Tutorial ~~6~~ 8: Panel Data Fixed and Random Effects Estimation

Hammad Shaikh

March 11, 2021

Fixed Effects (FE) Introduction

- $\overbrace{\text{Panel}}^{\text{Follow unit i over t}}$ regression: $\underbrace{Y_{it}}_{\substack{\text{Outcome} \\ \text{(i,t) index}}} = \beta_0 + \beta_1 \underbrace{X_{it}}_{\text{Treatment}} + \overbrace{\overbrace{\alpha_i}^{\text{unit FE}} + \underbrace{\delta_t}_{\text{time FE}}}^{\text{Panel data Components}} + \underbrace{\epsilon_{it}}_{\text{error}}$

- Period-level FE: Leverage within period variation across units

  ↳ Controls for time effects constant across units

- Unit-level FE: Leverage within unit over time variation

  ↳ Control for unit effects constant over time

Within Estimator (Fixed Effects Estimator)

- Fixed effect estimator assumes $Cov(\alpha_i, \epsilon_{it}) \neq 0$

  $\hookrightarrow Y_{it} = \beta_0 + \beta_1 X_{it} + \alpha_i + \varepsilon_{it} \Rightarrow \overline{Y_i} = \frac{1}{T}\sum_{t=1}^{T} Y_{it} = \beta_0 + \beta_1 \overline{X_i} + \alpha_i + \overline{\varepsilon_i}$

- Control for participant FE: regress $\underbrace{Y_{it} - \bar{Y}_i}_{\text{demean } Y}$ on $\underbrace{X_{it} - \bar{X}_i}_{\text{demean } X}$

  $\hookrightarrow Y_{it} - \overline{Y_i} = \beta_1 (X_{it} - \overline{X_i}) + (\varepsilon_{it} - \overline{\varepsilon_i}) \Rightarrow \hat{\beta}_1^{OLS}$ unbias if $Cov(X_{it} - \overline{X_i}, \varepsilon_{it} - \overline{\varepsilon_i}) = 0$

Demean leverages over time variation within unit

- Can alternatively include dummy variable for each unit

  $\hookrightarrow$ Estimate $Y_{it} = \beta_0 + \beta_1 X_{it} + \underbrace{D_2 + D_3 + \ldots + D_N}_{\text{N-1 dummies, omit unit 1,}} + \varepsilon_{it}$

  $D_j = \begin{cases} 1, & j \\ 0, & \text{not } j \end{cases}$

Random Effects (RE) Estimation

Strong assumption ⇒ RE less commonly used

- Random effects estimator assumes $Cov(\alpha_i, \epsilon_{it}) = 0$

  ↳ $Y_{it} = \beta_0 + \beta_1 X_{it} + \alpha_i + \varepsilon_{it}$

  ↳ Since $\alpha_i \perp \varepsilon_{it}$, do not need to eliminate α_i

* RE need to estimate $\hat{\theta}$ using $\hat{\sigma}_u^2$, $\hat{\sigma}_a^2$

- Regress $\underbrace{y_{it} - \theta \bar{y}_i}_{\text{Quasi-demean}}$ on $x_{it} - \theta \bar{x}_i, \theta = 1 - [\frac{\sigma_u^2}{\sigma_u^2 + T\sigma_a^2}]^{1/2}$, $\sigma_u^2 = Var(\varepsilon_{it})$, $\sigma_a^2 = Var(\alpha_i)$

  ↳ $Y_{it} - \theta \bar{Y}_i = \beta_0(1-\theta) + \beta_1(X_{it} - \theta \bar{X}_i) + (1-\theta)\alpha_i + (\varepsilon_{it} - \theta \bar{\varepsilon}_i)$

- RE estimation works with time-invariant variables X_i

  ↳ $Y_{it} = \beta_0 + \beta_1 X_{it}^1 + \beta_2 X_i^2 + \alpha_i + \varepsilon_{it}$

  ↳ Cannot use FE estimation because demean removes X_i^2

  ↳ Can use RE since quasi-demean will not remove X_i^2

↳ $\beta_2 (X_i - \theta \bar{X}_i) = \beta_2 (1-\theta) X_i$

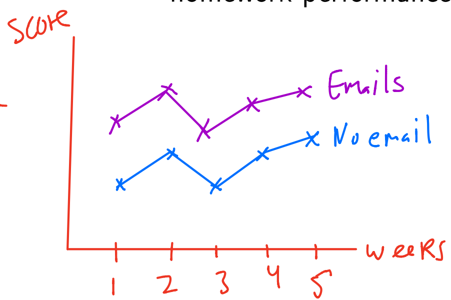
Panel data practice problem

Y_{iw} = score of student i in week w

$$\text{Email}_{iw} = \begin{cases} 1, & i \text{ gets reminder in week } w \\ 0, & \text{no reminder in week } w \end{cases}$$

Suppose you have panel data on students from a course with weekly online homework. At the start of each week, a randomly group of students receive an email reminder. What regression would you estimate to learn the impact of email reminder on homework performance?

* Panel reg. helps aggregate email effects across weeks



(1) Pool: $Y_{iw} = \beta_0 + \beta_1 \text{Email}_{iw} + \varepsilon_{iw}$

↳ Ignores panel data, but $\hat{\beta}_1$ unbias since $\text{Email}_{iw} \perp \varepsilon_{iw}$

(2) Week FE: $Y_{iw} = \beta_0 + \beta_1 \text{Email}_{iw} + \delta_w + \varepsilon_{iw}$

↳ Helps control for weekly-shocks like homework difficulty

3 Week FE + student FE:

$$Y_{iw} = \beta_0 + \beta_1 Email_{iw} + \delta_w + \alpha_i + \varepsilon_{iw}$$

↳ Helps control for student innate ability constant across weeks