Empirical Panel Data: Lecture 5

INSTRUCTOR: CHAOYI CHEN
NJE & MNB

April 5, 2023

Topic 3: A linear unobserved effects panel data models: error-component model

- A generalized setup for a linear unobserved (individual) effects panel data model consist of three components

$$y_{it} = \beta^\top x_{it} + u_{it}, 1 \le i \le n,\ 1 \le t \le T \quad (1)$$
$$u_{it} = \alpha_i + \lambda_t + \varepsilon_{it},$$

- α_i is the **individual** effect
- λ_t is the **time** effect
- ε_{it} is the **idiosyncratic error** term

- We make the following assumptions to support the model (1).

  Assumptions RE: The errors terms $u_{it} = \alpha_i + \lambda_t + \varepsilon_{it}$ are i.i.d. for all $1 \le i \le n$, $1 \le t \le T$ with

  1. $E(\alpha_i) = E(\lambda_t) = E(\varepsilon_{it}) = 0$
  2. $E(\alpha_i \lambda_t) = E(\lambda_t \varepsilon_{it}) = E(\alpha_i \varepsilon_{it}) = 0$
  3. $E(\alpha_i \alpha_j) = \begin{cases} \sigma_\alpha^2, & i = j \\ 0, & i \neq j \end{cases}$
  4. $E(\lambda_t \lambda_s) = \begin{cases} \sigma_\lambda^2, & i = j \\ 0, & i \neq j \end{cases}$
  5. $E(\varepsilon_{it} \varepsilon_{js}) = \begin{cases} \sigma_\varepsilon^2, & t = s, \; i = j \\ 0, & \text{otherwise} \end{cases}$
  6. $E(\alpha_i x_{it}^\top) = E(\lambda_t x_{it}^\top) = E(\varepsilon_{it} x_{it}^\top) = 0$

Topic 3: Remarks on Assumptions RE

- Assumptions RE implies α_i is uncorrelated with x_{it}.

- Under Assumptions RE, the variance of y_{it} conditional on x_{it} is equal to:

$$\sigma^2_{y|x} = \sigma^2_u = \sigma^2_\alpha + \sigma^2_\lambda + \sigma^2_\varepsilon$$

- Assumptions RE can be extended to **non-zero-mean** unobserved individual effect, α_i^*, with $E(\alpha_i^*) = \mu$. Then, we can define $\alpha_i^* = \mu + \alpha_i$ and the new error-component model is

$$y_{it} = \mu + \beta^\top x_{it} + u_{it}, 1 \leq i \leq n,\ 1 \leq t \leq T$$
$$u_{it} = \alpha_i + \lambda_t + \varepsilon_{it},$$

- W.l.o.g., in the following, for simplicity, we do not introduce any **time** effects and consider a simple random effect model with a non-zero-mean α_i, where $u_{it} = \alpha_i + \varepsilon_{it}$.

Topic 3: A vector form of a random effect model

- For the random effect model

$$y_{it} = \mu + \beta^\top x_{it} + u_{it}, 1 \le i \le n, \ 1 \le t \le T$$
$$u_{it} = \alpha_i + \varepsilon_{it},$$

we can use the following vectorial expression to redefine it

$$\underset{(T\times 1)}{\boldsymbol{y}_i} = \underset{(T\times k+1)}{\boldsymbol{X}_i} \underset{(k+1\times 1)}{\gamma} + \underset{(T\times 1)}{\boldsymbol{u_i}}, \qquad (2)$$
$$\underset{(T\times 1)}{\boldsymbol{u_i}} = \underset{(T\times 1)}{\boldsymbol{e}} \alpha_i + \underset{(T\times 1)}{\boldsymbol{\varepsilon_i}} \qquad (3)$$

- $\underset{(k+1\times 1)}{\boldsymbol{X}_i} = \left(\underset{(T\times 1)}{\boldsymbol{e}}, \underset{(T\times k)}{\boldsymbol{x_i}} \right)$
- $\underset{(k+1\times 1)}{\gamma} = \left(\mu, \boldsymbol{\beta}^\top \right)^\top$

- Under Assumptions RE, the variance-covariance matrix of $\boldsymbol{u_i}$ is equal to

$$\Omega = E(\boldsymbol{u_i}\boldsymbol{u_i}^\top) = E\left[(\boldsymbol{e}\alpha_i + \boldsymbol{\varepsilon_i})(\boldsymbol{e}\alpha_i + \boldsymbol{\varepsilon_i})^\top\right] = \sigma_\alpha^2 \boldsymbol{e}\boldsymbol{e}^\top + \sigma_\varepsilon^2 I_T$$

$$\underset{T\times T}{\Omega} = \begin{bmatrix} \sigma_\alpha^2 + \sigma_\varepsilon^2 & \sigma_\alpha^2 & \dots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_\varepsilon^2 & \dots & \sigma_\alpha^2 \\ \dots & \dots & \dots & \dots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \dots & \sigma_\alpha^2 + \sigma_\varepsilon^2 \end{bmatrix}$$

- The off-diagonal elements are **non-zero** due to the presence of α_i produces a correlation among errors of the same cross-sectional unit (autocorrelation)!

- The inverse matrix of Ω is

$$\Omega^{-1} = \frac{1}{\sigma_\varepsilon^2}\left[I_T - \left(\frac{\sigma_\alpha^2}{\sigma_\varepsilon^2 + T\sigma_\alpha^2} \right) \boldsymbol{e}\boldsymbol{e}^\top \right]$$

- Let $\underset{(nT\times 1)}{\boldsymbol{u}} = \left[\boldsymbol{u_1}^\top, \boldsymbol{u_2}^\top, \ldots, \boldsymbol{u_n}^\top \right]^\top$, we have

$$\underset{(nT\times nT)}{\boldsymbol{\Omega}(\boldsymbol{u})} = E(\boldsymbol{u}\boldsymbol{u}^\top) = \Omega \otimes I_n$$

$$\boldsymbol{\Omega}(\boldsymbol{u}) = \begin{bmatrix} \Omega & 0 & \ldots & 0 \\ 0 & \Omega & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \Omega \end{bmatrix}$$

Topic 3: Within transformation in random effects model

- In the last lecture, we used an idempotent matrix $Q = I_T - \frac{1}{T}\boldsymbol{e}\boldsymbol{e}^T$ to eliminate the individual effect α_i in the fixed effect model. Similarly, this technique can be used to obtain the within-group (or LSDV) estimator in the random effect model.
- Under Assumptions RE, where α_i is random and correlated with x_{it}, the within-group estimator is unbiased and consistent as either n, or T, or both tend to infinity.
- However, the within-group estimator is **not** the Best Linear Unbiased Estimator (BLUE).
- In this case, the is the Generalized Least Squares (GLS) estimator is the BLUE estimator. [1]

[1] The next few slides provide a review of prior-knowledge on GLS.

Review: Weighted least squares (WLS)

- Consider the classical linear regression model as studied in Lecture 1:

$$\boldsymbol{y} = \boldsymbol{x}\boldsymbol{\beta} + \mu$$

- Assume that Gauss-Markov Assumptions 1-4 hold, but Assumption 5 does **not** hold (i.e., we assume heteroscedastic errors). Specifically, we assume $E(\mu\mu^\top) = \Omega$, where Ω may depend on i or be correlated with x_i. However, we assume there is **no** autocorrealtion and Ω is a diagonal matrix.

- The weighted least squares (WLS) estimator with a diagonal weighting matrix W can be obtained by minimizing

$$\widehat{\boldsymbol{\beta}}^{WLS} = \underset{\boldsymbol{\beta}\in\Theta}{argmin}(\boldsymbol{y} - \boldsymbol{x}\boldsymbol{\beta})^\top W(\boldsymbol{y} - \boldsymbol{x}\boldsymbol{\beta}).$$

- Therefore, OLS is a special case of WLS with $W = I_n$, which suggests that each observation has the same weight.

Review: Vairance-covariance matrix of WLS and OLS

- Taking the derivative w.r.t. $\boldsymbol{\beta}$ we have

$$\widehat{\boldsymbol{\beta}}^{WLS} = \left(\boldsymbol{x}^\top W \boldsymbol{x}\right)^{-1} \boldsymbol{x}^T W \boldsymbol{y} = \boldsymbol{\beta} + \left(\boldsymbol{x}^\top W \boldsymbol{x}\right)^{-1} \boldsymbol{x}^T W \mu$$

- **Remark**: Under Gauss-Markov assumptions 1-4, given W is a positive semidefinite matrix, $\widehat{\boldsymbol{\beta}}^{WLS}$ is unbiased and consistent as $n \longrightarrow \infty$
- The $\text{Var}(\widehat{\boldsymbol{\beta}}^{WLS})$ is in the "sandwich" form with

$$\text{Var}(\widehat{\boldsymbol{\beta}}^{WLS}) = \left(\boldsymbol{x}^\top W \boldsymbol{x}\right)^{-1} \boldsymbol{x}^\top W \Omega W^\top \boldsymbol{x} \left(\boldsymbol{x}^\top W \boldsymbol{x}\right)^{-1}$$

- If $W = I_n$, $\widehat{\boldsymbol{\beta}}^{WLS}$ becomes $\widehat{\boldsymbol{\beta}}^{OLS}$ and

$$\text{Var}(\widehat{\boldsymbol{\beta}}^{OLS}) = \left(\boldsymbol{x}^\top \boldsymbol{x}\right)^{-1} \boldsymbol{x}^\top \Omega \boldsymbol{x} \left(\boldsymbol{x}^\top \boldsymbol{x}\right)^{-1}$$

Review: Generalized least squares (GLS)

- Question: As all $\widehat{\beta}^{WLS}$ unbiased and consistent, among them, which one is the most efficient, i.e., has the lowest variance? In other words, which W should we use to obtain the BLUE estimator?
- Intuition: Assign lower weights to higher variance error terms and higher weights to lower variance error terms!
- Solution: Each observation should be given a weight proportional to the inverse of the variance of its error term. Using Ω^{-1} as the weighting matrix, we have

$$\text{Var}(\widehat{\beta}^{\text{GLS}}) = \left(\boldsymbol{x}^\top \Omega^{-1} \boldsymbol{x}\right)^{-1} \boldsymbol{x}^\top \Omega^{-1} \Omega \Omega^{-1} \boldsymbol{x}^\top \left(\boldsymbol{x}^\top \Omega^{-1} \boldsymbol{x}\right)^{-1}$$

$$= \boxed{\left(\boldsymbol{x}^\top \Omega^{-1} \boldsymbol{x}\right)^{-1}}$$

Review: A simple proof of the GLS's efficiency

- To show GLS is BLUE estimator, we only need to show, for all positive semidefinite weighting matrix W,

$$\boldsymbol{x}^{\top}\Omega^{-1}\boldsymbol{x} - \boldsymbol{x}^{\top}W\boldsymbol{x}\left(\boldsymbol{x}^{\top}W\Omega W^{\top}\boldsymbol{x}\right)^{-1}\boldsymbol{x}^{\top}W\boldsymbol{x}$$

is positive semidefinite.

Proof.

$$\begin{aligned}
&\boldsymbol{x}^{\top}\Omega^{-1}\boldsymbol{x} - \boldsymbol{x}^{\top}W\boldsymbol{x}\left(\boldsymbol{x}^{\top}W\Omega W^{\top}\boldsymbol{x}\right)^{-1}\boldsymbol{x}^{\top}W\boldsymbol{x}\\
&= \boldsymbol{x}^{\top}\left[\Omega^{-1} - W\boldsymbol{x}\left(\boldsymbol{x}^{\top}W\Omega W^{\top}\boldsymbol{x}\right)^{-1}\boldsymbol{x}^{\top}W\right]\boldsymbol{x}\\
&= \boldsymbol{x}^{\top}\Omega^{-1/2}\left[I_n - W\boldsymbol{x}\Omega^{1/2}\left(\boldsymbol{x}^{\top}W\Omega W^{\top}\boldsymbol{x}\right)^{-1}\Omega^{1/2}\boldsymbol{x}^{\top}W\right]\Omega^{-1/2}\boldsymbol{x}\\
&= \boldsymbol{x}^{\top}\Omega^{-1/2}\left[I_n - W\boldsymbol{x}\left(\boldsymbol{x}^{\top}WW^{\top}\boldsymbol{x}\right)^{-1}\boldsymbol{x}^{\top}W\right]\Omega^{-1/2}\boldsymbol{x}
\end{aligned}$$

Review: A simple proof of the GLS's efficiency

Proof.

- Let $\boldsymbol{x}^* = W\boldsymbol{x}$. We have

$$W\boldsymbol{x}\left(\boldsymbol{x}^\top WW^\top \boldsymbol{x}\right)^{-1}\boldsymbol{x}^\top W = \boldsymbol{x}^*(\boldsymbol{x}^{*\top}\boldsymbol{x}^*)^{-1}\boldsymbol{x}^{*\top} = P_{x^*}.$$

- P_{x^*} is an idempotent projection matrix!!! $M_{x^*} = I_n - P_{x^*}$ is also an idempotent projection matrix!
- This implies

$$\begin{aligned}&\boldsymbol{x}^\top\Omega^{-1/2}\left[I_n - W\boldsymbol{x}\left(\boldsymbol{x}^\top WW^\top\boldsymbol{x}\right)^{-1}\boldsymbol{x}^\top W\right]\Omega^{-1/2}\boldsymbol{x}\\ &= \boldsymbol{x}^\top\Omega^{-1/2}M_{x^*}\Omega^{-1/2}\boldsymbol{x} = \boldsymbol{x}^\top\Omega^{-1/2}M_{x^*}M_{x^*}\Omega^{-1/2}\boldsymbol{x} = \boldsymbol{c}^\top\boldsymbol{c},\end{aligned}$$

where $\boldsymbol{c} = M_{x^*}\Omega^{-1/2}\boldsymbol{x}$.

$\mathcal{Q}.\mathcal{E}.\mathcal{D}.$

Topic 3: GLS estimator for the random effect model

- Recall, we assume
  $\Omega = E(\boldsymbol{u_i}\boldsymbol{u_i}^\top) = E\left[(\boldsymbol{e}\alpha_i + \boldsymbol{\varepsilon_i})(\boldsymbol{e}\alpha_i + \boldsymbol{\varepsilon_i})^\top\right] = \sigma_\alpha^2 \boldsymbol{e}\boldsymbol{e}^\top + \sigma_\varepsilon^2 I_T.$

- Closely following our discussion in GLS review for the classical linear regression model, if the variance covariance matrix Ω is known, the GLS estimator of the γ for the random effect model (2) is

$$\widehat{\gamma}^{GLS} = \left(\sum_{i=1}^{n} \boldsymbol{X_i}^\top \Omega^{-1} \boldsymbol{X_i}\right)\left(\sum_{i=1}^{n} \boldsymbol{X_i}^\top \Omega^{-1} \boldsymbol{y_i}\right).$$

- Under Assumptions RE, $\widehat{\gamma}^{GLS}$ is BLUE estimator!

- **At home**: Derive $\widehat{\gamma}^{GLS}$ and demonstrate that it is BLUE estimator. We will show this in the next lecture in a nutshell.