Advanced Econometrics II

School of Economics and Management - University of Geneva

Christophe Hurlin, Université d'Orléans

University of Orléans

February 2018

Introduction

"Econometrics is the quantitative analysis of actual economic phenomena based on the concurrent development of theory and observation, related by appropriate methods of inference", P. A. Samuelson, T. C. Koopmans, and J. R. N. Stone (1954)

Introduction

Econometrics is fundamentally based on four elements:

1. A **sample** of data
2. An econometric **model**
3. An **estimation** method
4. Some **inference** methods

Introduction

In econometrics, data come from one of the two sources: experiments and non-experimental observations

1. **Experimental** data are based on (randomized controlled) experiments designed to evaluate a treatment or policy or to investigate a causal effect.

2. Data obtained outside an experimental setting are called **observational data** (issued from survey, administrative records etc...)

All of this lecture is devoted to methods for handling real-world observational data

Introduction

Whether the data is experimental or observational, data sets can be mainly distinguished in three types:

1. Cross-sectional data
2. Time series data
3. Panel data

Introduction

Cross-sectional data:

- Data for different entities: workers, households, firms, cities, countries, and so forth.
- No time dimension (even if date of data collection varies somewhat across units, it is ignored).
- Order of data does not matter!

Introduction

Time series data:

- Data for a single entity (person, firm, country) collected at multiple time periods. Repeated observations of the same variables (GDP, prices).
- Order of data is important!
- Observations are typically not independent over time;
- In this case the notion of population corresponds to the **Data Generating Process (DGP)**.

Introduction

Panel data or longitudinal data:

- Data for multiple entities (individuals, firms, countries) in which outcomes and characteristics of each entity are observed at multiple points in time.
- Combine cross-sectional and time series issues.
- Present several advantages with respect to cross-sectional and time series data (depending on the question of interest!).

Introduction

Objectives of the course

The objectives of the course are the following:

1. to understand the specification, estimation, and inference in the context of models that include individual (firm, person, etc.) and/or time effects.
2. to review the standard linear regression model, then to apply it to panel data settings involving 'fixed', 'random', and 'mixed' effects.
3. to extend this linear panel data models to dynamic models with GMM and instrumental variables methods.
4. to extend this linear panel data models to non-linear panel data models

Section 2

Baseline Definitions

Definitions

Definition (Panel data set)

A longitudinal, or panel, data set is one that follows a given sample of individuals over time, and thus provides multiple observations on each individual in the sample (Hsiao 2003, page 2).

Definitions

Terminology and notations:

- **Individual or cross section unit** : country, region, state, firm, consumer, individual, couple of individuals or countries (gravity models), etc.

- **Double index** : i (for cross-section unit) and t (for time)

$$y_{it} \quad \text{for } i = 1, .., n \quad \text{and } t = 1, .., T$$

Definitions

Definition (micro-panel)

A **micro-panel** data set is a panel for which the time dimension T is largely less important than the individual dimension n:

$$T << n$$

Example (micro-panel)

The University of Michigan's Panel Study of Income Dynamics, PSID with 15,000 individuals observed since 1968 is a micro-panel.

Definitions

Definition (macro-panel)

A **macro-panel** data set is a panel for which the time dimension T is similar to the individual dimension n :

$$T \simeq n$$

Example (macro-panel)

A panel of 100 countries with quaterly data since the WW2 is considered as a macro-panel.

Definitions

Remark: some econometric issues are specific to micro or macro panels.

Example (heterogeneity issue)

The heterogeneity issue cannot be tackled with if the time dimension is too small.

Example (non stationarity issue)

The non-stationarity issue (non-stationarity and cointegration tests, VECM, etc.) is only relevant for macro-panel or for panel with a time dimension sufficiently large.

Definitions

Definition (balanced vs. unbalanced panels)

A panel is said to be **balanced** if we have the same time periods, $t = 1, .., T$, for each cross section observation. For an **unbalanced panel**, the time dimension, denoted T_i, is specific to each individual.

Introduction

	A	B	C	D	E
1	i	t	GDP	Consumption	Price
2	1	1980	9,14206153	4,70048037	2,34535718
3	1	1981	9,24657586	4,85203026	2,36755789
4	1	1982	9,29807649	4,89034913	2,52852473
5	1	1983	9,36383369	4,96284463	2,69311929
6	1	1984	9,44208681	5,06259503	2,70394847
7	1	1985	9,50286072	5,14749448	2,7263877
8	1	1986	9,55215545	5,18178355	2,7292636
9	1	1987	9,60326045	5,25749537	2,73954887
10	1	1988	9,6718707	5,35185813	2,76579067
11	1	1989	9,74020362	5,36597602	2,9385268
12	1	1990	9,81743935	5,42934563	2,93015855
13	2	1980	9,15207546	3,95124372	2,74373261
14	2	1981	9,27977315	4,11087386	2,85481366
15	2	1982	9,29155163	4,15888308	2,97297529
16	2	1983	9,36280392	4,24849524	2,91086042
17	2	1984	9,43755562	4,30406509	2,72110462
18	2	1985	9,49829741	4,38202663	2,83238948
19	2	1986	9,53481232	4,4543473	3,00848067
20	2	1987	9,61132881	4,49980967	3,06625131
21	2	1988	9,67325644	4,58496748	3,18265987
22	2	1989	9,71595243	4,67282883	3,25064965
23	2	1990	9,72579526	4,76217393	3,33581228
24	3	1980	9,10052551	4,60517019	2,80486833
25	3	1981	9,19136141	4,68213123	2,78833903
26	3	1982	9,26198361	4,75359019	2,859798
27	3	1983	9,3064684	4,83628191	2,94090115
28	3	1984	9,37822488	4,84418709	3,0435553
29	3	1985	9,41743584	4,93447393	3,08017911
30	3	1986	9,4526589	4,99043259	3,12348764
31	3	1987	9,50315923	5,07517382	3,15734504
32	3	1988	9,58231764	5,18178355	3,18263084
33	3	1989	9,66205273	5,23644196	3,21215328
34	3	1990	9,72752573	5,26785816	3,23254991

Balanced panel with $T = 11$ and $n = 3$

Introduction

	A	B	C	D	E
1	i	t	GDP	Consumption	Price
2	1	1980	9,14206153	4,70048037	2,34535718
3	1	1981	9,24657586	4,85203026	2,36755789
4	1	1982	9,29807649	4,89034913	2,52852473
5	1	1983	9,36383369	4,96284463	2,69311929
6	1	1984	9,44208681	5,06259503	2,70394847
7	1	1985	9,50286072	5,14749448	2,7263877
8	1	1986	9,55215545	NaN	2,7292636
9	1	1987	9,60326045	5,25749537	2,73954887
10	1	1988	9,6718707	5,35185813	2,76579067
11	1	1989	9,74020362	5,36597602	2,9385268
12	1	1990	9,81743935	5,42934563	2,93015855
13	2	1980	9,15207546	3,95124372	2,74373261
14	2	1981	9,27977315	4,11087386	2,85481366
15	2	1982	9,29155163	4,15888308	2,97297529
16	2	1983	9,36280392	4,24849524	2,91086042
17	2	1984	9,43755562	4,30406509	2,72110462
18	2	1985	9,49829741	4,38202663	2,83238948
19	2	1986	9,53481232	4,4543473	3,00848067
20	2	1987	9,61132881	4,49980967	3,06625131
21	2	1988	9,67325644	4,58496748	3,18265987
22	2	1989	NaN	4,67282883	3,25064965
23	2	1990	9,72579526	4,76217393	3,33581228
24	3	1980	9,10052551	4,60517019	2,80486833
25	3	1981	9,19136141	4,68213123	2,78833903
26	3	1982	9,26198361	4,75359019	2,859798
27	3	1983	9,3064684	4,83628191	2,94090115
28	3	1984	9,37822488	4,84418709	3,0435553
29	3	1985	9,41743584	4,93447393	3,08017911
30	3	1986	9,4526589	4,99043259	3,12348764
31	3	1987	9,50315923	5,07517382	3,15734504
32	3	1988	9,58231764	5,18178355	3,18263084
33	3	1989	9,66205273	5,23644196	3,21215328
34	3	1990	9,72752573	5,26785816	3,23254991

Balanced panel with **missing values**

Introduction

	A	B	C	D	E
1	i	t	GDP	Consumption	Price
2	1	1980	9,14206153	NaN	2,34535718
3	1	1981	9,24657586	NaN	2,36755789
4	1	1982	9,29807649	4,89034913	2,52852473
5	1	1983	9,36383369	4,96284463	2,69311929
6	1	1984	9,44208681	5,06259503	2,70394847
7	1	1985	9,50286072	5,14749448	2,7263877
8	1	1986	9,55215545	5,18178355	2,7292636
9	1	1987	9,60326045	5,25749537	2,73954887
10	1	1988	9,6718707	5,35185813	2,76579067
11	1	1989	9,74020362	5,36597602	NaN
12	1	1990	9,81743935	5,42934563	NaN
13	2	1980	9,15207546	3,95124372	2,74373261
14	2	1981	9,27977315	4,11087386	2,85481366
15	2	1982	9,29155163	4,15888308	2,97297529
16	2	1983	9,36280392	4,24849524	2,91086042
17	2	1984	9,43755562	4,30406509	2,72110462
18	2	1985	9,49829741	4,38202663	2,83238948
19	2	1986	9,53481232	4,4543473	3,00848067
20	2	1987	9,61132881	4,49980967	3,06625131
21	2	1988	9,67325644	4,58496748	3,18265987
22	2	1989	9,71595243	4,67282883	3,25064965
23	2	1990	9,72579526	4,76217393	3,33581228
24	3	1980	NaN	4,60517019	2,80486833
25	3	1981	NaN	4,68213123	2,78833903
26	3	1982	NaN	4,75359019	2,859798
27	3	1983	9,3064684	4,83628191	2,94090115
28	3	1984	9,37822488	4,84418709	3,0435553
29	3	1985	9,41743584	4,93447393	3,08017911
30	3	1986	9,4526589	4,99043259	3,12348764
31	3	1987	9,50315923	5,07517382	3,15734504
32	3	1988	9,58231764	5,18178355	3,18263084
33	3	1989	9,66205273	5,23644196	3,21215328
34	3	1990	9,72752573	5,26785816	3,23254991

Unbalanced panel with $T_1 = 9,\ T_2 = 11,\ T_3 = 8$ and $n = 3$

Definitions

Remark: While the mechanics of the unbalanced case are similar to the balanced case, a careful treatment of the unbalanced case requires a formal description of why the panel may be unbalanced, and the sample selection issues can be somewhat subtle.

=> **issues of sample selection and attrition**

Definitions

Definition (Panel data model)

A panel data regression model (or panel data model) is an econometric model specifically designed for panel data.

Section 3

Advantages of Panel Data Sets and Panel Data Models

Advantages of Panel Data

Panel data sets for economic research possess several major advantages over conventional cross-sectional or time-series data sets.

Hsiao, C., (2003, 2nd ed), Analysis of Panel Data, second edition, Cambridge University Press.

Wooldridge J.M., (2001), Econometric Analysis of Cross Section and Panel Data, The MIT Press.

Advantages of Panel Data

What are the main advantages of the panel data sets and the panel data models?

Advantage 1: the phantasm of a larger number of observations

Advantage 2: new economic questions (identification)

Advantage 3: unobservable components

Advantage 4: easier estimation and inference

Advantages of Panel Data

Advantage 1: the phantasm of a larger number of observations

- Panel data usually give the researcher a **large number of data points** ($n \times T$), increasing the degrees of freedom and reducing the collinearity among explanatory variables – hence improving the efficiency of econometric estimates
- But it is often of **phantasm**.... more data points doesn't necessarily imply more information $=>$ **heterogeneity issue !!**

Advantages of Panel Data

Advantage 2: new economic questions (identification)

Longitudinal data allow a researcher to analyze a number of important **economic questions** that cannot be addressed using cross-sectional or time-series data sets.

Advantages of Panel Data

Definition (identification)

The oft-touted power of panel data derives from their theoretical ability to **identify** the effects of specific actions, treatments, or more general policies.

Advantages of Panel Data

Example (Ben-Porath (1973), cited in Hsiao (2003))

Suppose that a cross-sectional sample of married women is found to have an average yearly labor-force participation rate of 50%.
1°) It might be interpreted as implying that each woman in a homogeneous population has a 50 percent chance of being in the labor force in any given year.

2°) It might imply that 50 percent of the women in a heterogeneous population always work and 50 percent never work.

To discriminate between these two stories, **we need to utilize individual labor-force histories (the time dimension)** to estimate the probability of participation in different subintervals of the life cycle.

Advantages of Panel Data

Advantage 3: unobservable components

- **Panel data allows to control for omitted (unobserved or mismeasured) variables.**
- Panel data provides a means of resolving the magnitude of econometric problems that often arises in empirical studies, namely the often heard assertion that the real reason one finds (or does not find) certain effects is the presence of omitted (mismeasured or unobserved) variables that are correlated with explanatory variables.

Advantages of Panel Data

Example: Let us consider a simple regression model.

$$y_{it} = \alpha + \beta' x_{it} + \rho' z_{it} + \varepsilon_{it} \quad i = 1, .., n \quad t = 1, .., T$$

where

- x_{it} and z_{it} are $k_1 \times 1$ and $k_2 \times 1$ vectors of exogenous variables
- α is a constant, β and ρ are $k_1 \times 1$ and $k_2 \times 1$ vectors of parameters
- ε_{it} is $i.i.d.$ over i and t, with $\mathbb{V}\left(\varepsilon_{it}\right) = \sigma_{\varepsilon}^2$
- Let us assume that z_{it} variables **are unobservable and correlated with** x_{it}

$$cov\left(x_{it}, z_{it}\right) \neq 0$$

Advantages of Panel Data

Example (ct'd): The model can be rewritten as

$$y_{it} = \alpha + \beta' x_{it} + \mu_{it}$$

$$\mu_{it} = \rho' z_{it} + \varepsilon_{it}$$

$$cov\left(x_{it}, \mu_{it}\right) \neq 0$$

It is well known that the least-squares regression coefficients of y_{it} on x_{it} are biased

=> **endogeneity bias**

Advantages of Panel Data

Example (ct'd): Let us assume that $z_{i,t} = z_i$, i.e. z values stay constant through time for a given individual but vary across individuals (**individual effects**).

$$y_{it} = \alpha + \beta' x_{it} + \mu_{it}$$

$$\mu_{it} = \rho' z_i + \varepsilon_{it} \quad \text{with} \quad cov\left(x_{it}, \mu_{it}\right) \neq 0$$

Then, if we take the first difference of individual observations over time:

$$y_{it} - y_{i,t-1} = \beta' \left(x_{it} - x_{i,t-1}\right) + \varepsilon_{it} - \varepsilon_{i,t-1}$$

Least squares regression now provides unbiased and consistent estimates of β.

Advantages of Panel Data

Example (ct'd): Let us assume that $z_{i,t} = z_t$, i.e. z values are common for all individuals but vary across time (**common factors**).

$$y_{it} = \alpha + \beta' x_{it} + \rho' z_t + \varepsilon_{it} \quad i = 1, .., n \quad t = 1, .., T$$

Then, if we consider deviation from the mean across individuals at a given time:

$$y_{it} - \overline{y}_t = \beta' \left(x_{it} - \overline{x}_t \right) + \varepsilon_{it} - \overline{\varepsilon}_t$$

where

$$\overline{y}_t = (1/n) \textstyle\sum_{i=1}^{n} y_{it} \quad \overline{x}_t = (1/n) \textstyle\sum_{i=1}^{n} x_{it} \quad \overline{\varepsilon}_t = (1/n) \textstyle\sum_{i=1}^{n} \varepsilon_{it}$$

Least squares regression now provides unbiased and consistent estimates of β.

Advantages of Panel Data

Advantage 4: easier estimation and inference

- Panel data involve two dimensions: a cross-sectional dimension n, and a time-series dimension T.
- We would expect that the computation of panel data estimators would be more complicated than the analysis of **cross-section data** alone (where $T = 1$) or **time series data** alone (where $n = 1$).
- However, in certain cases the availability of panel data can actually simplify the computation and inference.

Advantages of Panel Data

Example (time-series analysis of nonstationary data)

Let us consider a simple $AR(1)$ model.

$$x_t = \rho x_{t-1} + \varepsilon_t$$

where the innovation ε_t is $i.i.d.\left(0, \sigma_{\varepsilon}^2\right)$. Under the non-stationarity assumption $\rho = 1$, it is well known that the asymptotic distribution of the OLS estimator $\widehat{\rho}$ is given by:

$$T \ \left(\widehat{\rho} - 1\right) \underset{T \to \infty}{\overset{d}{\longrightarrow}} \frac{1}{2} \frac{W(1)^2 - 1}{\int_0^1 W(r)^2 \, dr}$$

where $W(.)$ denotes a standard Brownian motion.

Advantages of Panel Data

- Hence, the behavior of the usual test statistics in time series often have to be inferred through computer simulations.
- But if **panel data are available, and observations among cross-sectional units are independent**, then one can invoke the central limit theorem across cross-sectional units to show that
  - the limiting distributions of many estimators remain **asymptotically normal**
  - the **Wald type test statistics are asymptotically chi-square distributed.**
- See for instance Levin and Lin (1993); Im, Pesaran, Shin (1999), Phillips and Moon (1999, 2000), Quah (1994), etc.

Advantages of Panel Data

Example (time-series analysis of nonstationary data)

Let us consider the panel data model

$$x_{i,t} = \rho x_{i,t-1} + \varepsilon_{i,t}$$

where the innovation $\varepsilon_{i,t}$ is $i.i.d.$ $\left(0, \sigma_{\varepsilon}^{2}\right)$ over i and t, then under $H_0 : \rho = 1$:

$$T\sqrt{n}\left(\widehat{\rho}-1\right) \underset{n,T\rightarrow\infty}{\overset{d}{\longrightarrow}} \mathcal{N}\left(0,2\right)$$

Section 4

Issues Involved in using Panel Data

Issues with Panel Data

There are three main issues related to panel data:

1. **Heterogeneity bias => Chapter 1**
2. **Dynamic panel data models (Nickel's bias) => Chapter 2**
3. **Selectivity bias** (not specific to panel data models)

Issues with Panel Data

The heterogeneity issue

When important factors peculiar to a given individual are left out, the typical assumption that economic variable y is generated by a parametric probability distribution function $P(Y|\theta)$), **where θ is an m-dimensional real vector, identical for all individuals at all times**, may not be a realistic one.

Issues with Panel Data

Definition (Parameter heterogeneity issue)

The **parameter heterogeneity issue (in the model specification)** consists in specifying and estimating the individual and/or time-specific effects that exist among cross-sectional or time-series units but are not captured by the included explanatory variables.

Issues with Panel Data

Example: Let us consider a production function (Cobb Douglas) with two factors (labor and capital). We have n countries and T periods. Let us denote:

$$y_{i,t} = \alpha_i + \beta_i k_{i,t} + \gamma_i n_{i,t} + \varepsilon_{i,t}$$

with

- y_{it} the log of the GDP for country i at time t.
- n_{it} the log of the labor employment for country i at time t.
- k_{it} the log of the capital stock for country i at time t.
- $\varepsilon_{i,t}$ $i.i.d.$ $\left(0, \sigma_{\varepsilon}^2\right)$, $\forall\, i$, $\forall\, t$.

Issues with Panel Data

Example (ct'd): In this specification, the elasticities α_i and β_i are specific to each country

$$y_{i,t} = \alpha_i + \beta_i k_{i,t} + \gamma_i n_{i,t} + \varepsilon_{i,t}$$

- Several alternative specifications can be considered.
- First, we can assume that the production function is the same for all countries: in this case we have an **homogeneous specification:**

$$y_{i,t} = \alpha + \beta k_{i,t} + \gamma n_{i,t} + \varepsilon_{i,t}$$

$$\alpha_i = \alpha \quad \beta_i = \beta \quad \gamma_i = \gamma$$

Issues with Panel Data

Example (ct'd): However, an homogeneous specification of the production function for macro aggregated data is meaningless.

- Alternatively, we can consider an heterogeneous Total Factor Productivity (TFP), with $\mathbb{E}\left(\alpha_i + \varepsilon_{i,t}\right) = \alpha_i$, due to institutional organizational factors, etc.

- Thus, we can have a specification with **individual effects** α_i and common slope parameters (elasticities β and γ).

$$y_{i,t} = \alpha_i + \beta k_{i,t} + \gamma n_{i,t} + \varepsilon_{i,t}$$

$$\beta_i = \beta \quad \gamma_i = \gamma$$

Issues with Panel Data

Example (ct'd):

- Finally, we can assume that the labor and/or capital elasticities are different across countries.
- In this case, we will have an heterogeneous specification of the panel data model (**heterogeneous panel**).

$$y_{i,t} = \alpha_i + \beta_i k_{i,t} + \gamma_i n_{i,t} + \varepsilon_{i,t}$$

Issues with Panel Data

Example (ct'd):

$$y_{i,t} = \alpha_i + \beta_i k_{i,t} + \gamma_i n_{i,t} + \varepsilon_{i,t}$$

In this case, there are two solutions to estimate the parameters

1. The first solution consists in using n times series models to produce **some group-mean estimates of the elasticities**.
2. Consider a **random coefficient model**. In this case, we assume that parameters β_i and γ_i and randomly distributed, with for instance:

$$\beta_i \;\; i.i.i \;\left(\overline{\beta}, \sigma_{\beta}^2\right) \qquad \gamma_i \;\; i.i.i \;\left(\overline{\gamma}, \sigma_{\gamma}^2\right)$$

Issues with Panel Data

Fact (Heterogeneity bias)

Ignoring such heterogeneity (in slope and/or constant) could lead to inconsistent or meaningless estimates of interesting parameters.

Issues with Panel Data

The heterogeneity bias

- Let us consider a simple linear model with individual effects and only one explicative variable x_i (common slope) as a DGP.

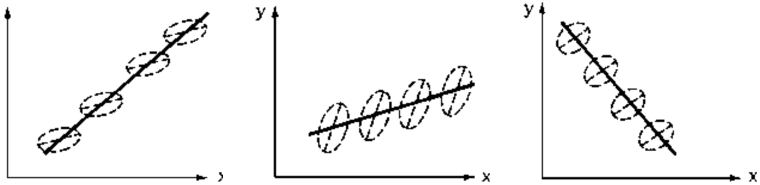
$$y_{it} = \alpha_i + \beta x_{it} + \varepsilon_{it}$$

- Let us assume that all NT observations $\{x_{it}, y_{it}\}$ are used to estimate the following homogeneous model.

$$y_{it} = \alpha + \beta x_{it} + \varepsilon_{it}$$

Issues with Panel Data

The heterogeneity bias



Source: Hsiao (2003)

Broken ellipses= point scatter for an individual over time

Broken straight lines = individual regressions.

Solid lines = least-squares regression using all NT observations

Issues with Panel Data

The heterogeneity bias

- All of these figures depict situations in which biases (on $\widehat{\beta}$) arise in pooled least-squares estimates because of heterogeneous intercepts.
- Obviously, in these cases, pooled regression ignoring heterogeneous intercepts should never be used.
- Moreover, the direction of the bias of the pooled slope estimates cannot be identified a priori; it can go either way.

Issues with Panel Data

The heterogeneity bias

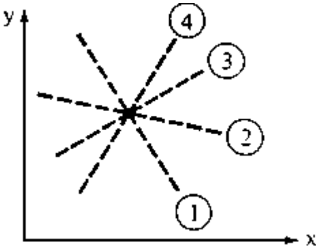
Let us consider another example. The true DGP is heterogeneous

$$y_{it} = \alpha_i + \beta_i x_{it} + \varepsilon_{it}$$

and we use all NT observations $\{x_{it}, y_{it}\}$ to estimate the homogeneous model.

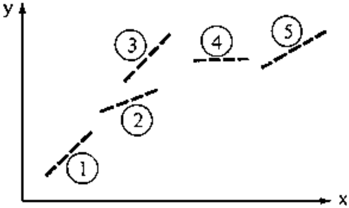
$$y_{it} = \alpha + \beta x_{it} + \varepsilon_{it}$$

Issues with Panel Data



- **Pooling the NT observations**, assuming identical parameters for all cross-sectional units, lead to **nonsensical results**
- It leads to estimate an **average** of coefficients that differ across individuals **(the phantasm of the NT observations)**

Issues with Panel Data



In this case, pooling gives rise to the **false inference** that the pooled relation is curvilinear.

Issues with Panel Data

Fact (Heterogeneity issue)

In both cases, the classic paradigm of the "representative agent" simply does not hold, and pooling the data under homogeneity assumption makes no sense.

Section 5

Course Information

Course Information

Course outline

Chapter 1: Linear Panel Models and Heterogeneity

Chapter 2: Dynamic Panel Data Models

Chapter 3: Non Stationarity and Panel Data Models

Chapter 4: Non Linear Panel Data Models

Course Information

Books: advanced econometrics (not specific to panel data)

Amemiya T. (1985), Advanced Econometrics. Harvard University Press.

Cameron A.C. and P.K. Trivedi (2005), Microeconometrics: Methods and Applications, Cambridge University Press, Cambridge, U.S.A.

Davidson R. (2000), Econometric Theory, Blackwell Publishers, Oxford.

Davidson R. and J. Mackinnon (2004), Econometric Theory and Methods, Oxford University Press, Oxford.

Greene W. (2007), Econometric Analysis, sixth edition, Pearson.

Johnston J. and J. Dinardo (1997), Econometric Methods, 4th ed., The McGraw-Hill Companies Inc., New York.

Course Information

Books: panel data econometrics (I/II)

- Arellano M. (2003), Panel Data Econometrics, Oxford University Press, U.K.
- Baltagi B. (2005), Econometric Analysis of Panel Data, John Wiley & Sons, New York, Third edition.
- Baltagi B. (2006), Panel Data Econometrics: Theoretical Contributions and Empirical Applications, Elsevier, Amsterdam.
- Hsiao (2003), Analysis of Panel Data, Cambridge University Press (**recommended**).
- Krishnakumar J. and E. Ronchetti (2000), Panel Data Econometrics: Future Directions, Elsevier, Amsterdam.
- Krishnakumar J. and E. Ronchetti (1983), Limited Dependent and Qualitative Variables in Econometrics, Cambridge University Press.

Course Information

Books: panel data econometrics (II/II)

- Matyas L. and P. Sevestre (2008), The Econometrics of Panel Data, Springer-Verlag, Berlin.
- Wooldridge J.M (2010), Econometric Analysis of Cross Section and Panel Data, MIT Press. (**recommended).**

Books: panel data econometrics (in French)

- Pirotte A. (2011), Econométrie des données de panel, Economica.
- Sevestre P. (2002), Econométrie des données de panel, Dunod, Paris.

Course Information

Additional references (articles and surveys) among many others...

Baltagi, B.H. and Kao, C. (2000), "Nonstationary panels, cointegration in panels and dynamic panels : a survey", in Advances in Econometrics, 15, edited by B. Baltagi et C. Kao, 7-51, Elsevier Science.

Dumitrescu E. and Hurlin C. (2012), "Testing for Granger Non-causality in Heterogeneous Panels", Economic Modelling, 29, 1450-1460.

Hurlin, C. and Mignon, V. (2005), "Une synthèse des tests de racine unitaire sur données de panel", Economie et Prévision, 169-171, 253-294

Hurlin C. et Mignon, V. (2007), "Une Synthèse des Tests de Cointégration sur Données de Panel", Economie et Prévision, 180-181, 241- 265

End of the general introduction

Christophe Hurlin (University of Orléans)