

Problem Set 5

Instrumental Variable Estimation

Read in the data set “fertility.raw” into TSP. It contains data on 4361 women aged between 15 and 49 in the African country of Botswana. In Botswana there is a “peculiar” institutional regulation that states that those born in the first half of the year (January-June) are put to school later (if at all). Accordingly the probability that they graduate after more than 7 years of education is smaller. In the following the effect of education on the fertility rate will be examined. The data contain the following variables in the given order:

Variable name	Variable description
mnthborn	Month of birth of woman
age	Women's age in years
electric	= 1, if electricity in the household
radio	= 1, if the household owns a radio
tv	= 1, if the household owns a TV
bicycle	= 1, if the household owns a bicycle
educ	Number of years of educ (woman)
children	Number of living children (outcome variable)
heduc	Number of years of education (husband)
age2	Second-order polynomial in age
urban	= 1, if place of residence in the city
catholic	= 1, if Catholic
frsthalf	= 1, if the woman is born in the first half of the year
evermarr	= 1, if ever married
educ7	= 1, if more than 7 years of education

1. Do some descriptive statistics in order to get a general idea of the data. What is the average number of children of a woman with more than 7 years education? And of a woman with less than 7 years of education? Is the difference in the number of children per woman with more or less than 7 years of education significant?
2. Estimate a linear regression model with the number of children as a dependent variable and *educ7* as a treatment dummy. As further explanatory variable use in addition to the constant *age*, *age2*, *evermarr*, *urban*, *electric*, and *tv*.
3. Think about which variable could be endogenous in the estimation equation for children and explain why.

Instrumental Variable Estimation

4. Assume in the following that *educ7* is endogenous. What conditions have to be fulfilled in the instrumental variable estimation? Which variable from the data set would be suitable for an instrument? Why?
5. Estimate a reduced form equation for *educ7* containing all explanatory variables from question 2 and the dummy variable *frsthalf*. Do *educ7* and *frsthalf* have a statistically significant partial correlation?
6. Estimate the effect of education on the fertility rate by IV, using *frsthalf* as an instrument for *educ7*. Compare the results with that obtained from question 2.
7. In the following we want to test whether the variable *educ7* is endogenous.
 - a) Explain the idea of the Hausman test for endogeneity.
 - b) Implement a regression-based form of the Hausman test.
 - c) Calculate the Hausman t statistic by comparing the coefficient of *educ7* obtained in the OLS estimation with that obtained in the IV regression.
8. In the following, an efficient IV-estimator has to be calculated using the following method:
 - Estimate the probability of more than 7 years of education by Probit and take into account all explanatory variables from question 2, as well as the instrument *frsthalf*.
 - Calculate the predicted probability for *educ7*.
 - Carry out a 2SLS-estimation that uses the predicted probability for *educ7* as an instrument.