Problem Set 5

Instrumental Variables Estimation

The data set card.raw contains the data used by David Card (1995) to study the returns to schooling.

a) Estimate a log($wage$) equation by OLS with $educ$, $exper$, $exper^2$, $black$, $south$, $smsa$, $reg661$ through $reg668$, and $smsa66$ as explanatory variables. What are the returns to one year of education? Is this a causal relationship?

b) Card proposed to use distance to college as an instrument for years of education. Which assumptions have to be fulfilled for this to be a valid instrument? How can you test these assumptions?

c) For the IV approach, estimate and interpret:
   - the first stage
   - the reduced form
   - the Wald estimator

d) What are possible reasons why the IV returns to schooling are higher than the OLS returns?

e) Estimate the returns to schooling by 2SLS using distance to a 4-year college ($nearc4$) and distance to a 2-year college ($nearc2$) as instruments.

f) For a subset of men in the sample, IQ scores are available. Regress $iq$ on $nearc4$. Is the IQ score uncorrelated with $nearc4$?

g) Now regress $iq$ on $nearc4$ along with $smsa66$, $reg661$, $reg662$, and $reg669$. Are $iq$ and $nearc4$ partially correlated? What do you conclude about the importance of controlling for the 1996 location and regional dummies in the log($wage$) equation when using $nearc4$ as an instrument for $educ$?