Ekaterina Kazak
University of Manchester

The purpose of this session is to introduce you to the instrumental variable approach.

Instrumental Variables

Use the Wooldridge dataset `wage 2` to estimate the returns to education.

$$\log(wage) = \beta_0 + \beta_1\, educ + u.$$

It is often argue that *educ* in such model is an endogenous regressor. In order to mitigate the problem we address the instrumental variables approach and use $sibs_i$, the number of siblings of individual i as an instrument for schooling.

a) To convince yourself that using *sibs* as an IV for *educ* is *not* the same as just plugging *sibs* in for *educ* and running an OLS regression, run the regression of $\log(wage)$ on *sibs* and explain your findings.

b) The variable *brthord* is birth order (*brthord* is 1 for the first-born, 2 for a second born child, and so on). Explain why *educ* and *brthord* might be negatively correlated. Regress *educ* on *brthord* to determine whether there is a statistically significant negative correlation.

c) Use *brthord* as an IV for education in the following equation:

$$\log(wage) = \beta_0 + \beta_1\, educ + u.$$

Report and interpret the results.

d) Now, suppose that we include number of siblings as an explanatory variable in the wage equation; this controls for family background, to some extent:

$$\log(wage) = \beta_0 + \beta_1\, educ + \beta_2\, sibs + u.$$

Suppose that we want to use *brthord* as an IV for *educ*, assuming that *sibs* is exogenous. The reduced form (first-stage) for *educ* is:

$$educ = \pi_0 + \pi_1\, sibs + \pi_2\, brthord + v.$$

State and test the identification assumption.

e) Estimate the equation in part (d) using *brthord* as an IV for *educ* (and *sibs* as its own IV). Comment on the standard errors for $\hat{\beta}_{educ}$ and $\hat{\beta}_{sibs}$.

f) Using the fitted values from part (d), $\widehat{educ}$, compute the correlation between $\widehat{educ}$ and *sibs*. Use this result to explain your findings from part (e).