

Advanced Applied Econometrics - IV Repetition

Felix Weinhardt

Exercise: Instrumental Variables I

In this exercise you are going to use the instrumental variables estimator to estimate the returns to education. The motivation for this is that the level of education of a person is likely to be endogenous. In particular most labor economists expect that more talented people will also receive more education, but may earn higher wages partly because of their higher talent. The data for this exercise are contained in the dataset `card.dta` which is an extract from the U.S. National Longitudinal Survey of Young Men (NLSY). The dataset contains information on 3010 young American males for the year 1976. This data is used in Card (1993) "Using Geographic Variation in College Proximity to Estimate the Return to Schooling", NBER Working Paper No. 4483. The data set contains the following variables which will be of interest for us:

<code>lwage</code>	log of the wage rate
<code>educ</code>	education (years of schooling)
<code>exper</code>	job market experience (years)
<code>expersq</code>	job market experience, squared
<code>black</code>	dummy = 1 if person is black
<code>smsa</code>	dummy = 1 if person lives in a Standard Metropolitan Statistical Area (SMSA)
<code>south</code>	dummy = 1 if person lives in a Southern state
<code>smsa66</code>	dummy = 1 if person lived in a SMSA in 1966
<code>reg661</code>	dummy = 1 if person lived in region 1 in 1966
<code>...</code>	<code>...</code>
<code>reg669</code>	dummy = 1 if person lived in region 9 in 1966
<code>nearc4</code>	dummy = 1 if person lived close to a 4-year college in 1966
<code>iq</code>	IQ test score

Questions:

1. Use the OLS estimator to regress the logarithm of the hourly wage (`lwage`) on the years of education (`educ`), years of experience (`exper`), the square of experience (`expersq`), ethnic background (`black`), the dummy for residents in a metropolitan statistical area in 1976 (`smsa`) and 1966 (`smsa66`), the dummy for a southern state (`south`) and different regions in 1966 (`reg662`–`reg669`). Do the results confirm your expectations? Would you expect the OLS estimate to be an over or under-estimate of the true effect of education?

2. The dataset contains a dummy variable `nearc4` which is equal to one if a person lives close to a four-year college. Use this instrument to re-estimate the returns to education. Does the instrument `nearc4` pass condition IV1? Compare the results of the IV estimation to the OLS results from the previous regression. Are the differences between the results what you would have expected? How plausible is it that instrument `nearc4` is exogenous?
3. For 2061 people in the data there is information about their results on a standardized IQ-test (variable `iq`). Regress `iq` on `nearc4`. Which implications do these results have for the likely exogeneity of the instrument `nearc4`?
4. Now include the variables `smsa66` and `reg662-reg669` in the regression of `iq` on `nearc4`. How do the results from this regression modify your conclusions in the previous question?