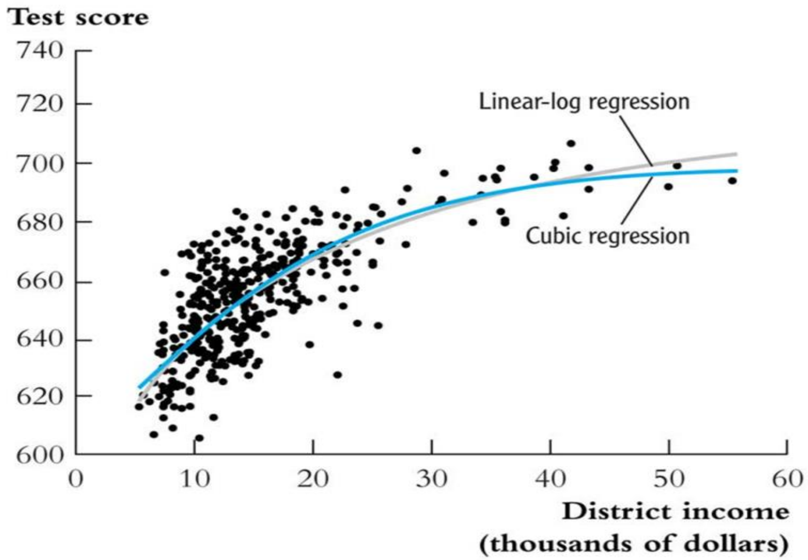
EC 3303: Econometrics I

Nonlinear Regression Functions



Kelvin Seah

Outline

1. Tying some loose ends from previous lecture

2. Interactions Between Independent Variables

Back to Polynomials

- Which degree polynomial should you use in practice?

Interactions Between Independent Variables

- Could reducing class size have a larger effect on test scores in districts with many ELL students than in districts with few?

- If so, $\frac{\Delta TestScore}{\Delta STR}$ would depend on the value of $PctEL$.

- More generally, $\frac{\Delta Y}{\Delta X_1}$ might depend on the value of X_2.

- How to model such "***interactions***"?

- 3 cases: (1) both independent variables are binary (2) one is binary, other is continuous (3) both are continuous.

(1) Interactions between Two Binary Variables

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + u_i \quad (20)$$

a) whether a district has a large class size (D_{1i}, where $D_{1i} = 1$ if the i^{th} district has a large class size ($STR \geq 20$) & $D_{1i} = 0$ otherwise)

b) whether a district has a large prevalence of ELL (D_{2i}, where $D_{2i} = 1$ if the percentage of ELL in the i^{th} district is 10% or more & $D_{2i} = 0$ otherwise)

<u>Limitation?</u>

- assumes that the effect of changing class size is the same for districts with a large and with a small prevalence of ELL students.

- Can allow the effect of changing class size to differ for districts with a large and with a small prevalence of ELL:

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 (D_{1i} \times D_{2i}) + u_i \quad (21)$$

- Let d_2 be the value of D_{2i}
- Assume LSA #1 holds: $[E(u_i|D_{1i}, D_{2i}) = 0]$

$\boldsymbol{D_{1i} = 1}$ ***(large class) ;*** $\boldsymbol{D_{1i} = 0}$ ***(small class)***

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

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

$\boldsymbol{D_{1i} = 1}$ ***(large class) ;*** $\boldsymbol{D_{1i} = 0}$ ***(small class)***

$$E(Y_i|D_{1i} = 1, D_{2i} = d_2)$$

$$= (\beta_0 + \beta_1) + (\beta_2 + \beta_3)d_2$$

$$E(Y_i|D_{1i} = 0, D_{2i} = d_2)$$

$$= \beta_0 + \beta_2 d_2$$

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

- effect of class size now depends on the prevalence of ELL.
- How to interpret β_3?

More generally

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

where D_{1i} , D_{2i} are binary

- β_1 is the effect on Y, on average, of changing $D_{1i} = 0$ to $D_{1i} = 1$
- In this specification, ***the effect of $\mathbf{D_1}$ does not depend on the value of $\mathbf{D_2}$***.
- To allow the effect of changing D_1 to depend on D_2, include $(D_{1i} \times D_{2i})$ as a regressor:

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

- Difference in conditional expectations is:

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

- effect on Y of a change in D_1 now depends on D_2.
- β_3 = difference in the effect on Y of D_1, when $D_2 = 1$ and when $D_2 = 0$.

Example: TestScore, HiSTR, HiEL

$$\widehat{TestScore} = 664.1 - 1.9HiSTR - 18.2HiEL - 3.5(HiSTR \times HiEL)$$
$$\begin{array}{cccc} (1.4) & (1.9) & (2.3) & (3.1) \end{array}$$

when $HiEL = 0$,

$$\widehat{TestScore} = 664.1 - 1.9HiSTR$$

when $HiEL = 1$,

$$\widehat{TestScore} = 664.1 - 1.9HiSTR - 18.2(1) - 3.5(HiSTR \times 1)$$
$$= 645.9 - 5.4HiSTR$$

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

- "Effect" of $HiSTR$ when $HiEL = 0$ is –1.9
- "Effect" of $HiSTR$ when $HiEL = 1$ is –1.9 – 3.5 = –5.4
- Class size change is estimated to have a bigger effect when the percent of English learners is large.
- But this interaction is not statistically significant:

$$|t\text{-}statistic| = |-3.5-0/3.1|=1.13<1.96.$$

(2) Interactions between a Continuous and a Binary Variable

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + u_i \quad (22)$$

where X_i is continuous & D_i is binary.

- Specified this way, the effect on Y of X = β_1.
- To allow the effect of X to depend on D,

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 (X_i \times D_i) + u_i \quad (23)$$

  - modified specification allows for 2 different regression lines relating Y_i & X_i, depending on the value of D_i.

Binary-Continuous Interactions: Two Regression Lines

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 (X_i \times D_i) + u_i \quad (23)$$

regression function:

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

If $D_i = 1$,

$$\beta_0 + \beta_1 X + \beta_2 (1) + \beta_3 (X \times 1)$$
$$= \underbrace{\beta_0 + \beta_2}_{\text{Intercept}} + \underbrace{(\beta_1 + \beta_3)X}_{\text{Slope}}$$

(***regression line for the*** $\boldsymbol{D_i = 1}$ ***group***)

If $D_i = 0$,

$$\underbrace{\beta_0}_{\text{Intercept}} + \underbrace{\beta_1 X}_{\text{Slope}}$$

(***regression line for the*** $\boldsymbol{D_i = 0}$ ***group***)

Consider the regression model without the interaction term again:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + u_i \quad (22)$$

regression function:

$$\beta_0 + \beta_1 X + \beta_2 D$$

when $D_i = 1$,

$$\beta_0 + \beta_1 X + \beta_2(1) = \underbrace{(\beta_0 + \beta_2)}_{\text{Intercept}} + \underbrace{\beta_1 X}_{\text{Slope}}$$

when $D_i = 0$,

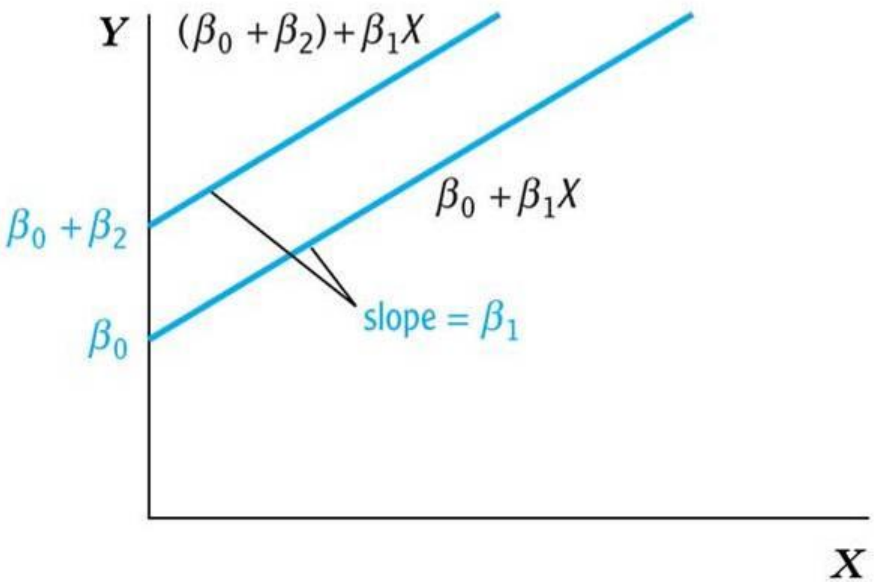
$$\underbrace{\beta_0}_{\text{Intercept}} + \underbrace{\beta_1 X}_{\text{Slope}}$$

Under this specification, the effect of a change in X is the same whether the value of D_i is 0 or 1

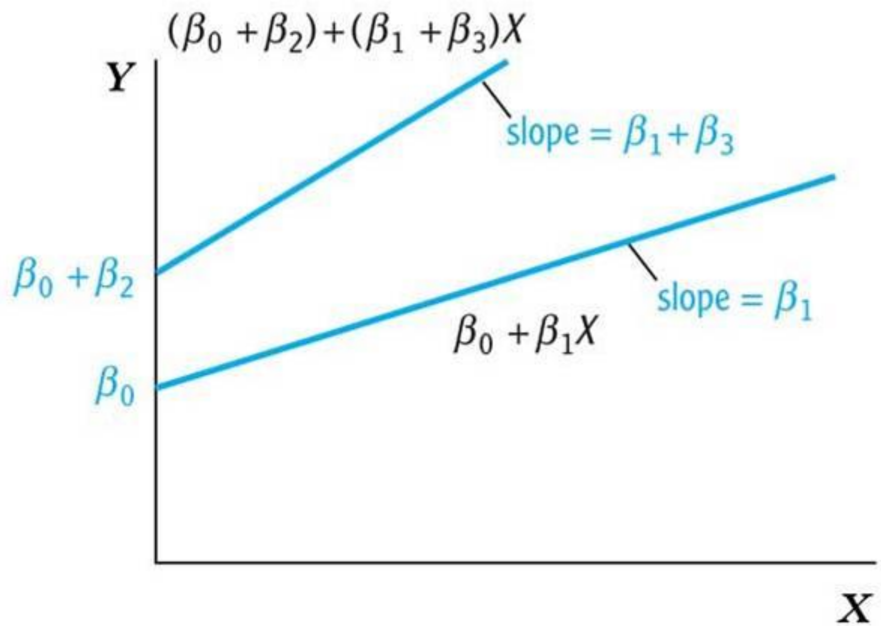
- the two regression lines have the ***same slope*** though ***different intercepts***.

- So inclusion of a continuous-binary interaction allows ***the effect of a change in X*** to be ***different depending on the value of the binary variable.***

  - slope of the regression line is allowed to be different under $D_i = 1$ & under $D_i = 0$.



(a) Different intercepts, same slope



(b) Different intercepts, different slopes

Interpreting the Coefficients: Binary-Continuous Interactions

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

regression function:

$$Y = \beta_0 + \beta_1 X + \beta_2 D + \beta_3 (X \times D) \quad (24)$$

Change X by ΔX. Value of the regression function becomes:

$$Y + \Delta Y = \beta_0 + \beta_1 (X + \Delta X) + \beta_2 D + \beta_3 ((X + \Delta X) \times D) \ \ (25)$$

(25)-(24): $\Delta Y = \beta_1 \Delta X + \beta_3 (\Delta X \times D)$

- So, $\frac{\Delta Y}{\Delta X} = \beta_1 + \beta_3 D$
- $\beta_3 =$ difference in the effect of a unit change in X, for observations in which $D_i = 1$ and observations in which $D_i = 0$.

Example: *TestScore, STR, HiEL* (=1 if *PctEL* ⍰ 10)

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

For districts with $HiEL_i = 0$,

$$\widehat{TestScore} = 682.2 - 0.97STR$$

For districts with $HiEL_i = 1$,

$$\widehat{TestScore} = 682.2 - 0.97STR + 5.6(1) - 1.28(STR \times 1) = 687.8 - 2.25STR$$

$$\widehat{TestScore} = 682.2 - 0.97STR \quad (HiEL_i = 0 \text{ districts})$$

$$\widehat{TestScore} = 687.8 - 2.25STR \quad (HiEL_i = 1 \text{ districts})$$

- A 1 unit reduction in *STR* is predicted to increase test scores:
  - by 0.97 points in districts with a low prevalence of ELL.
  - by 2.25 points in districts with a high prevalence of ELL.

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

Test the hypothesis that…

1) 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 < 0.001)
- reject the hypothesis that the two regression lines are the same.

2) the two regression lines have the ***same slope but different intercept*** <=> population coefficient on ($STR \times HiEL$) = 0.

- t-statistic testing $\beta_{STR \times HiEL} = 0$ is -1.28-0/0.97= -1.32
- *Since* $|t^{act}| = 1.32$ <1.64, do not reject the hypothesis that the two regression lines have the same slope at the 10% level.

(3) Interactions between Two Continuous Variables

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \quad (26)$$

where X_{1i} , X_{2i} are continuous.

- Specified this way, the effect on Y of X_1 does not depend on the value of X_2.

Also,

- the effect on Y of X_2 does not depend on the value of X_1.

To allow the effect of a unit change in X_1 to depend on the value of X_2 (and vice-versa),

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 (X_{1i} \times X_{2i}) + u_i \quad (27)$$

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

regression function is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 \times X_2) \qquad (28)$$

change X_1 by ΔX_1. Value of the regression function becomes:

$$Y + \Delta Y = \beta_0 + \beta_1 (X_1 + \Delta X_1) + \beta_2 X_2 + \beta_3 [(X_1 + \Delta X_1) \times X_2] \qquad (29)$$

(29)-(28): $\Delta Y = \beta_1 \Delta X_1 + \beta_3 (\Delta X_1 \times X_2)$

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

- effect of X_1 now depends on the value of X_2.

Example: TestScore, STR, PctEL

$$\widehat{TestScore} = 686.3 - 1.12STR - 0.67PctEL + 0.0012(STR \times PctEL)$$
$$(11.8) \quad (0.59) \quad (0.37) \quad (0.019)$$

estimated effect of class size reduction is nonlinear because the size of the effect depends on $PctEL$:

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

Value of PctEL	$\boldsymbol{\frac{\Delta TestScore}{\Delta STR}}$
0	*–1.12+(0.0012*$\times$ 0) $= -1.12$
20%	*–1.12+(0.0012×20) = –1.10*

Adding Control Variables

- To keep the analysis simple, we have not included control variables into our nonlinear regression models.

- All the tools we have discussed remain unchanged when we add control variables…

Example

- Suppose we want to know how class size affects testscores, controlling for student characteristics.

- Can estimate a regression model:

$$Testscore_i = \beta_0 + \beta_1 STR_i + \beta_2 STR_i{}^2 + \beta_3 STR_i{}^3 + \beta_4 HiEL_i + \beta_5 LchPct_i + \beta_6 ln(Income_i) + u_i \qquad (30)$$

- If we estimate (30) using our sample of data and obtain:

$$\widehat{Testscore} = 252.0 + 64.33STR - 3.42STR^2 + 0.059STR^3 - 5.47HiEL - 0.420LchPct + 11.75ln(Income)$$

(standard errors dropped for simplicity)

- How much is test score estimated to change when *STR* is reduced by 1 from 21 to 20?

$$\widehat{Testscore} = 252.0 + 64.33STR - 3.42STR^2 + 0.059STR^3 - 5.47HiEL - 0.420PctEL + 11.75ln(Income)$$

- How much is test score estimated to change, on average, when *STR* is reduced by 1 from 21 to 20?

$$\Delta\widehat{Testscore} = [64.33(20) - 3.42(20)^2 + 0.059(20)^3] - [64.33(21) - 3.42(21)^2 + 0.059(21)^3] = 1.49$$

$$\widehat{Testscore} = 252.0 + 64.33STR - 3.42STR^2 + 0.059STR^3 - 5.47HiEL - 0.420PctEL + 11.75ln(Income)$$

- How much is test score estimated to change, on average, when *STR* is reduced by 2 from 21 to 19?

$$\Delta\widehat{Testscore} = [64.33(19) - 3.42(19)^2 + 0.059(19)^3] - [64.33(21) - 3.42(21)^2 + 0.059(21)^3] = 3.22$$

An Application of the Interaction Specification

1. *Does beauty have an effect on earnings?*

Hamermesh, D. S. and Biddle, J. E. (1994). Beauty and the Labor Market. *American Economic Review*, 84(5):1174–94

2. Does the effect of beauty on earnings differ by gender?

- How to answer this?