extends naturally:

- OLS still minimizes $\sum \hat{e}_i^2$
- Software solves a system of equations (no matrix algebra needed)
- The interpretation of each β_k changes in one important way…

The Mincer Wage Equation

The classic labor economics model of wages:

$$\log(\text{wage}_i) = \beta_1 + \beta_2\,\text{educ}_i + \beta_3\,\text{exper}_i + \beta_4\,\text{exper}_i^2 + e_i$$

Why exper^2? Returns to experience diminish: the 2nd year of experience helps more than the 20th.

This is a multiple regression model with $K = 4$ parameters:

- β_1: intercept (baseline log wage)
- β_2: return to an extra year of education, *holding experience constant*
- β_3, β_4: together determine the return to experience (we'll revisit this when we cover interaction terms)

$\Longrightarrow$ Each coefficient measures a **partial effect**: the change in $E(y)$ when that variable changes by one unit, *holding all other variables fixed*.

Partial Effects: The Central Idea

In the general model $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_K x_{iK} + e_i$:

$$\beta_k = \frac{\partial\, \mathrm{E}(y \mid x_2, \ldots, x_K)}{\partial\, x_k}$$

Verbal interpretation: β_k is the change in the expected value of y when x_k increases by one unit, **holding all other x's constant**.

Compare to simple regression:

	SLR	**Multiple Regression**
β_2 measures	total association	partial effect
"Holding constant"	nothing	all other regressors

$\Longrightarrow$ In multiple regression, **always include "holding $x_3, x_4, \ldots$ constant"** when interpreting a coefficient. Without this qualifier, the interpretation is incomplete.

Interpreting the Wage Equation

Suppose OLS on 526 workers gives:

$$\widehat{\log(\text{wage})} = 0.083 + \underset{(0.007)}{0.092}\,\text{educ} + \underset{(0.005)}{0.039}\,\text{exper} - \underset{(0.0001)}{0.0006}\,\text{exper}^2$$

(standard errors in parentheses)

Education coefficient ($b_2 = 0.092$):

- One more year of education is associated with about a 9.2% increase in wages
- **Holding experience constant**

Compare to the simple regression: $b_2^{\text{SLR}} = 0.076$

- The return to education *increased* when we added experience
- Why? We'll see in the next section

Outline

Why Did the Coefficient Change?

	SLR (educ only)	**MR** (educ + exper + exper^2)
b_{educ}	0.076	0.092

The simple regression estimate of β_{educ} was **biased downward** (0.076 vs. 0.092). Why?

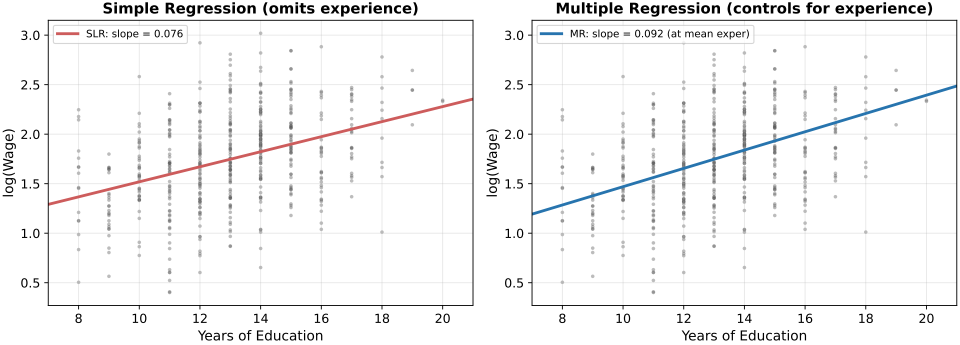
Two facts:

1. Experience increases wages ($\beta_{\text{exper}} > 0$)
2. Education and experience are **negatively correlated**: more schooling means fewer years in the labor market

When we omit experience, the education coefficient absorbs some of the negative correlation with experience, pulling the estimate down.

$\Longrightarrow$ This is **omitted variable bias** (OVB).

Visualizing OVB



Left: SLR picks up a mix of education and experience effects.
Right: MR isolates the partial effect of education, holding experience constant.

The OVB Formula

True model ("long regression"):

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i$$

What we run ("short regression", omitting x_3):

$$y_i = \gamma_1 + \gamma_2 x_{i2} + u_i$$

Auxiliary regression (relationship between included and omitted variable):

$$x_{i3} = \delta_0 + \delta_1 x_{i2} + v_i$$

The short-regression estimator satisfies:

$$\hat{\gamma}_2 = \underbrace{\beta_2}_{\text{true partial effect}} + \underbrace{\beta_3 \times \delta_1}_{\text{omitted variable bias}}$$

(This equality holds exactly in the population.)

Bias $= 0$ only when $\beta_3 = 0$ (omitted variable doesn't affect y) or $\delta_1 = 0$ (omitted variable is uncorrelated with the included one).

OVB in the Wage Equation

$$\hat{\gamma}_{\text{educ}} = \underbrace{\beta_{\text{educ}}}_{+0.092} + \underbrace{\beta_{\text{exper}}}_{\substack{>0 \\ \text{(experience} \\ \text{raises wages)}}} \times \underbrace{\delta_1}_{\substack{<0 \\ \text{(educ \& exper} \\ \text{negatively correlated)}}}$$

	β_{exper}	δ_1	**Bias**
Sign	$(+)$	$(-)$	$(-)$

Negative bias: the SLR estimate (0.076) **understates** the true return to education (0.092).

$\Longrightarrow$ Omitting a relevant variable that is correlated with an included variable biases the OLS estimate. The direction and size of the bias depend on two things: the effect of the omitted variable on y, and its correlation with the included variable.

When Is OVB a Problem?

OVB arises when **both** conditions hold:

1. The omitted variable affects y ($\beta_3 \neq 0$)
2. The omitted variable is correlated with an included regressor ($\delta_1 \neq 0$)

No bias if either condition fails:

- If $\beta_3 = 0$: the omitted variable doesn't belong in the model anyway
- If $\delta_1 = 0$: the omitted variable is uncorrelated with the included one, so the SLR estimate is unaffected

$\Longrightarrow$ The solution: **include the relevant variable in the regression.** That's exactly what multiple regression does.

Outline

How OLS Works with Multiple Regressors

OLS minimizes the sum of squared residuals, just as in SLR:

$$\min_{b_1, b_2, \ldots, b_K} \sum_{i=1}^{N} (y_i - b_1 - b_2 x_{i2} - \cdots - b_K x_{iK})^2$$

Setting partial derivatives to zero gives a system of K equations in K unknowns (the "normal equations").

In SLR ($K = 2$), we derived a closed-form formula for b_2. With more regressors, the formulas get complicated quickly.

Practical approach: Software handles the algebra. You type the command, software returns the estimates.

$\implies$ The mechanics of estimation generalize seamlessly. What changes is the **interpretation** (partial effects) and the **degrees of freedom** ($N - K$ instead of $N - 2$).

Degrees of Freedom: $N - K$

Recall from SLR: we used $N - 2$ degrees of freedom because we estimated 2 parameters (β_1 and β_2).

In multiple regression, we estimate K parameters ($\beta_1, \beta_2, \ldots, \beta_K$):

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} \hat{e}_i^2}{N - K} = \frac{SSE}{N - K}$$

$\Longrightarrow$ Everywhere you used $N - 2$, replace it with $N - K$.

Examples:

- Mincer equation ($K = 4$): df $= N - 4$
- Burger Barn model with PRICE, ADVERT ($K = 3$): df $= N - 3$
- SLR ($K = 2$): df $= N - 2$ (same as before)

Each parameter you estimate "costs" one degree of freedom.

What Changes, What Stays the Same

	SLR ($K = 2$)	**Multiple Regression**
Model	$y_i = \beta_1 + \beta_2 x_i + e_i$	$y_i = \beta_1 + \sum_{k=2}^{K} \beta_k x_{ik} + e_i$
Interpretation	total effect	partial effect
OLS	closed-form formula	software
$\hat{\sigma}^2$	$SSE/(N-2)$	$SSE/(N-K)$
t-distribution	$t_{(N-2)}$	$t_{(N-K)}$
R^2	same formula	same formula
Gauss-Markov	MR1–MR5 $\Longrightarrow$ BLUE	same

$\Longrightarrow$ Most of your SLR toolkit transfers directly. The main new ideas are **partial effects** and **omitted variable bias**.

Outline

R^2 in Multiple Regression

The coefficient of determination has the same formula as in SLR:

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum \hat{e}_i^2}{\sum (y_i - \bar{y})^2}$$

Interpretation: the proportion of variation in y explained by *all* the regressors together.

Problem: Adding any variable to a regression **cannot decrease** R^2, even if that variable is completely irrelevant.

Why? OLS finds the best fit. An extra variable gives OLS more flexibility, so SSE can only stay the same or shrink.

$\Longrightarrow$ R^2 rewards adding variables regardless of whether they belong in the model.

The Adjusted R^2

The adjusted R^2 penalizes for additional regressors:

$$\bar{R}^2 = 1 - \frac{SSE/(N-K)}{SST/(N-1)} = 1 - \frac{N-1}{N-K} \cdot \frac{SSE}{SST}$$

Equivalently:

$$\bar{R}^2 = 1 - \frac{N-1}{N-K}(1 - R^2)$$

Since $\frac{N-1}{N-K} > 1$ when $K > 1$, we always have $\bar{R}^2 \leq R^2$.

What the penalty does:

- Adding a useful variable: SSE drops enough to offset the penalty $\Longrightarrow$ $\bar{R}^2$ increases
- Adding a useless variable: SSE barely changes, but the penalty grows $\Longrightarrow$ $\bar{R}^2$ *decreases*

$\Longrightarrow$ Use $\bar{R}^2$ when comparing models with different numbers of regressors.

R^2 vs. Adjusted R^2: Example

Burger Barn sales data ($N = 75$):

Model	K	R^2	$\bar{R}^2$
$\text{SALES} = \beta_1 + \beta_2\,\text{PRICE} + e$	2	0.267	0.257
$\text{SALES} = \beta_1 + \beta_2\,\text{PRICE} + \beta_3\,\text{ADVERT} + e$	3	0.448	0.433

Adding advertising increased both R^2 ($0.267 \to 0.448$) and $\bar{R}^2$ ($0.257 \to 0.433$).

$\Longrightarrow$ Advertising genuinely improves the model. The $\bar{R}^2$ increase confirms this wasn't just the mechanical effect of adding a variable.

Outline

MR Assumptions: Same Spirit, Broader Scope

The assumptions generalize directly from SLR:

Assumption	Statement
MR1	$y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + e_i$
MR2	$\mathrm{E}(e_i \mid \mathrm{X}) = 0$ (strict exogeneity)
MR3	$\mathrm{Var}(e_i \mid \mathrm{X}) = \sigma^2$ (homoskedasticity)
MR4	$\mathrm{Cov}(e_i, e_j \mid \mathrm{X}) = 0$ for $i \neq j$ (no correlation)
MR5	No exact linear relationship among x's
MR6	$e_i \mid \mathrm{X} \sim N(0, \sigma^2)$ (optional, for finite-sample inference)

What's new?

- MR5 now applies to *all* regressors: no variable can be a perfect linear combination of the others
- Degrees of freedom change from $N - 2$ to $N - K$ throughout

Collinearity: Examples

What would it look like for regressors to carry redundant information?

Exact duplicates: Including both age and birth_year when age $= 2026 -$ birth_year. These carry identical information.

Exhaustive dummies: Including male and female plus an intercept. Since male $+$ female $= 1$ always, one dummy is a perfect function of the other and the constant. This is the **dummy variable trap**: include $J - 1$ dummies for J categories, not J.

Budget shares: If $x_2 + x_3 + x_4 = 1$ for all observations, including all three plus an intercept creates an exact dependency.

$\implies$ In each case, one regressor is a perfect linear function of the others. OLS cannot separate their individual effects.

MR5: The Formal Statement

No explanatory variable is an exact linear function of the others. Formally:

$$c_1 x_{i1} + c_2 x_{i2} + \cdots + c_K x_{iK} = 0 \text{ for all } i \quad \Longrightarrow \quad c_1 = c_2 = \cdots = c_K = 0$$

This also requires that each x_k has some variation across observations (not all the same value).

What happens if MR5 fails? OLS *cannot be computed*. There is no unique solution to the normal equations.

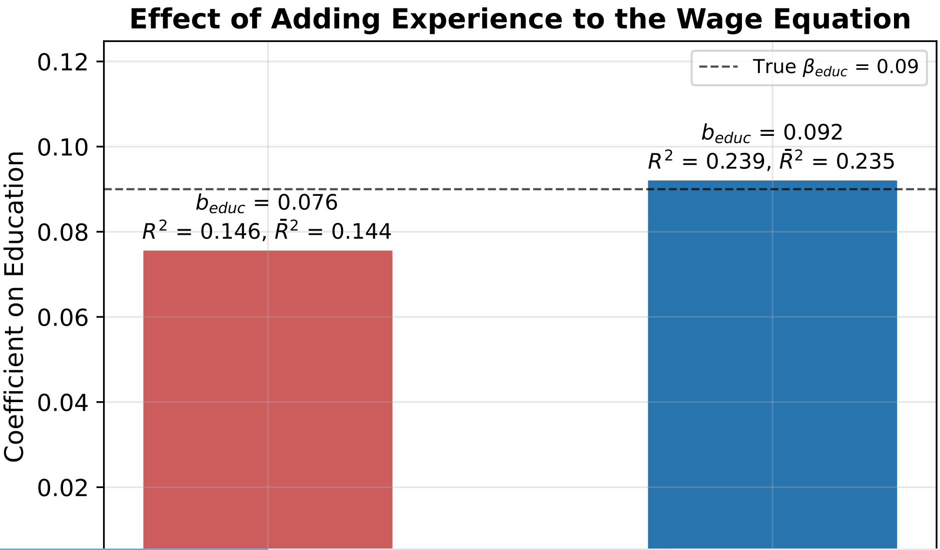
Near-violations (high but not perfect correlation):

- MR5 is technically satisfied, but estimates become imprecise
- This is the "collinearity" problem (Chapter 6)

$\Longrightarrow$ Each regressor must provide information that *no* combination of the other regressors already provides.

Outline

Side-by-Side Comparison



Reading Regression Output

A typical software output table for multiple regression:

Variable	Coefficient	Std. Error	t	p-value
Intercept	0.083	0.110	0.75	0.451
educ	0.092	0.007	12.80	0.000
exper	0.039	0.005	7.37	0.000
exper^2	−0.0006	0.0001	-4.93	0.000
$N = 526$, $K = 4$		$R^2 = 0.239$, $\bar{R}^2 = 0.235$		

What to report:

- Coefficient + standard error for each variable
- N, K, degrees of freedom ($N - K = 522$)
- R^2 and $\bar{R}^2$
- The t-statistics use the $t_{(N-K)}$ distribution (here $t_{(522)}$)

Checklist: Interpreting MR Output

For each coefficient b_k in a multiple regression:

1. **Sign:** Does the direction make economic sense?
2. **Magnitude:** A one-unit increase in x_k is associated with a b_k-unit change in y…
3. …**holding all other variables constant.** (Always say this.)
4. **Significance:** Is $|t| > t_c$? ($t = b_k / \text{se}(b_k)$, df $= N - K$)
5. **Units:** Match the units of b_k to the units of y and x_k.

For the model as a whole:

- R^2: how much variation is explained?
- $\bar{R}^2$: use this to compare models with different numbers of regressors

Outline

What You Should Take Away

1. The multiple regression model: $y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + e_i$
2. Each β_k is a **partial effect**: the change in $E(y)$ per unit change in x_k, *holding all other variables constant*
3. **Omitted variable bias**: leaving out a relevant variable that is correlated with an included regressor biases OLS. Direction: bias $= \beta_{\text{omitted}} \times \delta_1$
4. OLS in MR: same principle (minimize $\sum \hat{e}_i^2$), but degrees of freedom change to $N - K$
5. Use $\bar{R}^2$ (not R^2) to compare models with different numbers of regressors
6. Assumptions MR1–MR6 generalize directly from SLR; the new assumption (MR5) rules out exact linear dependence among regressors

Looking Ahead

Next lecture: Interpreting MR output and the full assumption set

- Confidence intervals and hypothesis tests with $t_{(N-K)}$
- What happens when regressors are highly correlated (collinearity preview)
- Nonlinear relationships: polynomials and interaction terms

The big picture: Multiple regression is the workhorse of applied econometrics. Nearly every empirical paper you read uses it. Getting the interpretation right (partial effects, "holding constant") is the single most important skill in this course.

Thank you!

jakeanderson@g.ucla.edu