Nonlinear Relationships: Quadratic and Polynomial Regression

19 Oct 2026 | Chapter 8

ECO 351 – Introduction to Econometrics

Shi Qiu – Fall 2026

The next restriction: a constant slope

Lecture 16 compared restricted fits and used causal diagrams to choose controls. We now examine a restriction on the *shape* of the regression.

Let x be district income in thousands of dollars. A fitted straight line has intercept a, slope b, and fitted score

$$\widehat{f}(x) = a + bx.$$

Discuss with a partner

1. Does this line predict the same score gain for \$10,000 to \$11,000 as for \$40,000 to \$41,000?
2. What would have to change if the observed relationship became flatter at higher incomes?

Equal income increases under a straight line

For the fitted line $\widehat{f}(x) = a + bx$, both \$1,000 increases give

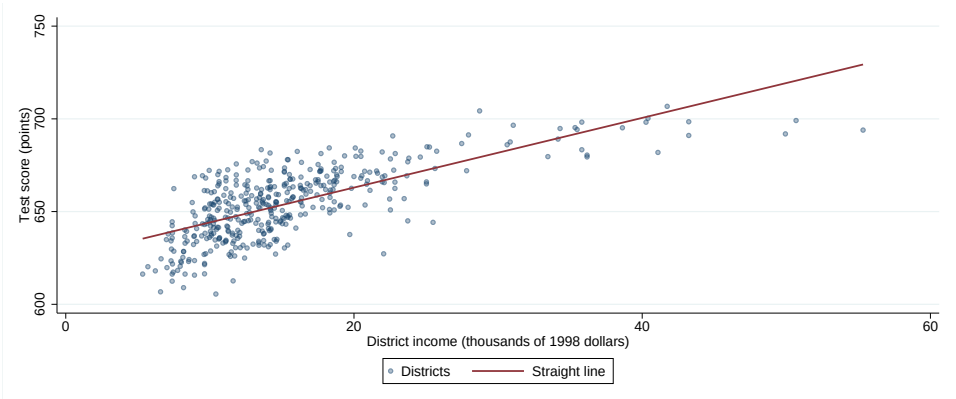
$$\widehat{f}(11) - \widehat{f}(10) = (a + 11b) - (a + 10b) = b,$$
$$\widehat{f}(41) - \widehat{f}(40) = (a + 41b) - (a + 40b) = b.$$

If the relationship flattens at higher incomes, one constant slope cannot capture it.

We will use the same 420 California districts to ask whether income and scores follow a curve. Income measures economic background, not school spending per pupil.

First inspect the data. Then fit a curve and interpret changes between predictions.

District income and scores: the straight-line fit



420 California districts; income is thousands of 1998 dollars per person. Compare observations at low, middle, and high incomes.

A quadratic regression with income and income squared

For district i, $Income_i$ is income in thousands of dollars, $Score_i$ is test-score points, and u_i contains other score determinants.

$$Score_i = \beta_0 + \beta_1 Income_i + \beta_2 Income_i^2 + u_i.$$

- Add the square of income as one more regressor. The coefficients are still estimated by OLS.
- β_0 is the intercept; β_1 and β_2 jointly determine the shape.

When income changes, its square changes too. We cannot interpret the income slope by holding income squared fixed.

Stata: estimate the quadratic on the same data

```
. regress score c.income##c.income, vce(robust)
Linear regression                               Number of obs     =        420
                                                R-squared         =     0.5562
------------------------------------------------------------------------------
             |               Robust
       score | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      income |   3.850995   .2680941    14.36   0.000      3.32401    4.377979
    c.income#|
    c.income |  -.0423085   .0047803    -8.85   0.000     -.051705   -.0329119
       _cons |   607.3017   2.901754   209.29   0.000     601.5978    613.0056
------------------------------------------------------------------------------
```

`c.income##c.income` includes income and its square. The fitted shape depends on both estimated slopes.

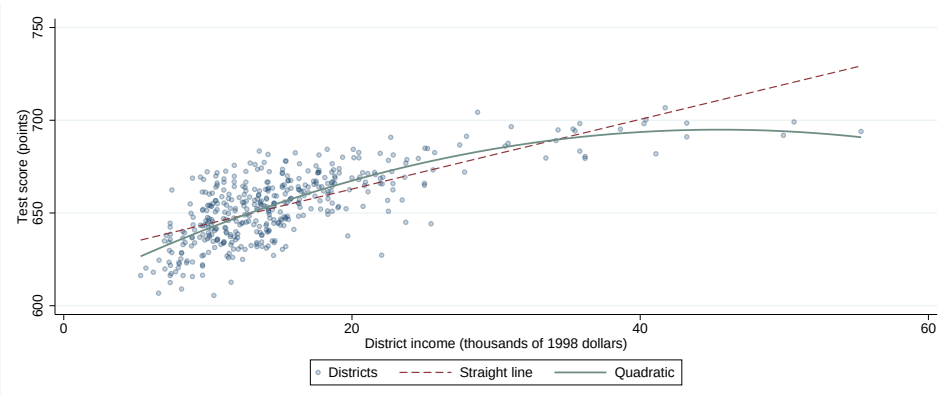
Linearity as a restriction on the quadratic

The restricted model sets $\beta_2 = 0$ and retains the intercept and income. The unrestricted quadratic also includes income squared.

```
. test c.income#c.income
 ( 1)  c.income#c.income = 0
       F(  1,   417) =   78.33
            Prob > F =    0.0000
```

Reject linearity against this quadratic alternative. This test does not establish that a quadratic is the best description among all possible shapes.

The fitted curve describes the middle of the data better



The dashed straight line misses the flattening at higher incomes. The solid quadratic allows the fitted gain to depend on the starting income. How should we measure that gain?

Interpret a curve by comparing two predictions

Let x be district income in thousands of dollars. Using rounded coefficients, the fitted score is

$$\widehat{Score}(x) = 607.3 + 3.85x - 0.0423x^2.$$

1. Choose the starting and ending incomes.
2. Calculate the predicted score at each income.
3. Subtract: predicted score after minus predicted score before.

The intercept cancels. Both income terms change. The starting income matters.

The same income increase can have different fitted gains

Income change	Fitted scores: before $\rightarrow$ after	Gain
\$10,000 to \$11,000	641.570 $\rightarrow$ 644.532	2.9617
\$40,000 to \$41,000	693.620 $\rightarrow$ 694.044	0.4237

For the first comparison,

$$\begin{aligned}\widehat{Score}(11) - \widehat{Score}(10) &= 3.85(11-10) - 0.0423(11^2 - 10^2) \\ &= 3.85 - 0.0423(21) = 2.9617.\end{aligned}$$

Both changes add \$1,000. The fitted gain is smaller at the higher starting income.

Your turn: training hours and a changing gain

Constructed practice model – 5 minutes

A fitted productivity score for workers is

$$\widehat{g}(H) = 50 + 4H - 0.10H^2,$$

where H is training hours and the outcome is score points.

1. Calculate the fitted change from 5 to 7 hours.
2. Calculate the fitted change from 15 to 17 hours.
3. Explain why these equal-sized changes give different answers.

Debrief: the starting value changes the answer

$$\widehat{g}(7) - \widehat{g}(5) = 8 - 0.10(24) = 5.6,$$
$$\widehat{g}(17) - \widehat{g}(15) = 8 - 0.10(64) = 1.6.$$

Both comparisons add 2 hours. At the higher starting value, the negative squared term offsets more of the linear contribution.

A fitted change needs a starting point, an ending point, and units. A raw polynomial coefficient is not enough.

Stata: uncertainty for the 10-to-11 comparison

Since $11 - 10 = 1$ and $11^2 - 10^2 = 21$, the fitted gain is $\widehat{\beta}_1 + 21\widehat{\beta}_2$. Stata can calculate its uncertainty:

```
. lincom income + 21*c.income#c.income
 ( 1)  income + 21*c.income#c.income = 0
------------------------------------------------------------------------------
       score | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         (1) |   2.962517   .1710575    17.32   0.000     2.626275    3.298759
------------------------------------------------------------------------------
```

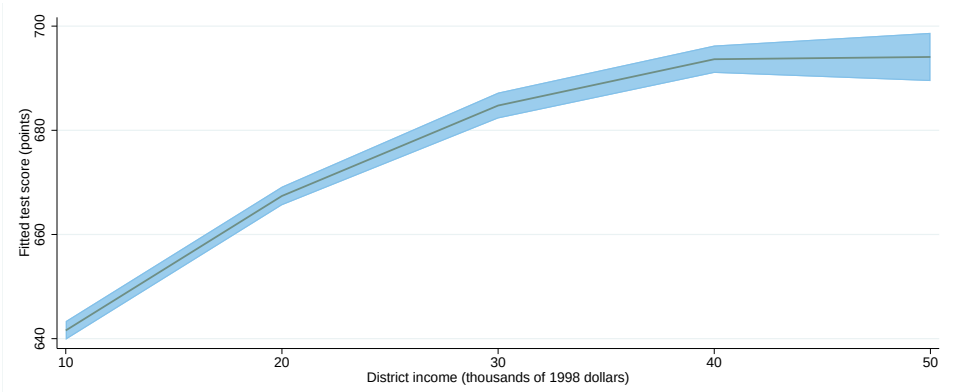
The gain is 2.963 points, with a 95% interval of $[2.626, 3.299]$. This uses full-precision estimates, not the rounded classroom equation.

Stata: predicted levels at specified incomes

```
. margins, at(income=(10 20 30 40 50))
------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   641.5808   .9095988   705.34   0.000     639.7929    643.3688
          2  |   667.3982   .9107629   732.79   0.000      665.608    669.1885
          3  |    684.754   1.264299   541.61   0.000     682.2688    687.2392
          4  |    693.648   1.338124   518.37   0.000     691.0177    696.2783
          5  |   694.0803   2.355118   294.71   0.000     689.4509    698.7097
------------------------------------------------------------------------------
```

Rows 1–5 correspond to incomes 10, 20, 30, 40, 50. Notice how the gains become smaller as income rises.

Stata graph: fitted levels and pointwise intervals



`marginsplot` displays the preceding predictions. The shaded band is a 95% interval for the fitted mean at each income, not a prediction interval for a new district.

A cubic adds the third power of income

For income x, the cubic keeps the lower powers and adds x^3:

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3.$$

```
. regress score c.income c.income#c.income ///
>     c.income#c.income#c.income, vce(robust)
------------------------------------------------------------------------------
             |               Robust
       score | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      income |   5.018677   .7073504     7.10   0.000      3.62825    6.409103
    c.income#|
    c.income |  -.0958052   .0289537    -3.31   0.001     -.152719   -.0388913
    c.income#|
    c.income#|
    c.income |   .0006855   .0003471     1.98   0.049     3.26e-06    .0013677
       _cons |    600.079   5.102062   117.61   0.000     590.0499     610.108
------------------------------------------------------------------------------
```

Quadratic or straight line: two restricted shapes

The unrestricted model contains an intercept, income, income squared, and income cubed.

Pair task – 3 minutes

1. Write the restriction that leaves a quadratic. Which income terms remain?
2. Write the restrictions that leave a straight line. Why should the income coefficient remain free?
3. If we fail to reject the cubic term alone, have we shown that a straight line is adequate? Explain.

Decide which simpler shape each test actually compares with the cubic.

Evidence against a line and evidence for a cubic

```
. test c.income#c.income#c.income
 ( 1)  c.income#c.income#c.income = 0
       F(  1,   416) =    3.90
            Prob > F =    0.0489

. test c.income#c.income c.income#c.income#c.income
 ( 1)  c.income#c.income = 0
 ( 2)  c.income#c.income#c.income = 0
       F(  2,   416) =   37.69
            Prob > F =    0.0000
```

The cubic term just passes a 5% test ($p = 0.0489$). The joint test gives much stronger evidence against a straight line.

Strong evidence of curvature can coexist with weak evidence for the extra cubic term. The tests compare different restricted shapes.

Choosing the polynomial degree

1. Start with a modest maximum degree, chosen for a substantive reason.
2. Test the highest power; if it is not needed, remove it and re-estimate.
3. Keep lower powers when a higher power remains. Inspect the fitted shape and where the data are sparse.
4. Report meaningful prediction differences, not just polynomial coefficients.

Starting from the cubic here, a 5% highest-power rule retains the cubic term. Report that this decision is close to the cutoff.

Lower SSR alone cannot choose the degree: adding powers gives OLS more possible fits. Check the shape and avoid extrapolating beyond the observed incomes.

A fitted gain at a new starting income

Exit calculation – 1 minute

$$\widehat{Score}(x) = 607.3 + 3.85x - 0.0423x^2,$$

with income x in thousands of dollars. Calculate the fitted change from \$20,000 to \$22,000. Why is $2(3.85)$ insufficient?

Debrief

$$3.85(22 - 20) - 0.0423(22^2 - 20^2) = 7.7 - 3.5532 = 4.1468.$$

The fitted gain is about 4.15 test-score points. Both income terms change.

This is a fitted association, not an identified causal income effect. Next: logarithms and proportional changes.

Optional B1: A local slope is a rate, not an exact finite change

Let $\widehat{f}(x)$ be the fitted score at income x; h is an income change. Both x and h are in thousands of dollars. For the quadratic,

$$\widehat{f}'(x) = 3.85 - 2(0.0423)x = 3.85 - 0.0846x.$$

Starting income	Local slope	Exact gain for $h = 1$
$x = 10$	3.004	2.9617
$x = 40$	0.466	0.4237

For a small change, the local slope times the change is an approximation. Subtract predictions for the exact finite comparison.

Optional B2: Stata: local slopes at two income levels

```
. margins, dydx(income) at(income=(10 40))
------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
income       |
         _at |
          1  |   3.004826   .1755929    17.11   0.000     2.659668    3.349983
          2  |   .4663179    .131894     3.54   0.000     .2070579    .7255779
------------------------------------------------------------------------------
```

Row 1 is $x = 10$; row 2 is $x = 40$. These are local slopes, not the changes from 10 to 11 or 40 to 41.

Optional B3: A fitted turning point is not an economic threshold

For the rounded quadratic, set its local slope to zero:

$$3.85 - 0.0846x = 0 \quad \Rightarrow \quad x \approx 45.51.$$

This is about \$45, 510 in the income units of the data. The negative squared term makes it a fitted maximum.

- Upper-income observations are sparse; another plausible function may differ there.
- Flexible shape does not remove confounding.

Inspect the data, estimate a plausible function, test its restrictions, and interpret changes between predictions. Next: logarithms and proportional changes.