Problem Set 1
Econometrics 718
Chris Taber
Due: Tues. Sept. 27

Problem 1. Take any data set you would like and verify the two part regression results (if you google “econometric data sets" there are a lot to choose from). That is, think about a multiple regression which you can separate the independent variables into to groups.

$$Y_i = X_{1i}'\beta_1 + X_{2i}'\beta_2 + u_i$$

Verify that you get exactly the same results for $\widehat{\beta}_1$ doing the two things:

a) A big multiple regression of Y_i on (X_{1i}, X_{2i}).

b) Two part regression where you first run Y_i and X_{1i} on X_{2i} and take residuals and then run the residuals from the Y regression on the residuals for the X regressions.

c) Now suppose that rather than the first set you just regressed Y_i on $\widetilde{X}_{1i}$. (where Y_i is the original data and $\widetilde{X}_{1i}$ is the residuals). Does that give you the same result? Why or why not?

Problem 2. Again with any data set you would like and with any software you would like think about the exactly identified IV problem. I would like you to produce the IV estimate in four different ways and show they are numerically equivalent (I don't care whether the instrument is really uncorrelated with the residual):

a) IV, that is $(Z'X)^{-1}Z'Y$ (you can use ivregress with the gmm option in stata)

b) Two staged least squares. That is first run the treatment variable on the instrument and the X's, form the predicted value, then run a regression of the outcome on the predicted value and the other X's.

c) Ratio of reduced form coefficients. Run the two reduced forms and take the ratio of the coefficients on the Z.

d) Literally use

$$\widehat{\alpha} = \frac{scov(\widetilde{Z}, \widetilde{Y})}{scov(\widetilde{Z}, \widetilde{T})}$$

where the tildes mean residuals after regressing on X, scov means sample covariance, and svar means sample variance.

Problem 3. Verify the measurement error result. That is I want you to use your statistical package to construct

- T_i to have whatever distribution you want (a uniform might be easy)
- $$Y_i = \beta_0 + \beta_1 T_i + u_i$$

  where u_i is $N(0, \sigma_u^2)$. You can choose these parameters to be anything you want.

- $$\tau_{1i} = T_i + \xi_i$$

  where ξ_i is $N(0, \sigma^2_\xi)$

- $$\tau_{2i} = T_i + \eta_i$$

  where η_i is $N(0, \sigma^2_\eta)$

Show that

a) If you run a regression of Y_i on T_i you get something close to β_1

b) If you run a regression of Y_i on τ_{1i} you get something close to

$$\beta_1 \frac{Var(T_i)}{Var(T_i) + \sigma^2_\xi}$$

c) Doing IV using τ_{2i} as an instrument for τ_{1i} gives an estimate close to β_1