HSE and University of London Double Degree Programme in Data Science and Business Analytics

Elements of Econometrics, 2023-2024

Classteacher: Ksenia Kasianova.

Class 24: Panel data

Problem 1

Explain how the covariance matrix of the random vector is structured in case of Feasible GLS for Random Effects Models.

Problem 2

A researcher analyzes the impact of a law allowing civilians to possess firearms on crime rates. He has panel data on 40 regions of a country over 20 years. D is a variable that is equal to one if in a given region in a given year there is a law allowing the possession of firearms, and zero otherwise. X and W are some control variables. Y - number of crimes in the region (thousands per year).

The researcher estimated 4 equations: equations 1-2 using a fixed effects model; Equations 3-4 - using a random effects model.

The results are presented in the table below.

Model evaluation results. Dependent variable - $\ln Y$

Model	Model 1	Model 2	Model 3	Model 4
Estimation method	FE	FE	RE	RE
D	−0.50 (0.04)	−0.50 (0.05)	−0.60 (0.01)	−0.40 (0.02)
X	0.32 (0.02)	0.21 (0.02)	0.05 (0.04)	0.04 (0.04)
W	−0.05 (0.01)	−0.06 (0.01)	−0.09 (0.02)	−0.10 (0.02)
Individual effects	Yes	Yes	Yes	Yes
Time dummy variables	No	Yes	No	Yes
Number of observations	800	800	800	800
R^2	0.657	0.780	-	-
P-Hausman test value	-	-	0.002	0.004
P-value of equality test zero coefficients for fixed-tive time variables	-	0.001	-	0.008

Note: Here and in all subsequent tables, robust standard errors are given in parentheses below the coefficient estimates.

a. Select the best model from the proposed list. Justify your choice.

b. For the selected model, check the statistical significance of the coefficient for the variable D and, if it turns out to be significant, give its meaningful interpretation.

Problem 3

Consider a model based on panel data:

$$y_{it} = \beta x_{it} + \mu_i + u_{it}; \quad i = 1, 2, \ldots, n; \quad t = 1, 2, \ldots, T.$$

Prove that, given data for two periods ($T = 2$), the within-group estimate (within estimator) of the parameter β coincides with the model estimate in first differences.

Problem 4

Consider a model based on panel data:

$$y_{it} = \theta x_{it} + \mu_i + u_{it}; \quad i = 1, 2, \ldots, 100; \quad t = 1, 2, 3,$$

where u_{it} are independent identically normally distributed quantities with zero mathematical expectation and variance equal to σ^2; μ_i is the individual fixed effect of the i-th object (unobserved variable).

The researcher obtains an estimate of the θ parameter using intragroup transformation of the initial data and application to the transformed data using the ordinary least squares method (within estimation).

a. Assuming x_{it} to be deterministic quantities, calculate the variance of the resulting parameter estimate θ (express it explicitly in terms of σ^2 and $x_{it}, i = 1, 2, \ldots, 100, \quad t = 1, 2, 3$).

b. Based on the results of the previous paragraph, explain in which of the two cases described below the accuracy of estimating the parameter θ will be higher:

1. The value of the regressor $\boldsymbol{x}$ for each object varies slightly from year to year, but within each year it differs greatly from object to object.

2. The value of the regressor $\boldsymbol{x}$ for each object varies greatly from year to year, but within each year it differs slightly from object to object.

Problem 5 Consider a model based on panel data:

$$y_{it} = \theta y_{it-1} + \mu_i + \varepsilon_{it}; \quad i = 1, 2, \ldots, n; \quad t = 1, 2, 3,$$

where ε_{it} are independent identically normally distributed quantities with zero mathematical expectation; μ_i is the individual fixed effect of the i-th object (unobserved variable). Such a model is called dynamic because it contains a dependent variable on the right side of the equation.

a. The researcher obtains an estimate of the θ parameter using a first difference model. Explain why this assessment would be invalid in this case.

b. Suggest a way to obtain a consistent estimate for the parameter θ.