

Elements of Econometrics, 2023-2024

Class teacher: Ksenia Kasianova.

Class 13: Instrumental variables

Problem 1

Show that

- (a) If random errors have zero conditional expectation for any given $x_i : E(\varepsilon_i | x_i) = 0$, then the regressor and random error are not correlated with each other: $\text{cov}(\varepsilon_i, x_i) = 0$.
(b) If observations are i.i.d. and $E(x_i^4) < \infty$, $E(y_i^4) < \infty$.

$$\widehat{\text{cov}}(x, y) \xrightarrow{p} \text{cov}(x_i, y_i).$$

Problem 2

Let the variables y_i and x_i^* be related by the exact relation:

$$y_i = \beta_1 + \beta_2 \cdot x_i^*.$$

Instead of the exact values of the regressor, we observe values measured with error:

$$x_i = x_i^* + \varepsilon_i, \text{cov}(x_i^*, \varepsilon_i) = 0$$

The following regression is estimated using the least squares method:

$$y_i = \beta_1 + \beta_2 \cdot x_i + u_i.$$

- (a) Show that the least squares estimate $\hat{\beta}_2$ will be inconsistent.
(b) What will the sign of a bias be?

Problem 3

- (a) What is a good instrument?
(b) We want to estimate the following regression using OLS, knowing that x is endogenous

$$y_i = \beta_1 + \beta_2 \cdot x_i + u_i.$$

What will the properties of $\hat{\beta}_2$ be?

- (c) Derive IV estimator given that z is a good instrument
(d) Compare the formula for standard error for the IV estimator

$$\text{se}(\hat{\beta}_2) = \sqrt{\frac{S^2}{\sum (x_i - \bar{x})^2} \cdot \frac{1}{\widehat{\text{corr}}^2(x, z)}}, \quad S^2 = \frac{\sum_{i=1}^n e_i^2}{n-2}.$$

with the OLS formula.

Problem 4

This problem is based on article by Acemoglu, Daron, Simon Johnson and James A. Robinson. Colonial Origins of Comparative Development: An Empirical Analysis Research, American Economic Review, 2001, vol. 91 (5). At the moment the article has about two thousand citations and ranks 8th in the list of the most cited articles from the top 5 scientific economic journals in the world.

The authors test the hypothesis that the quality of institutions is important for ensuring high rates of economic growth. More precisely, countries where property rights are better protected have an advantage in capital accumulation, which ultimately leads to higher long-term GDP per capita.

Source data file: “Acemoglu”.

The file contains information about the following variables:

“countryn” - country name, “shortnam” - short country code,

“lgdp” - logarithm of GDP per capita in 1995,

“logmort” - logarithm of the mortality rate of colonists (see details in the text of the article)

“latitude” - the latitude at which the capital of a given country is located (measured as a distance from the equator and normalized so as to vary from zero to one)

“prot” - an index of protection of property rights in a given country (the measure is arranged so that higher values correspond to better protection of property rights)

“euro” is the proportion of the population of European descent in a given country as of 1975.

In the task, you are invited to reproduce the main results of the authors of the article, following their logic.

(a) Construct a scatter diagram characterizing the dependence of the lgdp variable on the prot variable. Estimate the parameters of the corresponding pairwise regression using ordinary OLS. Why the resulting dependence cannot be interpreted as a cause-and-effect relationship?

(b) Estimate the new regression including latitude in the original equation. Based on the authors’ logic, explain why latitude might be an important control variable in this case. Interpret your results. Do they confirm that latitude is really an important variable?

(c) Before proceeding with 2SLS, estimate the equation of the reduced form: the regression of the lgdp on the mortality of settlers logmort. Make a scatterplot characterizing this relationship. How can you explain the results obtained?

(d) We now turn to assessing the impact of the quality of institutions on the logarithm of GDP using 2SLS. Use the logmort variable as an instrument. Explain meaningfully why the authors consider this tool valid, and why such an empirical strategy makes it possible to correctly test the presence of a causal relationship between the quality of institutions and GDP.

(e) Estimate the regression of the first step of 2SLS. Illustrate the result with a scatter plot. After calculating the appropriate F-statistic, explain whether the instrument used is strong (relevant).

(f) Estimate the second step regression. Interpret the results obtained. How has the estimate of the impact of the quality of institutions on GDP changed compared to the

estimate obtained with OLS? What does this change say about the cause of the endogeneity of the regressor (which of the causes of endogeneity prevailed in the original model)?

(g) Evaluate the alternative specification of 2SLS regression adding a control variable. Is the conclusion about the positive impact of the quality of institutions on GDP sustainable?

(h) Another way to test the robustness of the results is by analyzing individual subsamples. Construct the regressions from the two previous paragraphs using only data for all countries except African. Then build two similar models using a subsample of African countries only. Compare estimates of the impact of institutional quality on GDP in this subsample with those in the base sample. Is the analyzed effect heterogeneous? What problems with the quality of the models do you see?

(i) Suppose a critic asserts the following: “Settler mortality is correlated not only with the quality of institutions, but also with the proportion of the population of European origin in a given country (variable *euro*). Due to the peculiarities of the colonization process in some countries, the proportion of the European population is much higher than that of the indigenous population and sometimes approaches 100 percent. Moreover, this share itself affects the value of GDP: the more people with European origin in the country, the more GDP, even with the same quality of institutions.”

In other words, the critic argues that the high per capita GDP in former colonies such as the United States, Canada and Australia is primarily due not to the quality of institutions, but to the high proportion of the population of European descent in the total population of these countries.

If the critic is right, how does this affect the consistency of the 2SLS estimates in the regressions constructed above? Explain your answer in detail on an intuitive level.

Check the hypothesis formulated by the critic. If the hypothesis is confirmed, suggest an empirical approach that allows you to fix the problem. Implement it and interpret the results.

Problem 5

In the cigarette market in a certain country, the demand function in the i -th region has the form:

$$\ln Q_i = \beta_1 + \beta_2 \cdot \ln P_i + \varepsilon_i.$$

The supply function is described by the relation:

$$\ln Q_i = \gamma_1 + \gamma_2 \cdot \ln P_i + \gamma_3 \cdot \ln T_i + u_i,$$

where Q_i is the number of cigarettes in the i -th region; P_i is the price of cigarettes in the i -th region; T_i - sales tax in the i -th region; ε_i - independent and identically distributed random variables characterizing demand shocks (not correlated with taxes); u_i are independent and identically distributed random variables characterizing supply shocks.

(a) Explain why the OLS estimate of the price elasticity of demand in the model under consideration will be inconsistent. To do this, calculate the probability limit for the OLS estimate $\hat{\beta}_2^{\text{OLS}}$. Determine, if possible, whether OLS will overestimate or underestimate the elasticity of demand?

(b) Propose a procedure for estimating the price elasticity of demand. Formally justify your answer by calculating the appropriate probability limit.