Problem set 13: week 17

This problem set is to be completed prior to your Stata class. The data set used smoke.dta are available on the EC203 website, please copy it on to a memory stick or your H: drive. For help on each command use either: **help**, **findit** or the **search** function for guidance on each. This problem set relates to simultaneity and IV estimation.

Suppose you are interested in estimating the effect of Y_2 on Y_1, however, it is also known that Y_1 has an effect on Y_2. That is, we have a problem of simultaneity since Y_1 and Y_2 are jointly determined. As with omitted relevant variables, such situations can cause concerns for bias. Suppose the relationship between Y_1 and Y_2 is represented by the following;

$$Y_{i1} = \alpha_1 + \delta_1 Y_{i2} + u_{i1} \ (A)$$
$$Y_{i2} = \alpha_2 + \delta_2 Y_{i1} + \beta_2 Z_{i1} + u_{i2} \ (B)$$

Where we assume Z_1 is exogenous, such that it is uncorrelated with both u_1 and u_2.

1. Suppose we are interested in getting an unbiased estimate of δ_1. As usual this can be achieved if Y_2 is exogenous, that is, if $E[u_1|Y_2] = 0$. Solve the above equations for Y_2 and show that Y_2 is not exogenous if $\delta_2 \neq 0$.

2. Suppose we are interested in getting an unbiased estimate of δ_2. As usual this can be achieved if Y_1 is exogenous, that is, if $E[u_2|Y_1] = 0$. Solve the above equations for Y_1 and show that Y_1 is not exogenous if $\delta_1 \neq 0$.

3. Parts 1 and 2 show if our relationship suffers from simultaneity we cannot get unbiased estimates if we run OLS on either equation (A) or equation (B). One possible answer is to use IV (2SLS) estimation. Consider the following questions:

- suppose our equation of interest is (A), and in particular we want to estimate δ_1. That is, in the system of equations are there any possible IVs available for Y_2?
- suppose our equation of interest is (B), and in particular we want to estimate δ_2. That is, in the system of equations are there any possible IVs available for Y_1?

In answering the above 2 questions it will help to think in terms of **instrument exogeneity**.

4. Now we will cover an empirical question closely related to the above situation. Open smoke.dta in Stata. Run the two following regressions:

$$\begin{aligned} ln(income)_i =& \beta_0 + \beta_1 cigs_i + \beta_2 educ_i + \beta_3 age_i + \beta_4 age_i^2 + u_{i1} \ (C) \\ cigs_i =& \delta_0 + \delta_1 ln(income)_i + \delta_2 educ_i + \delta_3 age_i + \delta_4 age_i^2 \\ & + \delta_5 ln(cigpric)_i + \delta_6 restaurn_i + u_{i2} \ (D) \end{aligned}$$

Note, the equivalence between these equations and equations (A) and (B) above, that is, $ln(income) = Y_{i1}$ and $cigs = Y_{i2}$.

- In regression (C) interpret the coefficient on cigs.
- In regression (D) interpret the coefficient on income.
- Explain how income and cigarettes might be jointly determined.
- Suppose it is the case that cigs and income are jointly determined, then we known from parts (1) and (2) OLS estimation of (C) and (D) are biased.

5. Suppose we are primarily interested in estimating the effect of smoking on income. What are possible IVs for cigs? Do you think they are exogenous and relevant?

6. Re-estimate (C) using IV (2sls) with your selected instruments for the cigs variables.

7. Annotate and close your do file.