Department of Economics
Columbia University

UN3412
Fall 2021

Problem Set 7
Introduction to Econometrics
Seyhan Erden and Tamrat Gashaw
for all sections.

1. (33p) Do you think attendance to lectures affects performance on final exam? A model to explain the standardized outcome on a final exam (*stndfnl*) in terms of percentage of classes attended, prior college grade point average, and ACT score is

$$stndfnl = \beta_0 + \beta_1\, atndrte + \beta_2\, priGPA + \beta_3\, ACT + u$$

where variables are given in the following table:

Variable	**Definition**
stndfnl	Standardized final exam score
atndrte	Percentage of classes attended
priGPA	Prior college grade point average
ACT	Achievement Test score
priGPAxatndrte	Prior GPA times attendance rate

(a) (10p) Let *dist* be the distance from the students' living quarters to the lecture hall. Do you think *dist* is uncorrelated with u?

(b) (10p) Assuming that *dist* and u are uncorrelated, what other assumption must *dist* satisfy in order to be a valid IV for *atndrte*?

(c) (13p) Suppose we add the interaction term *priGPAxatndrte:*
$stndfnl = \beta_0 + \beta_1\, atndrte + \beta_2\, priGPA + \beta_3\, ACT + \beta_4\, priGPAxatndrte + u$
If *atndrte* is correlated with u, in general, so is *priGPAxatndrte*. What might be a good IV for *priGPAxatndrte*? [Hint: if $E(u \mid priGPA, ACT, dist)=0$, as happens when *priGPA*, *ACT* and *dist* are all exogenous, then any function of *priGPA* and *dist* is uncorrelated with u]

2. (34p) The purpose of this question is to compare the estimates and standard errors obtained by correctly using 2SLS with those obtained using inappropriate procedures. Use the data file WAGE2.dta, variables are:

Variable	**Definition**
wage	Monthly earnings
educ	Years of education
exper	Years of working experience
tenure	Years with current employment
black	=1 if the person is African-American, 0 otherwise
sibs	Number of siblings

(a) (12p) Use a 2SLS routine to estimate the equation

Log(*wage*) = $\beta_0 + \beta_1$ *educ* $+ \beta_2$ *exper* + β_3 *tenure* + β_4 *black* + *u*
using *sibs* as the IV for *educ*. Report your results

(b) (12p) Now, manually, carry out 2SLS. That is, first regress *educ* on *sibs, exper, tenure* and *black*, obtain the fitted values *educ_hat*, then run the 2nd stage regression log(*wage*) on *educ_hat, exper, tenure* and *black*. Verify that estimated sample coefficients are identical to those obtained form part (a) but that the standard errors are somewhat different. The standard errors obtained from the second stage regression when manually carrying out 2SLS are generally inappropriate. Are the standard errors larger or smaller than those from Part a? Explain.

(c) (10p) Now, use the following two-step procedure, which generally yields inconsistent parameter estimates of β_j and not just inconsistent standard errors. In step one, regress *educ* on *sibs* only and obtain the fitted values, say *educ_tilde* (Note that this is an incorrect first stage regression). Then, in the second step, run the regression of log(*wage*) on *educ_tilde*, *exper*, *tenure* and *black*. How does the estimate from this incorrect, two-step procedure compare with the correct 2SLS estimate of the return to education?

3. (33p) By studying the probability limit (plim) of the IV estimator we can see that when *Z* and *u* are possibly correlated, we can write

$$plim\ \hat{\beta}_{1,IV} = \beta_1 + \frac{Corr(Z,u)}{Corr(Z,X)} \frac{\sigma_u}{\sigma_x} \qquad (1)$$

where σ_u and σ_x are the standard deviation of u and *X* in the population, respectively. The interesting part of this equation involves the correlation terms. It shows that, even if Corr(*Z*,*u*) is small, the inconsistency in the IV estimator can be very large if Corr(*Z*,*X*) is also small. Thus, even if we focus only on consistency, it is not necessarily better to use IV than OLS if the correlation between Z and u are smaller than that between X and u. Using the fact that $Corr(X,u) = Cov(X,u)/(\sigma_u.\sigma_x)$ along with the fact that plim $\hat{\beta}_1 = \beta_1$ + Cov(*X*,*u*)/Var(*X*) = β when $Cov(X,u) = 0$, we can write the plim of OLS estimator – call it $plim\ \hat{\beta}_{1,OLS}$ - as

$$plim\ \hat{\beta}_{1,OLS} = \beta_1 + Corr(X,u)\ \frac{\sigma_u}{\sigma_x} \qquad (2)$$

Assume that $\sigma_u = \sigma_x$, so that the population variance in the error term is the same as it is in X. Suppose the instrumental variable, Z, is slightly correlated with u: $Cov(Z,u) = 0.1$. Suppose also that *Z* and *X* have somewhat stronger correlation: $Cov(Z,X) = 0.2$.

(i) (17p) What is the bias in the asymptotic IV estimator?

(ii) (16p) How much correlation would have to exist between *X* and u before OLS has more asymptotic bias than TSLS?

The following questions will not be graded, they are for you to practice and will be discussed at recitation:

1. SW Exercise 12.2
2. SW Exercise 12.5
3. SW Exercise 12.7
4. SW Exercise 12.8
5. SW Exercise 12.10
6. SW Empirical Exercise 12.1