An introduction to econometrics of panel data

F. Legendre

Université Paris-Est

Seminars given at UKMA departement of finance

Seminars organisation

- Two things to know about econometrics (among others)
- Panel data
- The random effect hypothesis
- A menagerie of estimators
- How to explain the discrepancies between estimators?

First thing to know: Frisch-Waugh theorem

$$\underline{y} = \hat{a}\,\underline{x} + \hat{b}\,\underline{z} + \underline{\hat{u}} \qquad \text{(a)}$$

where $\underline{x}$ is a variable of interest and $\underline{z}$ a control variable

$$p_z(\underline{y}) = p_z(\hat{a}\,\underline{x} + \hat{b}\,\underline{z} + \underline{\hat{u}}) = \hat{a}\,p_z(\underline{x}) + \hat{b}\,p_z(\underline{z}) + p_z(\underline{\hat{u}}) \qquad \text{(b)}$$

where

- $p_z(\cdot)$ is the orthogonal projection onto $\mathcal{V}(\underline{z})$
- $\mathcal{V}(\underline{z})$ is the vectorial space generated by $\underline{z}$

but

- $p_z(\underline{z}) = \underline{z}$
- $p_z(\underline{\hat{u}}) = \underline{0}$

$$\left[\underline{y} - p_z(\underline{y})\right] = \hat{a}\left[\underline{x} - p_z(\underline{x})\right] + \underline{\hat{u}} \qquad \text{(a) - (b)}$$

Example

$$\underline{y} = \hat{a}\,\underline{x} + \hat{b}\,\underline{1} + \underline{\hat{u}}$$

$$\left[\underline{y} - \overline{y}\,\underline{1}\right] = \hat{a}\left[\underline{x} - \overline{x}\,\underline{1}\right] + \underline{\hat{u}}$$

Second thing to know: what to do when $\mathrm{V}(\underline{u}) \neq \sigma^2 I_N$?

Two ways

- Use the best estimator i.e. the FIML estimator as OLS is neither BLUE nor the FIML estimator, but
  - BLUE is not always a strong property
  - FIML estimator is the best only asymptoticaly
  - FIML estimator is tedious to compute
- Use OLS but do not use the "wrong" formulas
  - $\hat{a} = (X'X)^{-1}X'\underline{y}$
  - $\mathrm{V}(\hat{a}) = (X'X)^{-1}X'\Omega X(X'X)^{-1}$ if $\mathrm{V}(\underline{u}) = \Omega$
  - Get $\widehat{\Omega}$ an convergent estimator of Ω,
  - $\widehat{\mathrm{V}}(\hat{a}) = (X'X)^{-1}X'\widehat{\Omega}X(X'X)^{-1}$ and the tests are valid

Panel data

- Cross section data and the limits of cross section data
- Panel data specificities
  - Individuals observed more than one time – repeated cross section
  - Time series from more than one individual – pooled time series
  - Double subscript: $y_{it}, x_{it}, \ldots \; i = 1, \ldots, N$ and $t = 1, \ldots, T$
- Different kinds of panel data
  - The cross dimension dominates
    - From survey: the same households are surveyed many times
    - From administrative data: social contributions of employers, ...
  - The time dimension dominates
    - Pooled data of countries
    - Pooled high frequencies financial data

The random effect hypothesis - I: Balestra and Nerlove 1966

Basic regression model

$$y_{it} = a x_{it} + b + u_{it}$$

- Estimate an average behaviour
- Taking account individual specificities

As always, random is the preferred way to capture ignorance

- On time series, the usual way to capture persistence is the "AR(1) hypothesis"

$$u_t = \rho u_{t-1} + \varepsilon_t$$

- On panel data, the usual way to capture individual specificities is the "random effect hypothesis"

$$u_{it} = \mu_i + v_{it}$$

Data layout

Matrix notations for the basic linear model (vectors are underlined)

$$\underline{y} = X\underline{a} + \underline{u}$$

Observations are sorted by individual then by period

$$\underline{z} = \begin{pmatrix} \underline{z}_1 \\ \underline{z}_2 \\ \vdots \\ \underline{z}_i \\ \vdots \\ \underline{z}_N \end{pmatrix} \qquad \underline{z}_i = \begin{pmatrix} z_{i1} \\ z_{i2} \\ \vdots \\ z_{it} \\ \vdots \\ z_{iT} \end{pmatrix}$$

The random effect hypothesis - II

Let us suppose

- no dependencies between individuals
- usual homoscedasticity hypothesis

Then

$$\begin{aligned}
\mathrm{E}(\mu_i) &= 0 \quad \forall\ i \\
\mathrm{E}(v_{it}) &= 0 \quad \forall\ i,t \\
\mathrm{V}(\mu_i) &= \sigma_\mu^2 \quad \forall\ i \\
\mathrm{V}(v_{it}) &= \sigma_v^2 \quad \forall\ i,t \\
\mathrm{Cov}(\mu_i,\mu_j) &= 0 \quad \forall\ i \neq j \\
\mathrm{Cov}(v_{it},v_{j\tau}) &= 0 \quad \forall\ i \neq j, t \neq \tau \\
\mathrm{Cov}(\mu_i,v_{jt}) &= 0 \quad \forall\ i,j,t
\end{aligned}$$

$$\mathrm{V}(\underline{u}_i) = \Sigma \ \ \forall\ i \quad \text{and} \quad \mathrm{E}(\underline{u}_i\underline{u}_j') = 0_T \ \ \forall\ i \neq j$$

Shape of variance-covariance matrix of $\underline{u}$

The variance-covariance matrix is bloc-diagonal

$$V(\underline{u}) = \begin{pmatrix} \Sigma & 0_T & \cdots & 0_T \\ 0_T & \Sigma & \cdots & 0_T \\ \vdots & \vdots & \ddots & \vdots \\ 0_T & 0_T & \cdots & \Sigma \end{pmatrix} = I_N \otimes \Sigma$$

with

$$\Sigma = \begin{pmatrix} \sigma_\mu^2 + \sigma_v^2 & \sigma_\mu^2 & \cdots & \sigma_\mu^2 \\ \sigma_\mu^2 & \sigma_\mu^2 + \sigma_v^2 & \cdots & \sigma_\mu^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\mu^2 & \sigma_\mu^2 & \cdots & \sigma_\mu^2 + \sigma_v^2 \end{pmatrix} = \sigma_v^2 W_T + (T\sigma_\mu^2 + \sigma_v^2) B_T$$

W et B matrix

$$\begin{cases} W_T = I_T - \frac{1}{T} J_T \\ B_T = \frac{1}{T} J_T \end{cases} \quad \text{with } J_T = \underline{1}_T \underline{1}_T' = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

$$\mathrm{V}(\underline{u}) = \sigma_v^2 W + (T\sigma_\mu^2 + \sigma_v^2) B$$

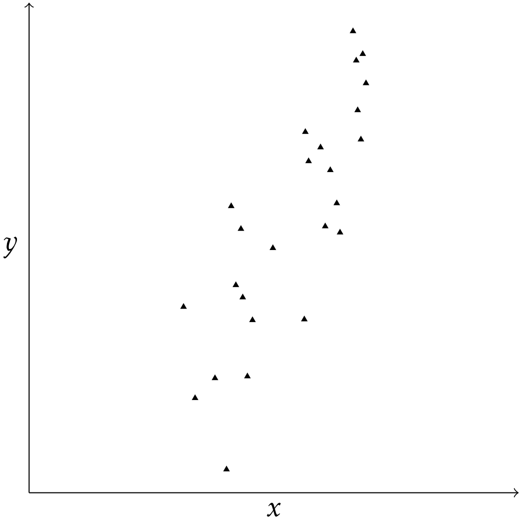
$$\begin{cases} W = I_N \otimes W_T \\ B = I_N \otimes B_T \end{cases}$$

$z_{it} - \overline{z}_i$ is the element of $W\underline{z}$
$\overline{z}_i$ is the element of $B\underline{z}$

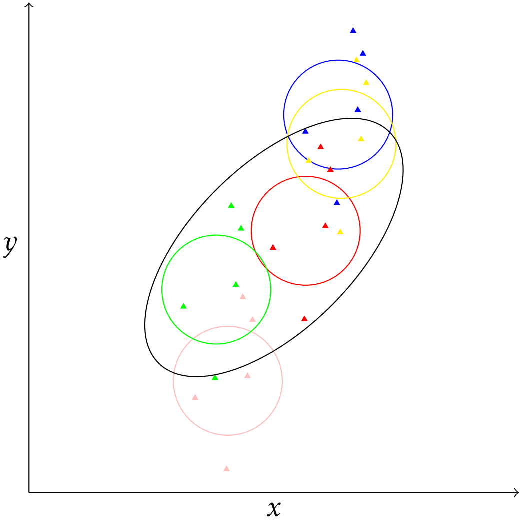
$$W^2 = W, \;\; W' = W, \;\; B^2 = B, \;\; B' = B,$$

$$W + B = I_{NT} \;\text{ and }\; W'B = 0_{NT}$$

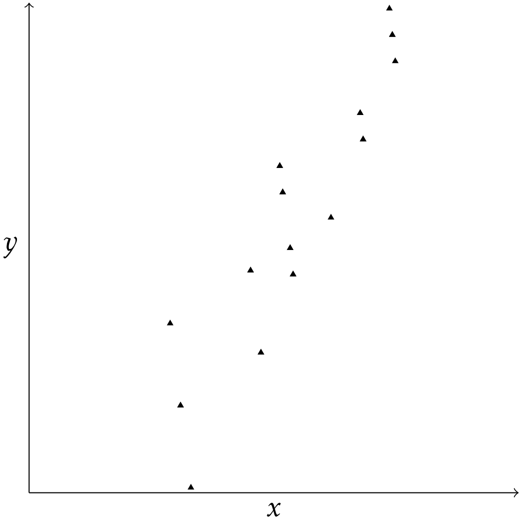
A strong link between x and y...



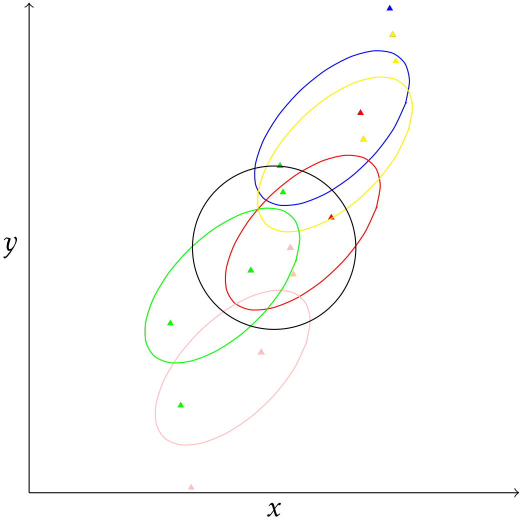
. . . obtained from ***between*** correlation



Another strong link between x and y



. . . obtained from *within* correlation



Application - I

Estimation on a production function on sectoral data from French national accounts
Departure equation, Cobb-Douglas specification

$$Q = AK^{\alpha} L^{\beta} e^{gt}$$

Estimated equation

$$\ln(q_{it}) = \alpha \ln(k_{it}) + \beta \ln(\ell_{it}) + g\,t + a + u_{it}$$