Econ 312: Introduction to Econometrics

NonLinear Regression

Sang-Yeob Lee

May 13, 2010

Application: Nonlinear Effects on Test Scores of the Student-Teacher Ratio

Nonlinear specifications let us examine more nuanced questions about the Test score - STR relation, such as:

1. 1. Are there nonlinear effects of class size reduction on test scores? (Does a reduction from 35 to 30 have same effect as a reduction from 20 to 15?)
2. 2. Are there nonlinear interactions between PctEL and STR? (Are small classes more effective when there are many English learners?)

Strategy for Question #1 (different effects for different STR?)

- Estimate linear and nonlinear functions of *STR*, holding constant relevant demographic variables
  - PctEL
  - Income (remember the nonlinear TestScore-Income relation!)
- See whether adding the nonlinear terms makes an "economically important" quantitative difference ("economic" or "real-world" importance is different than statistically significant)
- Test for whether the nonlinear terms are significant

Strategy for Question #2 (interactions between PctEL and STR?)

- Estimate linear and nonlinear functions of *STR*, interacted with *PctEL*.
- If the specification is nonlinear (with STR, STR^2, STR^3), then you need to add interactions with all the terms so that the entire functional form can be different, depending on the level of *PctEL*.

Interactions Between Independent Variables

- Perhaps a class size reduction is more effective in some circumstances than in others
- Perhaps smaller classes help more if there are many English learners, who need individual attention
- That is, $\frac{\triangle TestScore}{\triangle STR}$ might depend on $PctEL$
- More generally, $\frac{\triangle Y}{\triangle X_1}$ might depend on X_2
- How to model such "interactions" between X_1 and X_2?
- We first consider binary X's, then continuous X's

(a) Interactions between two binary variables

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \epsilon_i$$

- D_{1i}, D_{2i} are binary
- β_1 is the effect of changing $D_1 = 0$ to $D_1 = 1$. In this specification, this effect doesn't depend on the value of D_2.
- To allow the effect of changing D_1 to depend on D_2, include the "interaction term" $D_{1i} \times D_{2i}$ as a regressor:

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

Interpreting the coefficients

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

General rule: compare the various cases (b)
$E(Y_i | D_{1i} = 0, D_{2i} = d_2) = \beta_0 + \beta_2 d_2$ (a)
$E(Y_i | D_{1i} = 1, D_{2i} = d_2) = \beta_0 + \beta_1 + \beta_2 d_2 + \beta_3 d_2$
subtract (a) - (b):
$E(Y_i | D_{1i} = 1, D_{2i} = d_2) - E(Y_i | D_{1i} = 0, D_{2i} = d_2) = \beta_1 + \beta_3 d_2$

- The effect of D_1 depends on d_2 (what we wanted)
- β_3=increment to the effect of D_1, when $D_2 = 1$

Example: TestScore, STR, English learners

Let $HiSTR = 1$ if $STR \geq 20$, 0 otherwise and $HiEL = 1$ if $PecEL \geq 10$, 0 otherwise.

$$\widehat{TestScore} = 664.1 - 18.2HiEL - 1.9HiSTR - 3.5(HiSTR \times HiEL)$$

$$(1.4) \quad (2.3) \quad (1.9) \quad (3.1)$$

- "Effect" of $HiSTR$ when $HiEL = 0$ is -1.9
- "Effect" of $HiSTR$ when $HiEL = 1$ is $-1.9 - 3.5 = -5.4$
- Class size reduction is estimated to have a bigger effect when the percent of English learners is large
- This interaction isn't statistically significant: $t = 3.5/3.1$

(b) Interactions between continuous and binary variables

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 Xi + \epsilon_i$$

- D_i is binary, X is continuous
- As specified above, the effect on Y of X (holding constant D)= β_2, which does not depend on D
- To allow the effect of X to depend on D, include the "interaction term" $D_i \times X_i$ as a regressor:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + \beta_3 (D_i \times X_i) + \epsilon_i$$

Binary-continuous interactions: the two regression lines

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + \beta_3 (D_i \times X_i) + \epsilon_i$$

- Observations with $D_i = 0$ (the "$D = 0$" group):

$$Y_i = \beta_0 + \beta_2 X_i + \epsilon_i$$

- Observations with $D_i = 1$ (the "$D = 1$" group):

$$Y_i = (\beta_0 + \beta_1) + (\beta_2 + \beta_3) X_i + \epsilon_i$$

Interpreting the coefficients

General rule: compare the various cases

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

Now change X:

$$(a) Y + \triangle Y = \beta_0 + \beta_1 D + \beta_2 (X + \triangle X) + \beta_3 [D \times (X + \triangle X)]$$

subtract (a)-(b):

$$\triangle Y = \beta_2 \triangle X + \beta_3 D \triangle X \quad \text{or} \quad \frac{\triangle Y}{\triangle X} = \beta_2 + \beta_3 D$$

- The effect of X depends on D (what we wanted)
- β_3 = increment to the effect of X, when $D = 1$

Example: *TestScore*, *STR*, *HiEL* (= 1 if *PctEL* $\geq$ 10)

$$\widehat{TestScore} = 682.2 - 0.97STR - 5.6HiEL - 1.28(STR \times HiEL)$$

$$(11.9) \qquad (0.59) \qquad (19.5) \qquad (0.97)$$

- When $HiEL = 0$:
  $\widehat{TestScore} = 682.2 - 0.97STR$
- When $HiEL = 1$,
  $\widehat{TestScore} = 682.2 - 0.97STR - 5.6 - 1.28STR = 687.8 - 2.25STR$
- Two regression lines: one for each *HiSTR* group
- Class size reduction is estimated to have a larger effect when the percent of English learners is large.

Example, ctd: Testing hypotheses

$$\widehat{TestScore} = 682.2 - 0.97STR - 5.6HiEL - 1.28(STR \times HiEL)$$

$$(11.9) \qquad (0.59) \qquad (19.5) \qquad (0.97)$$

- The two regression lines have the same slope=the coefficient on $STR \times HiEL$ is zero: $t = -1.28/0.97 = -1.32$
- The two regression lines have the same intercept= the coefficient on $HiEL$ is zero: $t = -5.6/19.5 = 0.29$
- The two regression lines are the same population coefficient on $HiEL = 0$ and population coefficient on $STR \times HiEL = 0$: $F = 89.94$ (p-value $< .001$) !!
- We reject the joint hypothesis but neither individual hypothesis (how can this be?) These three tests produce seemingly contradictory results.

(c) Interactions between two continuous variables

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

- X_1, X_2 are continuous
- As specified, the effect of X_1 doesn't depend on X_2
- As specified, the effect of X_2 doesn't depend on X_1
- To allow the effect of X_1 to depend on X_2, include the "interaction term" $X_{1i} \times X_{2i}$ as a regressor:

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

Interpreting the coefficients

General rule: compare the various cases

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

Now change X_1:

$$(a) Y + \triangle Y = \beta_0 + \beta_1 (X_1 + \triangle X_1) + \beta_2 X_2 + \beta_3 [(X_1 + \triangle X_1) X_3]$$

subtract (a)-(b):

$$\triangle Y = \beta_1 \triangle X_1 + \beta_3 X_2 \triangle X_1 \quad \text{or} \quad \frac{\triangle Y}{\triangle X_1} = \beta_1 + \beta_3 X_2$$

- The effect of X_1 depends on X_2 (what we wanted)
- β_3 = increment to the effect of X_1,from a unit change in X_2

Example: *TestScore*, *STR*, *PctEL*

$$\widehat{TestScore} = 686.3 - 1.127STR - 0.67PctEL - 0.012(STR \times PctEL)$$

$$(11.8) \qquad (0.59) \qquad (0.37) \qquad (0.019)$$

The estimated effect of class size reduction is nonlinear because the size of the effect itself depends on *PctEL*:

$$\frac{\triangle TestScore}{\triangle STR} = -1.12 + 0.0012PctEL$$

- $PctEL = 0$, $\frac{\triangle TestScore}{\triangle STR} = -1.12$
- $PctEL = 20\%$, $\frac{\triangle TestScore}{\triangle STR} = -1.12 + 0.0012 \times 20 = -1.10$

Example, ctd: Testing hypotheses

$$\widehat{TestScore} = 686.3 - 1.127STR - 0.67PctEL - 0.012(STR \times PctEL)$$

$$(11.8) \qquad (0.59) \qquad (0.37) \qquad (0.019)$$

- Does population coefficient on $STR \times PctEL = 0$?
  $t = .0012/.019 = .06$ can't reject null at 5% level
- Does population coefficient on STR $= 0$?
  $t = -1.12/0.59 = -1.90$ can't reject null at 5% level
- Do the coefficients on both STR and $STR \times PctEL = 0$?
  $F = 3.89$ (p-value $= .021$) reject null at 5% level(!!) (Why? high but imperfect multicollinearity)