Econometrics 2

Nonlinear Regression Functions

Lasha Chochua

2026

Introduction

Why Nonlinear Regression?

- So far, we assumed the population regression function is **linear** – constant slope
- But the effect of X on Y may depend on the value of X itself or on another variable X_2
- Two groups of nonlinear methods:
  - Effect of X_1 on Y depends on the value of X_1 – polynomials, logarithms
  - Effect of X_1 on Y depends on the value of another variable X_2 – interaction terms
- Key insight: these models are **nonlinear in** X but **linear in the parameters** $\Rightarrow$ still estimated by OLS

The General Nonlinear Regression Model

General Nonlinear Population Regression Function

$$Y_i = f(X_{1i}, X_{2i}, \dots, X_{ki}) + u_i, \quad i = 1, \dots, n \tag{1}$$

where $f(\cdot)$ is a possibly nonlinear function of the independent variables.

- When f is linear: $f = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}$
- When f is nonlinear: the slope $\Delta Y / \Delta X_1$ depends on the values of $X_1, X_2, \dots, X_k$

Computing Effects in Nonlinear Models

Important

Key Concept 1: Effect of a Change in X_1
The expected change in Y from a change ΔX_1, holding $X_2, \dots, X_k$ constant:

$$\Delta Y = f(X_1 + \Delta X_1, X_2, \dots, X_k) - f(X_1, X_2, \dots, X_k) \tag{2}$$

Estimated effect:

$$\Delta \hat{Y} = \hat{f}(X_1 + \Delta X_1, X_2, \dots, X_k) - \hat{f}(X_1, X_2, \dots, X_k) \tag{3}$$

- This method **always** works – for large or small ΔX_1, continuous or discrete regressors

Polynomials

Polynomial Regression

Definition: Polynomial Regression of Degree r

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_r X_i^r + u_i \tag{4}$$

- $r = 2$: **quadratic** regression; $r = 3$: **cubic** regression
- Regressors are powers of the same variable X – still a multiple regression model $\Rightarrow$ estimate by OLS
- Individual coefficients $\beta_1, \beta_2, \ldots$ do **not** have simple interpretations – interpret by **plotting** and computing $\Delta\hat{Y}$

Quadratic Regression – Example

- Test scores vs. district income (California data, $n = 420$):

$$\widehat{TestScore} = \underset{(2.9)}{607.3} + \underset{(0.27)}{3.85} \cdot Income - \underset{(0.0048)}{0.0423} \cdot Income^2, \quad R^2 = 0.554 \tag{5}$$

- Testing linearity: $H_0 : \beta_2 = 0$
  - $t = -0.0423/0.0048 = -8.81$ – reject H_0 at all conventional levels
- Effect of income change from 10 to 11 (\$10K $\rightarrow$ \$11K):
  - $\Delta\hat{Y} = 644.53 - 641.57 = 2.96$ points
- Effect of income change from 40 to 41 (\$40K $\rightarrow$ \$41K):
  - $\Delta\hat{Y} = 694.04 - 693.62 = 0.42$ points
- Slope is **steeper at low income** – exactly what the quadratic captures

Testing the Degree of the Polynomial

- In the polynomial model (4), H_0: linear vs. H_1: polynomial of degree r

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

- Test using the F-**statistic** (joint hypothesis with $q = r - 1$ restrictions)

- **Sequential hypothesis testing** to choose degree:
  - Start with a maximum r (e.g., $r = 4$)
  - Test $H_0 : \beta_r = 0$ using t-statistic; if reject, use degree r
  - If fail to reject, drop X^r and test $\beta_{r-1} = 0$
  - Continue until the highest power is significant

Standard Errors of Estimated Effects

- For the quadratic model, the effect of income changing from 10 to 11 is:

$$\Delta\hat{Y} = \hat{\beta}_1(11-10) + \hat{\beta}_2(11^2-10^2) = \hat{\beta}_1 + 21\hat{\beta}_2 \tag{6}$$

- This is a linear combination of $\hat{\beta}_1$ and $\hat{\beta}_2$ – you know how to handle this! Apply $\widehat{\text{Var}}(\hat{\beta}) = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$:

$$\widehat{\text{Var}}(\Delta\hat{Y}) = \widehat{\text{Var}}(\hat{\beta}_1) + 21^2\widehat{\text{Var}}(\hat{\beta}_2) + 2\cdot 21\cdot\widehat{\text{Cov}}(\hat{\beta}_1,\hat{\beta}_2) \tag{7}$$

- Then $SE(\Delta\hat{Y}) = \sqrt{\widehat{\text{Var}}(\Delta\hat{Y})}$
- From the quadratic regression in (5): $\Delta\hat{Y} = 2.96,\ SE = 0.17$
  - 95% CI: $2.96 \pm 1.96 \times 0.17 = (2.63,\ 3.29)$

Logarithms

Logarithms and Percentages

- Key approximation: when $\Delta x/x$ is small,

$$\ln(x + \Delta x) - \ln(x) \approx \frac{\Delta x}{x}$$

- Changes in logarithms $\approx$ proportional (percentage) changes divided by 100
- Example: $x = 100$, $\Delta x = 1 \Rightarrow \Delta x/x = 0.01$; $\ln(101) - \ln(100) = 0.00995$

- Useful properties:

$$\ln(1/x) = -\ln(x), \quad \ln(ax) = \ln(a) + \ln(x), \quad \ln(x/a) = \ln(x) - \ln(a), \quad \ln(x^a) = a\ln(x)$$

Three Logarithmic Regression Models

Case	Specification	Interpretation of β_1
I: Linear-log	$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i$	A 1% change in $X \Rightarrow$ change in Y of $0.01\beta_1$
II: Log-linear	$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$	A unit change in $X \Rightarrow 100\beta_1\%$ change in Y
III: Log-log	$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$	A 1% change in $X \Rightarrow \beta_1\%$ change in Y (elasticity)

- Choose the specification based on **economic theory** – which relationship is most natural?
- R^2 can compare models with the **same** dependent variable (e.g., log-linear vs. log-log)
- R^2 **cannot** compare models with **different** dependent variables (e.g., linear-log vs. log-log)

Case I: Linear-Log – Example

$$\widehat{TestScore} = \underset{(3.8)}{557.8} + \underset{(1.40)}{36.42} \ln(Income), \quad R^2 = 0.561$$

- A 1% increase in income $\Rightarrow$ test scores increase by $0.01 \times 36.42 = 0.36$ points
- Effect of income change \$10K $\rightarrow$ \$11K:

$$\Delta\hat{Y} = 36.42 \times [\ln(11) - \ln(10)] = 3.47 \text{ points}$$

- Effect of income change \$40K $\rightarrow$ \$41K:

$$\Delta\hat{Y} = 36.42 \times [\ln(41) - \ln(40)] = 0.90 \text{ points}$$

- Larger effect at low income – similar to the quadratic

Case II: Log-Linear – Example

$$\ln(\widehat{Earnings}) = \underset{(0.019)}{2.876} + \underset{(0.0004)}{0.0095} \cdot Age, \quad R^2 = 0.033$$

- Each additional year of age $\Rightarrow$ earnings increase by 0.95%
- Natural specification when contracts specify percentage wage increases per year of service

Case III: Log-Log – Example

$$\widehat{\ln(TestScore)} = \underset{(0.006)}{6.336} + \underset{(0.0021)}{0.0554}\ln(Income), \quad R^2 = 0.557$$

- A 1% increase in income $\Rightarrow 0.0554\%$ increase in test scores
- $\beta_1 = 0.0554$ is the **elasticity** of test scores w.r.t. income

Interactions Between Independent Variables

Interaction Terms – The Idea

- **Interaction term**: the product $X_1 \times X_2$ included as a regressor
- Allows the effect of X_1 on Y to depend on the value of X_2 (and vice versa)
- Three cases:
    - Both X_1 and X_2 are **binary**
    - One **binary**, one **continuous**
    - Both **continuous**

Important

Key Concept 5
The coefficient on $X_1 \times X_2$ is the effect of a one-unit increase in **both** X_1 and X_2, above and beyond the sum of their individual effects.

Interactions Between Two Binary Variables

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

	$D_2 = 0$	$D_2 = 1$
$D_1 = 0$	β_0	$\beta_0 + \beta_2$
$D_1 = 1$	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_2 + \beta_3$

- Effect of D_1 (holding D_2 constant):
  - If $D_2 = 0$: effect $= \beta_1$
  - If $D_2 = 1$: effect $= \beta_1 + \beta_3$
- $\beta_3 =$ **difference** in the effect of D_1 between $D_2 = 1$ and $D_2 = 0$

Binary Interaction – Example

$$\widehat{TestScore} = \underset{(1.4)}{664.1} - \underset{(1.9)}{1.9}\,HiSTR - \underset{(2.3)}{18.2}HiEL - \underset{(3.1)}{3.5}\,(HiSTR \times HiEL),\ R^2 = 0.290$$

- $HiSTR = 1$ if STR ≥ 20; $HiEL = 1$ if % English learners $\geq 10\%$

	$HiEL = 0$	$HiEL = 1$
$HiSTR = 0$	664.1	645.9
$HiSTR = 1$	662.2	640.5

- Effect of high STR when few English learners: -1.9 points
- Effect of high STR when many English learners: $-1.9 - 3.5 = -5.4$ points

Interactions: Binary × Continuous

- Three possible specifications:

Model	Specification
Different intercept, same slope	$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + u_i$
Different intercept, different slope	$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 (X_i \times D_i) + u_i$
Same intercept, different slope	$Y_i = \beta_0 + \beta_1 X_i + \beta_2 (X_i \times D_i) + u_i$

- Most common: **different intercept and slope** – allows two fully separate regression lines
- When $D = 0$: regression line is $\beta_0 + \beta_1 X$
- When $D = 1$: regression line is $(\beta_0 + \beta_2) + (\beta_1 + \beta_3) X$
- β_3 = difference in slopes between $D = 1$ and $D = 0$

Binary × Continuous – Example

$$\widehat{TestScore} = \underset{(11.9)}{682.2} - \underset{(0.59)}{0.97} STR + \underset{(19.5)}{5.6} HiEL - \underset{(0.97)}{1.28} (STR \times HiEL), \; R^2 = 0.305$$

- Low % English learners ($HiEL = 0$): $\widehat{TestScore} = 682.2 - 0.97 \cdot STR$
- High % English learners ($HiEL = 1$): $\widehat{TestScore} = 687.8 - 2.25 \cdot STR$
- Reducing STR by 1:
  - Low EL districts: $+0.97$ points
  - High EL districts: $+2.25$ points
- Testing same slope: $t = -1.28/0.97 = -1.32$ – cannot reject at 10% level
- Testing same line ($\beta_2 = \beta_3 = 0$): $F = 89.9$ – reject at 1% level

Interactions Between Two Continuous Variables

- Model: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 (X_{1i} \times X_{2i}) + u_i$
- Effect of a change in X_1, holding X_2 constant:

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

- Effect of a change in X_2, holding X_1 constant:

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

- If both change simultaneously:

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

- The last term $\beta_3 \Delta X_1 \Delta X_2$ is the **extra** effect from changing both variables together

Continuous Interaction – Example

$$\widehat{TestScore} = \underset{(11.8)}{686.3} - \underset{(0.59)}{1.12}\,STR - \underset{(0.37)}{0.67}\,PctEL + \underset{(0.019)}{0.0012}(STR \times PctEL),\ R^2 = 0.422$$

- At median $PctEL = 8.85$: slope of STR $= -1.12 + 0.0012 \times 8.85 = -1.11$
- At 75th percentile $PctEL = 23.0$: slope $= -1.12 + 0.0012 \times 23.0 = -1.09$
- Very small difference – t-statistic on interaction: $0.0012/0.019 = 0.06$
- Conclusion: the effect of STR does **not** significantly depend on the % of English learners in this specification

Application: Nonlinear Effects on Test Scores

Three Questions

- After controlling for economic characteristics:
  - **Q1:** Does the effect of reducing STR depend on the fraction of English learners?
  - **Q2:** Does the effect depend on the **level** of STR itself?
  - **Q3:** What is the estimated effect of reducing STR by 2?
- Control variables: % eligible for subsidized lunch, $\ln(Income)$
- Approach: estimate multiple specifications and compare

Key Results from Table 8.3

- Adding $\ln(Income)$ as a control reduces the STR coefficient from -1.00 to -0.73 (both significant at 1%)
- **Interaction with HiEL**: $t = -0.58/0.50 = -1.16$ – not significant at 5%
- **Cubic specification** (regression 5):
  - F-test for $STR^2, STR^3 = 0$: $F = 6.17$ ($p < 0.001$) – reject linearity
  - Effect of STR $20 \to 18$: $\Delta\hat{Y} = 3.00$, 95% CI: $(1.64,\ 4.36)$
  - Effect of STR $22 \to 20$: $\Delta\hat{Y} = 1.93$, 95% CI: $(0.61,\ 3.25)$
- Nonlinear effect: reducing STR is **more effective for moderately sized classes**

Answers to the Three Questions

- **A1:** At most weak evidence that the STR effect depends on % English learners
  - Confidence intervals are wide; regression functions have similar slopes for $17 \leq STR \leq 23$
- **A2:** Yes – there is a nonlinear effect of STR on test scores
  - Linearity rejected at 1% level; effect is greater for moderate class sizes
- **A3:** Superintendent's problem – effect of reducing STR by 2:
  - Linear specification: $-0.73 \times (-2) = 1.46$ points (constant)
  - Cubic, $20 \to 18$: 3.00 points, CI $(1.64,\ 4.36)$
  - Cubic, $22 \to 20$: 1.93 points, CI $(0.61,\ 3.25)$
  - Nonlinear specs give a **more nuanced** answer

General Strategy and Summary

Five Steps for Modeling Nonlinearities

1. **Identify** a possible nonlinear relationship – use economic theory and intuition
2. **Specify** a nonlinear function and estimate by OLS (polynomials, logs, interactions)
3. **Test** whether the nonlinear model improves upon a linear model (t- and F-statistics)
4. **Plot** the estimated regression function – does it describe the data well?
5. **Estimate the effect** of a change in X using Key Concept 1

- The most important step: **use your head** – think about which nonlinearities matter for your substantive question before looking at the data

Chapter Summary

- **Polynomials**: include $X, X^2, X^3, \ldots$ as regressors – flexible but interpret by plotting
- **Logarithms**: convert to percentage changes
  - Linear-log: 1% $\Delta X \Rightarrow 0.01\beta_1$ change in Y
  - Log-linear: $\Delta X = 1 \Rightarrow 100\beta_1\%$ change in Y
  - Log-log: 1% $\Delta X \Rightarrow \beta_1\%$ change in Y (elasticity)
- **Interactions**: allow the effect of X_1 to depend on X_2
  - Binary $\times$ Binary, Binary $\times$ Continuous, Continuous $\times$ Continuous
- All these models are **linear in parameters** $\Rightarrow$ estimated by OLS
- Always compute $\Delta\hat{Y}$ via (3) and its SE via (7)

Required Reading

- Stock, J.H. and Watson, M.W. *Introduction to Econometrics*, 4th Global Edition, Chapter 8: Nonlinear Regression Functions (Sections 8.1–8.5).