BT2101 Tutorial

April 6, 2022

1. How does fertility affect labor supply? That is, how much does a woman's labor supply fall when she has an additional child? In this exercise you will estimate this effect using data for married women from the 1980 U.S. Census. Please refer to a set of data uploaded on luminus. The data contains information on married women aged 21-25 with two or more children. The description of data is as follows.

Variable	**Description**
morekids	=1 if mom had more than 2 children
boy1st	=1 if 1st child was a boy
boy2nd	=1 if 2nd child was a boy
samesex	=1 if 1st two children same sex
agem1	age of mom at census
black	=1 if mom is black
hispan	=1 if mom is Hispanic
othrace	=1 if mom is not black, Hispanic or white
weeksm1	mom's weeks worked in 1979

a. The result of regressing *weeksm1* on the indicator variable *morekids* using OLS is described as follows. On average, do women with more than two children work less than women with two children? How much less?

Regressor	OLS	IV	IV
Morekids	-5.387 (0.087)	-6.313 (1.275)	-5.821 (1.246)
Additional Regressor	Intercept	Intercept	Intercept, agem1, black, hispan, othrace
First Stage F-Statistic		1238.2	1280.9

b. Explain why the OLS regression estimated in (a) is inappropriate for estimating the causal effect of fertility (*morekids*) on labor supply (*weeksm1*).

c. Explain why *samesex* is a valid instrument for the instrumental variable regression of *weeksm1* on morekids.

d. Is samesex a weak instrument?

e. Do the results change when you include the variables *agem1*, *black*, *hispan*, and *othrace* in the labor supply regression (treating these variables as exogeneous)? Explain why or why not.

2. Card(1995) used wage and education data for a sample of men in 1976 to estimate the return to education. He used a dummy variable for whether someone grew up near a for-year college (nearc4) as an instrumental variable for education. In a log(wage) equation, he included other standard controls: experience, a black dummy variable for living in an SMSA and living in the South, and a full set of regional dummy variables and an SMSA dummy for where the man was living in 1966. In order for nearc4 to be a valid instrument, it must be uncorrelated with the error term in the wage equation – we assume this- and it must be partially correlated with educ. To check the latter requirement, we regress educ on nearc4 and all of the exogeneous variables appearing in the equation. (That is, we estimate the reduced form for educ) Using the data in CARD, we obtain, in condensed form,

$$educ = 16.64\ (.24) + .320(.088)nearc4 - .413(.034)exper + \cdots,\ R^2 = .477, n = 3{,}010$$

We are interested in the coefficient and t statistic on nearc4. The coefficient implies that in 1976, other things being fixed (experience, race, region, and so on), people who lived near a college in 1966 had, on average, about one-third of a year more education than those who did not grow up near a college. The t statistic on nearc4 is 3.64, which gives a p-value that is zero in the first three decimals. Therefore, if nearc4 is uncorrelated with unobserved factors in the error term, we can use nearc4 as an IV for educ.

Question) Run the model and explain if educ has endogenous or not.