Lecture 5: Nonlinear Regression Functions

Introduction to Econometrics,Spring 2023

Zhaopeng Qu

Business School,Nanjing University

March 29 2023

Review of previous lecture

OLS Regression and Hypothsis Testing

- OLS is the most basic and important tool in econometricians' toolbox.
- The OLS estimators is **unbiased**, **consistent** and **normal distributions** under *key assumptions*.
- Using the hypothesis testing and confidence interval in OLS regression, we could make a more reliable judgment about the relationship between the treatment and the outcomes.

Nonlinear Regression Functions:

Introduction

- Recall the assumption of Linear Regression Model

Linear Regression Model

The observations, (Y_i, X_i) come from a random sample(i.i.d) and satisfy the linear regression equation,

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \ldots + \beta_k X_{k,i} + u_i$$

- Everything what we have learned so far is under this assumption of linearity. But this linear approximation is not always a good one.

Introduction: Recall the whole picture what we want to do

- A general formula for a population regression model may be

$$Y_i = f(X_{1,i}, X_{2,i}, \ldots, X_{k,i}) + u_i$$

- **Parametric methods**: assume that the function form(families) is known, we just need to assure(estimate) some unknown parameters in the function.
  - **Linear**
  - **Nonlinear**
- **Nonparametric methods**: assume that the function form is unknown or unnecessary to known.

Nonlinear Regression Functions

- How to extend linear OLS model to be nonlinear?

1. **Nonlinear in Xs**(the lecture now)
    - **Polynomials,Logarithms and Interactions**
    - The multiple regression framework can be extended to handle regression functions that are nonlinear in one or more X.
    - the difference from a standard multiple OLS regression is *how to explain estimating coefficients*.
2. **Nonlinear in** β or **Nonlinear in Y**(the next lecture)
    - **Discrete Dependent Variables** or **Limited Dependent Variables**.
    - Linear function in Xs is not a good prediciton function or Y.
    - Need a function which parameters enter nonlinearly, such as logisitic or negative exponential functions.
    - Then the parameters can not obtained by OLS estimation any more but *Nonlinear Least Squres* or Maximum Likelyhood Estimation.

Marginal Effect of X in Nonlinear Regression

- If our regression model is linear: $Y_i = \beta_0 + \beta_1 X_{1,i} + ... + \beta_k X_{k,i} + u_i$
  - Then the **marginal effect** of X, thus *the effect of Y on a change in X_j by 1 (unit)* is **constant** and equals β_j:

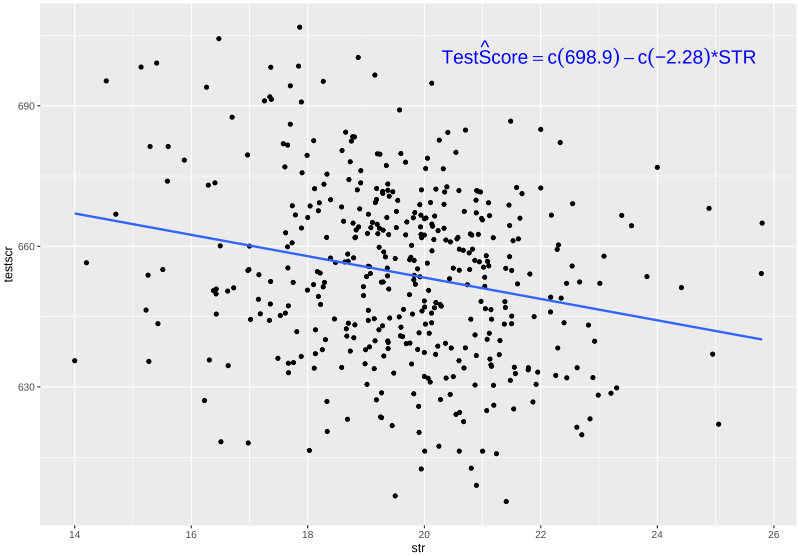
$$\beta_j = \frac{\partial Y_i}{\partial X_{ji}}$$

- But if a relation between Y and X is **nonlinear**, thus $Y_i = f(X_{1,i}, X_{2,i}, ..., X_{k,i}) + u_i$
  - Then the marginal effect of X is not constant, but depends on the value of Xs(including X_i itself or/and other X_js) because

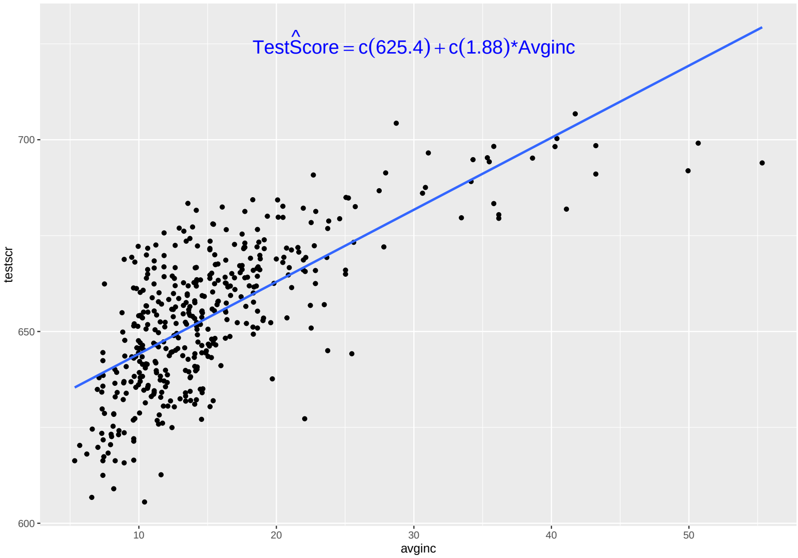
$$\frac{\partial Y_i}{\partial X_{ji}} = \frac{\partial f(X_{1,i}, X_{2,i}, ..., X_{k,i})}{\partial X_{ji}}$$

Nonlinear in Xs

The TestScore – STR relation looks linear (maybe)



But the TestScore – Income relation looks nonlinear

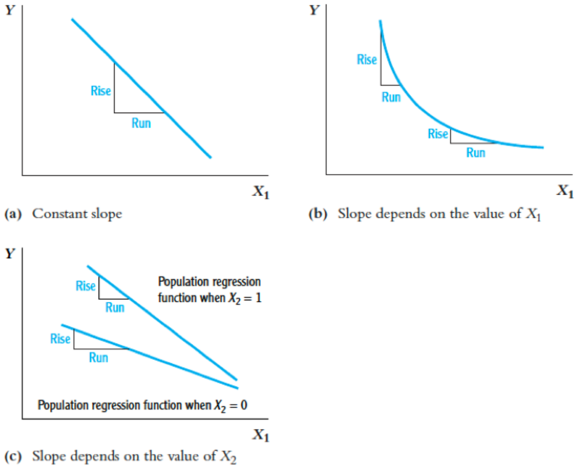


Three Complementary Approaches:

1. **Polynomials in X**
   - The population regression function is approximated by a quadratic, cubic, or higher-degree polynomial.
2. **Logarithmic transformations**
   - Y and/or X is transformed by taking its logarithm
   - this gives a *percentages* interpretation that makes sense in many applications
3. **Interactions**
   - the effect X on Y depends on the value of another independent variable
   - very often used in the analysis of hetergenous effects, some time used as analysis(channel).

Population Regression Functions with Different Slopes

FIGURE 8.1 Population Regression Functions with Different Slopes



In Figure 8.1a, the population regression function has a constant slope. In Figure 8.1b, the slope of the population regression function depends on the value of X_1. In Figure 8.1c, the slope of the population regression function

The Effect on Y of a Change in X in Nonlinear Functions

The Expected Change on Y of a Change in X_1 in the Nonlinear Regression Model (8.3)

KEY CONCEPT 8.1

The expected change in Y, ΔY, associated with the change in X_1, ΔX_1, holding $X_2, \ldots, X_k$ constant, is the difference between the value of the population regression function before and after changing X_1, holding $X_2, \ldots, X_k$ constant. That is, the expected change in Y is the difference:

$$\Delta Y = f(X_1 + \Delta X_1, X_2, \ldots, X_k) - f(X_1, X_2, \ldots, X_k). \tag{8.4}$$

The estimator of this unknown population difference is the difference between the predicted values for these two cases. Let $\hat{f}(X_1, X_2, \ldots, X_k)$ be the predicted value of Y based on the estimator $\hat{f}$ of the population regression function. Then the predicted change in Y is

$$\Delta \hat{Y} = \hat{f}(X_1 + \Delta X_1, X_2, \ldots, X_k) - \hat{f}(X_1, X_2, \ldots, X_k). \tag{8.5}$$

Polynomials in X

Example: the TestScore-Income relation

- If a straight line is NOT an adequate description of the relationship between district income and test scores, what is?
- Two options
  - Quadratic specification:

$$TestScore_i = \beta_0 + \beta_1 Income_i + \beta_2 (Income_i)^2 + u_i$$

  - Cubic specification:

$$TestScore_i = \beta_0 + \beta_1 Income_i + \beta_2 (Income_i)^2 + \beta_3 (Income_i)^3 + u_i$$

- How to estimate these models?
  - We can see **quadratic** and **cubic** terms as two independent variables.
  - Then the model turns into a **special form** of a multiple OLS regression model.

Estimation of the quadratic specification in R

```
##
## Call:
##    felm(formula = testscr ~ avginc + I(avginc^2), data = ca)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -44.416  -9.048   0.440   8.348  31.639
##
## Coefficients:
##              Estimate Robust s.e t value Pr(>|t|)
## (Intercept) 607.30174    2.90175 209.288   <2e-16 ***
## avginc        3.85100    0.26809  14.364   <2e-16 ***
## I(avginc^2)  -0.04231    0.00478  -8.851   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.72 on 417 degrees of freedom
## Multiple R-squared(full model): 0.5562   Adjusted R-squared: 0.554
## Multiple R-squared(proj model): 0.5562   Adjusted R-squared: 0.554
## F-statistic(full model, *iid*):261.3 on 2 and 417 DF, p-value: < 2.2e-16
```

Estimation of the cubic specification in R

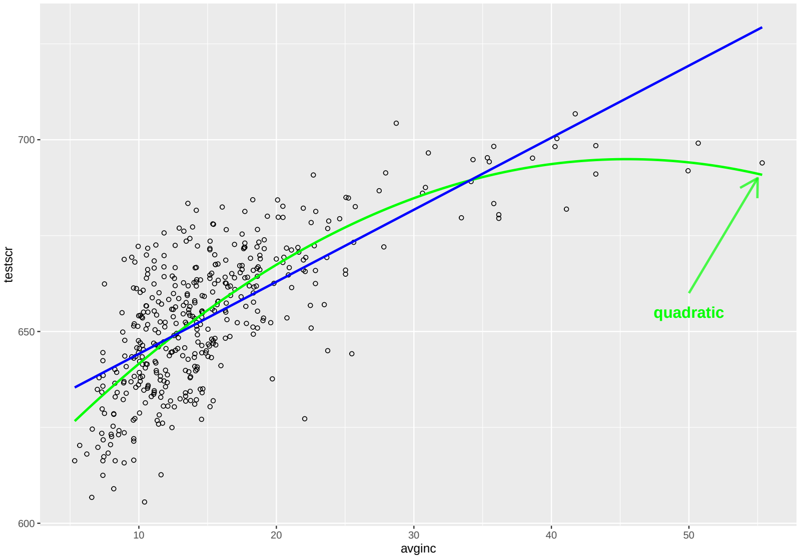
```
##
## Call:
##    felm(formula = testscr ~ avginc + I(avginc^2) + I(avginc3), data = ca)
##
## Residuals:
##    Min      1Q Median      3Q     Max
## -44.28  -9.21   0.20   8.32  31.16
##
## Coefficients:
##               Estimate Robust s.e t value Pr(>|t|)
## (Intercept)  6.001e+02  5.102e+00 117.615  < 2e-16 ***
## avginc       5.019e+00  7.074e-01   7.095 5.61e-12 ***
## I(avginc^2) -9.581e-02  2.895e-02  -3.309  0.00102 **
## I(avginc3)   6.855e-04  3.471e-04   1.975  0.04892 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.71 on 416 degrees of freedom
## Multiple R-squared(full model): 0.5584   Adjusted R-squared: 0.5552
## Multiple R-squared(proj model): 0.5584   Adjusted R-squared: 0.5552
```

TestScore and Income: OLS Regression Results

Table 1

	Dependent Variable: Test Score		
	(1)	(2)	(3)
avginc	1.879***	3.851***	5.019***
	(0.113)	(0.267)	(0.704)
I(avginc^2)		−0.042***	−0.096***
		(0.005)	(0.029)
I(avginc^3)			0.001**
			(0.0003)
Constant	625.384***	607.302***	600.079***
	(1.863)	(2.891)	(5.078)
Observations	420	420	420
Adjusted R^2	0.506	0.554	0.555
Residual Std. Error	13.387	12.724	12.707
F Statistic	430.830***	261.278***	175.352***

Figure: Linear and Quadratic Regression



Quadratic vs Linear

- **Question**: Is the quadratic model better than the linear model?
- We can test the null hypothesis that the regression function is linear against the alternative hypothesis that it is quadratic:

$$H_0 : \beta_2 = 0 \; and \; H_1 : \beta_2 \neq 0$$

- the t-statistic

$$t = \frac{(\hat{\beta}_2 - 0)}{SE(\hat{\beta}_2)} = \frac{-0.0423}{0.0048} = -8.81$$

- Since $8.81 > 2.58$, we reject the null hypothesis (the linear model) at a 1% significance level.
- Based on the F-test, we can also reject the null hypothesis

$$F - statistic_{q=2,d=417} = 261.3, p - value \cong 0.00$$

Interpreting the estimated quadratic regression function

- What is the **marginal effect** of X on Y in a quadratic regression function.
- The regression model now is

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i$$

- The marginal effect of X on Y

$$\frac{\partial Y_i}{\partial X_i} = \beta_1 + 2\beta_2 X_i$$

- It means that the **marginal effect** of X on Y depends on the specific value of X_i

Interpreting the estimated quadratic regression function

- The estimated regression function with a quadratic term of income is

$$\widehat{TestScore}_i = \underset{(2.90)}{607.3} + \underset{(0.27)}{3.85} \times income_i - \underset{(0.0048)}{0.0423} \times income_i^2.$$

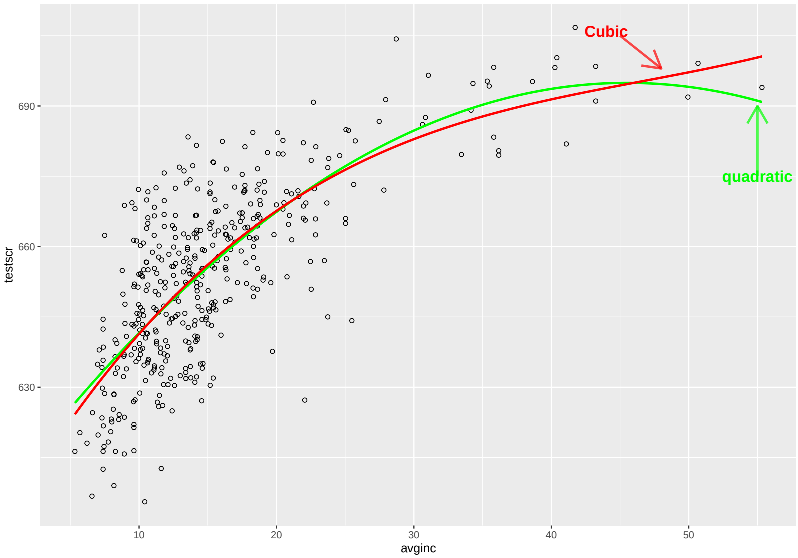
- Suppose *the effect of an $1000 increase on average income on test scores*
- A group: from $10,000 per capita to $11,000 per capita:

$$\begin{aligned}\Delta TestScore &= 607.3 + 3.85 \times 11 - 0.0423 \times (11)^2 \\ &- \left[607.3 + 3.85 \times 10 - 0.0423 \times (10)^2\right] \\ &= 2.96\end{aligned}$$

- B group: from $40,000 per capita to $41,000 per capita:

$$\begin{aligned}\Delta TestScore &= 607.3 + 3.85 \times 41 - 0.0423 \times (41)^2 \\ &- \left[607.3 + 3.85 \times 40 - 0.0423 \times (40)^2\right] \\ &= 0.42\end{aligned}$$

Figure: Cubic and Quadratic Regression



Quadratic vs Cubic

- **Question**: Is the cubic model better than the quadratic model?
- **Answer**: testing the null hypothesis that the regression function is *quadratic* against the alternative hypothesis that it is *cubic*:

$$H_0 : \beta_3 = 0 \; and \; H_1 : \beta_3 \neq 0$$

- the t-statistic

$$t = \frac{(\hat{\beta}_3 - 0)}{SE(\hat{\beta}_3)} = \frac{-0.001}{0.0003} = -3.33$$

- Since $3.33 > 2.58$, we reject the null hypothesis (the linear model) at a 1% significance level.
- the F-test also reject

$$F - statistic_{q=3,d=416} = 175.35, p - value \cong 0.00$$

Interpreting the estimated cubic regression function

- The regression model now is

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i$$

- The marginal effect of X on Y

$$\frac{\partial Y_i}{\partial X_i} = \beta_1 + 2\beta_2 X_i + 3\beta_3 X_i^2$$

Interpreting the estimated regression function

- The estimated cubic model is

$$\widehat{TestScore}_i = \underset{(5.83)}{600.1} + \underset{(0.86)}{5.02} \times income - \underset{(0.03)}{0.96} \times income^2 - \underset{(0.00047)}{0.00069} \times income^3.$$

- **A group**: from \$10,000 per capita to \$11,000 per capita:

$$\begin{aligned}\Delta TestScore &= 600.079 + 5.019 \times 11 - 0.96 \times (11)^2 + 0.001 \times (11)^3 \\ &\quad - [600.079 + 5.019 \times 10 - 0.96 \times (10)^2 + 0.001 \times (10)^3]\end{aligned}$$

- **B group**: from \$40,000 per capita to \$41,000 per capita:

$$\begin{aligned}\Delta TestScore &= 600.079 + 5.019 \times 41 - 0.96 \times (41)^2 + 0.001 \times (41)^3 \\ &\quad - [600.079 + 5.019 \times 40 - 0.96 \times (40)^2 + 0.001 \times (40)^3]\end{aligned}$$

Polynomials in X Regression Function

- Approximate the population regression function by a polynomial:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X^2 \ldots + \beta_r X_i^r + u_i$$

- This is just the multiple linear regression model – except that the regressors are **powers of X**!
- Estimation, hypothesis testing, etc. proceeds as in the multiple regression model using OLS.
- Although, the coefficients are difficult to interpret, the regression function itself is interpretable.

Testing the population regression function is linear

- If the population regression function is linear, then the higher-degree terms should not enter the population regression function.
- To perform hypothesis test

$$H_0 : \beta_2 = 0, \beta_3 = 0, ..., \beta_r = 0 \; and \; H_1 : at \; least \; one \; \beta_j \neq 0$$

- Because H_0 is a **joint null hypothesis** with $q = r - 1$ restrictions on the coefficients, it can be tested using the F-statistic.

Which degree polynomial should I use?

- How many powers of X should be included in a polynomial regression?
- The answer balances a **trade-off** between flexibility and statistical precision. (many ML or non-parametric or semi-parametric methods work on this)
    - Increasing the degree r introduces more flexibility into the regression function and allows it to match more shapes; a polynomial of degree r can have up to r - 1 bends (that is, inflection points) in its graph.
    - But increasing r means adding more regressors, which can reduce the precision of the estimated coefficients.

Which degree polynomial should I use?

- A practical way: asking whether the coefficients in the regression associated with the largest values of r are **zero**. If so, then these terms can be dropped from the regression.
- This procedure, which is called *sequential hypothesis testing*
  1. Pick a maximum value of r and estimate the polynomial regression for that r.
  2. Use the t-statistic to test whether the coefficient on X^r,β_r is **ZERO**.
  3. If reject, then the degree is r; if not then test the whether the coefficient on X^{r-1},β_{r-1} is **ZERO**.
  4. ...continue this procedure until the coefficient on the highest power in your polynomial is statistically significant.

Which degree polynomial should I use?

- The **initial degree** r of the polynomial is still missing.
- In many applications involving economic data, the nonlinear functions are smooth, that is, they do not have sharp jumps, or “spikes.”
- If so, then it is appropriate to choose a small maximum degree for the polynomial, such as 2, 3, or 4.

Which degree polynomial should I use?

- There are also several formal testing to determine the degree.
    - The F-statistic approach
    - The Akaike Information Criterion(AIC)
    - The Bayes Information Criterion(BIC)
- We will introduce them later on.

Wrap Up

- The nonlinear functions in Polynomials in Xs are just a special form of Multiple OLS Regression.
- If the true relationship between X and Y is nonlinear in polynomials in Xs, then a fully linear regression is misspecified – the functional form is wrong.
- The estimator of the effect on Y of X is biased(a special case of OVB).
- Estimation, hypothesis testing, etc. proceeds as in the multiple regression model using OLS, which can also help us to tell the degrees of polynomial functions.
- The big difference is how to explained the estimate coefficients and make the predicted change in Y with a change in Xs.

Logarithms

Logarithmic functions of Y and/or X

- Another way to specify a nonlinear regression model is to use the natural logarithm of Y and/or X.
- $Ln(x)$ = the natural logarithm of x is the inverse function of the exponential function e^x, here $e = 2.71828$.

$$x = ln(e^x)$$

Review of the Basic Logarithmic functions

- If X and a are variables, then we have

$$
\begin{aligned}
ln(1/x) &= -ln(x) \\
ln(ax) &= ln(a) + ln(x) \\
ln(x/a) &= ln(x) - ln(a) \\
ln(x^a) &= aln(x)
\end{aligned}
$$

Logarithms and percentages

- Because

$$ln(x+\Delta x) - ln(x) = ln\left(\frac{x+\Delta x}{x}\right)$$
$$\cong \frac{\Delta x}{x} \ (when\ \frac{\Delta x}{x}\ is\ very\ small)$$

- For example:

$$ln(1+0.01) = ln(101) - ln(100) = 0.00995 \cong 0.01$$

- Thus,logarithmic transforms permit modeling relations in **percentage** terms (like elasticities), rather than linearly.

The three log regression specifications:

Case	Population regression function
I.linear-log	$Y_i = \beta_0 + \beta_1 ln(X_i) + u_i$
II.log-linear	$ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$
III.log-log	$ln(Y_i) = \beta_0 + \beta_1 ln(X_i) + u_i$

- The interpretation of the slope coefficient **differs** in each case.
- The interpretation is found by applying the general "before and after" rule: "figure out the change in Y for a given change in X."(Key Concept 8.1 in S.W.pp301)

I. Linear-log population regression function

- Regression Model:

$$Y_i = \beta_0 + \beta_1 ln(X_i) + u_i$$

- Change X ΔX:

$$\begin{aligned}\Delta Y &= [\beta_0 + \beta_1 ln(X + \Delta X)] - [\beta_0 + \beta_1 ln(X)] \\ &= \beta_1 [ln(X + \Delta X) - ln(X)] \\ &\cong \beta_1 \frac{\Delta X}{X}\end{aligned}$$

- Note $100 \times \frac{\Delta X}{X} =$ *percentage change in X*, and

$$\beta_1 \cong \frac{\Delta Y}{\frac{\Delta X}{X}}$$

- Interpretation of β_1: *a 1 percent increase in X (multiplying X by 1.01 or* $100 \times \frac{\Delta X}{X}$*) is associated with a* $0.01\beta_1$ *or* $\frac{\beta_1}{100}$ *change in Y.*

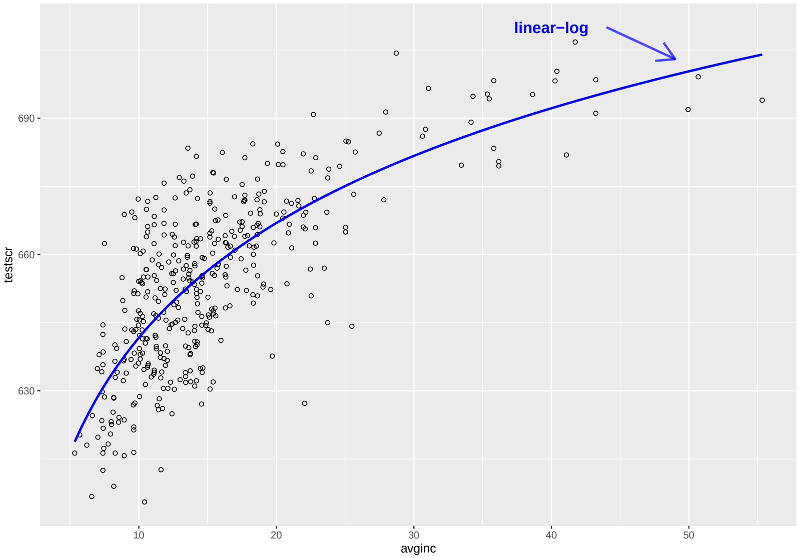
Example: the TestScore – log(Income) relation

- The OLS regression of ln(Income) on Testscore yields

$$\widehat{TestScore} = 557.8 + 36.42 \times ln(Income)$$
$$(3.8) \quad (1.4)$$

- Interpretation of β_1: a **1%** increase in Income is associated with an increase in TestScore of **0.3642** points on the test.

Test scores: linear-log function



Case II. Log-linear population regression function

- Regression model:

$$ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$$

- Change X:

$$ln(\Delta Y + Y) - ln(Y) = [\beta_0 + \beta_1(X + \Delta X)] - [\beta_0 + \beta_1 X]$$

$$ln(1 + \frac{\Delta Y}{Y}) = \beta_1 \Delta X$$

$$\Rightarrow \frac{\Delta Y}{Y} \cong \beta_1 \Delta X$$

- So $100\frac{\Delta Y}{Y} =$ *percentage change in Y* and

$$\beta_1 = \frac{\frac{\Delta Y}{Y}}{\Delta X}$$

- Then *a change in X by one unit is associated with a* $\beta_1 \times 100$ *percent change in Y*.

Mincer Earning Function: log-linear functions

- Example: Age(working experience) and Earnings
- The OLS regression of age on earnings yields

$$\widehat{ln(Earnings)} = \underset{(0.018)}{2.811} + \underset{(0.0004)}{0.0096} Age$$

- According to this regression, when one more year old, earnings are predicted to increase by $100 \times 0.0096 = 0.96\%$

Case III. Log-log population regression function

- the regression model is

$$ln(Y_i) = \beta_0 + \beta_1 ln(X_i) + u_i$$

- Change X:

$$\begin{aligned} ln(\Delta Y + Y) - ln(Y) &= [\beta_0 + \beta_1 ln(X + \Delta X)] - [\beta_0 + \beta_1 ln(X)] \\ ln(1 + \frac{\Delta Y}{Y}) &= \beta_1 ln(1 + \frac{\Delta X}{X}) \\ \Rightarrow \frac{\Delta Y}{Y} &\cong \beta_1 \frac{\Delta X}{X} \end{aligned}$$

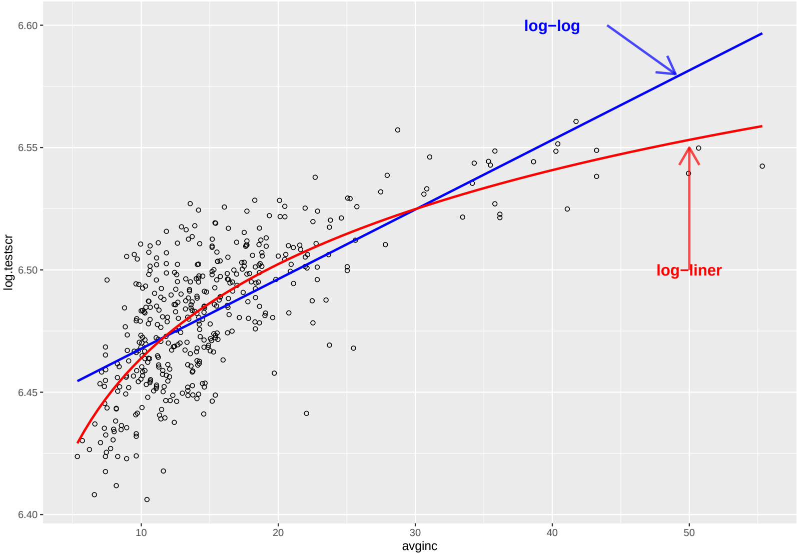
- Now $100\frac{\Delta Y}{Y} = \textit{percentage change in } Y$ and $100\frac{\Delta X}{X} = \textit{percentage change in } X$
- Therefore *a 1% change in X by one unit is associated with a* β_1*% change in Y*,thus β_1 has the interpretation of an **elasticity**.

Test scores and income: log-log specifications

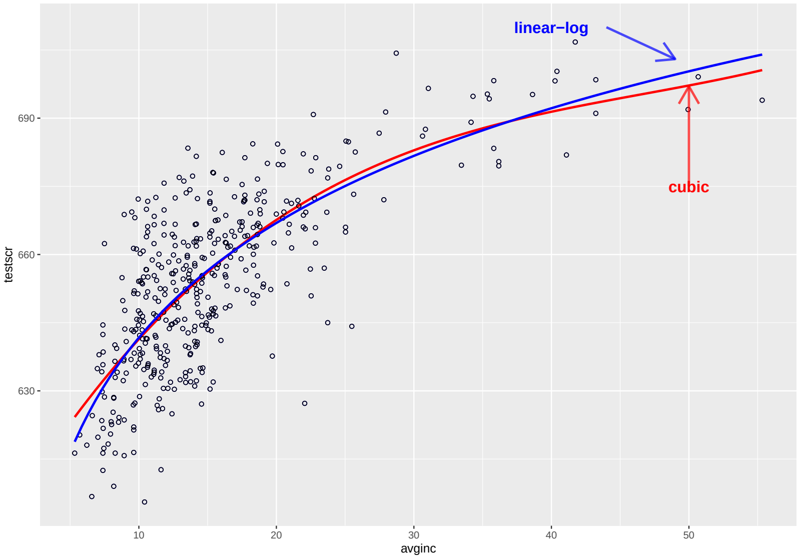
$$\widehat{ln(TestScore)} = 6.336 + 0.055 \times ln(Income)$$
$$(0.006) \quad (0.002)$$

- A 1% increase in Income is associated with an increase of 0.055% in TestScore.

Test scores: The log-linear and log-log functions



Test scores: The linear-log and cubic functions



Logarithmic and cubic functions

Table 3

	Dependent Variable: Test Score		
	testscr	log.testscr	testscr
	(1)	(2)	(3)
loginc	36.420***	0.055***	
		(0.002)	
avginc			5.019***
			(0.704)
I(avginc^2)			−0.096***
			(0.029)
I(avginc^3)			0.001**
			(0.0003)
Constant	557.832***	6.336***	600.079***
	(5.078)	(0.006)	(5.078)
Observations	420	420	420
Adjusted R^2	0.561	0.557	0.555

Choice of specification should be guided

- The two estimated regression functions are quite similar. So how to choose?
- The general rules:
  - By **economic logic or theories**(which interpretation makes the most sense in your application?).
  - There are several formal tests, while seldom used in reality. Actually t-test and F-test are enough.
  - Plotting predicted values and use $\overline{R^2}$ or *SER* can help to make further judgment.

Summary

- We can add polynomial terms of any significant variables to a model and to perform a single and joint test of significance. If the additional quadratics are significant, they can be added to the model.
- We can also change the variables values into logarithms to capture the nonlinear relationships.
- In reality, it can be difficult to pinpoint the precise reason for functional form misspecification.
- Fortunately, using **logarithms** of certain variables and adding **quadratic** or **cubic** functions are **sufficient** for detecting many(almost) important nonlinear relationships in Xs in economics.

Interactions Between Independent Variables

Introduction

- The product of two variables is called an **interaction term**.
- Try to answer *how the effect on Y of a change in an independent variable depends on the value of another independent variable.*
- Consider three cases:
    1. Interactions between **two binary** variables.
    2. Interactions between **a binary and a continuous variable**.
    3. Interactions between **two continuous variables**.

Interactions Between Two Binary Variables

- Assume we would like to study the earnings of worker in the labor market
- The population linear regression of Y_i is

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

- Dependent Variable: **log earnings**(Y_i,where $Y_i = ln(Earnings)$)
- Independent Variables: two binary variables
  - $D_{1i} = 1$ if the person graduate from college
  - $D_{2i} = 1$ if the worker's gender is female
- So β_1 is the effect on log earnings of having a college degree, *holding gender constant*, and β_2 is the effect of being female, *holding schooling constant*.

Interactions Between Two Binary Variables

- The effect of having a college degree in this specification, holding constant gender, is the **same** for men and women. No reason that this must be so.
- the effect on Y_i of D_{1i}, holding D_{2i} constant, could depend on the value of D_{2i}
- there could be an interaction between having a college degree and gender so that the value in the job market of a degree is different for men and women.
- The new regression model of Y_i is

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

- The new regressor, the product $D_{1i} \times D_{2i}$, is called an **interaction term** or an interacted regressor,

Interactions Between Two Binary Variables:

- The regression model of Y_i now is

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

- Then the *conditional expectation of Yi for* $D_{1i} = 0$, given a certain value of D_{2i},d_2

$$E(Y_i | D_{1i} = 0, D_{2i} = d_2) = \beta_0 + \beta_1 \times 0 + \beta_2 d_2 + \beta_3 (0 \times d_2) = \beta_0 + \beta_2 d_2$$

- Then the *conditional expectation of Yi for* $D_{1i} = 1$, given a certain value of D_{2i},d_2

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

Interactions Between Two Binary Variables:

- The effect of this change is the difference of expected values,which is

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

- In the binary variable interaction specification, the effect of acquiring a college degree (a unit change in D_{1i}) depends on the person's gender.
  - If the person is male,thus $D_{2i} = d_2 = 0$,then the effect is β_1
  - If the person is female,thus $D_{2i} = d_2 = 1$,then the effect is $\beta_1 + \beta_3$
- So the coefficient β_3 is **the difference in the effect of acquiring a college degree for women versus men.**

Application: the STR and the English learners

- Let $HiSTR_i$ be a binary variable for STR
  - $HiSTR_i = 1$ if the $STR > 20$
  - $HiSTR_i = 0$ otherwise
- Let $HiEL_i$ be a binary variable for the share of English learners
  - $HiEL_i = 1$ if the $el_{pct} > 10 percent$
  - $HiEL_i = 0$ otherwise

Application: the STR and the English learners

- the OLS regression result is

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

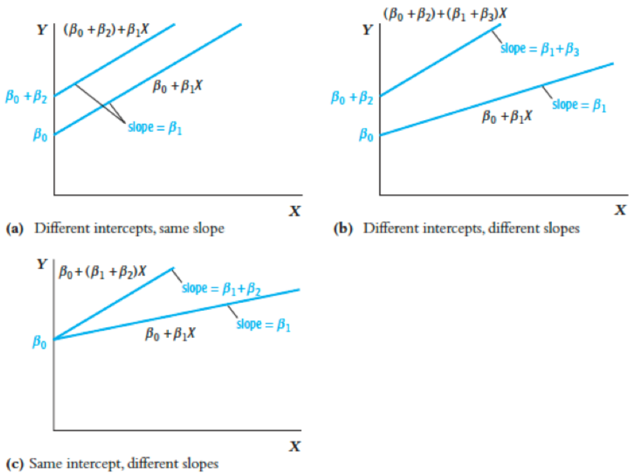
- The value of β_3 here(3.5) means that performance gap in test scores between large class($STR > 20$) and small class($STR \leq 20$) varies between the "higher-share-immigrant" class and the "lower-share immigrants" class.
- More precisely,the gap of test scores is positively related with the "higher-share-immigrant" class though insignificantly.

Interactions: a Continuous and a Binary Variable

- **Binary Variable**: eg, whether the worker has a college degree (D_i)
- **Continuous Variable**: eg, the individual's years of work experience (X_i)
- In this case, we can have three specifications:
  1. No interaction
  $$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + u_i$$
  2. a interaction and only one independent variable
  $$Y_i = \beta_0 + \beta_1 X_i + \beta_2 (D_i \times X_i) + u_i$$
  3. Interaction and two independent variables
  $$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 (D_i \times X_i) + u_i$$

A Continuous and a Binary Variable: Three Cases

FIGURE 8.8 Regression Functions Using Binary and Continuous Variables



Interactions of binary variables and continuous variables can produce three different population regression functions: (a) $\beta_0 + \beta_1 X + \beta_2 D$ allows for different intercepts but has the same slope, (b) $\beta_0 + \beta_1 X + \beta_2 D + \beta_3(X \times D)$ allows

A Continuous and a Binary Variable: Three Specifications

- All three specifications are just different versions of the multiple regression model.
- Different specifications are based on different assumptions of the relationships of X on Y depending on D.
- The **Model 3** is preferred, because it allows for both different intercepts and different slops.

Application: the STR and the English learners

- $HiEL_i$ is still a binary variable for English learner
- The estimated interaction regression

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

$$\overline{R^2} = 0.305$$

- For districts with a low fraction of English learners,the estimated regression line is $682.2 - 0.97STR_i$
- For districts with a high fraction of English learners,the estimated regression line is $682.2 + 5.6 - 0.97STR_i - 1.28STR_i = 687.8 - 2.25STR_i$
- The difference between these two effects, 1.28 points, is the coefficient on the interaction term.

Application: the STR and the English learners

- The value of β_3 here(-1.28) means that *the effect of class size on test scores varies between the "higher-share-immigrant" class and the "lower-share immigrants or more native" class.*
- More precisely,negatively related with the "higher-share-immigrant" class though insignificantly.

Hypotheses Testing

1. High fraction is the same as low fraction, thus the two line are in fact the same

- computing the **F-statistic** testing the joint hypothesis

$$\beta_2 = \beta_3 = 0$$

- This F-statistic is 89.9, which is significant at the 1% level.

2. The effects between two groups is the same,thus two lines have the same slope

- testing whether the coefficient on the interaction term is zero, which can be tested by using a **t-statistic**
- This t-statistic is -1.32, which is insignificant at the 10% level.

Hypotheses Testing

3. the lines have the same intercept

- Testing that the population coefficient on *HiEL* is zero,which can be tested by using a **t-statistic**.
- This t-statistic is 0.29, which is insignificant even at the 10% level.
- The reason is that the regressors, *HiEL* and $STR * HiEL$, are highly correlated. Then large standard errors on the individual coefficients.
- Even though it is impossible to tell which of the coefficients is nonzero, there is strong evidence against the hypothesis that both are zero.

Interactions Between Two Continuous Variables

- Now suppose that both independent variables (X_{1i} and X_{2i}) are continuous.
  - X_{1i} is his or her years of work experience
  - X_{2i} is the number of years he or she went to school.
- there might be an interaction between these two variables so that the effect on wages of an additional year of experience depends on the number of years of education.
- the population regression model

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

Interactions Between Two Continuous Variables

- Thus the effect on Y of a change in X_1, holding X_2 constant, is

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

- A similar calculation shows that the effect on Y of a change ΔX_1 in X_2, holding X_1 constant, is

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

- That is, if X_1 changes by ΔX_1 and X_2 changes by ΔX_2, then the expected change in Y

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

Application: the STR and the English learners

- The estimated interaction regression

$$
\widehat{ln(TestScore)} = \underset{(11.8)}{686.3} - \underset{(0.059)}{1.12}STR - \underset{(0.037)}{0.67}PctEL + \underset{(0.019)}{0.0012}(STR \times PctEL)
$$

- The value of β_3 here means *how the effect of class size on test scores varies along with the share of English learners in the class.*
- More precisely, *increase 1 unit of the share of English learners* make the effect of class size on test scores increase extra 0.0012 scores.

Application: the STR and the English learners

- when the percentage of English learners is at the **median**($PctEL = 8.85$), the slope of the line relating test scores and the STR is

$$\frac{\Delta Y}{\Delta X_1} = \beta_1 + \beta_3 X_2 = -1.12 + 0.0012 \times 8.85 = -1.11$$

- when the percentage of English learners is at the **75th percentile**($PctEL = 23.0$), the slope of the line relating test scores and the STR is

$$\frac{\Delta Y}{\Delta X_1} = \beta_1 + \beta_3 X_2 = -1.12 + 0.0012 \times 23.0 = -1.09$$

- The difference between these estimated effects is not statistically significant.Because?
  - The t-statistic testing whether the coefficient on the interaction term is zero $t = 0.0012/0/019 = 0.06$

Application: STR and Test Scores in a Summary

- Although these nonlinear specifications extend our knowledge about the relationship between STR and Testscore, it must be augmented with control variables such as **economic background** to avoid OVB bias.
- Two measures of the economic background of the students:
  1. the percentage of students eligible for a subsidized lunch
  2. the logarithm of average district income.

Application: STR and Test Scores in a Summary

- Then three specific questions about test scores and the student–teacher ratio.
  1. After controlling for differences in economic characteristics, does the effect on test scores of STR depend on the fraction of English learners?
  2. Does this effect depend on the value of the student–teacher ratio(STR)?
  3. Most important, after taking economic factors and nonlinearities into account,what is the estimated effect on test scores of reducing the student–teacher ratio by 2 students per teacher?

	score						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
str	-1.00^{***}	-0.73^{**}	-0.97	-0.53	64.34^{**}	83.70^{**}	65.29^{**}
	(0.27)	(0.26)	(0.59)	(0.34)	(24.86)	(28.50)	(25.26)
I(str^2)					-3.42^{**}	-4.38^{**}	-3.47^{**}
					(1.25)	(1.44)	(1.27)
I(str^3)					0.06^{**}	0.07^{**}	0.06^{**}
					(0.02)	(0.02)	(0.02)
str:HiEL			-1.28	-0.58		-123.28^{*}	
			(0.97)	(0.50)		(50.21)	
I(str^2):HiEL						6.12^{*}	
						(2.54)	
I(str^3):HiEL						-0.10^{*}	
						(0.04)	
english	-0.12^{***}	-0.18^{***}					-0.17^{***}
	(0.03)	(0.03)					(0.03)
HiEL			5.64	5.50	-5.47^{***}	816.08^{*}	
			(19.51)	(9.80)	(1.03)	(327.67)	
lunch	-0.55^{***}	-0.40^{***}		-0.41^{***}	-0.42^{***}	-0.42^{***}	-0.40^{***}
	(0.02)	(0.03)		(0.03)	(0.03)	(0.03)	(0.03)
log(income)		11.57^{***}		12.12^{***}	11.75^{***}	11.80^{***}	11.51^{***}
		(1.82)		(1.80)	(1.77)	(1.78)	(1.81)
Constant	700.15^{***}	658.55^{***}	682.25^{***}	653.67^{***}	252.05	122.35	244.81
	(5.57)	(8.64)	(11.87)	(9.87)	(163.63)	(185.52)	(165.72)
N	420	420	420	420	420	420	420
Adjusted R^2	0.77	0.79	0.31	0.79	0.80	0.80	0.80

$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Robust S.E. are shown in the parentheses

	score						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
str	-1.00^{***}	-0.73^{**}	-0.97	-0.53	64.34^{**}	83.70^{**}	65.29^{**}
	(0.27)	(0.26)	(0.59)	(0.34)	(24.86)	(28.50)	(25.26)
I(str^2)					-3.42^{**}	-4.38^{**}	-3.47^{**}
					(1.25)	(1.44)	(1.27)
I(str^3)					0.06^{**}	0.07^{**}	0.06^{**}
					(0.02)	(0.02)	(0.02)
str:HiEL			-1.28	-0.58		-123.28^{*}	
			(0.97)	(0.50)		(50.21)	
I(str^2):HiEL						6.12^{*}	
						(2.54)	
I(str^3):HiEL						-0.10^{*}	
						(0.04)	
english	-0.12^{***}	-0.18^{***}					-0.17^{***}
	(0.03)	(0.03)					(0.03)
HiEL			5.64	5.50	-5.47^{***}	816.08^{*}	
			(19.51)	(9.80)	(1.03)	(327.67)	
lunch	-0.55^{***}	-0.40^{***}		-0.41^{***}	-0.42^{***}	-0.42^{***}	-0.40^{***}
	(0.02)	(0.03)		(0.03)	(0.03)	(0.03)	(0.03)
log(income)		11.57^{***}		12.12^{***}	11.75^{***}	11.80^{***}	11.51^{***}
		(1.82)		(1.80)	(1.77)	(1.78)	(1.81)
Constant	700.15^{***}	658.55^{***}	682.25^{***}	653.67^{***}	252.05	122.35	244.81
	(5.57)	(8.64)	(11.87)	(9.87)	(163.63)	(185.52)	(165.72)
N	420	420	420	420	420	420	420
Adjusted R^2	0.77	0.79	0.31	0.79	0.80	0.80	0.80

$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Robust S.E. are shown in the parentheses

	score						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
str	−1.00***	−0.73**	−0.97	−0.53	64.34**	83.70**	65.29**
	(0.27)	(0.26)	(0.59)	(0.34)	(24.86)	(28.50)	(25.26)
I(str^2)					−3.42**	−4.38**	−3.47**
					(1.25)	(1.44)	(1.27)
I(str^3)					0.06**	0.07**	0.06**
					(0.02)	(0.02)	(0.02)
str:HiEL			−1.28	−0.58		−123.28*	
			(0.97)	(0.50)		(50.21)	
I(str^2):HiEL						6.12*	
						(2.54)	
I(str^3):HiEL						−0.10*	
						(0.04)	
english	−0.12***	−0.18***					−0.17***
	(0.03)	(0.03)					(0.03)
HiEL			5.64	5.50	−5.47***	816.08*	
			(19.51)	(9.80)	(1.03)	(327.67)	
lunch	−0.55***	−0.40***		−0.41***	−0.42***	−0.42***	−0.40***
	(0.02)	(0.03)		(0.03)	(0.03)	(0.03)	(0.03)
log(income)		11.57***		12.12***	11.75***	11.80***	11.51***
		(1.82)		(1.80)	(1.77)	(1.78)	(1.81)
Constant	700.15***	658.55***	682.25***	653.67***	252.05	122.35	244.81
	(5.57)	(8.64)	(11.87)	(9.87)	(163.63)	(185.52)	(165.72)
N	420	420	420	420	420	420	420
Adjusted R^2	0.77	0.79	0.31	0.79	0.80	0.80	0.80

$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Robust S.E. are shown in the parentheses

	score						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
str	-1.00^{***}	-0.73^{**}	-0.97	-0.53	64.34^{**}	83.70^{**}	65.29^{**}
	(0.27)	(0.26)	(0.59)	(0.34)	(24.86)	(28.50)	(25.26)
I(str^2)					-3.42^{**}	-4.38^{**}	-3.47^{**}
					(1.25)	(1.44)	(1.27)
I(str^3)					0.06^{**}	0.07^{**}	0.06^{**}
					(0.02)	(0.02)	(0.02)
str:HiEL			-1.28	-0.58		-123.28^{*}	
			(0.97)	(0.50)		(50.21)	
I(str^2):HiEL						6.12^{*}	
						(2.54)	
I(str^3):HiEL						-0.10^{*}	
						(0.04)	
english	-0.12^{***}	-0.18^{***}					-0.17^{***}
	(0.03)	(0.03)					(0.03)
HiEL			5.64	5.50	-5.47^{***}	816.08^{*}	
			(19.51)	(9.80)	(1.03)	(327.67)	
lunch	-0.55^{***}	-0.40^{***}		-0.41^{***}	-0.42^{***}	-0.42^{***}	-0.40^{***}
	(0.02)	(0.03)		(0.03)	(0.03)	(0.03)	(0.03)
log(income)		11.57^{***}		12.12^{***}	11.75^{***}	11.80^{***}	11.51^{***}
		(1.82)		(1.80)	(1.77)	(1.78)	(1.81)
Constant	700.15^{***}	658.55^{***}	682.25^{***}	653.67^{***}	252.05	122.35	244.81
	(5.57)	(8.64)	(11.87)	(9.87)	(163.63)	(185.52)	(165.72)
N	420	420	420	420	420	420	420
Adjusted R^2	0.77	0.79	0.31	0.79	0.80	0.80	0.80

$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$
Robust S.E. are shown in the parentheses

	score						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
str	−1.00***	−0.73**	−0.97	−0.53	64.34**	83.70**	65.29**
	(0.27)	(0.26)	(0.59)	(0.34)	(24.86)	(28.50)	(25.26)
I(str^2)					−3.42**	−4.38**	−3.47**
					(1.25)	(1.44)	(1.27)
I(str^3)					0.06**	0.07**	0.06**
					(0.02)	(0.02)	(0.02)
str:HiEL			−1.28	−0.58		−123.28*	
			(0.97)	(0.50)		(50.21)	
I(str^2):HiEL						6.12*	
						(2.54)	
I(str^3):HiEL						−0.10*	
						(0.04)	
english	−0.12***	−0.18***					−0.17***
	(0.03)	(0.03)					(0.03)
HiEL			5.64	5.50	−5.47***	816.08*	
			(19.51)	(9.80)	(1.03)	(327.67)	
lunch	−0.55***	−0.40***		−0.41***	−0.42***	−0.42***	−0.40***
	(0.02)	(0.03)		(0.03)	(0.03)	(0.03)	(0.03)
log(income)		11.57***		12.12***	11.75***	11.80***	11.51***
		(1.82)		(1.80)	(1.77)	(1.78)	(1.81)
Constant	700.15***	658.55***	682.25***	653.67***	252.05	122.35	244.81
	(5.57)	(8.64)	(11.87)	(9.87)	(163.63)	(185.52)	(165.72)
N	420	420	420	420	420	420	420
Adjusted R^2	0.77	0.79	0.31	0.79	0.80	0.80	0.80

$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Robust S.E. are shown in the parentheses

	score						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
str	−1.00***	−0.73**	−0.97	−0.53	64.34**	83.70**	65.29**
	(0.27)	(0.26)	(0.59)	(0.34)	(24.86)	(28.50)	(25.26)
I(str^2)					−3.42**	−4.38**	−3.47**
					(1.25)	(1.44)	(1.27)
I(str^3)					0.06**	0.07**	0.06**
					(0.02)	(0.02)	(0.02)
str:HiEL			−1.28	−0.58		−123.28*	
			(0.97)	(0.50)		(50.21)	
I(str^2):HiEL						6.12*	
						(2.54)	
I(str^3):HiEL						−0.10*	
						(0.04)	
english	−0.12***	−0.18***					−0.17***
	(0.03)	(0.03)					(0.03)
HiEL			5.64	5.50	−5.47***	816.08*	
			(19.51)	(9.80)	(1.03)	(327.67)	
lunch	−0.55***	−0.40***		−0.41***	−0.42***	−0.42***	−0.40***
	(0.02)	(0.03)		(0.03)	(0.03)	(0.03)	(0.03)
log(income)		11.57***		12.12***	11.75***	11.80***	11.51***
		(1.82)		(1.80)	(1.77)	(1.78)	(1.81)
Constant	700.15***	658.55***	682.25***	653.67***	252.05	122.35	244.81
	(5.57)	(8.64)	(11.87)	(9.87)	(163.63)	(185.52)	(165.72)
N	420	420	420	420	420	420	420
Adjusted R^2	0.77	0.79	0.31	0.79	0.80	0.80	0.80

$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

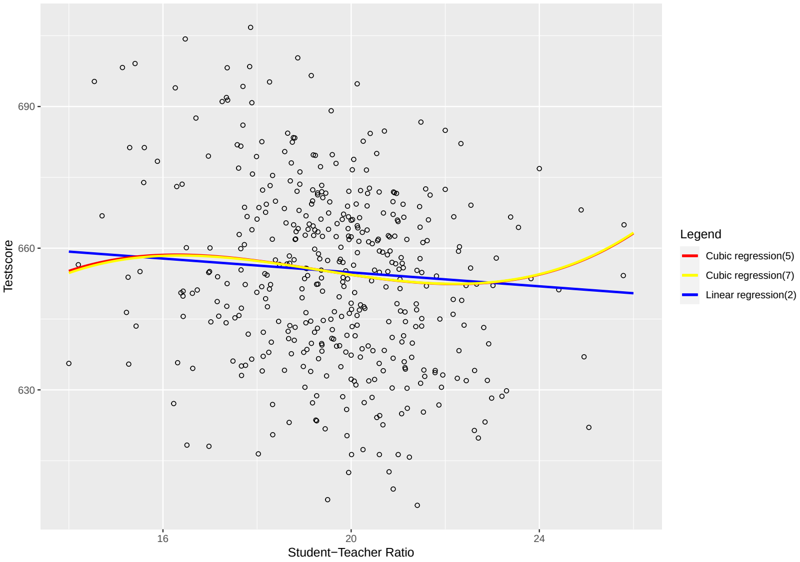
Robust S.E. are shown in the parentheses

	score						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
str	−1.00***	−0.73**	−0.97	−0.53	64.34**	83.70**	65.29**
	(0.27)	(0.26)	(0.59)	(0.34)	(24.86)	(28.50)	(25.26)
I(str^2)					−3.42**	−4.38**	−3.47**
					(1.25)	(1.44)	(1.27)
I(str^3)					0.06**	0.07**	0.06**
					(0.02)	(0.02)	(0.02)
str:HiEL			−1.28	−0.58		−123.28*	
			(0.97)	(0.50)		(50.21)	
I(str^2):HiEL						6.12*	
						(2.54)	
I(str^3):HiEL						−0.10*	
						(0.04)	
english	−0.12***	−0.18***					−0.17***
	(0.03)	(0.03)					(0.03)
HiEL			5.64	5.50	−5.47***	816.08*	
			(19.51)	(9.80)	(1.03)	(327.67)	
lunch	−0.55***	−0.40***		−0.41***	−0.42***	−0.42***	−0.40***
	(0.02)	(0.03)		(0.03)	(0.03)	(0.03)	(0.03)
log(income)		11.57***		12.12***	11.75***	11.80***	11.51***
		(1.82)		(1.80)	(1.77)	(1.78)	(1.81)
Constant	700.15***	658.55***	682.25***	653.67***	252.05	122.35	244.81
	(5.57)	(8.64)	(11.87)	(9.87)	(163.63)	(185.52)	(165.72)
N	420	420	420	420	420	420	420
Adjusted R^2	0.77	0.79	0.31	0.79	0.80	0.80	0.80

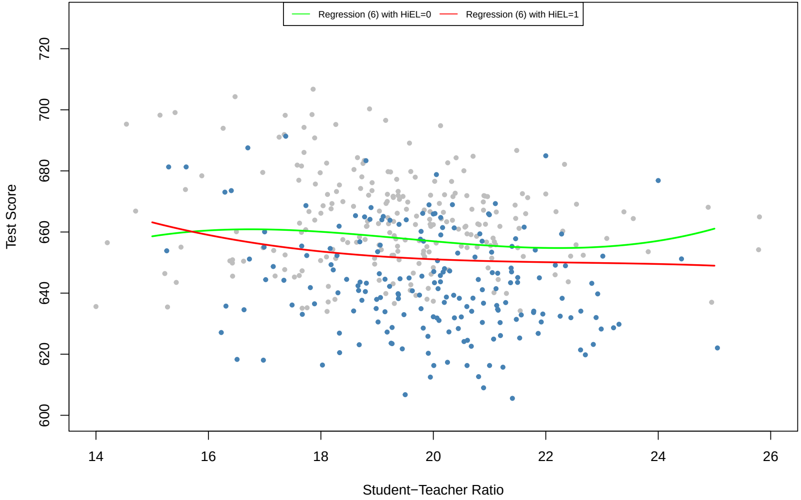
*p < .05; **p < .01; ***p < .001

Robust S.E. are shown in the parentheses

Three Regressions on graph



Interaction on graph



A Lastest and Smart Application: Jia and Ku(2019)

Jia and Ku(2019)

- Ruixue Jia and Hyejin Ku, "Is China's Pollution the Culprit for the Choking of South Korea?Evidence from the Asian Dust",The Economic Journal.
- **Main Question**: Whether the air pollution spillover from China to South Korea and affect the health of South Koreans?

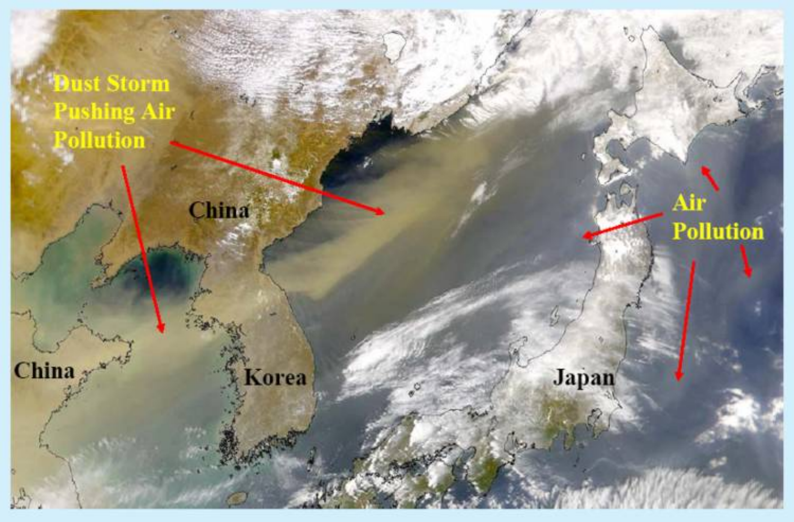
Empirical Strategy

- A naive strategy:
  - Dependent variable: **Deaths in South Korea**(respiratory and cardiovascular mortality)
  - Independt variable: **Chinese pollution**(Air Quality Index)
- Because the observed or measured air quality (i.e., pollution concentration) in Seoul or Tokyo increases in periods when China is more polluted does not mean that the pollution must have **originated from China**.

Jia and Ku(2019): Asian Dust as a carrier of pollutants

- **Asian Dust** (also yellow dust, yellow sand, yellow wind or China dust storms) is a **meteorological phenomenon** which affects much of East Asia year round but especially during the spring months.
  - The dust originates in China, the deserts of Mongolia, and Kazakhstan where high-speed surface winds and intense dust storms kick up dense clouds of fine, dry soil particles.
  - These clouds are then carried eastward by prevailing winds and pass over China, North and South Korea, and Japan, as well as parts of the Russian Far East.
  - In recent decades,Asian dust brings with it China's man-made pollution as well as its by-products.

Jia and Ku(2019): Asian Dust



Jia and Ku(2019): Asian Dust

1. A clear directional aspect in that the wind which transport Chinese pollutants to Korea but not vice versa.
2. Exogenous to South Korea's local activities. And wind patterns and topography generate rich spatial and temporal variation in the incidence.
3. The occurrence of Asian dust is monitored and recorded station by station in South Korea.(because of its visual salience)

Econometric Method: OLS Regressions with an interaction term

- **Dependent variable**: *Deaths in South Korea*(respiratory and cardiovascular mortality of South Koreans)
- **Treatment variable**: *Chinese pollution*(Air Quality Index in China)
- **Interaction Variable**: **Asian dust**(the number of Asian dust days in South Korea)
- **Control Variables**: Time, Regions, Weather,Local Economic Conditions...

Jia and Ku(2019): Estimation Strategy

- The impact of *Chinese pollution* on district-level *mortality* that operates via **Asian dust**

$$
\begin{aligned}
Mortality_{ijk} &= \beta_0 + \beta_1 AsianDust_{ijk} + \beta_2 ChinesePollution_{jk} \\
&+ \beta_3 AsianDust_{ijk} \times ChinesePollution_{jk} \\
&+ \delta_1 X_{ijk} + u_{ijk}
\end{aligned}
$$

- Main coefficient of interest is β_3, which measures the effect of Chinese pollution in year j and month k on mortality in district i of South Korea.

Jia and Ku(2019): the result of interaction terms

Table 2: The Impact of Dust*China's Pollution on Mortality Rates in South Korea

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				Baseline				Placebo Tests	
			Mortality rates: Respiratory and Cardiovascular					Cancers	Accidents
Mean Dependent Var.				12.23				16.30	4.21
#Dust*China's Mean AQI				0.038**	0.033*	0.043**	0.040**	-0.008	0.007
				(0.016)	(0.018)	(0.018)	(0.019)	(0.020)	(0.009)
#Dust	0.076***		0.039	-0.251*	-0.214	-0.313**	0.072	0.525*	-0.102
	(0.025)		(0.032)	(0.131)	(0.142)	(0.149)	(0.240)	(0.275)	(0.119)
China's Mean AQI		0.265***	0.193*	0.117	0.138	0.202*	0.200*	-0.080	0.005
		(0.081)	(0.104)	(0.105)	(0.107)	(0.110)	(0.111)	(0.121)	(0.060)
District FE*Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Province FE*Month FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Weather (cubic polynomial)					Y	Y	Y	Y	Y
Local Prod (export, energy)						Y	Y	Y	Y
Local Prod*Dust							Y	Y	Y
Observations	29,464	29,464	29,464	29,464	28,952	28,024	28,024	28,024	28,024
R-squared	0.695	0.695	0.695	0.696	0.703	0.717	0.717	0.718	0.473

Jia and Ku(2019): the result of interaction terms

Table 2: The Impact of Dust*China's Pollution on Mortality Rates in South Korea

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Baseline							Placebo Tests	
	Mortality rates: Respiratory and Cardiovascular							Cancers	Accidents
Mean Dependent Var.	12.23							16.30	4.21
#Dust*China's Mean AQI				0.038**	0.033*	0.043**	0.040**	-0.008	0.007
				(0.016)	(0.018)	(0.018)	(0.019)	(0.020)	(0.009)
#Dust	0.076***		0.039	-0.251*	-0.214	-0.313**	0.072	0.525*	-0.102
	(0.025)		(0.032)	(0.131)	(0.142)	(0.149)	(0.240)	(0.275)	(0.119)
China's Mean AQI		0.265***	0.193*	0.117	0.138	0.202*	0.200*	-0.080	0.005
		(0.081)	(0.104)	(0.105)	(0.107)	(0.110)	(0.111)	(0.121)	(0.060)
District FE*Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Province FE*Month FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Weather (cubic polynomial)					Y	Y	Y	Y	Y
Local Prod (export, energy)						Y	Y	Y	Y
Local Prod*Dust							Y	Y	Y
Observations	29,464	29,464	29,464	29,464	28,952	28,024	28,024	28,024	28,024
R-squared	0.695	0.695	0.695	0.696	0.703	0.717	0.717	0.718	0.473

Jia and Ku(2019): the result of interaction terms

Table 2: The Impact of Dust*China's Pollution on Mortality Rates in South Korea

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Baseline							Placebo Tests	
	Mortality rates: Respiratory and Cardiovascular							Cancers	Accidents
Mean Dependent Var.	12.23							16.30	4.21
#Dust*China's Mean AQI				0.038**	0.033*	0.043**	0.040**	-0.008	0.007
				(0.016)	(0.018)	(0.018)	(0.019)	(0.020)	(0.009)
#Dust	0.076***		0.039	-0.251*	-0.214	-0.313**	0.072	0.525*	-0.102
	(0.025)		(0.032)	(0.131)	(0.142)	(0.149)	(0.240)	(0.275)	(0.119)
China's Mean AQI		0.265***	0.193*	0.117	0.138	0.202*	0.200*	-0.080	0.005
		(0.081)	(0.104)	(0.105)	(0.107)	(0.110)	(0.111)	(0.121)	(0.060)
District FE*Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Province FE*Month FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Weather (cubic polynomial)					Y	Y	Y	Y	Y
Local Prod (export, energy)						Y	Y	Y	Y
Local Prod*Dust							Y	Y	Y
Observations	29,464	29,464	29,464	29,464	28,952	28,024	28,024	28,024	28,024
R-squared	0.695	0.695	0.695	0.696	0.703	0.717	0.717	0.718	0.473

Jia and Ku(2019): the result of interaction terms

Table 2: The Impact of Dust*China's Pollution on Mortality Rates in South Korea

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Baseline							Placebo Tests	
	Mortality rates: Respiratory and Cardiovascular							Cancers	Accidents
Mean Dependent Var.	12.23							16.30	4.21
#Dust*China's Mean AQI				0.038**	0.033*	0.043**	0.040**	-0.008	0.007
				(0.016)	(0.018)	(0.018)	(0.019)	(0.020)	(0.009)
#Dust	0.076***		0.039	-0.251*	-0.214	-0.313**	0.072	0.525*	-0.102
	(0.025)		(0.032)	(0.131)	(0.142)	(0.149)	(0.240)	(0.275)	(0.119)
China's Mean AQI		0.265***	0.193*	0.117	0.138	0.202*	0.200*	-0.080	0.005
		(0.081)	(0.104)	(0.105)	(0.107)	(0.110)	(0.111)	(0.121)	(0.060)
District FE*Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Province FE*Month FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Weather (cubic polynomial)					Y	Y	Y	Y	Y
Local Prod (export, energy)						Y	Y	Y	Y
Local Prod*Dust							Y	Y	Y
Observations	29,464	29,464	29,464	29,464	28,952	28,024	28,024	28,024	28,024
R-squared	0.695	0.695	0.695	0.696	0.703	0.717	0.717	0.718	0.473

Jia and Ku(2019): Placebo Test

Table 2: The Impact of Dust*China's Pollution on Mortality Rates in South Korea

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Baseline							Placebo Tests	
	Mortality rates: Respiratory and Cardiovascular							Cancers	Accidents
Mean Dependent Var.	12.23							16.30	4.21
#Dust*China's Mean AQI				0.038**	0.033*	0.043**	0.040**	-0.008	0.007
				(0.016)	(0.018)	(0.018)	(0.019)	(0.020)	(0.009)
#Dust	0.076***		0.039	-0.251*	-0.214	-0.313**	0.072	0.525*	-0.102
	(0.025)		(0.032)	(0.131)	(0.142)	(0.149)	(0.240)	(0.275)	(0.119)
China's Mean AQI		0.265***	0.193*	0.117	0.138	0.202*	0.200*	-0.080	0.005
		(0.081)	(0.104)	(0.105)	(0.107)	(0.110)	(0.111)	(0.121)	(0.060)
District FE*Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Province FE*Month FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Weather (cubic polynomial)					Y	Y	Y	Y	Y
Local Prod (export, energy)						Y	Y	Y	Y
Local Prod*Dust							Y	Y	Y
Observations	29,464	29,464	29,464	29,464	28,952	28,024	28,024	28,024	28,024
R-squared	0.695	0.695	0.695	0.696	0.703	0.717	0.717	0.718	0.473

Summary

Wrap up

- We extend our multiple ols model form linear to nonlinear in Xs(the independent variables)
    - Polynomials,Logarithms and Interactions
    - The multiple regression framework can be extended to handle regression functions that are nonlinear in one or more X.
    - the difference between a standard multiple OLS regression and a nonlinear OLS regression model in Xs is how to explain estimating coefficients.
- All are very useful and common tools with OLS regressions. You had better understand it very clear.