ECOM30003/ECOM90003: Applied Microeconometric Modelling
Tutorial 7

Please read Chapter 15 of Wooldridge before attempting the following.

1. The following is a simple model to measure the effect of a school choice program on standardized test performance

$$score = \beta_0 + \beta_1 choice + \beta_2 faminc + u_1 \tag{1}$$

where *score* is the score on a statewide test, *choice* is a binary variable indicating whether a student attended a choice school in the last year $faminc$ is family income. The IV for *choice* is *grant*, the dollar amount granted to students to use for tuition at *choice* schools. The grant amount differed by family income level, which is why we control for $faminc$ in the equation.

   (a) Even with $faminc$ in the equation why might *choice* be correlated with u_1?

   (b) If within each income class, the grant amounts were assigned randomly, is grant uncorrelated with u_1.

   (c) Write the reduced form equation for *choice*. What is needed for *grant* to be partially correlated with choice?

   (d) Write the reduced form equation for *score*. Explain why this is useful.

2. Use the data CARD.dta for this exercise.

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \cdots + u \tag{2}$$

where the other explanatory variables are listed in Table 15.1. (educ; exper; exper2;black; smsa; south).

   (a) In order for the IV estimator to be consistent, the IV for *educ*, *nearc*4, must be uncorrelated with u. Could *nearc*4 be correlated with things in the error term, such as unobserved ability? Explain.

   (b) An IQ score is available for a subsample of men in the data set. Regress IQ on *nearc*4 to check whether average IQ scores vary by whether the man grew up near a four-year college. What do you conclude?

   (c) Now, use OLS to estimate the regression model for which IQ as the dependent variable and the explanatory variables are *nearc*4, *smsa*66, and the 1966 regional dummy variables *reg*662, ..., *reg*669. Are IQ and *nearc*4 related after the geographic dummy variables have been partialled out? Reconcile this with your findings from part (b).

(d) From parts (b) and (c), what do you conclude about the importance of controlling for $smsa66$ and the 1966 regional dummies in the $log(wage)$ equation?

3. Use the data in HTV.dta for this exercise.

(a) Run a simple OLS regression of $log(wage)$ on *educ* without controlling for other factors. What is the 95% confidence interval for the return to another year of education.

(b) The variable *ctuit*, measured in thousands of dollars, is the change in college tuition facing students from age 17 to age 18. Show that *educ* and *ctuit* are essentially uncorrelated. What does this say about *ctuit* as a possible IV for *educ* in a simple regression analysis?

(c) Now add to the simple regression model in part (a) a quadratic in experience and a full set of regional dummy variables for place of current residence and place of residence at age 18. Also include the urban indicators for current and age 18 residence. What is the estimated return to a year of education?

(d) Again using *ctuit* as a potential IV for *educ*, estimate the reduce form for *educ*. Show that *ctuit* is now statistically significant in the reduced form for *educ*.

(e) Estimate the model from part (c) by IV using *ctuit* as an IV for *educ*. How does the confidence interval for the return to education compare with the OLS confidence interval from part *c*?

(f) Do you think the IV procedure from part *e* is convincing?