Problem Set 6
Due Thursday, June 2 in class
ECN 140—Spring 2016

Question 1

The data set PWT63.dta contains data on per capita consumption (cons) and GDP (rgdpl), for many countries for many years. You're interested in determining how much consumption rises due to an increase in output.

a. Estimate the consumption function (be sure to report robust standard errors):

$$\log(cons_{it}) = \alpha + \beta \log(rgdpl_{it}) + \epsilon_{it}$$

where i represents country and t represents time. You'll need to generate the log forms of these variables. Report your estimate for β. Now allow for country fixed effects so you're regression looks like this:

$$\log(const_{it}) = \alpha + \beta \log(rgdpl_{it}) + c_i + \epsilon_{it}$$

To do this, you'll need to transform the country variable into a dummy variable before you can use country fixed effects because currently it is coded as a continuous variable.

b. Describe (in a way that a non-statistician could understand) the *theoretical* differences (not the differences in your numerical output) between running your regression with country fixed effects and without. Basically, tell me what country fixed effects are controlling for in your regression.

c. Why might you also want to include time fixed effects? What happens if you do this?

Question 2

Let's try a system of simultaneous equations. Use the "Hw5" data. Suppose the system is

$$P = 6 - 4Q + S + \varepsilon \qquad \text{demand}$$
$$Q = 6 + 2P + u \qquad \text{supply}$$

where P and Q are prices and quantities of a certain book over time, S is the book's rating on the New York Times Bestseller list and u and ε are random shocks. Also assume that the book's rating (S) only affects people's demand for the book and not the supply.

a. First, try estimating the demand equation using OLS. What's your estimate for the coefficient on quantity? What is your estimate for the coefficient on the book's rating?

Do these look correct? In other words, compare your estimates for the slope and intercept to the ones in the above demand equation.

b. Second, try estimating the supply equation using OLS, what is your estimate for the coefficient on price? Does this look correct?

c. Why might using OLS (ordinary least squares) not be a good idea in this setting?

d. To overcome the problem highlighted in part (c), you decide to use S as an instrumental variable (IV). Think carefully about which curve S shifts. Based on the curve S shifts, what curve can you estimate (aka trace out), the supply curve or the demand curve?

e. Based on your answer from part (d), now try using instrumental variables (i.e., two-stage least squares, where the stata command is "ivreg") to estimate this equation. Do your answers look better than in parts (1) and (2)?

Question 3

1. Describe why you would need to use an instrumental variables estimation strategy?
2. What makes an instrument valid?
3. What is the intuition behind IV estimation?

Question 4

Use **card_minwages.dta**. This dataset is by a paper which tried to determine whether raising the minimum wage reduces employment. They used a natural experiment where New Jersey raised the minimum wage but Pennsylvania did not. The dataset has variables for employment in individual fast-food restaurants before and after the policy change.

(a) Explain why a difference-in-difference regression is appropriate here. Why can't we just regress employment on the value of the minimum wage?

(b) Create a table like we did in class with just means. What is effect if raising the minimum wage? (Hint: it's the difference-in-difference estimate)

(c) Estimate the same effect in a regression framework. Is the effect statistically significant?

Question 5

Use **INJURY.dta**. Our question is does increasing the weekly earnings cap on workers compensation increase the number of weeks that they claim the benefit?

- Kentucky and Michigan increased the length of time
  - Affected high-income workers
  - Did not affect low-income workers

A. Why would we expect that increasing the weekly earnings cap would increase the # of weeks workers claim the benefit?
B. Calculate the effect of the policy using only means (the difference-in-difference coefficient, but not from a regression).
C. Do it in a regression framework. Is the effect statistically significant?