Applied Statistics and Econometrics

Lecture 13 Nonlinearities

Saul Lach

October 2018

Outline of Lecture 13

1. **Nonlinear regression functions** (SW 8.1)
2. Polynomials (single regressor) (SW 8.2)
3. Logarithms (single regressor) (SW 8.2)
4. Interactions between variables (multiple regressors) (SW 8.3)
5. Application to California testscore data (SW 8.4)

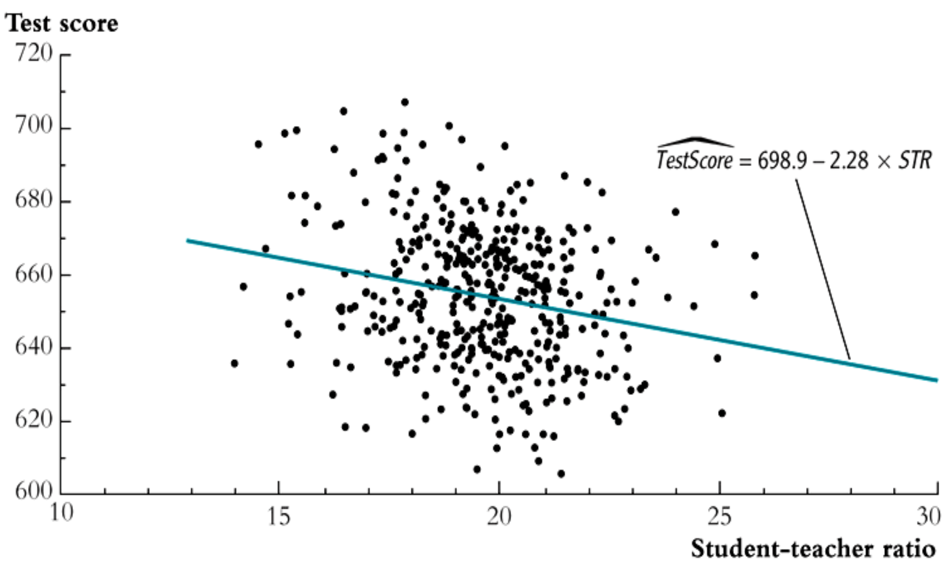
Nonlinear regression functions

- Everything so far has been linear in the X's.
- The approximation that the regression function is linear might be good for some variables, but not for others.
- The multiple regression framework can be extended to handle regression functions that are nonlinear in one or more of the X's.

The testscore – STR relation looks approximately linear...

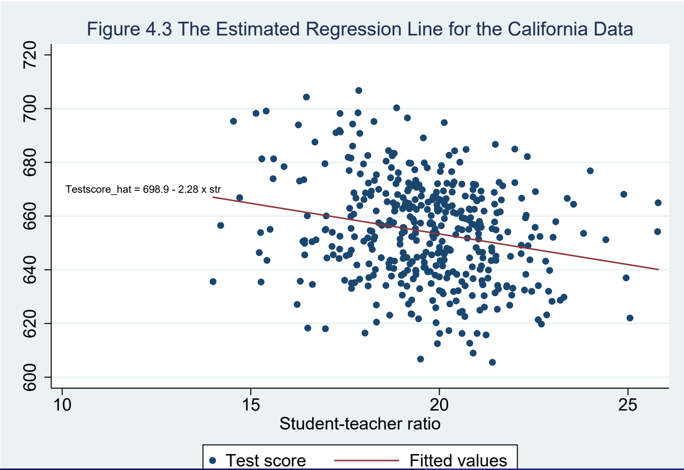
FIGURE 4.3 The Estimated Regression Line for the California Data

The estimated regression line shows a negative relationship between test scores and the student-teacher ratio. If class sizes fall by 1 student, the estimated regression predicts that test scores will increase by 2.28 points.



A Stata moment

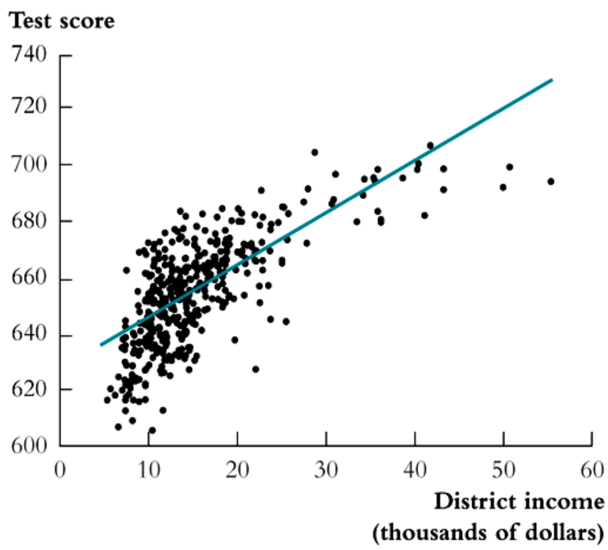
```
use "CASchool.dta"
label var testscr "Test score"
label var str "Student-teacher ratio"
twoway (scatter testscr str, sort msize(small)) (lfit
testscr str,sort), xlabel(10(5)25) ylabel(600(20)720)
text(670 10 "Testscore_hat = 698.9 - 2.28 x str",
placement(e) size(vsmall)) ti(Figure 4.3 The Estimated
regression Line for the California Data, size(medium))
```



But the testscore – income relation looks .... nonlinear.

FIGURE 6.2 **Scatterplot of Test Score vs. District Income with a Linear OLS Regression Function**

There is a positive correlation between test scores and district income (correlation = 0.71), but the linear OLS regression line does not adequately describe the relationship between these variables.



Nonlinear regression functions

- If the relation between Y and X is **nonlinear** then:
  - the effect on Y of a change in X depends on the value of X: the marginal effect of X is not constant.
  - A linear regression would be misspecified as it assumes the wrong functional form.
  - Because of this, the estimator of the effect on Y of X is biased in general; it even isn't right on average.
- The solution to this is to estimate a regression function that is **nonlinear** in X.

The general nonlinear population regression function

- The population regression function is

$$Y = f(X_1, X_2, \ldots, X_k) + u$$

  where $f(\cdot)$ is a possibly nonlinear function.
- The linear model is a special case where

$$f(X_1, X_2, \ldots, X_k) = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k$$

- In this course we assume the function $f(\cdot)$ is known.
- A topic of current research is "nonparametric econometrics" which seeks to estimate marginal effects without assuming a known functional form $f(\cdot)$.

Assumptions of the regression model

- We make the following assumptions:

1. $E(u|X_1, X_2, \ldots, X_k) = 0$ (same as LSA #1). It implies that $f(\cdot)$ is the **conditional expectation** of Y given the X's.
2. $(X_{1i}, X_{2i}, \ldots, X_{ki}, Y_i)$ are i.i.d. (same as LSA #2).
3. Big outliers are rare (same as LSA #3; the precise mathematical statement depends on specific function $f(\cdot)$).
4. No perfect multicollinearity (same idea as LSA #4; the precise mathematical statement depends on specific function $f(\cdot)$).

Marginal effect in general regression model

- The change in expected $Y - \Delta EY -$ associated with a change in X_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,

$$\Delta Y = f(X_1 + \Delta X_1, X_2, \ldots, X_k) - f(X_1, X_2, \ldots, X_k)$$

- This is very general as this "marginal" effect can depend on X_1 (i.e., it varies with X_1) and on other $X's$ besides X_1. This depends on the choice of function $f(\cdot)$.
  - Recall that in the linear model, $\Delta EY / \Delta X_1 = \beta_1$.
- We will study specific formulations of the function $f(X_1, X_2, \ldots, X_k)$.
- For simplicity, we do this in the context of a **single regressor** model but everything applies equally to the **multiple regression** model.

Two complementary choices of nonlinear functional form

1. **Polynomials in X**
   1. The population regression function is approximated by a quadratic, cubic, or higher-degree polynomial.
2. **Logarithmic transformations**
   1. Y and/or X is transformed by taking its (natural) logarithm. As we will see this gives a "percentages" interpretation that makes sense in many applications.

Where are we?

1. Nonlinear regression functions (SW 8.1)
2. **Polynomials (single regressor) (SW 8.2)**
3. Logarithms (single regressor) (SW 8.2)
4. Interactions between variables (multiple regressors) (SW 8.3)
5. Application to California testscore data (SW 8.4)

Polynomials in X

- Approximate the population regression function by a polynomial:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \ldots + \beta_r X^r + u$$

- This is just the linear multiple regression model – except that the regressors are powers of X!
- Estimation, hypothesis testing, etc. proceeds as in the multiple regression model using OLS.
- The coefficients are difficult to interpret (more on this below), but the regression function itself is interpretable.

A quadratic and cubic example

- Let `income` be the average income in the district (thousand dollars per capita).
- Quadratic specification:

$$Testscr = \beta_0 + \beta_1 income + \beta_2 (income)^2 + u$$

- Cubic specification:

$$Testscr = \beta_0 + \beta_1 income + \beta_2 (income)^2 + \beta_3 (income)^3 + u$$

Estimation of a quadratic specification in Stata

```
. rename avginc income   //reanme original variable

. g income2=income^2    //create square of income

. reg testscr income income2,r

Linear regression                               Number of obs     =        420
                                                F(2, 417)         =     428.52
                                                Prob > F          =     0.0000
                                                R-squared         =     0.5562
                                                Root MSE          =     12.724

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   3.850995   .2680941    14.36   0.000      3.32401    4.377979
     income2 |  -.0423085   .0047803    -8.85   0.000     -.051705   -.0329119
       _cons |   607.3017   2.901754   209.29   0.000     601.5978    613.0056
------------------------------------------------------------------------------
```


Testing the null hypothesis of linearity

- Testing the null hypothesis of linearity, against the alternative that the population regression is quadratic,

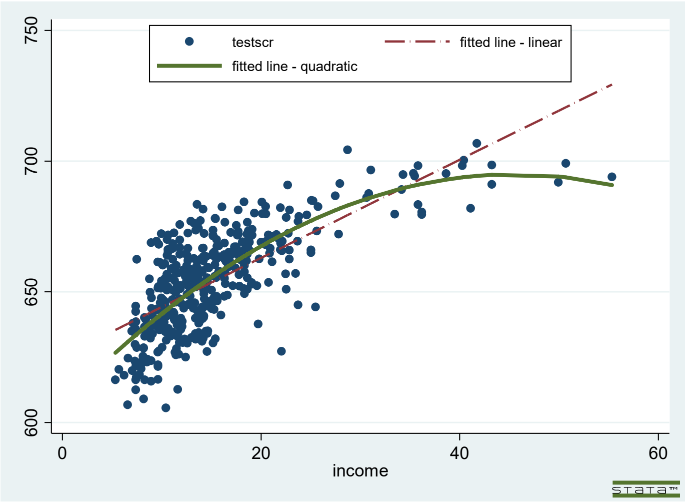
$$H_0 : \beta_2 = 0 \quad H_1 : \beta_2 \neq 0$$

- The t-statistic on income2 is -8.85, so the hypothesis of linearity is rejected against the quadratic alternative at the 1% significance level.

Quadratic model fits better than linear

Plot predicted (fitted) values of quadratic and linear models

$$\widehat{Testscr} = 625.38 + 1.879 income$$
$$\widehat{Testscr} = 607.3 + 3.851 income - 0.042(income)^2$$



Stata commands

```
reg testscr income income2,r
predict testscr_q
reg testscr income,r
predict testscr_l
label var testscr_q "fitted line - quadratic"
label var testscr_l "fitted line - linear"
twoway (scatter testscr income, sort) (line testscr_l
income, sort lwidth(medthick) lpattern(longdash_dot)) (line
testscr_q income, sort lwidth(thick)), legend(on
size(small) position(12) ring(0))
```

Marginal effects in nonlinear models

$$\widehat{Testscr} = \hat{\beta}_0 + \hat{\beta}_1 income + \hat{\beta}_2 (income)^2$$

- Predicted change in testscore of a change in income equal to Δinc:

$$\begin{aligned}
\Delta \widehat{Testscr} &= \left[\hat{\beta}_0 + \hat{\beta}_1 \left(income + \Delta inc\right) + \hat{\beta}_2 (income + \Delta inc)^2\right] \\
&\quad - \left[\hat{\beta}_0 + \hat{\beta}_1 income + \hat{\beta}_2 (income)^2\right] \\
&= \hat{\beta}_1 \Delta inc + \hat{\beta}_2 \left[(income + \Delta inc)^2 - (income)^2\right]
\end{aligned}$$

Marginal effect in quadratic model

- The effect of a **unit** change in income on test scores is

$$\Delta \widehat{Testscr} = \hat{\beta}_1 + \hat{\beta}_2 \left[(income + 1)^2 - (income)^2\right]$$

- Two implications:

1. The $\hat{\beta}'s$ (or the $\beta's$ themselves) **do not fully capture marginal effects in a quadratic model** (as they do in the linear model)!
2. The marginal effect of a change in income (X) **depends on the level of income (X).**
3. In fact, these implications hold for all nonlinear models (not just quadratic).

Marginal effect in quadratic models

- Given $\Delta inc = 1$(1 thousand dollars),

$$\begin{aligned}\Delta \widehat{Testscr} &= \hat{\beta}_1 + \hat{\beta}_2 \left[(income+1)^2 - (income)^2\right] \\ &= 3.851 - 0.042 \left[(income+1)^2 - (income)^2\right]\end{aligned}$$

depends on the level of income.

- We compute this at various levels of income

	$\Delta \widehat{Testscr}$
from 5 to 6	3.4
from 25 to 26	1.7
from 45 to 46	0.03

- The "effect" of a change in income is greater at low than high income levels (perhaps, due to a declining marginal benefit of an increase in school budgets?)

Marginal effect in quadratic models

- For infinitesimal changes in X we can just take the derivative (it is much simpler),

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + u \Longrightarrow \frac{dY}{dX} = \beta_1 + 2\beta_2 X$$

	$\Delta \widehat{Testscr}$
at income $= 5$	$3.851 - 2 \times 0.042 \times 5 = 3.431$
at income $= 25$	$3.851 - 2 \times 0.042 \times 25 = 1.75$
at income $= 45$	$3.851 - 2 \times 0.042 \times 45 = 0.071$

- Not a bad approximation... and much faster.

Estimation of a cubic specification in Stata

The cubic term is statistically significant at the 5%, but not 1%, level.

```
. g income3=income^3

. reg testscr income income2 income3,r

Linear regression                               Number of obs     =        420
                                                F(3, 416)         =     270.18
                                                Prob > F          =     0.0000
                                                R-squared         =     0.5584
                                                Root MSE          =     12.707

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   5.018677   .7073504     7.10   0.000     3.628251    6.409104
     income2 |  -.0958052   .0289537    -3.31   0.001     -.152719   -.0388913
     income3 |   .0006855   .0003471     1.98   0.049     3.27e-06    .0013677
       _cons |    600.079   5.102062   117.61   0.000     590.0499     610.108
------------------------------------------------------------------------------
```


Testing the null hypothesis of linearity

- Testing the null hypothesis of linearity, against the alternative that the population regression is quadratic and/or cubic, that is, is a polynomial of degree up to 3:

$$\begin{aligned} H_0 &: \beta_2 = \beta_3 = 0 \\ H_1 &: \text{ at least one of } \beta_2 \text{ and } \beta_3 \text{ is nonzero} \end{aligned}$$

- The null hypothesis is rejected at the 1% significance level.

```
. test income2 income3   //execute test command after running regression

 ( 1)  income2 = 0
 ( 2)  income3 = 0

       F(  2,   416) =   37.69
            Prob > F =    0.0000
```

Extension to multiple regression

- If we have a multiple regression

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + u$$

and we think there is a nonlinear relationship between Y and **one** of the $X's$, say the last one X_k, we can use a polynomial in that variable alone,

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + \beta_{k+1} X_k^2 + \beta_{k+2} X_k^3 + \ldots \beta_{k+r-1} X_k^r + u$$

- Estimation and hypotheses testing proceed as usual.

Extension to multiple regression

```
. reg testscr str el_pct income income2 income3

      Source |       SS           df       MS      Number of obs   =       420
-------------+----------------------------------   F(5, 414)       =    217.61
       Model |  110184.591         5  22036.9181   Prob > F        =    0.0000
    Residual |  41925.0031       414  101.268124   R-squared       =    0.7244
-------------+----------------------------------   Adj R-squared   =    0.7210
       Total |  152109.594       419  363.030056   Root MSE        =    10.063

------------------------------------------------------------------------------
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -.2257894   .2727757    -0.83   0.408    -.7619875    .3104087
      el_pct |  -.4645906   .0301539   -15.41   0.000    -.5238644   -.4053168
      income |    1.59157   .7188919     2.21   0.027     .1784367    3.004704
     income2 |   .0235483   .0307113     0.77   0.444    -.0368213    .0839178
     income3 |  -.0006129    .000384    -1.60   0.111    -.0013678    .0001419
       _cons |   638.9685   7.061529    90.49   0.000     625.0876    652.8494
------------------------------------------------------------------------------
```

Summary: polynomial regression functions

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \ldots + \beta_r X^r + u$$

- Estimation: by OLS after defining new regressors $X^2, \ldots, X^r$.
- Coefficients have complicated interpretations.
- To interpret the estimated regression function plot predicted values as functions of X and/or compute predicted $\Delta Y / \Delta X$ or $\frac{dY}{dX}$ at different values of X.
- Hypotheses concerning degree r can be tested by t- and F-tests on the appropriate (blocks of) variable(s).
- Choice of degree r: plot the data; t- and F-tests, check sensitivity of estimated effects; use judgment.

Where are we?

Logarithmic functions of Y and/or X

- $\ln(X)$ = the natural logarithm of X.
- We only deal with natural logarithms and often write also $\log(X)$ to mean $\ln(X)$ (so does Stata).
- Logarithms permit modeling relations in "percentage" terms (like elasticities), rather than linearly because **changes in logs are approximately equal to percentage changes.**

Changes in logs and percentage changes

- For any variable z, the change in logs is

$$\begin{aligned}\ln(z+\Delta z)-\ln(z) &= \ln\left(\frac{z+\Delta z}{z}\right)\\ &= \ln\left(1+\frac{\Delta z}{z}\right)\\ &\approx \frac{\Delta z}{z} \text{ for small } \frac{\Delta z}{z}\end{aligned}$$

- Thus, $100\times$ difference in the log of a variable z when it changes by Δz is approximately equal to the percentage difference, $100\frac{\Delta z}{z}$.

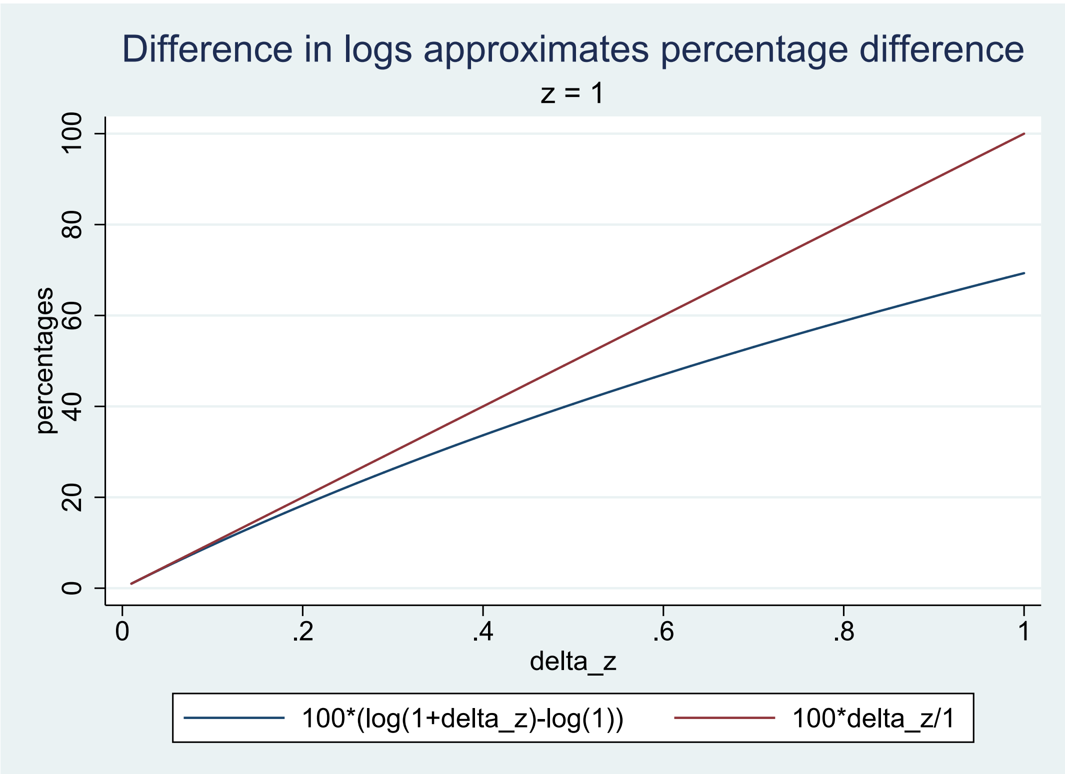
Changes in logs and percentage changes

- In calculus:
$$\frac{d \ln z}{dz} = \frac{1}{z} \Rightarrow d \ln z = \frac{dz}{z}$$
- And some numerical examples:

$$\begin{aligned}
\ln(1+.01)-\ln(1) &= \ln(1.01) = 0.009950 \approx \frac{.01}{1} = 0.01 \\
\ln(1+.1)-\ln(1) &= \ln(1.1) = 0.09531 \approx \frac{.1}{1} = 0.10 \\
\ln(1+.5)-\ln(1) &= \ln(1.5) = 0.40547 \not\approx \frac{.5}{1} = 0.5
\end{aligned}$$

so the approximation works for small relative increments $\frac{\Delta z}{z}$.

Plotting difference in logs and percentage difference



Three specifications using logs

Case	Population regression function
I. linear-log	$Y = \beta_0 + \beta_1 \ln(X) + u$
II. log-linear	$\ln(Y) = \beta_0 + \beta_1 X + u$
III. log-log	$\ln(Y) = \beta_0 + \beta_1 \ln(X) + u$

- 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.
- We use a single regressor for simplicity. Can extend to multiple regressors either with or without logs.

Logarithms permit non-linear relationship between Y and X

- For example,

$$\begin{aligned} \ln(Y) &= \beta_0 + \beta_1 X + u \Rightarrow Y = e^{\beta_0 + \beta_1 X + u} \\ \ln(Y) &= \beta_0 + \beta_1 \ln X + u \Rightarrow Y = e^{\beta_0 + \beta_1 \ln X + u} \end{aligned}$$

and, of course,

$$Y = \beta_0 + \beta_1 \ln(X) + u.$$

- But these models are still considered **linear** regression models since they just involve a transformation of the dependent and/or independent variables.
  - For example, in case I (linear-log), the model is linear in lnX (which we can just relabel with another name, say Z).

I. Linear-log population regression function

$$Y = \beta_0 + \beta_1 \ln(X) + u$$

- To **interpret** β_1 we calculate:

$$\begin{aligned}
\Delta Y &\equiv \overbrace{E\left[Y|X = x + \Delta x\right]}^{After} - \overbrace{E\left[Y|X = x\right]}^{Before} \\
&= \beta_0 + \beta_1 \ln(x + \Delta x) - \left[\beta_0 + \beta_1 \ln(x)\right] \\
&= \beta_1 \left(\ln(x + \Delta x) - \ln(x)\right) \\
&\approx \beta_1 \frac{\Delta x}{x} \text{ (for small } \Delta x|x)
\end{aligned}$$

- A 1% increase in X $\left(\frac{\Delta x}{x} = 0.01\right)$ is associated with $0.01\beta_1$ change in Y.
- A 10% increase in X $\left(\frac{\Delta x}{x} = 0.1\right)$ is associated with $0.1\beta_1$ change in Y.
- Or, differentiating, $dY = \beta_1 \frac{1}{X} dX$.

Example: testscores vs. log(income)

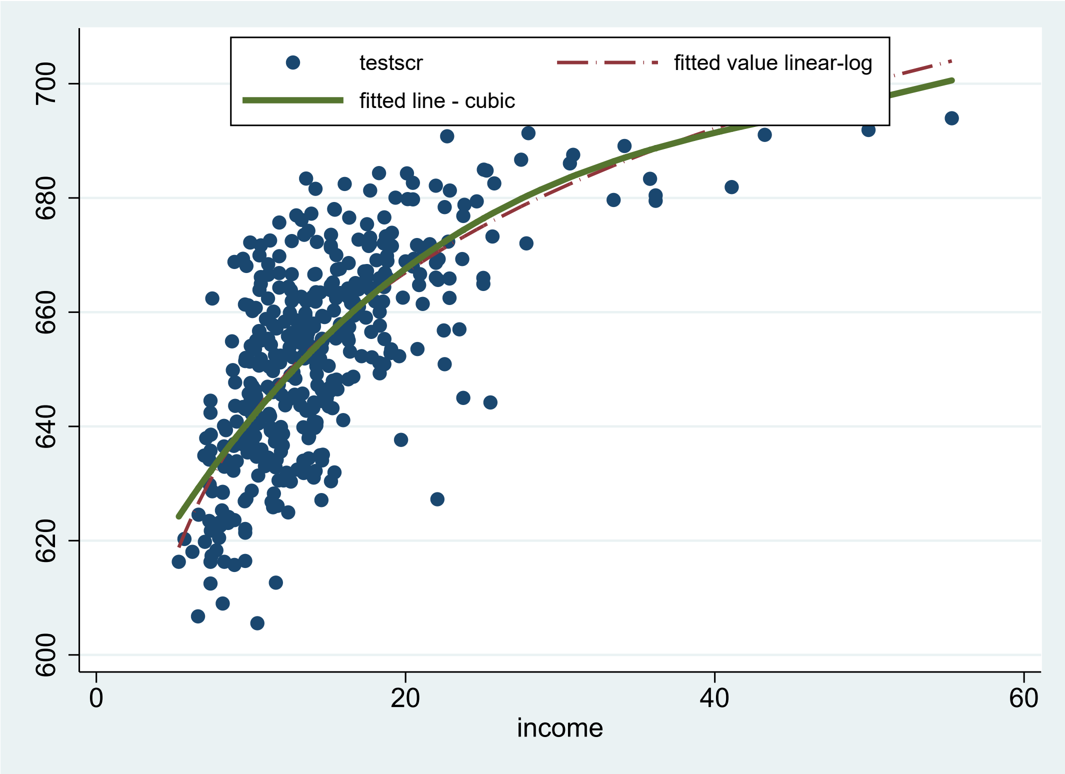
- First define the new regressor: `g lincome=ln(income)`
- The model is now linear in `ln(income)`, so the linear-log model can be estimated by OLS:

$$\widehat{Testscr} = \underset{(3.8)}{557.8} + \underset{(1.4)}{36.42} ln(income)$$

- A 1% increase in income is associated with an increase in test scores of 0.36 points.
- Standard errors, confidence intervals, R^2 – all the usual tools of regression apply here.
- How does this compare to the cubic model?

Cubic and linear-log models compared

In this sample, the linear-log and cubic specification are almost identical.



Stata commands

```
rename avginc income
g income2=income^2 //create square of income
g income3=income^3
reg testscr income income2 income3,r
predict testscr_c
label var testscr_c "fitted line - cubic"
reg testscr lincome,r
predict testscr_linlog
label var testscr_linlog "fitted value linear-log"
twoway (scatter testscr income, sort) (line testscr_linlog
income, sort lwidth(medthick) lpattern(longdash_dot)) (line
testscr_c income, sort lwidth(thick)), legend(on
size(small) position(12) ring(0))
```

II. Log-linear population regression function

$$\ln Y = \beta_0 + \beta_1 X + u$$

- To interpret β_1 we calculate

$$\begin{aligned} \Delta \ln Y &= E\left[\ln Y | X = x + \Delta x\right] - E\left[\ln Y | X = x\right] \\ &= \beta_0 + \beta_1 (x + \Delta x) - [\beta_0 + \beta_1 x] \\ &= \beta_1 \Delta x \end{aligned}$$

- But

$$\Delta \ln Y \approx \frac{\Delta Y}{Y} \Rightarrow \frac{\Delta Y}{Y} \approx \beta_1 \Delta x$$

- A unit increase in X is associated with a $100\beta_1\%$ change in Y.
- Differentiating, $\frac{1}{Y} dY = \beta_1 dX$.

Example: log (testscore) vs. income

After generating ln(testscore), and running the regression we get

$$\log\widehat{(Testscr)} = \underset{(0.0029)}{6.44} + \underset{(0.00018)}{0.0028}\, income$$

When (average per capita) income increases by \$1,000 ($\Delta income = 1$), testscore increase by 0.28 percent.
When (average per capita) income increases by \$10,000 ($\Delta income = 10$), testscore increase by 2.8 percent.

III. Log-log population regression function

$$\ln Y = \beta_0 + \beta_1 \ln(X) + u$$

- To interpret β_1 we calculate

$$\begin{aligned} & \Delta \ln Y = E[\ln Y|X = x + \Delta x] - E[\ln Y|X = x] \\ = \; & \beta_0 + \beta_1 \ln(x + \Delta x) - [\beta_0 + \beta_1 \ln x] \\ = \; & \beta_1 (\ln(x + \Delta x) - \ln(x)) \\ \Rightarrow \; & \frac{\Delta Y}{Y} \approx \beta_1 \frac{\Delta x}{x} \Rightarrow \underbrace{100\frac{\Delta Y}{Y}}_{\%\ change\ in\ Y} \approx \beta_1 \underbrace{100\frac{\Delta x}{x}}_{\%\ change\ in\ X} \end{aligned}$$

- A 1% change in X $\left(\frac{\Delta x}{x} = 0.01\right)$ is associated with a β_1 percent change in Y.
- In the log-log specification, β_1 has the interpretation of an **elasticity**.
- Differentiating, $\frac{1}{Y}dY = \beta_1 \frac{1}{X}dX \Longrightarrow \beta_1 = \frac{\frac{1}{Y}dY}{\frac{1}{X}dX}$.

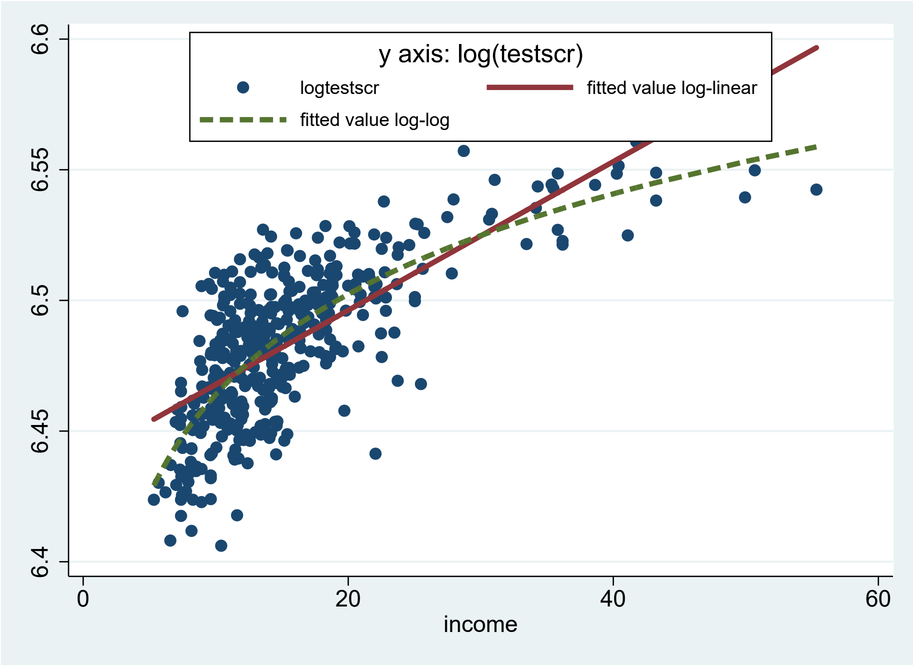
Example: log(testscore) vs. log(income)

$$\log\left(\widehat{Testscr}\right) = \underset{(0.006)}{6.336} + \underset{(0.0021)}{0.0554} ln(income)$$

- A 1% increase in income is associated with an increase of 0.0554% in test scores.
- A 10% increase in income is associated with an increase of 0.554% in test scores.

Comparing fitted values across models

- Models having the same dependent variable can be easily compared.
- The fitted values here are fitted values of log(testscr). The log-linear model is a straight line in the (Y, X) plane.



Comparing fitted values across models

- But if we want to compare models that have $\log(Y)$ and Y as dependent variable we need to be careful.
- Compute fitted values of Y for each model:

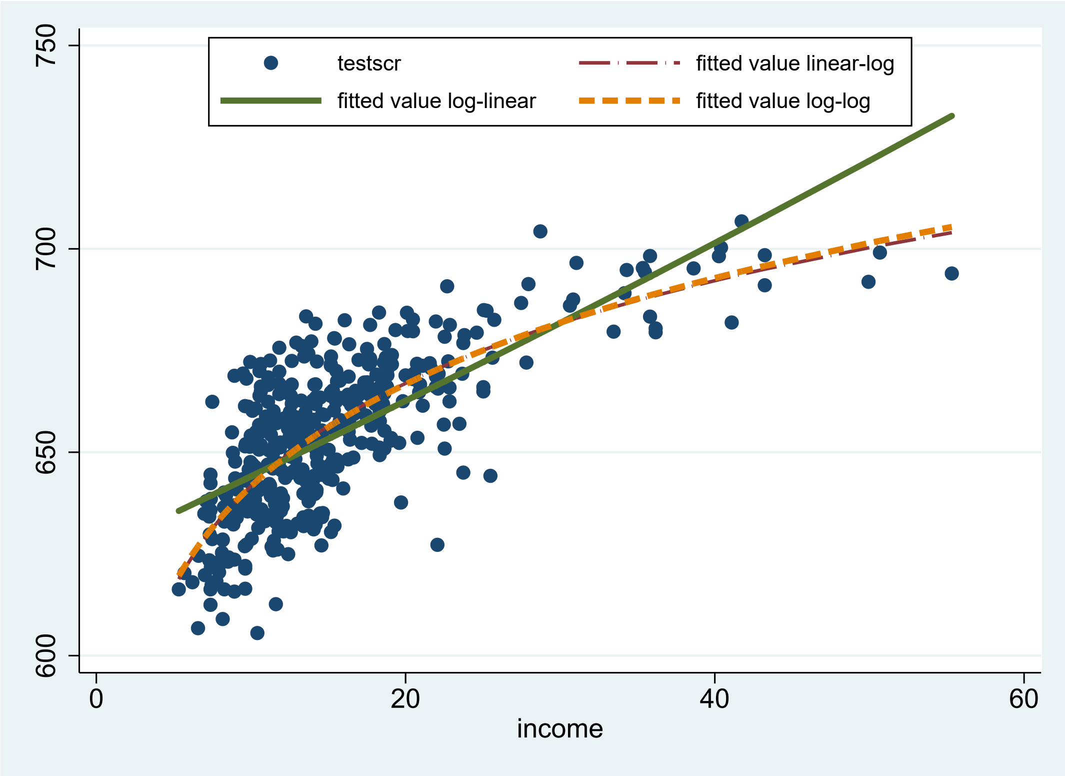
$$\widehat{\ln(Y)} = \hat{\beta}_0 + \hat{\beta}_1 X \text{ and define } \widehat{Y} = e^{\widehat{\ln(Y)}} = e^{\hat{\beta}_0 + \hat{\beta}_1 X}$$
$$\widehat{\ln(Y)} = \hat{\beta}_0 + \hat{\beta}_1 \ln X \text{ and define } \widehat{Y} = e^{\widehat{\ln(Y)}} = e^{\hat{\beta}_0 + \hat{\beta}_1 \ln X}$$

and, of course,

$$\widehat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \ln X$$

for which nothing has to be specially computed as it is the predicted value computed after the `regress` command.

Comparing fitted values of the three nonlinear models



Stata commands

```
g logtestscr=log(testscr)
reg logtestscr income,r
predict testscr_loglin
label var testscr_loglin "fitted value log-linear"
reg logtestscr lincome,r
predict testscr_loglog
label var testscr_loglog "fitted value log-log"
twoway (scatter logtestscr income, sort) (line
testscr_loglin income, sort lwidth(thick)) (line
testscr_loglog income, sort lpattern(dash) lwidth(thick)),
legend(on size(small) position(12) ring(0)subtitle(y axis:
log(testscr)))
```

Summary: Logarithmic transformations

- Three cases, differing in whether Y and/or X is transformed by taking logarithms.
- After creating the new variable(s) ln(Y) and/or ln(X), the regression is linear in the new variables and the coefficients can be estimated by OLS.
- Hypothesis tests and confidence intervals are implemented and interpreted as usual.
- The choice of specification should be guided by judgment (which interpretation makes the most sense in your application?), tests, and plotting predicted values.

Summary: Logarithmic transformations

- The interpretation of β_1 differs from case to case.
-

Model	Change in X	Change in Y	In words
linear-log	$\frac{\Delta X}{X}$	ΔY	$\frac{1}{100} \times \beta_1 = \frac{\text{Change in Y}}{\text{1 \% change in X}}$
log-linear	ΔX	$\frac{\Delta Y}{Y}$	$100 \times \beta_1 = \frac{\text{\% Change in Y}}{\text{1 unit change in X}}$
log-log	$\frac{\Delta X}{X}$	$\frac{\Delta Y}{Y}$	$\beta_1 = \frac{\text{\% Change in Y}}{\text{1 \% change in X}}$

Earning function estimation from Italian LFS

- We estimated "earning functions" – regressions of wages on education and other characteristics – using the level of wages as the dependent variable.
- The usual specification of an earning function differs in that:
  - we use logs of wages instead of wages.
  - we use a quadratic in age (or potential work experience).
- What are the implications of these modifications on the interpretation of the coefficient of education and of age?

The "old" specification

```
.  reg retric educ etam  female Center South,robust

Linear regression                               Number of obs     =     26,127
                                                F(5, 26121)       =    1535.35
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2647
                                                Root MSE          =      448.1
```

retric	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	58.36051	1.081296	53.97	0.000	56.24111	60.4799
etam	13.10358	.2559897	51.19	0.000	12.60183	13.60534
female	-342.248	5.613383	-60.97	0.000	-353.2506	-331.2455
Center	-95.4606	7.079332	-13.48	0.000	-109.3365	-81.58472
South	-174.0507	6.833489	-25.47	0.000	-187.4447	-160.6566
_cons	166.3771	19.17763	8.68	0.000	128.7879	203.9663

The new specification

```
.  g lretric=log(retric)
(75,789 missing values generated)

.  g etam2=etam^2

.  reg lretric educ etam etam2 female Center South,robust

Linear regression                               Number of obs     =     26,127
                                                F(6, 26120)       =    1256.03
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2395
                                                Root MSE          =     .39357

------------------------------------------------------------------------------
             |               Robust
     lretric |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0425932   .0008517    50.01   0.000     .0409239    .0442626
        etam |    .036764   .0017696    20.78   0.000     .0332956    .0402324
       etam2 |  -.0003106   .0000207   -15.04   0.000    -.0003511   -.0002701
      female |  -.2958357   .0049976   -59.20   0.000    -.3056314   -.2860401
      Center |  -.0866751   .0062467   -13.88   0.000     -.098919   -.0744313
       South |  -.1500943   .0063547   -23.62   0.000    -.1625497   -.1376388
       _cons |   5.725301   .0378177   151.39   0.000     5.651176    5.799426
------------------------------------------------------------------------------
```


Rate of return to education

- The coefficient of education in the new specification is the percentage change in wages associated with a one year change in education.
- It is a **rate of return** interpretation. In this sample, it is estimated to be 4.3%.
- Age (experience) has positive but **diminishing** (due to the negative quadratic coefficient) effects on wages.

Where are we?

1. Nonlinear regression functions (SW 8.1)
2. Polynomials (single regressor) (SW 8.2)
3. Logarithms (single regressor) (SW 8.2)
4. **Interactions between variables (multiple regressors) (SW 8.3)**
    1. Interaction between two binary variables.
    2. Interaction between a binary and a continuous variable.
    3. Interaction between two continuous variables.
5. Application to California testscore data (SW 8.4)

Interaction between regressors

- Perhaps a class size reduction is more effective in some circumstances than in others...
- Perhaps smaller classes are more effective when there are many English learners needing individual attention.
- This means that, perhaps, $\frac{\Delta Testscore}{\Delta STR}$ might depend on `el_pct` (% of English learners).
- More generally,

$$\frac{\Delta Y}{\Delta X_1} \text{ might depend on } X_2.$$

- How to model such "interactions" between X_1 and X_2?
    - We first consider binary X's, then continuous X's.

Interaction between two binary variables

$$Y = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + u$$

- D_1, D_2 are binary (dummy) variables.
- 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, we include the "**interaction term**"

  $$D_1 \times D_2$$

  as a regressor ($(D_1 \times D_2)$ represents the multiplication of D_1 and D_2):

  $$Y = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \beta_3 (D_1 \times D_2) + u \qquad \text{model with interaction}$$

- To run the regression we need to generate a new regressor (maybe call it $D1XD2$) equal to this multiplication.

Interpreting the coefficients

$$Y = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \beta_3 (D_1 \times D_2) + u$$

- General rule is to compare Y "before and after" a change in D_1, holding D_2 constant:

$$\begin{aligned} E(Y|D_1 &= 0, D_2 = d_2) = \beta_0 + \beta_2 d_2 && \text{(a)} \\ E(Y|D_1 &= 1, D_2 = d_2) = \beta_0 + \beta_1 + \beta_2 d_2 + \beta_3 d_2 && \text{(b)} \end{aligned}$$

- Subtract (a) from (b):

  $$\Delta Y \equiv E(Y|D_1 = 1, D_2 = d_2) - E(Y|D_1 = 0, D_2 = d_2) = \beta_1 + \beta_3 d_2$$

- The effect of D_1 equals β_1 when $D_2 = 0$ and $\beta_1 + \beta_3$ when $D_2 = 1$.
- β_3 is the increment to the effect of D_1 on Y when $D_2 = 1$.
- **The interaction term between the two variables allows for the effect of one variable on Y to depend on the value of the other variable.**

Example: testscores, STR, English learners

- We define 2 dummy variables

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

- One way to do this in Stata is

$$\begin{aligned} gen\ HiSTR &= (STR >= 20) \\ gen\ HiEL &= (el\_pct >= 10) \end{aligned}$$

and the interaction term is just the multiplication of these two dummies.

Dummies and their interaction

```
. gen Histr=(str>=20)

. gen Hiel = (el_pct>=10)

. gen HistrXHiel=Histr*Hiel

. reg testscr Histr Hiel HistrXHiel ,r

Linear regression                               Number of obs     =        420
                                                F(3, 416)         =      60.20
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2956
                                                Root MSE          =     16.049

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       Histr |  -1.907842   1.932215    -0.99   0.324    -5.705964    1.890279
        Hiel |  -18.16295   2.345952    -7.74   0.000    -22.77435   -13.55155
  HistrXHiel |  -3.494335   3.121226    -1.12   0.264    -9.629677    2.641006
       _cons |   664.1433   1.388089   478.46   0.000     661.4147    666.8718
------------------------------------------------------------------------------
```

Interpretation

$$\widehat{testscr} = \underset{(1.4)}{664.1} - \underset{(2.3)}{18.16} HiEL - \underset{(1.9)}{1.91} HiSTR - \underset{(3.1)}{3.49} (HiSTR \times HiEL)$$

- "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 **stronger** effect when the percent of English learners is **large**.
- The interaction isn't statistically significant in this sample: $t = -3.49/3.1 = -1.12$.

On dummy regressors and group ("cell") means

$$\widehat{testscr} = \underset{(1.4)}{664.1} - \underset{(2.3)}{18.16} HiEL - \underset{(1.9)}{1.91} HiSTR - \underset{(3.1)}{3.49} (HiSTR \times HiEL)$$

- The predicted values for each combination of the dummies is equal to the mean of testscore in the corresponding group (or "cell", e.g., HiEL=0 and HiSTR=1).
- Check it out! (differences due to rounding only)
-

```
. table Hiel Histr , c(mean testscr)
```

Hiel	Histr 0	Histr 1
0	**664.1433**	**662.2355**
1	**645.9803**	**640.5782**

Interactions between continuous and binary variables

$$Y = \beta_0 + \beta_1 D + \beta_2 X + u$$

where D is binary, X is continuous.

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

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

Interpreting the coefficients

- General rule is to compare Y "before and after" a change in X :

$$\begin{aligned} E(Y|D &= d, X = x) = \beta_0 + \beta_1 d + \beta_2 x + \beta_3 (d \times x) \\ E(Y|D &= d, X = x + \Delta x) = \beta_0 + \beta_1 d + \beta_2 (x + \Delta x) + \beta_3 (d \times (x + \Delta x)) \end{aligned}$$

- Subtracting the top from the bottom equation gives:

$$\begin{aligned} \Delta Y &\equiv E(Y|D = d, X = x + \Delta x) - E(Y|D = d, X = x) \\ &= \beta_2 \Delta x + \beta_3 d \Delta x \Rightarrow \frac{\Delta Y}{\Delta x} = \beta_2 + \beta_3 d \end{aligned}$$

- The effect of X depends on D (what we wanted)
- β_3 is the increment to the effect of X on Y when $D = 1$.
- **The interaction between a binary and a continuous variable allows for the (marginal) effect of the continuous variable to vary with the group defined by the binary variable.**

Binary-continuous interactions: two regression lines

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

- One way to understand what this interaction does is to realize that it allows for different regression lines for the two groups defined by the dummy variable.
- The population regression line when $D = 0$ (for observations with $D_i = 0$) is

$$Y = \beta_0 + \beta_2 X + u$$

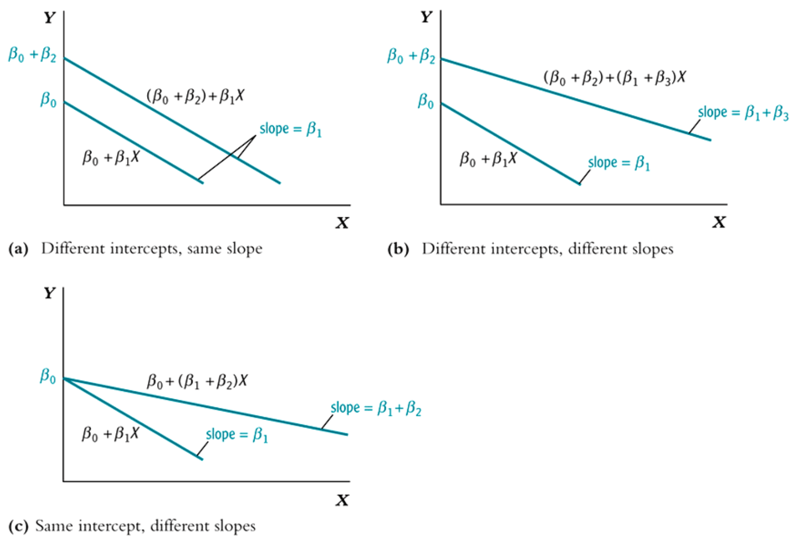
  and when $D = 1$ (for observations with $D_i = 1$) it is

$$Y = \beta_0 + \beta_1 + (\beta_2 + \beta_3) X + u$$

- These are two regression lines with different slopes and intercepts.

Binary-continuous interactions: two regression lines

FIGURE 6.8 **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 for different intercepts and different slopes; and (c) $\beta_0 + \beta_1 X + \beta_2 (X \times D)$ has the same intercept but allows for different slopes.

Example: interacting STR and HiEL

```
. gen strXHiel=str*Hiel

. reg testscr str Hiel strXHiel,r

Linear regression                               Number of obs     =        420
                                                F(3, 416)         =      63.67
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3103
                                                Root MSE          =      15.88
```

testscr	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
str	-.9684601	.5891016	-1.64	0.101	-2.126447	.1895268
Hiel	5.639141	19.51456	0.29	0.773	-32.72029	43.99857
strXHiel	-1.276613	.9669194	-1.32	0.187	-3.17727	.6240436
_cons	682.2458	11.86781	57.49	0.000	658.9175	705.5742

Example: interacting STR and HiEL

$$\widehat{testscr} = \underset{(11.87)}{682.2} - \underset{(0.59)}{0.97}\,STR + \underset{(19.52)}{5.6}\,HiEL - \underset{(0.97)}{1.28}\,(STR \times HiEL)$$

- Two regression lines: one for each HiSTR group. When $HiEL = 0$,

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

  and when $HiEL = 1$,

- $$\begin{aligned} \widehat{testscr} &= 682.2 - 0.97STR + 5.6 - 1.28STR \\ &= 687.8 - 2.25STR \end{aligned}$$

- Class size reduction is estimated to have a **larger** effect when the percent of English learners is **large**.

Testing hypotheses about the different regression lines

$$\widehat{testscr} = \underset{(11.87)}{682.2} - \underset{(0.59)}{0.97}\,STR + \underset{(19.52)}{5.6}\,HiEL - \underset{(0.97)}{1.28}\,(STR \times HiEL)$$

- There are various hypotheses of interest that can be tested:

1. Regressions lines have the same slope
2. Regressions lines have the same intercept
3. Regressions line are identical

Hypothesis 1: The two regression lines have the same slope

$$\widehat{testscr} = \underset{(11.87)}{682.2} - \underset{(0.59)}{0.97}\,STR + \underset{(19.52)}{5.6}\,HiEL - \underset{(0.97)}{1.28}\,(STR \times HiEL)$$

- H_0 : the coefficient on the interaction term $STR \times HiEL$ is zero.
- This implies no difference in slopes.
- The t-stastistic is:

$$t = \frac{-1.28}{0.97} = -1.32 \Rightarrow \text{do not reject } H_0$$

This hypothesis is very important since it tests that there is no difference in the response to class size between the two groups of school districts (with high and low % English learners).

Hypothesis 2: The two regression lines have the same intercept

$$\widehat{testscr} = \underset{(11.87)}{682.2} - \underset{(0.59)}{0.97}\,STR + \underset{(19.52)}{5.6}\,HiEL - \underset{(0.97)}{1.28}\,(STR \times HiEL)$$

- H_0 : the coefficient on *HiEL* is zero.
- This implies no difference in intercepts.
- The t-statistics is,

$$t = \frac{-5.6}{19.52} = 0.29 \Rightarrow \text{do not reject } H_0$$

Hypothesis 3: The two regression lines are identical

$$\widehat{testscr} = \underset{(11.87)}{682.2} - \underset{(0.59)}{0.97}\,STR + \underset{(19.52)}{5.6}\,HiEL - \underset{(0.97)}{1.28}\,(STR \times HiEL)$$

- H_0 : the coefficients on $HiEL = 0$ **and** on $STR \times HiEL$ are **both** zero. This is a **joint** hypothesis (of two coefficients).
- H_0 implies no differences in intercepts and slopes.
- The F-test is

$$F(2,416) = 89.94 \qquad Prob > F = 0.0000 \Rightarrow \text{reject } H_0!!!$$

- Note that we reject the joint hypothesis but do **not** reject the individual hypotheses.
- This "strange" result can happen when there is high (but not perfect!) multicollinearity: i. e., high correlation between *HiEL* and $STR \times HiEL$ (0.99 in our case).

Interactions between two continuous variables

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

where X_1 and X_2 are continuous.

- As specified above, the effect on Y of X_1 is β_1, which does not depend on X_2.
- And vice-versa: the effect of X_2 is β_2, which does not depend on X_1.
- To allow the effect of X_1 to depend on X_2, we include the "interaction term" $X_1 \times X_2$ as a regressor:

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

Interpreting the coefficients

- General rule is to compare Y "before and after" a change in, say, X_1 :

$$\begin{aligned} E(Y|X_1 &= x_1, X_2 = x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 \times x_2) \\ E(Y|X_1 &= x_1 + \Delta x_1, X_2 = x_2) = \beta_0 + \beta_1 (x_1 + \Delta x_1) + \beta_2 x_2 \\ &\quad + \beta_3 ((x_1 + \Delta x_1) \times x_2) \end{aligned}$$

- Subtracting the first from the second equation gives:

$$\begin{aligned} \Delta Y &\equiv E(Y|X_1 = x_1 + \Delta x_1, X_2 = x_2) - E(Y|X_1 = x_1, X_2 = x_2) \\ &= \beta_1 \Delta x_1 + \beta_3 x_2 \Delta x_1 \Rightarrow \frac{\Delta Y}{\Delta x_1} = \beta_1 + \beta_3 x_2 \end{aligned}$$

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

Example: interacting STR and EL_pct

```
. gen strXel_pct=str*el_pct

. reg testscr str el_pct strXel_pct,r

Linear regression                               Number of obs     =        420
                                                F(3, 416)         =     155.05
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4264
                                                Root MSE          =     14.482
```

testscr	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
str	-1.117018	.5875135	-1.90	0.058	-2.271884	.0378468
el_pct	-.6729116	.3741231	-1.80	0.073	-1.408319	.0624958
strXel_pct	.0011618	.0185357	0.06	0.950	-.0352736	.0375971
_cons	686.3385	11.75935	58.37	0.000	663.2234	709.4537

Example: interacting STR and HiEL

$$\widehat{testscr} = \underset{(11.8)}{686.3} - \underset{(0.59)}{1.12}\,STR - \underset{(0.37)}{0.67}\,PctEL + \underset{(0.019)}{0.0012}(STR \times El\_pct),$$

- The estimated effect of class size reduction is nonlinear because the size of the effect itself depends on El_pct,

$$\frac{\Delta testscore}{\Delta str} = -1.12 + 0.0012 El\_pct$$

El_pct	$\frac{\Delta testscore}{\Delta str}$	
0	$-1.12 + 0.0012 \times 0 =$	-1.12
10%	$-1.12 + 0.0012 \times 10 =$	-1.108
11%	$-1.12 + 0.0012 \times 11 =$	-1.1068
50%	$-1.12 + 0.0012 \times 50 =$	-1.06

Increasing El_pct from 10% to 11% changes the marginal effect of STR by $-1.1068 - (-1.108) = 0.0012$ as expected!

Testing hypotheses

$$\widehat{testscr} = \underset{(11.8)}{686.3} - \underset{(0.59)}{1.12}\, STR - \underset{(0.37)}{0.67}\, PctEL + \underset{(0.019)}{0.0012}(STR \times El\_pct),$$

- Does population coefficient on $STR \times El\_pct = 0$?
  - $t = 0.0012/0.019 = .06 \Rightarrow$ do not reject null at 5% level.
- Does population coefficient on $STR = 0$?
  - $t = -1.12/0.59 = -1.90 \Rightarrow$ do not reject null at 5% level
- Do the coefficients on **both** STR **and** $STR \times El\_pct = 0$?

$$F(2, 416) = 3.89 \qquad Prob > F = 0.0212 \Rightarrow \text{reject null at 5\% level !!}$$

As before, high but imperfect multicollinearity between STR and $STR \times El\_pct$ (0.25 in this sample) can lead to this "non-intuitive" result.

Where are we?

1. Nonlinear regression functions (SW 8.1)
2. Polynomials (single regressor) (SW 8.2)
3. Logarithms (single regressor) (SW 8.2)
4. Interactions between variables (multiple regressors) (SW 8.3)
5. **Application to California testscore data (SW 8.4)**

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

Focus on two questions:

1. Does a reduction from, say, 25 to 20 have same effect as a reduction from, say, 20 to 15? More generally, are there nonlinear effects of class size reduction on test scores?
2. Are small classes more effective when there are many English learners? More generally, are there interactions between EL_pct and STR?

Strategy for answering Question #1 (different effect of STR at different STR levels?)

- Estimate linear and nonlinear functions of STR, holding constant relevant demographic variables:
  - % of English learner (EL_pct)
  - Income (entered in logs because of previous work – recall the linear-log model).
  - % on free/subsidized lunch (mean_pct).
  - Expenditures per pupil are **not** included in the regression so as to allow for increases in expenditures when decreasing STR.
- See whether adding the nonlinear terms makes an "economically important" quantitative difference.
- Test for whether the nonlinear terms are significant.

Regression results for question 1

Nonlinear regression models

VARIABLES	(1) reg1 Test scores	(2) reg2 Test scores	(3) reg3 Test scores	(4) reg4 Test scores
Student-Teacher Ratio (STR)	-1.00***	-0.73***	65.3**	64.3***
	(0.27)	(0.26)	(25.3)	(24.9)
STR^2			-3.47***	-3.42***
			(1.27)	(1.25)
STR^3			0.060***	0.059***
			(0.021)	(0.021)
% English learners	-0.12***	-0.18***	-0.17***	
	(0.033)	(0.034)	(0.034)	
Binary for %English learners >= 10%)				-5.47***
				(1.03)
% Eligible for subsidized lunch	-0.55***	-0.40***	-0.40***	-0.42***
	(0.024)	(0.033)	(0.033)	(0.029)
Average district income (in logs)		11.6***	11.5***	11.7***
		(1.82)	(1.81)	(1.77)
Constant	700***	659***	245	252
	(5.57)	(8.64)	(166)	(164)
Observations	420	420	420	420
R-squared	0.775	0.796	0.801	0.801

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Stata commands

```
g str_sq=str^2
g str_cu=str^3
gen Hiel = (el_pct>=10)
g lincome=log(avginc)
label var el_pct "% English learners"
label var meal_pct "% Eligible for subsidized lunch"
label var lincome "Average district income (in logs)"
label var testscr "Test scores"
label var str "STR"
label var str_sq "STR^2"
label var str_cu "STR^3"
label var Hiel "Binary for %English learners >= 10%)"
label var str "Student-Teacher Ratio (STR)"
```

Stata commands

```
/////QUESTION 1
reg testscr str el_pct meal_pct,r
estimate store reg1
reg testscr str el_pct meal_pct lincome,r //adding income
estimate store reg2
reg testscr str str_sq str_cu el_pct meal_pct lincome,r
estimate store reg3
test str_sq str_cu
reg testscr str str_sq str_cu Hiel meal_pct lincome,r //dummy for %
english learners
estimate store reg4
test str_sq str_cu
outreg2 [reg1 reg2 reg3 reg4] using table1.xls, auto(2) sortvar(str str_sq
str_cu el_pct Hiel meal_pct lincome) label replace see
```


Testing significance of nonlinear terms

F-statistics and p-values on joint hypotheses

model	H_0:	F	p-value
reg 3	$STR^2 = STR^3 = 0$	5.96	0.0028
reg 4	$STR^2 = STR^3 = 0$	6.17	0.0023

- Nonlinear terms are significantly different from zero.
- But do they matter economically?

Economic significance of nonlinear terms

- After taking economic factors and nonlinearities into account, what is the estimated effect on test scores of reducing the student–teacher ratio by **one** student per teacher?
- Strong "diminishing returns": Cutting STR has a greater effect at lower student-teacher ratios.
- Not much difference in marginal effects between cols 3 and 4.
- But big difference with linear specifications (cols 1 and 2).

Model	STR	Effect of reducing STR by 1 student	
		formula	value
linear (col. 2)	all	- 0.73x(-1)	0.73
nonlinear (col.3)	20	65.3×(-1)-3.47×(19²-20²)+0.060×(19³-20³)	1.57
nonlinear (col.3)	22	65.3×(-1)-3.47×(21²-22²)+0.060×(21³-22³)	0.69
nonlinear (col.4)	20	64.3x(-1)-3.42×(19²-20²)+0.059×(19³-20³)	1.761
nonlinear (col.4)	22	64.3x(-1)-3.42×(21²-22²)+0.059×(21³-22³)	0.927

Strategy for answering Question #2 (differential effect of changing STR by % English learners)

- Question 2: Are smaller classes more effective when there are many English learners? More generally, are there interactions between EL_pct and STR?
- Estimate linear and nonlinear functions of STR, interacted with % English learners.
- If the specification is nonlinear (with STR, STR^2, STR^3), then you need to add interactions with **all** the nonlinear terms.
- We will use binary-continuous interactions by adding $HiEL \times STR$, $HiEL \times STR^2$ and $HiEL \times STR^3$.

Regression results for question 2

Nonlinear regression models

VARIABLES	(1) reg1 Test scores	(2) reg2 Test scores	(3) reg3 Test scores	(4) reg4 Test scores
Student-Teacher Ratio (STR)	-0.734***	-0.772***	64.34***	83.70***
	(0.257)	(0.256)	(24.86)	(28.50)
STR^2			-3.424***	-4.381***
			(1.250)	(1.441)
STR^3			0.0593***	0.0749***
			(0.0208)	(0.0240)
% English learners	-0.176***			
	(0.0337)			
Binary for %English learners >= 10%		-5.791***	-5.474***	816.1**
		(1.027)	(1.034)	(327.7)
HiELxSTR				-123.3**
				(50.21)
HiELxSTR^2				6.121**
				(2.542)
HiELxSTR^3				-0.101**
				(0.0425)
% Eligible for subsidized lunch	-0.398***	-0.417***	-0.420***	-0.418***
	(0.0332)	(0.0283)	(0.0285)	(0.0287)
Average district income (in logs)	11.57***	11.82***	11.75***	11.80***
	(1.819)	(1.775)	(1.771)	(1.778)
Constant	658.6***	659.4***	252.0	122.3
	(8.642)	(8.421)	(163.6)	(185.5)
Observations	420	420	420	420
R-squared	0.796	0.797	0.801	0.803

Tests of hypotheses

F-statistics and p-values on joint hypotheses for model 4

$\mathbf{H}_0$:	**F**	**p-value**
$HiEL \times STR = HiEL \times STRSTR^2 = HiEL \times STRSTR^3 = 0$	2.69	0.046
All coefficients involving STR (6 coeffs.)	4.96	0.0001
$STR^2 = STR^3 = 0$	5.81	0.0033

- Interactions of STR with HiEl are significantly different from zero at the 5% (but not 1%) significance level.
- But do they matter economically?

Economic significance of interaction terms

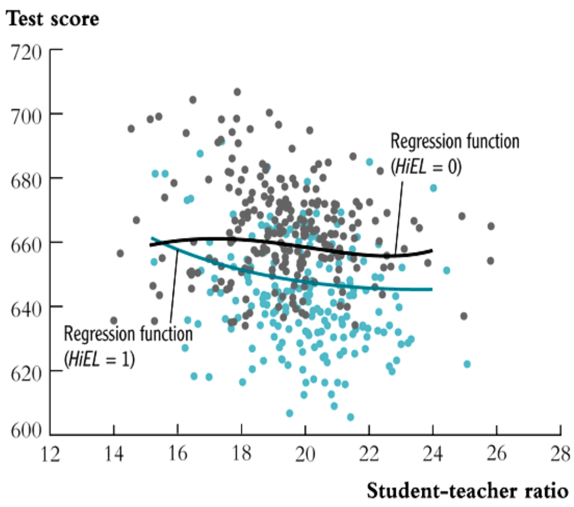
- Use model 4 to compute change in expected testscore when STR is reduced by one student at STR=20 for schools with a high ($\geq$ 10%)% of English learners ($HiEL = 1$) and for schools with a low percentage ($HiEL = 0$).
- The marginal effect of STR is not very different between the two groups of schools (1.7 and 1.56).

Model	STR	Effect of reducing STR by 1 student	
		formula	value
Hiel=0	20	83.7×(-1) - 4.381 × (19^2-20^2) + 0.0749 × (19^3-20^3)	1.6981
Hiel=1	20	1.6981 (effect when Hiel=0) - 123.3 × (-1) + 6.12 × (19^2-20^2) - 0.101 × (19^3-20^3)	1.5591

Two regressions: low and high % of English learners (model 4)

FIGURE 6.11 Regression Functions for Districts with High and Low Percentages of English Learners

Districts with low percentages of English learners ($HiEL = 0$) are shown by gray dots and districts with $HiEL = 1$ are shown by colored dots. The cubic regression function for $HiEL = 1$ from regression (6) in Table 6.2 is approximately 10 points below the cubic regression function for $HiEL = 0$ for $17 \leq STR \leq 23$, but otherwise the two functions have similar shapes and slopes in this range. The slopes of the regression functions differ most for very large and small values of STR, where there are few observations.



Summary of the empirical application

- The empirical analysis tried to provide answers to the following questions:
- Does the effect on test scores of reducing STR depends on the value of STR, after controlling for the observables?
  - After controlling for economic background, there is evidence of a nonlinear effect on test scores of the student–teacher ratio. This effect is statistically significant at the 1% level (the coefficients on STR^2 and STR^2 are always significant at the 1% level).
- Does the effect on test scores of reducing STR depends on the % of English learners, after controlling for the observables?
  - After controlling for economic background, whether there are many or few English learners in the district does not have a substantial influence on the effect on final test scores of a change in the student–teacher ratio. Although statistically significant in the nonlinear specifications, the effect is minimal in the region of STR comprising most of the data.

Summary of the empirical application

- After taking economic factors and nonlinearities into account, what is the estimated effect on test scores of reducing the student–teacher ratio by **one** student per teacher?
  - In the linear specification, this effect does not depend on the student–teacher ratio itself, and the estimated effect of this reduction is to improve test scores by 0.73 points.
  - In the nonlinear specifications, this effect depends on the value of the student–teacher ratio. If the district currently has a STR of 20, then cutting it to 19 has an estimated effect, based on regression (3) in the first table, of improving test scores by 1.57 points, while based on regression (4) the estimate is 1.76 points. If the district currently has a STR student–teacher ratio of 22, then cutting it to 21 has sharply more modest effect: 0.69 and 0.93 points, respectively. The estimates from the nonlinear specifications suggest that cutting the student–teacher ratio has a greater effect if this ratio is already small.

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 proceeds 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 the change in Y "before and after" a change in X.
- Many nonlinear specifications are possible, so you must use judgment:
  - What nonlinear effect you want to analyze?
  - What makes sense in your application?