

Эконометрика-2 ММАЭ

Семинар 12

Эндогенность

Лекции: А.А. Пересецкий

Семинары: Е.С. Вакуленко

Задача 1.

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^* ($n \times k$):

$$E(\varepsilon | Z^*) = 0.$$

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

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

где u - ошибка измерения, ε, u - независимы и $E(u | X, Z^*) = 0$.

Докажите, что в данном случае ошибки измерения в инструментальных переменных не влияют на состоятельность IV оценки $\hat{\beta}_{IV}$. Означает ли это, что ошибки измерения не важны?

Задача 2.

Рассмотрим модель

$$y = X\beta + \varepsilon, \quad X - n \times k, \quad E(\varepsilon | X) \neq 0$$

Пусть Z_1 и Z_2 представляют собой матрицы инструментов размерности $n \times l_1$ и $n \times l_2$ соответственно, и пусть матрица Z_2 состоит из Z_1 плюс $l_2 - l_1$ дополнительных столбцов.

Докажите, что оценка метода инструментальных переменных (IV), полученная с использованием инструментов Z_2 , асимптотически более эффективна по сравнению с IV оценкой, полученной на основе инструментов Z_1 .

Hint 1: $AsyVar(\hat{\beta}_{IV}) = \sigma^2 \lim_{n \rightarrow \infty} (n^{-1} X^T P_Z X)^{-1}$ для произвольной матрицы Z .

Hint 2: Чтобы это доказать, Вам необходимо показать, что матрица

$$(X^T P_{Z_1} X)^{-1} - (X^T P_{Z_2} X)^{-1} \text{ положительно полуопределена.}$$

Задача 3. Wooldridge Intro 15.2

Suppose that you wish to estimate the effect of class attendance on student performance. 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 *priGPA*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 $E(u | priGPA, ACT, dist) = 0$, as happens when *priGPA*, *ACT*, and *dist* are all exogenous, then any function of *priGPA* and *dist* is uncorrelated with *u*.]

Задача 4. Wooldridge – CrossSectionPanel 5.4

Use the data in “CARD.dta” for this problem. See Card (1995): “Using Geographic Variation in College Proximity to Estimate the Return to Schooling,” in *Aspects of Labour Market Behavior: Essays in Honour of John Vanderkamp*, ed. L. N. Christophides, E.K. Grant, and R. Swidinsky. Toronto: University of Toronto Press, 201–222.

WP: <http://www.nber.org/papers/w4483.pdf>

Problem: whether individual's choice of educational attainment is related to their potential earnings.

- a. Estimate a $\log(wage)$ equation by OLS with *educ*, *exper*, *exper 2*, *black*, *south*, *smsa*, *reg661* through *reg668*, and *smsa66* as explanatory variables.

```
reg lwage educ exper expersq black south smsa reg661 reg662 reg663 reg664
reg665 reg666 reg667 reg668
```

```
est store ols
```

```
*The Breusch-Godfrey test
```

```
estat hettest
```

```
*The White test
```

```
estat imtest, white
```

```
reg lwage educ exper expersq black south smsa reg661 reg662 reg663 reg664
reg665 reg666 reg667 reg668, robust
```

Пошаговое исключение незначимых переменных (встроенный алгоритм)

```
stepwise, pr(.05): reg lwage educ exper expersq black south smsa reg661
reg662 reg663 reg664 reg665 reg666 reg667 reg668
```

b. Estimate a reduced form equation for *educ* containing all explanatory variables from part a and the dummy variable *nearc4*. Do *educ* and *nearc4* have a practically and statistically significant partial correlation?

```
reg educ nearc4 exper expersq black south smsa reg661 reg662 reg663 reg664
reg665 reg666 reg667 reg668

test nearc4=0
```

c. Estimate the $\log(\text{wage})$ equation by IV, using *nearc4* as an instrument for *educ*. Compare the 95 percent confidence interval for the return to education with that obtained from part (a).

```
ivregress 2sls lwage (educ=nearc4) exper expersq black south smsa reg661
reg662 reg663 reg664 reg665 reg666 reg667 reg668

estat firststage
```

For the case of a single endogenous regressor, Staiger and Stock (1997) suggested declaring instruments to be weak if the First-stage F-statistic is less than 10.

d. Now use *nearc2* along with *nearc4* as instruments for *educ*. First estimate the reduced form for *educ*, and comment on whether *nearc2* or *nearc4* is more strongly related to *educ*. How do the 2SLS estimates compare with the earlier estimates?

```
reg educ nearc4 nearc2 exper expersq black south smsa reg661 reg662 reg663
reg664 reg665 reg666 reg667 reg668

test nearc4=nearc2=0
```

```
ivregress 2sls lwage (educ=nearc4 nearc2) exper expersq black south smsa
reg661 reg662 reg663 reg664 reg665 reg666 reg667 reg668

est store iv

estat firststage
```

e. For a subset of the men in the sample, IQ score is available. Regress *iq* on *nearc4*. Is IQ score uncorrelated with *nearc4*?

```
reg iq nearc4
```

f. 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 1966 location and regional dummies in the *log(wage)* equation when using *nearc4* as an IV for *educ*?

```
reg iq nearc4 smsa reg661 reg662 reg663 reg664 reg665 reg666 reg667 reg668
```

The Hausman test

	H_0	H_1
$\hat{\theta}_0$	consistent, efficient	inconsistent
$\hat{\theta}_1$	consistent, inefficient	consistent

$$H = n(\hat{\theta}_1 - \hat{\theta}_0)(\hat{V}_1 - \hat{V}_0)^{-1}(\hat{\theta}_1 - \hat{\theta}_0) \sim \chi^2(\dim(\theta))$$

```
hausman estimates_1 estimates_0
```

```
hausman iv ols
```