

Panel Data Regression Models

(c) Pongsa Pornchaiwiseskul, Faculty of Economics, Chulalongkorn University

1

What is Panel Data? (1)

- Multiple dimensioned
- Dimensions, e.g.,
 - cross-section and time
 - node-to-node

(c) Pongsa Pornchaiwiseskul, Faculty of Economics, Chulalongkorn University

3

Covered Topics

- What is Panel Data?
- Pooled Regression
- Fixed Effect Models
- Random Effect Models
- Other Panel Data Models

(c) Pongsa Pornchaiwiseskul, Faculty of Economics, Chulalongkorn University

2

What is Panel Data? (2)

Node-to-Node Example

Y_{ij} = flow from node i to node j

$X2_{ij}$ = unit cost between node i
and node j

$X3_{ij}$ = capacity between node i
and node j

(c) Pongsa Pornchaiwiseskul, Faculty of Economics, Chulalongkorn University

4

What is Panel Data?(3)

with cross-section(i) and time(t) indices

i	t	Y_{it}	$X1_{it}$	XK_{it}
1	1	Y_{11}	$X1_{11}$...	XK_{11}
:	:	:	:	...	:
1	T_1	Y_{1T_1}	$X1_{1T_1}$...	XK_{1T_1}
:	:	:	:	:
N	1	Y_{N1}	$X1_{N1}$...	XK_{N1}
:	:	:	:	...	:
N	T_N	Y_{NTN}	$X1_{NTN}$	XK_{NTN}

(c) Pongsa Pornchaisweskul, Faculty of Economics, Chulalongkorn University

5

Linear Model for Panel Data (1)

$$Y_{it} = \beta_{1it}X_{1it} + \beta_{2it}X_{2it} + \dots + \beta_{Kit}X_{Kit} + \varepsilon_{it}$$

- beta coefficients could be

time - invariant : $\beta_{kit} = \beta_{ki}$ for $\forall t$

section - invariant : $\beta_{kit} = \beta_{kt}$ for $\forall i$

both : $\beta_{kit} = \beta_k$ for $\forall i, t$

(c) Pongsa Pornchaisweskul, Faculty of Economics, Chulalongkorn University

7

Balanced Panel Data

- Each cross-sections has equal number of time periods
or $T_1 = T_2 = \dots = T_N$
- Simple Data Structure
- Less complicated computation

(c) Pongsa Pornchaisweskul, Faculty of Economics, Chulalongkorn University

6

Linear Model for Panel Data (2)

- variance of error terms could be

time - invariant : $V(\varepsilon_{it}) = \sigma_i^2$ for $\forall t$

section - invariant : $V(\varepsilon_{it}) = \sigma_t^2$ for $\forall i$

both : $V(\varepsilon_{it}) = \sigma^2$ for $\forall i, t$

(c) Pongsa Pornchaisweskul, Faculty of Economics, Chulalongkorn University

8

Pooled Regression (1)

General Assumption:

$$Y_{it} = \beta_1 X_{1it} + \beta_2 X_{2it} + \dots + \beta_K X_{Kit} + \varepsilon_{it}$$

where

- 1) all the β coefficients are both time and cross-sectional invariant

Pooled Regression (2)

i	t	Y_{it}	$X1_{it}$...	XK_{it}
1	1	Y_{11}	$X1_{11}$...	XK_{11}
1	2	Y_{12}	$X1_{12}$...	XK_{12}
:	:	:	:	...	:
1	T	Y_{1T}	$X1_{1T}$...	XK_{1T}
:	:	\vdots	\vdots	X	\vdots
N	1	Y_{N1}	$X1_{N1}$...	XK_{N1}
N	2	Y_{N2}	$X1_{N2}$...	XK_{N2}
:	:	:	:	...	:
N	T	Y_{NT}	$X1_{NT}$...	XK_{NT}

Pooled Regression (2)

- 2) homoscedastic error terms

$$V(\varepsilon_{it}) = \sigma^2 \text{ for all } i, t$$

or it is also time-invariant and section-invariant

=> OLS applies.

Fixed Effect Models (1)

Also called LS Dummy Variable (LSDV) Model. X_1 is constant.

$$Y_{it} = \beta_{1i} + \beta_2 X_{2it} + \dots + \beta_K X_{Kit} + \varepsilon_{it}$$

- 1) all the β coefficients are time-invariant and cross-sectional invariant except that β_1 is section-variant but time-invariant

Fixed Effect Models (2)

2) homoscedastic error terms

$$V(\varepsilon_{it}) = \sigma^2 \text{ for all } i,t$$

or it is time-invariant and section-invariant

=> OLS applies if dummy variables are introduced.

Fixed Effect Models (4)

i	t	Y_{it}	D_{1it}	...	D_{Nit}
1	1	Y_{11}	1	...	0
:	:	:	:	...	:
1	T_1	Y_{1T_1}	1	...	0
:	:	:	:	...	:
N	1	Y_{N1}	0	...	1
:	:	:	:	...	:
N	T_N	Y_{NTN}	0	...	1

Fixed Effect Models (3)

Equivalent LSDV

$$Y_{it} = \beta_{11}D_{1it} + \beta_{12}D_{2it} + \dots + \beta_{1N}D_{Nit} + \beta_2X_{2it} + \dots + \beta_KX_{Kit} + \varepsilon_{it}$$

where

$$D_{jit} = 1 \text{ if } i=j, j=1,\dots,N \\ = 0 \text{ otherwise}$$

Note that

$$D_{1it} + D_{2it} + \dots + D_{Nit} = 1 \text{ for all } i,t$$

Fixed Effect Models (5)

Alternative form of FEM

$$Y_{it} = \beta_{1t} + \beta_2X_{2it} + \dots + \beta_KX_{Kit} + \varepsilon_{it}$$

Note that β_1 is time-variant but section-invariant, instead

Equivalent LSDV for this FEM is as follows:

Fixed Effect Models (6)

Equivalent LSDV

$$Y_{it} = \beta_{11}D_{1it} + \beta_{12}D_{2it} + \dots + \beta_{1T}D_{Tit} \\ + \beta_2X_{2it} + \dots + \beta_KX_{Kit} + \varepsilon_{it}$$

where

$$D_{jit} = 1 \text{ if } t=j, j=1, \dots, T \\ = 0 \text{ otherwise}$$

Note that

$$D_{1it} + D_{2it} + \dots + D_{Tit} = 1 \text{ for all } i, t$$

(c) Pongsa Pornchaiwiseskul, Faculty of Economics, Chulalongkorn University

17

Fixed Effect Models (8)

Heteroscedasticity

-cross-section weight

$$V(\varepsilon_{it}) = \sigma_i^2 \text{ for } \forall t$$

-time weight

$$V(\varepsilon_{it}) = \sigma_t^2 \text{ for } \forall i$$

=>WLS applies

(c) Pongsa Pornchaiwiseskul, Faculty of Economics, Chulalongkorn University

19

Fixed Effect Models (7)

Common Problems

- many dummy variables required
- multi-collinearity problem likely
- interpretation of variant coefficients
- What if error terms are heteroscedastic?

(c) Pongsa Pornchaiwiseskul, Faculty of Economics, Chulalongkorn University

18

Fixed Effect Models (9)

Heteroscedasticity

-cross-section covariance

$$\text{COV}(\varepsilon_{it}, \varepsilon_{jt}) = \sigma_{ij} = \sigma_{ji} \text{ for } \forall t$$

-auto-correlation

$$\text{COV}(\varepsilon_{is}, \varepsilon_{it}) = \sigma_{st} = \sigma_{ts} \text{ for } \forall i$$

=>FGLS applies

(c) Pongsa Pornchaiwiseskul, Faculty of Economics, Chulalongkorn University

20

Random Effect Models (1)

or REM for short. Also known as Error Component Models (ECM). X_1 is also constant.

$$Y_{it} = \beta_{1i} + \beta_2 X_{2it} + \dots + \beta_K X_{Kit} + \varepsilon_{it}$$

Similar to FEM except that

$$\beta_{1i} = \beta_1 + \xi_i$$

where ξ_i is cross-sectional variation

Random Effect Models (3)

Note that

$$V(\varepsilon'_{it}) = V(\varepsilon_{it}) + V(\xi_i) + 2C(\varepsilon_{it}, \xi_i)$$

$V(\varepsilon'_{it})$ is invariant.

=> OLS applies. Trivial.

Random Effect Models (2)

Case 1 $V(\xi_i)$ is section-invariant and $C(\varepsilon_{it}, \xi_i)$ is invariant

$$Y_{it} = \beta_1 + \beta_2 X_{2it} + \dots + \beta_K X_{Kit} + \varepsilon'_{it}$$

where $\varepsilon'_{it} = \varepsilon_{it} + \xi_i$

Random Effect Models (4)

Case 2 $V(\xi_i)$ is section-variant and/or $C(\varepsilon_{it}, \xi_i)$ is section-variant

Note that $V(\varepsilon'_{it})$ is section - variant.

Error term is heteroscedastic

=> FGLS applies. How?

$$V(\varepsilon') = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & \ddots & 0 & 0 & \ddots & 0 \\ 0 & 0 & \sigma_1^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & \ddots & 0 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & \ddots & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & \sigma_N^2 & 0 & 0 \\ 0 & \ddots & 0 & 0 & \ddots & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_N^2 \end{bmatrix}$$

$$V(\varepsilon') = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & \ddots & 0 & 0 & \ddots & 0 \\ 0 & 0 & \sigma_1^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & \ddots & 0 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & \ddots & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & \sigma_N^2 & 0 & 0 \\ 0 & \ddots & 0 & 0 & \ddots & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_N^2 \end{bmatrix}$$

Random Effect Models (5)

Case 3 $V(\xi_i)$ is section-variant

and/or $C(\varepsilon_{it}, \xi_i)$ is section-variant

but $\varepsilon_{it} = \varepsilon_i$

Note that $V(\varepsilon'_{it})$ is section - variant.

Error term is general

=> FGLS applies. How?

FEM vs REM (1)

They are substitute if no theoretical preference. Note that

- FEM is preferred when T is large and N is small.

2) T is small but N is large. Degree of freedom for FEM is small.

REM is more efficient.

FEM vs REM (2)

3) For T is small and N is large,
FEM is preferred if
cross-sectional variation (ξ_i) is
non-random. Otherwise, REM
is preferred.

Cross-sectional Heteroscedasticity (1)

$$Y_{it} = \beta_{1i} + \beta_{2i}X_{2it} + \dots + \beta_{Ki}X_{Kit} + \varepsilon_{it}$$

Assume time-invariant
variance-covariance for error terms.

In addition to possibility of different
cross-section weights (variances),
covariances between errors of cross
sections could be non-zero.

FEM vs REM (3)

4) ξ_i and X_{kit} are correlated. FEM
yields unbiased estimator but
REM yields biased estimator

$$V(\varepsilon) = \begin{array}{c} \begin{array}{ccc} \xleftarrow{\text{Cross-section\#1}} & & \xrightarrow{\text{Cross-section\#N}} \\ \hline \begin{matrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_1^2 \end{matrix} & \begin{matrix} \sigma_{12} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_{12} \end{matrix} & \begin{matrix} \sigma_{1N} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_{1N} \end{matrix} \\ \hline \begin{matrix} \sigma_{12} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_{12} \end{matrix} & \begin{matrix} \ddots & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \ddots \end{matrix} & \begin{matrix} \ddots & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \ddots \end{matrix} \\ \hline \begin{matrix} \sigma_{1N} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_{1N} \end{matrix} & \begin{matrix} \ddots & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \ddots \end{matrix} & \begin{matrix} \sigma_N^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_N^2 \end{matrix} \end{array} \end{array}$$

Cross-sectional Heteroscedasticity (2)

$$V(\boldsymbol{\varepsilon}_{it}) = \boldsymbol{\sigma}_j^2$$

$$\text{Cov}(\boldsymbol{\varepsilon}_{it}, \boldsymbol{\varepsilon}_{jt}) = \boldsymbol{\sigma}_{jj} \text{ for all } t$$

or they are time-invariant but section-variant

=> WLS will not apply as there are non-zero covariances between observations. Need GLS or FGLS.