

Econ 103: Introduction to Econometrics

Lecture 13 — The Multiple Regression Model

Ryan Longmuir

UCLA

Summer Session C, 2026

Reading: Hill, Griffiths & Lim (5th ed.), §5.1–5.2; Stock & Watson (4th ed.), §6.1–6.3.

Where we are

The simple regression model has an [Achilles heel](#). Its causal reading rests on SR2:

$$\mathbb{E}(e | x) = 0.$$

But e holds *everything else* — and if any omitted factor is correlated with x , SR2 fails and b_2 is **biased** (the ability-in-the-wage-equation problem from Lecture 7).

The cure is to stop hiding confounders in e and instead put them [in the regression](#). With more than one regressor we can finally hold other factors constant — *ceteris paribus* for real. Today:

- the **multiple regression model** and its [partial](#) coefficients,
- the assumptions **MR1–MR6** (one is genuinely new),
- OLS estimation, applied to [Big Andy's Burger Barn](#).

Today's plan

- ① Why more than one regressor?
- ② The model and its partial coefficients
- ③ Assumptions MR1–MR6
- ④ OLS estimation & Big Andy's results

Part 1

Why more than one regressor?

Omitted-variable bias

When does omitting a variable bias OLS? (two conditions)

A left-out variable biases b_2 only if it is **both**

- ① correlated with the included regressor x , *and*
- ② a determinant of the outcome y (so it sits in e).

Example (Class size and test scores (S&W))

*Regress test scores on the student–teacher ratio (STR) alone. Districts with larger classes also have more **English learners** (corr ≈ 0.19), and English learners score lower. “% English learners” is correlated with STR and drives scores — so omitting it **biases** the class-size effect. (It roughly halves once controlled.)*

Both conditions matter: the time-of-day of the test affects scores but is *uncorrelated* with class size — omitting it is harmless.

The fix: put the confounder in the regression

Omitted-variable bias is just SR2 failing: a determinant of y that is correlated with x lives in e , so $\mathbb{E}(e | x) \neq 0$.

Solution. Move it out of e and **into the model as its own regressor**. Then OLS can estimate the effect of x *holding that variable constant* — and the bias it caused disappears.

This is what “control for” means

Adding % English learners lets us compare districts *as if* they had the same share of English learners. Multiple regression does, with continuous data, what we wished we could do by hand: hold the other factors fixed.

Caveat for later: you can only control for what you *observe*. Unobservable confounders (Lecture 7's “ability”) still threaten SR2 — the deeper fix waits for treatment effects (Lecture 20).

Part 2

The model and its partial coefficients

Big Andy's Burger Barn

Running example (HGL). A burger chain in 75 small cities sets a different **price** and **advertising** budget in each, and observes monthly **sales**. How does revenue respond to each — holding the other fixed?

$$\text{SALES} = \beta_1 + \beta_2 \text{PRICE} + \beta_3 \text{ADVERT} + e.$$

(SALES, ADVERT in \$1000; PRICE a \$ price index.) The error e collects everything else: competitors, local demographics, location.

The general model with $K - 1$ regressors and an intercept:

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

Under strict exogeneity the regression function is $\mathbb{E}(y \mid \mathbf{X}) = \beta_1 + \beta_2 x_2 + \cdots + \beta_K x_K$ — now a **plane** (or hyperplane), not a line.

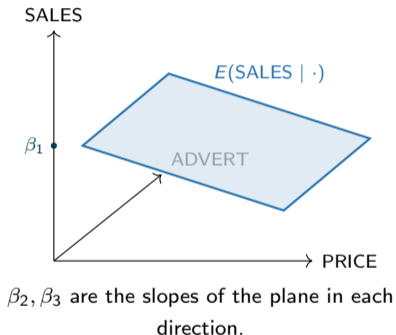
Partial coefficients: ceteris paribus at last

Each slope is a **partial effect** — the change in $\mathbb{E}(y)$ from a one-unit change in *that* regressor, **holding all others fixed**:

$$\beta_k = \frac{\partial \mathbb{E}(y | \mathbf{X})}{\partial x_k} \quad (\text{other } x\text{'s held constant}).$$

- β_2 : effect of PRICE on sales, ADVERT fixed.
- β_3 : effect of ADVERT on sales, PRICE fixed.
- The intercept $\beta_1 = \mathbb{E}(y)$ when all $x = 0$ — often not meaningful, but keep it.

What does “held constant” precisely mean? The Frisch–Waugh–Lovell theorem: β_3 is the effect of ADVERT after the linear influence of PRICE has been *partialled out* of both SALES and ADVERT.



Part 3

Assumptions MR1–MR6

The assumptions, mostly familiar

The multiple-regression assumptions mirror SR1–SR6 — with one true newcomer.

MR1	$y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + e_i$
MR2	$\mathbb{E}(e_i \mathbf{X}) = 0$ (strict exogeneity — now for <i>all</i> regressors)
MR3	$\text{Var}(e_i \mathbf{X}) = \sigma^2$ (homoskedastic)
MR4	$\text{Cov}(e_i, e_j \mathbf{X}) = 0, i \neq j$
MR5	no exact linear relationship among the regressors (new)
MR6	$e_i \mathbf{X} \sim N(0, \sigma^2)$ (optional)

- MR2 must now hold for **every** regressor — the bar for “no confounders” is higher.
- MR2 implies $\mathbb{E}(e_i) = 0$ and $\text{Cov}(e_i, x_{jk}) = 0$ for all regressors k .

MR5: no exact linear relationship

MR5 — no perfect collinearity

No regressor may be written as an **exact linear combination** of the others (including the constant). If one can, OLS **cannot** separate their effects — the formulas divide by zero.

Why it's required — you'd be asking an impossible question:

- Include both % and *fraction* of English learners ($\text{Pct} = 100 \times \text{Frac}$)? OLS can't find “the effect of Pct holding Frac constant” — they move together perfectly.
- It generalizes SR5: “ x must take ≥ 2 values” is the one-regressor special case (a constant regressor equals the intercept's).

MR5 rules out *perfect* collinearity only. *Near*-collinear regressors are allowed — but, as Lecture 14 shows, they inflate standard errors.

What the assumptions buy: Gauss–Markov, again

Gauss–Markov for multiple regression

If **MR1–MR5** hold, the OLS estimators b_1, \dots, b_K are the **Best Linear Unbiased Estimators (BLUE)** of β_1, \dots, β_K .

- Everything from Lecture 7 carries over: linear, unbiased ($\mathbb{E}(b_k | \mathbf{X}) = \beta_k$), minimum variance in that class.
- Add **MR6** (normal errors) and each b_k is normal — giving exact t -based inference (Lecture 15). Without it, the same inference holds *approximately* in large samples (CLT).

The conceptual machinery is unchanged — only the *bookkeeping* (more coefficients, degrees of freedom $N - K$) grows.

Part 4

OLS estimation & Big Andy's results

Least squares, same principle

OLS chooses b_1, \dots, b_K to minimize the sum of squared residuals — the identical idea as Lecture 6, just with more terms:

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

- Setting the K partial derivatives to zero gives K normal equations in K unknowns — solved once, by the computer.
- The formulas are messy by hand (matrix algebra in advanced courses), so we [let software do it](#) and focus on *reading* the output.
- As always: b_1, \dots, b_K are random-variable [estimators](#); one sample's numbers are [estimates](#).

Big Andy's: the fitted equation

OLS on the 75 cities gives (standard errors below):

$$\widehat{\text{SALES}} = 118.91 - 7.908 \text{ PRICE} + 1.863 \text{ ADVERT}, \quad R^2 = 0.448.$$

(6.35) (1.096) (0.683)

Price: $b_2 = -7.908$

Holding advertising fixed, a \$1 price rise lowers mean monthly revenue by **\$7,908** (a realistic 10-cent cut \rightarrow +\$791). Revenue *falls* when price rises \Rightarrow demand is **price-elastic**.

Advertising: $b_3 = 1.863$

Holding price fixed, \$1,000 more on ads raises mean revenue by **\$1,863**. Whether that's *profitable* (is $\beta_3 > 1$?) is a hypothesis test — Lecture 15.

Intercept \$118,914 = predicted sales at zero price *and* zero advertising — economically impossible; keep it only to pin the plane.

Error variance, fit, and a prediction

Error variance (degrees of freedom
 $N - K = 75 - 3 = 72$):

$$\hat{\sigma}^2 = \frac{\text{SSE}}{N - K} = \frac{1718.94}{72} = 23.87,$$

$$\hat{\sigma} = \sqrt{23.87} = 4.89.$$

Fit: $R^2 = 1 - \text{SSE}/\text{SST} = 0.448$ — price and advertising explain **44.8%** of the variation in sales.

Prediction. At PRICE = 5.50,
ADVERT = 1.2:

$$\widehat{\text{SALES}} = 118.91 - 7.908(5.5) + 1.863(1.2) = 77.66.$$

Predicted revenue \$77,656.

Note the only change from simple regression: $\hat{\sigma}^2$ divides by $N - K$ (here $K = 3$), not $N - 2$ — one df per estimated coefficient.

A standing caution

The negative price coefficient does *not* say “cut price to zero.” Estimated models describe the data's neighborhood — **extrapolating** to extreme values invites disaster.

Recap

Why MR

- omitted-variable bias: confounder (i) correlated with x , (ii) drives y
- fix: include it \Rightarrow hold it constant

The model

- $y = \beta_1 + \beta_2 x_2 + \cdots + \beta_K x_K + e$
- $\beta_k = \partial \mathbb{E}(y) / \partial x_k$ (partial / *ceteris paribus*)
- regression function is a plane

Assumptions

- MR1–MR4, MR6 as before; MR5 = no perfect collinearity (new)
- MR1–MR5 \Rightarrow OLS is BLUE

Big Andy's

- $\widehat{\text{SALES}} = 118.9 - 7.91 \text{ PRICE} + 1.86 \text{ ADVERT}$
- demand elastic;
 $\hat{\sigma}^2 = \text{SSE} / (N - K) = 23.87$; $R^2 = 0.448$

Next time (Lecture 14): precision & collinearity

How reliable are these slopes? We build the **variance–covariance matrix**, see what drives the standard errors, and meet the regression headache of **collinearity** — when regressors move together.

Questions?