Эконометрика-2 ММАЭ
Семинар 11

Лекции: А.А. Пересецкий
Семинары: Е.С. Вакуленко

Эндогенность. Метод инструментальных переменных

Задача 1. *(c) R. Enikolopov, NES, 2008*

In her paper "Does Competition Among Public Schools Benefit Students and Taxpayers" (1994), Hoxby investigates whether a wider choice of schools for children provides incentives for schools to increase their quality. Her outcome variable (denoted by y) is a measure of children's reading ability at age 16, and her measure of how much choice of school a child has is the number of different school districts within a certain distance of their home (denote this variable by X). Because of the potential endogeneity of the number of school districts she instruments this with the number of rivers in the local area (denoted by Z), arguing that when school districts were defined (mostly in the 19th century) rivers were often used as natural boundaries for them.

(a) What is the first-stage regression for this model? What should you look for to decide whether there is a problem of 'weak instruments' in this case?

(b) Describe how would you obtain the IV estimate of the effect of X on y? What are the likely problems with this estimate if the instruments are 'weak'?

(c) Another researcher uses an alternative valid instrument. Why might they get different results from Hoxby?

(d) Another researcher argues that rivers represent, even today, a barrier to travel so that their presence in an area limits the choice of schools for children. What problems does this argument pose for Hoxby's methodology? Can you test whether this is an important criticism in practice?

Задача 2. Wooldridge – Introduction. *Instrumental variables vs. OLS*

Consider the regression model $y = \alpha + \beta x + u$, where x and u are correlated, $Corr(x,u) \neq 0$. Assume for the simplicity that $\sigma_u = \sigma_x$, so that the population variation in the error term is the same as it is in x. Suppose that the instrumental variable, z, is slightly correlated with u: $Corr(z,u) = 0.1$. Suppose also that z and x have a somewhat stronger correlation: $Corr(z,x) = 0.2$.

a. What is the asymptotic bias in the IV estimator?

b. How much correlation would have to exist between x and u before OLS has more asymptotic bias than 2SLS?

Задача 3.

Предположим, исследователя интересует влияние X $(n \times k)$ на y $(n \times 1)$

$$y = X\beta + \varepsilon$$

X эндогенны, однако существует набор «валидных» инструментов Z^* $(n \times k)$:

$$E\left(\varepsilon \mid Z^*\right) = 0.$$

Однако исследователь наблюдает только Z $n \times k$, измеренные с ошибкой:

$$Z = Z^* + u,$$

где u - ошибка измерения, ε, u - независимы и $E\left(u \mid X, Z^*\right)=0$.
Докажите, что в данном случае ошибки измерения в инструментальных переменных не влияют на состоятельность IVоценки $\hat{\beta}_{IV}$. Означает ли это, что ошибки измерения не важны?

Задача 4. *Wooldridge Intro 15.2*
Suppose that you wish to estimate the effect of class attendance on student performance, as in Problem 3. A basic model is

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + u$$

where the variables are defined in "ATTEND.des".

(a) Let *dist* be the distance from the students' living quarters to the lecture hall. Do you think dist is uncorrelated with *u*?
(b) Assuming that *dist* and *u* are uncorrelated, what other assumption must *dist* satisfy in order to be a valid IV for *atndrte*?
(c) Suppose we add the interaction term *priGP*atndrte*

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + \beta_4 priGPA * atndrte + u$$

If *atndrte* is correlated with *u*, then, in general, so is *priGPA*atndrte*. What might be a good IV for *priGPA*atndrte*? [Hint: If $E\left(u \mid priGPA, ACT, dist\right)=0$, as happens when *priGPA*, *ACT*, and *dist* are all exogenous, then any function of *priGPA* and *dist* is uncorrelated with *u*.]

Задача 5. *(2SLS, (c) Ruben Enikolopov, NES, 2008)*
Let the true model be given by:

$$\log(e) = \alpha + \beta \log(a) + \varepsilon \quad (1)$$

where e is the expense ratio, $e = \frac{x}{a}$, x - total expenditures (USD), a - the size of a firm (in terms of net assets, USD). $E\left(\varepsilon \mid a\right)=0$.

β is the elasticity of the expense with respect to size. $\beta > 0$ - decreasing returns to scale, $\beta < 0$ - increasing returns to scale.

The data is a sample of 1876 mutual fund firms is in the "Funds.dta".

(a) Run the OLS regression (1)

(b) Run the IV regression (1) with variable "age"" as an instrumental variable for log size for an endogenous regressor $\log(a)$.

(c) Construct the IV estimate in STATA in two steps. In the first step, run OLS and use the predict option to create the appropriate "hat" variable. In the second step use this "hat" variable as a regressor to generate the IV estimate. What do you notice about the standard error of the IV estimate computed this way, and using the **ivreg** command in STATA?

Hint: See “problem5.do”