Lecture 1

Panel Data Models (I)

– revised, 2010 –

Read:
- Wooldridge ch10*
- CT ch21,22
- Verbeek ch10
- Greene ch10

EC 671

Lee

Panel data Models : outline

(I) Basic Linear Models .. Lecture 1

(II) Further Issues (dynamic...).. Lecture 2

(III) Nonlinear Panel data Models ... Part II and beyond

Basic Models

$$y_{it} = \alpha_i + X_{it}'\beta + u_{it} \qquad \begin{cases} i=1,\dots N & (\text{Big } N) \\ t=1,\dots T & (\text{small } T) \end{cases}$$

$$E(u_{it} \mid X_{it}, \alpha_i) = 0$$

Features

1. Combination of cross-section & time series data

   eg) earnings (y_{it}) of 500 firms over 10 yrs.

   X_{it} includes # of employed, sales, industry, stock prices, location,

   eg) crime rates of 50 states, 1981 – 2007

   X_{it} includes population, # policemen, income, race, location ..

2. More observations! (NT obs for balanced panel)

3. X_{it} can include :

   i) time invariant variables (Z_i)

   ii) common factors invariant to individuals (W_t) (year dummies, ..)

eg) cigarette demand of consumers (y_{it})

z_i = race, gender, $\underbrace{\text{religion, education}}_{\text{little variation}}$

w_t = nationwide effects (TV advertisement..), time trends

4. Can control for <u>unobserved heterogeneity</u> (α_i)

; only with panel data (but not with indep pooled data)

⇒ <u>key point</u> : Can control omitted variables bias if the omitted variables are <u>time-invariant</u>. They are included in α_i.

α_i = { unobserved heterogeneity
individual effect, individual heterogeneity
unobserved component, ..

⇒ there are three methods to control for the effects of α_i.

{ FE (Fixed effects)
RE (Random effects)
FD (First differences)

eg) y_{it} = output of soybeans for farm i at t.

x_{it} = capital, labor, materials, rainfall, location ..

α_i = average quality of land, managerial ability
time constant factors ; <u>NOT</u> observed

More on α_i: (Direction of Bias)

α_i represents unobserved omitted variables.

Thus, pooling OLS (ignoring α_i) has the problem of omitted variables.

$\Rightarrow$ biased!

eg) $trade_{(ij),t} = \alpha_{ij} + X_{(ij),t}'\beta + d\,FTA_{(ij),t} + u_{ij,t}$

ij = bilateral (i = export country, j = import country)

$X_{ij,t}$ = sum of GDP, distance, … factor prices, …

α_{ij} = unobserved (not included in $X_{ij,t}$) factors that affect trade (say, barriers to trade)

$FTA_{ij,t} = \begin{cases} 1 & \text{if free trade agreements} \\ 0 & \text{o/w} \end{cases}$

$\Rightarrow$ direction of bias (of OLS estimates) = ?

$$\mathrm{Corr}(\underset{\substack{\uparrow \\ (\text{trade barriers})}}{\alpha_{ij}}, \underset{\substack{\uparrow \\ (\text{trade}_{ij,t})}}{u_{ij,t}}) = (-)$$

$$\mathrm{Corr}(FTA_{ij,t}, \underset{\substack{\uparrow \\ (\text{trade barriers})}}{\alpha_{ij}}) = (+)$$

Recall

$y = X_1\beta_1 + X_2\beta_2 + u$ is a true model

$y = X_1\beta_1 + \varepsilon$ is your model.

what is the bias of $\hat{\beta}_1$?

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'y = (X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2 + u)$$

$$= \beta_1 + \underbrace{(X_1'X_1)^{-1}X_1'X_2}\beta_2 + (X_1'X_1)^{-1}X_1'u$$

$$E(\hat{\beta}_1) = \beta_1 + \underbrace{\gamma \cdot \beta_2}_{\text{bias}} + 0$$

bias = (coef of X_1 in the reg. of X_2 on X_1)

• (coef of X_2 if X_2 is included)

Thus, bias of $\hat{d}$ in the previous example of FTA

$= (+)(-)$

that is, the (pooling) OLS estimates tend to be biased downward.

(under-estimated)

then, panel estimates (FE or RE) could be higher than OLS estimates.

; see Baier & Bergstrand (2007, JIE)

Distinguish

1. Panel data vs independent pooled data

- panel data : the <u>same</u> observations are used over time (d_i can be controlled)
- independent pooled data : <u>Different</u> individuals over time

eg1) crime rates of 50 states
eg2) political opinions of freshmen at UA, past 5 yrs
eg3) PSID (panel study of income dynamics)
eg4) IPO firms from 2001 to 2007

- Problem of panel data ⇒ have to trace individuals (died, moved, no response..)

  then we have <u>unbalanced</u> panel.

  point: As long as missing observations are random, usual methods using balanced panels are also valid. Otherwise, <u>selection bias</u> or <u>attrition issues</u> arise.

- Independent pooling data can also control for the effects of w_i and z_t (but not d_i)

  ; w_i, z_t, time dummies, industry dummies and so on.

2. Panel data vs system of equations

- System of equations = multivariate regressions

$\Rightarrow$ # of dependent variables is more than one. (G)

eg)
$$\begin{cases} y_{1t} = \alpha_1 + X_{1t}\beta_1 + e_{1t} & \text{GM} \\ y_{2t} = \alpha_2 + X_{2t}\beta_2 + e_{2t} & \text{Toyota} \\ y_{3t} = \alpha_3 + X_{3t}\beta_3 + e_{3t} & \text{Ford} \end{cases}$$

: $i = 1, 2, 3$ ($N = 3$ or $G = 3$)
$t = 1970, \dots 2007$ ($T = 28$)

This can be written as

$$y_{it} = \alpha_i + X_{it}\boxed{\beta_i} + e_{it} \quad \begin{cases} i = 1, \dots G \\ t = 1 \dots T \end{cases}$$

not β

differences

i) $N (= G)$ is small.

ii) Heterogeneous intercepts (α_i) and slopes (β_i)

iii) e_{it} are contemporaneously correlated

$$Var\begin{pmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \end{pmatrix} = \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ & \sigma_2^2 & \sigma_{23} \\ & & \sigma_3^2 \end{bmatrix}$$

$\Rightarrow$ GLS is used. otherwise, same as OLS.

$\Rightarrow$ N should be small. why?

think about:

i) $y_{it} = \alpha_i + X_{it}'\beta_{it} + u_{it}$ ⇒ stupid. why?

ii) $y_{it} = \alpha_i + X_{it}'\beta_i + u_{it}$ ⇒ System of equations

⇒ stupid if N is big.

"incidental parameter problem"

⇒ If N is small, there are big advantages

- Heterogeneous parameters even in slopes
- can utilize Σ to have more efficient estimators (GLS)

iii) $y_{it} = \alpha_i + X_{it}'\beta + u_{it}$ ⇒ Panel data

⇒ Even if N is big, we're ok. we need big N.

iv) $y_{it} = \alpha + X_{it}'\beta + u_{it}$

⇒ independent pooled data (no α_i's).

<u>Examples</u> SUR or Panel or independent pooled data?

i) Trade data of 100 countries over 5 years

ii) STD rates of 50 states

iii) CAPM models of 4 firms over 10 years, weekly data

iv) ACT scores of AL students, in 1990, 1995, 2000, 2005

v) your examples?

pool or not to pool?

1) $y_{it} = \alpha_i + X_{it}'\beta_i + u_{it}$ SUR

2) $y_{it} = \alpha_i + X_{it}'\beta + u_{it}$ panel

3) $y_{it} = \alpha + X_{it}'\beta + u_{it}$ pooled OLS

(1) vs (2) : Test $\beta_1 = \cdots = \beta_N$

(2) vs (3) : Test $\alpha_i = \alpha$

point: If $\alpha_i = \alpha$, using (3) is more efficient.
If $\alpha_i \neq \alpha$, (3) leads to biased estimators

More on α_i

Equation (2) is equivalently expressed as

$$y_{it} = \underbrace{c + \alpha_1 D_{1t} + \cdots + \alpha_{N-1} D_{N-1,t}}_{(= \alpha_i)} + X_{it}'\beta + u_{it}$$

Thus testing $\alpha_i = \alpha$ means

$$\alpha_1 = \cdots = \alpha_{N-1} = 0$$ (N-1 restrictions. why?)

or, equation (2) can be given as

$$y_{it} = \underbrace{\alpha_1 D_{1t} + \cdots + \alpha_N D_{Nt}}_{N \text{ dummies without a constant}} + X_{it}'\beta + u_{it}$$

Also, test on $\alpha_1 = \cdots = \alpha_{N-1} = 0$.

Panel Data Models (Linear Models only)

The key is how to control for α_i (unobserved heterogeneity)
(eliminate the effect of)

There are three methods

1. Fixed Effects Model
2. Random Effects "
3. First difference "

1. Fixed Effect Models (FE)

$$y_{it} = \alpha_0 + \boxed{\alpha_i} + \beta X_{it} + u_{it} \quad \cdots (1)$$

individual fixed effects (unobserved heterogeneity)

1. Omitting α_i leads to a pooled OLS.
2. when $\beta = \beta_i$ (different parameters over t), it's a system of equations (SUR). (when N is big, there may be too many parameters. Thus, we do not consider this case.)
3. One can add γ_t (time fixed effects)

Estimation

1. De-mean all variables

$$y_{it}^* = y_{it} - \bar{y}_{i\cdot} \quad \text{where } \bar{y}_{i\cdot} \text{ is the mean over time.}$$

$$\bar{y}_{i\cdot} = \alpha_0 + \alpha_i + \beta \bar{X}_{i\cdot} + \bar{u}_{i\cdot} \quad \cdots (2)$$

then (1) − (2) gives

$$(y_{it} - \bar{y}_{i\cdot}) = (\cancel{\alpha_0} - \cancel{\alpha_0}) - (\cancel{\alpha_i} - \cancel{\alpha_i}) + \beta(X_{it} - \bar{X}_{i\cdot}) + (u_{it} - \bar{u}_{i\cdot}) \quad \cdots (3)$$

the FE term α_i is cancelled out!

then do OLS in (3). this is the FE Model.!

② this time-demeaning process is equivalent to using (N−1) dummy variables. then it's called LS Dummy variables (LSDV) model

It's also called within estimator

$(y_{it}^* = y_{it} - \bar{y}_{i\cdot}, \quad X_{it}^* = X_{it} - \bar{X}_{i\cdot})$

Notes on FE

① Time invariant variables cannot be included.

$X_{it}^* = X_{it} - \bar{X}_{i\cdot} = 0$ eg) X_{it} = Gender $= \begin{cases} 1 & \text{male} \\ 0 & \text{female} \end{cases}$

$(X_{it} = \bar{X}_{i\cdot})$

; Gender, Race, location, area, ..

edu (a little variation) → std. error will be small ; t big!

② $\hat{\alpha}_i = \bar{Y}_{i\cdot} - \hat{\beta}\bar{X}_{i\cdot}$ FE can be estimated.

inconsistent unless T is big.

③* FE is valid even if X_{it} and α_i are correlated.

(RE is invalid in this case.)

④ We can test for FE

$H_0: \alpha_1 = \cdots = \alpha_{N-1} = 0$ (No FE effects)

$\Rightarrow$ F-test or LR-test (assuming homoskedasticity)

- If H_0 is rejected, then FE exists.
  $\Rightarrow$ OLS (pooling) is biased.
- If H_0 is not rejected, FE is inefficient relative to OLS (losing d.f.)

Example) Stata wagepan.do (wagepan.dta)
(wagepan.log)

Dep = log(wage)

	Pooled OLS	FE	RE
educ	.099	— (dropped)	0.101
black	−.144	— "	−0.144
hisp	.016	— "	0.020
exper	.089	0.117	0.112
expersq	−0.003	−0.004	−0.004
married	.108	.045	0.063
union	.181	.082	0.107
const	−0.035	1.06	−.107
R^2	0.187	0.064	0.181

- F-stat for $H_0: \alpha_1 = \cdots = \alpha_{N-1}$ = 7.98 (p-val = 0.000)
  $\sim F_{544, 3811}$
  $\Rightarrow$ We reject H_0. So, we favor FE over OLS

later
- LM test $H_0: \sigma_\alpha^2 = 0$ LM = 3216.00 (p-value = 0.000)
  $\Rightarrow$ We reject H_0. So, we favor RE over OLS
- Hausman test W = 31.45 (p-value = 0.000)
  $\Rightarrow$ We reject H_0 of RE model $\Rightarrow$ We favor ~~RE~~ FE.

Algebra of FE Models

(Notation) $i_T = (1, \dots, 1)'$ vector of ones $T \times 1$

① $i_T (i_T' i_T)^{-1} i_T' = \frac{1}{T} i_T i_T' \stackrel{\text{let}}{=} P_i \quad (T \times T)$

$$P_i x = \begin{pmatrix} \bar{x} \\ \bar{x} \\ \vdots \\ \bar{x} \end{pmatrix} \quad T \times 1 \qquad \text{since} \quad \frac{1}{T} i_T \underbrace{i_T' x}_{\Sigma x_i} = i_T \frac{i_T' x}{T} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \bar{x}$$

: P_i makes means

② $I_T - i_T (i_T' i_T)^{-1} i_T' \stackrel{\text{let}}{=} M_i \quad (T \times T)$

$$M_i x = (I - P_i) x = x - P_i x = \begin{pmatrix} x_1 - \bar{x} \\ \vdots \\ x_T - \bar{x} \end{pmatrix}$$

; M_i makes deviations

③ $\underset{(NT \times NT)}{P} = \underset{N \times N}{I_N} \otimes \underset{T \times T}{\frac{1}{T} i_T i_T'}$

Py makes <u>individual</u> means $= \begin{bmatrix} \begin{pmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_1 \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \bar{y}_N \\ \vdots \\ \bar{y}_N \end{pmatrix} \end{bmatrix}$ $(NT \times 1)$

N blocks, each $T \times 1$

④ $Q = I_{NT} - P = I_N \otimes (I_T - \frac{1}{T} i_T i_T')$

Qy makes <u>individual</u> deviations $= \begin{bmatrix} \begin{pmatrix} y_{11} - \bar{y}_1 \\ \vdots \\ y_{1T} - \bar{y}_1 \end{pmatrix} \\ \vdots \\ \begin{pmatrix} y_{N1} - \bar{y}_N \\ \vdots \\ y_{NT} - \bar{y}_N \end{pmatrix} \end{bmatrix}$

N blocks, each $T \times 1$

: time-demeaning matrix (using the mean over time for each i)

<u>Note</u> $Q^2 = Q, \quad P^2 = P$

$P + Q = I, \quad PQ = 0$

FE model

$$y = X\beta + D\alpha + u$$

where $\underset{NT\times N}{D} = I_N \otimes i_T$ = dummy variables

$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)$

<u>Note</u> use N-1 dummies and a constant or
use N dummies without a constant.

FE (Within) estimator

$$\hat{\beta}_{FE} = \left[X'[I - D(D'D)^{-1}D']X\right]^{-1} X'[I - D(D'D)^{-1}D']y$$

... like $(X'M_D X)^{-1} X'M_D y$

where $D = I_N \otimes i_T$, $D'D = I_N \otimes i_T' i_T = I_N \otimes T = T \cdot I_N$

$$= (X'QX)^{-1} X'Qy$$

this is the regression coeff of β from

$$(Qy) = (QX)\beta + u$$

... Each is deviation from individual means.

$$\sqrt{NT}\,(\hat{\beta}_{FE} - \beta) \to N\left[0,\ \sigma_u^2 \lim_{NT\to\infty}\left(\frac{1}{NT} X'QX\right)^{-1}\right]$$

ie $\mathrm{var}(\hat{\beta}_{FE}) = \hat{\sigma}_u^2 (X'QX)^{-1} = \hat{\sigma}_u^2 \sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it} - \bar{X}_i)'(X_{it} - \bar{X}_i)$

Also, $\hat{\alpha}_i = \bar{y}_i - \bar{X}_i \hat{\beta}_{FE}$ is consistent if T is big.
(can be biased if T is small)

$\hat{\alpha}_i$ = estimates of FE.
(can rank them!)

$$\hat{\sigma}_u^2 = \frac{1}{N(T-1) - K} \sum_i^N \sum_t^T \hat{u}_{it}^2$$

<u>Note</u> It is not NT−K.
The df = NT−N−K.

Robust variance for FE estimator

$$Var(\hat{\beta}_{FE}) = (X'QX)^{-1}\left(\sum_{i=1}^{N} X_i'Q_i\hat{u}_i\hat{u}_i'QX_i\right)(X'QX)^{-1}$$

$$\text{or} = (\ddot{X}'\ddot{X})^{-1}\left(\sum_{i=1}^{N} \ddot{X}_i'\hat{u}_i\hat{u}_i'\ddot{X}_i\right)$$

$T \geq 3$

see Arellano (1987, Oxford Bulletin)

... robust to any heteroskedasticity or serial correlation

(Note: We do not need to use HAC estimator.
Robust variance is valid for arbitrary serial correlation
(if N is big and T is small)

point The SUR estimator corrects for cross-correlation when N is small but T is big. On the contrary, robust variance correct for serial-correlation when T is small but N is big.
(see Wooldridge p.275. Also eq 7.26)

This can be used if serial correlation or heteroskedasticity is an issue.

Summary
- FE = WI = LSDV estimator
- α_i is inconsistent if T is small
- F-test for OLS vs FE
- Robust variance correct for arbitrary serial correlation.
- $Var(\hat{\beta}_{FE}) = \sigma_u^2 (X'QX)^{-1}$

2. Random Effect Models (RE)

Key Required assumption : X_{it} & α_i are not correlated.
(unobserved)

; Any indep. var. should not be correlated with unobserved heterogeneity (omitted).

This assumption can be tested (later).

Rewrite (1) as

$$y_{it} = \alpha_0 + \beta X_{it} + V_{it}$$

$$V_{it} = \alpha_i + u_{it}$$

That is, treat α_i as a random term included in the error term V_{it}.

$$\left[Var(V_{it}) = Var(\alpha_i + u_{it}) = \sigma_\alpha^2 + \sigma_u^2 \right]$$

$$\left[Corr(V_{it}, V_{is}) = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_u^2} \quad t \neq s \right]$$

Let $\hat{\lambda} = 1 - \sqrt{\dfrac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + T\hat{\sigma}_\alpha^2}}$

then

$$y_{it}^* = y_{it} - \hat{\lambda}\,\overline{y_{i\cdot}} \qquad \text{quasi-demeaning}$$

$$X_{it}^* = X_{it} - \hat{\lambda}\,\overline{X_{i\cdot}}$$

Then do OLS ⇒ this is GLS.
RE estimator
(details ; later)

Notes on RE

① More efficient than FE

As long as the key assumption is satisfied, RE is superior.

② Time invariant variables can be included.

③ When $\lambda = 0$ (ie. $\sigma_\alpha^2 = 0$), it's OLS.

when $\lambda = 1$ (ie $\sigma_\alpha^2 = \infty$, $T = \infty$), it's FE

In between, it's RE.

(Eg, wooldridge p471 $\hat{\lambda} = .643$)

④ Between estimator (BE) requires the same assumption, $Cov(X_{it}, \alpha_i) = 0$.
(If this condition is met, RE is better than BE)

⑤ Test OLS vs RE

LM test (Lagrange Multiplier Test)

Eg) Stata output (p.8)

LM test $H_0 : \sigma_\alpha^2 = 0$, $LM = 3216$
(OLS) (p-value = .0000)

⇒ we favor RE over OLS.
(Ha) (Ho)

Algebra of RE estimator

$$y_{it} = \beta X_{it} + V_{it} \quad \Rightarrow \quad y = X\beta + V$$

$$V_{it} = \alpha_i + u_{it} \qquad V = \alpha + u$$

$$Var(V) \equiv \Omega = \underbrace{\sigma_\alpha^2 (I_N \otimes i_T i_T')}_{T\sigma_\alpha^2 P \text{ since } P = I_N \otimes \frac{1}{T} i_T i_T'} + \sigma_u^2 I_{NT} = \begin{bmatrix} \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \vdots & & \vdots \\ \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \end{bmatrix} + \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \ddots \end{pmatrix}$$

$$= T\sigma_\alpha^2 P + \sigma_u^2 I_{NT}$$

$$\Omega^{-1} = \frac{1}{\sigma_u^2}(Q + \theta^2 P) \quad \text{where } \theta^2 = \frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}$$

Approximate transformation $\Omega^{-\frac{1}{2}} \simeq \frac{1}{\sigma_u}(Q + \theta P)$

since $(Q+\theta P)^2 = Q^2 + \theta^2 P^2 + 2\theta \underbrace{PQ}_{0}$
$= Q + \theta^2 P$

Thus GLS is OLS on

$y^* = X^*\beta + V^*$

$\Rightarrow \Omega^{-\frac{1}{2}} y = \Omega^{-\frac{1}{2}} X\beta + \Omega^{-\frac{1}{2}} V \Rightarrow (Q+\theta P) y = (Q+\theta P) X\beta + (Q+\theta P) V$

this amounts to $(1-\theta)$ differencing

$$[(Q+\theta P) y]_{it} = (y_{it} - \bar{y}_{i\cdot}) + \theta \bar{y}_{i\cdot}$$

$$= y_{it} - (1-\theta)\bar{y}_{i\cdot}$$

$$= y_{it} - \lambda \bar{y}_{i\cdot}$$

where $\lambda = 1 - \theta = 1 - \sqrt{\frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}}$

therefore

$$y_{it}^* = X_{it}^* \beta + V_{it}^*$$

where $y_{it}^* = y_{it} - \lambda \bar{y}_{i\cdot}$ see p. 13

$X_{it}^* = X_{it} - \lambda \bar{X}_{i\cdot}$

This expression is the same as the above

$$\hat{\beta}_{RE} = \left(\sum_{i=1}^{N} X_i' \hat{\Omega}^{-1} X_i\right)^{-1} \left(\sum_{i=1}^{N} X_i' \hat{\Omega}^{-1} y_i\right) \quad — (*)$$

where $\hat{\Omega}^{-1} = \frac{1}{\hat{\sigma}_u^2}(Q + \hat{\theta}^2 P)$

with $\hat{\theta}^2 = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + T\hat{\sigma}_\alpha^2}$

$Q = I - P$, $P = I_N \otimes \frac{1}{T} \iota_T \iota_T'$

How to estimate error variances $[\sigma_v^2 = \sigma_\alpha^2 + \sigma_u^2]$

$$\hat{\sigma}_v^2 = \frac{1}{NT-k} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{v}_{it}^2$$ where $\hat{v}_{it}$ is the pooled OLS residuals.

stata: uses FE residuals

$$\hat{\sigma}_\alpha^2 = \frac{1}{NT(T-1)/2 - k} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \hat{v}_{it} \hat{v}_{is}$$

$$\hat{\sigma}_u^2 = \hat{\sigma}_v^2 - \hat{\sigma}_\alpha^2$$ (this can be negative, when u_{it} is negatively correlated) → then use feasible GLS.

stata: $\max(0, \hat{\sigma}_u^2)$

- Feasible GLS

$$\hat{\Omega} = \frac{1}{N} \sum_{i=1}^{N} \hat{v}_i \hat{v}_i'$$ where $\hat{v}_{it}$ is from the pooled OLS.

and use $\hat{\beta}_{RE}$ (* above).

<u>Alternatively</u>,

- MLE (maximum likelihood estimator) is often used. <u>Note</u> this is an option in stata.

$$\mathcal{L}_i = -\frac{1}{2}\left[\frac{1}{\sigma_\alpha^2}\left\{\sum_{t=1}^{T_i}(y_{it} - x_{it}\beta)^2 - \frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}\left\{\sum_{t=1}^{T_i}(y_{it} - x_{it}\beta)\right\}^2\right\} + \ln\left(T_i \frac{\sigma_u^2}{\sigma_\alpha^2} + 1\right) + T_i \ln(2\pi\sigma_\alpha^2)\right]$$

Time Fixed Effects

we may add time dummies (in both FE & RE)

$$y_{it} = \alpha + X_{it}\beta + \underbrace{\gamma_t}_{T-1 \text{ dummies } (D_{1t}, \dots D_{T-1,t})} + V_{it}$$

- Omitting them can induce serial correlation in V_{it}.
- One may test for time FE (F-test)
- Time Fixed Effects measure changes over time. (in the mean)
- One often adds a time trend, individual-trend, if needed

$$y_{it} = \alpha + X_{it}\beta + \underbrace{c\cdot t}_{\substack{t=1,2,\dots T \\ \text{trend}}} + \gamma_t + d\underbrace{(\alpha_i \cdot t)}_{} + V_{it}$$

α_i dummies for each cross-section
$\alpha_i \cdot t$ interaction terms

FE or RE ?

i) Common suggestion : RE is appropriate if the same is drawn from a large population by random draw.

eg) Panel study on individuals (consumption)

$$C_{it} = \alpha + X_{it}\beta + V_{it}, \quad V_{it} = \alpha_i + u_{it}$$

eg) How about # of wins of SEC football teams each year?

ii) Hausman test

$$W = (\hat{\beta}_{FE} - \hat{\beta}_{RE})'(Var(\hat{\beta}_{FE}) - Var(\hat{\beta}_{RE}))^{-1}(\hat{\beta}_{FE} - \hat{\beta}_{RE}) \sim \chi^2_K$$

H_0: RE is fine ($Cov(X_{it}, \alpha_i) = 0$)
H_a: RE is not ok ($Cov(X_{it}, \alpha_i) \neq 0$)

<u>Note</u> $Var(\hat{\beta}_{RE}) = Cov(\hat{\beta}_{FE}, \hat{\beta}_{RE})$ under H_0.

Note on Hausman test

i) Strict exogeneity is assumed for the Hausman test.
RE & FE can differ (so that Ho is rejected), just because the assumption of strict exogeneity is not satisfied.

ii) Homoskedasticity is assumed in the Hausman test.
thus, rejection of the null can imply the presence of heteroskedasticity ⇒ Robust form can be used in this case.

iii) It is possible to focus on just <u>one</u> coefficient

$$(\hat{\beta}_{FEi} - \hat{\beta}_{REi}) / [Var(\hat{\beta}_{FEi}) - Var(\hat{\beta}_{REi})]^{\frac{1}{2}} \sim Z$$

or can use F-tests on a few coefficients (but not all coefficients)

⇒ The null may not be rejected if $Var(\hat{\beta}_{FE})$ or $Var(\hat{\beta}_{RE})$ or both are large.

iv) Consider an alternative Hausman test

$$y_{it}^* = x_{it}^{*\prime}\beta + \ddot{w}_{it}'\delta + u_{it}$$

where y_{it}^* is quasi-demeaned (RE)

x_{it}^* "

$\ddot{w}_{it}$ is FE demeaned (only time-varying variables)

Test Ho: $\delta = 0$. (or $\bar{w}_{i\cdot}$ can be used)

More questions on RE vs FE

what if a researcher is interested in examing the coefficients of time-invariant variables?

- FE cannot be used.
- RE cannot be used if $Cov(X_{it}, \alpha_i) \neq 0$

there are a few solutions, if X_{it} & α_i are correlated

i) Add group dummies and use RE. (Wooldridge p 288)

eg) add city dummies in a wage equation
The city dummies can be correlated with time-constant elements of X_{it}. Thus their effects are mitigated. then use RE.

ii) Mundlak's approach ; next page
$\Rightarrow$ Add $\bar{X}_{i\cdot}$ to the regression.

iii) Hausman - Taylor method : next lecture.

$\Rightarrow$ IV estimation using $\overline{W_{i\cdot}}$

(more on this, later)

FE or RE : Mundlak's view (1978)

From Hsiao book (2003)

Mundlak (1978) argues that the RE models are mis-specified, and suggests using

$$\alpha_i = \bar{X}_{i\cdot}' a + w_i \ , \qquad w_i \sim N(0, \sigma_w^2)$$

where $\bar{X}_{i\cdot}$ = time demeans; $K\times 1$ vector.

Recall:

(1) FE $\quad y_{it} = \alpha_i + X_{it}\beta + u_{it} \overset{\text{let}}{\Rightarrow} \hat{\beta}_{FE}$: Within

(2) RE $\quad y_{it} = \alpha + X_{it}\beta + \underbrace{(\alpha_i + u_{it})}_{V_{it}} \Rightarrow \hat{\beta}_{RE}$: GLS

Mundlak's equation

(3) Mundlak $\quad y_{it} = \alpha + X_{it}\beta + \overbrace{\boxed{\bar{X}_{i\cdot}' a} + (w_i}^{\alpha_i} + u_{it})$ (new V_{it})

$\overset{\text{let}}{\Rightarrow} \hat{\beta}^*_{RE}$ is obtained : GLS on this

(with $\bar{X}_{i\cdot}$ in the model)

also, $\hat{\alpha}^*_{RE}$ and $\hat{a}^*_{RE}$ are obtained.

Mundlak's results

(a) $\hat{\alpha}^*_{RE} = \bar{y} - \bar{x}'\hat{\beta}_{BE}$ where $\hat{\beta}_{BE}$ is obtained from the regression of $\bar{y}_{i\cdot}$ on $\bar{X}_{i\cdot}$ (between estimator)

* (b) $\hat{\beta}^*_{RE} = \hat{\beta}_{FE}$ (GLS = FE!)

(c) $\hat{a}^*_{RE} = \hat{\beta}_{BE} - \hat{\beta}_{FE}$

(d) $\hat{\beta}_{RE} = \delta\hat{\beta}_{BE} + (I-\delta)\hat{\beta}_{FE} \Rightarrow E(\hat{\beta}_{RE}) = \beta + \delta a$ if (3) is true.

$\neq \beta$ biased

using (2) $[E(\hat{\beta}_{BE}) = \beta + a$ & $E(\hat{\beta}_{FE}) = \beta]$

inconsistent unless T is big

(If T is big, $a \to 0$)

<u>Points</u> (Mundlak)

i) the distinction on α_i (FE or random) is not meaningful. what matters is whether X_{it} and α_i are correlated.

If correlated, call it FE : $\hat{\beta}^*_{RE} = \hat{\beta}_{FE}$

If uncorrelated, call it RE : $\hat{\beta}^*_{RE}$ is unbiased

ii) $\hat{\beta}_{RE}$ is biased due to omitting $\bar{X}_{i\cdot}$, if correlated.

iii) But, if $a=0$, $\hat{\beta}_{RE}$ is unbiased. Then

testing $a=0$ $\underset{\text{same}}{\Longleftrightarrow}$ $\hat{\beta}_{RE}$ is unbiased.

$H_0: a=0$ ($\hat{\beta}_{RE}$ is unbiased) $\quad H_a: a \neq 0$ ($\hat{\beta}_{RE}$ is biased) $\Rightarrow$ F-tests using SSRs.

If H_0 is not rejected, use equation (2) rather than (3).

<u>Note</u> $\hat{\beta}_{FE}$ is consistent under either H_0 or H_a.

<u>Note</u> this is essentially the same as the Hausman test comparing $\hat{\beta}^*_{RE}$ with $\hat{\beta}_{RE}$.

iv) However, Mundlak's results are <u>not</u> general;

$\hat{\beta}^*_{RE} \neq \hat{\beta}_{FE}$ in nonlinear models or dynamic models

More on assumptions

$$y_{it} = \alpha_i + X_{it}'\beta + u_{it}$$

strictly exogeneity assumption

$$E(y_{it} \mid x_{i1}, x_{i2}, \dots, x_{iT}, \alpha_i) = E(y_{it} \mid x_{it}, \alpha_i)$$

$$= X_{it}'\beta + \alpha_i \qquad \text{— (1)}$$

$$\Rightarrow E(u_{it} \mid x_{i1}, x_{i2}, \dots, x_{iT}, \alpha_i) = 0, \quad t = 1, \dots, T \qquad (2)$$

Note this assumption is different from

$$E(y_{it} \mid x_{i1}, \dots x_{iT}) = E(y_{it} \mid x_{it}) = X_{it}\beta$$

; (1) or (2) allows for the role of α_i since inputs (x_{it}) can depend on α_i

The strictly exogeneity assumption implies the followings.

① $E(X_{is}'u_{it}) = 0, \quad s, t = 1, \dots, T$

i.e. $E(X_{i5}'u_{i6}) = 0$ future errors

$E(X_{i5}'u_{i4}) = 0$ past errors (☆)

eg) $\log(wage_{it}) = \alpha_i + \gamma_t + z_{it}\gamma + \delta\,\underline{PROG_{it}} + u_{it}$

where $PROG_{it} = \begin{cases} 1 & \text{if participated in job training} \\ 0 & \text{o/w} \end{cases}$

All three methods will fail! $\Longrightarrow$: If there occur shocks to wages in the past people choose to participate in the program in the future! $E(PROG_{i,t+1}\, u_{it}) \neq 0$

② $E(X_{it}' u_{it}) = 0$, $t = 1, .. T$

.. usual assumption of no endogeneity
If not satisfied, use IV estimators.

In the previous example,

$E(PROG_{it} u_{it}) \neq 0$: selection bias
(The choice is not random)

③ However, we do NOT assume $E(X_{it}' \alpha_i) = 0$
thus, X_{it} & c_i can be correlated.

In the above example,

$E(PROG_{it} \cdot \alpha_i) \neq 0$: Program participation depends on your unobserved skills (α_i).

Even if $E(X_{it}' \alpha_i) \neq 0$, our FE or FD estimators are still valid. But NOT RE. why?

Another example (Wooldridge, p 255)

$$Patents_{it} = \alpha_i + \gamma_t + z_{it}\gamma + \delta_0 RD_{it} + \delta_1 RD_{it-1} + \cdots + \delta_5 RD_{it-5} + u_{it}$$

: R&D expenditures of 5 years affect patents.

i) R&D and α_i can be correlated. Fine.

ii) shocks to patents today (change in u_{it}) can affect future R&D spendings.

$E(RD_{it+1} u_{it}) \neq 0$

Thus the strictly exg. assumption fails.
⇒ All three methods fail.

Other Methods

(3) First Difference (FD) Method

$$
\begin{aligned}
y_{it} &= \alpha_0 + \alpha_i + \beta X_{it} + u_{it} \\
- \; y_{i,t-1} &= \alpha_0 + \alpha_i + \beta X_{i,t-1} + u_{i,t-1} \\
\hline
\Delta y_{it} &= (\cancel{\alpha_0 - \alpha_0}) + (\cancel{\alpha_i - \alpha_i}) + \beta \Delta X_{it} + \Delta u_{it}
\end{aligned}
$$

cancelled out

(α_i is controlled for!)

OLS of Δy_{it} on ΔX_{it} = FD method. $\Rightarrow \hat{\beta}_{FD}$

: $\hat{\beta}_{FD} = (\Delta X' \Delta X)^{-1} \Delta X' \Delta y$

Note

① FD works ok even if X_{it} & α_i are correlated.

② If $T=2$, FD = FE (Time invariant variables cannot be included.)

③ It is often troublesome to interpret the results since FD is based on the differenced data.

④ This idea is extended to quasi-differencing method. (e.g. twin data.

⑤ A strict exogeneity assumption is imposed.

$\Delta y = \Delta X \beta + \Delta u$, $T=2$

Assumption $E(\Delta X' \Delta u) = 0$

This implies $E(X_2' u_2) + E(X_1' u_1) - \underbrace{E(X_1' u_2) - E(X_2' u_1)}_{\text{They are uncorrelated over different time periods.}} = 0$

$\Rightarrow Cov(X_t, u_s) = 0$ for all t & s.

(strict exogeneity.)

Testing for strict exogeneity

When $T=2$,

$$\Delta y_i = \Delta x_i'\beta + x_{i2}\delta + \Delta u_i$$

Test $\delta = 0$ (since x_{i2} does not belong here if strict exogeneity holds)

When $T>2$,

$$\Delta y_{it} = \Delta x_{it}\beta + w_{it}\delta + \Delta u_{it}, \quad t=2,\dots T$$

Test $\delta = 0$. w_{it} is a subset of x_{it}

In the FE models,

$$y_{it} = x_{it}'\beta + w_{i,t+1}\delta + \alpha_i + u_{it}, \quad t=1,\dots,T-1$$

where $w_{i,t+1}$ is a subset of $x_{i,t+1}$

Test $\delta = 0$

Note use robust wald tests if heteroskedasticity or arbitrary serial correlation can exist.

Variance of $\hat{\beta}_{FD}$

$$Var(\hat{\beta}_{FD}) = \begin{cases} \hat{\sigma}^2(\Delta X'\Delta X)^{-1} & \text{where } \hat{\sigma}^2 = \frac{1}{NT-N-K}\sum^N\sum^T \hat{u}_{it}^2 \\ (\Delta X'\Delta X)^{-1}\left[\sum_i^N \Delta X_i \hat{u}_i \hat{u}_i' \Delta X_i\right](\Delta X'\Delta X)^{-1} & \\ \text{robust.} & \end{cases}$$

D-i-D in FD

Assume $T=2$, $\text{Prog}_{it} = \begin{cases} 1 & \text{treated} \\ 0 & \text{o/w} \end{cases}$

$$\Delta \text{PROG}_{it} = \text{PROG}_{i2}$$

$$\Delta y_{i2} = \theta_2 + \underbrace{\Delta X_{i2}\beta}_{\text{control variables}} + \underbrace{\delta \, \text{Prog}_{i2}}_{\hat{\delta} = \text{treatment effect}} + \Delta u_{i2} \quad , t=2$$

$\hat{\delta}$ = treatment effect $= \overline{\Delta Y}_{\text{treated}} - \overline{\Delta Y}_{\text{controlled}}$

$\Rightarrow$ If $T>2$ (or some participated in the prog at $t=1$)
use ΔPROG_{it} instead of PROG_{it}.

FD or FE?

i) FE is more efficient when u_{it} are serially uncorrelated.
FD " if u_{it} follow a random walk.
In fact, the truth is between these.

ii) FE is more popular since it's easy to interpret the results. But FD is used when FE is not feasible (eg. dynamic panel data models)

iii) Both are inconsistent if $\text{cov}(X_{it}, u_{it}) \neq 0$ (endogeneity)
$\Rightarrow$ IV version of panel data models.

* Extensions (Applying the idea of panel data models to interesting topics)

eg) "pairs" of sisters to study the effects of teen childbearing
Geronimus & Korenman (1992, QJE). FD is used.

eg) 'twin' study by Ashenfelter & Kruger (1994, 1998) AER
FD over twins.

(4) Between Estimator

Regression of <u>Means</u>

$$\begin{pmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_N \end{pmatrix} \text{ on } \begin{pmatrix} 1 & \bar{x}_1 \\ \vdots & \vdots \\ 1 & \bar{x}_N \end{pmatrix}$$

- Biased, inconsistent unless $(\alpha_i - \bar{\alpha})$ is uncorrelated with $\bar{X}_i$ (same requirement as in <u>RE</u>).
- Time invariant variables are excluded.
- Less efficient than RE.

<u>Exercises</u> H/W

1) Wooldridge, p. 292 Ex 10.3 a) show FE & FD are identical
   b) " $\hat{\beta}$'s are "

2) Wooldridge, p. 296 Ex 10.12 Empirical applications (wagepan.raw)
   <u>Note</u> (e) gen unionp1 = union[_n+1] if year < 1987
   ... Create the new variable "unionp1" using this code.

More on RE (from CT ch21)

α_i : a random intercept
or one-way individual-specific r.e. model

$$\mathrm{cov}[(\alpha_i + u_{it}), (\alpha_i + u_{is})] = \begin{cases} \sigma_\alpha^2 & t \neq s \\ \sigma_\alpha^2 + \sigma_u^2 & t = s \end{cases}$$

: equi-correlated

$$\mathrm{corr}(v_{it}, v_{is}) = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_u^2} \quad t \neq s$$

this does not depend on $(t-s)$.

Thus, the RE model is also called as the equicorrelated model or exchangeable errors model.

Panel-Robust Sandwich Std. Errors

Let
$$\tilde{y}_{it} = \begin{cases} y_{it} - \bar{y}_{i\cdot} & \text{FE} \\ y_{it} - \hat{\lambda}\bar{y}_{i\cdot} & \text{RE} \\ y_{it} - y_{i,t-1} & \text{FD} \end{cases}$$

Also, for $\tilde{w}_{it}$ using x_{it}

$\Rightarrow$ these can cause autocorrelations!

$$\tilde{y}_i = \tilde{W}_i \theta + \tilde{u}_i \quad \text{or} \quad \tilde{y} = \tilde{W}\theta + \tilde{u}$$

$$\hat{\theta} = (\tilde{W}'\tilde{W})^{-1}\tilde{W}'\tilde{y} = \left(\sum_{i=1}^N \tilde{W}_i'\tilde{W}_i\right)^{-1} \sum_{i=1}^N \tilde{W}_i'\tilde{y}_i$$

$$= \left(\sum_i^N \sum_t^T \tilde{w}_{it}'\tilde{w}_{it}\right)^{-1} \left(\sum_{i=1}^N \sum_{t=1}^T \tilde{w}_{it}'\tilde{y}_{it}\right)$$

$$\mathrm{Var}(\hat{\theta}) = \left(\sum_{i=1}^N \tilde{W}_i'\tilde{W}_i\right)^{-1} \sum_{i=1}^N \tilde{W}_i'\hat{u}_i\hat{u}_i'\tilde{W}_i \left[\sum_{i=1}^N \tilde{W}_i'\tilde{W}_i\right]^{-1}$$

$$= \left(\sum_{i=1}^N \sum_{t=1}^T \tilde{w}_{it}'\tilde{w}_{it}\right)^{-1} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T \tilde{w}_{it}\tilde{w}_{is}'\hat{u}_{it}\hat{u}_{is} \left(\sum_{i=1}^N \sum_{t=1}^T \tilde{w}_{it}'\tilde{w}_{it}\right)^{-1}$$

Note this permits arbitrary autocorrelations & heteroskedasticity.

Note If we choose i as a cluster variable, it's also a cluster-robust std. error.

$$Var(\hat{\theta}) = \left(\sum_{c=1}^{C} X_c' X_c\right)^{-1} \sum_{c=1}^{C} X_c' \hat{u}_c \hat{u}_c' X_c \left(\sum_{c=1}^{C} X_c' X_c\right)^{-1}$$

More on Hausman Test (from CT 21.4.3, p.717)

Alternative form :

$$(y_{it} - \hat{\lambda}\bar{y}_{i\cdot}) = (1-\hat{\lambda})\mu + (X_{it} - \hat{\lambda}\bar{X}_{i\cdot})'\beta + (X_{it} - \bar{X}_{i\cdot})\gamma + \text{error}$$

$H_0: \gamma = 0$ (RE is valid) Wald test using robust std. errors

point
- If RE is valid, $\gamma = 0$ is obtained.
- If RE is not valid, the error v_{it} will be correlated with X_{it} (α_i & X_{it} are correlated). then γ is significant.

Effects of heteroskedasticity & autocorrelation on the Hausman test

The Hausman test is valid only when α_i & ε_{it} are iid

Then one may obtain bootstrap variances of $Var(\hat{\beta}_{FE} - \hat{\beta}_{RE})$

(Wooldridge showed the case of heteroskedasticity.
Jing Li has a paper on this.

Random Coefficients Model

(Added to Panel models, Lecture 5)

$$y_{it} = X_{it}\boxed{\beta_i} + e_{it}$$

$\beta_i = \beta + u_i$; β varies but # of parameters remains the same (except for $Var(u_i) = \Gamma$; which needs to be estimated)

$\left(\begin{array}{c}\text{with } E(u_i)=0 \\ Var(u_i) = \Gamma\end{array}\right)$

the issue is how to estimate β.

Rewrite it as

$$\begin{aligned} y_{it} &= X_{it}(\beta + u_i) + e_{it} \\ &= X_{it}\beta + (X_{it}u_i + e_{it}) \\ &= X_{it}\beta + \varepsilon_{it} \end{aligned}$$

where $\varepsilon_{it} = X_{it}u_i + e_{it}$

Then do GLS

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$

where $\Omega = Var(e)$

Assume: no autocorrelation or cross-correlation

then Ω is block-diagonal with Ω_{ii}

$$\Omega_{ii} = E(e_i e_i') = \sigma_\varepsilon^2 I + X_i \Gamma X_i'$$

where $\Gamma = Var(u_i)$ can be estimated by following the procedure of Swamy (1971)

; see Greene, p. 224

Note i) If the constant term varies but the slope coefficients do not vary, it becomes the usual RE model

$$y_{it} = x_{it}\beta + (\beta_{0i} + e_{it}) \Rightarrow RE$$

$\uparrow$ constant

ii) MLE is possible

$$\ln L = \sum_{i=1}^{N}\left[-\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma_e^2 - \boxed{\frac{1}{2}\sum_{t=1}^{T}(y_{it} - x_{it}(\beta + u_i))^2/\sigma_i^2}\right]$$

$\Rightarrow$ But, Γ is not defined.

Let $u_i = \Lambda v_i$ where $\Lambda\Lambda' = \Gamma$ (like $\Lambda = \Gamma^{\frac{1}{2}}$ cholesky decomposition

$v_i \sim N(0,1)$

$$\Rightarrow \boxed{-\frac{1}{2}\sum_{t=1}^{T}(y_{it} - x_{it}(\beta + \underbrace{\Lambda v_i}_{u_i}))^2/\sigma_i^2}$$

$\Rightarrow$ How to take out v_i? (like joint density of e_i & v_i)

$$\ln L = \sum_{i=1}^{N}\int_{v_i}\left[\qquad\qquad\right] f(v_i)\, dv_i$$

Not easy : then rely on simulations

$\Rightarrow$ Max. simulated log likelihood estimator

$$\ln L = \sum_{i=1}^{N}\frac{1}{R}\sum_{r=1}^{R}\left[\qquad\qquad\right]$$

; Integration $\Rightarrow$ "summation" of mixture of distributions (R diff. dist.)

$\Rightarrow$ EM algorithm can be used.

iii) Mixed models

Multi-level mixed models

(xtmixed : stata)

g) Hierarchical Linear models

- $y_{it} = x_{it}\beta + c_i + e_{it}$ $\quad$ i = individual, t = groups.

  $c_i = z_i\alpha + u_i$

- $y_i = x_i\beta_i + e_i$

  $\beta_i = \beta + \Delta z_i + u_i$

- Many extensions if you believe you know the structure of the hierachical models.

  (perhaps yes, in management & marketing studies : structural equations models)

iv) State Space Models : not panel (time series)

$$y_t = A x_t + H \xi_t + w_t \quad : \text{observation equation}$$

$$\xi_t = F \xi_{t-1} + v_t \quad : \text{state equation}$$

Estimation by the Kalman filtering method

many important applications!

Review Questions for Panel Data Models

Part I. Basic Models

1. Discuss the difference between independent pooled data and panel data.

2. Discuss about advantages and disadvantages of panel data models.

3. Suppose that you have pooled data, and consider four different models, where e_{it} satisfies ideal conditions for all cases.

   (eq. 1) $y_{it} = \alpha + \beta X_{it} + e_{it}$

   (eq. 2) $y_{it} = \alpha_i + \beta_i X_{it} + e_{it}$

   (eq. 3) $y_{it} = \alpha_i + \beta_i X_{it} + e_{it}$, where $Cov(e_{it}, e_{jt}) = \sigma_{ij} \neq 0$ for $i \neq j$

   (eq. 4.1) $y_{it} = \alpha_i + \beta X_{it} + e_{it}$ Individual FE
   (eq. 4.2) $y_{it} = \alpha + \gamma_t + \beta X_{it} + e_{it}$ Time FE
   (eq. 4.3) $y_{it} = \alpha + \beta X_{it} + u_{it}$ where $u_{it} = \alpha_i + e_{it}$ Individual RE

(a) Discuss advantage(s) and disadvantage(s) of each model.

(b) When can we use each of them? When can't we use each of them?

(c) Consider equation (4), which allows for different intercepts for either different individuals (α_i) or different time periods (γ_t).

   (i) Which model and which terms measure *individual specific factors* of each individual? Are they assumed to be *time invariant*? Are these terms considered as measures for unobserved heterogeneity?
   (ii) Which model and which terms measure *time trend factors*? Are they assumed to affect all individuals commonly, that is, are they *individual invariant*? Are these terms considered as measures for unobserved factors that vary over time?

Part II. Panel Data Estimation

4. There are three main approaches in estimating panel data models. FD, FE, and RE. All of these methods control for or eliminate the effect of unobserved heterogeneity.

(a) Discuss how each of these eliminates the effect of unobserved heterogeneity.

(b) Which method allows for including time invariant regressors?

(c) Which methods require the assumption that regressors and the unobserved heterogeneity are independent?

(d) Which method allows us to estimate the magnitude of individual fixed effect terms and thus the differences between individuals?

(e) Which method gives the most efficient estimator? Under what condition?

(f) One can consider a between estimator (BE). Under what condition, is it valid? If the condition were met, what other method would you recommend to use? Why?

(g) How can you decide to choose between (i) FE and pooled OLS (ii) RE and pooled OLS and (iii) FE and RE?

37

```
-------------------------------------------------------------------------------------------
----
       log:  C:\Documents and Settings\jlee\My Documents\EC671\wagepan.log
  log type:  text
 opened on:  13 Sep 2004, 21:07:33

.
. clear

. set memory 40m
(40960k)

. set more off

. set matsize 350

.
. use wagepan, clear

. *tsset nr year
.
. iis nr

. tis year

.
. xtreg  lwage  educ black hisp exper expersq married union, fe

Fixed-effects (within) regression               Number of obs      =      4360
Group variable (i): nr                          Number of groups   =       545

R-sq:  within  = 0.1780                         Obs per group: min =         8
       between = 0.0005                                        avg =       8.0
       overall = 0.0638                                        max =         8

                                                F(4,3811)          =    206.38
corr(u_i, Xb)  = -0.1139                        Prob > F           =    0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |  (dropped)
       black |  (dropped)
        hisp |  (dropped)
       exper |   .1168467   .0084197    13.88   0.000     .1003392    .1333542
     expersq |  -.0043009   .0006053    -7.11   0.000    -.0054876   -.0031142
```



```
     married |   .0453033   .0183097     2.47   0.013     .0094056     .081201
       union |   .0820871   .0192907     4.26   0.000      .044266    .1199083
       _cons |    1.06488   .0266607    39.94   0.000     1.012609     1.11715
-------------+----------------------------------------------------------------
     sigma_u |   .4000539
     sigma_e |  .35125535
         rho |   .5646785   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(544, 3811) =     7.98         Prob > F = 0.0000

. regress  lwage  educ black hisp exper expersq married union

      Source |       SS       df       MS              Number of obs =    4360
-------------+------------------------------           F(  7,  4352) =  142.61
       Model |  230.719766     7  32.9599665           Prob > F      =  0.0000
    Residual |  1005.80988  4352  .231114402           R-squared     =  0.1866
-------------+------------------------------           Adj R-squared =  0.1853
       Total |  1236.52964  4359  .283672779           Root MSE      =  .48074

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0993878   .0046776    21.25   0.000     .0902173    .1085583
       black |  -.1438417   .0235595    -6.11   0.000    -.1900303   -.0976531
        hisp |    .015698   .0208112     0.75   0.451    -.0251026    .0564985
       exper |   .0891791    .010111     8.82   0.000     .0693563    .1090019
     expersq |  -.0028487   .0007074    -4.03   0.000    -.0042354   -.0014619
     married |   .1076656   .0156965     6.86   0.000     .0768925    .1384387
       union |   .1800726   .0171205    10.52   0.000     .1465076    .2136375
       _cons |  -.0347057    .064569    -0.54   0.591    -.1612938    .0918824
------------------------------------------------------------------------------

. xtreg  lwage  educ black hisp exper expersq married union, re

Random-effects GLS regression                   Number of obs      =      4360
Group variable (i): nr                          Number of groups   =       545

R-sq:  within  = 0.1774                         Obs per group: min =         8
       between = 0.1837                                        avg =       8.0
       overall = 0.1808                                        max =         8

Random effects u_i ~ Gaussian                   Wald chi2(7)       =    943.95
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
```

38

```
-------------+----------------------------------------------------------------
        educ |   .1012246   .0089133    11.36   0.000     .0837549    .1186943
       black |  -.1441307   .0476148    -3.03   0.002     -.237454   -.0508073
        hisp |   .0201511   .0426011     0.47   0.636    -.0633456    .1036477
       exper |   .1121195   .0082609    13.57   0.000     .0959285    .1283105
     expersq |  -.0040689   .0005918    -6.88   0.000    -.0052288   -.0029089
     married |   .0627951   .0167729     3.74   0.000     .0299209    .0956693
       union |   .1073789     .01783     6.02   0.000     .0724327     .142325
       _cons |  -.1074643   .1107057    -0.97   0.332    -.3244435    .1095149
-------------+----------------------------------------------------------------
     sigma_u |  .32456727
     sigma_e |  .35125535
         rho |  .46057172   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. xttest0

Breusch and Pagan Lagrangian multiplier test for random effects:

        lwage[nr,t] = Xb + u[nr] + e[nr,t]

        Estimated results:
                         |       Var     sd = sqrt(Var)
                ---------+-----------------------------
                   lwage |   .2836728       .5326094
                       e |   .1233803       .3512553
                       u |   .1053439       .3245673

        Test:   Var(u) = 0
                                  chi2(1) =  3216.73
                              Prob > chi2 =     0.0000

. xthausman
(Warning:  xthausman is no longer a supported command; use -hausman-.  For instructions, see help hausman.)

Hausman specification test

                 ---- Coefficients ----
             |      Fixed       Random
       lwage |    Effects      Effects       Difference
-------------+-----------------------------------------
       exper |   .1168467     .1121195         .0047272
     expersq |  -.0043009    -.0040689         -.000232
     married |   .0453033     .0627951        -.0174918
       union |   .0820871     .1073789        -.0252917
```



```
    Test:  Ho:  difference in coefficients not systematic

                 chi2(  4) = (b-B)'[S^(-1)](b-B), S = (S_fe - S_re)
                           =     31.45
                 Prob>chi2 =      0.0000

. xtreg  lwage  educ black hisp exper expersq married union, fe

Fixed-effects (within) regression               Number of obs      =      4360
Group variable (i): nr                          Number of groups   =       545

R-sq:  within  = 0.1780                         Obs per group: min =         8
       between = 0.0005                                        avg =       8.0
       overall = 0.0638                                        max =         8

                                                F(4,3811)          =    206.38
corr(u_i, Xb)  = -0.1139                        Prob > F           =    0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |  (dropped)
       black |  (dropped)
        hisp |  (dropped)
       exper |   .1168467   .0084197    13.88   0.000     .1003392    .1333542
     expersq |  -.0043009   .0006053    -7.11   0.000    -.0054876   -.0031142
     married |   .0453033   .0183097     2.47   0.013     .0094056     .081201
       union |   .0820871   .0192907     4.26   0.000      .044266    .1199083
       _cons |    1.06488   .0266607    39.94   0.000     1.012609     1.11715
-------------+----------------------------------------------------------------
     sigma_u |   .4000539
     sigma_e |  .35125535
         rho |   .5646785   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(544, 3811) =     7.98           Prob > F = 0.0000

. est store fixed

. xtreg  lwage  educ black hisp exper expersq married union, re

Random-effects GLS regression                   Number of obs      =      4360
Group variable (i): nr                          Number of groups   =       545

R-sq:  within  = 0.1774                         Obs per group: min =         8
```

39

```
      between = 0.1837                                        avg =        8.0
      overall = 0.1808                                        max =          8

Random effects u_i ~ Gaussian                   Wald chi2(7)       =    943.95
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .1012246   .0089133    11.36   0.000     .0837549    .1186943
       black |  -.1441307   .0476148    -3.03   0.002     -.237454   -.0508073
        hisp |   .0201511   .0426011     0.47   0.636    -.0633456    .1036477
       exper |   .1121195   .0082609    13.57   0.000     .0959285    .1283105
     expersq |  -.0040689   .0005918    -6.88   0.000    -.0052288   -.0029089
     married |   .0627951   .0167729     3.74   0.000     .0299209    .0956693
       union |   .1073789     .01783     6.02   0.000     .0724327     .142325
       _cons |  -.1074643   .1107057    -0.97   0.332    -.3244435    .1095149
-------------+----------------------------------------------------------------
     sigma_u |  .32456727
     sigma_e |  .35125535
         rho |  .46057172   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. hausman fixed .

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     fixed          .          Difference          S.E.
-------------+----------------------------------------------------------------
       exper |    .1168467     .1121195        .0047272        .0016276
     expersq |   -.0043009    -.0040689        -.000232        .0001269
     married |    .0453033     .0627951       -.0174918        .0073427
       union |    .0820871     .1073789       -.0252917        .0073636
------------------------------------------------------------------------------
                            b = consistent under Ho and Ha; obtained from xtreg
             B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(4) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =       31.45
                Prob>chi2 =      0.0000

.
.
.
```



```
end of do-file

. exit, clear
```

40

5.4. Use the data in CARD.RAW for this problem.

a. Estimate a log(*wage*) equation by OLS with *educ*, *exper*, *exper*^{2}, *black*, *south*, *smsa*, *reg661* through *reg668*, and *smsa66* as explanatory variables. Compare your results with Table 2, Column (2) in Card (1995).

b. Estimate a reduced form equation for *educ* containing all explanatory variables from part a and the dummy variable *nearc4*. Do *educ* and *nearc4* have a practically and statistically significant partial correlation? [See also Table 3, Column (1) in Card (1995).]

c. Estimate the log(*wage*) equation by IV, using *nearc4* as an instrument for *educ*. Compare the 95 percent confidence interval for the return to education with that obtained from part a. [See also Table 3, Column (5) in Card (1995).]

d. Now use *nearc2* along with *nearc4* as instruments for *educ*. First estimate the reduced form for *educ*, and comment on whether *nearc2* or *nearc4* is more strongly related to *educ*. How do the 2SLS estimates compare with the earlier estimates?

e. For a subset of the men in the sample, IQ score is available. Regress *iq* on *nearc4*. Is IQ score uncorrelated with *nearc4*?

f. Now regress *iq* on *nearc4* along with *smsa66*, *reg661*, *reg662*, and *reg669*. Are *iq* and *nearc4* partially correlated? What do you conclude about the importance of controlling for the 1966 location and regional dummies in the log(*wage*) equation when using *nearc4* as an IV for *educ*?

6.1. a. In Problem 5.4d, test the null hypothesis that *educ* is exogenous.

b. Test the the single overidentifying restriction in this example.

- Wooldridge ch6, Ex 6.1 (p.135)
  "CARD.RAW" by Card (1995) Endogeneity test

Note use two different tests for endogeneity

① Hausman test

② LM test

Note Calculating LM stat

"display 3010 * 0.0004"

ans = 1.204

"display chiprob(1, 1.204)

ans = 0.2733 ... This is the p-value for the overidentifying restriction test (df = 1) why?

Ex 10.3 Show FD = FE when T=2. (H/W)

Let $\bar{x}_i = (x_{i1} + x_{i2})/2$, $\bar{y}_i = (y_{i1} + y_{i2})/2$

$x_{i1}^* = x_{i1} - \bar{x}_i$, $x_{i2}^* = x_{i2} - \bar{x}_i$, $y_{it}^* = y_{it} - \bar{y}_i$

$$\hat{\beta}_{FE} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)^2\right]^{-1} \left[\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)\right]$$

$$= \left[\sum_{i=1}^{N} (x_{i1}^{*\prime} x_{i1}^* + x_{i2}^{*\prime} x_{i2}^*)\right]^{-1} \left(\sum_{i=1}^{N} (x_{i1}^{*\prime} y_{it}^* + x_{i2}^{*\prime} y_{it}^*)\right]$$

where $x_{i1}^* = x_{i1} - \bar{x}_i = x_{i1} - \frac{1}{2}(x_{i1} + x_{i2})$

$= \frac{1}{2}(x_{i1} - x_{i2}) = -\frac{1}{2}\Delta x_i$

$x_{2i}^* = x_{2i} - \bar{x}_i = x_{2i} - \frac{1}{2}(x_{i1} + x_{i2})$

$= \frac{1}{2}(x_{i2} - x_{i1}) = \frac{1}{2}\Delta x_i$

$y_{i1}^* = y_{i1} - \frac{1}{2}(y_{i1} + y_{i2}) = \frac{1}{2}(y_{i1} - y_{i2}) = -\frac{1}{2}\Delta y_i$

$y_{i2}^* = y_{i2} - \frac{1}{2}(y_{i1} + y_{i2}) = \frac{1}{2}(y_{i2} - y_{i1}) = \frac{1}{2}\Delta y_i$

thus

$$\hat{\beta}_{FE} = \left(\sum_{i=1}^{N} \Delta x_i' \Delta x_i / 2\right)^{-1} \left(\sum_{i=1}^{N} \Delta x_i \Delta y_i / 2\right)$$

$$= \left(\sum \Delta x_i' \Delta x_i\right)^{-1} \left(\sum \Delta x_i \Delta y_i\right) = \hat{\beta}_{FD}$$

$$Var(\hat{\beta}_{FE}) = \hat{\sigma}_u^2 \left[\sum_i^N (x_{i1}^{*\prime} x_{i1}^* + x_{i2}^{*\prime} x_{i2})\right]^{-1}$$

$$= \frac{\hat{\sigma}_u^2}{2} \left[\sum \Delta x_i' \Delta x_i / 2\right]^{-1} = \hat{\sigma}_u^2 \left(\sum \Delta x_i' \Delta x_i\right)^2$$

43

Homework
Panel Data Models (I)

There is a Stata dataset on my website called panel_hw.dta. This is a dataset examining voter turnout in 49 US states (Louisiana is omitted because of an unusual election in 1982) plus the District of Columbia over 11 elections. In other words, this dataset is time series cross section (TSCS) data on 50 units over 11 time periods. The variables in this dataset are:

year. The year of the election.

stcode: The ICPSR state code number.

state: The state name.

vaprate: The turnout rate as a percentage of the voting age population.

midterm: A dummy variable for midterm election years.

regdead: The number of days before the general election by which an individual needs to register.

gsp: State per capita income in 1000s of dollars.

There are also 9 regional dummy variables.

Download this data set from the web site. Turn in your write-ups for these problems as well as the log files.

1. Regress turnout as a percent of voting age population on the number of days before the general election by which an individual needs to register, state per capita income, the dummy variable for midterm elections, and the dummy variables for West North Central, the South, and the Border states.

   ```
   regress vaprate gsp midterm regdead WNCentral South Border
   ```

   Which coefficients are significant? Are there any regional effects of these regions? Use F-test to determine this.

2. Obviously, we assumed that <u>pooling</u> our TSCS data was valid in Q. 1. Now let us test this assumption by estimating a fixed-effects model. We must first declare our data to be panel data. Type

   ```
   iis stcode
   ```

   to let Stata know that our units are indicated by *stcode*, and type

44

```
tis year
```

to let Stata know our time periods are indicated by *year*. Then run a fixed-effects model by typing

```
xtreg vaprate midterm gsp regdead WNCentral South Border, fe
est store fixed
```

The bottom line of the estimation results give us an F test for pooling—Can we conclude the 50 unit—specific dummy variables are all equal to zero? Is pooling appropriate in light of the results of this test? Why do some variables drop out of our estimation?

3. Concerned with what happened to some variables in question 2, you decide to try a random-effects model. Type

```
xtreg vaprate midterm gsp regdead WNCentral South Border, re
```

to get a random effects model. Why do these results differ from those in Question 2?

Then, type

```
xttest0
```

What does this specification test tell you about the appropriateness of the random-effects model?

Finally, type

```
Xthausman          (old)
hausman fixed .    (new)
```

What does this test tell you? Discuss the tradeoffs between using pooled OLS, fixed-effects, and random-effects for this model.

Summarize your estimation results in Q.1 – 3 with a few paragraphs.

45

Homework
Panel Data Models (II)

The data file, crime4.dta, contains the follow data set, for the study on the county crime rates in North Carolina. Cornwell and Trumbell (1994) used data on 90 counties in North Carolina for the years 1981 through 1987, to estimate an unobserved effects model of crime.

Obs: 630

```
 1. county           county identifier
 2. year             81 to 87
 3. crmrte           crimes committed per person
 4. prbarr           'probability' of arrest
 5. prbconv          'probability' of conviction
 6. prbpris          'probability' of prison sentenc
 7. avgsen           avg. sentence, days
 8. polpc            police per capita
 9. density          people per sq. mile
10. taxpc            tax revenue per capita
11. west             =1 if in western N.C.
12. central          =1 if in central N.C.
13. urban            =1 if in SMSA
14. pctmin80         perc. minority, 1980
15. wcon             weekly wage, construction
16. wtuc             wkly wge, trns, util, commun
17. wtrd             wkly wge, whlesle, retail trade
18. wfir             wkly wge, fin, ins, real est
19. wser             wkly wge, service industry
20. wmfg             wkly wge, manufacturing
21. wfed             wkly wge, fed employees
22. wsta             wkly wge, state employees
23. wloc             wkly wge, local gov emps
24. mix              offense mix: face-to-face/other
25. pctymle          percent young male
26. d82              =1 if year == 82
27. d83              =1 if year == 83
28. d84              =1 if year == 84
29. d85              =1 if year == 85
30. d86              =1 if year == 86
31. d87              =1 if year == 87
32. lcrmrte          log(crmrte)
33. lprbarr          log(prbarr)
34. lprbconv         log(prbconv)
35. lprbpris         log(prbpris)
36. lavgsen          log(avgsen)
37. lpolpc           log(polpc)
38. ldensity         log(density)
39. ltaxpc           log(taxpc)
40. lwcon            log(wcon)
41. lwtuc            log(wtuc)
42. lwtrd            log(wtrd)
43. lwfir            log(wfir)
```

44. lwser	log(wser)
45. lwmfg	log(wmfg)
46. lwfed	log(wfed)
47. lwsta	log(wsta)
48. lwloc	log(wloc)
49. lmix	log(mix)
50. lpctymle	log(pctymle)
51. lpctmin	log(pctmin)
52. clcrmrte	lcrmrte - lcrmrte[t-1]
53. clprbarr	lprbarr - lprbarr[t-1]
54. clprbcon	lprbconv - lprbconv[t-1]
55. clprbpri	lprbpri - lprbpri[t-1]
56. clavgsen	lavgsen - lavgsen[t-1]
57. clpolpc	lpolpc - lpolpc[t-1]
58. cltaxpc	ltaxpc - ltaxpc[t-1]
59. clmix	lmix - lmix[t-1]

We want to take logs for most variables when it is desired to do so. Consider a regression of lcrmrte on d82-d87, lprbarr, lprbconv, lprbpris, lavgsen, lpolpc, ltaxpc, pctymle, west, central, urban, pctmin80 and lmix.

(a) Use a pooling regression to estimate the above model.

(b) Use a FE model to estimate the above model. Is there evidence of unobserved heterogeneity? Show your testing hypothesis, and decide on it. Which variables are omitted from the estimation? Why?

(c) In the FE estimation, carefully interpret each of the coefficients of d82, lprbarr and lpolpc.

(d) In the FE estimation, is there evidence of the time fixed effects? What is your testing hypothesis? Test the hypothesis. [You may need to calculate the F-statistic by yourself.]

(e) Use a RE model to estimate the above model. Is there evidence of unobserved heterogeneity? (hint: xttest0) Show your testing hypothesis, and decide on it. Are some variables omitted from the estimation? Why or why not?

(f) Which model is more appropriate, FE or RE? What is your underlying testing hypothesis? What implication does the null hypothesis have? Explain intuitively why the null hypothesis is rejected; provide an example story to explain this.

(g) Using the RE model estimation, determine if there is evidence of the regional effect of three variables, west, central, and urban. [You may need to calculate the F-statistic by yourself.]

Questions on Lecture 5, Panel Data Models (I)

1. On *unobserved heterogeneity.*
   a. Define it.
   b. Can we control for *unobserved heterogeneity* in the independent pooling OLS regression?
   c. Discuss how each of three panel data estimation methods can control for the effects of *unobserved heterogeneity.*

2. On *time invariant variables*
   a. Define *time invariant variables* and illustrate an example of your choice of your example models.
   b. What is the difference between *time invariant variables* and *unobserved heterogeneity*?
   c. Define *common factors invariant to individuals (cross-section units)* and illustrate an example of your choice of your example models.

3. What is the unbalanced panel? When do we need to care for this?

4. Can we control for *time fixed effects* in the independent pooling models using the OLS estimation? Can we control for the effects of *time invariant variables* in the independent pooling models using the OLS estimation?

5. Discuss the differences between the system of equations and the panel data models.

6. On whether a regressor is correlated with *unobserved heterogeneity*

   a. *Why* is the RE estimator inconsistent if any regressor is correlated with *unobserved heterogeneity*? Explain your reasoning.
   b. *Why* is the FE estimator consistent if any regressor is correlated with *unobserved heterogeneity*? Explain your reasoning.
   c. *Why* is the FD estimator consistent if any regressor is correlated with *unobserved heterogeneity*? Explain your reasoning.

7. On Choosing the best fitting model

   a. Discuss briefly how one can choose the best fitting model among OLS, FE and RE models. (half page)
   b. Why can't we test for the RE over the FD models using usual tests? Why can't we test for the FE over the FD models?
   c. Which estimator has a higher variance, FE or RE? **WHY?**
   d. Discuss how one can choose between FE and RE, without having to use the Hausman test. When is RE preferred? When is FE preferred?
   e. What are the limitations of the Hausman test in choosing the FE or RE estimators?

8. Can we include time dummies in the RE estimation? How about FE and FD?

9. Consider the following data.

id	t	Y_{it}	X_{it}	i	D_1	D_2	D_3	DT_1	DT_2	Y_{it}^*	X_{it}^*	Y_{it}^+	X_{it}^+	P_iX	P_iX	PX	QX
1	1	3	1	1	1			1									
1	2	5	4	1	1			0									
1	3	3	3	1	1			0									
1	4	6	5	1	1			0									
2	1	7	6		0			1									
2	2	8	4		0			0									
2	3	5	3		0			0									
2	4	3	7		0			0									
3	1	7	8		0			1									
3	2	9	3		0			0									
3	3	3	5		0			0									
3	4	5	8		0			0									

(a) Fill out the columns of D_2, D_3, and DT_2.

(b) Consider a within transformation of Y_{it} for the usual FE model. That is, find $Yit^* = Y_{it} - \bar{Y}_{i,\bullet}$ and fill out the column of Y_{it}^* in the table. Also find Yit^* and fill out the column of X_{it}^* in the table.

(c) Consider a FE model that controls for the time FE in addition to (b) and find the transformed variable as $Y_{it}^+ = Y_{it} - \bar{Y}_{i,\bullet} - \bar{Y}_{\bullet,t}$. Then fill out the columns of Y_{it}^+. Do the same for X_{it}^+.

(d) Let $i = (1, 1, 1, 1)'$ be a vector of ones and $P_i = i(i'i)^{-1}i'$. Then find P_iX_1, where X_1 denotes $\{X_{it}, t = 1, 2, 3, 4\}$ of the first cross section unit. That is, $X_1 = (1, 4, 3, 5)'$. Specifically, what is the dimension of P_iX_1? Also, find P_iX_2, P_iX_3, and P_iX_4 and fill out the last column of the table.

(e) Let $P = I_N \otimes P_i$. Find PX, where $X = \{X_{it}\}$, i=1,..4, and t=1,,.3; thus, X denotes the third column of the table. Fill out the table.

(f) Let $Q = I_{NT} - P$. Find QX, where $X = \{X_{it}\}$, i=1,..4, and t=1,,.3. Fill out the table.

(g) Let $Y^* = QY$, and $X^* = QX$. Find the OLS estimator of Y^* on $X^*$and show that it is the FE estimator.

(h) Find the variance of the FE estimator in (g).

10. What is the strictly exogeneity assumption in panel data models?

11. What is the feedback effect? Define it.

12. Sketch the algebra of the RE estimator. Be sure to explain how the GLS procedure is adopted, and show some details of the GLS procedures. Explain how a quasi-demeaning procedure is adopted in the RE estimator. (half page)

13. Can STATA use an MLE to estimate the RE estimator? What is the command for this?

14. Refer to the textbooks of your choice and write down the assumptions on each components of the error term in the RE model.

15. On the Mundlak's method

    a. Discuss the suggested method of Mundlak when the usual assumption required for the RE.
    b. How can you test for the validity of the usual RE using the suggested method of Mundlak? What is the null hypothesis?
    c. What is your suggested solution if the null hypothesis is rejected in (b)? What is the implication on the usual RE?
    d. What is your suggested solution if the null hypothesis is not rejected in (b)? What is the implication on the usual RE?
    e. Is the FE estimator still useful if the null is rejected or not?

16. Explain how you can test for strict exogeneity in the FD estimation.

17. On the D-i-D method

    a. What is the D-i-D estimator?
    b. We usually include a dummy variable in the treatment effect models. One wishes to use the FD estimator using it. How can we capture the treatment effect from the FD model?
    c. If $T = 2$, should we take the first difference of the dummy variable in the FD model? Or, is it the case that it does not matter even when we do not take the FD of the dummy variable?
    d. If $T > 2$, should we take the first difference of the dummy variable in the FD model? Or, is it the case that it does not matter even when we do not take the FD of the dummy variable?

18. What is the BE estimator? What is the required assumption? Why is it less popular in empirical applications? Illustrate two reasons.

19. The standard errors can be obtained differently if the usual assumption on the error term is not satisfied regarding the efficiency issue. Explain how we can control for each of (i) heteroskedasticity, (ii) autocorrelation, (iii) common factors affecting all cross sectional units uniformly, (iv) cross-correlation across cross sectional units, and (v) cross-correlation across clusters.