Hayden C. Hancock

GBA 6230, Assignment 7

Dr. Hwu

1. Consider the following model,

colGP A = β0 + β1skipped + β2sat + u

1. Discuss why skipped might be an endogenous variable here.

**Ans:** By definition, endogenous variables are those whose measure is determined by the model. Skipped in this case is a measure of the amount of classes skipped. Typically, in a college environment, there are in class assignments, quizzes, and attendance. If that is the case, it could be directly correlated with the dependent variable colGPA.

(b) Suppose we want to use dist, which denotes the distance from student’s

house to lecture hall, as our instrument variable, what assumption we need

for dist? Discuss your answer in detail.

**Ans:** In order for dist to be an instrument variable, we have to satisfy 2 requirements. Firstly, dist needs to be an observable variable and is exogenous, meaning it should have no partial effect on y and uncorrelated with the omitted variables. Secondly, it must be related to the endogenous variable skipped. Lastly dist only affects colGPA only through skipped.

2. Consider the following simple regression model,

y1 = β0 + α1y2 + u

where y2 is endogenous.

(a) Suppose I have an instrument variable z1, which turns out to be correlated

with u. Suppose I applied two stage least square, what kind of problem

we will have in the second stage regression? Prove your answer.

**Ans:** To get to the second stage regression we want to run a regression on the above equation and get the fitted value above for y2hat, once we replace the original y2 with y2hat we get the second equation. We then run a regression on that equation and get a consistent alpha 1 estimator. The problem here with the second stage regression is the fact that the rsquare can be negative. It concludes that rsquared in this case is not a useful model selection tool and the F test needs to be adjusted.

(b) Suppose I have another instrument variable z2, which is uncorrelated with

u. Suppose I applied control function approach and found that δb1 = 0 in

the following first stage regression,

y2 = δ0 + δ1z2 + u

what kind of problem we will have in the second stage regression? Prove

your answer.

**Ans:** Our second equation would produce y1= Bo+b1Y2hat+u+b1v. The issue with this is the composite error in 2SLS, u + b1v.

Computer Exercise

3. The purpose of this exercise is to compare the estimates and standard errors

obtained by correctly using 2SLS with those obtained using inappropriate procedures. Use the data file WAGE2.

(a) Use 2SLS package (ivreg function) to estimate the equation,

log(wage) = β0 + β1educ + β2exper + β3tenure + u

where sibs is the IV for educ. Report the results in the usual form.

A close-up of a document

Description automatically generated with medium confidence

(b) Now, manually carry out 2SLS. That is, first regress educ on sibs, exper,

and tenure and obtain the fitted values, educ d . Then run the second stage

regression log(wage) on educ d , exper, and tenure. Verify that βbj are identical to those obtained from 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.

Text

Description automatically generated with medium confidence

(c) Now, use the following two-step procedure, which generally yields inconsistent parameter estimates of the βj, and not just inconsistent standard errors. In step one, regress educ on sibs only and obtain the fitted values, educ d . Then, in the second step, run the regression of log(wage) on educ d , exper, and tenure. How does the estimate from this incorrect, twostep procedure compare with the correct 2SLS estimate of the return to education.

Ans: A screenshot of a computer

Description automatically generated with low confidence

It is different from the correct 2SLS estimate of the return on education because the

4. The data in FERTIL2 include, for women in Botswana during 1988, information

on number of children, years of education, age, and religious and economic status

variables.

(a) Estimate the model

children = β0 + β1educ + β2age + β3age2 + u

by OLS, and interpret the estimates. 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?

**Ans:** Based on the estimate of the equation above. There will be an expectation of 9 fewer children being born per 100 women gaining an extra year of education

(b) The variable frsthalf is a dummy variable equal to one if the woman was

born during the first six months of the year. Assuming that frsthalf is

uncorrelated with the error term from part (a), Run a first stage regression

to show that frsthalf is a reasonable IV candidate for educ.

**Ans:** Running the first stage regression, the coefficient we get for firsthalf is roughly around -.85, we have a strong correlation between frsthalf and educ

(c) Estimate the model from part (a) by using frsthalf as an IV for educ.

Compare the estimated effect of education with the OLS estimate from

part (a).

**Ans:** We can compare the effects of OLS model. After we run the estmates of OLS in part C we can see the educ coefficient actually changes -.17