Econ 103: Introduction to Econometrics

Lecture 14 — Interpreting MR: Variance & Collinearity

Ryan Longmuir

UCLA

Summer Session C, 2026

Reading: Hill, Griffiths & Lim (5th ed.), §5.3, 6.5; Stock & Watson (4th ed.), §6.4, 6.6–6.7.

Where we are

Last lecture we estimated Big Andy's plane:

$$\widehat{\text{SALES}} = 118.91 - 7.908\,\text{PRICE} + 1.863\,\text{ADVERT}.$$

As in simple regression, these slopes are random variables — so the next question is precision: *how much would they wobble across samples?*

Today we answer that, and meet a problem unique to multiple regression:

- the **variance–covariance matrix** and where standard errors come from,
- the four drivers of precision — one of them brand new,
- **collinearity**: what happens when regressors move together,
- and **perfect** collinearity / the **dummy-variable trap**.

Today's plan

1. The variance–covariance matrix
2. What drives precision
3. Collinearity
4. Perfect collinearity & the dummy-variable trap

Part 1

The variance–covariance matrix

From one variance to a whole matrix

With K coefficients we track not just each variance but every covariance between estimators. Stack them in the **variance–covariance matrix** — a diagonal of variances, off-diagonal of covariances:

$$\widehat{\text{Cov}}(b_1, b_2, b_3) = \begin{bmatrix} \widehat{\text{Var}}(b_1) & \widehat{\text{Cov}}(b_1, b_2) & \widehat{\text{Cov}}(b_1, b_3) \\ \widehat{\text{Cov}}(b_1, b_2) & \widehat{\text{Var}}(b_2) & \widehat{\text{Cov}}(b_2, b_3) \\ \widehat{\text{Cov}}(b_1, b_3) & \widehat{\text{Cov}}(b_2, b_3) & \widehat{\text{Var}}(b_3) \end{bmatrix}.$$

- It is built, as always, by plugging $\hat{\sigma}^2 = \text{SSE}/(N - K)$ into the variance formulas.
- The diagonal gives standard errors; the off-diagonal covariances are needed for any *linear combination* of coefficients (Lecture 15).

Big Andy's variance–covariance matrix

With $\hat{\sigma}^2 = 23.87$, software returns

$$\widehat{\text{Cov}}(b_1, b_2, b_3) = \begin{bmatrix} 40.343 & -6.795 & -0.748 \\ -6.795 & 1.201 & -0.020 \\ -0.748 & -0.020 & 0.467 \end{bmatrix}.$$

Standard errors are square roots of the diagonal:

$$\text{se}(b_1) = \sqrt{40.343} = 6.35, \quad \text{se}(b_2) = \sqrt{1.201} = 1.10, \quad \text{se}(b_3) = \sqrt{0.467} = 0.68.$$

These are exactly the standard errors printed under the coefficients last lecture. As a rough reading: across resampled cities we'd expect b_2 within about $\pm 2(1.10) = \pm 2.2$ of β_2.

Part 2

What drives precision

The variance of a slope, dissected

For two regressors, the variance of the price coefficient is

$$\boxed{\mathrm{Var}(b_2 \mid \mathbf{X}) = \frac{\sigma^2}{(1 - r_{23}^2)\,\sum_i (x_{i2} - \bar{x}_2)^2}}$$

where r_{23} is the sample correlation between the two regressors x_2 and x_3.

Three of the four levers are the *same* as simple regression — plus one genuinely new term, $(1 - r_{23}^2)$, in the denominator. Read it carefully on the next slide.

Four drivers of precision

1. Error variance σ^2 (numerator). Noisier model $\Rightarrow$ larger variance. Same as before.
2. Sample size N. More observations enlarge the sum $\sum(x_{i2} - \bar{x}_2)^2 \Rightarrow$ smaller variance.
3. Variation in x_2, $\sum(x_{i2} - \bar{x}_2)^2$. More spread in price $\Rightarrow$ smaller variance.
4. Correlation between regressors r_{23} (new!). The factor $(1 - r_{23}^2)$ shrinks toward 0 as $|r_{23}| \to 1$, so the variance explodes.

The new tension

Items 1–3 are familiar from Lecture 7. Item 4 is the price of multiple regression: when two regressors carry overlapping information, it is hard to pin down *either* one's separate effect.

Part 3

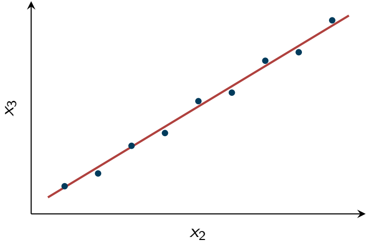
Collinearity

Near collinearity: the precision killer

Collinearity (a.k.a. multicollinearity) = regressors that are highly correlated with one another. By the variance formula, $|r_{23}|$ near 1 makes $(1 - r_{23}^2)$ tiny and the standard errors huge.

Why: if x_2 and x_3 almost always move together, the data contain little independent variation in x_2 to identify its *own* effect. You cannot separate effects that never separate.

Example: add "% immigrants" to a scores regression already containing "% English learners." The two track each other; the coefficient on either becomes imprecise.



Collinear regressors: $x_3 \approx x_2$.

Recognizing and living with collinearity

Symptoms

- large standard errors, wide confidence intervals
- coefficients insignificant even when the group clearly matters
- high overall R^2 but few significant t's
- estimates swing wildly when a variable is added/dropped

What to do

- get more / better data with independent variation (Andy could vary price and ads more independently across cities)
- drop a redundant regressor — carefully
- or simply accept it: imperfect collinearity is *not an error*, just a limit of what the data can tell you

Key point: collinearity does *not* bias OLS — it stays BLUE. It only makes the (still-unbiased) estimates imprecise.

Part 4

Perfect collinearity & the dummy-variable trap

Perfect collinearity: OLS breaks down

If a regressor is an exact linear function of the others, $r_{23}^2 = 1$, the variance formula divides by zero — MR5 is violated and OLS cannot be computed at all.

- % vs. fraction of English learners: $\text{Pct} = 100 \times \text{Frac}$ — perfectly redundant.
- % English speakers vs. % learners: $\text{PctES} = 100 - \text{PctEL}$ (with the constant) — again exact.

It's a logical error, not a data problem

OLS is being asked an impossible question — "the effect of Pct holding Frac constant" when they *always* move together. The fix is to respecify: drop the redundant regressor. Software will warn you (or silently drop one).

The dummy-variable trap

A classic way to stumble into perfect collinearity. Partition cities into mutually exclusive categories — say `Rural`, `Suburban`, `Urban` — each a 0/1 indicator. Every city is in exactly one, so

$$\texttt{Rural}_i + \texttt{Suburban}_i + \texttt{Urban}_i = 1 = \text{the intercept's regressor.}$$

Including all G dummies and an intercept $\Rightarrow$ perfect collinearity.

The fix: drop one

Include only $G - 1$ dummies (keep the intercept), leaving one category as the **base/reference** group. Each included coefficient is then the difference from the base. (Or: keep all G dummies and drop the intercept.)

Indicator variables — intercept shifts, slope dummies, reference groups — get a full treatment in Lecture 19. Here it's just the trap to avoid.

Recap

Precision machinery

- $\hat{\sigma}^2 = \mathrm{SSE}/(N-K)$; var–cov matrix holds variances (diag) & covariances
- $\mathrm{se}(b_k) = \sqrt{\text{diagonal}}$; Andy: 1.10, 0.68
- $\mathrm{Var}(b_2) = \dfrac{\sigma^2}{(1-r_{23}^2)\sum(x_{i2}-\bar{x}_2)^2}$

Four drivers

- small σ^2, big N, spread-out x_2, low r_{23}

Collinearity

- near: huge SE, wide CIs, insignificance — but *still unbiased*
- fix: more independent variation, or accept the limit
- perfect ($r^2 = 1$): MR5 violated, OLS undefined
- dummy-variable trap: drop one dummy (or the intercept)

Next time (Lecture 15): inference in multiple regression

With standard errors in hand, we test and build intervals — t-tests on single coefficients (df $= N-K$), CIs, and tests of linear combinations of several coefficients (Andy's price-and-ad strategy).

Questions?