Time-Series-Cross-Section Data 2

Jason Qiang Guo, New York

Fixed Effects

- Basic Setup (motivation: unobserved time-invariant heterogeneity):
$$y_{it} = \alpha + \beta x_{it} + \delta u_i + \epsilon_{it}, i = 1, \ldots N, t = 1, \ldots T$$

Fixed Effects

- Basic Setup (motivation: unobserved time-invariant heterogeneity):
$$y_{it} = \alpha + \beta x_{it} + \delta u_i + \epsilon_{it}, i = 1, \ldots N, t = 1, \ldots T$$
- Exogeneity given x_i and u_i: $\mathbb{E}(\epsilon_{it}|x_i, u_i) = 0$

Fixed Effects

- Basic Setup (motivation: unobserved time-invariant heterogeneity):
$$y_{it} = \alpha + \beta x_{it} + \delta u_i + \epsilon_{it}, i = 1, \dots N, t = 1, \dots T$$
- Exogeneity given x_i and u_i: $\mathbb{E}(\epsilon_{it}|x_i, u_i) = 0$
- Homoskedasticity and independence across time and units: $Var(\epsilon_{it}) = \sigma_e^2$

Fixed Effects

- Basic Setup (motivation: unobserved time-invariant heterogeneity):
$$y_{it} = \alpha + \beta x_{it} + \delta u_i + \epsilon_{it}, i = 1, \ldots N, t = 1, \ldots T$$
- Exogeneity given x_i and u_i: $\mathbb{E}(\epsilon_{it}|x_i, u_i) = 0$
- Homoskedasticity and independence across time and units: $Var(\epsilon_{it}) = \sigma_e^2$
- LSDV representation: $\mathbf{Y} = \beta\mathbf{X} + \delta_{\mathbf{u}} u + \epsilon$, where

$$\mathbf{X} = \begin{bmatrix} x_{11}^{(1)} & x_{11}^{(2)} & \ldots & x_{11}^{(K)} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1T}^{(1)} & x_{1T}^{(2)} & \ldots & x_{1T}^{(K)} \\ x_{21}^{(1)} & x_{21}^{(2)} & \ldots & x_{21}^{(K)} \\ \vdots & \vdots & \vdots & \vdots \\ x_{2T}^{(1)} & x_{2T}^{(2)} & \ldots & x_{2T}^{(K)} \\ \ldots & \ldots & \ldots & \ldots \\ x_{N1}^{(1)} & x_{N1}^{(2)} & \ldots & x_{N1}^{(K)} \\ \vdots & \vdots & \vdots & \vdots \\ x_{NT}^{(1)} & x_{NT}^{(2)} & \ldots & x_{NT}^{(K)} \end{bmatrix}_{NT \times K} \text{ and } \delta_{\mathbf{u}} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \ldots & 0 \\ 0 & 0 & \ldots & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix}_{NT \times N}$$

Fixed Effects

- Basic Setup (motivation: unobserved time-invariant heterogeneity):
$$y_{it} = \alpha + \beta x_{it} + \delta u_i + \epsilon_{it}, i = 1, \dots N, t = 1, \dots T$$
- Exogeneity given x_i and u_i: $\mathbb{E}(\epsilon_{it}|x_i, u_i) = 0$
- Homoskedasticity and independence across time and units: $Var(\epsilon_{it}) = \sigma_e^2$
- LSDV representation: $\mathbf{Y} = \beta\mathbf{X} + \delta_{\mathbf{u}} u + \epsilon$, where

$$\mathbf{X} = \begin{bmatrix} x_{11}^{(1)} & x_{11}^{(2)} & \dots & x_{11}^{(K)} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1T}^{(1)} & x_{1T}^{(2)} & \dots & x_{1T}^{(K)} \\ x_{21}^{(1)} & x_{21}^{(2)} & \dots & x_{21}^{(K)} \\ \vdots & \vdots & \vdots & \vdots \\ x_{2T}^{(1)} & x_{2T}^{(2)} & \dots & x_{2T}^{(K)} \\ \dots & \dots & \dots & \dots \\ x_{N1}^{(1)} & x_{N1}^{(2)} & \dots & x_{N1}^{(K)} \\ \vdots & \vdots & \vdots & \vdots \\ x_{NT}^{(1)} & x_{NT}^{(2)} & \dots & x_{NT}^{(K)} \end{bmatrix}_{NT \times K} \quad \text{and } \delta_{\mathbf{u}} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}_{NT \times N}$$

- There are in total $(N + K)$ parameters to be estimated, and as a result the estimation is inefficient when T is small

Fixed Effects

- Let $\mathbf{Z} = [\mathbf{X}\delta_{\mathbf{u}}]$, and the vector of parameters becomes $\tau = [\beta u]$, and we get least square estimator $\hat{\tau}_{FE} = (\mathbf{Z}'\mathbf{Z})^{-\mathbf{1}}\mathbf{Z}'\mathbf{Y}$

Fixed Effects

- Let $\mathbf{Z} = [\mathbf{X}\delta_{\mathbf{u}}]$, and the vector of parameters becomes $\tau = [\beta u]$, and we get least square estimator $\hat{\tau}_{FE} = (\mathbf{Z}'\mathbf{Z})^{-\mathbf{1}}\mathbf{Z}'\mathbf{Y}$
- The dimension of Z gets really large when N is large, so . . .

Fixed Effects

- Let $\mathbf{Z} = [\mathbf{X}\delta_{\mathbf{u}}]$, and the vector of parameters becomes $\tau = [\beta u]$, and we get least square estimator $\hat{\tau}_{FE} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$
- The dimension of Z gets really large when N is large, so . . .
- Demeaned estimation: Regress Y and X on δ_u and then regress the residuals $\tilde{Y}$ and the residuals $\tilde{X}$, where

$$\tilde{Y} = \begin{pmatrix} y_{11} - \bar{y_1} \\ \vdots \\ y_{it} - \bar{y}_i \\ \vdots \\ y_{NT} - \bar{Y}_N \end{pmatrix} \text{ and } \tilde{X} = \begin{pmatrix} x_{11} - \bar{x_1} \\ \vdots \\ x_{it} - \bar{x}_i \\ \vdots \\ x_{NT} - \bar{x}_N \end{pmatrix}$$

Fixed Effects

- Let $\mathbf{Z} = [\mathbf{X}\delta_\mathbf{u}]$, and the vector of parameters becomes $\tau = [\beta u]$, and we get least square estimator $\hat{\tau}_{FE} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$
- The dimension of Z gets really large when N is large, so ...
- Demeaned estimation: Regress Y and X on δ_u and then regress the residuals $\tilde{Y}$ and the residuals $\tilde{X}$, where

$$\tilde{Y} = \begin{pmatrix} y_{11} - \bar{y_1} \\ \vdots \\ y_{it} - \bar{y_i} \\ \vdots \\ y_{NT} - \bar{Y_N} \end{pmatrix} \text{ and } \tilde{X} = \begin{pmatrix} x_{11} - \bar{x_1} \\ \vdots \\ x_{it} - \bar{x_i} \\ \vdots \\ x_{NT} - \bar{x_N} \end{pmatrix}$$

- Then $\hat{\beta_{FE}} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y} = \hat{\tau}_{FE}$, the within-group estimator

Random Effects

- Basic Setup

$$y_{it} = \alpha + \beta x_{it} + \delta u_i + \epsilon_{it}, \text{where } i = 1, \ldots N, t = 1, \ldots T$$

Random Effects

- Basic Setup

$$y_{it} = \alpha + \beta x_{it} + \delta u_i + \epsilon_{it}, \text{where } i = 1, \ldots N, t = 1, \ldots T$$

$$u_i \sim i.i.d.(0, \sigma_u^2)$$

$$\epsilon_{it} \sim i.i.d.(0, \sigma_e^2)$$

Random Effects

- Basic Setup

$$y_{it} = \alpha + \beta x_{it} + \delta u_i + \epsilon_{it}, \text{where } i = 1, \ldots N, t = 1, \ldots T$$

$$u_i \sim i.i.d.(0, \sigma_u^2)$$

$$\epsilon_{it} \sim i.i.d.(0, \sigma_e^2)$$

- The critical step is to derive the variance-covariance matrix of the augmented errors $(u + \epsilon)$, and we have $V = \Omega \otimes \mathbf{I_N}$ where

$$\Omega = \begin{pmatrix} \sigma_u^2 + \sigma_e^2 & \sigma_u^2 & \ldots & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 + \sigma_e^2 & \ldots & \sigma_u^2 \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_u^2 & \sigma_u^2 & \ldots & \sigma_u^2 + \sigma_e^2 \end{pmatrix}_{T \times T}$$

Random Effects

- Basic Setup

$$y_{it} = \alpha + \beta x_{it} + \delta u_i + \epsilon_{it}, \text{where } i = 1, \ldots N, t = 1, \ldots T$$

$$u_i \sim i.i.d.(0, \sigma_u^2)$$

$$\epsilon_{it} \sim i.i.d.(0, \sigma_e^2)$$

- The critical step is to derive the variance-covariance matrix of the augmented errors $(u + \epsilon)$, and we have $V = \Omega \otimes \mathbf{I_N}$ where

$$\Omega = \begin{pmatrix} \sigma_u^2 + \sigma_e^2 & \sigma_u^2 & \ldots & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 + \sigma_e^2 & \ldots & \sigma_u^2 \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_u^2 & \sigma_u^2 & \ldots & \sigma_u^2 + \sigma_e^2 \end{pmatrix}_{T \times T}$$

- MLE now can be implemented with $Y|X \sim N(Z\gamma, \Omega)$ where $Z = [1X]_{NT \times K}$ and $\gamma = [\alpha\beta]^T_{K \times 1})$

Random Effects

- Eventually, we can show that
$$\hat{\beta_{RE}} = (\tilde{X}'\tilde{X} + \frac{\sigma_e^2}{\sigma_e^2 + \sigma_u^2 T}\dot{X}'\dot{X})^{-1}(\tilde{X}'\tilde{X}\hat{\beta_W} + \frac{\sigma_e^2}{\sigma_e^2 + \sigma_u^2 T}\dot{X}'\dot{X}\hat{\beta}_B)$$

Random Effects

- Eventually, we can show that
  $\hat{\beta_{RE}} = (\tilde{X}'\tilde{X} + \frac{\sigma_e^2}{\sigma_e^2 + \sigma_u^2 T}\dot{X}'\dot{X})^{-1}(\tilde{X}'\tilde{X}\hat{\beta_W} + \frac{\sigma_e^2}{\sigma_e^2 + \sigma_u^2 T}\dot{X}'\dot{X}\hat{\beta_B})$
- Recall within-group estimator $\beta_W = \beta_{FE} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y}$

Random Effects

- Eventually, we can show that
$$\hat{\beta_{RE}} = (\tilde{X}'\tilde{X} + \frac{\sigma_e^2}{\sigma_e^2 + \sigma_u^2 T}\dot{X}'\dot{X})^{-1}(\tilde{X}'\tilde{X}\hat{\beta_W} + \frac{\sigma_e^2}{\sigma_e^2 + \sigma_u^2 T}\dot{X}'\dot{X}\hat{\beta_B})$$
- Recall within-group estimator $\beta_W = \beta_{FE} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y}$
- $\beta_B = (\dot{X}'\dot{X})^{-1}\dot{X}'\dot{Y}$, where $\dot{Y}_i = \bar{y}_i - \bar{y}$, $\dot{X}_i = \bar{x}_i - \bar{x}$, and
$$\dot{Y}_{NT\times 1} = \begin{pmatrix} 1_T(\bar{y_1} - \bar{y}) \\ \vdots \\ 1_T(\bar{y_N} - \bar{y}) \end{pmatrix} \text{ and } \dot{X}_{NT\times 1} = \begin{pmatrix} 1_T(\bar{x_1} - \bar{x}) \\ \vdots \\ 1_T(\bar{x_N} - \bar{x}) \end{pmatrix}$$
. We call it between-group estimator
- Clearly, we can see that the random effects estimator is just a weighted average of within-in group estimator and between-group estimator. If $T \to \infty$ or $\sigma_u^2 \to \infty$, then it is the case of fixed effects. If $\sigma_u^2 = 0$, then it becomes a pooled OLS estimator. (Will come back to this again when we study Bayesian!)

Fixed Effects

- Incidental Parameter Problem (Neyman and Scott): Fixed effects through MLE procedure are biased and poorly estimated when T is small, because the number of regressors increases with N while in MLE framework it is fixed.

Fixed Effects

- Incidental Parameter Problem (Neyman and Scott): Fixed effects through MLE procedure are biased and poorly estimated when T is small, because the number of regressors increases with N while in MLE framework it is fixed.
- Going to contaminate the rest of the coefficients, and for linear case it is not a big problem because we can use demeaned estimation such that the estimation of β does not depend on u_i.

Fixed Effects

- Incidental Parameter Problem (Neyman and Scott): Fixed effects through MLE procedure are biased and poorly estimated when T is small, because the number of regressors increases with N while in MLE framework it is fixed.
- Going to contaminate the rest of the coefficients, and for linear case it is not a big problem because we can use demeaned estimation such that the estimation of β does not depend on u_i.
- But for non-linear case we have a big problem (bivariate dependent variable panel data)

Fixed Effects

- Incidental Parameter Problem (Neyman and Scott): Fixed effects through MLE procedure are biased and poorly estimated when T is small, because the number of regressors increases with N while in MLE framework it is fixed.
- Going to contaminate the rest of the coefficients, and for linear case it is not a big problem because we can use demeaned estimation such that the estimation of β does not depend on u_i.
- But for non-linear case we have a big problem (bivariate dependent variable panel data)
- Assume a logit model with two independent time periods, fixed effects and one explanatory variable x_{it} s.t. $\forall i$, $x_{i1} = 0$ and $x_{i2} = 1$

$$\mathbf{P}(y_{it=1} | x, \alpha) = \frac{e^{\alpha_i + x_{it}\beta}}{1 + e^{\alpha_i + x_{it}\beta}}$$

if $y_{i1} = 0$ and $y_{i2} = 0$ then $\hat{\alpha}_i = -\infty$
if $y_{i1} = 1$ and $y_{i2} = 1$ then $\hat{\alpha}_i = \infty$
if $y_{i1} + y_{i2} = 1$ then $\hat{\alpha}_i = -\hat{\beta}/2$ and $\hat{\beta} \xrightarrow{p} 2\beta$

Random Effects

- Basic setup: $P(y_{it} = 1 | x_{it}, u_i) = \Phi(x_{it}\beta + u_i)$ where $y_{i,1}, \ldots y_{it}$ independent conditional on (x_i, u_i)

Random Effects

- Basic setup: $P(y_{it} = 1 | x_{it}, u_i) = \Phi(x_{it}\beta + u_i)$ where $y_{i,1}, \dots y_{it}$ independent conditional on (x_i, u_i)
- then we have $f(y_{i1}, \dots y_{iT} | X, u_i, \beta) = \prod_{t=1}^{T} f(y_{it} | X, u_i, \beta)$

Random Effects

- Basic setup: $P(y_{it} = 1|x_{it}, u_i) = \Phi(x_{it}\beta + u_i)$ where $y_{i,1}, \ldots y_{it}$ independent conditional on (x_i, u_i)
- then we have $f(y_{i1}, \ldots y_{iT}|X, u_i, \beta) = \prod_{t=1}^{T} f(y_{it}|X, u_i, \beta)$
- assuming $u_i|X \sim N(0, \sigma_u^2)$, we can integrate u_i out by

$$f(y_{i1}, \ldots, y_{iT}|X, \beta, \sigma_u^2) = \int [\prod_{t=1}^{T} f(y_{it}|X, u_i, \beta)] \frac{1}{\sigma_u} \phi(\frac{u}{\sigma_u}) du$$

Random Effects

- Basic setup: $P(y_{it} = 1|x_{it}, u_i) = \Phi(x_{it}\beta + u_i)$ where $y_{i,1}, \ldots y_{it}$ independent conditional on (x_i, u_i)
- then we have $f(y_{i1}, \ldots y_{iT}|X, u_i, \beta) = \prod_{t=1}^{T} f(y_{it}|X, u_i, \beta)$
- assuming $u_i|X \sim N(0, \sigma_u^2)$, we can integrate u_i out by

$$f(y_{i1}, \ldots, y_{iT}|X, \beta, \sigma_u^2) = \int [\prod_{t=1}^{T} f(y_{it}|X, u_i, \beta)] \frac{1}{\sigma_u} \phi(\frac{u}{\sigma_u}) du$$

- Already implemented in R and Stata

Random Effects

- Basic setup: $P(y_{it} = 1 | x_{it}, u_i) = \Phi(x_{it}\beta + u_i)$ where $y_{i,1}, \ldots y_{it}$ independent conditional on (x_i, u_i)
- then we have $f(y_{i1}, \ldots y_{iT} | X, u_i, \beta) = \prod_{t=1}^{T} f(y_{it} | X, u_i, \beta)$
- assuming $u_i | X \sim N(0, \sigma_u^2)$, we can integrate u_i out by

$$f(y_{i1}, \ldots, y_{iT} | X, \beta, \sigma_u^2) = \int [\prod_{t=1}^{T} f(y_{it} | X, u_i, \beta)] \frac{1}{\sigma_u} \phi(\frac{u}{\sigma_u}) du$$

- Already implemented in R and Stata
- The independence assumption of u_i and x_i is very strong

Conditional logit: removing the fixed effects

- Example with $T = 2$:

$$\Lambda_i = \begin{cases} 1 & \text{if } y_{i1} = 1, y_{i2} = 0 \\ 0 & \text{if } y_{i1} = 0, y_{i2} = 1 \end{cases}$$

Conditional logit: removing the fixed effects

- Example with $T = 2$:

$$\Lambda_i = \begin{cases} 1 & \text{if } y_{i1} = 1, y_{i2} = 0 \\ 0 & \text{if } y_{i1} = 0, y_{i2} = 1 \end{cases}$$

- $P(\Lambda_i = 1 | y_{i1} + y_{i2} = 1) = \dfrac{exp(\beta(x_{i1} - x_{i2}))}{1 + exp(\beta(x_{i1} - x_{i2}))} = \dfrac{exp(\beta x_{i1})}{exp(\beta x_{i1}) + exp(\beta x_{i2})}$
  (familiar?)

Conditional logit: removing the fixed effects

- Example with $T = 2$:
$$\Lambda_i = \begin{cases} 1 & \text{if } y_{i1} = 1, y_{i2} = 0 \\ 0 & \text{if } y_{i1} = 0, y_{i2} = 1 \end{cases}$$
- $P(\Lambda_i = 1 | y_{i1} + y_{i2} = 1) = \dfrac{exp(\beta(x_{i1} - x_{i2}))}{1 + exp(\beta(x_{i1} - x_{i2}))} = \dfrac{exp(\beta x_{i1})}{exp(\beta x_{i1}) + exp(\beta x_{i2})}$ (familiar?)
- So we have $L = \prod_{i=1}^{N} logit^{-1}(\beta(x_{i1} - x_{i2}))^{\Lambda_i}[1 - logit^{-1}(\beta(x_{i1} - x_{i2}))]^{1-\Lambda_i}$, and u_i disappears from the estimation.

Conditional logit: removing the fixed effects

- Example with $T = 2$:

$$\Lambda_i = \begin{cases} 1 & \text{if } y_{i1} = 1, y_{i2} = 0 \\ 0 & \text{if } y_{i1} = 0, y_{i2} = 1 \end{cases}$$

- $P(\Lambda_i = 1 | y_{i1} + y_{i2} = 1) = \dfrac{exp(\beta(x_{i1} - x_{i2}))}{1 + exp(\beta(x_{i1} - x_{i2}))} = \dfrac{exp(\beta x_{i1})}{exp(\beta x_{i1}) + exp(\beta x_{i2})}$ (familiar?)
- So we have $L = \prod_{i=1}^{N} logit^{-1}(\beta(x_{i1} - x_{i2}))^{\Lambda_i}[1 - logit^{-1}(\beta(x_{i1} - x_{i2}))]^{1-\Lambda_i}$, and u_i disappears from the estimation.
- In general:

$$\prod_{i=1}^{N} \frac{exp(\beta \sum_{t=1}^{T} x_{it} y_{it})}{\sum_{d \in D_i} exp(\beta \sum_{t=1}^{T} x_{it} d_t)}$$

where $D_i = \{d = (d_1, \ldots, d_T) | d_t = 0 \text{ or } 1, \text{and } \sum_{t=1}^{T} d_t = \sum_{t=1}^{T} y_{it}\}$

Conditional logit: removing the fixed effects

- Example with $T = 2$:

$$\Lambda_i = \begin{cases} 1 & \text{if } y_{i1} = 1, y_{i2} = 0 \\ 0 & \text{if } y_{i1} = 0, y_{i2} = 1 \end{cases}$$

- $P(\Lambda_i = 1 | y_{i1} + y_{i2} = 1) = \dfrac{exp(\beta(x_{i1} - x_{i2}))}{1 + exp(\beta(x_{i1} - x_{i2}))} = \dfrac{exp(\beta x_{i1})}{exp(\beta x_{i1}) + exp(\beta x_{i2})}$ (familiar?)
- So we have $L = \prod_{i=1}^{N} logit^{-1}(\beta(x_{i1} - x_{i2}))^{\Lambda_i}[1 - logit^{-1}(\beta(x_{i1} - x_{i2}))]^{1-\Lambda_i}$, and u_i disappears from the estimation.
- In general:

$$\prod_{i=1}^{N} \frac{exp(\beta \sum_{t=1}^{T} x_{it} y_{it})}{\sum_{d \in D_i} exp(\beta \sum_{t=1}^{T} x_{it} d_t)}$$

where $D_i = \{d = (d_1, \ldots, d_T) | d_t = 0 \text{ or } 1, \text{and } \sum_{t=1}^{T} d_t = \sum_{t=1}^{T} y_{it}\}$

- Equivalent to the partial likelihood function of the stratified Cox model in discrete time
  - single discrete time period
  - $y_{it} = 1$ observations are taken as failures at time 1
  - $y_{it} = 0$ observations are taken as alive/censored at time 1
  - units correspond to strata

Conditional logit: removing the fixed effects

Drawbacks:

Conditional logit: removing the fixed effects

Drawbacks:

- Loss of information: the identification uses only the individuals who change state

Conditional logit: removing the fixed effects

Drawbacks:

- Loss of information: the identification uses only the individuals who change state
- Heavily relies on functional form (does not work for Probit)

Conditional logit: removing the fixed effects

Drawbacks:

- Loss of information: the identification uses only the individuals who change state
- Heavily relies on functional form (does not work for Probit)
- Although $\hat{\beta}$ is fine, predicted probabilities, odds ratio and other quantities of interest can not be obtained