Interactions and Nonlinearities

Gov 2001: Quantitative Social Science Methods I

Week 11, Lecture 21

Spring 2026

For Today

Required Reading

- Blackwell, Chapter 7 (pp. 139–157)
- Aronow & Miller, §4.2 (pp. 156–170)

Today: Making regression more flexible with interactions and transformations.

Roadmap

1. Interaction terms
2. Interpreting interactions
3. Logarithmic transformations
4. Polynomial regression
5. When to use what

Part I: Interaction Terms

Why Interactions?

So far: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

This assumes the effect of X_1 on Y is the **same** regardless of X_2.

But what if the effect depends on context?

- Does education boost wages more for men or women?
- Does a drug work better for younger or older patients?
- Does democracy improve growth more in rich or poor countries?

Interaction terms let the effect of X_1 vary with X_2.

The Interaction Model

Add a product term:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 \times X_2) + \varepsilon$$

The partial effect of X_1 on Y:

$$\frac{\partial Y}{\partial X_1} = \beta_1 + \beta_3 X_2$$

The effect of X_1 now **depends on the value of X_2**!

Interaction with a Binary Variable

Let $D \in \{0, 1\}$ be a dummy variable (e.g., female = 1).

$$Y = \beta_0 + \beta_1 X + \beta_2 D + \beta_3 (X \times D) + \varepsilon$$

For $D = 0$:

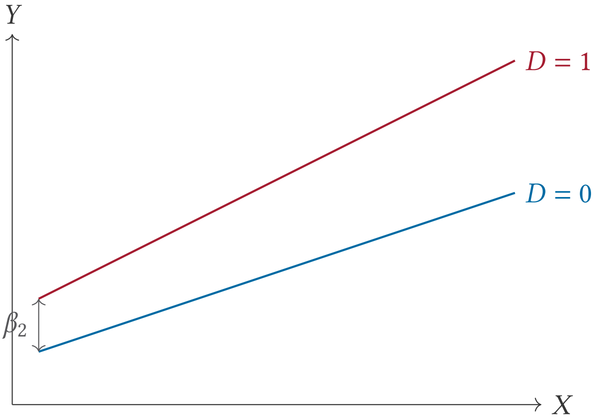
$$Y = \beta_0 + \beta_1 X$$

For $D = 1$:

$$Y = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) X$$

Two different lines: different intercept AND different slope.

Visualizing the Interaction



β_2 = difference in intercepts
β_3 = difference in slopes

Interpreting the Coefficients

$$Y = \beta_0 + \beta_1 X + \beta_2 D + \beta_3 (X \times D) + \varepsilon$$

- β_0: Intercept when $D = 0$
- β_1: Slope of X when $D = 0$
- β_2: Difference in intercept between $D = 1$ and $D = 0$
- β_3: Difference in slope between $D = 1$ and $D = 0$

Test for interaction: $H_0 : \beta_3 = 0$
If $\beta_3 = 0$, the effect of X doesn't depend on D.

Example: Returns to Education by Gender

$$\widehat{\log(\text{wage})} = 1.2 + 0.10 \times \text{Educ} + 0.15 \times \text{Female} - 0.02 \times (\text{Educ} \times \text{Female})$$

For men (Female = 0):

$$\widehat{\log(\text{wage})} = 1.2 + 0.10 \times \text{Educ}$$

For women (Female = 1):

$$\widehat{\log(\text{wage})} = 1.35 + 0.08 \times \text{Educ}$$

Women have higher baseline but lower returns to education.

Political Science Example: Campaign Spending by Competitiveness

$$\widehat{\text{Vote share}} = 45 + 0.8 \times \text{Ads} + 5 \times \text{Competitive} + 1.2 \times (\text{Ads} \times \text{Competitive})$$

In safe seats (Competitive = 0):

$$\widehat{\text{Vote share}} = 45 + 0.8 \times \text{Ads}$$

In competitive races (Competitive = 1):

$$\widehat{\text{Vote share}} = 50 + 2.0 \times \text{Ads}$$

Campaign ads have **larger effects** in competitive districts.
The interaction term ($\beta_3 = 1.2$) captures this differential effect.

Interaction Between Two Continuous Variables

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 \times X_2) + \varepsilon$$

Marginal effect of X_1:

$$\frac{\partial Y}{\partial X_1} = \beta_1 + \beta_3 X_2$$

Effect depends on value of X_2.

Marginal effect of X_2:

$$\frac{\partial Y}{\partial X_2} = \beta_2 + \beta_3 X_1$$

Effect depends on value of X_1. (Symmetric!)

Part II: Interpreting Interactions

Common Mistakes with Interactions

Mistake 1: Only including the interaction, not the main effects.
Wrong: $Y = \beta_0 + \beta_3(X_1 \times X_2) + \varepsilon$
Right: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3(X_1 \times X_2) + \varepsilon$

Mistake 2: Interpreting β_1 as "the effect of X_1"
With interactions, β_1 is the effect of X_1 **when** $X_2 = 0$.

Mistake 3: Forgetting that the effect varies.
There's no single "effect of X_1"—it depends on X_2.

Centering Variables

The interpretation of β_1 depends on where $X_2 = 0$.

Problem: $X_2 = 0$ might be outside the data range or meaningless.
(e.g., "effect of education when experience = 0 years")

Solution: Center variables at meaningful values.

$$\tilde{X}_2 = X_2 - c$$

Now β_1 is the effect of X_1 when $X_2 = c$.
Common choices: $c = \bar{X}_2$ (mean) or a substantively interesting value.

Part III: Logarithmic Transformations

Why Use Logarithms?

Many variables are right-skewed:

- Income, wealth, firm size, population, GDP
- A few very large values, many smaller values

Taking logs:

- Compresses large values
- Makes distribution more symmetric
- Allows percentage interpretation

Note: Log only defined for positive values!

Log-Level Model

$$\log(Y) = \beta_0 + \beta_1 X + \varepsilon$$

Interpretation of β_1:
A one-unit increase in X is associated with a $100 \times \beta_1$ percent change in Y.

Why?

$$\frac{\partial \log(Y)}{\partial X} = \beta_1 \quad \Rightarrow \quad \frac{\partial Y / Y}{\partial X} = \beta_1$$

Example: $\beta_1 = 0.05$ means a one-unit increase in X is associated with 5% higher Y.

Level-Log Model

$$Y = \beta_0 + \beta_1 \log(X) + \varepsilon$$

Interpretation of β_1:
A 1% increase in X is associated with a $\beta_1/100$ unit change in Y.

Why?

$$\frac{\partial Y}{\partial \log(X)} = \beta_1 \quad \Rightarrow \quad \frac{\partial Y}{\partial X/X} = \beta_1$$

Example: $\beta_1 = 0.5$ means a 1% increase in X is associated with 0.005 units higher Y.

Log-Log Model

$$\log(Y) = \beta_0 + \beta_1 \log(X) + \varepsilon$$

Interpretation of β_1:
A 1% increase in X is associated with a β_1 percent change in Y.

β_1 is an **elasticity**.

Example: If $\beta_1 = 0.3$ in a regression of log(wage) on log(experience):
A 10% increase in experience is associated with 3% higher wages.

Summary: Log Interpretations

Model	**Equation**	**Interpretation of β_1**
Level-Level	$Y = \beta_0 + \beta_1 X$	$\Delta X = 1 \Rightarrow \Delta Y = \beta_1$
Log-Level	$\log Y = \beta_0 + \beta_1 X$	$\Delta X = 1 \Rightarrow \%\Delta Y \approx 100\beta_1$
Level-Log	$Y = \beta_0 + \beta_1 \log X$	$\%\Delta X = 1 \Rightarrow \Delta Y = \beta_1 / 100$
Log-Log	$\log Y = \beta_0 + \beta_1 \log X$	$\%\Delta X = 1 \Rightarrow \%\Delta Y = \beta_1$

Part IV: Polynomial Regression

Why Polynomial Terms?

Linear regression: $Y = \beta_0 + \beta_1 X + \varepsilon$
Assumes a **straight line** relationship.

What if the relationship is curved?

- Diminishing returns (e.g., experience and wages)
- U-shaped relationships (e.g., age and happiness)
- Threshold effects

Solution: Add polynomial terms.

Quadratic Regression

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$$

Marginal effect of X:

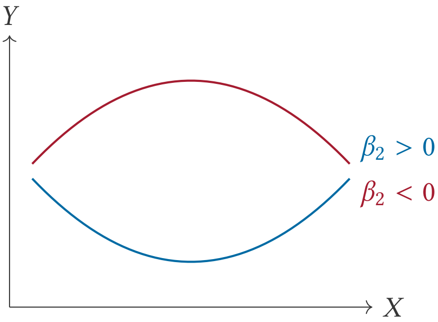
$$\frac{\partial Y}{\partial X} = \beta_1 + 2\beta_2 X$$

The effect of X depends on the level of X!

Shape:

- $\beta_2 > 0$: U-shaped (convex)
- $\beta_2 < 0$: Inverted-U (concave)

Visualizing Quadratic Relationships



Quadratic terms allow for curvature and turning points.

Finding the Turning Point

For $Y = \beta_0 + \beta_1 X + \beta_2 X^2$:

Set the derivative to zero:

$$\frac{\partial Y}{\partial X} = \beta_1 + 2\beta_2 X = 0$$

Solve for X:

$$X^* = -\frac{\beta_1}{2\beta_2}$$

Example: If $\hat{\beta}_1 = 0.06$ and $\hat{\beta}_2 = -0.001$:

$$X^* = -\frac{0.06}{2 \times (-0.001)} = 30$$

The effect of X is positive below 30, negative above 30.

Higher-Order Polynomials

Cubic:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \varepsilon$$

General polynomial of degree p:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \cdots + \beta_p X^p + \varepsilon$$

Trade-off:

- Higher degree = more flexible fit
- But: harder to interpret, risk of overfitting
- Usually $p = 2$ (quadratic) is enough

Part V: When to Use What

Practical Guidelines

Interactions:

- Use when you believe the effect of X_1 differs across groups or levels of X_2
- Binary $\times$ continuous: allows different slopes by group
- Theory should guide: why would effects differ?

Logs:

- Use for right-skewed positive variables
- Use when percentage changes are more meaningful than absolute changes
- Common for: income, prices, population, GDP

Polynomials:

- Use when you expect diminishing returns or turning points
- Common for: age, experience, time trends

It's Still "Linear" Regression

All of these models are **linear in parameters**:

- $Y = \beta_0 + \beta_1 X + \beta_2 D + \beta_3 (X \cdot D)$
- $\log(Y) = \beta_0 + \beta_1 \log(X)$
- $Y = \beta_0 + \beta_1 X + \beta_2 X^2$

We can always write: $Y = \mathbf{X}\boldsymbol{\beta} + \varepsilon$
where $\mathbf{X}$ might include products, logs, or powers of the original variables.

OLS still applies!

Summary

Key tools for flexible regression:

1. **Interactions**: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 \times X_2)$
   Effect of X_1 varies with X_2
2. **Logs**: $\log(Y) = \beta_0 + \beta_1 X$ or $Y = \beta_0 + \beta_1 \log(X)$
   Percentage interpretations; handles skewed data
3. **Polynomials**: $Y = \beta_0 + \beta_1 X + \beta_2 X^2$
   Curved relationships; diminishing returns

Looking Ahead

Next lecture: Matrix Notation and F-tests

- OLS in matrix form: $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$
- Testing multiple coefficients simultaneously
- F-tests for joint hypotheses

Interactions let the effect of one variable
depend on another variable.

Log transformations give percentage interpretations.

Polynomial terms capture curved relationships.

All are still linear regression—OLS applies.