Panel Data Regression Models

Covered Topics

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

What is Panel Data? (1)

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

What is Panel Data? (2)

<u>Node-to-Node Example</u>

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

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_{1T1}	$X1_{1T1}$	...	XK_{1T1}
:	:	:	:	....	:
N	1	Y_{N1}	$X1_{N1}$	...	XK_{N1}
:	:	:	:	...	:
N	T_N	Y_{NTN}	$X1_{NTN}$	....	XK_{NTN}

Balanced Panel Data

- Each cross-sections has equal number of time periods

  or $T_1=T_2=...=T_N$
- Simple Data Structure
- Less complicated computation

Linear Model for Panel Data (1)

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

- beta coefficients could be

$$\begin{aligned} &\text{time-invariant:} && \beta_{kit} = \beta_{ki} \text{ for } \forall t \\ &\text{section-invariant:} && \beta_{kit} = \beta_{kt} \text{ for } \forall i \\ &\text{both:} && \beta_{kit} = \beta_{k} \text{ for } \forall i, t \end{aligned}$$

Linear Model for Panel Data (2)

- variance of error terms could be

$$\begin{aligned} &\text{time-invariant:} && V(\varepsilon_{it}) = \sigma_i^2 \text{ for } \forall t \\ &\text{section-invariant:} && V(\varepsilon_{it}) = \sigma_t^2 \text{ for } \forall i \\ &\text{both:} && V(\varepsilon_{it}) = \sigma^2 \text{ for } \forall i, t \end{aligned}$$

Pooled Regression (1)

General Assumption:

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

where

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

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.

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}
:	:	:	:		:
N	1	Y_{N1}	$X1_{N1}$	...	XK_{N1}
N	2	Y_{N2}	$X1_{N2}$	....	XK_{N2}
:	:	:	:	...	:
N	T	Y_{NT}	$X1_{NT}$	....	XK_{NT}

Y **X**

Fixed Effect Models (1)

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

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

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 (3)

Equivalent LSDV

$$Y_{it} = \beta_{11} D_{1it} + \beta_{12} D_{2it} + \ldots + \beta_{1N} D_{Nit}$$
$$+ \beta_2 X_{2it} + \ldots + \beta_K X_{Kit} + \varepsilon_{it}$$

where

D_{jit} = 1 if i=j, j=1,...,N
= 0 otherwise

Note that

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

Fixed Effect Models (4)

i	t	Y_{it}	D_{1it}	....	D_{Nit}
1	1	Y_{11}	1	...	0
:	:	:	:	...	:
1	T_1	Y_{1T1}	1	...	0
:	:	:	:	....	:
N	1	Y_{N1}	0	...	1
:	:	:	:	...	:
N	T_N	Y_{NTN}	0	....	1

Fixed Effect Models (5)

Alternative form of FEM

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

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

$$Y_{it} = \beta_{11} D_{1it} + \beta_{11} D_{2it} + ... + \beta_{1T} D_{Tit}$$
$$+ \beta_2 X_{2it} + ... + \beta_K X_{Kit} + \varepsilon_{it}$$

where

D_{jit} = 1 if t=j, j=1,...,T
= 0 otherwise

Note that

D_{1it}+ D_{2it}+...+ D_{Tit}=1 for all i,t

Fixed Effect Models (7)

Common Problems

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

Fixed Effect Models (8)

Heteroscedasticity

-cross-section weight

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

-time weight

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

=>WLS applies

Fixed Effect Models (9)

Heteroscedasticity

-cross-section covarince

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

-auto-correlation

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

=>FGLS applies

Random Effect Models (1)

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

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

Similar to FEM except that

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

where ξ_i is cross-sectional variation

Random Effect Models (2)

<u>Case 1</u> $V(\xi_i)$ is section-invariant

and $C(\varepsilon_{it},\xi_i)$ is invariant

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

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

Random Effect Models (3)

Note that

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

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

=> OLS applies. Trivial.

Random Effect Models (4)

<u>Case 2</u> $\mathrm{V}(\xi_i)$ is section-variant

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

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

Error term is heteroscedastic

=> FGLS applies. How?

$$V(\varepsilon') = \left[\begin{array}{ccc|ccc|ccc}
\sigma_1^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \ddots & 0 & 0 & \ddots & 0 & 0 & \ddots & 0 \\
0 & 0 & \sigma_1^2 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & \ddots & 0 & 0 & 0 & 0 & 0 \\
0 & \ddots & 0 & 0 & \ddots & 0 & 0 & \ddots & 0 \\
0 & 0 & 0 & 0 & 0 & \ddots & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & \sigma_N^2 & 0 & 0 \\
0 & \ddots & 0 & 0 & \ddots & 0 & 0 & \ddots & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_N^2
\end{array}\right]$$

Random Effect Models (5)

<u>Case 3</u> V(ξ_i) is section-variant

and/or C(ε_{it},ξ_i) is section-variant

but $\varepsilon_{it} = \varepsilon_i$

Note that V(ε'_{it}) is section - variant.

Error term is general

=> FGLS applies. How?

$$
V(\varepsilon') = \left[\begin{array}{ccc|ccc|ccc}
\sigma_1^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \ddots & 0 & 0 & \ddots & 0 & 0 & \ddots & 0 \\
0 & 0 & \sigma_1^2 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & \ddots & 0 & 0 & 0 & 0 & 0 \\
0 & \ddots & 0 & 0 & \ddots & 0 & 0 & \ddots & 0 \\
0 & 0 & 0 & 0 & 0 & \ddots & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & \sigma_N^2 & 0 & 0 \\
0 & \ddots & 0 & 0 & \ddots & 0 & 0 & \ddots & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_N^2
\end{array}\right]
$$

FEM vs REM (1)

They are substitute if no theoretical preference. Note that

1) 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.

FEM vs REM (3)

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

Cross-sectional Heteroscedasticity (1)

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

Assume time-invariant variance-covarinace for error terms.

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

Cross-section#1 Cross-section#N

$$V(\varepsilon) = \left[\begin{array}{ccc|ccc|ccc}
\sigma_1^2 & 0 & 0 & \sigma_{12} & 0 & 0 & \sigma_{1N} & 0 & 0 \\
0 & \ddots & 0 & 0 & \ddots & 0 & 0 & \ddots & 0 \\
0 & 0 & \sigma_1^2 & 0 & 0 & \sigma_{12} & 0 & 0 & \sigma_{1N} \\
\hline
\sigma_{12} & 0 & 0 & \ddots & 0 & 0 & \ddots & 0 & 0 \\
0 & \ddots & 0 & 0 & \ddots & 0 & 0 & \ddots & 0 \\
0 & 0 & \sigma_{12} & 0 & 0 & \ddots & 0 & 0 & \ddots \\
\hline
\sigma_{1N} & 0 & 0 & \ddots & 0 & 0 & \sigma_N^2 & 0 & 0 \\
0 & \ddots & 0 & 0 & \ddots & 0 & 0 & \ddots & 0 \\
0 & 0 & \sigma_{1N} & 0 & 0 & \ddots & 0 & 0 & \sigma_N^2
\end{array}\right]$$

Cross-sectional Heteroscedasticity (2)

$V(\varepsilon_{it}) = \sigma_j^2$

$Cov(\varepsilon_{it}, \varepsilon_{jt}) = \sigma_{jj}$ for all t

or they are time-invariant but section-variant

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