Lecture 5: Nonlinear Regression Functions

Introduction to Econometrics, Spring 2026

Zhaopeng Qu

Nanjing University Business School

April 15 2026

1. Review of the previous lecture
2. Nonlinear Regression Functions
3. Nonlinear in Xs
4. Polynomials in X
5. Logarithms
6. Interactions Between Independent Variables
7. Application: Jia and Ku (2019)
8. Summary

Review of the previous lecture

OLS Regression and Hypothesis Testing

- Our estimates $\hat{\beta}$ come from **one sample**, so they contain **sampling error**. Statistical inference lets us use that sample to learn about the relationship between X and Y **in the population**.
    - **Hypothesis testing** uses *sample evidence* to evaluate claims about *population parameters*.
    - A **confidence interval** is a range of plausible values for a parameter (at a chosen confidence level).

OLS Regression and Hypothesis Testing

- **Hypothesis testing** in OLS regressions
  - One coefficient at a time: the **t-statistic**.
  - In large samples, the t-statistic is approximately **normal** under the null.
- The **standard error** (S.E.) measures the sampling variability of $\hat{\beta}$—roughly the **standard deviation** we would see if we could redraw the sample many times.
- The **t-statistic** is

$$t = \frac{\hat{\beta} - \beta}{SE(\hat{\beta})}.$$

- A **95% confidence interval** is often written as

$$\hat{\beta} \pm t_{\text{crit}} \times SE(\hat{\beta}).$$

With **large** n, $|t_{\text{crit}}| \approx 1.96$ is a widely used approximation.

OLS Regression and Hypothesis Testing

- **Assumption 4**: If errors are **homoskedastic** in OLS,

$$\mathrm{Var}(u_i \mid X_i) = \sigma_u^2,$$

then $\hat{\beta}^{OLS}$ is the **best linear unbiased estimator (BLUE)** among the class of linear unbiased estimators.

- In practice, homoskedasticity often fails.
- **Homoskedasticity** is only a special case of **heteroskedasticity**, so heteroskedasticity-robust standard errors remain valid—even when errors happen to be homoskedastic.
- **In summary:** in applications, report **heteroskedasticity-robust standard errors** unless theory or the empirical design clearly justifies an alternative. Later we will adjust SEs for other problems (for example, **clustered** errors).

OLS Regression and Hypothesis Testing

- Two or more coefficients: the **F-statistic**
  - Individual t-tests do not tell you whether a *set of coefficients is jointly zero*.
  - The **F-statistic** is closely related to **Wald**, **likelihood ratio**, and **Lagrange multiplier** tests(beyond the scope of this course).
- Tests and confidence intervals discipline claims about population relationships, keeping inference **transparent** and appropriately cautious.

Nonlinear Regression Functions

Introduction

- Recall the **linear regression model**.

Linear Regression Model

The observations (Y_i, X_i) are i.i.d., and each satisfies the linear regression equation

$$Y_i = \beta_0 + \beta_1 X_{1,i} + ... + \beta_k X_{k,i} + u_i.$$

- So far we have treated $E[Y \mid X]$ as linear in the regressors we include. In practice, many conditional means **bend**, especially when variables enter in **levels** rather than logs.

Introduction: the big picture

- A general population regression model is

$$Y_i = f(X_{1,i}, X_{2,i}, ..., X_{k,i}) + u_i$$

- **Parametric methods** fix a **functional form** (a family of functions) and **estimate** unknown parameters.
  - **Linear in** X (previous lectures).
  - **Nonlinear in** X (this lecture).
- **Nonparametric methods** flexibly estimate $f(\cdot)$ without fixing a narrow parametric family (beyond mild smoothness assumptions).

Nonlinear Regression Functions

- **To what extent can we retain OLS while relaxing linearity in the regressors?**
  Two broad extensions:

1. **Nonlinear in X (this lecture)**
   - **Polynomials, logarithms, and interactions**
   - Multiple regression still applies; the innovation is that the **CEF** can curve or shift with X.
   - Relative to introductory OLS exercises, the main change is **how we read** marginal effects and individual coefficients.
2. **Nonlinear in β or Y (later)**
   - **Discrete** or **limited dependent variables**
   - A regression line in X may be a poor approximation to $E[Y \mid X]$.
   - Parameters often enter nonlinearly (e.g., logistic or exponential links); estimation uses **nonlinear least squares (NLS)** or **maximum likelihood (MLE)** rather than linear OLS.

Marginal effects in nonlinear regression

- In a linear model, $Y_i = \beta_0 + \beta_1 X_{1,i} + ... + \beta_k X_{k,i} + u_i$:
  - The **marginal effect** of X_j—*the effect on Y of increasing X_j by one unit, holding the other regressors fixed*—is **constant** and equals β_j:

$$\beta_j = \frac{\partial Y_i}{\partial X_{ji}}.$$

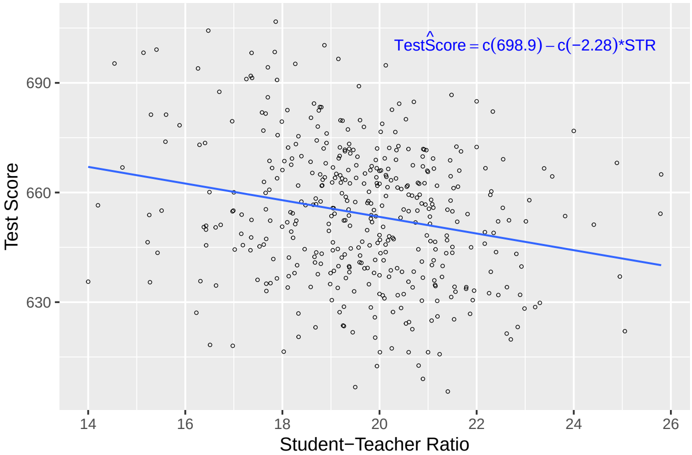
- In a **nonlinear** CEF, $Y_i = f(X_{1,i}, X_{2,i}, ..., X_{k,i}) + u_i$:
  - Marginal effects **vary** with the values of the regressors (X_i itself and sometimes other X_j) because

$$\frac{\partial Y_i}{\partial X_{ji}} = \frac{\partial f(X_{1,i}, X_{2,i}, ..., X_{k,i})}{\partial X_{ji}}.$$

- So **interpreting** individual slope coefficients is **less direct** than in the linear case; we often report **marginal effects** at representative values of X.

Nonlinear in Xs

The TestScore – STR relation looks linear (maybe)



But the TestScore – Income relation looks nonlinear

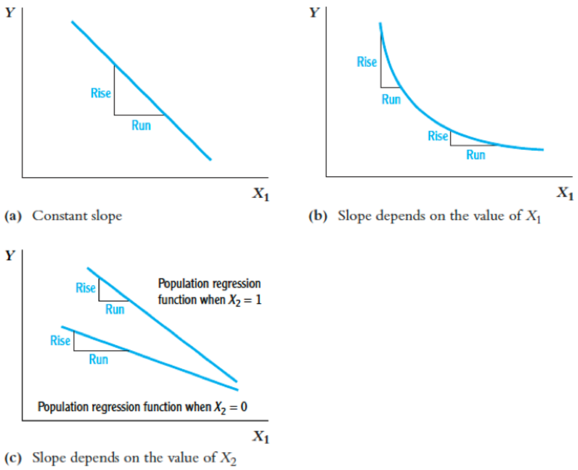
- A straight line **overstates** the relationship when income is very high or very low and **understates** it for middle-income districts.

Three complementary tools

1. **Polynomials in X**
   - Approximate $E[Y \mid X]$ with a quadratic, cubic, or higher-order polynomial in X.
2. **Logarithmic transformations**
   - Take logs of Y and/or X.
   - You then read slopes as **percentage changes**, **elasticities**, or **semi-elasticities** (depending on which side is logged).
3. **Interactions**
   - Let the effect of one regressor depend on the level of another.
   - Natural when studying **heterogeneous effects** or **channels**.

Population Regression Functions with Different Slopes

FIGURE 8.1 Population Regression Functions with Different Slopes



In Figure 8.1a, the population regression function has a constant slope. In Figure 8.1b, the slope of the population regression function depends on the value of X_1. In Figure 8.1c, the slope of the population regression function

The Effect of a Change in X in Nonlinear Functions

The Expected Change on Y of a Change in X_1 in the Nonlinear Regression Model (8.3)

KEY CONCEPT 8.1

The expected change in Y, ΔY, associated with the change in X_1, ΔX_1, holding $X_2, \ldots, X_k$ constant, is the difference between the value of the population regression function before and after changing X_1, holding $X_2, \ldots, X_k$ constant. That is, the expected change in Y is the difference:

$$\Delta Y = f(X_1 + \Delta X_1, X_2, \ldots, X_k) - f(X_1, X_2, \ldots, X_k). \qquad (8.4)$$

The estimator of this unknown population difference is the difference between the predicted values for these two cases. Let $\hat{f}(X_1, X_2, \ldots, X_k)$ be the predicted value of Y based on the estimator $\hat{f}$ of the population regression function. Then the predicted change in Y is

$$\Delta \hat{Y} = \hat{f}(X_1 + \Delta X_1, X_2, \ldots, X_k) - \hat{f}(X_1, X_2, \ldots, X_k). \qquad (8.5)$$

Polynomials in X

Example: the test score–income relationship

- If a straight line does not fit district income and test scores, what should we try next?
- Two simple building blocks:
  - **Quadratic in income**:

$$TestScore_i = \beta_0 + \beta_1 Income_i + \beta_2 (Income_i)^2 + u_i$$

  - **Cubic in income**:

$$TestScore_i = \beta_0 + \beta_1 Income_i + \beta_2 (Income_i)^2 + \beta_3 (Income_i)^3 + u_i$$

- How do we estimate them?
  - Treat **powers of income** as additional **regressors** in an otherwise standard multiple regression.
  - Estimation is still OLS; inference is the usual story (with heteroskedasticity-robust SEs).

Estimation of the quadratic specification in R

```
#>
#> Call:
#>    felm(formula = testscr ~ avginc + I(avginc^2), data = ca)
#>
#> Residuals:
#>     Min      1Q  Median      3Q     Max
#> -44.416  -9.048   0.440   8.348  31.639
#>
#> Coefficients:
#>              Estimate Robust s.e t value Pr(>|t|)
#> (Intercept) 607.30174    2.90175 209.288   <2e-16 ***
#> avginc        3.85100    0.26809  14.364   <2e-16 ***
#> I(avginc^2)  -0.04231    0.00478  -8.851   <2e-16 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Residual standard error: 12.72 on 417 degrees of freedom
#> Multiple R-squared(full model): 0.5562   Adjusted R-squared: 0.554
#> Multiple R-squared(proj model): 0.5562   Adjusted R-squared: 0.554
#> F-statistic(full model, *iid*):261.3 on 2 and 417 DF, p-value: < 2.2e-16
```

Estimation of the cubic specification in R

```
#>
#> Call:
#>    felm(formula = testscr ~ avginc + I(avginc^2) + I(avginc3), data = ca)
#>
#> Residuals:
#>    Min      1Q Median      3Q     Max
#> -44.28  -9.21   0.20    8.32   31.16
#>
#> Coefficients:
#>                Estimate Robust s.e t value Pr(>|t|)
#> (Intercept)  6.001e+02  5.102e+00 117.615  < 2e-16 ***
#> avginc       5.019e+00  7.074e-01   7.095 5.61e-12 ***
#> I(avginc^2) -9.581e-02  2.895e-02  -3.309  0.00102 **
#> I(avginc3)   6.855e-04  3.471e-04   1.975  0.04892 *
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Residual standard error: 12.71 on 416 degrees of freedom
#> Multiple R-squared(full model): 0.5584   Adjusted R-squared: 0.5552
#> Multiple R-squared(proj model): 0.5584   Adjusted R-squared: 0.5552
```

Table 1: Test Score and Income: Nonlinear OLS Regression

	Dependent Variable: Test Score		
	(1)	(2)	(3)
avginc	1.879***	3.851***	5.019***
	(0.113)	(0.267)	(0.704)
I(avginc^2)		−0.042***	−0.096***
		(0.005)	(0.029)
I(avginc^3)			0.001**
			(0.0003)
Constant	625.384***	607.302***	600.079***
	(1.863)	(2.891)	(5.078)
Observations	420	420	420
Adjusted R^2	0.506	0.554	0.555
F Statistic	430.830***	261.278***	175.352***

Note: *p<0.1; **p<0.05; ***p<0.01
Robust standard errors are in parentheses.

Quadratic vs linear

- **Question**: does adding a quadratic term improve the fit in a meaningful way?
- Test linearity against a quadratic alternative:

$$H_0 : \beta_2 = 0 \quad \textbf{vs.} \quad H_1 : \beta_2 \neq 0.$$

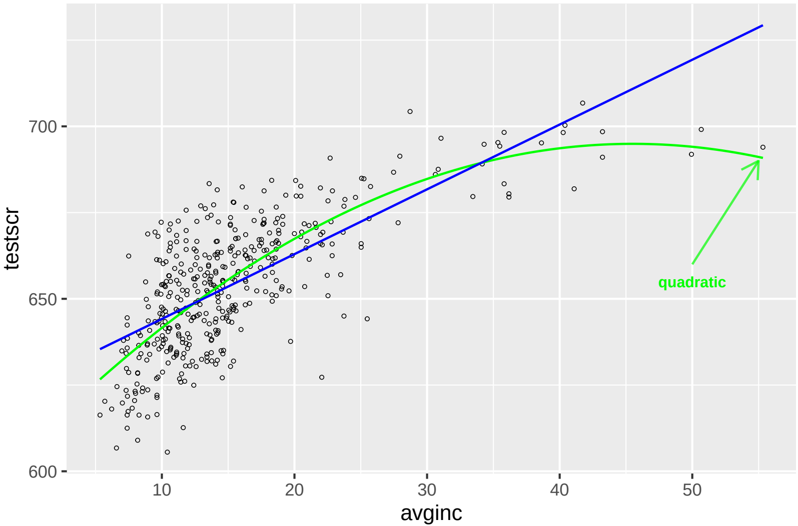
- The t-statistic is

$$t = \frac{(\hat{\beta}_2 - 0)}{SE(\hat{\beta}_2)} = \frac{-0.0423}{0.0048} = -8.81.$$

- Because $|8.81| > 2.58$, we reject H_0 at the 1% level—**income enters nonlinearly** (the quadratic term matters).
- A **joint** F-test of the polynomial terms likewise rejects with

$$F_{2,417} = 261.3, \quad p\textbf{-value} \approx 0.$$

Figure: Linear and Quadratic Regression



Interpreting the estimated quadratic regression function

- What is the **marginal effect** of X on Y in a quadratic specification?
- Write

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i.$$

- The marginal effect of X on Y is

$$\frac{\partial Y_i}{\partial X_i} = \beta_1 + 2\beta_2 X_i.$$

- So the **marginal effect depends on where** you evaluate X_i.

Interpreting the quadratic regression function

- The estimated regression function with a quadratic term of income is

$$\widehat{TestScore_i} = \underset{(2.90)}{607.3} + \underset{(0.27)}{3.85} \times income_i - \underset{(0.0048)}{0.0423} \times income_i^2.$$

- Consider a **$1,000** increase in average district income (in the units of the dataset).
- **Low-income districts** (from $10,000 to $11,000 per capita):

$$\begin{aligned} \Delta TestScore &= 607.3 + 3.85 \times 11 - 0.0423 \times (11)^2 \\ &\quad - [607.3 + 3.85 \times 10 - 0.0423 \times (10)^2] \\ &= 2.96 \end{aligned}$$

- **Higher-income districts** (from $40,000 to $41,000 per capita):

$$\begin{aligned} \Delta TestScore &= 607.3 + 3.85 \times 41 - 0.0423 \times (41)^2 \\ &\quad - [607.3 + 3.85 \times 40 - 0.0423 \times (40)^2] \\ &= 0.42 \end{aligned}$$

Quadratic vs cubic

- **Question**: should we add a **cubic** term on top of the quadratic?
- Test *quadratic vs. cubic*:

$$H_0 : \beta_3 = 0 \quad \textbf{vs.} \quad H_1 : \beta_3 \neq 0.$$

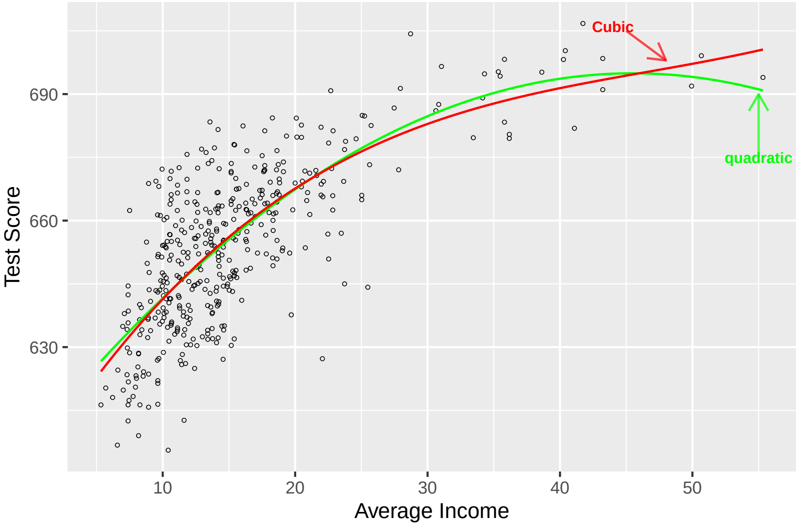
- The t-statistic is

$$t = \frac{(\hat{\beta}_3 - 0)}{SE(\hat{\beta}_3)} = \frac{-0.001}{0.0003} = -3.33.$$

- Because $|3.33| > 2.58$, we reject H_0 at the 1% level—the **cubic term is statistically detectable** in this sample.
- A joint F-test of $\beta_1 = \beta_2 = \beta_3 = 0$ also rejects decisively:

$$F_{3,416} = 175.35, \quad p\text{-value} \approx 0.$$

Figure: Cubic and Quadratic Regression



Interpreting the estimated cubic regression function

- **With a cubic term,**

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i.$$

- **The marginal effect of X on Y is**

$$\frac{\partial Y_i}{\partial X_i} = \beta_1 + 2\beta_2 X_i + 3\beta_3 X_i^2.$$

Interpreting the estimated regression function

- **The estimated cubic model is**

$$\widehat{TestScore_i} = \underset{(5.83)}{600.1} + \underset{(0.86)}{5.02} \times income - \underset{(0.03)}{0.096} \times income^2 + \underset{(0.00047)}{0.00069} \times income^3.$$

- **Lower income** (from \$10,000 to \$11,000 per capita):

$$\begin{aligned}\Delta TestScore = & \ 600.079 + 5.019 \times 11 - 0.096 \times (11)^2 + 0.001 \times (11)^3 \\ & - [600.079 + 5.019 \times 10 - 0.096 \times (10)^2 + 0.001 \times (10)^3]\end{aligned}$$

- **Higher income** (from \$40,000 to \$41,000 per capita):

$$\begin{aligned}\Delta TestScore = & \ 600.079 + 5.019 \times 41 - 0.096 \times (41)^2 + 0.001 \times (41)^3 \\ & - [600.079 + 5.019 \times 40 - 0.096 \times (40)^2 + 0.001 \times (40)^3]\end{aligned}$$

Polynomial regression in X

- Approximate the CEF with

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_r X_i^r + u_i.$$

- This **is** multiple linear regression—the regressors are **powers of** X.
- Estimation and inference proceed as usual (OLS with robust SEs).
- **Raw coefficients** are awkward to interpret; focus on the **fitted curve** and **marginal effects**.

Which polynomial degree?

- **Higher-order** terms add **flexibility** but cost **precision** (more parameters).
  - A degree-r polynomial can have up to $r-1$ **turning points**.
  - Each added power is another regressor—coefficient estimates become less precise.
- Modern machine-learning or nonparametric econometric methods face the same trade-off(**Bias-Variance Trade-off**)

Which polynomial degree?

- A practical rule: start high enough, then ask whether the **leading** terms are needed.
- **Sequential testing** (informal but common):
  1. Choose a maximum degree r and estimate the rth-order polynomial.
  2. Use a t-test on β_r (the coefficient on X^r).
  3. If you **cannot** reject $\beta_r = 0$, drop X^r and test β_{r-1}, and so on.
  4. Stop when the highest retained power is statistically distinguishable from zero.
- **Caution**: sequential tests are not guaranteed to control overall size; **AIC/BIC** and out-of-sample checks are complements, not substitutes.

Which polynomial degree?

- You still need a **starting** r, thus the initial degree of the polynomial is still missing.
- In many applications involving economic data, the nonlinear functions are smooth, no sharp jumps or spikes.
- If so, then it is appropriate to choose a small **maximum degree** for the polynomial, such as **2, 3, or 4**.

Joint testing of polynomial terms

- If the population regression function is **linear in** X, higher powers have population coefficients equal to **zero**.
- Joint test:

$$H_0 : \beta_2 = \beta_3 = \cdots = \beta_r = 0 \quad \textbf{vs.} \quad H_1 : \text{ at least one } \beta_j \neq 0.$$

- Count the restrictions: if you include powers $X^2, \ldots, X^r$, that is $q = r - 1$ restrictions—an **F** test is the textbook tool (robust F when appropriate).
- The same nesting logic compares *quadratic* vs. *cubic* specifications.
- **Model selection** also uses penalized fit:
  - Akaike information criterion (**AIC**)
  - Bayesian information criterion (**BIC**)

Wrap-up: polynomials

- Polynomials in X are still **linear OLS**—with cleverly chosen regressors multiple linear regression.
- Omitting curvature when the CEF bends is a **functional-form** mistake—conceptually akin to **OVB** from leaving out a relevant part of $E[Y \mid X]$.
- Mechanical pieces stay the same: OLS, t- and F-tests (with robust SEs).
- What changes is **interpretation**: read results through **marginal effects** and **predicted changes**, not only through tabulated slopes.

Logarithms

Logarithmic functions of Y and/or X

- Logs turn **multiplicative** stories into **additive** regressions and deliver **percentage** interpretations.
- **The natural log $\ln(x)$ is the inverse of the exponential: $\ln(e^x) = x$, with** $e \approx 2.71828$.

Properties of logarithms

- For positive x and y:

Properties of logarithms

- For positive x and y:

$$ln(1/x) = -ln(x)$$

Properties of logarithms

- **For positive x and y:**

$$ln(1/x) = -ln(x)$$

$$ln(xy) = ln(y) + ln(x)$$

Properties of logarithms

- For positive x and y:

$$ln(1/x) = -ln(x)$$

$$ln(xy) = ln(y) + ln(x)$$

$$ln(x/y) = ln(x) - ln(y)$$

Properties of logarithms

- **For positive x and y:**

$$
\begin{aligned}
ln(1/x) &= -ln(x) \\
ln(xy) &= ln(y) + ln(x) \\
ln(x/y) &= ln(x) - ln(y) \\
ln(x^y) &= yln(x)
\end{aligned}
$$

Logarithms and percentages

- **For small x, $\ln(1+x) \approx x$ (first-order Taylor approximation).**
- **Hence**

$$\ln(x+\Delta x) - \ln(x) = \ln\left(\frac{x+\Delta x}{x}\right)$$

Logarithms and percentages

- **For small x, $\ln(1+x) \approx x$ (first-order Taylor approximation).**
- **Hence**

$$\ln(x+\Delta x) - \ln(x) = \ln\left(\frac{x+\Delta x}{x}\right)$$
$$\approx \frac{\Delta x}{x} \quad \text{(when } \Delta x/x \text{ is small).}$$

- **Example:**

$$\ln(1+0.01) = \ln(101) - \ln(100) = 0.00995 \approx 0.01.$$

- That is why log specifications map naturally onto **percent changes** and **elasticities.**

Three common log specifications

Case	Population regression function
I.linear-log	

Three common log specifications

Case	Population regression function
I.linear-log	$Y_i = \beta_0 + \beta_1 ln(X_i) + u_i$
II.log-linear	

Three common log specifications

Case	Population regression function
I.linear-log	$Y_i = \beta_0 + \beta_1 ln(X_i) + u_i$
II.log-linear	$ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$
III.log-log	

Three common log specifications

Case	Population regression function
I.linear-log	$Y_i = \beta_0 + \beta_1 ln(X_i) + u_i$
II.log-linear	$ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$
III.log-log	$ln(Y_i) = \beta_0 + \beta_1 ln(X_i) + u_i$

- The meaning of the slope **depends on the case**—always ask: "**how does Y move when X moves?**" (Stock & Watson, Key Concept 8.1).

I. Linear–log (level–log) model

- **Specification:**

$$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i$$

- **For a small change ΔX:**

$$\Delta Y = [\beta_0 + \beta_1 ln(X + \Delta X)] - [\beta_0 + \beta_1 ln(X)]$$

I. Linear–log (level–log) model

- **Specification:**

$$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i$$

- **For a small change ΔX:**

$$\begin{aligned} \Delta Y &= [\beta_0 + \beta_1 ln(X + \Delta X)] - [\beta_0 + \beta_1 ln(X)] \\ &= \beta_1 [ln(X + \Delta X) - ln(X)] \end{aligned}$$

I. Linear–log (level–log) model

- **Specification:**

$$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i$$

- **For a small change ΔX:**

$$\begin{aligned}\Delta Y &= [\beta_0 + \beta_1 ln(X + \Delta X)] - [\beta_0 + \beta_1 ln(X)] \\ &= \beta_1 [ln(X + \Delta X) - ln(X)] \\ &\cong \beta_1 \frac{\Delta X}{X}\end{aligned}$$

- Here $100 \times \Delta X / X$ is the **percent change in** X, and $\beta_1 \approx \Delta Y / (\Delta X / X)$.
- **Interpretation**: a **1% increase** in X shifts Y by about $\beta_1 / 100$ **units** on the scale of Y.

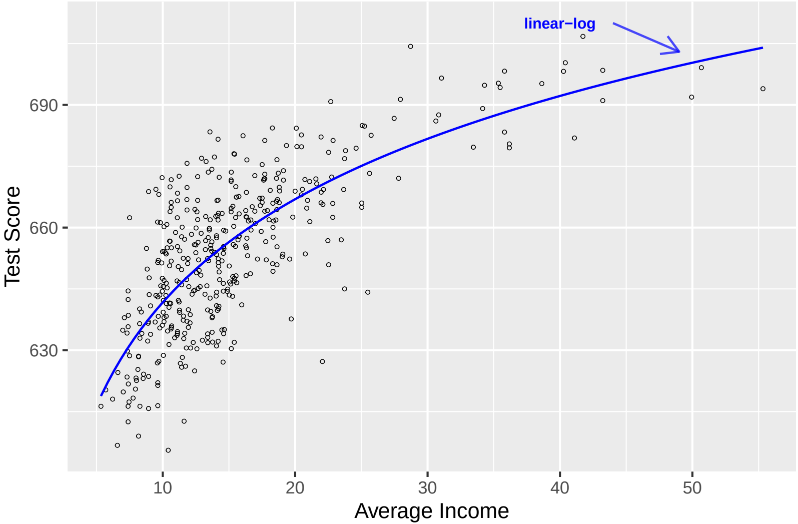
Example: test scores vs. $\ln(\text{income})$

- OLS of **test score on** $\ln(\textbf{income})$ **gives**

$$\widehat{TestScore} = 557.8 + 36.42 \times ln(Income)$$
$$(3.8) \quad (1.4)$$

- Interpretation: **a 1% increase in district income** is associated with roughly **0.36** additional test-score points ($36.42/100$).
- **In this sample**, mean district income is about **$15,300** (`avginc` $\approx$ **15.3 in $1,000s**), so a 1% bump is about **$153** on the dollar scale; the fitted equation says test scores rise by about **0.36 points** for that proportional change.

Test scores: linear-log function



II. Log–linear (log–level) model

- **Specification:**

$$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$$

- **Increase X by ΔX:**

$$ln(\Delta Y + Y) - ln(Y) = [\beta_0 + \beta_1(X + \Delta X)] - [\beta_0 + \beta_1 X]$$

II. Log–linear (log–level) model

- **Specification:**

$$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$$

- **Increase X by ΔX:**

$$ln(\Delta Y + Y) - ln(Y) = [\beta_0 + \beta_1(X + \Delta X)] - [\beta_0 + \beta_1 X]$$
$$\Rightarrow ln(1 + \frac{\Delta Y}{Y}) = \beta_1 \Delta X$$

II. Log–linear (log–level) model

- **Specification:**

$$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$$

- **Increase X by ΔX:**

$$ln(\Delta Y + Y) - ln(Y) = [\beta_0 + \beta_1(X + \Delta X)] - [\beta_0 + \beta_1 X]$$
$$\Rightarrow ln(1 + \frac{\Delta Y}{Y}) = \beta_1 \Delta X$$
$$\Rightarrow \frac{\Delta Y}{Y} \cong \beta_1 \Delta X$$

- **So $100(\Delta Y/Y)$ is the percent change in Y, and $\beta_1 \approx (\Delta Y/Y)/\Delta X$ for small changes.**
- **One-unit increase in $X \Rightarrow$ roughly $100\beta_1$ percent change in Y (the scaling of X must be interpreted accordingly).**

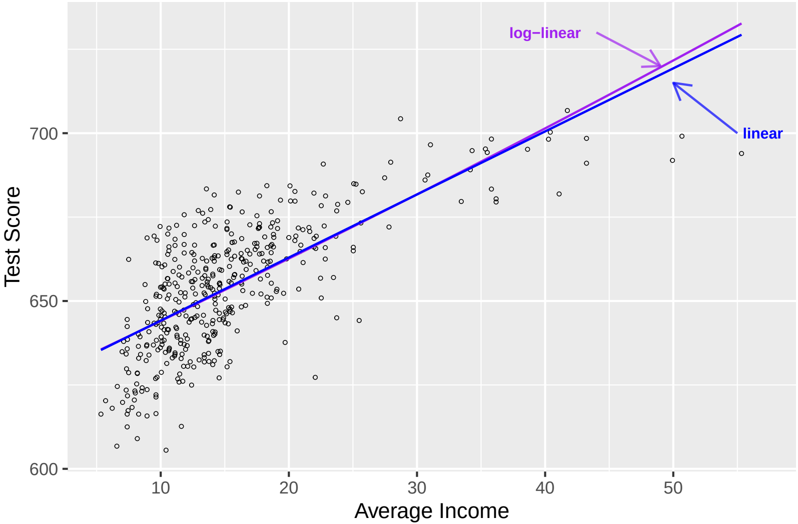
Example: $\ln($test score$)$ vs. income

- OLS of $\ln($**test score**$)$ **on income** yields

$$ln(\widehat{TestScore}) = 6.439 + 0.003 \times (Income)$$
$$\qquad\qquad\qquad (0.0028) \quad (0.0002)$$

- Interpretation: if income is measured in **$1,000s**, one more unit (**+$1,000**) is associated with roughly a **0.3%** increase in test scores (because $100 \times 0.003 = 0.3$).
- **At the mean** test score of 650, 0.3% of 650 is about 1.95 **points**.

Test scores: log-linear function



A canonical log–linear specification: the Mincer equation

- **Mincer earnings equation**:

$$\ln(\textbf{Earnings}) = \beta_0 + \beta_1 S_i + \beta_2 \textbf{Exper}_i + \beta_3 \textbf{Exper}_i^2 + u_i.$$

- Standard in empirical **human-capital** work.
  - β_1 is the **return to schooling** in the log–linear sense.
- Illustrative coefficients:

$$\ln(\textbf{Earnings}) = 2.881 + 0.06\, S_i + 0.01\, \textbf{Exper}_i - 0.0003\, \textbf{Exper}_i^2 + u_i.$$

- Question: how should we read $\hat{\beta}_1 = 0.06$?
- Answer: one more year of schooling is associated with roughly **6% higher** earnings (because $100 \times 0.06 = 6$ is the usual percent interpretation of the log–linear coefficient—**not 0.06%**).

III. Log–log model

- **Specification:**

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$$

- **For small proportional moves:**

$$\ln(\Delta Y + Y) - \ln(Y) = [\beta_0 + \beta_1 \ln(X + \Delta X)] - [\beta_0 + \beta_1 \ln(X)]$$

III. Log–log model

- **Specification:**

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$$

- **For small proportional moves:**

$$\ln(\Delta Y + Y) - \ln(Y) = [\beta_0 + \beta_1 \ln(X + \Delta X)] - [\beta_0 + \beta_1 \ln(X)]$$
$$\Rightarrow \ln\left(1 + \frac{\Delta Y}{Y}\right) = \beta_1 \ln\left(1 + \frac{\Delta X}{X}\right)$$

III. Log–log model

- **Specification:**

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$$

- **For small proportional moves:**

$$\begin{aligned}
\ln(\Delta Y + Y) - \ln(Y) &= [\beta_0 + \beta_1 \ln(X + \Delta X)] - [\beta_0 + \beta_1 \ln(X)] \\
\Rightarrow \ln\Big(1 + \frac{\Delta Y}{Y}\Big) &= \beta_1 \ln\Big(1 + \frac{\Delta X}{X}\Big) \\
\Rightarrow \frac{\Delta Y}{Y} &\approx \beta_1 \frac{\Delta X}{X}
\end{aligned}$$

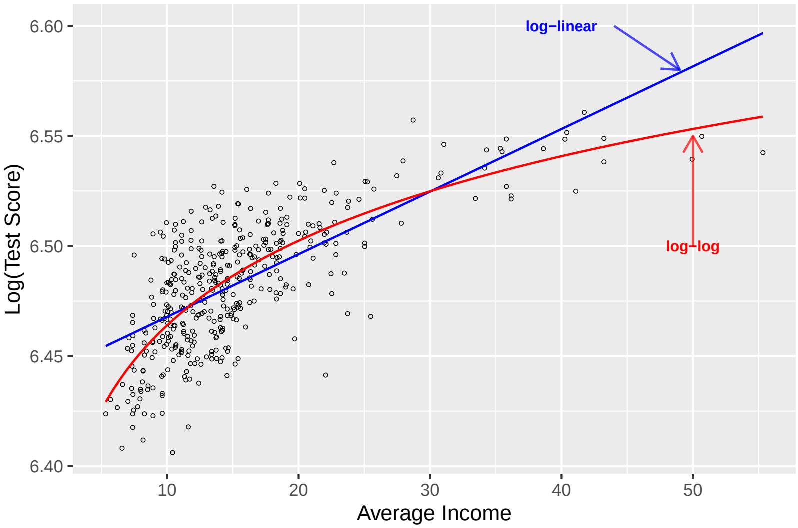
- **So β_1 is an elasticity: a 1% increase in X is associated with about a β_1% change in Y (for small changes, and holding the rest of the model fixed).**

Test scores and income: log-log specifications

$$\widehat{ln(TestScore)} = 6.336 + 0.055 \times ln(Income)$$
$$(0.006) \quad (0.002)$$

- **Interpretation: a 1% increase in income** is associated with about a **0.055% increase in test scores**. The slope on $\ln(\text{income})$ is therefore the **elasticity** of test scores with respect to income.
- **Example:** if mean district income is about **\$15,300**, a **1%** rise is roughly **\$153**. A **0.055%** increase in test scores, evaluated at a mean score near **654**, is about $654 \times 0.00055 \approx 0.36$ points.

Test scores: The log-linear and log-log functions



Test scores: The linear-log and cubic functions

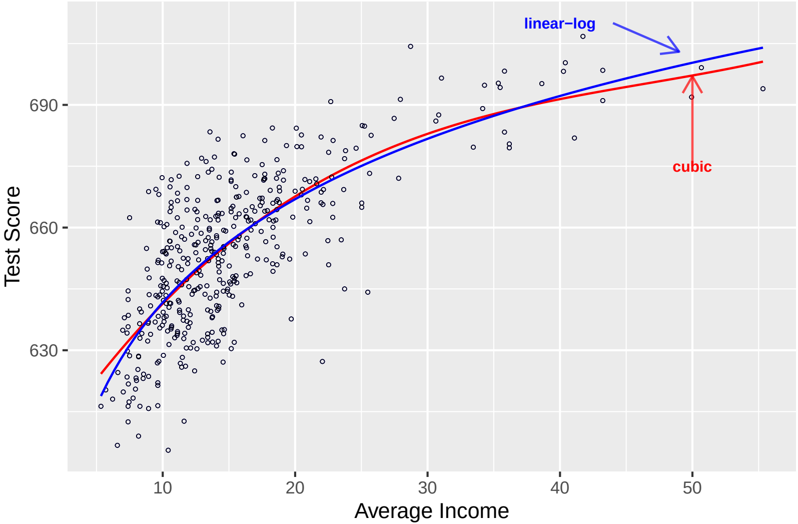


Table 2: Test Score and Income

	Dependent Variable: Test Score		
	testscr	log.testscr	testscr
	(1)	(2)	(3)
loginc	36.420***	0.055***	
		(0.002)	
avginc			5.019***
			(0.704)
I(avginc^2)			−0.096***
			(0.029)
I(avginc^3)			0.001**
			(0.0003)
Constant	557.832***	6.336***	600.079***
	(0.003)	(0.006)	(5.078)
Observations	420	420	420
Adjusted R^2	0.561	0.557	0.555
Residual Std. Error	12.618	0.019	12.707
F Statistic	537.444***	527.238***	175.352***

Choosing among specifications

- When several specifications fit **similarly well** statistically, **substantive reasoning**—economic interpretation and plausibility—should guide the choice.
  1. Prefer the functional form with the clearest **economic** interpretation.
  2. Use **t-** and **F-tests** for nested comparisons; more elaborate selection procedures are beyond the scope of this course.
  3. Examine **fitted values** and standard **goodness-of-fit** measures ($\overline{R}^2$, SER)—**no single statistic** is decisive on its own.

Wrap-up: logs and polynomials

- **Polynomials** and **logs** offer simple, transparent ways to **relax linearity** in the CEF and to summarize relationships in **elasticities** or percentage terms.
- Identifying the precise reason a linear specification is inadequate is often impossible; **robustness** across reasonable specifications therefore carries greater weight.
- For many cross-section applications in economics, **quadratic or cubic terms** together with **logs** suffice to capture **most** of the flexibility typically required.

Interactions Between Independent Variables

Introduction

- Core question: *how does the effect of X on Y depend on the level of another regressor Z?*
- Answer: include their **product** $X \times Z$—an interaction term.
- Three workhorses:
    1. Interactions between two binary variables.
    2. Interactions between a binary and a continuous variable.
    3. Interactions between two continuous variables.

Interactions Between Two Binary Variables

- Suppose we study **log earnings** for workers in the labor market.
- Baseline specification:

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + u_i, \quad Y_i = \ln(\textbf{Earnings}_i).$$

- Two dummies:
  - $D_{1i} = 1$ if worker i **graduated from college** (0 otherwise);
  - $D_{2i} = 1$ if worker i is **female** (0 otherwise).

- **Reading** β_1 and β_2 (with **female** coded as 1):
  - β_1: **college premium for men** ($D_2 = 0$), *holding gender fixed*;
  - β_2: **female minus male gap** among workers **without** a college degree (*holding the degree indicator fixed*).

Interactions Between Two Binary Variables

- This **additive** model forces the college premium to be **the same** for men and women.
- That restriction is often **too strong**.
  - Returns to college can **differ by gender**.
- Allow **different college premia** by adding a product term:

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 (D_{1i} \times D_{2i}) + u_i.$$

- The product $D_{1i} \times D_{2i}$ is the **interaction** (or **interacted regressor**).

Interactions Between Two Binary Variables:

- The population regression model now is

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 (D_{1i} \times D_{2i}) + u_i$$

- Let $d_2 \in \{0, 1\}$ denote the value of D_{2i} (men vs. women).
- Then the expected earnings of men and women **without a college degree** are

$$E(Y_i | D_{1i} = 0, D_{2i} = d_2) = \beta_0 + \beta_1 \times 0 + \beta_2 d_2 + \beta_3 (0 \times d_2) = \beta_0 + \beta_2 d_2$$

- Then the expected earnings of men and women **with a college degree** are

$$\begin{aligned} E(Y_i | D_{1i} = 1, D_{2i} = d_2) &= \beta_0 + \beta_1 \times 1 + \beta_2 d_2 + \beta_3 (1 \times d_2) \\ &= \beta_0 + \beta_1 + \beta_2 d_2 + \beta_3 d_2 \end{aligned}$$

Interactions Between Two Binary Variables:

- **The college premium** is

$$E(Y_i \mid D_{1i} = 1, D_{2i} = d_2) - E(Y_i \mid D_{1i} = 0, D_{2i} = d_2) = \beta_1 + \beta_3 d_2.$$

- Going from $D_{1i} = 0$ to $D_{1i} = 1$ therefore **depends on gender**:
  - Men ($d_2 = 0$): premium equals β_1.
  - Women ($d_2 = 1$): premium equals $\beta_1 + \beta_3$.
- So β_3 is the extra college premium for women relative to men.

Application: student–teacher ratio and English learners

- $HiSTR_i$: high **student–teacher ratio**,
  - $HiSTR_i = 1$ if STR$_i \geq 20$;
  - $HiSTR_i = 0$ otherwise.
- $HiEL_i$: high share of **English learners**,
  - $HiEL_i = 1$ if the percent English learners exceeds **10%**;
  - $HiEL_i = 0$ otherwise.

Application: STR and English learners

- An OLS fit yields

$$\widehat{TestScore} = \underset{(11.9)}{682.2} - \underset{(0.59)}{0.97}STR + \underset{(19.5)}{5.6}HiEL - \underset{(0.97)}{1.28}(STR \times HiEL)$$

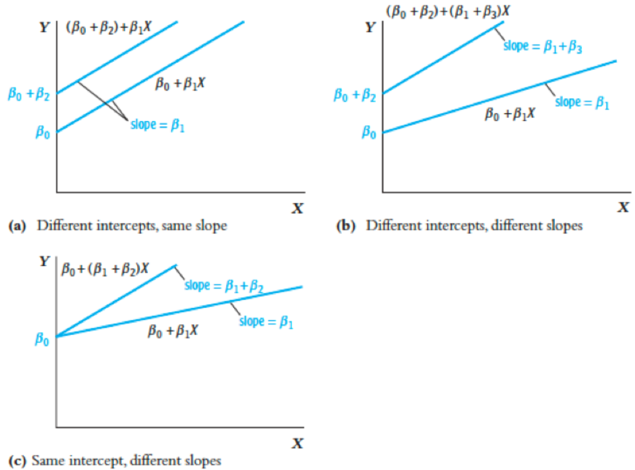
- **Question**: what does the interaction coefficient (here, on STR $\times$ $HiEL$) tell us?
- **Answer:** the **effect of STR on test scores** is allowed to differ between districts with **high** versus **low** shares of English learners. The interaction term captures that difference in slopes.
- In this particular fit, the interaction estimate is **small and imprecise**; interpret it **cautiously** rather than as definitive evidence.

Interactions: a continuous and a binary regressor

- **Binary** D_i (e.g., college degree).
- **Continuous** X_i (e.g., experience).
- Three common templates:
  1. Parallel lines (different intercepts, common slope):
  $$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + u_i$$
  2. Lines through the **same intercept** but **different slopes**:
  $$Y_i = \beta_0 + \beta_1 X_i + \beta_2 (D_i \times X_i) + u_i$$
  3. **General** (different intercepts **and** slopes):
  $$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 (D_i \times X_i) + u_i$$

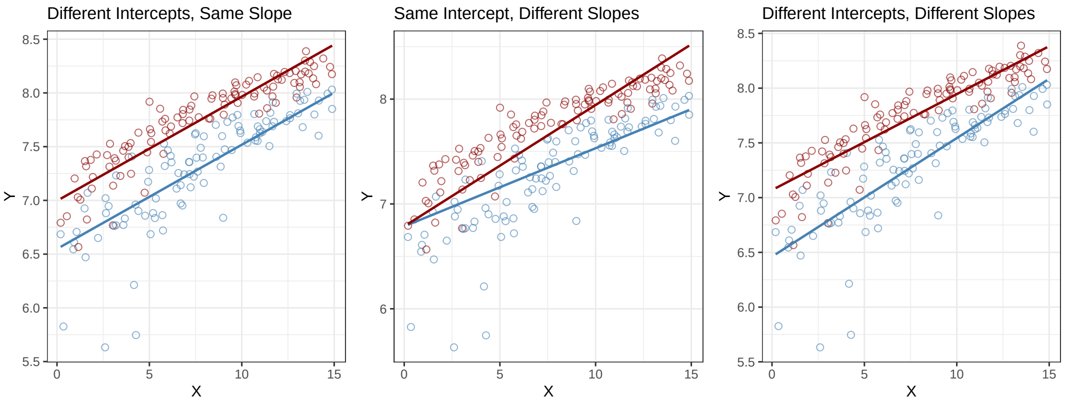
A Continuous and a Binary Variable: Three Cases

FIGURE 8.8 Regression Functions Using Binary and Continuous Variables



Interactions of binary variables and continuous variables can produce three different population regression functions: (a) $\beta_0 + \beta_1 X + \beta_2 D$ allows for different intercepts but has the same slope, (b) $\beta_0 + \beta_1 X + \beta_2 D + \beta_3 (X \times D)$ allows

A Continuous and a Binary Variable: Three Cases



A Continuous and a Binary Variable: Specifications

- All three specifications are just different versions of the multiple regression model.
- Each model **embeds different restrictions** on how D shifts the X–Y relationship.
- **Model 3** is the **flexible default**—it allows **both** separate intercepts **and** separate slopes.

Application: STR and English learners

- $HiEL_i$ is **binary**; **STR** is **continuous**.
- The estimated interaction regression

$$\widehat{TestScore} = \underset{(11.9)}{682.2} - \underset{(0.59)}{0.97}STR + \underset{(19.5)}{5.6}HiEL - \underset{(0.97)}{1.28}(STR \times HiEL)$$

- **Low** EL share ($HiEL_i = 0$):

$$\widehat{TestScore} = 682.2 - 0.97\,\mathbf{STR}_i.$$

- **High** EL share ($HiEL_i = 1$):

$$\widehat{TestScore} = 682.2 + 5.6 - 0.97\,\mathbf{STR}_i - 1.28\,\mathbf{STR}_i = 687.8 - 2.25\,\mathbf{STR}_i.$$

Application: STR and English learners

- The **1.28** difference between the two STR slopes is exactly the coefficient on the **interaction term** in this specification.
- **Economic reading:** when $HiEL = 1$, the marginal effect of STR on scores is **more negative**—class size is associated with **larger test-score losses** where a greater share of students are English learners—but the point estimate is **imprecise** in the baseline table.

Testing Model Specifications

- We can evaluate the three specifications using **F-tests** and **t-tests**.

1. Testing whether test scores are identical between groups (same intercept and slope)
    - $H_0 : \beta_2 = \beta_3 = 0$
    - F-test result: F-statistic = 89.9, significant at the 1% level, allowing us to reject this hypothesis.
2. Testing whether effects are identical between groups (same slope but potentially different intercepts)
    - $H_0 : \beta_3 = 0$
    - t-statistic = -1.32, not significant at the 10% level.

Further Testing

3. Testing whether groups have the same intercept but different slopes
   - $H_0 : \beta_2 = 0$
   - t-statistic = 0.29, not significant at the 10% level.

- The high correlation between regressors $HiEL$ and $STR * HiEL$ leads to inflated standard errors for individual coefficients.
- **Conclusion: Multicollinearity** can make each **separate** t-ratio uninformative even when an **F**-test rejects a **joint** restriction. Read the table accordingly.

Interactions between two continuous regressors

- Let both X_{1i} and X_{2i} be continuous—e.g., experience and schooling in a wage equation.
- Allow **experience returns to depend on schooling** via

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 (X_{1i} \times X_{2i}) + u_i.$$

Marginal effects with a continuous interaction

- **Effect of X_1 holding X_2 fixed:**

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

- **Effect of X_2 holding X_1 fixed:**

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

- **If both X_1 and X_2 move, the total effect to first order includes a cross term:**

$$\Delta Y \approx (\beta_1 + \beta_3 X_2)\Delta X_1 + (\beta_2 + \beta_3 X_1)\Delta X_2 + \beta_3 \Delta X_1 \Delta X_2.$$

Application: STR and English learners

- Now both **STR** and **PctEL** are continuous.
- Fitted specification (levels, as in Stock & Watson's CA schools discussion):

$$\widehat{TestScore} = \underset{(11.8)}{686.3} - \underset{(0.059)}{1.12}\,\textbf{STR} - \underset{(0.037)}{0.67}\,\textbf{PctEL} + \underset{(0.019)}{0.0012}\,(\textbf{STR} \times \textbf{PctEL})$$

- **Meaning of** β_3: the **STR slope** shifts with the **percent English learners**.
- Each **percentage-point increase** in PctEL **tilts** the STR slope by **+0.0012** test-score points per student–teacher unit.

Application: STR and English learners

- **At the median** ($PctEL = 8.85$):

$$\frac{\partial Y}{\partial \text{STR}} = \beta_1 + \beta_3 \cdot \textbf{PctEL} = -1.12 + 0.0012 \times 8.85 \approx -1.11.$$

- **At the 75th percentile** ($PctEL = 23.0$):

$$\frac{\partial Y}{\partial \text{STR}} = -1.12 + 0.0012 \times 23.0 \approx -1.09.$$

- The effect of class size on test scores depends on the share of English learners.
- The gap between those two slopes is **economically modest** in this table—and **not statistically distinguishable** from zero at conventional levels.
  - The reported t-**statistic on the interaction** is small, so we would **fail to reject** that the interaction is zero using a standard **t-test.

Application: STR and Test Scores in a Summary

- Flexible terms in STR are informative, but you still need **background controls** (family resources, immigration intensity, etc.) to mitigate **omitted-variable bias**.
- Two stylized **socioeconomic** controls in the table:
  1. the fraction of students on **subsidized lunch**;
  2. **log average district income**.

Application: STR and test scores—summary

- Three guiding questions:
  1. After conditioning on economics, does the STR effect **interact** with the EL share?
  2. Does that interaction itself **vary with STR** (higher-order terms)?
  3. Most important for policy: holding those controls and nonlinearities fixed, what is the **predicted gain** from cutting STR by **one student per teacher**?

	score						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
str	-1.00^{***}	-0.73^{**}	-0.97	-0.53	64.34^{**}	83.70^{**}	65.29^{**}
	(0.27)	(0.26)	(0.59)	(0.34)	(24.86)	(28.50)	(25.26)
I(str^2)					-3.42^{**}	-4.38^{**}	-3.47^{**}
					(1.25)	(1.44)	(1.27)
I(str^3)					0.06^{**}	0.07^{**}	0.06^{**}
					(0.02)	(0.02)	(0.02)
str:HiEL			-1.28	-0.58		-123.28^{*}	
			(0.97)	(0.50)		(50.21)	
I(str^2):HiEL						6.12^{*}	
						(2.54)	
I(str^3):HiEL						-0.10^{*}	
						(0.04)	
english	-0.12^{***}	-0.18^{***}					-0.17^{***}
	(0.03)	(0.03)					(0.03)
HiEL			5.64	5.50	-5.47^{***}	816.08^{*}	
			(19.51)	(9.80)	(1.03)	(327.67)	
lunch	-0.55^{***}	-0.40^{***}		-0.41^{***}	-0.42^{***}	-0.42^{***}	-0.40^{***}
	(0.02)	(0.03)		(0.03)	(0.03)	(0.03)	(0.03)
log(income)		11.57^{***}		12.12^{***}	11.75^{***}	11.80^{***}	11.51^{***}
		(1.82)		(1.80)	(1.77)	(1.78)	(1.81)
Constant	700.15^{***}	658.55^{***}	682.25^{***}	653.67^{***}	252.05	122.35	244.81
	(5.57)	(8.64)	(11.87)	(9.87)	(163.63)	(185.52)	(165.72)
N	420	420	420	420	420	420	420
Adjusted R^2	0.77	0.79	0.31	0.79	0.80	0.80	0.80

$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$
Robust standard errors in parentheses.

	score						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
str	-1.00^{***}	-0.73^{**}	-0.97	-0.53	64.34^{**}	83.70^{**}	65.29^{**}
	(0.27)	(0.26)	(0.59)	(0.34)	(24.86)	(28.50)	(25.26)
I(str^2)					-3.42^{**}	-4.38^{**}	-3.47^{**}
					(1.25)	(1.44)	(1.27)
I(str^3)					0.06^{**}	0.07^{**}	0.06^{**}
					(0.02)	(0.02)	(0.02)
str:HiEL			-1.28	-0.58		-123.28^{*}	
			(0.97)	(0.50)		(50.21)	
I(str^2):HiEL						6.12^{*}	
						(2.54)	
I(str^3):HiEL						-0.10^{*}	
						(0.04)	
english	-0.12^{***}	-0.18^{***}					-0.17^{***}
	(0.03)	(0.03)					(0.03)
HiEL			5.64	5.50	-5.47^{***}	816.08^{*}	
			(19.51)	(9.80)	(1.03)	(327.67)	
lunch	-0.55^{***}	-0.40^{***}		-0.41^{***}	-0.42^{***}	-0.42^{***}	-0.40^{***}
	(0.02)	(0.03)		(0.03)	(0.03)	(0.03)	(0.03)
log(income)		11.57^{***}		12.12^{***}	11.75^{***}	11.80^{***}	11.51^{***}
		(1.82)		(1.80)	(1.77)	(1.78)	(1.81)
Constant	700.15^{***}	658.55^{***}	682.25^{***}	653.67^{***}	252.05	122.35	244.81
	(5.57)	(8.64)	(11.87)	(9.87)	(163.63)	(185.52)	(165.72)
N	420	420	420	420	420	420	420
Adjusted R^2	0.77	0.79	0.31	0.79	0.80	0.80	0.80

$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$
Robust standard errors in parentheses.

	score						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
str	-1.00^{***}	-0.73^{**}	-0.97	-0.53	64.34^{**}	83.70^{**}	65.29^{**}
	(0.27)	(0.26)	(0.59)	(0.34)	(24.86)	(28.50)	(25.26)
I(str^2)					-3.42^{**}	-4.38^{**}	-3.47^{**}
					(1.25)	(1.44)	(1.27)
I(str^3)					0.06^{**}	0.07^{**}	0.06^{**}
					(0.02)	(0.02)	(0.02)
str:HiEL			-1.28	-0.58		-123.28^{*}	
			(0.97)	(0.50)		(50.21)	
I(str^2):HiEL						6.12^{*}	
						(2.54)	
I(str^3):HiEL						-0.10^{*}	
						(0.04)	
english	-0.12^{***}	-0.18^{***}					-0.17^{***}
	(0.03)	(0.03)					(0.03)
HiEL			5.64	5.50	-5.47^{***}	816.08^{*}	
			(19.51)	(9.80)	(1.03)	(327.67)	
lunch	-0.55^{***}	-0.40^{***}		-0.41^{***}	-0.42^{***}	-0.42^{***}	-0.40^{***}
	(0.02)	(0.03)		(0.03)	(0.03)	(0.03)	(0.03)
log(income)		11.57^{***}		12.12^{***}	11.75^{***}	11.80^{***}	11.51^{***}
		(1.82)		(1.80)	(1.77)	(1.78)	(1.81)
Constant	700.15^{***}	658.55^{***}	682.25^{***}	653.67^{***}	252.05	122.35	244.81
	(5.57)	(8.64)	(11.87)	(9.87)	(163.63)	(185.52)	(165.72)
N	420	420	420	420	420	420	420
Adjusted R^2	0.77	0.79	0.31	0.79	0.80	0.80	0.80

$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$
Robust standard errors in parentheses.

	score						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
str	-1.00^{***}	-0.73^{**}	-0.97	-0.53	64.34^{**}	83.70^{**}	65.29^{**}
	(0.27)	(0.26)	(0.59)	(0.34)	(24.86)	(28.50)	(25.26)
I(str^2)					-3.42^{**}	-4.38^{**}	-3.47^{**}
					(1.25)	(1.44)	(1.27)
I(str^3)					0.06^{**}	0.07^{**}	0.06^{**}
					(0.02)	(0.02)	(0.02)
str:HiEL			-1.28	-0.58		-123.28^{*}	
			(0.97)	(0.50)		(50.21)	
I(str^2):HiEL						6.12^{*}	
						(2.54)	
I(str^3):HiEL						-0.10^{*}	
						(0.04)	
english	-0.12^{***}	-0.18^{***}					-0.17^{***}
	(0.03)	(0.03)					(0.03)
HiEL			5.64	5.50	-5.47^{***}	816.08^{*}	
			(19.51)	(9.80)	(1.03)	(327.67)	
lunch	-0.55^{***}	-0.40^{***}		-0.41^{***}	-0.42^{***}	-0.42^{***}	-0.40^{***}
	(0.02)	(0.03)		(0.03)	(0.03)	(0.03)	(0.03)
log(income)		11.57^{***}		12.12^{***}	11.75^{***}	11.80^{***}	11.51^{***}
		(1.82)		(1.80)	(1.77)	(1.78)	(1.81)
Constant	700.15^{***}	658.55^{***}	682.25^{***}	653.67^{***}	252.05	122.35	244.81
	(5.57)	(8.64)	(11.87)	(9.87)	(163.63)	(185.52)	(165.72)
N	420	420	420	420	420	420	420
Adjusted R^2	0.77	0.79	0.31	0.79	0.80	0.80	0.80

$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Robust standard errors in parentheses.

	score						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
str	-1.00^{***}	-0.73^{**}	-0.97	-0.53	64.34^{**}	83.70^{**}	65.29^{**}
	(0.27)	(0.26)	(0.59)	(0.34)	(24.86)	(28.50)	(25.26)
I(str^2)					-3.42^{**}	-4.38^{**}	-3.47^{**}
					(1.25)	(1.44)	(1.27)
I(str^3)					0.06^{**}	0.07^{**}	0.06^{**}
					(0.02)	(0.02)	(0.02)
str:HiEL			-1.28	-0.58		-123.28^{*}	
			(0.97)	(0.50)		(50.21)	
I(str^2):HiEL						6.12^{*}	
						(2.54)	
I(str^3):HiEL						-0.10^{*}	
						(0.04)	
english	-0.12^{***}	-0.18^{***}					-0.17^{***}
	(0.03)	(0.03)					(0.03)
HiEL			5.64	5.50	-5.47^{***}	816.08^{*}	
			(19.51)	(9.80)	(1.03)	(327.67)	
lunch	-0.55^{***}	-0.40^{***}		-0.41^{***}	-0.42^{***}	-0.42^{***}	-0.40^{***}
	(0.02)	(0.03)		(0.03)	(0.03)	(0.03)	(0.03)
log(income)		11.57^{***}		12.12^{***}	11.75^{***}	11.80^{***}	11.51^{***}
		(1.82)		(1.80)	(1.77)	(1.78)	(1.81)
Constant	700.15^{***}	658.55^{***}	682.25^{***}	653.67^{***}	252.05	122.35	244.81
	(5.57)	(8.64)	(11.87)	(9.87)	(163.63)	(185.52)	(165.72)
N	420	420	420	420	420	420	420
Adjusted R^2	0.77	0.79	0.31	0.79	0.80	0.80	0.80

$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$
Robust standard errors in parentheses.

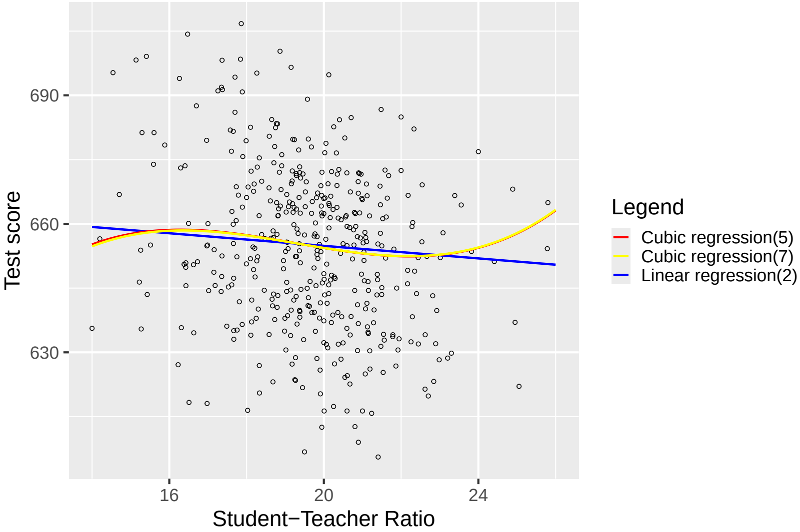
	score						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
str	-1.00^{***}	-0.73^{**}	-0.97	-0.53	64.34^{**}	83.70^{**}	65.29^{**}
	(0.27)	(0.26)	(0.59)	(0.34)	(24.86)	(28.50)	(25.26)
I(str^2)					-3.42^{**}	-4.38^{**}	-3.47^{**}
					(1.25)	(1.44)	(1.27)
I(str^3)					0.06^{**}	0.07^{**}	0.06^{**}
					(0.02)	(0.02)	(0.02)
str:HiEL			-1.28	-0.58		-123.28^{*}	
			(0.97)	(0.50)		(50.21)	
I(str^2):HiEL						6.12^{*}	
						(2.54)	
I(str^3):HiEL						-0.10^{*}	
						(0.04)	
english	-0.12^{***}	-0.18^{***}					-0.17^{***}
	(0.03)	(0.03)					(0.03)
HiEL			5.64	5.50	-5.47^{***}	816.08^{*}	
			(19.51)	(9.80)	(1.03)	(327.67)	
lunch	-0.55^{***}	-0.40^{***}		-0.41^{***}	-0.42^{***}	-0.42^{***}	-0.40^{***}
	(0.02)	(0.03)		(0.03)	(0.03)	(0.03)	(0.03)
log(income)		11.57^{***}		12.12^{***}	11.75^{***}	11.80^{***}	11.51^{***}
		(1.82)		(1.80)	(1.77)	(1.78)	(1.81)
Constant	700.15^{***}	658.55^{***}	682.25^{***}	653.67^{***}	252.05	122.35	244.81
	(5.57)	(8.64)	(11.87)	(9.87)	(163.63)	(185.52)	(165.72)
N	420	420	420	420	420	420	420
Adjusted R^2	0.77	0.79	0.31	0.79	0.80	0.80	0.80

$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$
Robust standard errors in parentheses.

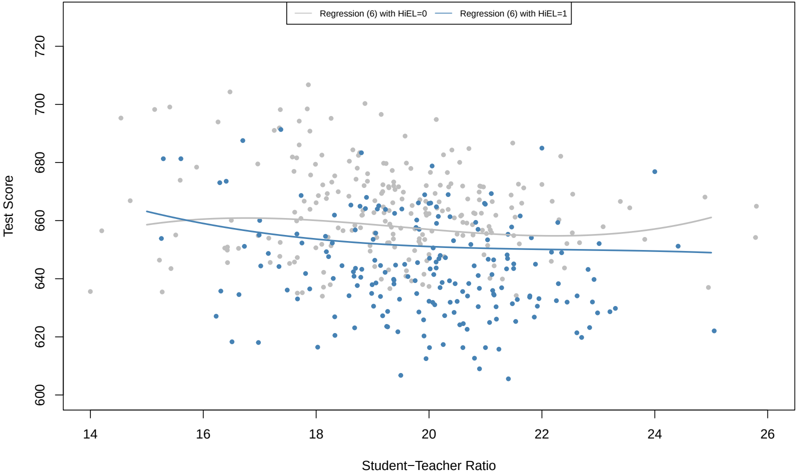
	score						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
str	-1.00^{***}	-0.73^{**}	-0.97	-0.53	64.34^{**}	83.70^{**}	65.29^{**}
	(0.27)	(0.26)	(0.59)	(0.34)	(24.86)	(28.50)	(25.26)
I(str^2)					-3.42^{**}	-4.38^{**}	-3.47^{**}
					(1.25)	(1.44)	(1.27)
I(str^3)					0.06^{**}	0.07^{**}	0.06^{**}
					(0.02)	(0.02)	(0.02)
str:HiEL			-1.28	-0.58		-123.28^{*}	
			(0.97)	(0.50)		(50.21)	
I(str^2):HiEL						6.12^{*}	
						(2.54)	
I(str^3):HiEL						-0.10^{*}	
						(0.04)	
english	-0.12^{***}	-0.18^{***}					-0.17^{***}
	(0.03)	(0.03)					(0.03)
HiEL			5.64	5.50	-5.47^{***}	816.08^{*}	
			(19.51)	(9.80)	(1.03)	(327.67)	
lunch	-0.55^{***}	-0.40^{***}		-0.41^{***}	-0.42^{***}	-0.42^{***}	-0.40^{***}
	(0.02)	(0.03)		(0.03)	(0.03)	(0.03)	(0.03)
log(income)		11.57^{***}		12.12^{***}	11.75^{***}	11.80^{***}	11.51^{***}
		(1.82)		(1.80)	(1.77)	(1.78)	(1.81)
Constant	700.15^{***}	658.55^{***}	682.25^{***}	653.67^{***}	252.05	122.35	244.81
	(5.57)	(8.64)	(11.87)	(9.87)	(163.63)	(185.52)	(165.72)
N	420	420	420	420	420	420	420
Adjusted R^2	0.77	0.79	0.31	0.79	0.80	0.80	0.80

$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$
Robust standard errors in parentheses.

Three specifications plotted against STR



Interaction plotted against STR



Wrap-up: interactions

- Interactions let one regressor **change the slope** attached to another—often the clearest way to model **heterogeneity**.
- When paired with **polynomials**, interaction terms sometimes act mainly as **flexible controls**; then you may care about **fit**, not each interaction coefficient on its own.
- In other designs—for example, **difference-in-differences**—the interaction coefficient **is** the estimand of primary interest.

Application: Jia and Ku (2019)

Jia and Ku (2019)

- Ruixue Jia and Hyejin Ku, "Is China's Pollution the Culprit for the Choking of South Korea? Evidence from the Asian Dust", *The Economic Journal*, 129(624), November 2019, 3154–3188.
- **Central question**:
  - Does air pollution from China spill over to South Korea, and if so, how does it affect South Koreans' health?

Empirical strategy

- **Benchmark specification**:
  - Outcome: mortality in South Korea (respiratory and cardiovascular deaths);
  - Regressor: pollution in China (air-quality index):

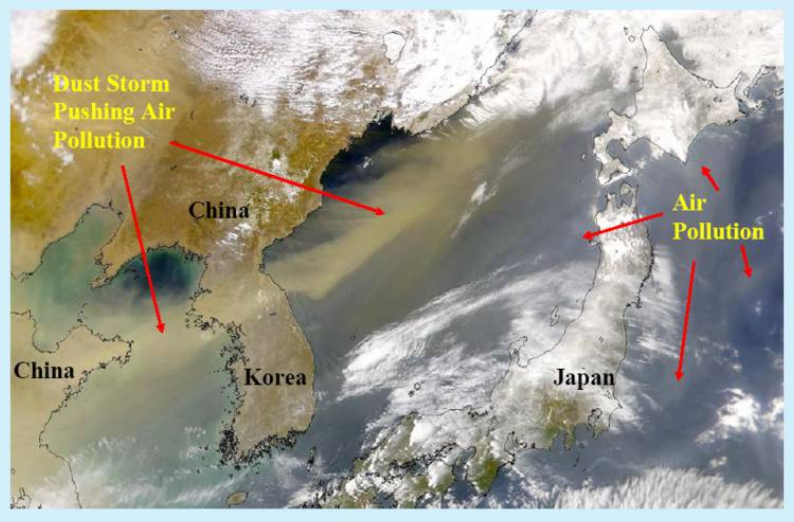
$$Mortality_{ijk} = \beta_0 + \beta_1 ChinesePollution_{jk} + \delta X_{ijk} + u_{ijk}$$

- **Co-movement** of Seoul pollution with Chinese pollution **does not prove** the health damage is **externally imported** rather than locally generated.
- **Confounders**: Korea's own emissions and weather can correlate with both mortality and measured Chinese pollution—classic endogeneity.
- **Search for variation**: is there something that **shifts exposure** to Chinese pollution without being a direct **health determinant** on its own?
- **Idea:** Asian dust episodes—days when long-range transport can **carry Chinese pollution** toward Korea along with desert dust.

Jia and Ku (2019): Asian dust as a carrier of pollutants

- Asian dust—yellow sand storms—is a **meteorological** shock affecting East Asia year-round but **especially in spring**.
  - Dust lifts from deserts in Mongolia, Kazakhstan, and **China (Inner Mongolia)**, where surface winds loft fine particles.
  - Easterly flow carries plumes across China, the Koreas, Japan, and the Russian Far East.
  - In recent decades these air masses increasingly **co-transport industrial pollutants**, much of it **sourced in China**.

Jia and Ku (2019): Asian Dust as a carrier of pollutants



Jia and Ku (2019): Asian Dust as a carrier of pollutants

- Why Asian dust **works for identification** in their design:
  1. **Directional**: pollution travels **toward** Korea on dust days—not a symmetric back-and-forth "swap."
  2. It may be **plausibly exogenous** to Korean **local** industrial conditions once you condition on detailed **weather**—yet it still generates **sharp** spatial and temporal variation.
  3. **Observable**: visibility and PM monitoring yield **well-defined indicators** of dust conditions at Korean stations.

Jia and Ku (2019): estimation strategy

- **Moderator**: Asian-dust days in South Korea.
- **Controls**: time, region, weather, local economic activity, and more.
- **Interactive specification** (pollution matters **more** when dust is present):

$$\begin{aligned} Mortality_{ijk} &= \beta_0 + \beta_1 AsianDust_{ijk} + \beta_2 ChinesePollution_{jk} \\ &\quad + \beta_3\, AsianDust_{ijk} \times ChinesePollution_{jk} + \delta_1 X_{ijk} + u_{ijk} \end{aligned}$$

- **Target estimand**: β_3—the **extra mortality response** on dust days to a given Chinese pollution reading (the **carrier channel**).

Jia and Ku (2019): results for the interaction terms

Table 2: The Impact of Dust*China's Pollution on Mortality Rates in South Korea

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Baseline							Placebo Tests	
	Mortality rates: Respiratory and Cardiovascular							Cancers	Accidents
Mean Dependent Var.	12.23							16.30	4.21
#Dust*China's Mean AQI				0.038**	0.033*	0.043**	0.040**	-0.008	0.007
				(0.016)	(0.018)	(0.018)	(0.019)	(0.020)	(0.009)
#Dust	0.076***		0.039	-0.251*	-0.214	-0.313**	0.072	0.525*	-0.102
	(0.025)		(0.032)	(0.131)	(0.142)	(0.149)	(0.240)	(0.275)	(0.119)
China's Mean AQI		0.265***	0.193*	0.117	0.138	0.202*	0.200*	-0.080	0.005
		(0.081)	(0.104)	(0.105)	(0.107)	(0.110)	(0.111)	(0.121)	(0.060)
District FE*Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Province FE*Month FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Weather (cubic polynomial)					Y	Y	Y	Y	Y
Local Prod (export, energy)						Y	Y	Y	Y
Local Prod*Dust							Y	Y	Y
Observations	29,464	29,464	29,464	29,464	28,952	28,024	28,024	28,024	28,024
R-squared	0.695	0.695	0.695	0.696	0.703	0.717	0.717	0.718	0.473

Jia and Ku (2019): results for the interaction terms

Table 2: The Impact of Dust*China's Pollution on Mortality Rates in South Korea

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Baseline							Placebo Tests	
	Mortality rates: Respiratory and Cardiovascular							Cancers	Accidents
Mean Dependent Var.	12.23							16.30	4.21
#Dust*China's Mean AQI				0.038**	0.033*	0.043**	0.040**	-0.008	0.007
				(0.016)	(0.018)	(0.018)	(0.019)	(0.020)	(0.009)
#Dust	0.076***		0.039	-0.251*	-0.214	-0.313**	0.072	0.525*	-0.102
	(0.025)		(0.032)	(0.131)	(0.142)	(0.149)	(0.240)	(0.275)	(0.119)
China's Mean AQI		0.265***	0.193*	0.117	0.138	0.202*	0.200*	-0.080	0.005
		(0.081)	(0.104)	(0.105)	(0.107)	(0.110)	(0.111)	(0.121)	(0.060)
District FE*Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Province FE*Month FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Weather (cubic polynomial)					Y	Y	Y	Y	Y
Local Prod (export, energy)						Y	Y	Y	Y
Local Prod*Dust							Y	Y	Y
Observations	29,464	29,464	29,464	29,464	28,952	28,024	28,024	28,024	28,024
R-squared	0.695	0.695	0.695	0.696	0.703	0.717	0.717	0.718	0.473

Jia and Ku (2019): results for the interaction terms

Table 2: The Impact of Dust*China's Pollution on Mortality Rates in South Korea

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Baseline							Placebo Tests	
	Mortality rates: Respiratory and Cardiovascular							Cancers	Accidents
Mean Dependent Var.	12.23							16.30	4.21
#Dust*China's Mean AQI				0.038**	0.033*	0.043**	0.040**	-0.008	0.007
				(0.016)	(0.018)	(0.018)	(0.019)	(0.020)	(0.009)
#Dust	0.076***		0.039	-0.251*	-0.214	-0.313**	0.072	0.525*	-0.102
	(0.025)		(0.032)	(0.131)	(0.142)	(0.149)	(0.240)	(0.275)	(0.119)
China's Mean AQI		0.265***	0.193*	0.117	0.138	0.202*	0.200*	-0.080	0.005
		(0.081)	(0.104)	(0.105)	(0.107)	(0.110)	(0.111)	(0.121)	(0.060)
District FE*Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Province FE*Month FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Weather (cubic polynomial)					Y	Y	Y	Y	Y
Local Prod (export, energy)						Y	Y	Y	Y
Local Prod*Dust							Y	Y	Y
Observations	29,464	29,464	29,464	29,464	28,952	28,024	28,024	28,024	28,024
R-squared	0.695	0.695	0.695	0.696	0.703	0.717	0.717	0.718	0.473

Jia and Ku (2019): results for the interaction terms

Table 2: The Impact of Dust*China's Pollution on Mortality Rates in South Korea

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Baseline							Placebo Tests	
	Mortality rates: Respiratory and Cardiovascular							Cancers	Accidents
Mean Dependent Var.	12.23							16.30	4.21
#Dust*China's Mean AQI				0.038**	0.033*	0.043**	0.040**	-0.008	0.007
				(0.016)	(0.018)	(0.018)	(0.019)	(0.020)	(0.009)
#Dust	0.076***		0.039	-0.251*	-0.214	-0.313**	0.072	0.525*	-0.102
	(0.025)		(0.032)	(0.131)	(0.142)	(0.149)	(0.240)	(0.275)	(0.119)
China's Mean AQI		0.265***	0.193*	0.117	0.138	0.202*	0.200*	-0.080	0.005
		(0.081)	(0.104)	(0.105)	(0.107)	(0.110)	(0.111)	(0.121)	(0.060)
District FE*Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Province FE*Month FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Weather (cubic polynomial)					Y	Y	Y	Y	Y
Local Prod (export, energy)						Y	Y	Y	Y
Local Prod*Dust							Y	Y	Y
Observations	29,464	29,464	29,464	29,464	28,952	28,024	28,024	28,024	28,024
R-squared	0.695	0.695	0.695	0.696	0.703	0.717	0.717	0.718	0.473

Jia and Ku (2019): placebo checks

Table 2: The Impact of Dust*China's Pollution on Mortality Rates in South Korea

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Baseline							Placebo Tests	
	Mortality rates: Respiratory and Cardiovascular							Cancers	Accidents
Mean Dependent Var.	12.23							16.30	4.21
#Dust*China's Mean AQI				0.038**	0.033*	0.043**	0.040**	-0.008	0.007
				(0.016)	(0.018)	(0.018)	(0.019)	(0.020)	(0.009)
#Dust	0.076***		0.039	-0.251*	-0.214	-0.313**	0.072	0.525*	-0.102
	(0.025)		(0.032)	(0.131)	(0.142)	(0.149)	(0.240)	(0.275)	(0.119)
China's Mean AQI		0.265***	0.193*	0.117	0.138	0.202*	0.200*	-0.080	0.005
		(0.081)	(0.104)	(0.105)	(0.107)	(0.110)	(0.111)	(0.121)	(0.060)
District FE*Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Province FE*Month FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Weather (cubic polynomial)					Y	Y	Y	Y	Y
Local Prod (export, energy)						Y	Y	Y	Y
Local Prod*Dust							Y	Y	Y
Observations	29,464	29,464	29,464	29,464	28,952	28,024	28,024	28,024	28,024
R-squared	0.695	0.695	0.695	0.696	0.703	0.717	0.717	0.718	0.473

Summary

Wrap-up

- We moved from **strict linearity in X** to **flexible** specifications—still estimated by **OLS**:
  - **Polynomials**, **logs**, and **interactions**;
  - the toolkit is the same multiple-regression algebra you already know;
  - what changes is **interpretation**: marginal effects, elasticities, and **differences across groups**.
- These tools are **standard in applied econometrics**. The objective is to move from **reported estimates** to **clear, substantive interpretations** that can be explained and defended in coursework or applied writing.