ECON3360 Causal Inference for Microeconometrics

Tutorial 4: Stata application of IVs

Instructor: Julie Moschion

Problem I: the effect of education on fertility

Use the data in fertil2.dta for the following questions. These data include women in Botswana during 1988.

(1) Describe the variables in the data.

(2) Estimate the following model by OLS and interpret the estimates.

$$children = \beta_0 + \beta_1 educ + \beta_2 age + \beta_1 age^2 + u$$

In particular, holding *age* fixed, what is the estimated effect of another year of education on fertility? If 100 women receive another year of education, how many fewer children are they expected to have?

(3) The variable $frsthalf$ is a dummy variable equal to one if the woman was born during the first six months of the year. Is $frsthalf$ a reasonable IV candidate for *educ*?

(4) We want to estimate the effect of *educ* on *children*. Implement the IV method by 2SLS and compare IV estimates to the OLS estimates in (2).

Problem II: the effect of smoking on birth weight

Use the data in bwght.dta for the following questions. These data include observations for pregnant women in the US.

(1) Describe the variables in the data.

(2) Estimate the following model by OLS and interpret the estimates.

$$log(bwght) = \beta_0 + \beta_1 packs + \beta_2 faminc + u$$

We might worry that *packs* is correlated with other health behaviours (e.g. drinking alcohol), or the local availability of good hospital care, so that *packs* and the error term u might be correlated. To solve this potential issue, we want to examine *cigprice* as a possible instrumental variable for *packs*.

(3) Suppose that we assume that *cigprice* and u are uncorrelated. Is it good assumption? Can you somehow test this? (hint: use robust standard errors)

(4) Draw a scatter plot of *packs* on *cigprice*.

(5) Perform a reduced form regression. In other words, regress *lbwght* on *cigprice* (hint: use robust standard errors). Interpret the regression result. What may this imply for the first stage?

(6) Implement the IV method by 2SLS. How do the estimates compare with the OLS?

Problem III (Conditional IV): the effects of education on wages

Use the data in card.dta for the following questions. These data include men in the US during 1976. Card(1995) used wage and education data to estimate the return to education.

(1) Describe the variables in the data.

(2) Estimate the following model by OLS and interpret the estimates (hint: use robust standard errors).

$$ln(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \beta_4 smsa + \beta_5 south + u$$

Education is likely to be endogenous. Card proposes to use $near4$ as an instrument for the endogenous variable $educ$.

(3) Use a simple OLS regression to check whether $near4$ has an effect on educ (hint: use robust standard errors). Could $near4$ be correlated with factors in the error term?

(4) Implement the IV method by 2SLS. Interpret the coefficient of $educ$. Do you find a different coefficient when you use robust standard errors compared to when you don't use robust standard errors?

(5) For a sub-sample of the men in the data set, an IQ score is available. Do IQ scores vary by whether a man grew up near a four-year college? What do you conclude from that?

(6) Now use the OLS to regress IQ on $nearc4$, $smsa66$, and 1966 regional dummy variables $reg662$,...., $reg669$ (hint: use robust standard errors). Are IQ and $near4$ related after the geographic dummy variables have been controlled for?

(7) What do you conclude about the importance of controlling for $smsa66$ and the 1966 regional dummies in the wage equation?

(8) Now implement the IV method by 2SLS again while you control for the regional dummies (hint: use robust standard errors) and interpret the coefficient of $educ$.