Panel Data Analysis
Introduction

- Model Representation
  - N-first or T-first representation
    - Pooled Model
    - Fixed Effects Model
    - Random Effects Model
  - Asymptotic Theory
    - $N\rightarrow\infty$, or $T\rightarrow\infty$
    - $N\rightarrow\infty$, $T\rightarrow\infty$
    - Panel-Robust Inference

Panel Data Analysis
Introduction

- The Model

$$\left.\begin{aligned} y_{it} &= \mathbf{x}_{it}^{'}\beta + \varepsilon_{it} \\ \varepsilon_{it} &= u_i + v_t + e_{it} \end{aligned}\right\} \Rightarrow y_{it} = \mathbf{x}_{it}^{'}\beta + u_i + v_t + e_{it}$$

  - One-Way (Individual) Effects: $y_{it} = \mathbf{x}_{it}^{'}\beta + u_i + e_{it}$
    - Unobserved Heterogeneity
    - Cross Section and Time Series Correlation

$$Cov(u_i, u_j) \neq 0, Cov(e_{it}, e_{jt}) \neq 0, i \neq j$$

$$Cov(e_{it}, e_{i\tau}) \neq 0, t \neq \tau$$

Panel Data Analysis

Introduction

- N-first Representation

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + u_i + e_{it}$$

$$i = 1, 2, ..., N; t = 1, 2, ..., T$$

$$\Downarrow$$

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + u_i\mathbf{i}_T + \mathbf{e}_i$$

$$\Downarrow$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{I}_N \otimes \mathbf{i}_T)\mathbf{u} + \mathbf{e}$$

- T-first Representation

$$y_{ti} = \mathbf{x}_{ti}'\boldsymbol{\beta} + u_i + e_{ti}$$

$$t = 1, 2, ..., T; i = 1, 2, ..., N$$

$$\Downarrow$$

$$\mathbf{y}_t = \mathbf{X}_t\boldsymbol{\beta} + \mathbf{u} + \mathbf{e}_t$$

$$\Downarrow$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{i}_T \otimes \mathbf{I}_N)\mathbf{u} + \mathbf{e}$$

- Dummy Variables Representation

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}\mathbf{u} + \mathbf{e}$$

$$\mathbf{D} = \mathbf{I}_N \otimes \mathbf{i}_T \text{ or } \mathbf{D} = \mathbf{i}_T \otimes \mathbf{I}_N$$

Panel Data Analysis
Introduction

- Notations

$$\mathbf{y}_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix}, \mathbf{X}_i = \begin{bmatrix} \mathbf{x}_{i1}' \\ \mathbf{x}_{i2}' \\ \vdots \\ \mathbf{x}_{iT}' \end{bmatrix} = \begin{bmatrix} x_{1,i1} & x_{2,i1} & \cdots & x_{K,i1} \\ x_{1,i2} & x_{2,i2} & \cdots & x_{K,i2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1,iT} & x_{2,iT} & \cdots & x_{K,iT} \end{bmatrix}, \mathbf{e}_i = \begin{bmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{iT} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}$$

$$\mathbf{y}_t = \begin{bmatrix} y_{t1} \\ y_{t2} \\ \vdots \\ y_{tN} \end{bmatrix}, \mathbf{X}_t = \begin{bmatrix} \mathbf{x}_{t1}' \\ \mathbf{x}_{t2}' \\ \vdots \\ \mathbf{x}_{tN}' \end{bmatrix} = \begin{bmatrix} x_{1,t1} & x_{2,t1} & \cdots & x_{K,t1} \\ x_{1,t2} & x_{2,t2} & \cdots & x_{K,t2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1,tN} & x_{2,tN} & \cdots & x_{K,tN} \end{bmatrix}, \mathbf{e}_t = \begin{bmatrix} e_{t1} \\ e_{t2} \\ \vdots \\ e_{tN} \end{bmatrix}, \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}$$

Example: Investment Demand

- Grunfeld and Griliches [1960]

$$I_{it} = \alpha_i + \beta F_{it} + \gamma C_{it} + \varepsilon_{it}$$

  - i = 10 firms: GM, CH, GE, WE, US, AF, DM, GY, UN, IBM; t = 20 years: 1935-1954
  - I_{it} = Gross investment
  - F_{it} = Market value
  - C_{it} = Value of the stock of plant and equipment

Pooled (Constant Effects) Model

$$y_{it} = \mathbf{x}_{it}^{'}\boldsymbol{\beta} + u_i + e_{it} \; (i = 1, 2, ..., N; t = 1, 2, ..., T)$$

$$\Downarrow \text{assuming } u = u_i \;\; \forall i$$

$$y_{it} = \mathbf{x}_{it}^{'}\boldsymbol{\beta} + u + e_{it}, \; or$$

$$y_{it} = \begin{bmatrix} \mathbf{x}_{it}^{'} & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ u \end{bmatrix} + e_{it} \Rightarrow \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

$$E(\mathbf{e} \mid \mathbf{X}) = \mathbf{0}, Var(\mathbf{e} \mid \mathbf{X}) = \sigma_e^2 \mathbf{I}$$

Fixed Effects Model

$$y_{it} = \mathbf{x}_{it}^{'}\boldsymbol{\beta} + u_i + e_{it} \quad (i = 1, 2, \ldots, N; t = 1, 2, \ldots, T)$$

- u_i is fixed, independent of e_{it}, and may be correlated with $\mathbf{x}_{it}$.

$$Cov(u_i, e_{it}) = 0, Cov(u_i, \mathbf{x}_{it}) \neq 0$$

$$\Rightarrow \begin{cases} \mathbf{y}_i = \mathbf{X}_i \beta + u_i i_T + \mathbf{e}_i, \; i = 1, 2, \ldots, N \\ \mathbf{y}_t = \mathbf{X}_t \beta + \mathbf{u} + \mathbf{e}_t, \; t = 1, 2, \ldots, T \end{cases}$$

Fixed Effects Model

- Fixed Effects Model
  - Classical Assumptions
    - Strict Exogeneity: $E(e_{it} \mid \mathbf{u}, \mathbf{X}) = 0$
    - Homoschedasticity: $Var(e_{it} \mid \mathbf{u}, \mathbf{X}) = \sigma_e^2$
    - No cross section and time series correlation:

$$Var(\mathbf{e} \mid \mathbf{u}, \mathbf{X}) = \sigma_e^2 \mathbf{I}_{NT}$$

  - Extensions: $Var(\mathbf{e} \mid \mathbf{u}, \mathbf{X}) = \Omega$
    - Panel Robust Variance-Covariance Matrix

Random Effects Model

- Error Components

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

$$\varepsilon_{it} = u_i + e_{it} \ (i = 1, 2, \ldots, N; t = 1, 2, \ldots, T)$$

  - u_i is random, independent of e_{it} and $\mathbf{x}_{it}$.

$$Cov(u_i, e_{it}) = 0, Cov(u_i, \mathbf{x}_{it}) = 0, Cov(e_{it}, \mathbf{x}_{it}) = 0$$

  - Define the error components as $\varepsilon_{it} = u_i + e_{it}$

$$\Rightarrow \begin{cases} \mathbf{y}_i = \mathbf{X}_i\beta + (u_i i_T + \mathbf{e}_i), i = 1, 2, \ldots, N \\ \mathbf{y}_t = \mathbf{X}_t\beta + (\mathbf{u} + \mathbf{e}_t), t = 1, 2, \ldots, T \end{cases}$$

Random Effects Model

- Random Effects Model
  - Classical Assumptions
    - Strict Exogeneity

$$E(e_{it} \mid \mathbf{X}) = 0,\ E(u_i \mid \mathbf{X}) = 0 \Rightarrow E(\varepsilon_{it} \mid \mathbf{X}) = 0$$

      - **X** includes a constant term, otherwise $E(u_i|\mathbf{X})=u$.
    - Homoschedasticity

$$Var(e_{it} \mid \mathbf{X}) = \sigma_e^2, Var(u_i \mid \mathbf{X}) = \sigma_u^2, Cov(u_i, e_{it}) = 0$$

$$\Rightarrow Var(\varepsilon_{it} \mid \mathbf{X}) = \sigma_\varepsilon^2 = \sigma_e^2 + \sigma_u^2$$

    - Constant Auto-covariance (within panels)

$$Var(\boldsymbol{\varepsilon}_i \mid \mathbf{X}) = \sigma_e^2 \mathbf{I}_T + \sigma_u^2 \mathbf{i}_T \mathbf{i}_T'$$

Random Effects Model

- Random Effects Model
  - Classical Assumptions (Continued)
    - Cross Section Independence

$$Var(\boldsymbol{\varepsilon}_i \mid \mathbf{X}) = \Omega = \sigma_e^2 \mathbf{I}_T + \sigma_u^2 \mathbf{i}_T \mathbf{i}_T'$$

$$Var(\boldsymbol{\varepsilon} \mid \mathbf{X}) = \boldsymbol{\Omega} = \mathbf{I}_N \otimes \Omega$$

  - Extensions:
    - Panel Robust Variance-Covariance Matrix

Fixed Effects Model
Estimation

- Within Model Representation

$$y_{it} = \mathbf{x}_{it}^{'}\boldsymbol{\beta} + u_i + e_{it}$$

$$\bar{y}_i = \bar{\mathbf{x}}_i^{'}\boldsymbol{\beta} + u_i + \bar{e}_i$$

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it}^{'} - \bar{\mathbf{x}}_i^{'})\boldsymbol{\beta} + (e_{it} - \bar{e}_i)$$

$$\Rightarrow \tilde{y}_{it} = \tilde{\mathbf{x}}_{it}^{'}\boldsymbol{\beta} + \tilde{e}_{it}$$

$$\tilde{\mathbf{y}}_i = \tilde{\mathbf{X}}_i\boldsymbol{\beta} + \tilde{\mathbf{e}}_i \quad or$$

$$Q\mathbf{y}_i = Q\mathbf{X}_i\boldsymbol{\beta} + Q\mathbf{e}_i$$

$$where \quad Q = \mathbf{I}_T - \frac{1}{T}\mathbf{i}_T\mathbf{i}_T^{'}, \quad (Q\mathbf{i}_T = 0, Q'Q = Q)$$

Fixed Effects Model
Estimation

- Model Assumptions

$$E(\tilde{e}_{it} \mid \tilde{\mathbf{x}}_{it}) = 0$$

$$Var(\tilde{e}_{it} \mid \tilde{\mathbf{x}}_{it}) = (1 - 1/T)\sigma_e^2$$

$$Cov(\tilde{e}_{it}, \tilde{e}_{is} \mid \tilde{\mathbf{x}}_{it}, \tilde{\mathbf{x}}_{is}) = (-1/T)\sigma_e^2 \neq 0, t \neq s$$

$$\Downarrow$$

$$Var(\tilde{\mathbf{e}}_i \mid \tilde{\mathbf{X}}_i) = \Omega = \sigma_e^2 Q = \sigma_e^2 (\mathbf{I}_T - \frac{1}{T}\mathbf{i}_T \mathbf{i}_T^{'})$$

$$Var(\tilde{\mathbf{e}} \mid \tilde{\mathbf{X}}) = \mathbf{\Omega} = \mathbf{I}_N \otimes \Omega$$

Fixed Effects Model
Estimation: OLS

- Within Estimator: OLS

$$\tilde{\mathbf{y}}_i = \tilde{\mathbf{X}}_i \boldsymbol{\beta} + \tilde{\mathbf{e}}_i \quad \Rightarrow \tilde{\mathbf{y}} = \tilde{\mathbf{X}} \boldsymbol{\beta} + \tilde{\mathbf{e}}$$

$$\hat{\boldsymbol{\beta}}_{OLS} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}'\tilde{\mathbf{y}} = \left[\sum_{i=1}^{N} \tilde{\mathbf{X}}_i' \tilde{\mathbf{X}}_i \right]^{-1} \sum_{i=1}^{N} \tilde{\mathbf{X}}_i' \tilde{\mathbf{y}}_i$$

$$\hat{V}ar(\hat{\boldsymbol{\beta}}_{OLS}) = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}'\boldsymbol{\Omega}\tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}$$

$$= \hat{\sigma}_e^2 \left[\sum_{i=1}^{N} \tilde{\mathbf{X}}_i' \tilde{\mathbf{X}}_i \right]^{-1} \left[\sum_{i=1}^{N} \tilde{\mathbf{X}}_i' Q \tilde{\mathbf{X}}_i \right] \left[\sum_{i=1}^{N} \tilde{\mathbf{X}}_i' \tilde{\mathbf{X}}_i \right]^{-1}$$

$$= \hat{\sigma}_e^2 \left[\sum_{i=1}^{N} \tilde{\mathbf{X}}_i' \tilde{\mathbf{X}}_i \right]^{-1}$$

$$\hat{\sigma}_e^2 = \hat{\tilde{\mathbf{e}}}'\hat{\tilde{\mathbf{e}}} / (NT - N - K), \quad \hat{\tilde{\mathbf{e}}} = \tilde{\mathbf{y}} - \tilde{\mathbf{X}}\hat{\boldsymbol{\beta}}$$

Fixed Effects Model

Estimation: ML

- Normality Assumption

$$y_{it} = \mathbf{x}_{it}^{'}\boldsymbol{\beta} + u_i + e_{it} \quad (t = 1, 2, ..., T)$$

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + u_i\mathbf{i}_{T_i} + \mathbf{e}_i \quad (i = 1, 2, ..., N)$$

$$\mathbf{e}_i \sim normal\ iid(\mathbf{0}, \sigma_e^2\mathbf{I}_T)$$

$$\Downarrow$$

$$\tilde{\mathbf{y}}_i = \tilde{\mathbf{X}}_i\boldsymbol{\beta} + \tilde{\mathbf{e}}_i \ with\ \tilde{\mathbf{y}}_i = Q\mathbf{y}_i, \tilde{\mathbf{X}}_i = Q\mathbf{X}_i, \tilde{\mathbf{e}}_i = Q\mathbf{e}_i,$$

$$Q = \mathbf{I}_T - \frac{1}{T}\mathbf{i}_T\mathbf{i}_T^{'}$$

$$\tilde{\mathbf{e}}_i \sim normal(0, \Omega),\ where\ \Omega = \sigma_e^2 QQ' = \sigma_e^2 Q$$

Fixed Effects Model

Estimation: ML

- Log-Likelihood Function

$$ll_i(\boldsymbol{\beta}, \sigma_e^2 \mid \mathbf{y}_i, \mathbf{X}_i) = -\frac{T}{2}\ln\left(2\pi\right) - \frac{1}{2}\ln\left|\Omega\right| - \frac{1}{2}\tilde{\mathbf{e}}_i'\Omega^{-1}\tilde{\mathbf{e}}_i$$

$$= -\frac{T}{2}\ln\left(2\pi\right) - \frac{T}{2}\ln(\sigma_e^2) - \frac{1}{2}\ln\left|Q\right| - \frac{1}{2\sigma_e^2}\tilde{\mathbf{e}}_i'Q^{-1}\tilde{\mathbf{e}}_i$$

- Since Q is singular and |Q|=0, we maximize

$$l\tilde{l}_i(\boldsymbol{\beta}, \sigma_{\tilde{e}}^2 \mid \mathbf{y}_i, \mathbf{X}_i) = -\frac{T}{2}\ln\left(2\pi\right) - \frac{T}{2}\ln(\sigma_{\tilde{e}}^2) - \frac{1}{2\sigma_{\tilde{e}}^2}\tilde{\mathbf{e}}_i'\tilde{\mathbf{e}}_i$$

Fixed Effects Model

Estimation: ML

- ML Estimator

$$(\hat{\boldsymbol{\beta}}, \tilde{\sigma}_{\tilde{e}}^{2})_{ML} = \arg\max \sum_{i=1}^{N} l\tilde{l}_i(\boldsymbol{\beta}, \sigma_{\tilde{e}}^{2} \mid \mathbf{y}_i, \mathbf{X}_i)$$

$$\hat{\sigma}_{\tilde{e}}^{2} = \frac{\sum_{i=1}^{N} \hat{\tilde{\mathbf{e}}}_i' \hat{\tilde{\mathbf{e}}}_i}{NT} = \left(1 - \frac{1}{T}\right)\hat{\sigma}_{e}^{2},\ \ \hat{\tilde{\mathbf{e}}}_i = \tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i \hat{\boldsymbol{\beta}}$$

$$\Rightarrow \hat{\sigma}_{e}^{2} = \left(\frac{T}{T-1}\right)\hat{\sigma}_{\tilde{e}}^{2} = \frac{\hat{\tilde{\mathbf{e}}}'\hat{\tilde{\mathbf{e}}}}{N(T-1)}$$

Fixed Effects Model

Hypothesis Testing

- Pool or Not Pool
  - F-Test based on dummy variable model: constant or zero coefficients for D w.r.t F(N-1,NT-N-K)
  - F-test based on fixed effects (unrestricted) model vs. pooled (restricted) model

$$y_{it} = \mathbf{x}_{it}^{'}\boldsymbol{\beta} + u_i + e_{it}$$

$$vs. \quad (u_i = u, \forall i)$$

$$y_{it} = \mathbf{x}_{it}^{'}\boldsymbol{\beta} + u + e_{it}$$

$$F = \frac{(RSS_R - RSS_{UR}) / N - 1}{RSS_{UR} / (NT - N - K)} \sim F(N-1, NT-N-K)$$

$$RSS_{UR} = \hat{\mathbf{e}}_{FE}^{'}\hat{\mathbf{e}}_{FE}, \; RSS_R = \hat{\mathbf{e}}_{PO}^{'}\hat{\mathbf{e}}_{PO}$$

Random Effects Model

Estimation: GLS

- The Model

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \;\; \boldsymbol{\varepsilon}_i = u_i\mathbf{i}_T + \mathbf{e}_i$$

$$E(\boldsymbol{\varepsilon}_i \mid \mathbf{X}_i) = \mathbf{0}$$

$$Var(\boldsymbol{\varepsilon}_i \mid \mathbf{X}_i) = \Omega = \sigma_e^2\mathbf{I}_T + \sigma_u^2\mathbf{i}_T\mathbf{i}_T^{'}$$

$$= \sigma_e^2\left[Q + \frac{\sigma_e^2 + T\sigma_u^2}{\sigma_e^2}\left(\mathbf{I}_T - Q\right)\right]$$

$$where\; Q = \mathbf{I}_T - \frac{1}{T}\mathbf{i}_T\mathbf{i}_T^{'}, \; \mathbf{I}_T - Q = \frac{1}{T}\mathbf{i}_T\mathbf{i}_T^{'}$$

Random Effects Model
Estimation: GLS

- GLS

$$\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}\boldsymbol{\Omega}^{-1}\mathbf{y} = \left[\sum_{i=1}^{N}\mathbf{X}_i\Omega^{-1}\mathbf{X}_i\right]^{-1}\sum_{i=1}^{N}\mathbf{X}_i\Omega^{-1}\mathbf{y}_i$$

$$Var(\hat{\boldsymbol{\beta}}_{GLS}) = (\mathbf{X}\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1} = \left[\sum_{i=1}^{N}\mathbf{X}_i\Omega^{-1}\mathbf{X}_i\right]^{-1}$$

$$where\ \Omega^{-1} = \frac{1}{\sigma_e^2}\left[\mathbf{I}_T - \frac{\sigma_u^2}{\sigma_e^2 + T\sigma_u^2}\mathbf{i}_T\mathbf{i}_T'\right] = \frac{1}{\sigma_e^2}\left[Q + \frac{\sigma_e^2}{\sigma_e^2 + T\sigma_u^2}\left(\mathbf{I}_T - Q\right)\right]$$

$$and\ \Omega^{-\frac{1}{2}} = \frac{1}{\sigma_e}\left[Q + \sqrt{\frac{\sigma_e^2}{\sigma_e^2 + T\sigma_u^2}}\left(\mathbf{I}_T - Q\right)\right]$$

Random Effects Model

Estimation: GLS

- Feasible GLS
  - Based on estimated residuals of fixed effects model

$$\hat{\sigma}_e^2 = \hat{\tilde{\mathbf{e}}}'\hat{\tilde{\mathbf{e}}} / N(T-1)$$

$$\hat{\sigma}_1^2 = T\hat{\sigma}_u^2 + \hat{\sigma}_e^2 = T\bar{\hat{\tilde{\mathbf{e}}}}'\bar{\hat{\tilde{\mathbf{e}}}} / N, \; where \; \bar{\hat{\tilde{e}}}_i = \frac{1}{T}\sum_{t=1}^{T}\hat{\tilde{e}}_{it}$$

$$\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X})^{-1}\mathbf{X}\hat{\boldsymbol{\Omega}}^{-1}\mathbf{y}$$

$$Var(\hat{\boldsymbol{\beta}}_{GLS}) = (\mathbf{X}\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X})^{-1}$$

$$where \; \hat{\Omega}^{-1} = \left[\frac{1}{\hat{\sigma}_e^2}Q + \frac{1}{\hat{\sigma}_1^2}\left(\mathbf{I}_T - Q\right)\right], \hat{\sigma}_1^2 = \hat{\sigma}_e^2 + T\hat{\sigma}_u^2$$

Random Effects Model

Estimation: GLS

- Feasible GLS
  - Within Model Representation

$$\theta = 1 - \sqrt{\frac{\sigma_e^2}{\sigma_1^2}} = 1 - \sqrt{\frac{\sigma_e^2}{T\sigma_u^2 + \sigma_e^2}}$$

$$y_{it} - \theta \bar{y}_i = (\mathbf{x}_{it}' - \theta \bar{\mathbf{x}}_i')\beta + (\varepsilon_{it} - \theta \bar{\varepsilon}_i)$$

$$\tilde{y}_{it} = \tilde{\mathbf{x}}_{it}'\beta + \tilde{\varepsilon}_{it}$$

$$\tilde{\varepsilon}_{it} = \varepsilon_{it} - \theta \bar{\varepsilon}_i = (1-\theta)u_i + (e_{it} - \theta \bar{e}_i)$$

$$E(\tilde{\varepsilon}_{it}) = 0, Var(\tilde{\varepsilon}_{it}) = \sigma_e^2$$

$$Cov(\tilde{\varepsilon}_{it}, \tilde{\varepsilon}_{i\tau}) = Cov(\tilde{\varepsilon}_{it}, \tilde{\varepsilon}_{jt}) = 0$$

Random Effects Model

Estimation: ML

- Log-Likelihood Function

$$y_{it} = \mathbf{x}_{it}^{'}\boldsymbol{\beta} + (u_i + e_{it}) = \mathbf{x}_{it}^{'}\boldsymbol{\beta} + \varepsilon_{it} \quad (t = 1, 2, ..., T)$$

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \quad (i = 1, 2, ..., N)$$

$$\boldsymbol{\varepsilon}_i \sim normal\ iid(\mathbf{0}, \Omega)$$

$$\Downarrow$$

$$ll_i(\boldsymbol{\beta}, \sigma_e^2, \sigma_u^2 \mid \mathbf{y}_i, \mathbf{X}_i) = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln|\Omega| - \frac{1}{2}\boldsymbol{\varepsilon}_i^{'}\Omega^{-1}\boldsymbol{\varepsilon}_i$$

Random Effects Model

Estimation: ML

- where

$$\Omega = \sigma_e^2 \mathbf{I}_T + \sigma_u^2 \mathbf{i}_T \mathbf{i}_T' = \left[Q - \frac{\sigma_e^2 + T\sigma_u^2}{\sigma_e^2} (\mathbf{I}_T - Q) \right]$$

$$\Omega^{-1} = \frac{1}{\sigma_e^2} \left(\mathbf{I}_T - \frac{\sigma_u^2}{T\sigma_u^2 + \sigma_e^2} \mathbf{i}_T \mathbf{i}_T' \right) = \frac{1}{\sigma_e^2} \left[Q + \frac{\sigma_e^2}{T\sigma_u^2 + \sigma_e^2} (\mathbf{I}_T - Q) \right]$$

$$|\Omega| = (\sigma_e^2)^T \left| \mathbf{I}_T + \frac{\sigma_u^2}{\sigma_e^2} \mathbf{i}_T \mathbf{i}_T' \right| = (\sigma_e^2)^T \left(1 + \frac{T\sigma_u^2}{\sigma_e^2} \right)$$

Random Effects Model

Estimation: ML

- ML Estimator

$$(\hat{\boldsymbol{\beta}}, \hat{\sigma}_e^2, \hat{\sigma}_u^2)_{ML} = \arg\max \sum_{i=1}^{N} ll_i(\boldsymbol{\beta}, \sigma_e^2, \sigma_u^2 \mid \mathbf{y}_i, \mathbf{X}_i)$$

where

$$ll_i(\boldsymbol{\beta}, \sigma_e^2, \sigma_u^2 \mid \mathbf{y}_i, \mathbf{X}_i) = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln|\Omega| - \frac{1}{2}\boldsymbol{\varepsilon}_i \Omega^{-1}\boldsymbol{\varepsilon}_i$$

$$= -\frac{T}{2}\ln\left(2\pi\sigma_e^2\right) - \frac{1}{2}\ln\left(\frac{\sigma_e^2 + T\sigma_u^2}{\sigma_e^2}\right)$$

$$-\frac{1}{2\sigma_e^2}\left\{\left[\sum_{t=1}^{T}(y_{it} - \mathbf{x}_{it}^{'}\boldsymbol{\beta})^2\right] - \frac{\sigma_u^2}{\sigma_e^2 + T\sigma_u^2}\left[\sum_{t=1}^{T}(y_{it} - \mathbf{x}_{it}^{'}\boldsymbol{\beta})\right]^2\right\}$$

Random Effects Model

Hypothesis Testing

- Pool or Not Pool
  - Test for $Var(u_i) = 0$, that is

$$Cov(\varepsilon_{it}, \varepsilon_{is}) = Cov(u_i + e_{it}, u_i + e_{is}) = Cov(e_{it}, e_{is})$$

  - For balanced panel data, the Lagrange-multiplier test statistic (Breusch-Pagan, 1980) is:

Random Effects Model

Hypothesis Testing

- Pool or Not Pool (Cont.)

$$LM = \frac{NT}{2(T-1)}\left[\frac{\hat{\mathbf{e}}'(\mathbf{J}_T \otimes \mathbf{I}_N)\hat{\mathbf{e}}}{\hat{\mathbf{e}}'\hat{\mathbf{e}}} - 1\right] \sim \chi^2(1)$$

$$= \frac{NT}{2(T-1)}\left[\frac{\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\hat{e}_{it}\right)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{e}_{it}^2} - 1\right]^2$$

$$where\ \hat{e}_{it} = y_{it} - \begin{bmatrix}\mathbf{x}_{it}' & 1\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{\beta}} \\ \hat{u}\end{bmatrix}_{Pooled}$$

Random Effects Model

Hypothesis Testing

- Fixed Effects vs. Random Effects

$$H_0 : Cov(u_i, \mathbf{x}_{it}^{'}) = 0 \ (random\ effects)$$

$$H_1 : Cov(u_i, \mathbf{x}_{it}^{'}) \neq 0 \ (fixed\ effects)$$

Estimator	Random Effects $E(u_i\|X_i) = 0$	Fixed Effects $E(u_i\|X_i)$ =/= 0
GLS or RE-OLS (Random Effects)	Consistent and Efficient	Inconsistent
LSDV or FE-OLS (Fixed Effects)	Consistent Inefficient	Consistent Possibly Efficient

Random Effects Model

Hypothesis Testing

- Fixed effects estimator is consistent under H_0 and H_1; Random effects estimator is efficient under H_0, but it is inconsistent under H_1.
- Hausman Test Statistic

$$H = \left(\hat{\boldsymbol{\beta}}_{RE} - \hat{\boldsymbol{\beta}}_{FE}\right)' \left[Var(\hat{\boldsymbol{\beta}}_{RE}) - Var(\hat{\boldsymbol{\beta}}_{FE})\right]^{-1} \left(\hat{\boldsymbol{\beta}}_{RE} - \hat{\boldsymbol{\beta}}_{FE}\right)$$

$$\sim \chi^2(\#\hat{\boldsymbol{\beta}}_{FE}), \quad provided\ \#\hat{\boldsymbol{\beta}}_{FE} = \#\hat{\boldsymbol{\beta}}_{RE}\ (no\ intercept)$$

Random Effects Model

Hypothesis Testing

- Alternative Hausman Test
  - Estimate the random effects model

$$y_{it} = \mathbf{x}_{it}^{'}\boldsymbol{\beta} + \bar{\mathbf{x}}_{i}^{'}\boldsymbol{\gamma} + (u_i + e_{it})$$

  - F Test that $\gamma = 0$

$$H_0 : \boldsymbol{\gamma} = 0 \quad \Leftrightarrow \quad H_0 : Cov(u_i, \mathbf{x}_{it}) = 0$$

Extensions

- Random Coefficients Model

$$\left.\begin{aligned} y_{it} &= \mathbf{x}_{it}^{'}\boldsymbol{\beta}_i + e_{it} \\ \boldsymbol{\beta}_i &= \boldsymbol{\beta} + \mathbf{u}_i \end{aligned}\right\} y_{it} = \mathbf{x}_{it}^{'}\boldsymbol{\beta} + (\mathbf{x}_{it}^{'}\mathbf{u}_i + e_{it})$$

- Mixed Effects Model

$$y_{it} = \mathbf{x}_{it}^{'}\boldsymbol{\beta} + (\mathbf{z}_{it}^{'}\boldsymbol{\gamma}_i + e_{it})$$

- Two-Way Effects $y_{it} = \mathbf{x}_{it}^{'}\boldsymbol{\beta} + u_i + v_t + e_{it}$
- Nested Random Effects $y_{ijt} = \mathbf{x}_{ijt}^{'}\boldsymbol{\beta} + u_i + v_{ij} + e_{ijt}$

Example: U. S. Productivity

- Munnell [1988] Productivity Data
  48 Continental U.S. States, 17 Years:1970-1986
  - STATE = State name,
  - ST ABB=State abbreviation,
  - YR =Year, 1970, . . . ,1986,
  - PCAP =Public capital,
  - HWY =Highway capital,
  - WATER =Water utility capital,
  - UTIL =Utility capital,
  - PC =Private capital,
  - GSP =Gross state product,
  - EMP =Employment,

References

- B. H. Baltagi, *Econometric Analysis of Panel Data*, 4th ed., John Wiley, New York, 2008.
- W. H. Greene, *Econometric Analysis*, 7th ed., Chapter 11: Models for Panel Data, Prentice Hall, 2011.
- C. Hsiao, *Analysis of Panel Data*, 2nd ed., Cambridge University Press, 2003.
- J. M. Wooldridge, *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, 2002.