Empirical Panel Data: Lecture 1

INSTRUCTOR: CHAOYI CHEN
NJE & MNB

March 7, 2023

General information

- The syllabus is the main source of information for the course. Please check the syllabus before asking questions.
- Assignments:
  - ASSN 1 Due Date: Tuesday, 25 April, In-class
  - ASSN 2 Due Date: Monday, 29 May
- Exams:
  - Final: Date TBA, In-class
- Grades:
  - 15% Each Assignment, 10% Class participation, 20% Group Project, 40% Final Exam

General information continue

- Book:
  - Hsiao, C. (2010). *Analysis of panel data*. Vol. Series Number 34. Econometric Society Monographs.
  - Baltagi, B. H. (2021). *Econometric analysis of panel data*. Springer.
  - Wooldridge, J. M. (2011). *Econometric analysis of cross section and panel data*. MIT Press.
- Lecture room: Info park campus, Lámfalussy lecture room, 4th floor.
- Office Hours: 1:30-2:30pm Tuesday, Info park campus, 1st floor.
- Email: chaoyi.chen@uni-neumann.hu

Term project

- Group number: you will form groups of 3-4 people and each group will turn in one project. (If there is no group available for you by the end of next week, please let me know).
- The paper should consist of **a simple empirical analysis** using **panel data**, and it should be at most **15 pages** long (incl. tables, figures, and bibliography). You can use either *word* or LaTeX to write your essay. Formal requirements will be strictly enforced. a Harvard style is recommended.
- All authors should have **same** contributions to the project formulation and paper preparation
- Project Due Date: Monday, 19 June. **NO EXTENSION**

Term project schedule suggestion

- Choose your group topic by the end of March
- A complete literature review with respect to your topic by the end of April
- After literature review, start to work on your topic. This includes
  - Introduction and a short literature review summary.
  - Specification of the econometric model
  - Obtain data
  - Estimation of the econometric model
  - Test hypothesis
  - Conclusion
- Your first version of the paper should be available by the end of May.
- The final project deadline is 19 June.

What is econometrics

- Definition from Jeffrey M. Wooldridge:

  Econometrics is based upon the development of statistical methods for estimating economic relationships, testing economic theories, and evaluating and implementing government and business policy.

- Econometrics is fundamentally based on four elements:
  - A sample of data
  - An econometric model
  - An estimation method
  - Some inference methods

Learning outcomes

At the end of this course you should

- have knowledge of regression analysis relevant for analyzing panel data.
- be able to interpret and critically evaluate outcomes of an empirical analysis
- know the theoretical background and assumptions for standard panel regression models and estimation methods.
- be able to use Stata to perform an empirical analyses.
- be able to read and understand journal articles that make use of the methods introduced in this course.
- be able to make use of econometric models in your own academic/practical work.

Course outline

- Topic 1: Reviews of Basic Econometrics
  - Linear model with cross-sectional data
  - OLS estimator
  - Matrix form of OLS estimator
- Topic 2: Panel Data Analysis: Basic
  - Panel data
  - Pooled panel model
  - Random effects model
  - Fixed effects model
- Topic 3: Panel Data Analysis: Estimation
  - LS dummy variable (LSDV) estimator
  - Within estimator
  - First-difference estimator
- Topic 4: Panel model with Endogeneity
- Topic 5: Panel unit-root and cointegration testings

- Cross-sectional Data
  - Sample of agents taken at one point in time.
  - Ideally, the data is a random sample and observations are independet.

TABLE 1.2 A Data Set on Economic Growth Rates and Country Characteristics

obsno	country	gpcrgdp	govcons60	second60
1	Argentina	0.89	9	32
2	Austria	3.32	16	50
3	Belgium	2.56	13	69
4	Bolivia	1.24	18	12
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
61	Zimbabwe	2.30	17	6

Reviews of basic econometrics: MLR model

- The multivariate linear regression (MLR) model is defined as

$$y_i = \beta_0 + \beta_1 x_{1i} + .. + \beta_k x_{ki} + \mu_i, \tag{1}$$

  where $i = 1, \ldots, n$.
- This model tries to explain variable y in terms of variables $x_1, x_2, \ldots, x_k$.
- y_i is the dependent/explained variable.
- $x_1, x_2, \ldots, x_k$ are the independent/explanatory variables.
- μ_i is the error term.

- The ordinary least square (OLS) estimator of model (1) is

$$\widehat{\beta} = \underset{\beta \in \Theta}{argmin} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - .. - \beta_k x_{ki})^2 .$$

- Let $\mathbf{y} = [y_1,\ y_2,\ \ldots, y_n]^T$, $\boldsymbol{\beta} = [\beta_0,\ \beta_1,\ \ldots, \beta_k]^T$, $\boldsymbol{\mu} = [\mu_1,\ \mu_2,\ \ldots,\ \mu_n]^T$, and

$$\mathbf{x} = \begin{bmatrix} 1 & x_{11} & \ldots & x_{1k} \\ \ldots & & & \\ 1 & x_{n1} & \ldots & x_{nk} \end{bmatrix} .$$

- Thus, in matrix form, we can rewrite model (1) as

$$\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \mu.$$

Reviews of basic econometrics: Algebra for deriving OLS estimator

- Note that OLS estimator minimizes

$$SSR = \mu^T \mu = (\mathbf{y} - \mathbf{x}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{x}\boldsymbol{\beta}) .$$

- SSR is concave in $\boldsymbol{\beta}$. Thus, FOC is sufficient to solve the minimizer!
- By using $\mathbf{y}^T \mathbf{x}\boldsymbol{\beta} = (\mathbf{y}^T \mathbf{x}\boldsymbol{\beta})^T = \boldsymbol{\beta}^T \mathbf{x}^T \mathbf{y}$, we have

$$\mu^T \mu = \mathbf{y}^T \mathbf{y} - 2\boldsymbol{\beta}^T \mathbf{x}^T \mathbf{y} + \boldsymbol{\beta}^T \mathbf{x}^T \mathbf{x}\boldsymbol{\beta}.$$

- The FOC gives

$$\frac{\partial \mu^T \mu}{\partial \boldsymbol{\beta}} = -2\mathbf{x}^T \mathbf{y} + 2\mathbf{x}^T \mathbf{x}\widehat{\boldsymbol{\beta}} = 0$$

$$\Longrightarrow \widehat{\boldsymbol{\beta}} = \left(\mathbf{x}^T \mathbf{x}\right)^{-1} \mathbf{x}^T \mathbf{y}.$$

Reviews of basic econometrics: The Gauss-Markov assumptions

1. Model is linear in Parameters. This assumption states that there is a linear relationship between **y** and **x**.
2. The data is a random sample drawn from the population. This assumption states the random sampling. i.i.d. data satisfies with this assumption.
3. **x** is an n by k matrix of **full** rank. This assumption states that there is no perfect multicollinearity.
4. $E(\mu|\mathbf{x}) = 0$. This assumption - the zero conditional mean assumption - states that the disturbances (error term) average out to 0 for any value of **x**. The assumption implies that $E(\mathbf{y}) = \mathbf{x}\boldsymbol{\beta}$.
5. $E(\mu\mu^T|\mathbf{x}) = \sigma^2 I$. This assumption assumes the homoskedasticity and no autocorrelation.

Reviews of basic econometrics: Unbiasedness

- Under Gauss-Markov assumptions 1-4, $\widehat{\beta}$ is an unbiased estimator of β. That is $E\left(\widehat{\beta}\right) = \beta$

Proof.

$$E\left(\widehat{\beta}\right) = E\left[\left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T\mathbf{y}\right] = E\left[\left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T\left(\mathbf{x}\beta + \mu\right)\right]$$
$$= E\left[\left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T\mathbf{x}\beta\right] + E\left[\left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T\mu\right]$$
$$= \beta,$$

where we use $E(\mu|\mathbf{x}) = 0$ to vanish the second term. ∎

Reviews of basic econometrics: Gauss-Markov theorem

- In class, I will show, under Gauss-Markov assumptions 1-5, $Var\left(\widehat{\boldsymbol{\beta}}\right) = \sigma^2\left(\mathbf{x}^T\mathbf{x}\right)^{-1}$.
- The Gauss-Markov Theorem: conditional on assumptions 1-5, there will be no other linear and unbiased estimator of $\boldsymbol{\beta}$ that has a smaller sampling variance. In other words, the OLS estimator is the Best Linear, Unbiased and Efficient estimator (BLUE).