Econ 103: Introduction to Econometrics

Lecture 16 — Interaction Terms

Ryan Longmuir

UCLA

Summer Session C, 2026

Reading: Hill, Griffiths & Lim (5th ed.), §5.6; Stock & Watson (4th ed.), §8.2–8.3.

Where we are

Every multiple-regression coefficient so far has been a constant partial effect: β_3 is the effect of advertising, the *same* at every level of advertising, for every firm.

But economics is full of effects that *change*:

- advertising shows diminishing returns — the next dollar buys less than the last;
- the return to experience may depend on education;
- how income drives spending may depend on age.

Today we let marginal effects vary, using two devices that are still ordinary OLS:

- **polynomials** (x^2) — an effect that depends on *its own* level,
- **interactions** ($x_2 \times x_3$) — an effect that depends on *another* variable,
- and we cash them in for economic optimization (where MB $=$ MC).

Today's plan

1. Polynomials in multiple regression
2. Interaction terms
3. Economic optimization

Part 1

Polynomials in multiple regression

Diminishing returns to advertising

A linear model forces a *constant* advertising effect β_3 — but the 10th $1,000 of ads surely does less than the 1st. Add a squared term so the slope can bend:

$$\text{SALES} = \beta_1 + \beta_2\,\text{PRICE} + \beta_3\,\text{ADVERT} + \beta_4\,\text{ADVERT}^2 + e.$$

Now the marginal effect of advertising is *not* a number — it's a function:

$$\frac{\partial\,\mathbb{E}(\text{SALES})}{\partial\,\text{ADVERT}} = \beta_3 + 2\beta_4\,\text{ADVERT}.$$

For diminishing returns we expect $\beta_3 > 0$ (helps at first) and $\beta_4 < 0$ (tapering).

Crucially, this is a *multiple* regression: ADVERT and ADVERT^2 are two regressors. In simple regression (Lecture 12) we could only fit one of them at a time.

Big Andy's, with diminishing returns

OLS (HGL) gives

$$\widehat{SALES} = 109.72 - 7.640\,PRICE$$

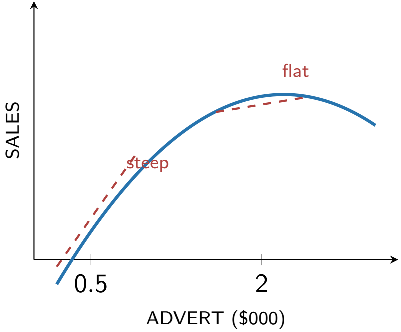
$$+\underset{(3.556)}{12.151}\,ADVERT - \underset{(0.941)}{2.768}\,ADVERT^2.$$

Both signs as expected; $ADVERT^2$ is significant. The estimated marginal effect is

$$\frac{\partial\widehat{SALES}}{\partial ADVERT} = 12.151 - 5.536\,ADVERT:$$

$$\text{at } \$500: 9.38, \qquad \text{at } \$2{,}000: 1.08.$$

An extra $1,000 of ads is worth far less once you already advertise heavily.



Curve flattens: diminishing returns.

Polynomials are everywhere in economics

Cost and product curves are inherently curved — polynomials capture them while staying linear in the parameters:

Average cost (U-shaped): quadratic

$AC = \beta_1 + \beta_2 Q + \beta_3 Q^2 + e$, slope $\beta_2 + 2\beta_3 Q$.
Expect $\beta_2 < 0$, $\beta_3 > 0$.

Total cost (S-shaped): cubic

$TC = \alpha_1 + \alpha_2 Q + \alpha_3 Q^2 + \alpha_4 Q^3 + e$,
marginal cost $= \alpha_2 + 2\alpha_3 Q + 3\alpha_4 Q^2$.

The interpretation habit

A polynomial coefficient is *not* a slope. Always report the marginal effect dy/dx evaluated at chosen values of x (low/median/high), or plot the curve. "The effect of x" is now a moving target.

One practical wrinkle: x and x^2 can be highly correlated, sometimes inflating their standard errors (collinearity, Lecture 14).

Part 2

Interaction terms

When the effect of one variable depends on another

A polynomial lets an effect depend on *its own* level. An **interaction** lets it depend on *another* variable — include the product $x_2 x_3$:

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 (x_2 \times x_3) + e.$$

Differentiate: the marginal effect of x_2 now *slides* with x_3:

$$\frac{\partial\, \mathbb{E}(y)}{\partial x_2} = \beta_2 + \beta_4\, x_3, \qquad \frac{\partial\, \mathbb{E}(y)}{\partial x_3} = \beta_3 + \beta_4\, x_2.$$

What β_4 means (S&W)

β_4 is the effect of raising *both* x_2 and x_3, above and beyond the sum of their separate effects. If $\beta_4 = 0$, the effects are additive and separable; if not, they reinforce ($\beta_4 > 0$) or offset ($\beta_4 < 0$) each other.

The motivating case: age $\times$ income $\rightarrow$ pizza

Does the effect of income on pizza spending depend on age? Model it:

$$\text{PIZZA} = \beta_1 + \beta_2\,\text{AGE} + \beta_3\,\text{INCOME} + \beta_4\,(\text{AGE} \times \text{INCOME}) + e.$$

The marginal effect of income is

$$\frac{\partial\,\mathbb{E}(\text{PIZZA})}{\partial\,\text{INCOME}} = \beta_3 + \beta_4\,\text{AGE}.$$

- If $\beta_4 < 0$: an extra dollar of income raises pizza spending less for older people — the income effect *fades* with age.
- Without the interaction you'd report one income effect for everyone — hiding exactly the pattern of interest.

Note: β_3 alone is now the income effect *only at* $\text{AGE} = 0$ — rarely meaningful. With interactions, never read a "main effect" in isolation.

A worked interaction: education and experience

Do education and experience reinforce each other in the labor market?

$$\text{WAGE} = \beta_1 + \beta_2\,\text{EDUC} + \beta_3\,\text{EXPER} + \beta_4\,(\text{EDUC} \times \text{EXPER}) + e.$$

OLS on the CPS data (HGL, `cps5_small`):

$$\widehat{\text{WAGE}} = -18.76 + 2.656\,\text{EDUC} + 0.2384\,\text{EXPER} - 0.002747\,(\text{EDUC} \times \text{EXPER}).$$

Return to an extra year of experience, $\dfrac{\partial \text{WAGE}}{\partial \text{EXPER}} = 0.2384 - 0.002747\,\text{EDUC}$:

$$\text{at EDUC} = 8 : \$0.22/\text{hr}, \qquad \text{at EDUC} = 16 : \$0.19/\text{hr}.$$

The (small, here insignificant) negative β_4 hints that more schooling makes an extra year of experience slightly *less* valuable — a substitutes story.

Binary interactions — a preview

Interactions are even more common with a 0/1 indicator: interacting a dummy with a continuous x gives the two groups *different slopes* (and a dummy alone shifts the *intercept*).

$y = \beta_1 + \beta_2 x + \beta_3 D$	different intercepts, same slope
$y = \beta_1 + \beta_2 x + \beta_3 D + \beta_4 (x \times D)$	different intercepts *and* slopes

e.g. does the return to the student–teacher ratio differ in districts with many vs. few English learners? Interact STR with a high-EL dummy.

Held for Lecture 19

Indicator variables — intercept shifts, slope dummies, reference groups, the house-price UTOWN/POOL example — get their own lecture. Today's tools (the product term, the moving marginal effect) are *exactly* the machinery you'll use there.

Part 3

Economic optimization

Varying marginal effects let us optimize

A constant slope can never have an interior optimum — but a varying marginal effect can. Economic logic: push a choice until **marginal benefit = marginal cost**.

Big Andy's optimal advertising. From the ADVERT^2 model, the marginal revenue of \$1 more advertising is $\beta_3 + 2\beta_4\,\text{ADVERT}$. The marginal cost of \$1 of advertising is \$1. Set them equal:

$$\beta_3 + 2\beta_4\,\text{ADVERT}_0 = 1 \quad \Longrightarrow \quad \text{ADVERT}_0 = \frac{1-\beta_3}{2\beta_4}.$$

Plug in the estimates:

$$\widehat{\text{ADVERT}}_0 = \frac{1-12.151}{2(-2.768)} = 2.014 \Rightarrow \text{ optimal } \approx \$2{,}014/\text{month}.$$

The optimum is a nonlinear function of coefficients

$\widehat{\text{ADVERT}}_0 = (1 - b_3)/(2b_4)$ divides one estimator by another — a nonlinear function, so the tidy linear-combination variance rule no longer applies exactly.

The delta method (in brief)

Approximate the standard error of a smooth function $g(b_3, b_4)$ using its derivatives and the var–cov matrix — valid in *large samples*. Software does it automatically. For Andy: $\mathrm{se}(\widehat{\text{ADVERT}}_0) = 0.129$, giving an approximate 95% interval

$$2.014 \pm t_c(0.129) = [\,1.757,\ 2.271\,] \ \Rightarrow\ \$1{,}757 \text{ to } \$2{,}271.$$

Same idea answers "how many years of experience maximize wages?" (set $\partial\text{WAGE}/\partial\text{EXPER} = 0$ in a quadratic) — regression doing genuine economic optimization, with an honest margin of error.

Recap

Polynomials

- add x^2: marginal effect $\beta_3 + 2\beta_4 x$ (varies with own level)
- Andy: ads 9.38 $\to$ 1.08 — diminishing returns
- cost/product curves; report slope at chosen x

Interactions

- add $x_2 \times x_3$: $\partial y / \partial x_2 = \beta_2 + \beta_4 x_3$
- age$\times$income$\to$pizza; educ$\times$exper$\to$wage
- never read a "main effect" alone

Optimization

- set MB = MC: $\beta_3 + 2\beta_4 \text{ADVERT}_0 = 1$
- Andy's optimum \$2,014, CI [\$1757, \$2271]
- nonlinear function $\Rightarrow$ delta method SE

Binary interactions

- dummy $\times$ continuous $\to$ different slopes (Lecture 19)

Next time (Lecture 17): *F*-tests

Is the whole curvature/interaction block worth keeping? Testing *several* coefficients at once ($\beta_4 = 0$ and $\beta_5 = 0$) needs the *F*-test — restricted vs. unrestricted models, overall significance,

Questions?