Lecture 6

OLS Extension: Nonlinear Regression Function

Zhaopeng Qu

2017 Fall

Dummy Variables

- In some cases, we must also incorporate qualitative factors into regression models.

- Gender/race of an individual
- The industry of an firm
- The region in China where a city is located

Interactions Between Independent Variables

- Perhaps a class size reduction is more effective in some circumstances than in others…such as smaller classes help more if there are many English learners, who need individual attention.
- That is, $\frac{\Delta TestScore}{\Delta STR}$ might depend on *PctEL*(percentage of English learners).
- More generally, $\frac{\Delta Y}{\Delta X_1}$ might depend on X_2(Another Important variable have impact on Y)

Interactions: Three Types

1. Interactions Between 2 Binary Variables

1. Interactions Between Continuous and Binary

2. Interactions Between 2 Continuous Variables

1. Two Binary(Discrete)Variables

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

- 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:

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

Interpreting the coefficients

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

General rule: compare the various cases

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

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

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

Let

$$HiSTR = \begin{cases} 1 \text{ if } STR \geq 20 \\ 0 \text{ if } STR < 20 \end{cases} \qquad HiEL = \begin{cases} 1 \text{ if } PctEL \geq 10 \\ 0 \text{ if } PctEL < 10 \end{cases}$$

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

- "Effect" of *HiSTR* when *HiEL* = 0 is –1.9
- "Effect" of *HiSTR* when *HiEL* = 1 is –1.9 – 3.5 = –5.4
- Class size reduction is estimated to have a bigger effect when the percent of English learners is large.

Example: TestScore, STR, English

- Can you relate these coefficients to the following table of group ("cell") means?

	Low STR	*High STR*
Low EL	664.1	662.2
High EL	645.9	640.5

2. Continuous and Binary Variables

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + u_i$$

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

Binary-continuous interactions:

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

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

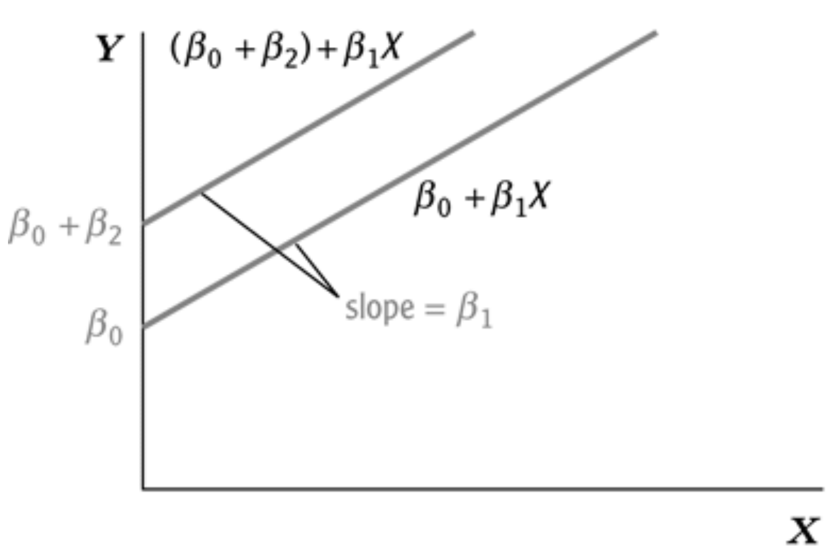
$$Y_i = \beta_0 + \beta_2 X_i + u_i \qquad \textbf{\textit{The D=0 regression line}}$$

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

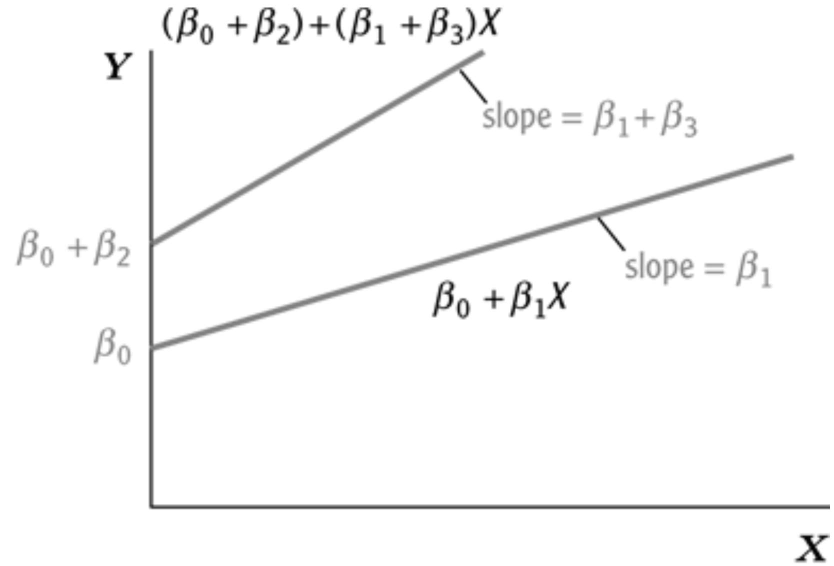
$$Y_i = \beta_0 + \beta_1 + \beta_2 X_i + \beta_3 X_i + u_i$$

$$= (\beta_0+\beta_1) + (\beta_2+\beta_3)X_i + u_i \quad \textbf{\textit{The D=1 regression line}}$$

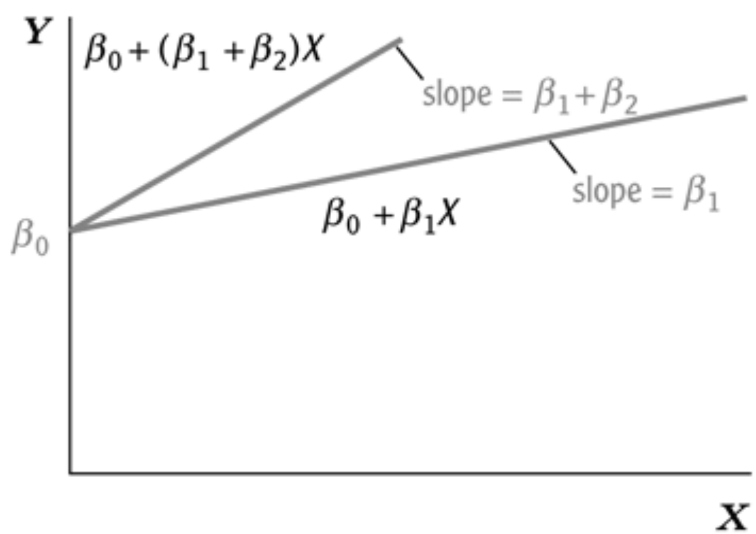
Binary-Continuous Interactions



(a) Different intercepts, same slope



(b) Different intercepts, different slopes



(c) Same intercept, different slopes

Interpreting the coefficients

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

General rule: compare the various cases

$$Y = \beta_0 + \beta_1 D + \beta_2 X + \beta_3 (D \times X) \qquad \text{(b)}$$

Now change *X*:

$$Y + \Delta Y = \beta_0 + \beta_1 D + \beta_2 (X + \Delta X) + \beta_3 [D \times (X + \Delta X)] \qquad \text{(a)}$$

subtract (a) – (b):

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

- 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 ≥ 10)

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

- 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.

3. Two Continuous Cariables

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

- 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:

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

Interpreting the coefficients:

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

General rule: compare the various cases

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3(X_1 \times X_2) \qquad \text{(b)}$$

Now change X_1:

$$Y + \Delta Y = \beta_0 + \beta_1(X_1 + \Delta X_1) + \beta_2 X_2 + \beta_3[(X_1 + \Delta X_1) \times X_2] \qquad \text{(a)}$$

subtract (a) – (b):

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

- 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

$$\widehat{TestScore} = \underset{(11.8)}{686.3} - \underset{(0.59)}{1.12}STR - \underset{(0.37)}{0.67}PctEL + \underset{(0.019)}{.0012}(STR \times PctEL),$$

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

$$\frac{\Delta TestScore}{\Delta STR} = -1.12 + .0012PctEL$$

PctEL	$\frac{\Delta TestScore}{\Delta STR}$
0	–1.12
20%	–1.12+.0012 × 20 = –1.10

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. 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. 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

1. *PctEL(percentage of English learners)*
2. *Income* (remember the nonlinear *TestScore-Income* relation!)
3. *LunchPCT* (fraction on free/subsidized lunch)

- 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: 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*.
- We will use a binary-continuous interaction specification by adding $HiEL \times STR$, $HiEL \times STR^2$, and $HiEL \times STR^3$.

What is a good “base” specification?

- The *TestScore* – *Income* relation:
- The logarithmic specification is better behaved near the extremes of the sample, especially for large values of income.

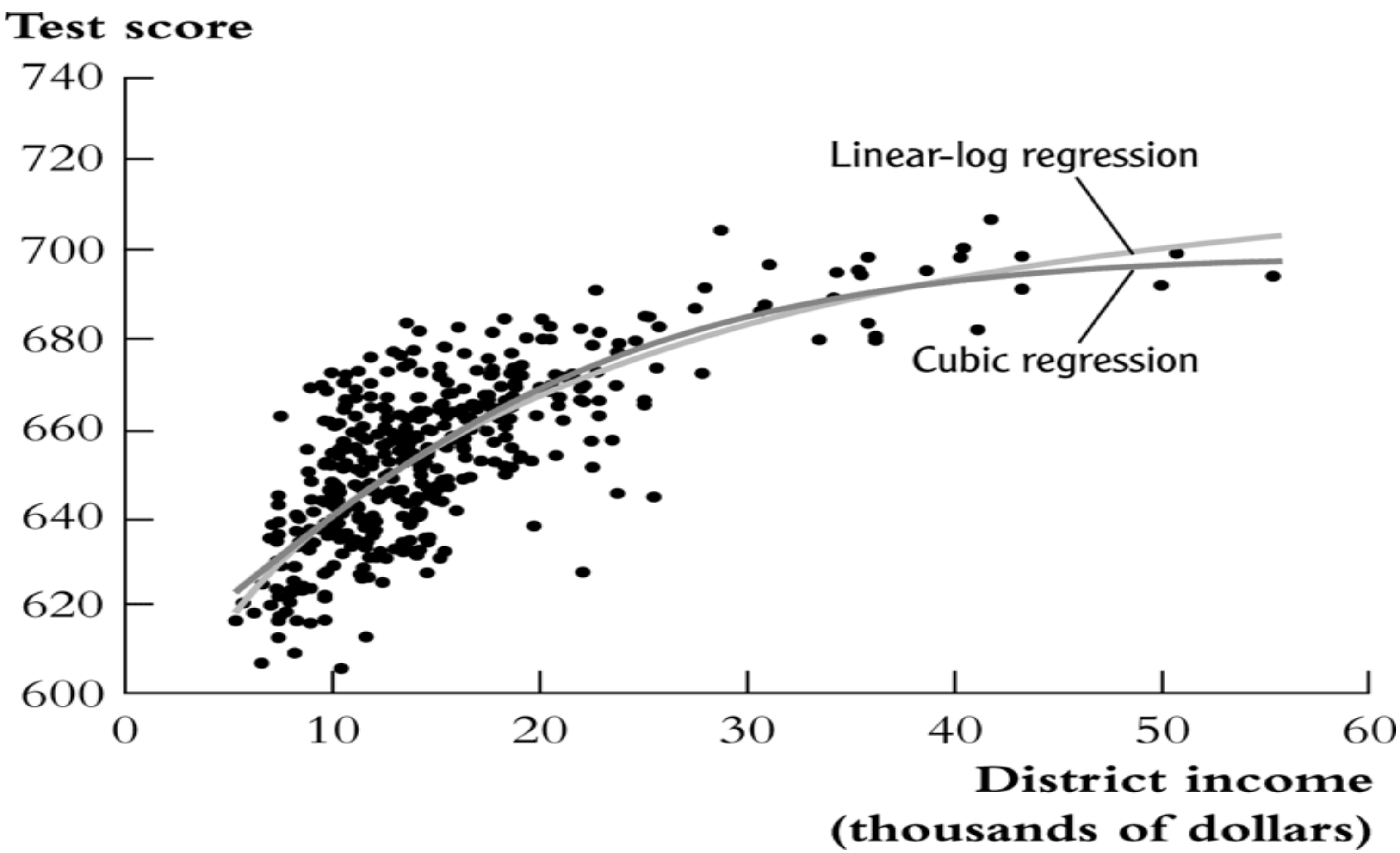


TABLE 8.3 Nonlinear Regression Models of Test Scores

Dependent variable: average test score in district; 420 observations.

Regressor	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Student–teacher ratio (*STR*)	−1.00** (0.27)	−0.73** (0.26)	−0.97 (0.59)	−0.53 (0.34)	64.33** (24.86)	83.70** (28.50)	65.29** (25.26)
STR^2					−3.42** (1.25)	−4.38** (1.44)	−3.47** (1.27)
STR^3					0.059** (0.021)	0.075** (0.024)	0.060** (0.021)
% English learners	−0.122** (0.033)	−0.176** (0.034)					−0.166** (0.034)
% English learners ≥ 10%? (Binary, *HiEL*)			5.64 (19.51)	5.50 (9.80)	−5.47** (1.03)	816.1* (327.7)	
$HiEL \times STR$			−1.28 (0.97)	−0.58 (0.50)		−123.3* (50.2)	
$HiEL \times STR^2$						6.12* (2.54)	
$HiEL \times STR^3$						−0.101* (0.043)	
% Eligible for subsidized lunch	−0.547** (0.024)	−0.398** (0.033)		−0.411** (0.029)	−0.420** (0.029)	−0.418** (0.029)	−0.402** (0.033)
Average district income (logarithm)		11.57** (1.81)		12.12** (1.80)	11.75** (1.78)	11.80** (1.78)	11.51** (1.81)
Intercept	700.2** (5.6)	658.6** (8.6)	682.2** (11.9)	653.6** (9.9)	252.0 (163.6)	122.3 (185.5)	244.8 (165.7)

Tests of joint hypotheses:

F-Statistics and p-Values on Joint Hypotheses

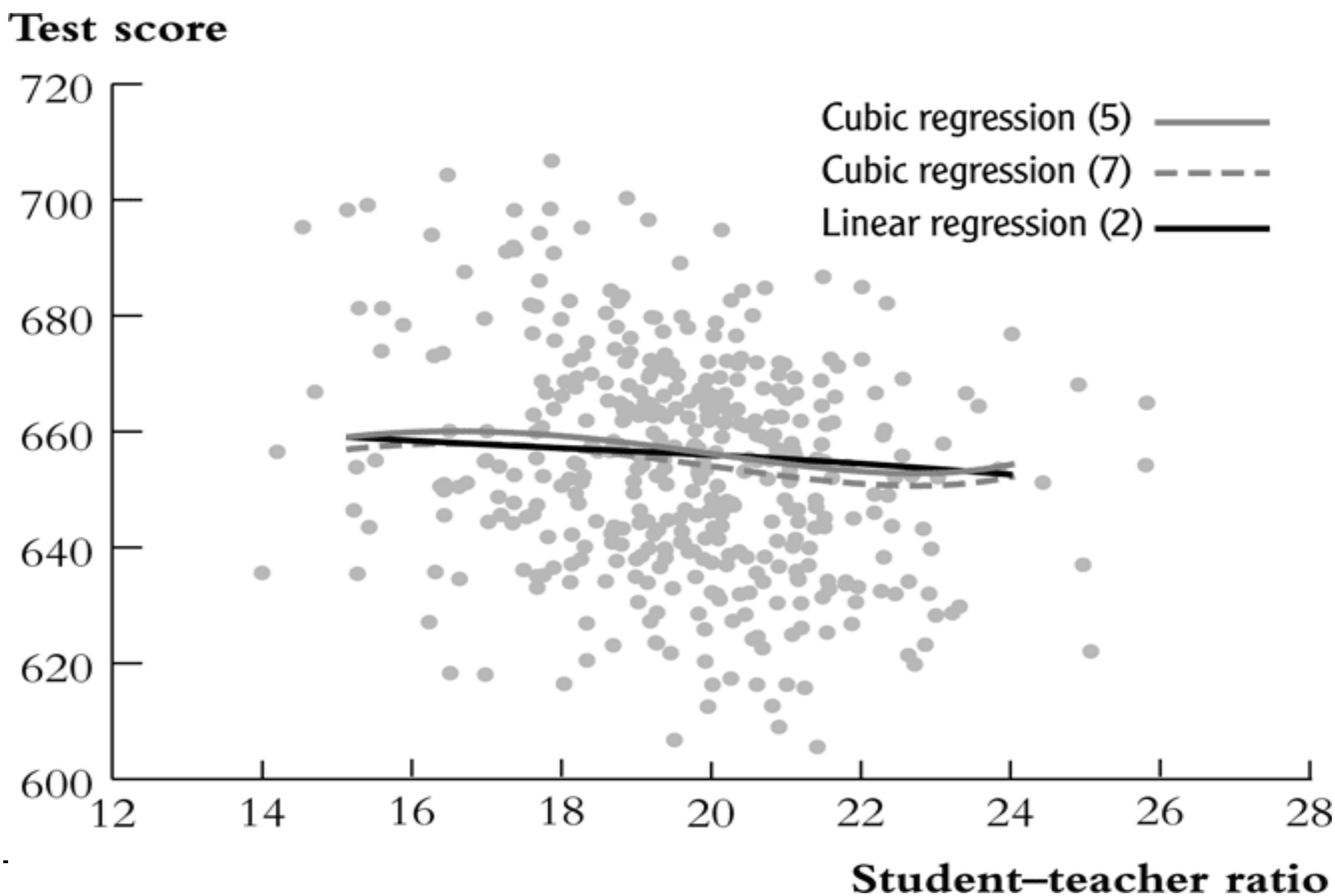
(a) All STR variables and interactions = 0			5.64 (0.004)	5.92 (0.003)	6.31 (< 0.001)	4.96 (< 0.001)	5.91 (0.001)
(b) $STR^2, STR^3 = 0$					6.17 (< 0.001)	5.81 (0.003)	5.96 (0.003)
(c) $HiEL \times STR, HiEL \times STR^2, HiEL \times STR^3 = 0$						2.69 (0.046)	
SER	9.08	8.64	15.88	8.63	8.56	8.55	8.57
$\overline{R}^2$	0.773	0.794	0.305	0.795	0.798	0.799	0.798

These regressions were estimated using the data on K–8 school districts in California, described in Appendix 4.1. Standard errors are given in parentheses under coefficients, and p-values are given in parentheses under F-statistics. Individual coefficients are statistically significant at the *5% or **1% significance level.

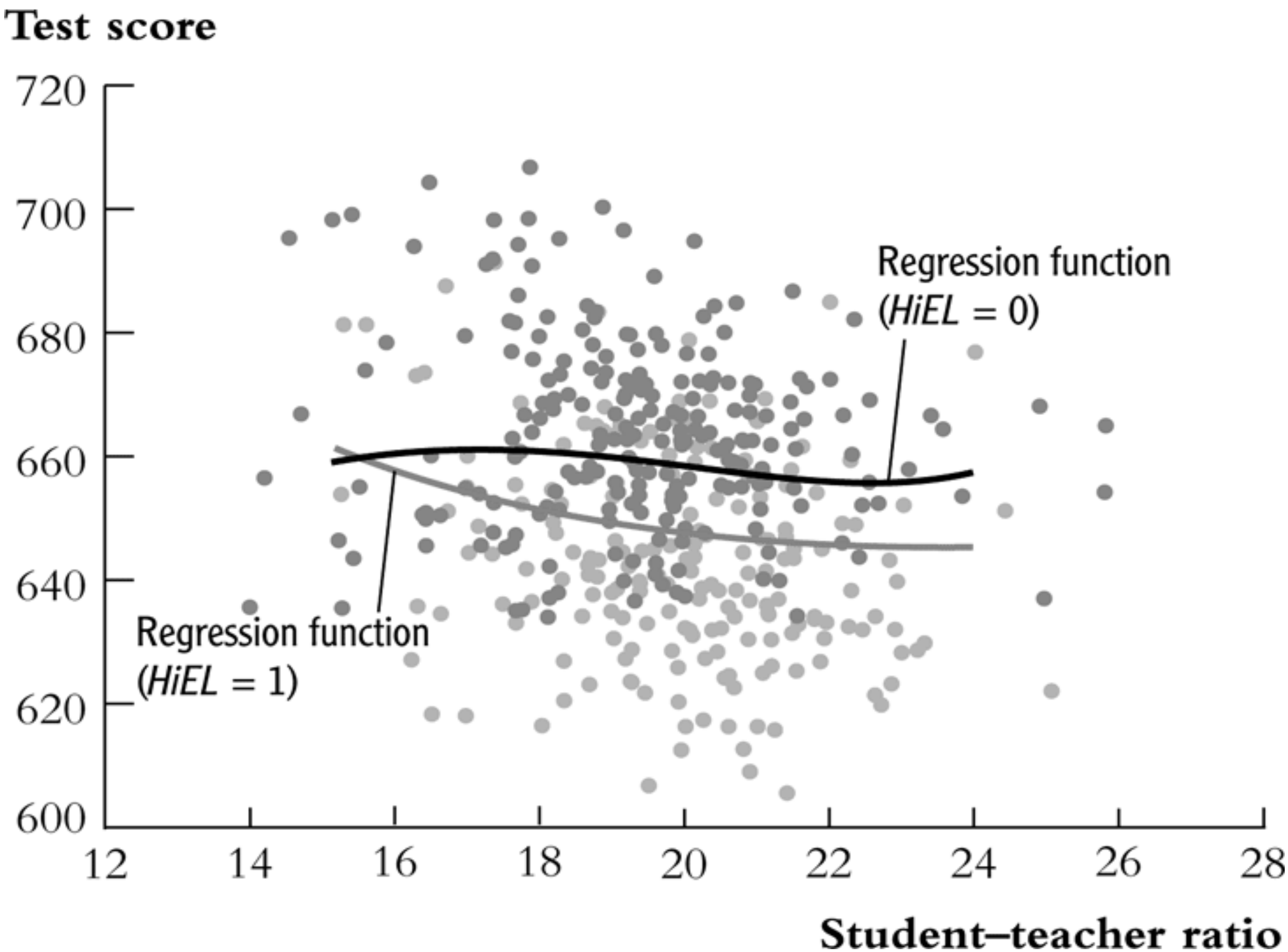
- ***What can you conclude about question #1?***
- ***About question #2?***

the regression functions via plots:

First, compare the linear and nonlinear specifications:



Next, compare the regressions with interactions:



Summary: Nonlinear Regression Functions

- Using functions of the independent variables such as ln(*X*) or $X_1 \times X_2$, allows recasting a large family of nonlinear regression functions as multiple regression.
- Estimation and inference proceed in the same way as in the linear multiple regression model.
- Interpretation of the coefficients is model-specific, but the general rule is to compute effects by comparing different cases (different value of the original *X*'s)
- Many nonlinear specifications are possible, so you must use judgment:
  – What nonlinear effect you want to analyze?
  – What makes sense in your application?