Education and Adult Mortality

David Zynda

November 13, 2017

The Ideal (and Simple) Model

When considering a model assessing the impact of education on health, a first try might implement the following:

$$H_i = X_i\beta + E_i\pi + \epsilon_i$$

where H_i is individual i's health stock, E is his education level, and X is a vector of individual characteristics that affect health, e.g. smoking and genetic factors. Then, one could use OLS to determine if π is significantly different that 0.

However, there may be a couple concerns (impart based on previous research):

- Omitted Variable Bias
  - Health during childhood could affect educational attainment
  - Or improved health increases investments in education
  - Or the same people who value health also value education

IV Estimates

This paper employs an instrumental variable - a variable which shares a common association with another independent variable but does not share an association with the unobserved term. Sometimes, the zero-conditional mean assumption cannot be met due to constraints of the data. Like differencing, IV, in short, is another way to get around that.

For some $y = \beta_0 + \beta_1 x + u$ assume $Cov(x, u) \neq 0$

let z be an observable variable that satisfies two assumptions:

$$Cov(z, u) = 0$$

(Compulsory Edu. Laws and unobserved terms in equation predicting health) and

$$Cov(z, x) \neq 0$$

(Compulsory Edu. Laws and Educational attainment)

Then z is called an instrumental variable for x or an instrument for x

IV Estimates

Here, we define education as:

$$E_{ics} = b + CL_{cs}\pi + X_{ics}\beta + W_{cs}\delta + \gamma_c + \alpha_s + \epsilon_{ics}$$

where CL_{cs} is the compulsory education laws that serve as instruments to identify the education equation.

Now, take E_{ics} and put it into the equation we have seen before:

$$D_{itcs} = b + E_{ics}\pi + X_{ics}\beta + W_{cs}\delta + \gamma_c + \alpha_s + \epsilon_{ics}$$

and estimated by a common IV estimator called the Wald Estimate.

Instead of using a health stock indicator H_i, use D_{itcs} with 1 represeting the case that individual i at time t in cohort c of state s is deceased. E is education and X still represents time invarainte individual characteristics like gender. *W represents individuals characteristics at age 14* γ is a set of cohort dummies *α is set of state dummies

Two Stage Least Square Estimates

As the name implies, 2SLS model estimates our model of interest... in two steps.

First, estimate (in aggregate/averages)

$$\overline{E}_{gtcs} = b_1 + CL_{gcs}\pi + \overline{X}_{gcs}\beta + W_{cs}\delta + \gamma_c + \alpha_s + \epsilon_{gcs}$$

And using those estimates, run the following regression:

$$\overline{D}_{gtcs} = b + \overline{E}_{gcs}\pi + \overline{X}_{gcs}\beta + W_{cs}\delta + \gamma_c + \alpha_s + \epsilon_{gcs}$$

This is done first using the aggregate data, like above. The author implements a second model too. Using the same method above, another model is run. However, $\overline{E}$ is estimated now using the individual data. The only difference, then, is that the standard 2SLS uses predicted average education whereas Mixed-2SLS uses average predicted education.

In Short:

- Step 1: Look at the direct effect of compulsory education laws on education attainment using Compulsory Educations laws as an instrumental variable
- Step 2: Using Education Regression from Step 1, estimate death rate for cohorts