Homework 3

Introduction ot Econometrics,Fall 2018

11/30/2018

1 Due Date:

- ***Due to Dec.29 12:00 pm***

2 Learning Objectives

- Use *Stata* or *R* to understand and practice following Econometric Methods
  - **Oaxaca-Blinder Decomposition**
  - **Instrumental Variable**

3 Main Content

3.1 Instrumental Variable

- Joshua Angrist "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," American Economic Review, June 1998, Vol. 88, No. 3, 450-477.
- 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 problem set you will estimate this effect using data for married women from the 1980 U.S. Census.
- Data and Variable Descriptions: The file fertility.dta (in Stata format) contains data on 254,654 women between the age of 21 and 35 with two or more children.

Variables	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

Variables	Description
othrace	=1 if mom is not black, Hispanic or white
weeksm1	mom's weeks worked in 1979

1. Our outcome variable is *weeksm1*. Regress *weeksm1* on the dummy variable *morekids*, which is our treatment variable. On average, do women with more than two children work less than women with two children? How much less?
2. Explain why the OLS regression estimated in (1) is inappropriate for estimating the causal effect of fertility (morekids) on labor supply (weeksm1).
3. The data set contains the variable *samesex*, which is equal to 1 if the first two children are of the same sex (boy-boy or girl-girl) and equal to 0 otherwise. Are couples whose first two children are of the same sex more likely to have a third child? Is the effect large? Is it statistically significant?
4. Explain why *samesex* is a valid instrument for the instrumental variable regression of *weeksm1* on *morekids*.
5. Run the reduced form regression.
6. Estimate the regression of *weeksm1* on *morekids*, using *samesex* as an instrument. How large is the fertility effect on labor supply?
7. How does the results change when you include the variables *agem1*, *black*, *hispan*, and *othrace* in the labor supply regression?
8. Report the five regressions you just run in a table. Note that standard errors should be shown in parenthesis.
    - The first column: the OLS regression.
    - The second column: the first stage regression.
    - The third column: the reduced form regression.
    - The fourth column: the 2SLS regression
    - The fifth column: the 2SLS regression with controls

3.2 Oaxaca-Blinder Decomposition

- In this exercise, you are asked to produce some results about Rural-urban migrants v.s Urban natives in China using a cleaned subset of *Rural to Urban Migrant Suvey in China(RUMiC)*.

1. Generate basic summary statistics such as mean, standard deviation, and total number of observations for all variables(including **age,gend,enth,masta,edu,hwage and exper**) in the data sets by migrate status(thus one for migrants and the other for urban natives) in two years(2002 and 2007) respectively.(**Table 1**)

2. The classical Mincer wage equation is

$$lny_i = \beta_0 + \beta_1 edu + \beta_2 exper + \beta_3 exper^2 + \gamma X + u_i$$

- where X include marital status and minority variable. We put some dummy variables to run a series of OLS regression to measure the log wage differentials as following one by one
  1) a dummy variable of migrating status(migrants/urban residents) in 2002 and 2007 respectively.
  2) a dummy variable of gender (male/female) in 2002 and 2007 repectively
  3) a dummy variable of selecting year(2002/2007).
- Please put all regressions above into one table(**Table 2**), report the amounts and the significance of these estimate coefficients of dummies and interactions.And try to present the economic implications of these results respectively.

3. Use Stata/R command to perform an Oaxaca-Blinder decomposition(after being sure that the oaxaca package has been installed on your computer) males and females in 2002 and 2007 wage gaps and the dynamics across 2002 and 2007.

   1) Please use in turn the urban coefficients, the migrants coefficients and two weighted(0.5 and sample-adjusted) average of the two in 2002 answer the following questions: Report the amount and significance of explained effects and unexplained effects, and calculate the shares of these effects to total wage gaps. Put all results in one table(**Table 3**) and presents which one dominate more in 2002.

   - Based on **Table 3**,answer the following questions: a) how to explain the fact? b) Is there some differences among these results using different weights? Please explain why?

   2) Please use in turn the Oaxaca and Ransom(1994) or Jann(2008) weighted average of male and female in 2002 and 2007 respectively.Report the amount and significance of explained effect and unexplained effect for three specific variables such as education,age and experience. Calculate the shares of these effects to total explained/unexplained effects and to total wage gaps,receptively. Put all results in one table(**Table 4**)

   - Based on **Table 4** answer the following questions: a)which one dominate more for three variables in 2002 and 2007,respectively? b)Please explain these results one by one in 2002 and 2007. c) does anything change during 2002-2007? Explain what is going on.

   3) Please use OB decomposition with Jan(2008)'s weight to distengle unexplained and explained effect to the wage change across 2002 and 2007. Report the amount and significance of explained effect and unexplained effect for total *AND* three specific variables such as education,age and experience. Calculate the shares of these effects to total explained/unexplained effects and to total wage changes during 2002-2007,repectively. Put all results in one table(**Table 5**)

   - Based on **Table 5** answer the following questions:a) explain these results in detail. 2) make

a comparsion with the result of the last subquestion in question 3(thus "And does anything change during 2002-2007? Explain what is going on?) and tell what is the difference of implications between these two decomposition results.