Model Based on Panel Data

Yusep Suparman

Introduction

Panel data: data of repeated observations over the same observations collected over a number of periods.

- Allowing for estimating more complicated and realistic models
  - Analyzing changes on individual level
- Increasing efficiency
- Reducing omitted variable bias

Introduction

Long data structure

1	1	y_{11}	$\mathbf{x}_{11}^T$
1	2	y_{12}	$\mathbf{x}_{12}^T$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	T	y_{1T}	$\mathbf{x}_{1T}^T$
2	1	y_{21}	$\mathbf{x}_{21}^T$
2	2	y_{22}	$\mathbf{x}_{22}^T$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
2	T	y_{2T}	$\mathbf{x}_{2T}^T$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	1	y_{N1}	$\mathbf{x}_{N1}^T$
N	2	y_{N2}	$\mathbf{x}_{N2}^T$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	T	y_{NT}	$\mathbf{x}_{NT}^T$

Wide data Structure

1	y_{11}	y_{12}	$\cdots$	y_{1T}	$\mathbf{x}_{11}^T$	$\mathbf{x}_{12}^T$	$\cdots$	$\mathbf{x}_{1T}^T$
2	y_{21}	y_{22}	$\cdots$	y_{2T}	$\mathbf{x}_{21}^T$	$\mathbf{x}_{22}^T$	$\cdots$	$\mathbf{x}_{2T}^T$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
N	y_{N1}	y_{N2}	$\cdots$	y_{NT}	$\mathbf{x}_{N1}^T$	$\mathbf{x}_{N2}^T$	$\cdots$	$\mathbf{x}_{NT}^T$

Introduction

General model

$$y_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta}_{it} + \varepsilon_{it},\ i = 1,2,\ldots,N,\ t = 1,2,\ldots,T$$

This is too general to be useful. More structured models is required.

Commonly used models

$$y_{it} = \alpha_i + \mathbf{x}_{it}^T \boldsymbol{\beta} + \varepsilon_{it}$$

$$E(\varepsilon_{it}) = 0,\ \ \operatorname{var}(\varepsilon_{it}) = \sigma^2,\ \ \operatorname{cov}(\varepsilon_{it}, \varepsilon_{js}) = 0$$
$$i, j = 1,2,\ldots,N;\ t, s = 1,2,\ldots,T;\ i \neq j,\ t \neq s$$

Fixed Effects Model

Specification

$$y_{it} = \alpha_i + \mathbf{x}_{it}^T \boldsymbol{\beta} + \varepsilon_{it}$$

$$E(\varepsilon_{it}) = 0, \ \ \text{var}(\varepsilon_{it}) = \sigma^2, \ \ \text{cov}(\varepsilon_{it}, \varepsilon_{js}) = 0 \quad \text{or} \quad \varepsilon_{it} \sim IID(0, \sigma_\varepsilon^2)$$
$$i, j = 1, 2, \ldots, N; \ t, s = 1, 2, \ldots, T; \ i \neq j, \ t \neq s$$

$\mathbf{x}_{it}$ does not contain unit constant. The model without intercept

α_i captures the **fixed (unknown parameter) effects** of variables being peculiar to the i-th individual and **constant over time.** It may **correlate** to the explanatory variables in $\mathbf{x}_{it}$

Fixed Effects Model

Introducing individual dummy variables for estimating individual specific fixed effect

$$d_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

$$y_{it} = \sum_{j=1}^{N} \alpha_i d_{ij} + \mathbf{x}_{it}^T \boldsymbol{\beta} + \varepsilon_{it}$$

Fixed Effects Model

Alternative formulation, by defining individual reference (say, h-th individual as the reference

$$y_{it} = \mu + \sum_{\substack{j=1 \\ j \neq h}}^{N} \alpha_i d_{ij} + \mathbf{x}_{it}^T \boldsymbol{\beta} + \varepsilon_{it}$$

μ is the individual specific fixed effect for the h-th individual. While α_i is **the deviance** of individual specific fixed effect for the i-th individual from the h-th individual.

Fixed Effects Model

Estimation

Under

$$y_{it} = \sum_{j=1}^{N} \alpha_i d_{ij} + \mathbf{x}_{it}^T \boldsymbol{\beta} + \varepsilon_{it}$$

$$d_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

OLS can be applied for estimating $\alpha_1, \ldots, \alpha_N$ and $\boldsymbol{\beta}$, least square dummy variable (LSDV) estimator.

Fixed Effects Model

An alternative to LSDV, introducing

$$\bar{y}_i = \alpha_i + \bar{\mathbf{x}}_i^T \boldsymbol{\beta} + \bar{\varepsilon}_i$$

We have the regression in deviations from individual means

$$y_{it} - \bar{y}_i = \left(\mathbf{x}_{it}^T - \bar{\mathbf{x}}_i^T\right)\boldsymbol{\beta} + \left(\varepsilon_{it} - \bar{\varepsilon}_i\right)$$

OLS is applied to the transformed data i.e. as deviances from individual means (within transformation).

$$\hat{\boldsymbol{\beta}}_{FE} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right)\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right)^T\right]^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right)\left(y_{it} - \bar{y}_i\right)$$

Fixed Effects Model

The estimate of individual specific fixed effect

$$\hat{\alpha}_i = \bar{y}_i - \bar{x}_i^T \hat{\boldsymbol{\beta}}_{FE}$$

The covariance of $\hat{\boldsymbol{\beta}}_{FE}$ is

$$\operatorname{cov}\left\{\hat{\boldsymbol{\beta}}_{FE}\right\} = \sigma_\varepsilon^2 \left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)^T\right]^{-1}$$

With the consistent estimate of σ_ε^2

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - \bar{y}_i - (\mathbf{x}_{it} - \bar{\mathbf{x}}_{it})^T \hat{\boldsymbol{\beta}}_{FE}\right)^2$$

Correction can be applied by subtracting the denominator by number of beta.

Random Effects Model

Specification

$$y_{it} = \mu + \alpha_i + \mathbf{x}_{it}^T \boldsymbol{\beta} + \varepsilon_{it}$$

$$\varepsilon_{it} \sim IID\left(0, \sigma_\varepsilon^2\right) \quad \text{and} \quad \alpha_i \sim IID\left(0, \sigma_\alpha^2\right)$$

$\mathbf{x}_{it}$ does not contain unit constant. μ is the intercept of the model, and unit constant may be integrated into $\mathbf{x}_{it}$

α_i captures the **random effects** of variables being peculiar to the i-th individual and **constant over time.** It does **not correlate** to the explanatory variables.

Random Effects Model

Estimation

Assuming alphas as a random variable results in an autocorrelated error component structure $\varepsilon_{it}^{*} = \alpha_i + \varepsilon_{it}$

Implying incorrect OLS estimates of standard errors, recalling for GLS estimator.

Defining $\boldsymbol{\varepsilon}_i^{*} = \alpha_i \boldsymbol{\iota}_T + \boldsymbol{\varepsilon}_i$ with $\boldsymbol{\iota}$ is a T dimensional vector unit and $\boldsymbol{\varepsilon}_i^{T} = \begin{bmatrix} \varepsilon_{i1} & \varepsilon_{i2} & \cdots & \varepsilon_{iT} \end{bmatrix}$

The covariance matrix can be expressed as

$$\operatorname{cov}\left\{\boldsymbol{\varepsilon}_i^{*}\right\} = \boldsymbol{\Omega} = \sigma_\alpha^2 \boldsymbol{\iota}_T \boldsymbol{\iota}_T^{T} + \sigma_\varepsilon^2 \mathbf{I}_T$$

Random Effects Model

The GLS estimator is obtained by transforming individual vector variable by

$$\mathbf{\Omega}^{-1} = \sigma_\varepsilon^{-2}\left[\mathbf{I}_T - \frac{\sigma_\alpha^2}{\sigma_\varepsilon^2+\sigma_\alpha^2}\boldsymbol{\iota}_T\boldsymbol{\iota}_T^T\right] = \sigma_\varepsilon^{-2}\left[\left(\mathbf{I}_T - \frac{1}{T}\boldsymbol{\iota}_T\boldsymbol{\iota}_T^T\right) + \psi\frac{1}{T}\boldsymbol{\iota}_T\boldsymbol{\iota}_T^T\right]$$

$$\psi = \frac{\sigma_\alpha^2}{\sigma_\varepsilon^2+\sigma_\alpha^2}$$

It results in

$$\hat{\boldsymbol{\beta}}_{GLS} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{x}_{it}-\overline{\mathbf{x}}_i\right)\left(\mathbf{x}_{it}-\overline{\mathbf{x}}_i\right)^T + \psi T\sum_{i=1}^{N}\left(\overline{\mathbf{x}}_i-\overline{\mathbf{x}}\right)\left(\overline{\mathbf{x}}_i-\overline{\mathbf{x}}\right)^T\right)^{-1}$$

$$\times\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{x}_{it}-\overline{\mathbf{x}}_{it}\right)\left(y_{it}-\overline{y}_i\right)^T + \psi T\sum_{i=1}^{N}\left(\overline{\mathbf{x}}_i-\overline{\mathbf{x}}\right)\left(\overline{y}_i-\overline{y}\right)^T\right)^T$$

Random Effects Model

From the formula above we can derive

$$\hat{\boldsymbol{\beta}}_{GLS} = \Delta\hat{\boldsymbol{\beta}}_B + \left(\mathbf{I}_K - \Delta\right)\hat{\boldsymbol{\beta}}_{FE}$$

with

$$\hat{\boldsymbol{\beta}}_B = \left(\sum_{i=1}^{N}\left(\overline{\mathbf{x}}_i - \overline{\mathbf{x}}\right)\left(\overline{\mathbf{x}}_i - \overline{\mathbf{x}}\right)^T\right)^{-1}\sum_{i=1}^{N}\left(\overline{\mathbf{x}}_i - \overline{\mathbf{x}}\right)\left(\overline{y}_i - \overline{y}\right)^T$$

the between estimator, i.e. the OLS estimator in the individual means model

$$\overline{y}_i = \mu + \overline{\mathbf{x}}_i^T\boldsymbol{\beta} + \alpha_i + \overline{\varepsilon}_i,\ i = 1, 2, \ldots, N$$

$\boldsymbol{\Delta}$ is proportional to the invers matrix of the covariance matrix of between estimator

Random Effects Model

EGLS: subtitute the varainces by their estimates -> random effect estimator $\hat{\boldsymbol{\beta}}_{RE}$

$$\hat{\sigma}_{\alpha}^{2} = \hat{\sigma}_{B}^{2} + \frac{1}{T}\hat{\sigma}_{\varepsilon}^{2}$$

$$\hat{\sigma}_{B}^{2} = \frac{1}{N}\sum_{i=1}^{N}\left(\bar{y}_{i} - \hat{\mu}_{B} - \bar{\mathbf{x}}_{i}^{T}\hat{\boldsymbol{\beta}}_{B}\right)^{2}$$

Estimated covariance matrix is

$$\operatorname{cov}\left\{\hat{\boldsymbol{\beta}}_{RE}\right\} = \hat{\sigma}_{\varepsilon}^{2}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i}\right)\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i}\right)^{T} + \hat{\psi}T\sum_{i=1}^{N}\left(\bar{\mathbf{x}}_{i} - \bar{\mathbf{x}}\right)\left(\bar{\mathbf{x}}_{i} - \bar{\mathbf{x}}\right)^{T}\right)$$

FEM or REM?

- Substantive

the individual specific effect basically represents regressors not including in the model. If those factors substantively correlate to the regressors in the model, we should chose FEM. If we have no information on the correlations we may adopt REM.

the use of FEM may reduce endogeneity due to correlated omitted variable, omitted variable bias

FEM or REM?

- Hausman's test (empirical)
  - Hyphotheses

$$H_0 : \boldsymbol{\beta}_{REM} = \boldsymbol{\beta}_{FEM}$$
$$H_1 : \boldsymbol{\beta}_{REM} \neq \boldsymbol{\beta}_{FEM}$$

  - Statistical test

$$\chi^2 = \left(\hat{\boldsymbol{\beta}}_{FEM} - \hat{\boldsymbol{\beta}}_{REM}\right)^T \left(\hat{\text{cov}}\left(\hat{\boldsymbol{\beta}}_{FEM}\right) - \hat{\text{cov}}\left(\hat{\boldsymbol{\beta}}_{FEM}\right)\right)^{-1} \left(\hat{\boldsymbol{\beta}}_{FEM} - \hat{\boldsymbol{\beta}}_{REM}\right) \sim \chi^2_K$$

Title Lorem Ipsum

- Dolor Sit Amet
- Consectetuer Elit
- Nunc Viverra
- Pellentesque Habitant
- Lorem Ipsum