## Example #1: Effect of Studying on Grades

What is the effect on grades of studying for an additional hour per day?

$Y$ = GPA

$X$ = study time (hours per day)

Data: grades and study hours of college freshmen.

*Would you expect the OLS estimator of $\beta_1$ (the effect on GPA of studying an extra hour per day) to be unbiased? Why or why not?*

## Studying on grades, ctd.

Stinebrickner, Ralph and Stinebrickner, Todd R. (2008) "The Causal Effect of Studying on Academic Performance," *The B.E. Journal of Economic Analysis & Policy*: Vol. 8: Iss. 1 (Frontiers), Article 14.

- $n$ = 210 freshman at Berea College (Kentucky) in 2001
- $Y$ = first-semester GPA
- $X$ = average study hours per day (time use survey)
- Roommates were randomly assigned
- $Z$ = 1 if roommate brought video game, = 0 otherwise

Do you think $Z_i$ (whether a roommate brought a video game) is a valid instrument?

1. Is it relevant (correlated with $X$)?
2. Is it exogenous (uncorrelated with $u$)?

**Studying on grades, ctd.**

$$X = \pi_0 + \pi_1 Z + v_i$$

$$Y_i = \beta_0 + \beta_1 \hat{X}_i + u_i$$

*Y* = *GPA (*4 point scale)

*X = time spent studying (hours per day)*

*Z = 1 if roommate brought video game, = 0 otherwise*

<u>Stinebrinckner and Stinebrinckner's findings</u>

$$\hat{\pi}_1 = -.668$$

$$\hat{\beta}_1^{IV} = 0.360$$

Do these estimates make sense in a real-world way? (*Note*: They actually ran the regressions including additional regressors – more on this later.)

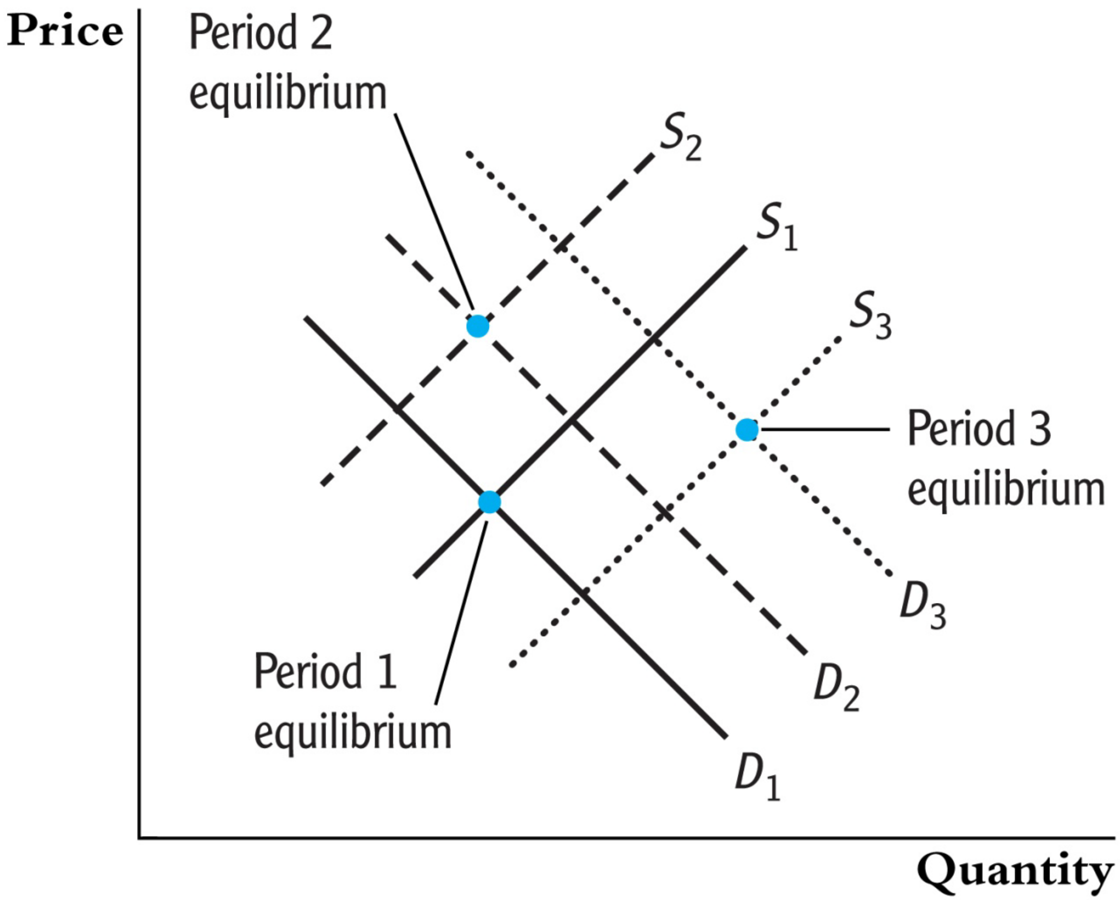
**Example #2: Supply and demand for butter**

IV regression was first developed to estimate demand elasticities for agricultural goods, for example, butter:

$$\ln(Q_i) = \beta_0 + \beta_1 \ln(P_i) + u_i$$

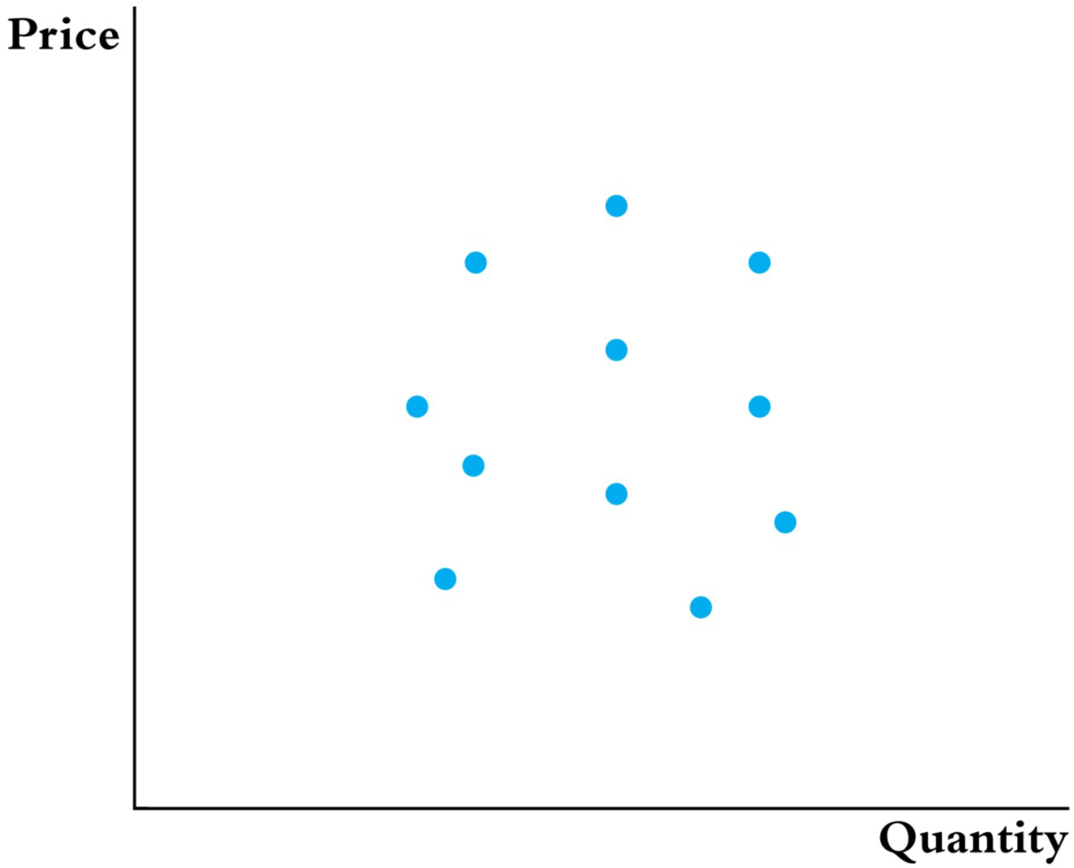
- $\beta_1$ = price elasticity of butter = percent change in quantity for a 1% change in price (recall log-log specification discussion)
- Data: observations on price and quantity of butter for different years
- The OLS regression of $\ln(Q_i)$ on $\ln(P_i)$ suffers from simultaneous causality bias (*why*?)

Simultaneous causality bias in the OLS regression of $\ln(Q_i)$ on $\ln(P_i)$ arises because price and quantity are determined by the interaction of demand *and* supply:




**(a)** Demand and supply in three time periods

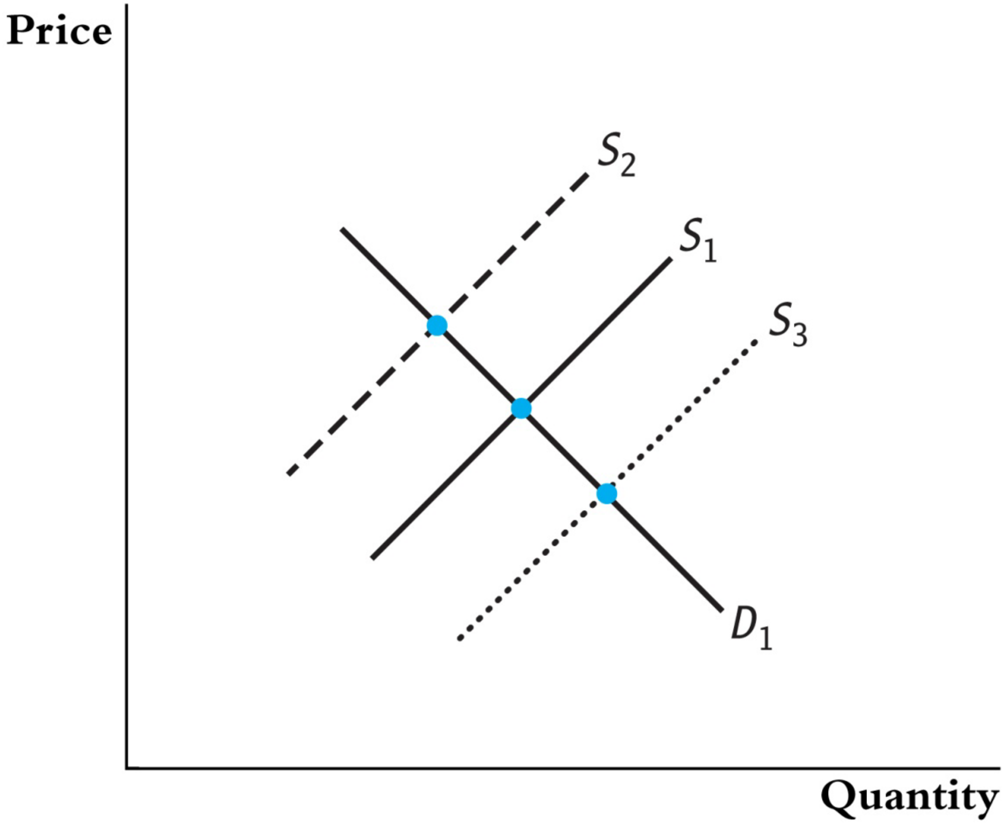
This interaction of demand and supply produces data like…




**(b)** Equilibrium price and quantity for 11 time periods

*Would a regression using these data produce the demand curve?*

But…what would you get if only supply shifted?




**(c)** Equilibrium price and quantity when only the supply curve shifts

- TSLS estimates the demand curve by isolating shifts in price and quantity that arise from shifts in supply.
- $Z$ is a variable that shifts supply but not demand.

**TSLS in the supply-demand example:**

$$\ln(Q_i) = \beta_0 + \beta_1 l \ln(P_i) + u_i$$

Let $Z$ = rainfall in dairy-producing regions.
Is $Z$ a valid instrument?

(1) Relevant? $\text{corr}(rain_i, \ln(P_i)) \neq 0$?
*Plausibly*: insufficient rainfall means less grazing means less butter means higher prices

(2) Exogenous? $\text{corr}(rain_i, u_i) = 0$?
*Plausibly*: whether it rains in dairy-producing regions should not affect demand for butter

# TSLS in the supply-demand example, ctd.

$$\ln(Q_i) = \beta_0 + \beta_1 \ln(P_i) + u_i$$

$Z_i$ = $rain_i$ = rainfall in dairy-producing regions.

Stage 1: regress $\ln(P_i)$ on *rain*, get $\widehat{ln(P\imath)}$

$\widehat{ln(P\imath)}$ isolates changes in log price that arise from supply (part of supply, at least)

Stage 2: regress $\ln(Q_i)$ on $\widehat{ln(P\imath)}$

The regression counterpart of using shifts in the supply curve to trace out the demand curve.