Nonlinear Applications and an Introduction to IV

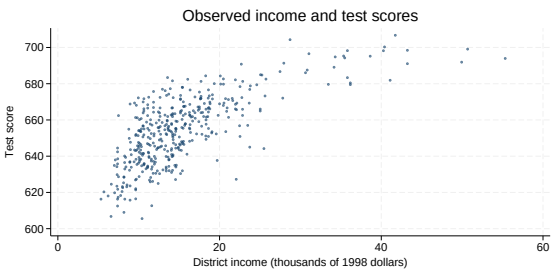
28 Oct 2026 | Chapters 8 and 12

ECO 351 – Introduction to Econometrics

Shi Qiu – Fall 2026

Two different questions about district income

We observe 420 California school districts. Scores are points; income I is annual income per person in thousands of 1998 dollars.



Fitted comparison
How do predicted scores differ at incomes 10 and 15, with the same controls?

Causal question
What would change if income rose while other score determinants stayed fixed?

Last class, interactions allowed a comparison to differ across groups. Here a square allows it to differ with starting income.

Write the fitted comparison first

Let S be students per teacher and E the percentage of English learners. Our fitted score is

$$\widehat{f}(I, S, E) = \widehat{a}_0 + \widehat{a}_1 I + \widehat{a}_2 I^2 + \widehat{a}_3 S + \widehat{a}_4 E.$$

The $\widehat{a}$'s are estimated coefficients. Compare predictions at the *same* S and E:

$$\widehat{f}(15, S, E) - \widehat{f}(10, S, E) = 5\widehat{a}_1 + 125\widehat{a}_2.$$

Regressor	Before	After	Change
I	10	15	5
I^2	100	225	125
S, E	Same	Same	0

Both income regressors change. The intercept and additive controls cancel.

Read the estimates before calculating

```
. regress score c.income##c.income str el, vce(robust)
Linear regression                               Number of obs     =        420
------------------------------------------------------------------------------
                     |               Robust
               score | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
---------------------+--------------------------------------------------------------
              income |   2.665477   .2335836    11.41   0.000     2.206323    3.124632
 c.income#c.income |  -.0247843   .0042885    -5.78   0.000    -.0332143   -.0163543
                str |  -.2564803   .2952173    -0.87   0.385     -.836788    .3238274
                 el |  -.4536707   .0305133   -14.87   0.000    -.5136505   -.3936908
              _cons |   632.6268   5.828567   108.54   0.000     621.1696     644.084
------------------------------------------------------------------------------
```

`c.income##c.income` includes income and income squared.

Use *both* income rows. Scores remain in points, and income remains in thousands of dollars.

Work in pairs: calculate and rewrite the claim

Use $\widehat{a}_1 = 2.665477$ and $\widehat{a}_2 = -0.0247843$, holding S and E fixed.

Income comparison	Change in I	Change in I^2
10 to 15	5	125
30 to 35	?	?

1. Calculate the fitted score gain for each comparison.
2. Explain why equal dollar increases need not give equal fitted gains.
3. Rewrite: "Every \$5,000 income increase causes a 10.23-point score gain."

One partner checks the arithmetic; the other checks units and the causal language.

First comparison: income 10 to 15

```
. lincom 5*income + 125*c.income#c.income
 ( 1)  5*income + 125*c.income#c.income = 0
------------------------------------------------------------------------------
       score | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         (1) |   10.22935   .6624466    15.44   0.000     8.927185    11.53152
------------------------------------------------------------------------------
```

$$5(2.665477) + 125(-0.0247843) \approx 10.23 \text{ points.}$$

Holding S and E fixed, the fitted gain is 10.23 points; its 95% interval is about 8.93 to 11.53.

This interval concerns a fitted mean difference, not the score of a new district.

Second comparison: income 30 to 35

$$\widehat{\Delta} = 5\widehat{a}_1 + (35^2 - 30^2)\widehat{a}_2 = 5\widehat{a}_1 + 325\widehat{a}_2.$$

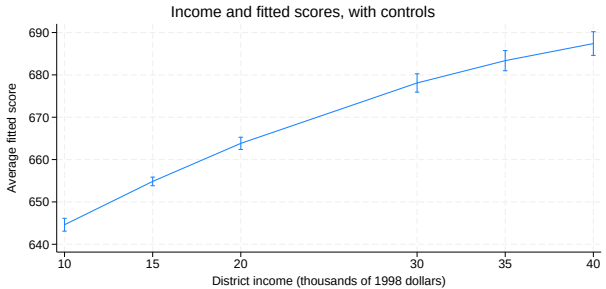
```
. lincom 5*income + 325*c.income#c.income
 ( 1)  5*income + 325*c.income#c.income = 0
------------------------------------------------------------------------------
       score | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         (1) |   5.272501   .3923998    13.44   0.000     4.501162     6.04384
------------------------------------------------------------------------------
```

The fitted gain is 5.27 points, with a 95% interval of about 4.50 to 6.04.

A defensible report: "The fitted score is about 5.27 points higher at income 35 than at 30, holding S and E fixed."

The curve shows why the gains differ

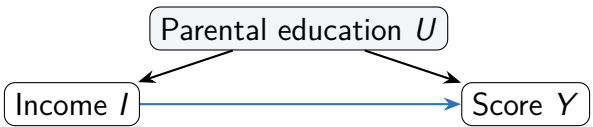
```
margins, at(income=(10 15 20 30 35 40))
marginsplot
```



The curve averages fitted scores over the *same* observed controls at each income. The bars are 95% intervals for fitted means.

Discuss: what if an important factor remains unmeasured?

Suppose parental education U raises scores and remains associated with district income after holding S and E fixed.



1. Name the open back-door path in this *hypothetical* causal story.
2. Would adding I^2 block it? Would a narrower confidence interval block it?
3. What information or design would you need before making a causal claim?

Debrief: the structural error can still move with income

For a causal model, let u contain other score determinants, including unmeasured parental education:

$$Y = \beta_0 + \beta_1 I + \beta_2 I^2 + \beta_3 S + \beta_4 E + u.$$

The β's describe the structural model. They need not equal the coefficients of the fitted association.

The identifying requirement

A familiar sufficient condition is that other score determinants average to zero at each income and control combination: $E(u \mid I, S, E) = 0$.

Income is **endogenous** when it remains related to the structural error in the relevant comparison. Curvature and precision do not remove that relationship.

Workshop: could a demand relationship look positive?

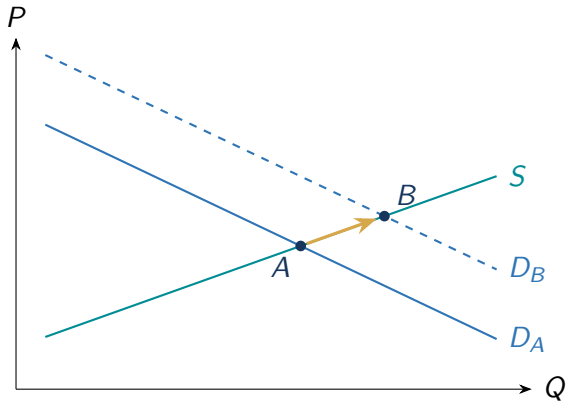
Two constructed markets have the following equilibrium outcomes. Price and quantity are in illustrative units.

Market	Price	Quantity
A	2.06	2.43
B	2.49	3.14

1. What sign does the observed price–quantity comparison have?
2. Could both observations be consistent with downward-sloping demand?
3. Sketch a change that raises *both* equilibrium price and quantity. Which curve shifts?

First describe an economic mechanism. The two observations alone cannot tell you which curve moved.

Debrief: a demand shock also changes the market price



Constructed markets; supply stays fixed

For this constructed example, demand is

$$Q = 5 - 1.25P + u_D.$$

u_D is an omitted demand shift.

Raising u_D shifts demand out. Equilibrium price and quantity rise together.

The structural price slope is negative even though the two observed points slope upward.

Price moves with the demand error. Comparing these prices does not hold demand determinants fixed.

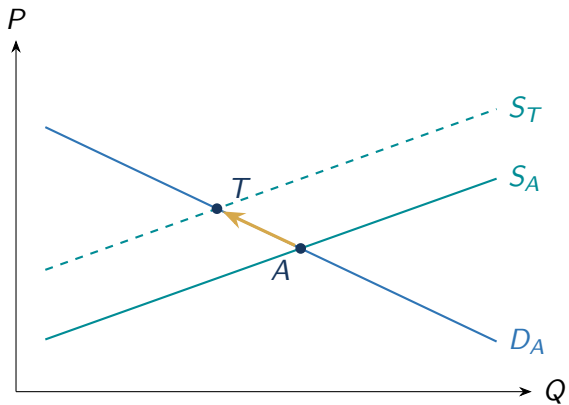
Which price change would be useful?

The target is the response of demand to price. Consider three possible changes:

A	A new advertising campaign makes the product more attractive.
B	A per-unit tax raises sellers' costs; preferences and other demand determinants stay unchanged.
C	The same tax arrives with a campaign discouraging consumption.

1. Which change could trace a fixed demand curve? Explain which curve moves.
2. Why does merely observing a price change not settle the issue?
3. For the actual tax proposal, what part of the story would need evidence?

Debrief: move price while holding demand fixed



Constructed markets; demand stays fixed

A cost increase shifts supply up.

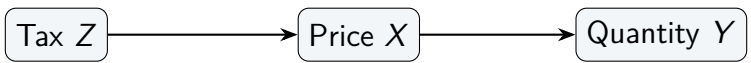
With demand unchanged, the market moves from A to T *along the same demand curve*.

The tax is useful only if it leaves other demand determinants unchanged.

The advertising stories shift demand too. A tax's effect on price does not prove that it supplies a valid demand comparison.

Name the idea: an instrumental variable

Let Y be quantity demanded, X price, and Z a proposed tax instrument.



Relevance Z changes or predicts X.

Exogeneity Z is unrelated to the structural demand error.

Exclusion Z has no outcome route outside the modeled price channel.

An instrument supplies a *defended source* of price variation. Predicting price is only one part of that argument.

From a fitted answer to a causal comparison

Question	What the workshop established
How much is the fitted gain?	Use both income terms and state the controls and units.
Why might it be noncausal?	The regressor can move with omitted outcome determinants.
What could help?	A relevant source of regressor variation that leaves those determinants unchanged.

Next class: evaluate a cigarette-tax instrument, measure the price and sales responses, and connect them to an IV estimate.