Computer Homework 6

This homework is concerned with IV estimation. Use the *Stata* data file card.dta to estimate the following model:

 $lwage = \alpha + \beta educ + \gamma exper + \delta expersq + \theta black + \lambda smsa + \tau south + \zeta married + \pi IQ + \kappa KWW + \varepsilon$

The variable names used here are identical to those in the data file. Note the use of CAPs in some of these labels. The variables are log of wages (*lwage*), education in years (*educ*), potential work experience and its square (*exper* and *expersq*), binaries for *black*, those living in an *smsa*, those living in the *south*, and those that are *married*, *IQ* score, and Knowledge of World of Work score (*KWW*).

It has been argued that *educ* is potentially correlated with ε because of difficulties in accurately measuring innate ability, quality of education, and other latent factors. Unmeasured factors related to ability are likely to be positively correlated with both educational attainment and wages. There are a number of commonly used instruments for education in this data set. Specifically, binaries for proximity to a 4 year college (*nearc4*) and proximity to a 2 year college (*nearc2*), mother's educational attainment (*motheduc*), and father's educational attainment (*fatheduc*). Note that *motheduc* and *fatheduc* are missing for some observations. Conduct all of the analysis requested below using only the subsample of complete observations.

1. Discuss the strengths and weaknesses of each of these instruments with respect to its relevance and its exogeniety.

Estimate the model with OLS. Then estimate the model with IV using each of the potential instruments individually.

- 2. Compare the point estimates and standard errors of β .
- 3. Construct and discuss 95% confidence intervals for β using OLS and each of the IV estimates.
- 4. Discuss the costs and benefits of IV estimation?

One would think that the arguments that would make *nearc4* a valid instrument would also apply to *nearc2*. The same can be said of *motheduc* and *fatheduc*.

- 5. How robust are the IV estimates of β to the choice of instruments in this example?
- 6. Compare the fit of the OLS and IV estimates. Are the results a surprise?
- 7. There are no missing observations for *nearc4* and *nearc2*. Use the entire sample to estimate the model with OLS, and with IV using *nearc4* and *nearc2* individually as instruments for *educ*. Compare the point estimates and standard errors of β obtained with the subsample of complete observations versus the entire sample. Is there some reason to expect the validity of the college proximity instruments to differ depending on whether the individual provides a response to a question on parental education?

Stata Tip

In *Stata*, if we want IV estimates of a simple model where *lwage* is regressed on *educ* and *exper*, and where *nearc4* is used as an instrument for *educ*, we would use the command line:

ivreg lwage exper (educ = nearc4)