Econometrics

Practical Session 11

Non-linear Specifications

Ricardo Gouveia-Mendes
rgouveiamendes@ucp.pt

Spring 2025-26

Católica-Lisbon School of Business and Economics

Theoretical Wrap-up

Can OLS be applied to nonlinear specifications?

- **No and... yes**
  - The model must still be **linear in the parameters**
  - But we may **apply transformations to the variables** that can make the model **nonlinear in the data space**
  - Example with a classic Cobb-Douglas production function:

$$Y_i = AK_i^{\alpha} L_i^{1-\alpha} \quad \rightarrow \quad \underbrace{\ln\left(\frac{Y_i}{L_i}\right)}_{Y_i^*} = \overbrace{\ln(A)}^{\beta_0} + \overbrace{\alpha}^{\beta_1} \underbrace{\ln\left(\frac{K_i}{L_i}\right)}_{X_i^*}$$

Can OLS be applied to nonlinear specifications?

- **Polynomials**:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{1i}^2 + \ldots + \beta_k X_{ki} + u_i \quad \rightarrow \quad \frac{\partial Y}{\partial X_1} = \beta_1 + 2 \cdot \beta_2 X_1$$

- **linear-log**:

$$Y_i = \beta_0 + \beta_1 \log(X_{1i}) + \ldots + \beta_k X_{ki} + u_i \quad \rightarrow \quad \frac{\partial Y}{\partial X_1} = \frac{\partial Y}{\partial \log(X_1)} \frac{\mathrm{d} \log(X_1)}{\mathrm{d} X_1} = \frac{\beta_1}{X_1}$$

which also means that:

$$\beta_1 = \frac{\partial Y}{\partial X_1 / X_1} \Leftrightarrow \frac{\beta_1}{100} = \frac{\partial Y}{100 \cdot \partial X_1 / X_1} \Leftrightarrow \frac{\beta_1}{100} = \frac{\partial Y}{\Delta\% \, X_1}$$

Can OLS be applied to nonlinear specifications?

- **log-linear** $\longrightarrow$ semi-elasticities:

$$\log(Y_i) = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki} + u_i \quad \rightarrow \quad \frac{\partial \log(Y)}{\partial X_1} = \beta_1 \Leftrightarrow 100 \cdot \beta_1 \approx \frac{\Delta\% \, Y}{\partial X_1}$$

- **log-log** $\longrightarrow$ elasticities:

$$\log(Y_i) = \beta_0 + \beta_1 \log(X_{1i}) + \ldots + \beta_k \log(X_{ki}) + u_i \quad \rightarrow \quad \frac{\partial \log(Y)}{\partial \log(X_1)} = \beta_1 \approx \frac{\Delta\% \, Y}{\Delta\% \, X_1}$$

- **Interaction variables**:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 (X_{1i} \cdot D) + \ldots + \beta_k X_i + u_i \quad \rightarrow \quad \begin{cases} D = 0 \Rightarrow \partial Y / \partial X_1 = \beta_1 \\ D = 1 \Rightarrow \partial Y / \partial X_1 = \beta_1 + \beta_2 \end{cases}$$

Exercises

Stock & Watson – Exercise 8.3

After reading this chapter's analysis of test scores and class size, an educator comments:

> *In my experience, student performance depends on class size, but not in the way your regressions say. Rather, students do well when class size is less than 20 students and do very poorly when class size is greater than 25. There are no gains from reducing class size below 20 students, the relationship is constant in the intermediate region between 20 and 25 students, and there is no loss to increasing class size when it is already greater than 25.*

Stock & Watson – Exercise 8.3

The educator is describing a threshold effect, in which performance is constant for class sizes less than 20, jumps and is constant for class sizes between 20 and 25, and then jumps again for class sizes greater than 25. To model these threshold effects, define the binary variables:

- $STRsmall = 1$ if $STR < 20$, and $STRsmall = 0$ otherwise;
- $STRmoderate = 1$ if $20 \leq STR \leq 25$, and $STRmoderate = 0$ otherwise;
- $STRlarge = 1$ if $STR > 25$, and $STRlarge = 0$ otherwise.

Stock & Watson – Exercise 8.3

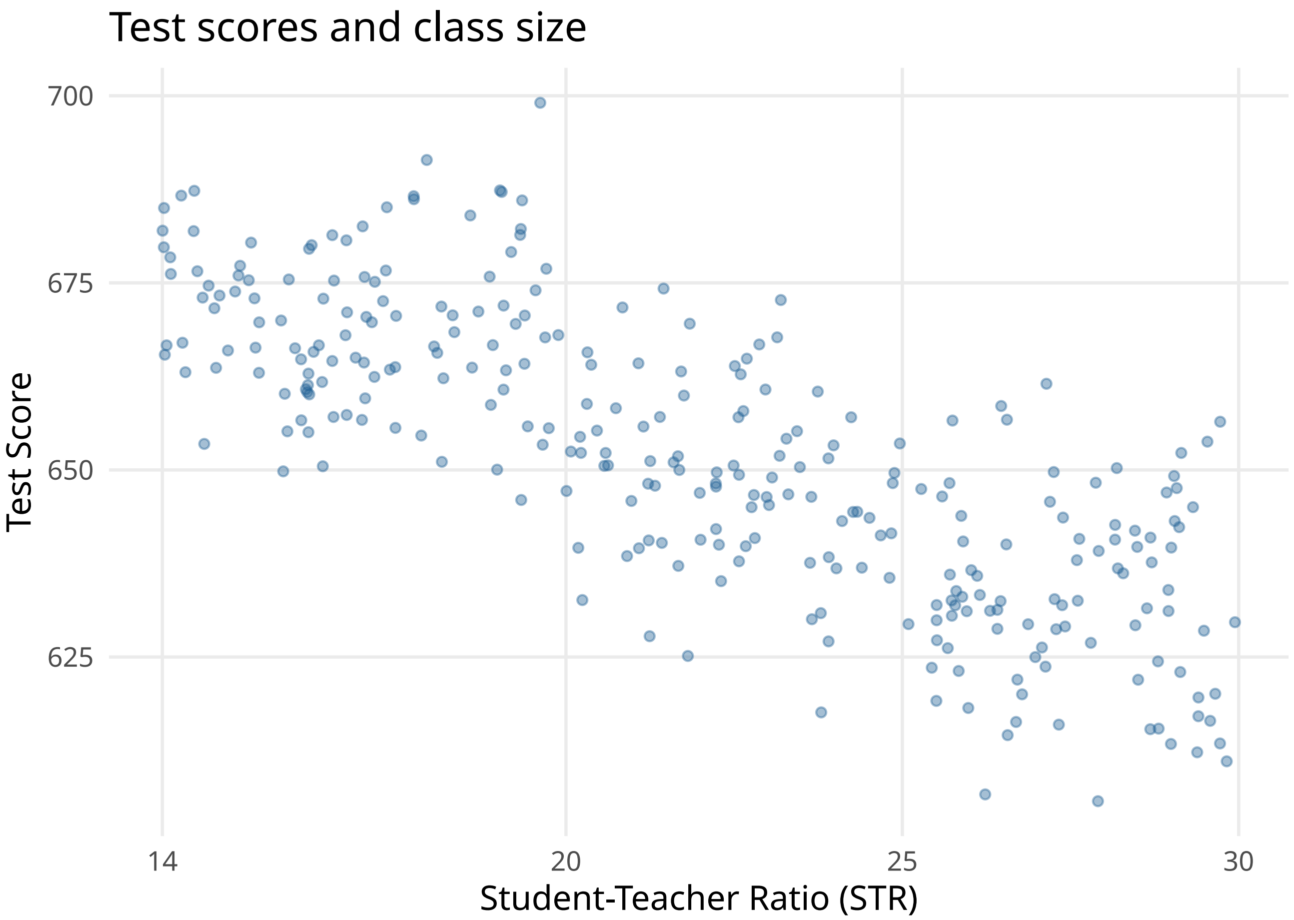
a. Consider the regression

$$\text{TestScore}_i = \beta 0 + \beta 1\,\text{STRsmall}_i + \beta 2\,\text{STRlarge}_i + u_i.$$

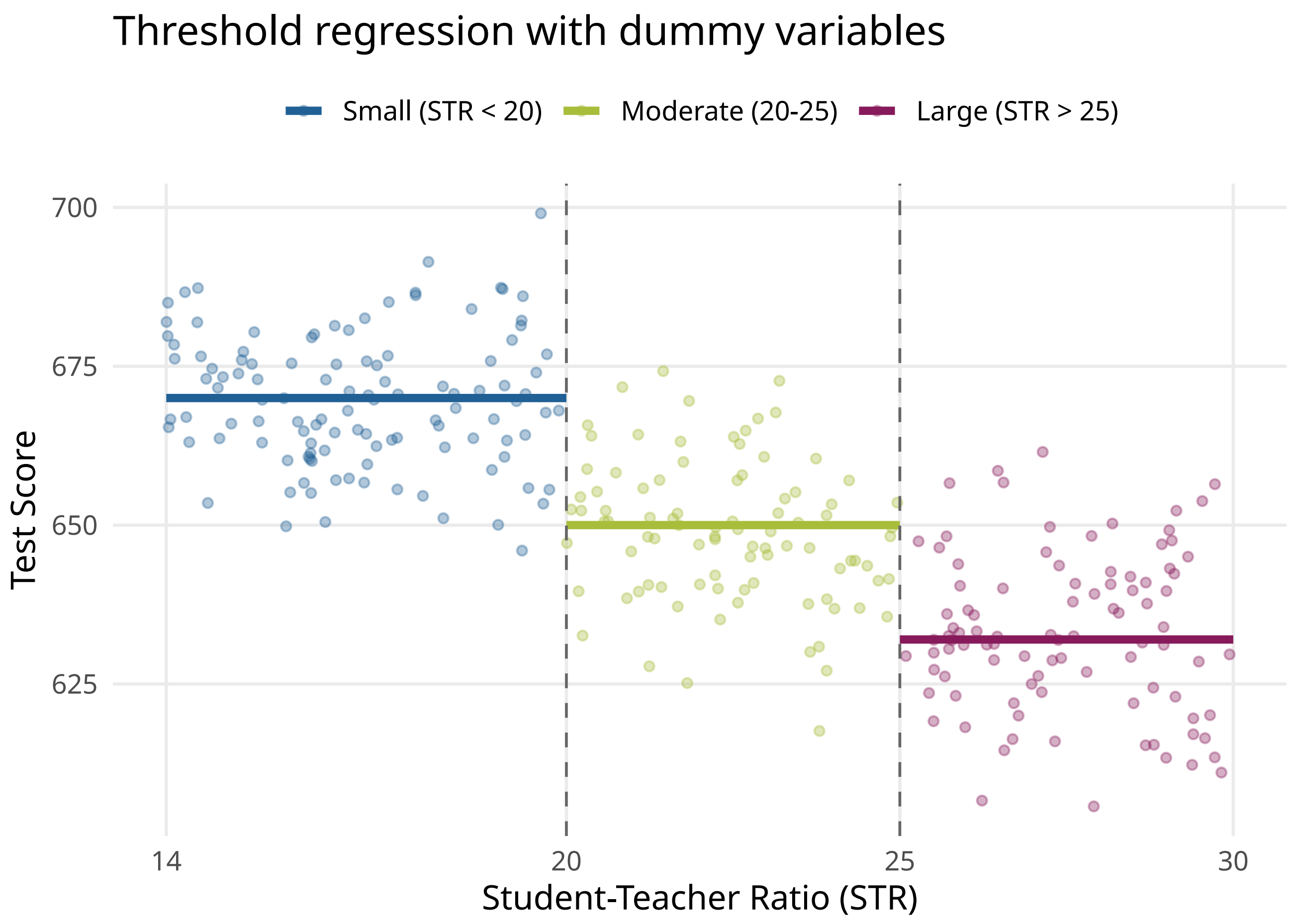
Sketch the regression function relating *TestScore* to *STR* for hypothetical values of the regression coefficients that are consistent with the educator's statement.

> - β_0 is the average test score for the intermediate class size, while β_1 and β_2 are the average difference in test score for small and large class sizes relative to the intermediate class size
> - The educator's claim is: $\beta_1 > 0$ and $\beta_2 < 0$
> - Graphically, the regression function is a step function

Stock & Watson – Exercise 8.3



Stock & Watson – Exercise 8.3



Stock & Watson – Exercise 8.3

b. A researcher tries to estimate the regression

$$TestScore_i = \beta_0 + \beta_1\, STRsmall_i + \beta_2\, STRmoderate_i + \beta_3\, STRlarge_i + u_i$$

and finds that the software gives an error message. Why?

- The error message indicates a problem of multicollinearity
- With three dummy variables mutually exclusive and exhaustive for every observation, we can only use two of them in the regression

Stock & Watson – Exercise 8.6

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)	

Stock & Watson – Exercise 8.6

Regressor	(1)	(2)	(3)	(4)	(5)	(6)	(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)	
Included Economic Control Variables							
%Eligible for subsidized lunch	Y	Y	N	Y	Y	Y	Y
Average district income (log)	N	Y	N	Y	Y	Y	Y
SER	9.08	8.64	15.88	8.63	8.56	8.55	8.57
$\bar{R}^2$	0.773	0.794	0.305	0.795	0.798	0.799	0.798

Stock & Watson – Exercise 8.6

TABLE 8.3 *(continued)* **95% Confidence Intervals for the Effect of Reducing STR by 2**

Regressor	(1)	(2)	(3)	(4)	(5)	(6)	(7)
No *HiEL* interaction	[0.93, 3.06]	[0.46, 2.48]					
22 to 20					[0.61, 3.25]		[0.54, 3.26]
20 to 18					[1.64, 4.36]		[1.55, 4.30]
HiEL = 0			[−0.38, 4.25]	[−0.28, 2.41]			
22 to 20						[0.40, 3.98]	
20 to 18						[1.22, 4.99]	
HiEL = 1			[1.48, 7.50]	[0.80, 3.63]			
22 to 20						[−0.98, 2.91]	
20 to 18						[−0.72, 4.01]	

Stock & Watson – Exercise 8.6

TABLE 8.3 *(continued)* **F-Statistics and p-Values on Joint Hypotheses**

Regressor	(1)	(2)	(3)	(4)	(5)	(6)	(7)
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)
STR^2, STR^3 = 0					6.17 (<0.001)	5.81 (0.003)	5.96 (0.003)
$HiEL \times STR$, $HiEL \times STR^2$, $HiEL \times STR^3$ = 0						2.69 (0.046)	

Stock & Watson – Exercise 8.6

Refer to Table 8.3.

a. A researcher suspects that the effect of *%Eligible for subsidized lunch* has a nonlinear effect on test scores. In particular, he conjectures that increases in this variable from 10% to 20% have little effect on test scores but that changes from 50% to 60% have a much larger effect.

Stock & Watson – Exercise 8.6

i. Describe a nonlinear specification that can be used to model this form of nonlinearity.

- There are several ways to model this form of nonlinearity
- One way is to use a second order polynomial term for *%Eligible for subsidized lunch*

$$\text{TestScore}_i = \boldsymbol{\beta}'\boldsymbol{X}_i + \gamma_1\,\%\text{ESL}_i + \gamma_2\,\%\text{ESL}_i^2 + u_i$$

- Another way is to use dummy variables to code the three levels of *%Eligible for subsidized lunch*

$$\text{TestScore}_i = \boldsymbol{\beta}'\boldsymbol{X}_i + \gamma_1\,\%\text{ESL}_i + \gamma_2\left(\%\text{ESL}_i^{\text{low}} \cdot \%\text{ESL}_i\right) + \gamma_3\left(\%\text{ESL}_i^{\text{high}} \cdot \%\text{ESL}_i\right) + u_i$$

Stock & Watson – Exercise 8.6

ii. How would you test whether the researcher's conjecture was better than the linear specification in column (7) of Table 8.3?

- If we were to use a second-order polynomial term for *%Eligible for subsidized lunch*, we would run a t-test on the coefficient of the square term, γ_2, to see if it is significantly different from zero
- If using the dummy variables approach, we would run an F-test on the two coefficient, γ_2 and γ_3, to see if they are jointly significantly different from zero

Stock & Watson – Exercise 8.6

b. A researcher suspects that the effect of income on test scores is different in districts with small classes than in districts with large classes.

i. Describe a nonlinear specification that can be used to model this form of nonlinearity.

> $\text{TestScore}_i = \boldsymbol{\beta}'\boldsymbol{X}_i + \gamma_1 \log(\text{AvgDistrictIncome}_i) + \gamma_2 (\log(\text{AvgDistrictIncome}_i) \cdot \text{STR}) + u_i$

ii. How would you test whether the researcher's conjecture was better than the linear specification in column (7) of Table 8.3?

> We would perform a t-test on the coefficient of the interaction term, γ_2, to see if it is significantly different from zero

Extra Exercises

Refer to Table 8.3.

a. Using column (5), write the expression for the marginal effect of *STR* on test scores. Evaluate it at STR = 20.

- The marginal effect is:

$$\frac{\partial\ \text{TestScore}}{\partial\ \text{STR}} = \hat{\beta}_1 + 2\hat{\beta}_2\ \text{STR} + 3\hat{\beta}_3\text{STR}^2 = 64.33 - 6.84\ \text{STR} + 0.177\ \text{STR}^2$$

- Evaluated at STR = 20:

$$64.33 - 6.84 \times 20 + 0.177 \times 400 = 64.33 - 136.80 + 70.80 = -1.67$$

- A one-unit increase in *STR* at STR = 20 is associated with a decrease of about 1.67 points in test scores

Extra Exercises

b. Using the F-statistics reported in Table 8.3, test at the 1% significance level whether the nonlinear terms STR^2 and STR^3 are jointly significant in column (5). What do you conclude about the appropriate functional form?

> - $H_0 : \beta_{STR^2} = \beta_{STR^3} = 0$ vs. H_1 : at least one $\neq 0$
> - $F = 6.17$, p-value $< 0.001 < 0.01$
> - We reject H_0 at the 1% level: the nonlinear terms are jointly significant
> - The cubic specification is preferred over the linear model

Extra Exercises

c. In column (6), test at the 5% and 1% significance levels whether the effect of *STR* on test scores differs between districts with high and low fractions of English learners. What do you conclude?

- $H_0 : \beta_{\mathbf{HiEL} \times \mathbf{STR}} = \beta_{\mathbf{HiEL} \times \mathbf{STR^2}} = \beta_{\mathbf{HiEL} \times \mathbf{STR^3}} = 0$
- $F = 2.69$, p-value $= 0.046$
- At 5%: $p < 0.05 \longrightarrow$ reject H_0 — the regression function differs between high and low *HiEL* districts
- At 1%: $p > 0.01 \longrightarrow$ fail to reject H_0
- Conclusion: tentative evidence of a difference — significant at 5% but not at 1%

Extra Exercises

d. A colleague prefers the simpler linear model in column (2) over the cubic in column (5), arguing that the gain in $\bar{R}^2$ is negligible (0.794 vs. 0.798). What do you think of this argument?

- The difference in $\bar{R}^2$ is indeed tiny
- But the F-test for $\text{STR}^2, \text{STR}^3 = 0$ gives $F = 6.17$ ($p < 0.001$): we strongly reject the hypothesis that the nonlinear terms are zero
- So the colleague is using the right intuition (parsimony) but the wrong tool ($\bar{R}^2$ instead of an F-test)
- Lesson: $\bar{R}^2$ can fail to detect significant nonlinearities when the linear and cubic fits are close over most of the data range — the F-test is the appropriate tool for comparing nested models

Extra Exercises

e. The superintendent proposes reducing *STR* by 2. Using the confidence intervals in Table 8.3 for column (7), assess whether this policy is likely to improve test scores, both for a reduction from 22 to 20 and from 20 to 18.

> - Reduction from 22 to 20: 95% CI = $[0.54, 3.26]$ — excludes 0, statistically significant positive effect
> - Reduction from 20 to 18: 95% CI = $[1.55, 4.30]$ — also excludes 0, larger and more precise effect
> - The policy is estimated to improve test scores regardless of the starting class size
> - The gain is larger when reducing from an already moderate class size (20 to 18) than from a larger one (22 to 20)