Notes on Linear Panel Model

Max Leung

May 18, 2024

1 Fixed Effect Model

$$y_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta} + \alpha_i + \varepsilon_{it} \qquad \text{Level 1}$$

$$\begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}'_{i1} \\ \vdots \\ \boldsymbol{x}'_{iT} \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \alpha_i + \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iT} \end{pmatrix}$$

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \underbrace{(\boldsymbol{e}\alpha_i + \boldsymbol{\varepsilon}_i)}_{\boldsymbol{u}_i} \qquad \text{Level 2}$$

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_N \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_N \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{e} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{e} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_N \end{pmatrix}$$

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I}_N \otimes \boldsymbol{e})\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \qquad \text{Level 3}$$

where α_i is unobserved heterogeneity, $\boldsymbol{\varepsilon}_i$ is idiosyncratic error, $\boldsymbol{u}_i$ is composite error.

1.1 Assumption

1.1.1 Strong/strict exogeneity of regressors

For all t,

$$\mathbb{E}(\varepsilon_{it}|\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iT}) = 0$$

Equivalently,

$$\mathbb{E}(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i) = \boldsymbol{0}$$

1.2 OLS estimator is inconsistent and biased

The necessary condition for OLS estimator to be consistent is $\mathbb{E}(\boldsymbol{X}'_i\boldsymbol{u}_i) = \boldsymbol{0}$.

$$\begin{aligned}
\mathbb{E}(\boldsymbol{X}'_i\boldsymbol{u}_i) &= \mathbb{E}(\mathbb{E}(\boldsymbol{X}'_i\boldsymbol{u}_i|\boldsymbol{X}_i)) \\
&= \mathbb{E}(\boldsymbol{X}'_i\mathbb{E}(\boldsymbol{u}_i|\boldsymbol{X}_i)) \\
&= \mathbb{E}(\boldsymbol{X}'_i\mathbb{E}(\boldsymbol{e}\alpha_i + \boldsymbol{\varepsilon}_i|\boldsymbol{X}_i)) \\
&= \mathbb{E}(\boldsymbol{X}'_i\mathbb{E}(\boldsymbol{e}\alpha_i|\boldsymbol{X}_i) + \boldsymbol{X}'_i\underbrace{\mathbb{E}(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i)}_{\boldsymbol{0}}) && \text{because of strict exogeneity} \\
&= \mathbb{E}(\boldsymbol{X}'_i\boldsymbol{e}\mathbb{E}(\alpha_i|\boldsymbol{X}_i)) \\
&= \mathbb{E}(\boldsymbol{X}'_i\boldsymbol{e}\alpha_i)
\end{aligned}$$

$\mathbb{E}(\boldsymbol{X}'_i\boldsymbol{e}\alpha_i) \neq 0 \iff \mathbb{E}(\boldsymbol{X}'_i\boldsymbol{u}_i) \neq \boldsymbol{0}$. Thus, OLS estimator is inconsistent if $\mathbb{E}(\boldsymbol{X}'_i\boldsymbol{e}\alpha_i) \neq \boldsymbol{0}$.

The necessary condition for OLS estimator to be unbiased is $\mathbb{E}(\boldsymbol{u}_i|\boldsymbol{X}_i) = \boldsymbol{0}$. However, $\mathbb{E}(\boldsymbol{u}_i|\boldsymbol{X}_i) = \boldsymbol{0} \implies \mathbb{E}(\boldsymbol{X}'_i\boldsymbol{u}_i) = \boldsymbol{0}$ as $\mathbb{E}(\boldsymbol{X}'_i\boldsymbol{u}_i) = \mathbb{E}(\mathbb{E}(\boldsymbol{X}'_i\boldsymbol{u}_i|\boldsymbol{X}_i)) = \mathbb{E}(\boldsymbol{X}'_i\mathbb{E}(\boldsymbol{u}_i|\boldsymbol{X}_i)) = \mathbb{E}(\boldsymbol{X}'_i\boldsymbol{0}) = \boldsymbol{0}$. Thus, $\mathbb{E}(\boldsymbol{X}'_i\boldsymbol{u}_i) \neq \boldsymbol{0} \implies \mathbb{E}(\boldsymbol{u}_i|\boldsymbol{X}_i) \neq \boldsymbol{0}$. As a result, OLS estimator is biased if $\mathbb{E}(\boldsymbol{X}'_i\boldsymbol{e}\alpha_i) \neq \boldsymbol{0}$.

$\mathbb{E}(\alpha_i|\boldsymbol{X}_i) = 0 \implies \mathbb{E}(\boldsymbol{X}_i'\boldsymbol{e}\alpha_i) = \boldsymbol{0}$ as $\mathbb{E}(\boldsymbol{X}_i'\boldsymbol{e}\alpha_i) = \mathbb{E}(\mathbb{E}(\boldsymbol{X}_i'\boldsymbol{e}\alpha_i|\boldsymbol{X}_i)) = \mathbb{E}(\boldsymbol{X}_i'\boldsymbol{e}\mathbb{E}(\alpha_i|\boldsymbol{X}_i)) = \mathbb{E}(\boldsymbol{X}_i'\boldsymbol{e}\boldsymbol{0}) = \boldsymbol{0}$. Thus, $\mathbb{E}(\boldsymbol{X}_i'\boldsymbol{e}\alpha_i) \neq \boldsymbol{0} \implies \mathbb{E}(\alpha_i|\boldsymbol{X}_i) \neq 0$

OLS estimator of $\boldsymbol{\beta}$ is inconsistent and biased if α_i is correlated with $\boldsymbol{X}_i$ ($\boldsymbol{u}_i$ is also correlated with $\boldsymbol{X}_i$). This is called omitted variable bias. To tackle this, we simply eliminate α_i by using different methods.

1.3 Fixed Effect Estimator / Within Estimator

1.3.1 Demean operator

$$\boldsymbol{Q} = \boldsymbol{I}_T - T^{-1}\boldsymbol{e}\boldsymbol{e}'$$

This $\boldsymbol{Q}$ is symmetric and idempotent,

$$\begin{aligned}
\boldsymbol{Q}' &= (\boldsymbol{I}_T - T^{-1}\boldsymbol{e}\boldsymbol{e}')' \\
&= \boldsymbol{I}_T' - T^{-1}\boldsymbol{e}''\boldsymbol{e}' \\
&= \boldsymbol{I}_T - T^{-1}\boldsymbol{e}\boldsymbol{e}' = \boldsymbol{Q}
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{Q}\boldsymbol{Q}' &= \boldsymbol{Q}\boldsymbol{Q} \\
&= (\boldsymbol{I}_T - T^{-1}\boldsymbol{e}\boldsymbol{e}')(\boldsymbol{I}_T - T^{-1}\boldsymbol{e}\boldsymbol{e}') \\
&= \boldsymbol{I}_T\boldsymbol{I}_T - \boldsymbol{I}_T T^{-1}\boldsymbol{e}\boldsymbol{e}' - T^{-1}\boldsymbol{e}\boldsymbol{e}'\boldsymbol{I}_T + T^{-1}\boldsymbol{e}\boldsymbol{e}'T^{-1}\boldsymbol{e}\boldsymbol{e}' \\
&= \boldsymbol{I}_T - T^{-1}\boldsymbol{e}\boldsymbol{e}' - T^{-1}\boldsymbol{e}\boldsymbol{e}' + T^{-2}\boldsymbol{e}\boldsymbol{e}'\boldsymbol{e}\boldsymbol{e}' \\
&= \boldsymbol{I}_T - 2T^{-1}\boldsymbol{e}\boldsymbol{e}' + T^{-2}\boldsymbol{e}T\boldsymbol{e}' \\
&= \boldsymbol{I}_T - 2T^{-1}\boldsymbol{e}\boldsymbol{e}' + T^{-1}\boldsymbol{e}\boldsymbol{e}' \\
&= \boldsymbol{I}_T - T^{-1}\boldsymbol{e}\boldsymbol{e}' = \boldsymbol{Q}
\end{aligned}$$

1.3.2 Demean transformed model

$$\begin{aligned}
\boldsymbol{Q}\boldsymbol{y}_i &= \boldsymbol{Q}(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{e}\alpha_i + \boldsymbol{\varepsilon}_i) \\
&= \boldsymbol{Q}\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Q}\boldsymbol{e}\alpha_i + \boldsymbol{Q}\boldsymbol{\varepsilon}_i \\
&= \boldsymbol{Q}\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{0}\alpha_i + \boldsymbol{Q}\boldsymbol{\varepsilon}_i \\
&= \boldsymbol{Q}\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Q}\boldsymbol{\varepsilon}_i
\end{aligned} \qquad \text{Level 2}$$

It is because

$$\begin{aligned}
\boldsymbol{Q}\boldsymbol{e} &= (\boldsymbol{I}_T - T^{-1}\boldsymbol{e}\boldsymbol{e}')\boldsymbol{e} \\
&= \boldsymbol{I}_T\boldsymbol{e} - T^{-1}\boldsymbol{e}\boldsymbol{e}'\boldsymbol{e} \\
&= \boldsymbol{e} - T^{-1}\boldsymbol{e}T \\
&= \boldsymbol{e} - \boldsymbol{e} = \boldsymbol{0}
\end{aligned}$$

It can be written as $\boldsymbol{y}_i - \boldsymbol{e}\bar{y}_i = (\boldsymbol{X}_i - \boldsymbol{e}\bar{\boldsymbol{x}}_i')\boldsymbol{\beta} + (\boldsymbol{\varepsilon}_i - \boldsymbol{e}\bar{\varepsilon}_i)$ because

$$\begin{aligned}
\boldsymbol{Q}\boldsymbol{X}_i &= (\boldsymbol{I}_T - T^{-1}\boldsymbol{e}\boldsymbol{e}')\boldsymbol{X}_i \\
&= \boldsymbol{I}_T\boldsymbol{X}_i - T^{-1}\boldsymbol{e}\boldsymbol{e}'\boldsymbol{X}_i \\
&= \boldsymbol{X}_i - T^{-1}\boldsymbol{e}\begin{pmatrix}1 & \cdots & 1\end{pmatrix}\begin{pmatrix}\boldsymbol{x}_{i1}' \\ \vdots \\ \boldsymbol{x}_{iT}'\end{pmatrix} \\
&= \boldsymbol{X}_i - \boldsymbol{e}T^{-1}\sum_{t=1}^{T}\boldsymbol{x}_{it}' \\
&= \boldsymbol{X}_i - \boldsymbol{e}\bar{\boldsymbol{x}}_i'
\end{aligned}$$

$$
\begin{aligned}
\boldsymbol{Q}\boldsymbol{y}_i &= (\boldsymbol{I}_T - T^{-1}\boldsymbol{e}\boldsymbol{e}')\boldsymbol{y}_i \\
&= \boldsymbol{I}_T\boldsymbol{y}_i - T^{-1}\boldsymbol{e}\boldsymbol{e}'\boldsymbol{y}_i \\
&= \boldsymbol{y}_i - T^{-1}\boldsymbol{e}\begin{pmatrix}1 & \cdots & 1\end{pmatrix}\begin{pmatrix}y_{i1}\\ \vdots \\ y_{iT}\end{pmatrix} \\
&= \boldsymbol{y}_i - \boldsymbol{e}T^{-1}\sum_{t=1}^{T} y_{it} \\
&= \boldsymbol{y}_i - \boldsymbol{e}\bar{y}_i
\end{aligned}
$$

$\boldsymbol{y}_i - \boldsymbol{e}\bar{y}_i = (\boldsymbol{X}_i - \boldsymbol{e}\bar{\boldsymbol{x}}_i')\boldsymbol{\beta} + (\boldsymbol{\varepsilon}_i - \boldsymbol{e}\bar{\varepsilon}_i)$ can be written as

$$
\begin{aligned}
\begin{pmatrix}y_{i1}\\ \vdots \\ y_{iT}\end{pmatrix} - \begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}\bar{y}_i &= (\begin{pmatrix}\boldsymbol{x}_{i1}'\\ \vdots \\ \boldsymbol{x}_{iT}'\end{pmatrix} - \begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}\bar{\boldsymbol{x}}_i')\boldsymbol{\beta} + (\begin{pmatrix}\varepsilon_{i1}\\ \vdots \\ \varepsilon_{iT}\end{pmatrix} - \begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}\bar{\varepsilon}_i) \\
\begin{pmatrix}y_{i1}-\bar{y}_i\\ \vdots \\ y_{iT}-\bar{y}_i\end{pmatrix} &= \begin{pmatrix}\boldsymbol{x}_{i1}'-\bar{\boldsymbol{x}}_i'\\ \vdots \\ \boldsymbol{x}_{iT}'-\bar{\boldsymbol{x}}_i'\end{pmatrix}\boldsymbol{\beta} + \begin{pmatrix}\varepsilon_{i1}-\bar{\varepsilon}_i\\ \vdots \\ \varepsilon_{iT}-\bar{\varepsilon}_i\end{pmatrix} \\
\begin{pmatrix}y_{i1}-\bar{y}_i\\ \vdots \\ y_{iT}-\bar{y}_i\end{pmatrix} &= \begin{pmatrix}(\boldsymbol{x}_{i1}-\bar{\boldsymbol{x}}_i)'\\ \vdots \\ (\boldsymbol{x}_{iT}-\bar{\boldsymbol{x}}_i)'\end{pmatrix}\boldsymbol{\beta} + \begin{pmatrix}\varepsilon_{i1}-\bar{\varepsilon}_i\\ \vdots \\ \varepsilon_{iT}-\bar{\varepsilon}_i\end{pmatrix} \\
y_{it} - \bar{y}_i &= (\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)'\boldsymbol{\beta} + (\varepsilon_{it}-\bar{\varepsilon}_i) && \text{Level 1}
\end{aligned}
$$

1.3.3 OLS estimator of the demean transformed model / Fixed Effect (FE) estimator

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}_{within}^{ols} &= [\sum_{i=1}^{N}(\boldsymbol{Q}\boldsymbol{X}_i)'\boldsymbol{Q}\boldsymbol{X}_i]^{-1}\sum_{i=1}^{N}(\boldsymbol{Q}\boldsymbol{X}_i)'\boldsymbol{Q}\boldsymbol{y}_i && \text{Level 2} \\
&= [\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)']^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(y_{it}-\bar{y}_i) && \text{Level 1}
\end{aligned}
$$

It is because

$$
\begin{aligned}
(\boldsymbol{Q}\boldsymbol{X}_i)'\boldsymbol{Q}\boldsymbol{X}_i &= (\boldsymbol{X}_i - \boldsymbol{e}\bar{\boldsymbol{x}}_i')'(\boldsymbol{X}_i - \boldsymbol{e}\bar{\boldsymbol{x}}_i') \\
&= (\begin{pmatrix}\boldsymbol{x}_{i1}'\\ \vdots \\ \boldsymbol{x}_{iT}'\end{pmatrix} - \begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}\bar{\boldsymbol{x}}_i')'(\begin{pmatrix}\boldsymbol{x}_{i1}'\\ \vdots \\ \boldsymbol{x}_{iT}'\end{pmatrix} - \begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}\bar{\boldsymbol{x}}_i') \\
&= \begin{pmatrix}\boldsymbol{x}_{i1}'-\bar{\boldsymbol{x}}_i'\\ \vdots \\ \boldsymbol{x}_{iT}'-\bar{\boldsymbol{x}}_i'\end{pmatrix}'\begin{pmatrix}\boldsymbol{x}_{i1}'-\bar{\boldsymbol{x}}_i'\\ \vdots \\ \boldsymbol{x}_{iT}'-\bar{\boldsymbol{x}}_i'\end{pmatrix} \\
&= \begin{pmatrix}(\boldsymbol{x}_{i1}-\bar{\boldsymbol{x}}_i)'\\ \vdots \\ (\boldsymbol{x}_{iT}-\bar{\boldsymbol{x}}_i)'\end{pmatrix}'\begin{pmatrix}(\boldsymbol{x}_{i1}-\bar{\boldsymbol{x}}_i)'\\ \vdots \\ (\boldsymbol{x}_{iT}-\bar{\boldsymbol{x}}_i)'\end{pmatrix} \\
&= \begin{pmatrix}(\boldsymbol{x}_{i1}-\bar{\boldsymbol{x}}_i) & \cdots & (\boldsymbol{x}_{iT}-\bar{\boldsymbol{x}}_i)\end{pmatrix}\begin{pmatrix}(\boldsymbol{x}_{i1}-\bar{\boldsymbol{x}}_i)'\\ \vdots \\ (\boldsymbol{x}_{iT}-\bar{\boldsymbol{x}}_i)'\end{pmatrix} \\
&= \sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)'
\end{aligned}
$$

$$
\begin{aligned}
(\boldsymbol{QX}_i)'\boldsymbol{Qy}_i &= (\boldsymbol{X}_i - \boldsymbol{e}\bar{\boldsymbol{x}}_i')'(\boldsymbol{y}_i - \boldsymbol{e}\bar{y}_i) \\
&= (\begin{pmatrix} \boldsymbol{x}_{i1}' \\ \vdots \\ \boldsymbol{x}_{iT}' \end{pmatrix} - \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \bar{\boldsymbol{x}}_i')'(\begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix} - \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \bar{y}_i) \\
&= \begin{pmatrix} \boldsymbol{x}_{i1}' - \bar{\boldsymbol{x}}_i' \\ \vdots \\ \boldsymbol{x}_{iT}' - \bar{\boldsymbol{x}}_i' \end{pmatrix}' \begin{pmatrix} y_{i1} - \bar{y}_i \\ \vdots \\ y_{iT} - \bar{y}_i \end{pmatrix} \\
&= \begin{pmatrix} (\boldsymbol{x}_{i1} - \bar{\boldsymbol{x}}_i) & \cdots & (\boldsymbol{x}_{iT} - \bar{\boldsymbol{x}}_i) \end{pmatrix} \begin{pmatrix} y_{i1} - \bar{y}_i \\ \vdots \\ y_{iT} - \bar{y}_i \end{pmatrix} \\
&= \sum_{t=1}^{T} (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)(y_{it} - \bar{y}_i)
\end{aligned}
$$

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}_{within}^{ols} &= [\sum_{i=1}^{N} (\boldsymbol{QX}_i)'\boldsymbol{QX}_i]^{-1} \sum_{i=1}^{N} (\boldsymbol{QX}_i)'\boldsymbol{Qy}_i && \text{Level 2} \\
&= [\sum_{i=1}^{N} (\boldsymbol{X}_i'\boldsymbol{X}_i - \bar{\boldsymbol{x}}_i T \bar{\boldsymbol{x}}_i')]^{-1} \sum_{i=1}^{N} (\boldsymbol{X}_i'\boldsymbol{y}_i - \bar{\boldsymbol{x}}_i T \bar{y}_i) \\
&= [\sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{X}_i - T \sum_{i=1}^{N} \bar{\boldsymbol{x}}_i \bar{\boldsymbol{x}}_i']^{-1} (\sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{y}_i - T \sum_{i=1}^{N} \bar{\boldsymbol{x}}_i \bar{y}_i) \\
&= [\begin{pmatrix} \boldsymbol{X}_1' & \cdots & \boldsymbol{X}_N' \end{pmatrix} \begin{pmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_N \end{pmatrix} - T \begin{pmatrix} \bar{\boldsymbol{x}}_1 & \cdots & \bar{\boldsymbol{x}}_N \end{pmatrix} \begin{pmatrix} \bar{\boldsymbol{x}}_1' \\ \vdots \\ \bar{\boldsymbol{x}}_N' \end{pmatrix}]^{-1} (\begin{pmatrix} \boldsymbol{X}_1' & \cdots & \boldsymbol{X}_N' \end{pmatrix} \begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_N \end{pmatrix} - T \begin{pmatrix} \bar{\boldsymbol{x}}_1 & \cdots & \bar{\boldsymbol{x}}_N \end{pmatrix} \begin{pmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_N \end{pmatrix}) \\
&= [\boldsymbol{X}'\boldsymbol{X} - T \bar{\boldsymbol{X}}'\bar{\boldsymbol{X}}]^{-1} (\boldsymbol{X}'\boldsymbol{y} - T \bar{\boldsymbol{X}}'\bar{\boldsymbol{y}}) && \text{Level 3}
\end{aligned}
$$

It is because

$$
\begin{aligned}
(\boldsymbol{QX}_i)'\boldsymbol{QX}_i &= (\boldsymbol{X}_i - \boldsymbol{e}\bar{\boldsymbol{x}}_i')'(\boldsymbol{X}_i - \boldsymbol{e}\bar{\boldsymbol{x}}_i') \\
&= (\boldsymbol{X}_i' - \bar{\boldsymbol{x}}_i'' \boldsymbol{e}')(\boldsymbol{X}_i - \boldsymbol{e}\bar{\boldsymbol{x}}_i') \\
&= \boldsymbol{X}_i'\boldsymbol{X}_i - \boldsymbol{X}_i'\boldsymbol{e}\bar{\boldsymbol{x}}_i' - \bar{\boldsymbol{x}}_i \boldsymbol{e}'\boldsymbol{X}_i + \bar{\boldsymbol{x}}_i \boldsymbol{e}'\boldsymbol{e}\bar{\boldsymbol{x}}_i' \\
&= \boldsymbol{X}_i'\boldsymbol{X}_i - \boldsymbol{X}_i'\boldsymbol{e}\bar{\boldsymbol{x}}_i' - \bar{\boldsymbol{x}}_i \boldsymbol{e}'\boldsymbol{X}_i + \bar{\boldsymbol{x}}_i T \bar{\boldsymbol{x}}_i' \\
&= \boldsymbol{X}_i'\boldsymbol{X}_i - (\boldsymbol{e}'\boldsymbol{X}_i)'\bar{\boldsymbol{x}}_i' - \bar{\boldsymbol{x}}_i \boldsymbol{e}'\boldsymbol{X}_i + \bar{\boldsymbol{x}}_i T \bar{\boldsymbol{x}}_i' \\
&= \boldsymbol{X}_i'\boldsymbol{X}_i - (\sum_{t=1}^{T} \boldsymbol{x}_{it}')'\bar{\boldsymbol{x}}_i' - \bar{\boldsymbol{x}}_i \sum_{t=1}^{T} \boldsymbol{x}_{it}' + \bar{\boldsymbol{x}}_i T \bar{\boldsymbol{x}}_i' \\
&= \boldsymbol{X}_i'\boldsymbol{X}_i - (\sum_{t=1}^{T} \boldsymbol{x}_{it}/T) T \bar{\boldsymbol{x}}_i' - \bar{\boldsymbol{x}}_i T \sum_{t=1}^{T} \boldsymbol{x}_{it}'/T + \bar{\boldsymbol{x}}_i T \bar{\boldsymbol{x}}_i' \\
&= \boldsymbol{X}_i'\boldsymbol{X}_i - \bar{\boldsymbol{x}}_i T \bar{\boldsymbol{x}}_i' - \bar{\boldsymbol{x}}_i T \bar{\boldsymbol{x}}_i' + \bar{\boldsymbol{x}}_i T \bar{\boldsymbol{x}}_i' \\
&= \boldsymbol{X}_i'\boldsymbol{X}_i - \bar{\boldsymbol{x}}_i T \bar{\boldsymbol{x}}_i'
\end{aligned}
$$

$$
\begin{aligned}
(\boldsymbol{QX}_i)'\boldsymbol{Qy}_i &= (\boldsymbol{X}_i - \boldsymbol{e}\bar{\boldsymbol{x}}_i')'(\boldsymbol{y}_i - \boldsymbol{e}\bar{y}_i) \\
&= \boldsymbol{X}_i'\boldsymbol{y}_i - \boldsymbol{X}_i'\boldsymbol{e}\bar{y}_i - \bar{\boldsymbol{x}}_i \boldsymbol{e}'\boldsymbol{y}_i + \bar{\boldsymbol{x}}_i \boldsymbol{e}'\boldsymbol{e}\bar{y}_i \\
&= \boldsymbol{X}_i'\boldsymbol{y}_i - (\boldsymbol{e}'\boldsymbol{X}_i)'\bar{y}_i - \bar{\boldsymbol{x}}_i \boldsymbol{e}'\boldsymbol{y}_i + \bar{\boldsymbol{x}}_i T \bar{y}_i \\
&= \boldsymbol{X}_i'\boldsymbol{y}_i - (\sum_{t=1}^{T} \boldsymbol{x}_{it}')'\bar{y}_i - \bar{\boldsymbol{x}}_i \sum_{t=1}^{T} y_{it} + \bar{\boldsymbol{x}}_i T \bar{y}_i \\
&= \boldsymbol{X}_i'\boldsymbol{y}_i - (\sum_{t=1}^{T} \boldsymbol{x}_{it}/T) T \bar{y}_i - \bar{\boldsymbol{x}}_i T \sum_{t=1}^{T} y_{it}/T + \bar{\boldsymbol{x}}_i T \bar{y}_i \\
&= \boldsymbol{X}_i'\boldsymbol{y}_i - \bar{\boldsymbol{x}}_i T \bar{y}_i - \bar{\boldsymbol{x}}_i T \bar{y}_i + \bar{\boldsymbol{x}}_i T \bar{y}_i \\
&= \boldsymbol{X}_i'\boldsymbol{y}_i - \bar{\boldsymbol{x}}_i T \bar{y}_i
\end{aligned}
$$

1.3.4 The necessary condition for consistency and unbiasedness

The necessary condition for FE estimator (OLS estimator of the demean transformed model) to be consistent is $\mathbb{E}((\boldsymbol{Q}\boldsymbol{X}_i)'\boldsymbol{Q}\boldsymbol{\varepsilon}_i) = \mathbf{0}$.

$$\begin{aligned}
\mathbb{E}((\boldsymbol{Q}\boldsymbol{X}_i)'\boldsymbol{Q}\boldsymbol{\varepsilon}_i) &= \mathbb{E}(\boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{\varepsilon}_i) \\
&= \mathbb{E}(\boldsymbol{X}_i'\boldsymbol{Q}\boldsymbol{\varepsilon}_i) && \text{as } \boldsymbol{Q} \text{ is idempotent and symmetric} \\
&= \mathbb{E}(\mathbb{E}(\boldsymbol{X}_i'\boldsymbol{Q}\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i)) \\
&= \mathbb{E}(\boldsymbol{X}_i'\boldsymbol{Q}\underbrace{\mathbb{E}(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i)}_{\mathbf{0}}) && \text{because of strict exogeneity} \\
&= \mathbf{0}
\end{aligned}$$

Thus, FE estimator satisfies the necessary condition for consistency given strict exogeneity assumption. Indeed, strict exogeneity is stronger than what is required. To see this, first note that for any t,

$$\mathbb{E}(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i) = \mathbb{E}(\varepsilon_{it}|\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{iT}) = \mathbf{0} \implies \mathbb{E}(\boldsymbol{x}_{is}\varepsilon_{it}) = \mathbf{0} \qquad \text{for all } s$$

It is because for any t and s,

$$\begin{aligned}
\mathbb{E}(\boldsymbol{x}_{is}\varepsilon_{it}) &= \mathbb{E}(\mathbb{E}(\boldsymbol{x}_{is}\varepsilon_{it}|\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{iT})) \\
&= \mathbb{E}(\boldsymbol{x}_{is}\underbrace{\mathbb{E}(\varepsilon_{it}|\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{iT})}_{0}) \\
&= \mathbf{0}
\end{aligned}$$

The necessary condition for FE estimator to be consistent can also be written as $\mathbb{E}((\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)(\varepsilon_{it} - \bar{\varepsilon}_i)) = \mathbf{0}$.

$$\mathbb{E}((\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)(\varepsilon_{it} - \bar{\varepsilon}_i)) = \mathbb{E}(\boldsymbol{x}_{it}\varepsilon_{it}) - \mathbb{E}(\boldsymbol{x}_{it}\bar{\varepsilon}_i) - \mathbb{E}(\bar{\boldsymbol{x}}_i\varepsilon_{it}) + \mathbb{E}(\bar{\boldsymbol{x}}_i\bar{\varepsilon}_i) = \mathbf{0}$$

It is because $\mathbb{E}(\boldsymbol{x}_{is}\varepsilon_{it}) = \mathbf{0}$ for any t and s implies

$$\begin{aligned}
\mathbb{E}(\boldsymbol{x}_{it}\varepsilon_{it}) &= \mathbf{0} \\
\mathbb{E}(\boldsymbol{x}_{it}\bar{\varepsilon}_i) &= \mathbb{E}(\boldsymbol{x}_{it}T^{-1}\sum_{s=1}^{T}\varepsilon_{is}) = T^{-1}\sum_{s=1}^{T}\underbrace{\mathbb{E}(\boldsymbol{x}_{it}\varepsilon_{is})}_{\mathbf{0}} = \mathbf{0} \\
\mathbb{E}(\bar{\boldsymbol{x}}_i\varepsilon_{it}) &= \mathbb{E}(T^{-1}\sum_{s=1}^{T}\boldsymbol{x}_{is}\varepsilon_{it}) = T^{-1}\sum_{s=1}^{T}\underbrace{\mathbb{E}(\boldsymbol{x}_{is}\varepsilon_{it})}_{\mathbf{0}} = \mathbf{0} \\
\mathbb{E}(\bar{\boldsymbol{x}}_i\bar{\varepsilon}_i) &= \mathbb{E}(T^{-1}\sum_{s=1}^{T}\boldsymbol{x}_{is}T^{-1}\sum_{t=1}^{T}\varepsilon_{it}) = T^{-2}\sum_{s=1}^{T}\sum_{t=1}^{T}\underbrace{\mathbb{E}(\boldsymbol{x}_{is}\varepsilon_{it})}_{\mathbf{0}} = \mathbf{0}
\end{aligned}$$

Thus, the weaker assumption $\mathbb{E}(\boldsymbol{x}_{is}\varepsilon_{it}) = \mathbf{0}$ for any t and s is sufficient for $\mathbb{E}((\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)(\varepsilon_{it} - \bar{\varepsilon}_i)) = \mathbf{0}$

The necessary condition for FE estimator to be unbiased is $\mathbb{E}(\boldsymbol{Q}\boldsymbol{\varepsilon}_i|\boldsymbol{Q}\boldsymbol{X}_i) = \mathbf{0}$.

$$\begin{aligned}
\mathbb{E}(\boldsymbol{Q}\boldsymbol{\varepsilon}_i|\boldsymbol{Q}\boldsymbol{X}_i) &= \boldsymbol{Q}\underbrace{\mathbb{E}(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i)}_{\mathbf{0}} && \text{as } \boldsymbol{Q} \text{ is constant and strict exogeneity} \\
&= \mathbf{0}
\end{aligned}$$

1.3.5 Conditional variance of $\widehat{\boldsymbol{\beta}}^{ols}_{within}$

Given independence of i,

$$\begin{aligned}
Var(\widehat{\boldsymbol{\beta}}^{ols}_{within}|\boldsymbol{X}_i) &= Var([\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{X}_i]^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{y}_i|\boldsymbol{X}_i) \\
&= [\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{X}_i]^{-1}Var(\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{y}_i|\boldsymbol{X}_i)[\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{X}_i]^{-1'} \\
&= [\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{X}_i]^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}'Var(\boldsymbol{Q}\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i)\boldsymbol{Q}\boldsymbol{X}_i[\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{X}_i]^{-1}
\end{aligned}$$

It is because

$$
\begin{aligned}
Var(\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{y}_i|\boldsymbol{X}_i) &= \sum_{i=1}^{N} Var(\boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{y}_i|\boldsymbol{X}_i)\\
&= \sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}' Var(\boldsymbol{Q}\boldsymbol{y}_i|\boldsymbol{X}_i)(\boldsymbol{X}_i'\boldsymbol{Q}')'\\
&= \sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}' Var(\boldsymbol{Q}\boldsymbol{X}_i\boldsymbol{\beta}+\boldsymbol{Q}\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i)\boldsymbol{Q}''\boldsymbol{X}_i''\\
&= \sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}' Var(\boldsymbol{Q}\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i)\boldsymbol{Q}\boldsymbol{X}_i
\end{aligned}
$$

1.3.6 $Var(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i)=\sigma_\varepsilon^2\boldsymbol{I}_T$

If ε_{it} is homoskedasticity and serially uncorrelated across t i.e., $Var(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i)=\sigma_\varepsilon^2\boldsymbol{I}_T$ (further assume independence of i and strict exogeneity), we have $\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i \sim iid\ [\boldsymbol{0},\sigma_\varepsilon^2\boldsymbol{I}_T]$.

$$
Var(\boldsymbol{Q}\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i) = \boldsymbol{Q}Var(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i)\boldsymbol{Q}' = \boldsymbol{Q}\sigma_\varepsilon^2\boldsymbol{I}_T\boldsymbol{Q}' = \sigma_\varepsilon^2\boldsymbol{Q}\boldsymbol{Q}' = \sigma_\varepsilon^2\boldsymbol{Q} = \sigma_\varepsilon^2(\boldsymbol{I}_T - T^{-1}\boldsymbol{e}\boldsymbol{e}') = \sigma_\varepsilon^2\begin{pmatrix} 1-\frac{1}{T} & -\frac{1}{T} & \cdots & -\frac{1}{T}\\ -\frac{1}{T} & 1-\frac{1}{T} & \cdots & -\frac{1}{T}\\ \vdots & \vdots & \vdots & \vdots\\ -\frac{1}{T} & -\frac{1}{T} & \cdots & 1-\frac{1}{T}\end{pmatrix}.
$$

Thus, $\boldsymbol{Q}\boldsymbol{\varepsilon}_i$ is homoskedasticity but negatively serially correlated. For any t,

$$
\begin{aligned}
Var(\varepsilon_{it}-\bar{\varepsilon}_i) = \sigma_\varepsilon^2(1-\frac{1}{T}) \iff \sigma_\varepsilon^2 &= \frac{T}{T-1}Var(\varepsilon_{it}-\bar{\varepsilon}_i)\\
\hat{\sigma}_\varepsilon^2 &= \frac{T}{T-1}\widehat{Var}(\widehat{\varepsilon_{it}-\bar{\varepsilon}_i})\\
&= \frac{T}{T-1}\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-\bar{y}_i-(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)'\widehat{\boldsymbol{\beta}}_{within}^{ols})^2}{NT-(K+N)}\\
&= \frac{T}{T-1}\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-\boldsymbol{x}_{it}'\widehat{\boldsymbol{\beta}}_{within}^{ols}-\overbrace{(\bar{y}_i-\bar{\boldsymbol{x}}_i'\widehat{\boldsymbol{\beta}}_{within}^{ols})}^{\widehat{\alpha}_i})^2}{\underbrace{NT-(K+N)}_{N(T-1)-K}} \qquad \frac{T}{T-1}\approx 1 \text{ if } T \text{ is large.}
\end{aligned}
$$

$$
\begin{aligned}
Var(\widehat{\boldsymbol{\beta}}_{within}^{ols}|\boldsymbol{X}_i) &= [\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{X}_i]^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}'\sigma_\varepsilon^2\underbrace{\boldsymbol{Q}\boldsymbol{Q}}_{\boldsymbol{Q}}\boldsymbol{X}_i[\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{X}_i]^{-1}\\
&= \sigma_\varepsilon^2[\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{X}_i]^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{X}_i[\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{X}_i]^{-1}\\
&= \sigma_\varepsilon^2\boldsymbol{I}_T[\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{X}_i]^{-1}\\
&= \sigma_\varepsilon^2[\sum_{i=1}^{N}(\boldsymbol{Q}\boldsymbol{X}_i)'\boldsymbol{Q}\boldsymbol{X}_i]^{-1} && \text{Level 2}\\
&= \sigma_\varepsilon^2[\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)']^{-1} && \text{Level 1}
\end{aligned}
$$

1.3.7 $Var(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i) = \boldsymbol{\Omega}_i$

We have $\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i \sim inid\ [\boldsymbol{0}, \boldsymbol{\Omega}_i]$.

$$\begin{aligned}
Var(\widehat{\boldsymbol{\beta}}^{ols}_{within}|\boldsymbol{X}_i) &= [\sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{X}_i]^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{Q}'\mathbb{E}[(\boldsymbol{Q}\boldsymbol{\varepsilon}_i - \mathbb{E}(\boldsymbol{Q}\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i))(\boldsymbol{Q}\boldsymbol{\varepsilon}_i - \mathbb{E}(\boldsymbol{Q}\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i))'|\boldsymbol{X}_i]\boldsymbol{Q}\boldsymbol{X}_i[\sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{X}_i]^{-1} \\
&= [\sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{X}_i]^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{Q}'\mathbb{E}[(\boldsymbol{Q}\boldsymbol{\varepsilon}_i - \boldsymbol{Q}\mathbb{E}(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i))(\boldsymbol{Q}\boldsymbol{\varepsilon}_i - \boldsymbol{Q}\mathbb{E}(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i))'|\boldsymbol{X}_i]\boldsymbol{Q}\boldsymbol{X}_i[\sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{X}_i]^{-1} \\
&= [\sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{X}_i]^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{Q}'\mathbb{E}[(\boldsymbol{Q}\boldsymbol{\varepsilon}_i - \boldsymbol{Q}\boldsymbol{0})(\boldsymbol{Q}\boldsymbol{\varepsilon}_i - \boldsymbol{Q}\boldsymbol{0})'|\boldsymbol{X}_i]\boldsymbol{Q}\boldsymbol{X}_i[\sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{X}_i]^{-1} \\
&= [\sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{X}_i]^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{Q}'\mathbb{E}[\boldsymbol{Q}\boldsymbol{\varepsilon}_i(\boldsymbol{Q}\boldsymbol{\varepsilon}_i)'|\boldsymbol{X}_i]\boldsymbol{Q}\boldsymbol{X}_i[\sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{X}_i]^{-1} \\
&= [\sum_{i=1}^{N} (\boldsymbol{Q}\boldsymbol{X}_i)'\boldsymbol{Q}\boldsymbol{X}_i]^{-1} \sum_{i=1}^{N} (\boldsymbol{Q}\boldsymbol{X}_i)'\mathbb{E}[\boldsymbol{Q}\boldsymbol{\varepsilon}_i(\boldsymbol{Q}\boldsymbol{\varepsilon}_i)'|\boldsymbol{X}_i]\boldsymbol{Q}\boldsymbol{X}_i[\sum_{i=1}^{N} (\boldsymbol{Q}\boldsymbol{X}_i)'\boldsymbol{Q}\boldsymbol{X}_i]^{-1} \\
&= [\sum_{i=1}^{N}\sum_{t=1}^{T} (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)']^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)\mathbb{E}[\dot{\varepsilon}_{it}\dot{\varepsilon}_{is}|\boldsymbol{X}_i](\boldsymbol{x}_{is} - \bar{\boldsymbol{x}}_i)'[\sum_{i=1}^{N}\sum_{t=1}^{T} (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)']^{-1}
\end{aligned}$$

It is because

$$\begin{aligned}
\sum_{i=1}^{N} (\boldsymbol{Q}\boldsymbol{X}_i)'\mathbb{E}[\boldsymbol{Q}\boldsymbol{\varepsilon}_i(\boldsymbol{Q}\boldsymbol{\varepsilon}_i)'|\boldsymbol{X}_i]\boldsymbol{Q}\boldsymbol{X}_i &= \sum_{i=1}^{N} (\boldsymbol{Q}\boldsymbol{X}_i)'\mathbb{E}[\dot{\boldsymbol{\varepsilon}}_i\dot{\boldsymbol{\varepsilon}}_i'|\boldsymbol{X}_i]\boldsymbol{Q}\boldsymbol{X}_i \\
&= \sum_{i=1}^{N} \begin{pmatrix} (\boldsymbol{x}_{i1} - \bar{\boldsymbol{x}}_i)' \\ \vdots \\ (\boldsymbol{x}_{iT} - \bar{\boldsymbol{x}}_i)' \end{pmatrix}' \begin{pmatrix} \mathbb{E}[\dot{\varepsilon}_{i1}^2|\boldsymbol{X}_i] & \cdots & \mathbb{E}[\dot{\varepsilon}_{i1}\dot{\varepsilon}_{iT}|\boldsymbol{X}_i] \\ \vdots & \ddots & \vdots \\ \mathbb{E}[\dot{\varepsilon}_{iT}\dot{\varepsilon}_{i1}|\boldsymbol{X}_i] & \cdots & \mathbb{E}[\dot{\varepsilon}_{iT}^2|\boldsymbol{X}_i] \end{pmatrix} \begin{pmatrix} (\boldsymbol{x}_{i1} - \bar{\boldsymbol{x}}_i)' \\ \vdots \\ (\boldsymbol{x}_{iT} - \bar{\boldsymbol{x}}_i)' \end{pmatrix} \\
&= \sum_{i=1}^{N} \begin{pmatrix} (\boldsymbol{x}_{i1} - \bar{\boldsymbol{x}}_i) & (\boldsymbol{x}_{iT} - \bar{\boldsymbol{x}}_i) \end{pmatrix} \begin{pmatrix} \mathbb{E}[\dot{\varepsilon}_{i1}^2|\boldsymbol{X}_i] & \cdots & \mathbb{E}[\dot{\varepsilon}_{i1}\dot{\varepsilon}_{iT}|\boldsymbol{X}_i] \\ \vdots & \ddots & \vdots \\ \mathbb{E}[\dot{\varepsilon}_{iT}\dot{\varepsilon}_{i1}|\boldsymbol{X}_i] & \cdots & \mathbb{E}[\dot{\varepsilon}_{iT}^2|\boldsymbol{X}_i] \end{pmatrix} \begin{pmatrix} (\boldsymbol{x}_{i1} - \bar{\boldsymbol{x}}_i)' \\ \vdots \\ (\boldsymbol{x}_{iT} - \bar{\boldsymbol{x}}_i)' \end{pmatrix} \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)\mathbb{E}[\dot{\varepsilon}_{it}\dot{\varepsilon}_{is}|\boldsymbol{X}_i](\boldsymbol{x}_{is} - \bar{\boldsymbol{x}}_i)'
\end{aligned}$$

Finite sample adjustment can also be added. In Stata, $\frac{N}{N-1}\frac{NT-1}{NT-(K-1)}$ is multiplied.

1.3.8 GLS estimator of the demean transformed model if $Var(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i) = \sigma_\varepsilon^2\boldsymbol{I}_T$

$\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i \sim iid\ [\boldsymbol{0}, \sigma_\varepsilon^2\boldsymbol{I}_T]$ implies $\boldsymbol{Q}\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i \sim iid\ [\boldsymbol{0}, \sigma_\varepsilon^2\boldsymbol{Q}]$, we want to find a GLS transformer $\boldsymbol{T}_{GLS}$ such that

$$\begin{aligned}
Var(\boldsymbol{T}_{GLS}\boldsymbol{Q}\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i) &= \sigma_\varepsilon^2\boldsymbol{I}_T \\
\boldsymbol{T}_{GLS}Var(\boldsymbol{Q}\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i)\boldsymbol{T}_{GLS}' &= \sigma_\varepsilon^2\boldsymbol{I}_T \\
\boldsymbol{T}_{GLS}\sigma_\varepsilon^2\boldsymbol{Q}\boldsymbol{T}_{GLS}' &= \sigma_\varepsilon^2\boldsymbol{I}_T \\
\boldsymbol{T}_{GLS}\boldsymbol{Q}^{1/2}\boldsymbol{Q}^{1/2}\boldsymbol{T}_{GLS}' &= \boldsymbol{I}_T \\
\boldsymbol{T}_{GLS}\boldsymbol{Q}^{1/2}\boldsymbol{Q}'^{1/2}\boldsymbol{T}_{GLS}' &= \boldsymbol{I}_T \\
\boldsymbol{T}_{GLS}\boldsymbol{Q}^{1/2}(\boldsymbol{T}_{GLS}\boldsymbol{Q}^{1/2})' &= \boldsymbol{I}_T
\end{aligned}$$

So, $\boldsymbol{T}_{GLS} = \boldsymbol{Q}^{-1/2}$

$$\boldsymbol{Q}^{-1/2}\boldsymbol{Q}\boldsymbol{y}_i = \boldsymbol{Q}^{-1/2}(\boldsymbol{Q}\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Q}\boldsymbol{\varepsilon}_i) = \boldsymbol{Q}^{-1/2}\boldsymbol{Q}\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Q}^{-1/2}\boldsymbol{Q}\boldsymbol{\varepsilon}_i$$

Thus, we have $Var(\boldsymbol{Q}^{-1/2}\boldsymbol{Q}\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i) = \boldsymbol{Q}^{-1/2}Var(\boldsymbol{Q}\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i)\boldsymbol{Q}'^{-1/2} = \boldsymbol{Q}^{-1/2}\sigma_\varepsilon^2\boldsymbol{Q}\boldsymbol{Q}^{-1/2} = \sigma_\varepsilon^2\boldsymbol{Q}^{-1/2}\boldsymbol{Q}^{1/2}\boldsymbol{Q}^{1/2}\boldsymbol{Q}^{-1/2} = \sigma_\varepsilon^2\boldsymbol{I}_T$.

By Gauss-Markov Theorem, GLS estimator is efficient.

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}^{gls}_{within} &= [\sum_{i=1}^{N}(\boldsymbol{Q}^{-1/2}\boldsymbol{Q}\boldsymbol{X}_i)'\boldsymbol{Q}^{-1/2}\boldsymbol{Q}\boldsymbol{X}_i]^{-1}\sum_{i=1}^{N}(\boldsymbol{Q}^{-1/2}\boldsymbol{Q}\boldsymbol{X}_i)'\boldsymbol{Q}^{-1/2}\boldsymbol{Q}\boldsymbol{y}_i \\
&= [\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}'^{-1/2}\boldsymbol{Q}^{-1/2}\boldsymbol{Q}\boldsymbol{X}_i]^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}'^{-1/2}\boldsymbol{Q}^{-1/2}\boldsymbol{Q}\boldsymbol{y}_i \\
&= [\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}^{-1/2}\boldsymbol{Q}^{-1/2}\boldsymbol{Q}\boldsymbol{X}_i]^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}^{-1/2}\boldsymbol{Q}^{-1/2}\boldsymbol{Q}\boldsymbol{y}_i \\
&= [\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}^{-}\boldsymbol{Q}\boldsymbol{X}_i]^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}^{-}\boldsymbol{Q}\boldsymbol{y}_i \\
&= [\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{X}_i]^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{y}_i = \widehat{\boldsymbol{\beta}}^{ols}_{within} \qquad \text{So, FE estimator is also efficient}
\end{aligned}$$

For generalized inverse, $\boldsymbol{Q}'\boldsymbol{Q}^{-}\boldsymbol{Q} = \boldsymbol{Q}$. As $\boldsymbol{Q}$ is idempotent and symmetry, $\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{Q}' = \boldsymbol{Q}'\boldsymbol{Q}$. Therefore, $\boldsymbol{Q}'\boldsymbol{Q}^{-}\boldsymbol{Q} = \boldsymbol{Q}'\boldsymbol{Q}$.

1.4 First-Difference Estimator

1.4.1 First-difference operator

$$\boldsymbol{\Delta} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ 0 & 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}$$

$$\begin{aligned}
\boldsymbol{\Delta}\boldsymbol{\Delta}' &= \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ 0 & 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ 0 & 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}' \\
&= \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ 0 & 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 2 \end{pmatrix}
\end{aligned}$$

1.4.2 First difference transformed model

$$\begin{aligned}
\boldsymbol{\Delta}\boldsymbol{y}_i &= \boldsymbol{\Delta}(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{e}\alpha_i + \boldsymbol{\varepsilon}_i) \\
&= \boldsymbol{\Delta}\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\Delta}\boldsymbol{e}\alpha_i + \boldsymbol{\Delta}\boldsymbol{\varepsilon}_i \\
&= \boldsymbol{\Delta}\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{0}\alpha_i + \boldsymbol{\Delta}\boldsymbol{\varepsilon}_i \\
&= \boldsymbol{\Delta}\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\Delta}\boldsymbol{\varepsilon}_i \qquad \text{Level 2}
\end{aligned}$$

It is because

$$\boldsymbol{\Delta e} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ 0 & 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix} = \mathbf{0}$$

It can be written as

$$\begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ 0 & 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix} \begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \\ \vdots \\ \vdots \\ y_{iT} \end{pmatrix} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ 0 & 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{x}'_{i1} \\ \boldsymbol{x}'_{i2} \\ \boldsymbol{x}'_{i3} \\ \vdots \\ \vdots \\ \boldsymbol{x}'_{iT} \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ 0 & 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \vdots \\ \vdots \\ \varepsilon_{iT} \end{pmatrix}$$

$$\begin{pmatrix} y_{i2} - y_{i1} \\ y_{i3} - y_{i2} \\ y_{i4} - y_{i3} \\ \vdots \\ \vdots \\ y_{iT} - y_{i,T-1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}'_{i2} - \boldsymbol{x}'_{i1} \\ \boldsymbol{x}'_{i3} - \boldsymbol{x}'_{i2} \\ \boldsymbol{x}'_{i4} - \boldsymbol{x}'_{i3} \\ \vdots \\ \vdots \\ \boldsymbol{x}'_{iT} - \boldsymbol{x}'_{i,T-1} \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \varepsilon_{i2} - \varepsilon_{i1} \\ \varepsilon_{i3} - \varepsilon_{i2} \\ \varepsilon_{i4} - \varepsilon_{i3} \\ \vdots \\ \vdots \\ \varepsilon_{iT} - \varepsilon_{i,T-1} \end{pmatrix}$$

$$\begin{pmatrix} y_{i2} - y_{i1} \\ y_{i3} - y_{i2} \\ y_{i4} - y_{i3} \\ \vdots \\ \vdots \\ y_{iT} - y_{i,T-1} \end{pmatrix} = \begin{pmatrix} (\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1})' \\ (\boldsymbol{x}_{i3} - \boldsymbol{x}_{i2})' \\ (\boldsymbol{x}_{i4} - \boldsymbol{x}_{i3})' \\ \vdots \\ \vdots \\ (\boldsymbol{x}_{iT} - \boldsymbol{x}_{i,T-1})' \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \varepsilon_{i2} - \varepsilon_{i1} \\ \varepsilon_{i3} - \varepsilon_{i2} \\ \varepsilon_{i4} - \varepsilon_{i3} \\ \vdots \\ \vdots \\ \varepsilon_{iT} - \varepsilon_{i,T-1} \end{pmatrix}$$

$$y_{it} - y_{i,t-1} = (\boldsymbol{x}_{it} - \boldsymbol{x}_{i,t-1})'\boldsymbol{\beta} + (\varepsilon_{it} - \varepsilon_{i,t-1}) \qquad \text{Level 1}$$

1.4.3 OLS estimator of the first difference transformed model

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{fd}^{ols} &= [\sum_{i=1}^{N} (\boldsymbol{\Delta X}_i)' \boldsymbol{\Delta X}_i]^{-1} \sum_{i=1}^{N} (\boldsymbol{\Delta X}_i)' \boldsymbol{\Delta y}_i & \text{Level 2} \\ &= [\sum_{i=1}^{N} \sum_{t=2}^{T} (\boldsymbol{x}_{it} - \boldsymbol{x}_{i,t-1})(\boldsymbol{x}_{it} - \boldsymbol{x}_{i,t-1})']^{-1} \sum_{i=1}^{N} \sum_{t=2}^{T} (\boldsymbol{x}_{it} - \boldsymbol{x}_{i,t-1})(y_{it} - y_{i,t-1}) & \text{Level 1} \end{aligned}$$

It is because

$$(\boldsymbol{\Delta X}_i)'\boldsymbol{\Delta X}_i = \begin{pmatrix} (\boldsymbol{x}_{i2}-\boldsymbol{x}_{i1})' \\ (\boldsymbol{x}_{i3}-\boldsymbol{x}_{i2})' \\ (\boldsymbol{x}_{i4}-\boldsymbol{x}_{i3})' \\ \vdots \\ \vdots \\ (\boldsymbol{x}_{iT}-\boldsymbol{x}_{i,T-1})' \end{pmatrix}' \begin{pmatrix} (\boldsymbol{x}_{i2}-\boldsymbol{x}_{i1})' \\ (\boldsymbol{x}_{i3}-\boldsymbol{x}_{i2})' \\ (\boldsymbol{x}_{i4}-\boldsymbol{x}_{i3})' \\ \vdots \\ \vdots \\ (\boldsymbol{x}_{iT}-\boldsymbol{x}_{i,T-1})' \end{pmatrix}$$

$$= \begin{pmatrix} (\boldsymbol{x}_{i2}-\boldsymbol{x}_{i1}) & (\boldsymbol{x}_{i3}-\boldsymbol{x}_{i2}) & (\boldsymbol{x}_{i4}-\boldsymbol{x}_{i3}) & \cdots & \cdots & (\boldsymbol{x}_{iT}-\boldsymbol{x}_{i,T-1}) \end{pmatrix} \begin{pmatrix} (\boldsymbol{x}_{i2}-\boldsymbol{x}_{i1})' \\ (\boldsymbol{x}_{i3}-\boldsymbol{x}_{i2})' \\ (\boldsymbol{x}_{i4}-\boldsymbol{x}_{i3})' \\ \vdots \\ \vdots \\ (\boldsymbol{x}_{iT}-\boldsymbol{x}_{i,T-1})' \end{pmatrix}$$

$$= \sum_{t=2}^{T} (\boldsymbol{x}_{it}-\boldsymbol{x}_{i,t-1})(\boldsymbol{x}_{it}-\boldsymbol{x}_{i,t-1})'$$

$$(\boldsymbol{\Delta X}_i)'\boldsymbol{\Delta y}_i = \begin{pmatrix} (\boldsymbol{x}_{i2}-\boldsymbol{x}_{i1})' \\ (\boldsymbol{x}_{i3}-\boldsymbol{x}_{i2})' \\ (\boldsymbol{x}_{i4}-\boldsymbol{x}_{i3})' \\ \vdots \\ \vdots \\ (\boldsymbol{x}_{iT}-\boldsymbol{x}_{i,T-1})' \end{pmatrix}' \begin{pmatrix} y_{i2}-y_{i1} \\ y_{i3}-y_{i2} \\ y_{i4}-y_{i3} \\ \vdots \\ \vdots \\ y_{iT}-y_{i,T-1} \end{pmatrix}$$

$$= \begin{pmatrix} (\boldsymbol{x}_{i2}-\boldsymbol{x}_{i1}) & (\boldsymbol{x}_{i3}-\boldsymbol{x}_{i2}) & (\boldsymbol{x}_{i4}-\boldsymbol{x}_{i3}) & \cdots & \cdots & (\boldsymbol{x}_{iT}-\boldsymbol{x}_{i,T-1}) \end{pmatrix} \begin{pmatrix} y_{i2}-y_{i1} \\ y_{i3}-y_{i2} \\ y_{i4}-y_{i3} \\ \vdots \\ \vdots \\ y_{iT}-y_{i,T-1} \end{pmatrix}$$

$$= \sum_{t=2}^{T} (\boldsymbol{x}_{it}-\boldsymbol{x}_{i,t-1})(y_{it}-y_{i,t-1})$$

1.4.4 The necessary condition for consistency and unbiasedness

The necessary condition for FD estimator (OLS estimator of the FD transformed model) to be consistent is $\mathbb{E}(\boldsymbol{\Delta X}_i)'\boldsymbol{\Delta \varepsilon}_i) = \mathbf{0}$

$$\begin{aligned} \mathbb{E}((\boldsymbol{\Delta X}_i)'\boldsymbol{\Delta \varepsilon}_i) &= \mathbb{E}(\boldsymbol{X}_i'\boldsymbol{\Delta}'\boldsymbol{\Delta \varepsilon}_i) \\ &= \mathbb{E}(\mathbb{E}(\boldsymbol{X}_i'\boldsymbol{\Delta\Delta}'\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i)) \\ &= \mathbb{E}(\boldsymbol{X}_i'\boldsymbol{\Delta\Delta}'\underbrace{\mathbb{E}(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i)}_{\mathbf{0}}) && \text{because of strict exogeneity} \\ &= \mathbf{0} \end{aligned}$$

Thus, FD estimator satisfies the necessary condition for consistency given strict exogeneity assumption. Indeed, strict exogeneity is stronger than what is required. To see this, first note that for any t

$$\mathbb{E}(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i) = \mathbb{E}(\varepsilon_{it}|\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{iT}) = 0 \implies \mathbb{E}(\boldsymbol{x}_{is}\varepsilon_{it}) = \mathbf{0} \qquad \text{for all } s$$

The necessary condition for FD estimator to be consistent can also be written as $\mathbb{E}((\boldsymbol{x}_{it}-\boldsymbol{x}_{i,t-1})(\varepsilon_{it}-\varepsilon_{i,t-1})) = \mathbf{0}$

$$\mathbb{E}((\boldsymbol{x}_{it}-\boldsymbol{x}_{i,t-1})(\varepsilon_{it}-\varepsilon_{i,t-1})) = \mathbb{E}(\boldsymbol{x}_{it}\varepsilon_{it}) - \mathbb{E}(\boldsymbol{x}_{it}\varepsilon_{i,t-1}) - \mathbb{E}(\boldsymbol{x}_{i,t-1}\varepsilon_{it}) + \mathbb{E}(\boldsymbol{x}_{i,t-1}\varepsilon_{i,t-1}) = \mathbf{0}$$

It is because $\mathbb{E}(\boldsymbol{x}_{is}\varepsilon_{it}) = \mathbf{0}$ for any t and s implies

$$\mathbb{E}(\boldsymbol{x}_{it}\varepsilon_{it}) = \mathbb{E}(\boldsymbol{x}_{it}\varepsilon_{i,t-1}) = \mathbb{E}(\boldsymbol{x}_{i,t-1}\varepsilon_{it}) = \mathbb{E}(\boldsymbol{x}_{i,t-1}\varepsilon_{i,t-1}) = \mathbf{0}$$

Thus, the weaker assumption $\mathbb{E}(\boldsymbol{x}_{is}\varepsilon_{it}) = \mathbf{0}$ for any t and s is sufficient for $\mathbb{E}((\boldsymbol{x}_{it} - \boldsymbol{x}_{i,t-1})(\varepsilon_{it} - \varepsilon_{i,t-1})) = \mathbf{0}$

The necessary condition for FD estimator to be unbiased is $\mathbb{E}(\boldsymbol{\Delta\varepsilon}_i|\boldsymbol{\Delta X}_i) = \mathbf{0}$

$$\begin{aligned}
\mathbb{E}(\boldsymbol{\Delta\varepsilon}_i|\boldsymbol{\Delta X}_i) &= \boldsymbol{\Delta}\underbrace{\mathbb{E}(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i)}_{\mathbf{0}} && \text{as } \boldsymbol{\Delta} \text{ is constant and strict exogeneity} \\
&= \mathbf{0}
\end{aligned}$$

1.4.5 Conditional variance of $\widehat{\boldsymbol{\beta}}_{fd}^{ols}$

$$\begin{aligned}
Var(\widehat{\boldsymbol{\beta}}_{fd}^{ols}|\boldsymbol{X}_i) &= Var([\sum_{i=1}^{N}(\boldsymbol{\Delta X}_i)'\boldsymbol{\Delta X}_i]^{-1}\sum_{i=1}^{N}(\boldsymbol{\Delta X}_i)'\boldsymbol{\Delta y}_i|\boldsymbol{X}_i) \\
&= [\sum_{i=1}^{N}(\boldsymbol{\Delta X}_i)'\boldsymbol{\Delta X}_i]^{-1}Var(\sum_{i=1}^{N}(\boldsymbol{\Delta X}_i)'\boldsymbol{\Delta y}_i|\boldsymbol{X}_i)[\sum_{i=1}^{N}(\boldsymbol{\Delta X}_i)'\boldsymbol{\Delta X}_i]^{-1'} \\
&= [\sum_{i=1}^{N}(\boldsymbol{\Delta X}_i)'\boldsymbol{\Delta X}_i]^{-1}\sum_{i=1}^{N}(\boldsymbol{\Delta X}_i)'Var(\boldsymbol{\Delta\varepsilon}_i|\boldsymbol{X}_i)\boldsymbol{\Delta X}_i[\sum_{i=1}^{N}(\boldsymbol{\Delta X}_i)'\boldsymbol{\Delta X}_i]^{-1}
\end{aligned}$$

1.4.6 $Var(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i) = \sigma_\varepsilon^2\boldsymbol{I}_T$

If ε_{it} is homoskedasticity and serially uncorrelated across t i.e., $Var(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i) = \sigma_\varepsilon^2\boldsymbol{I}_T$ (further assume independence of i and strict exogeneity), we have $\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i \sim iid\ [\mathbf{0}, \sigma_\varepsilon^2\boldsymbol{I}_T]$.

$Var(\boldsymbol{\Delta\varepsilon_i}|\boldsymbol{X}_i) = \boldsymbol{\Delta}Var(\boldsymbol{\varepsilon}|\boldsymbol{X}_i)\boldsymbol{\Delta'} = \boldsymbol{\Delta}\sigma_\varepsilon^2\boldsymbol{I}_T\boldsymbol{\Delta'} = \sigma_\varepsilon^2\boldsymbol{\Delta\Delta'} = \sigma_\varepsilon^2\begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 2 \end{pmatrix}$ Thus, $\boldsymbol{\Delta\varepsilon}_i$ is homoskedasticity but not serially uncorrelated e.g. $Cov(\varepsilon_{it}-\varepsilon_{i,t-1}, \varepsilon_{i,t-1}-\varepsilon_{i,t-2}|\boldsymbol{X}_i) = -\sigma_\varepsilon^2 < 0$. Therefore, we cannot apply Gauss-Markov Theorem.

$$\begin{aligned}
Var(\widehat{\boldsymbol{\beta}}_{fd}^{ols}|\boldsymbol{X}_i) &= [\sum_{i=1}^{N}(\boldsymbol{\Delta X}_i)'\boldsymbol{\Delta X}_i]^{-1}\sum_{i=1}^{N}(\boldsymbol{\Delta X}_i)'\sigma_\varepsilon^2\boldsymbol{\Delta\Delta'\Delta X}_i[\sum_{i=1}^{N}(\boldsymbol{\Delta X}_i)'\boldsymbol{\Delta X}_i]^{-1} \\
&= \sigma_\varepsilon^2[\sum_{i=1}^{N}(\boldsymbol{\Delta X}_i)'\boldsymbol{\Delta X}_i]^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{\Delta'\Delta\Delta'\Delta X}_i[\sum_{i=1}^{N}(\boldsymbol{\Delta X}_i)'\boldsymbol{\Delta X}_i]^{-1}
\end{aligned}$$

1.4.7 $Var(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i) = \boldsymbol{\Omega}_i$

We have $\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i \sim inid\ [\mathbf{0}, \boldsymbol{\Omega}_i]$,

$Var(\boldsymbol{\Delta\varepsilon_i}|\boldsymbol{X}_i) = \boldsymbol{\Delta}Var(\boldsymbol{\varepsilon}|\boldsymbol{X}_i)\boldsymbol{\Delta'} = \boldsymbol{\Delta}\mathbb{E}[(\boldsymbol{\varepsilon}_i - \mathbb{E}[\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i])(\boldsymbol{\varepsilon}_i - \mathbb{E}[\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i])'|\boldsymbol{X}_i]\boldsymbol{\Delta'} = \boldsymbol{\Delta}\mathbb{E}[(\boldsymbol{\varepsilon}_i - \mathbf{0})(\boldsymbol{\varepsilon}_i - \mathbf{0})'|\boldsymbol{X}_i]\boldsymbol{\Delta'} = \boldsymbol{\Delta}\mathbb{E}[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'|\boldsymbol{X}_i]\boldsymbol{\Delta'} = \mathbb{E}[\boldsymbol{\Delta\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{\Delta'}|\boldsymbol{X}_i] = \mathbb{E}[\boldsymbol{\Delta\varepsilon}_i(\boldsymbol{\Delta\varepsilon}_i)'|\boldsymbol{X}_i]$

$$Var(\widehat{\boldsymbol{\beta}}_{fd}^{ols}|\boldsymbol{X}_i) = [\sum_{i=1}^{N}(\boldsymbol{\Delta X}_i)'\boldsymbol{\Delta X}_i]^{-1}\sum_{i=1}^{N}(\boldsymbol{\Delta X}_i)'E[\boldsymbol{\Delta\varepsilon}_i(\boldsymbol{\Delta\varepsilon}_i)'|\boldsymbol{X}_i]\boldsymbol{\Delta X}_i[\sum_{i=1}^{N}(\boldsymbol{\Delta X}_i)'\boldsymbol{\Delta X}_i]^{-1}$$

If ε_{it} follows random walk process i.e., $\varepsilon_{it} = \varepsilon_{i,t-1} + v_{it}$ where v_{it} follows white noise process, $\varepsilon_{it} - \varepsilon_{i,t-1} = v_{it}$ follows white noise process. Thus, $\varepsilon_{it} - \varepsilon_{i,t-1}$ is homoskedasticity and serially uncorrelated as they are the properties of white noise process. As a result, FD estimator is efficient in this case by applying Gauss-Markov Theorem.

1.5 Least-Squares Dummy Variable Estimator

$$\boldsymbol{y} = (\boldsymbol{I}_N \otimes \boldsymbol{e})\boldsymbol{\alpha} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \begin{pmatrix}(\boldsymbol{I}_N \otimes \boldsymbol{e}) & \boldsymbol{X}\end{pmatrix}\begin{pmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\beta}\end{pmatrix} + \boldsymbol{\varepsilon} \qquad \text{Level 3}$$

$$\begin{aligned}
\begin{pmatrix}\widehat{\boldsymbol{\alpha}}_{dv}^{ols}\\ \widehat{\boldsymbol{\beta}}_{dv}^{ols}\end{pmatrix} &= [\begin{pmatrix}(\boldsymbol{I}_N \otimes \boldsymbol{e}) & \boldsymbol{X}\end{pmatrix}' \begin{pmatrix}(\boldsymbol{I}_N \otimes \boldsymbol{e}) & \boldsymbol{X}\end{pmatrix}]^{-1} \begin{pmatrix}(\boldsymbol{I}_N \otimes \boldsymbol{e}) & \boldsymbol{X}\end{pmatrix}' \boldsymbol{y}\\
&= \begin{pmatrix}(\boldsymbol{I}_N \otimes \boldsymbol{e})'(\boldsymbol{I}_N \otimes \boldsymbol{e}) & (\boldsymbol{I}_N \otimes \boldsymbol{e})'\boldsymbol{X}\\ \boldsymbol{X}'(\boldsymbol{I}_N \otimes \boldsymbol{e}) & \boldsymbol{X}'\boldsymbol{X}\end{pmatrix}^{-1} \begin{pmatrix}(\boldsymbol{I}_N \otimes \boldsymbol{e})'\boldsymbol{y}\\ \boldsymbol{X}'\boldsymbol{y}\end{pmatrix}\\
&= \begin{pmatrix}T\boldsymbol{I}_N & T\bar{\boldsymbol{X}}\\ T\bar{\boldsymbol{X}}' & \boldsymbol{X}'\boldsymbol{X}\end{pmatrix}^{-1} \begin{pmatrix}T\bar{\boldsymbol{y}}\\ \boldsymbol{X}'\boldsymbol{y}\end{pmatrix}\\
\widehat{\boldsymbol{\beta}}_{dv}^{ols} &= [\boldsymbol{X}'\boldsymbol{X} - T\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}}]^{-1}(\boldsymbol{X}'\boldsymbol{y} - T\bar{\boldsymbol{X}}'\bar{\boldsymbol{y}}) = \widehat{\boldsymbol{\beta}}_{within}^{ols}
\end{aligned}$$

It is because

$$\begin{aligned}
\begin{pmatrix}(\boldsymbol{I}_N \otimes \boldsymbol{e}) & \boldsymbol{X}\end{pmatrix}' \begin{pmatrix}(\boldsymbol{I}_N \otimes \boldsymbol{e}) & \boldsymbol{X}\end{pmatrix} &= \begin{pmatrix}(\boldsymbol{I}_N \otimes \boldsymbol{e})'\\ \boldsymbol{X}'\end{pmatrix} \begin{pmatrix}(\boldsymbol{I}_N \otimes \boldsymbol{e}) & \boldsymbol{X}\end{pmatrix}\\
&= \begin{pmatrix}(\boldsymbol{I}_N \otimes \boldsymbol{e})'(\boldsymbol{I}_N \otimes \boldsymbol{e}) & (\boldsymbol{I}_N \otimes \boldsymbol{e})'\boldsymbol{X}\\ \boldsymbol{X}'(\boldsymbol{I}_N \otimes \boldsymbol{e}) & \boldsymbol{X}'\boldsymbol{X}\end{pmatrix}
\end{aligned}$$

$$\begin{aligned}
(\boldsymbol{I}_N \otimes \boldsymbol{e})'(\boldsymbol{I}_N \otimes \boldsymbol{e}) &= \begin{pmatrix}\boldsymbol{e} & \cdots & \boldsymbol{0}\\ \vdots & \ddots & \vdots\\ \boldsymbol{0} & \cdots & \boldsymbol{e}\end{pmatrix}' \begin{pmatrix}\boldsymbol{e} & \cdots & \boldsymbol{0}\\ \vdots & \ddots & \vdots\\ \boldsymbol{0} & \cdots & \boldsymbol{e}\end{pmatrix}\\
&= \begin{pmatrix}\boldsymbol{e}' & \cdots & \boldsymbol{0}\\ \vdots & \ddots & \vdots\\ \boldsymbol{0} & \cdots & \boldsymbol{e}'\end{pmatrix} \begin{pmatrix}\boldsymbol{e} & \cdots & \boldsymbol{0}\\ \vdots & \ddots & \vdots\\ \boldsymbol{0} & \cdots & \boldsymbol{e}\end{pmatrix}\\
&= \begin{pmatrix}\boldsymbol{e}'\boldsymbol{e} & \cdots & \boldsymbol{0}\\ \vdots & \ddots & \vdots\\ \boldsymbol{0} & \cdots & \boldsymbol{e}'\boldsymbol{e}\end{pmatrix}\\
&= \begin{pmatrix}T & \cdots & \boldsymbol{0}\\ \vdots & \ddots & \vdots\\ \boldsymbol{0} & \cdots & T\end{pmatrix}\\
&= T\boldsymbol{I}_N
\end{aligned}$$

$$\begin{aligned}
(\boldsymbol{I}_N \otimes \boldsymbol{e})'\boldsymbol{X} &= \begin{pmatrix}\boldsymbol{e}' & \cdots & \boldsymbol{0}\\ \vdots & \ddots & \vdots\\ \boldsymbol{0} & \cdots & \boldsymbol{e}'\end{pmatrix} \begin{pmatrix}\boldsymbol{X}_1\\ \vdots\\ \boldsymbol{X}_N\end{pmatrix}\\
&= \begin{pmatrix}\boldsymbol{e}'\boldsymbol{X}_1\\ \vdots\\ \boldsymbol{e}'\boldsymbol{X}_N\end{pmatrix}\\
&= \begin{pmatrix}\sum_{t=1}^T \boldsymbol{x}_{1t}'\\ \vdots\\ \sum_{t=1}^T \boldsymbol{x}_{Nt}'\end{pmatrix}\\
&= \begin{pmatrix}T\sum_{t=1}^T \boldsymbol{x}_{1t}'/T\\ \vdots\\ T\sum_{t=1}^T \boldsymbol{x}_{Nt}'/T\end{pmatrix}\\
&= \begin{pmatrix}T\bar{\boldsymbol{x}}_1'\\ \vdots\\ T\bar{\boldsymbol{x}}_N'\end{pmatrix}\\
&= T\bar{\boldsymbol{X}}
\end{aligned}$$

$$
\begin{aligned}
((\boldsymbol{I}_N \otimes \boldsymbol{e}) \quad \boldsymbol{X})' \boldsymbol{y} &= \begin{pmatrix} (\boldsymbol{I}_N \otimes \boldsymbol{e})' \\ \boldsymbol{X}' \end{pmatrix} \boldsymbol{y} \\
&= \begin{pmatrix} (\boldsymbol{I}_N \otimes \boldsymbol{e})'\boldsymbol{y} \\ \boldsymbol{X}'\boldsymbol{y} \end{pmatrix}
\end{aligned}
$$

Another way to show the equivalence of within estimator and dummy variable estimator by using Frisch-Waugh-Lovell Theorem,

$$
\begin{aligned}
\boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I}_N \otimes \boldsymbol{e})\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \\
\boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{E}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{X} &= \boldsymbol{E}\boldsymbol{\alpha}_{XE} + \boldsymbol{\varepsilon}_{XE} \\
\boldsymbol{y} &= \boldsymbol{E}\boldsymbol{\alpha}_{yE} + \boldsymbol{\varepsilon}_{yE}
\end{aligned}
$$

$$
\begin{aligned}
\widehat{\boldsymbol{\alpha}}_{yE} &= (\boldsymbol{E}'\boldsymbol{E})^{-1}\boldsymbol{E}'\boldsymbol{y} \\
&= \begin{pmatrix} \boldsymbol{e} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{e} \end{pmatrix}' \begin{pmatrix} \boldsymbol{e} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{e} \end{pmatrix})^{-1} \begin{pmatrix} \boldsymbol{e} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{e} \end{pmatrix}' \begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_N \end{pmatrix} \\
&= (\begin{pmatrix} \boldsymbol{e}' & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{e}' \end{pmatrix} \begin{pmatrix} \boldsymbol{e} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{e} \end{pmatrix})^{-1} \begin{pmatrix} \boldsymbol{e}' & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{e}' \end{pmatrix} \begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_N \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{e}'\boldsymbol{e} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{e}'\boldsymbol{e} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{e}'\boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{e}'\boldsymbol{y}_N \end{pmatrix} \\
&= \begin{pmatrix} (\boldsymbol{e}'\boldsymbol{e})^{-1} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & (\boldsymbol{e}'\boldsymbol{e})^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{e}'\boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{e}'\boldsymbol{y}_N \end{pmatrix} \\
&= \begin{pmatrix} (\boldsymbol{e}'\boldsymbol{e})^{-1}\boldsymbol{e}'\boldsymbol{y}_1 \\ \vdots \\ (\boldsymbol{e}'\boldsymbol{e})^{-1}\boldsymbol{e}'\boldsymbol{y}_N \end{pmatrix} = \begin{pmatrix} T^{-1}\sum_{t=1}^{T} y_{1t} \\ \vdots \\ T^{-1}\sum_{t=1}^{T} y_{Nt} \end{pmatrix} = \begin{pmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_N \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\widehat{\boldsymbol{\varepsilon}}_{yE} = \boldsymbol{y} - \boldsymbol{E}\widehat{\boldsymbol{\alpha}}_{yE} &= \begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_N \end{pmatrix} - \begin{pmatrix} \boldsymbol{e} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{e} \end{pmatrix} \begin{pmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_N \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_N \end{pmatrix} - \begin{pmatrix} \boldsymbol{e}\bar{y}_1 \\ \vdots \\ \boldsymbol{e}\bar{y}_N \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{y}_1 - \boldsymbol{e}\bar{y}_1 \\ \vdots \\ \boldsymbol{y}_N - \boldsymbol{e}\bar{y}_N \end{pmatrix} = \begin{pmatrix} \boldsymbol{Q}\boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{Q}\boldsymbol{y}_N \end{pmatrix} = \boldsymbol{Q}\boldsymbol{y}
\end{aligned}
$$

Similarly,

$$
\widehat{\boldsymbol{\varepsilon}}_{XE} = \boldsymbol{Q}\boldsymbol{X}
$$

By Frisch-Waugh-Lovell Theorem,

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}_{dv}^{ols} &= (\widehat{\boldsymbol{\varepsilon}}_{XE}'\widehat{\boldsymbol{\varepsilon}}_{XE})^{-1}\widehat{\boldsymbol{\varepsilon}}_{XE}'\widehat{\boldsymbol{\varepsilon}}_{yE} \\
&= [(\boldsymbol{Q}\boldsymbol{X})'\boldsymbol{Q}\boldsymbol{X}]^{-1}(\boldsymbol{Q}\boldsymbol{X})'\boldsymbol{Q}\boldsymbol{y} = \widehat{\boldsymbol{\beta}}_{within}^{ols}
\end{aligned}
$$

If $N \to \infty$, the number of α_i estimated goes to infinity. If T is fixed, the LSDV estimates are consistent for $\boldsymbol{\beta}$ (as FE estimates for $\boldsymbol{\beta}$ is consistent for fixed T and $N \to \infty$) but inconsistent for $\boldsymbol{\alpha}$. (There is no incidental parameters problem as the estimates for $\boldsymbol{\beta}$ are not contaminated). If T also $\to \infty$, then the LSDV estimates of $\boldsymbol{\alpha}$ are also consistent.

2 Random Effect Model

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

$$\begin{aligned}
\begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix} &= \begin{pmatrix} \boldsymbol{x}'_{i1} \\ \vdots \\ \boldsymbol{x}'_{iT} \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \alpha_i + \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iT} \end{pmatrix} \\
\boldsymbol{y}_i &= \boldsymbol{X}_i\boldsymbol{\beta} + \underbrace{(\boldsymbol{e}\alpha_i + \boldsymbol{\varepsilon}_i)}_{\boldsymbol{u}_i}
\end{aligned}$$

$$\begin{aligned}
\begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_N \end{pmatrix} &= \begin{pmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_N \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{e} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{e} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_N \end{pmatrix} \\
\boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I}_N \otimes \boldsymbol{e})\boldsymbol{\alpha} + \boldsymbol{\varepsilon}
\end{aligned}$$

2.1 Assumptions

2.1.1 Strong/strict exogeneity of regressors

For all t,

$$E(\varepsilon_{it}|\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iT}) = 0$$

Equivalently,

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

2.1.2 Covariance structure

$$\begin{aligned}
\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i &\sim iid\ [\boldsymbol{0}, \sigma^2_\varepsilon \boldsymbol{I}_T] \\
\alpha_i|\boldsymbol{X}_i &\sim iid\ [0, \sigma^2_\alpha] \\
\boldsymbol{\varepsilon}_i &\perp \alpha_i|\boldsymbol{X}_i
\end{aligned}$$

2.2 Moments of $\boldsymbol{u}_i|\boldsymbol{X}_i$

$$\begin{aligned}
\boldsymbol{\Omega} := Var(\boldsymbol{u}_i|\boldsymbol{X}_i) &= Var(\boldsymbol{e}\alpha_i + \boldsymbol{\varepsilon}_i|\boldsymbol{X}_i) && \text{because of } \boldsymbol{\varepsilon}_i\perp\alpha_i|\boldsymbol{X}_i \\
&= Var(\boldsymbol{e}\alpha_i|\boldsymbol{X}_i) + Var(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i) \\
&= \boldsymbol{e}Var(\alpha_i|\boldsymbol{X}_i)\boldsymbol{e}' + Var(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i) \\
&= \sigma^2_\alpha \boldsymbol{e}\boldsymbol{e}' + \sigma^2_\varepsilon \boldsymbol{I}_T \\
&= \begin{pmatrix} \sigma^2_\alpha & \cdots & \sigma^2_\alpha \\ \vdots & \ddots & \vdots \\ \sigma^2_\alpha & \cdots & \sigma^2_\alpha \end{pmatrix} + \begin{pmatrix} \sigma^2_\varepsilon & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2_\varepsilon \end{pmatrix} \\
&= \begin{pmatrix} \sigma^2_\alpha + \sigma^2_\varepsilon & \cdots & \sigma^2_\alpha \\ \vdots & \ddots & \vdots \\ \sigma^2_\alpha & \cdots & \sigma^2_\alpha + \sigma^2_\varepsilon \end{pmatrix}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(\boldsymbol{u}_i|\boldsymbol{X}_i) &= \mathbb{E}(\boldsymbol{e}\alpha_i + \boldsymbol{\varepsilon}_i|\boldsymbol{X}_i) \\
&= \boldsymbol{e}\underbrace{\mathbb{E}(\alpha_i|\boldsymbol{X}_i)}_{0} + \underbrace{\mathbb{E}(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i)}_{\boldsymbol{0}} \\
&= \boldsymbol{0}
\end{aligned}$$

Note that $\mathbb{E}(\alpha_i|\boldsymbol{X}_i) = \mathbb{E}(\alpha_i|\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iT}) = 0$ called orthogonality assumption. $\mathbb{E}(\alpha_i|\boldsymbol{X}_i) = 0 \implies Cov(\alpha_i, \boldsymbol{X}_i) = \boldsymbol{0}$. It is because $Cov(\alpha_i, \boldsymbol{X}_i) = \mathbb{E}(\alpha_i\boldsymbol{X}_i) - \mathbb{E}(\alpha_i)\mathbb{E}(\boldsymbol{X}_i) = \mathbb{E}(\mathbb{E}(\alpha_i\boldsymbol{X}_i|\boldsymbol{X}_i)) - \mathbb{E}(\mathbb{E}(\alpha_i|\boldsymbol{X}_i))\mathbb{E}(\boldsymbol{X}_i) = \mathbb{E}(\mathbb{E}(\alpha_i|\boldsymbol{X}_i)\boldsymbol{X}_i) = \boldsymbol{0}$. There is no OVB i.e., $\boldsymbol{u}_i$ is not correlated with $\boldsymbol{X}_i$.

$\mathbb{E}(\boldsymbol{u}_i|\boldsymbol{X}_i) = \mathbf{0}$ means that the necessary condition for OLS estimator to be unbiased is satisfied. Moreover, $\mathbb{E}(\boldsymbol{u}_i|\boldsymbol{X}_i) = \mathbf{0} \implies \mathbb{E}(\boldsymbol{X}_i'\boldsymbol{u}_i) = \mathbf{0}$ which means that the necessary condition for OLS estimator to be consistent is also satisfied. However, $\boldsymbol{u}_i|\boldsymbol{X}_i$ is homoskedasticity but serially correlated. Thus, the necessary condition for OLS estimator to be efficient is not satisfied. As a result, it is not efficient.

As we know the covariance structure of $\boldsymbol{u}_i|\boldsymbol{X}_i$ due to the strong assumptions in random effect model, we can use GLS estimation, which yields efficient estimates.

2.3 Random Effect Estimator (GLS Estimator)

2.3.1 GLS transformed model

We want to find a $\boldsymbol{T}_{GLS}$ such that

$$\begin{aligned}
Var(\boldsymbol{T}_{GLS}\boldsymbol{u}_i|\boldsymbol{X}_i) &= \sigma_\varepsilon^2\boldsymbol{I}_T \\
\boldsymbol{T}_{GLS}Var(\boldsymbol{u}_i|\boldsymbol{X}_i)\boldsymbol{T}_{GLS}' &= \sigma_\varepsilon^2\boldsymbol{I}_T \\
\boldsymbol{T}_{GLS}\boldsymbol{\Omega}\boldsymbol{T}_{GLS}' &= \sigma_\varepsilon^2\boldsymbol{I}_T \\
\boldsymbol{T}_{GLS}\boldsymbol{\Omega}^{1/2}\boldsymbol{\Omega}^{1/2}\boldsymbol{T}_{GLS}' &= \sigma_\varepsilon^2\boldsymbol{I}_T \\
\boldsymbol{T}_{GLS}\boldsymbol{\Omega}^{1/2}\boldsymbol{\Omega}'^{1/2}\boldsymbol{T}_{GLS}' &= \sigma_\varepsilon^2\boldsymbol{I}_T \\
\boldsymbol{T}_{GLS}\boldsymbol{\Omega}^{1/2}(\boldsymbol{T}_{GLS}\boldsymbol{\Omega}^{1/2})' &= \sigma_\varepsilon^2\boldsymbol{I}_T
\end{aligned}$$

So, $\boldsymbol{T}_{GLS} = \sigma_\varepsilon\boldsymbol{\Omega}^{-1/2}$. Define $\psi^2 := \frac{\sigma_\varepsilon^2}{T\sigma_\alpha^2+\sigma_\varepsilon^2}$.

$$\begin{aligned}
\boldsymbol{\Omega} &= \sigma_\varepsilon^2\boldsymbol{I}_T + \sigma_\alpha^2\boldsymbol{e}\boldsymbol{e}' \\
&= \sigma_\varepsilon^2(\boldsymbol{I}_T + \frac{\sigma_\alpha^2}{\sigma_\varepsilon^2}\boldsymbol{e}\boldsymbol{e}') \\
&= \sigma_\varepsilon^2(\boldsymbol{I}_T + \frac{T\sigma_\alpha^2}{\sigma_\varepsilon^2}T^{-1}\boldsymbol{e}\boldsymbol{e}') \\
&= \sigma_\varepsilon^2(\boldsymbol{I}_T + \frac{T\sigma_\alpha^2+\sigma_\varepsilon^2-\sigma_\varepsilon^2}{\sigma_\varepsilon^2}T^{-1}\boldsymbol{e}\boldsymbol{e}') \\
&= \sigma_\varepsilon^2(\boldsymbol{I}_T + (\frac{T\sigma_\alpha^2+\sigma_\varepsilon^2}{\sigma_\varepsilon^2}-1)T^{-1}\boldsymbol{e}\boldsymbol{e}') \\
&= \sigma_\varepsilon^2(\boldsymbol{I}_T + (\frac{1}{\psi^2}-1)T^{-1}\boldsymbol{e}\boldsymbol{e}') \\
&= \sigma_\varepsilon^2(\boldsymbol{I}_T + \frac{1}{\psi^2}T^{-1}\boldsymbol{e}\boldsymbol{e}' - T^{-1}\boldsymbol{e}\boldsymbol{e}') \\
&= \sigma_\varepsilon^2(\boldsymbol{I}_T - T^{-1}\boldsymbol{e}\boldsymbol{e}' + \frac{1}{\psi^2}(T^{-1}\boldsymbol{e}\boldsymbol{e}' - \boldsymbol{I}_T + \boldsymbol{I}_T)) \\
&= \sigma_\varepsilon^2(\boldsymbol{Q} + \frac{1}{\psi^2}(\boldsymbol{I}_T - \boldsymbol{Q}))
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\Omega}^{-1} &= [\sigma_\varepsilon^2(\boldsymbol{Q} + \frac{1}{\psi^2}(\boldsymbol{I}_T - \boldsymbol{Q}))]^{-1} \\
&= \sigma_\varepsilon^{-2}(\boldsymbol{Q}^- + \psi^2(\boldsymbol{I}_T^{-1} - \boldsymbol{Q}^-)) \\
&= \sigma_\varepsilon^{-2}(\boldsymbol{Q} + \psi^2(\boldsymbol{I}_T - \boldsymbol{Q}))
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\Omega}^{-1/2} &= \sigma_\varepsilon^{-1}(\boldsymbol{Q} + \psi(\boldsymbol{I}_T - \boldsymbol{Q})) \\
\sigma_\varepsilon\boldsymbol{\Omega}^{-1/2} &= (\boldsymbol{Q} + \psi(\boldsymbol{I}_T - \boldsymbol{Q}))
\end{aligned}$$

$$\sigma_\varepsilon\boldsymbol{\Omega}^{-1/2}\boldsymbol{y}_i = \sigma_\varepsilon\boldsymbol{\Omega}^{-1/2}(\boldsymbol{X}_i\boldsymbol{\beta} + (\boldsymbol{e}\alpha_i + \boldsymbol{\varepsilon}_i)) = \sigma_\varepsilon\boldsymbol{\Omega}^{-1/2}(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{u}_i) = \sigma_\varepsilon\boldsymbol{\Omega}^{-1/2}\boldsymbol{X}_i\boldsymbol{\beta} + \sigma_\varepsilon\boldsymbol{\Omega}^{-1/2}\boldsymbol{u}_i$$

So, $Var(\sigma_\varepsilon\boldsymbol{\Omega}^{-1/2}\boldsymbol{u}_i|\boldsymbol{X}_i) = \sigma_\varepsilon\boldsymbol{\Omega}^{-1/2}Var(\boldsymbol{u}_i|\boldsymbol{X}_i)\sigma_\varepsilon\boldsymbol{\Omega}'^{-1/2} = \sigma_\varepsilon^2\boldsymbol{\Omega}^{-1/2}\boldsymbol{\Omega}\boldsymbol{\Omega}^{-1/2} = \sigma_\varepsilon^2\boldsymbol{\Omega}^{-1/2}\boldsymbol{\Omega}^{1/2}\boldsymbol{\Omega}^{1/2}\boldsymbol{\Omega}^{-1/2} = \sigma_\varepsilon^2\boldsymbol{I}_T$

$$(\boldsymbol{Q} + \psi(\boldsymbol{I}_T - \boldsymbol{Q}))\boldsymbol{y}_i = (\boldsymbol{Q} + \psi(\boldsymbol{I}_T - \boldsymbol{Q}))\boldsymbol{X}_i\boldsymbol{\beta} + (\boldsymbol{Q} + \psi(\boldsymbol{I}_T - \boldsymbol{Q}))\boldsymbol{e}\alpha_i + (\boldsymbol{Q} + \psi(\boldsymbol{I}_T - \boldsymbol{Q}))\boldsymbol{\varepsilon}_i \qquad \text{Level 2}$$

It can also be written as

$$\boldsymbol{y}_i - \lambda \boldsymbol{e}\bar{y}_i = (\boldsymbol{X}_i - \lambda \boldsymbol{e}\bar{\boldsymbol{x}}_i')\boldsymbol{\beta} + (1-\lambda)\boldsymbol{e}\alpha_i + (\boldsymbol{\varepsilon}_i - \lambda \boldsymbol{e}\bar{\varepsilon}_i) \qquad \text{Level 2}$$

where $\lambda = 1 - \psi = 1 - \frac{\sigma_\varepsilon}{\sqrt{T\sigma_\alpha^2 + \sigma_\varepsilon^2}}$. It is because

$$\begin{aligned}
\sigma_\varepsilon \boldsymbol{\Omega}^{-1/2}\boldsymbol{y}_i = (\boldsymbol{Q} + \psi(\boldsymbol{I}_T - \boldsymbol{Q}))\boldsymbol{y}_i &= \boldsymbol{Q}\boldsymbol{y}_i + \psi(\boldsymbol{I}_T\boldsymbol{y}_i - \boldsymbol{Q}\boldsymbol{y}_i) \\
&= \boldsymbol{y}_i - \boldsymbol{e}\bar{y}_i + \psi(\boldsymbol{y}_i - \boldsymbol{y}_i + \boldsymbol{e}\bar{y}_i) \\
&= \boldsymbol{y}_i - \boldsymbol{e}\bar{y}_i + \psi\boldsymbol{e}\bar{y}_i \\
&= \boldsymbol{y}_i - \boldsymbol{e}\bar{y}_i(1-\psi) \\
&= \boldsymbol{y}_i - \lambda\boldsymbol{e}\bar{y}_i
\end{aligned}$$

$$\begin{aligned}
\sigma_\varepsilon \boldsymbol{\Omega}^{-1/2}\boldsymbol{X}_i\boldsymbol{\beta} = (\boldsymbol{Q} + \psi(\boldsymbol{I}_T - \boldsymbol{Q}))\boldsymbol{X}_i\boldsymbol{\beta} &= \boldsymbol{Q}\boldsymbol{X}_i\boldsymbol{\beta} + \psi(\boldsymbol{I}_T\boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{Q}\boldsymbol{X}_i\boldsymbol{\beta}) \\
&= (\boldsymbol{X}_i - \boldsymbol{e}\bar{\boldsymbol{x}}_i')\boldsymbol{\beta} + \psi(\boldsymbol{X}_i\boldsymbol{\beta} - (\boldsymbol{X}_i - \boldsymbol{e}\bar{\boldsymbol{x}}_i')\boldsymbol{\beta}) \\
&= (\boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{e}\bar{\boldsymbol{x}}_i'\boldsymbol{\beta}) + \psi(\boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{e}\bar{\boldsymbol{x}}_i'\boldsymbol{\beta}) \\
&= \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{e}\bar{\boldsymbol{x}}_i'\boldsymbol{\beta} + \psi\boldsymbol{e}\bar{\boldsymbol{x}}_i'\boldsymbol{\beta} \\
&= (\boldsymbol{X}_i - \boldsymbol{e}\bar{\boldsymbol{x}}_i' + \psi\boldsymbol{e}\bar{\boldsymbol{x}}_i')\boldsymbol{\beta} \\
&= (\boldsymbol{X}_i - \boldsymbol{e}\bar{\boldsymbol{x}}_i'(1-\psi))\boldsymbol{\beta} \\
&= (\boldsymbol{X}_i - \lambda\boldsymbol{e}\bar{\boldsymbol{x}}_i')\boldsymbol{\beta}
\end{aligned}$$

$$\begin{aligned}
\sigma_\varepsilon \boldsymbol{\Omega}^{-1/2}\boldsymbol{e}\alpha_i = (\boldsymbol{Q} + \psi(\boldsymbol{I}_T - \boldsymbol{Q}))\boldsymbol{e}\alpha_i &= \boldsymbol{Q}\boldsymbol{e}\alpha_i + \psi(\boldsymbol{I}_T\boldsymbol{e}\alpha_i - \boldsymbol{Q}\boldsymbol{e}\alpha_i) \\
&= \boldsymbol{0}\alpha_i + \psi(\boldsymbol{e}\alpha_i - \boldsymbol{0}\alpha_i) \\
&= \psi\boldsymbol{e}\alpha_i \\
&= (1-\lambda)\boldsymbol{e}\alpha_i
\end{aligned}$$

Random effect estimator is the OLS estimator of the beta in the transformed model $\boldsymbol{y}_i - \lambda\boldsymbol{e}\bar{y}_i = (\boldsymbol{X}_i - \lambda\boldsymbol{e}\bar{\boldsymbol{x}}_i')\boldsymbol{\beta} + (1-\lambda)\boldsymbol{e}\alpha_i + (\boldsymbol{\varepsilon}_i - \lambda\boldsymbol{e}\bar{\varepsilon}_i)$.

Fixed effect / within estimator is the OLS estimator of the beta in the transformed model $\boldsymbol{y}_i - \boldsymbol{e}\bar{y}_i = (\boldsymbol{X}_i - \boldsymbol{e}\bar{\boldsymbol{x}}_i')\boldsymbol{\beta} + (\boldsymbol{\varepsilon}_i - \boldsymbol{e}\bar{\varepsilon}_i)$.

Pooled OLS estimator is the OLS estimator of the beta in the original model $\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{e}\alpha_i + \boldsymbol{\varepsilon}_i$.

As $T \to \infty$, $\lambda \to 1$, $\boldsymbol{y}_i - \lambda\boldsymbol{e}\bar{y}_i = (\boldsymbol{X}_i - \lambda\boldsymbol{e}\bar{\boldsymbol{x}}_i')\boldsymbol{\beta} + (1-\lambda)\boldsymbol{e}\alpha_i + (\boldsymbol{\varepsilon}_i - \lambda\boldsymbol{e}\bar{\varepsilon}_i)$ converges to $\boldsymbol{y}_i - \boldsymbol{e}\bar{y}_i = (\boldsymbol{X}_i - \boldsymbol{e}\bar{\boldsymbol{x}}_i')\boldsymbol{\beta} + (\boldsymbol{\varepsilon}_i - \boldsymbol{e}\bar{\varepsilon}_i)$ Thus, random effect estimator converges to fixed effect / within estimator as $T \to \infty$.

As $\sigma_\alpha^2 \to 0$, $\lambda \to 0$, $\boldsymbol{y}_i - \lambda\boldsymbol{e}\bar{y}_i = (\boldsymbol{X}_i - \lambda\boldsymbol{e}\bar{\boldsymbol{x}}_i')\boldsymbol{\beta} + (1-\lambda)\boldsymbol{e}\alpha_i + (\boldsymbol{\varepsilon}_i - \lambda\boldsymbol{e}\bar{\varepsilon}_i)$ converges to $\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{e}\alpha_i + \boldsymbol{\varepsilon}_i$ Thus, random effect estimator converges to pooled OLS estimator as $\sigma_\alpha^2 \to 0$.

$$\begin{aligned}
\begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix} - \lambda \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \bar{y}_i &= (\begin{pmatrix} \boldsymbol{x}_{i1}' \\ \vdots \\ \boldsymbol{x}_{iT}' \end{pmatrix} - \lambda \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \bar{\boldsymbol{x}}_i')\boldsymbol{\beta} + (1-\lambda) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \alpha_i + (\begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iT} \end{pmatrix} - \lambda \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \bar{\varepsilon}_i) \\
\begin{pmatrix} y_{i1} - \lambda\bar{y}_i \\ \vdots \\ y_{iT} - \lambda\bar{y}_i \end{pmatrix} &= \begin{pmatrix} \boldsymbol{x}_{i1}' - \lambda\bar{\boldsymbol{x}}_i' \\ \vdots \\ \boldsymbol{x}_{iT}' - \lambda\bar{\boldsymbol{x}}_i' \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} (1-\lambda)\alpha_i \\ \vdots \\ (1-\lambda)\alpha_i \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} - \lambda\bar{\varepsilon}_i \\ \vdots \\ \varepsilon_{iT} - \lambda\bar{\varepsilon}_i \end{pmatrix} \\
\begin{pmatrix} y_{i1} - \lambda\bar{y}_i \\ \vdots \\ y_{iT} - \lambda\bar{y}_i \end{pmatrix} &= \begin{pmatrix} (\boldsymbol{x}_{i1} - \lambda\bar{\boldsymbol{x}}_i)' \\ \vdots \\ (\boldsymbol{x}_{iT} - \lambda\bar{\boldsymbol{x}}_i)' \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} (1-\lambda)\alpha_i \\ \vdots \\ (1-\lambda)\alpha_i \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} - \lambda\bar{\varepsilon}_i \\ \vdots \\ \varepsilon_{iT} - \lambda\bar{\varepsilon}_i \end{pmatrix} \\
y_{it} - \lambda\bar{y}_i &= (\boldsymbol{x}_{it} - \lambda\bar{\boldsymbol{x}}_i)'\boldsymbol{\beta} + \underbrace{(1-\lambda)\alpha_i + (\varepsilon_{it} - \lambda\bar{\varepsilon}_i)}_{v_{it}} \qquad \text{Level 1}
\end{aligned}$$

2.3.2 OLS estimator of the GLS transformed model i.e., Random Effect / GLS estimator

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{re}^{ols} &= [\sum_{i=1}^{N}(\boldsymbol{X}_i-\lambda\boldsymbol{e}\bar{\boldsymbol{x}}_i')'(\boldsymbol{X}_i-\lambda\boldsymbol{e}\bar{\boldsymbol{x}}_i')]^{-1}\sum_{i=1}^{N}(\boldsymbol{X}_i-\lambda\boldsymbol{e}\bar{\boldsymbol{x}}_i')'(\boldsymbol{y}_i-\lambda\boldsymbol{e}\bar{y}_i) && \text{Level 2}\\
&= [\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\lambda\bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{it}-\lambda\bar{\boldsymbol{x}}_i)']^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\lambda\bar{\boldsymbol{x}}_i)(y_{it}-\lambda\bar{y}_i) && \text{Level 1}
\end{aligned}$$

If $\boldsymbol{x}_{it}$ is replaced by $\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{x}}_i$ is replaced by $\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}}$,

$$\begin{aligned}
(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}})-\lambda(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}}) &= \boldsymbol{x}_{it}-\bar{\boldsymbol{x}}-\lambda\bar{\boldsymbol{x}}_i+\lambda\bar{\boldsymbol{x}}\\
&= \boldsymbol{x}_{it}-\bar{\boldsymbol{x}}-(1-\psi)\bar{\boldsymbol{x}}_i+(1-\psi)\bar{\boldsymbol{x}}\\
&= \boldsymbol{x}_{it}-\bar{\boldsymbol{x}}-\bar{\boldsymbol{x}}_i+\psi\bar{\boldsymbol{x}}_i+\bar{\boldsymbol{x}}-\psi\bar{\boldsymbol{x}}\\
&= (\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)+\psi(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})\\
\sum_{i=1}^{N}\sum_{t=1}^{T}((\boldsymbol{x}_{it}-\bar{\boldsymbol{x}})-\lambda(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}}))((\boldsymbol{x}_{it}-\bar{\boldsymbol{x}})-\lambda(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}}))' &= \sum_{i=1}^{N}\sum_{t=1}^{T}((\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)+\psi(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}}))((\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)+\psi(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}}))'\\
&= \sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)'+\psi\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})'+\\
&\quad \psi\sum_{i=1}^{N}\sum_{t=1}^{T}(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)'+\psi^2\sum_{i=1}^{N}\sum_{t=1}^{T}(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})'\\
&= \sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)'+\psi^2\sum_{i=1}^{N}\sum_{t=1}^{T}(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})'
\end{aligned}$$

It is because

$$\begin{aligned}
\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})' &= \sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})'-\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{\boldsymbol{x}}_i(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})'\\
&= \sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})'-\sum_{i=1}^{N}T\bar{\boldsymbol{x}}_i(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})'\\
&= \sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})'-\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})'\\
&= \boldsymbol{0}
\end{aligned}$$

Similarly, if y_{it} is replaced by $y_{it}-\bar{y}$ and $\bar{y}_i$ is replaced by $\bar{y}_i-\bar{y}$

$$\sum_{i=1}^{N}\sum_{t=1}^{T}((\boldsymbol{x}_{it}-\bar{\boldsymbol{x}})-\lambda(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}}))((y_{it}-\bar{y})-\lambda(\bar{x}_i-\bar{y})) = \sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(y_{it}-\bar{y}_i)+\psi^2\sum_{i=1}^{N}\sum_{t=1}^{T}(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})(\bar{y}_i-\bar{y})$$

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{re}^{ols} = (&\overbrace{\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)'}^{Within}+\psi^2\overbrace{\sum_{i=1}^{N}\sum_{t=1}^{T}(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})'}^{Between})^{-1}\\
&(\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(y_{it}-\bar{y}_i)+\psi^2\sum_{i=1}^{N}\sum_{t=1}^{T}(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})(\bar{y}_i-\bar{y}))
\end{aligned}$$

If $T\to\infty$, $\psi^2\to 0$, $\hat{\boldsymbol{\beta}}_{re}^{ols}\to\hat{\boldsymbol{\beta}}_{within}^{ols}$

If $\sigma_\alpha^2 \to 0$, $\psi^2 \to 1$, $\hat{\boldsymbol{\beta}}_{re}^{ols} \to \hat{\boldsymbol{\beta}}_{pool}^{ols}$ It is because

$$\begin{aligned}
\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}})(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}})' &= \sum_{i=1}^{N}\sum_{t=1}^{T}((\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)+(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}}))((\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)+(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}}))' \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)' + \sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})' + \\
&\quad \sum_{i=1}^{N}\sum_{t=1}^{T}(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)' + \sum_{i=1}^{N}\sum_{t=1}^{T}(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})' \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)' + \sum_{i=1}^{N}\sum_{t=1}^{T}(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})'
\end{aligned}$$

Similarly,

$$\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}})(y_{it}-\bar{y}) = \sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(y_{it}-\bar{y}_i) + \sum_{i=1}^{N}\sum_{t=1}^{T}(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})(\bar{y}_i-\bar{y})$$

Thus,

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{pool}^{ols} &= (\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}})(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}})')^{-1}(\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}})(y_{it}-\bar{y})) \\
&= (\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)' + \sum_{i=1}^{N}\sum_{t=1}^{T}(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})')^{-1} \\
&\quad (\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(y_{it}-\bar{y}_i) + \sum_{i=1}^{N}\sum_{t=1}^{T}(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})(\bar{y}_i-\bar{y}))
\end{aligned}$$

So, pooled OLS estimator is an inefficient weighted average of within and between effects. RE estimator is an efficient weighted average of within and between effects. As RE model assumes $\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i \sim iid\ [\boldsymbol{0}, \sigma_\varepsilon^2 \boldsymbol{I}_T]$,

$$Var(\widehat{\boldsymbol{\beta}}_{re}^{ols}) = \sigma_\varepsilon^2[\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\lambda\bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{it}-\lambda\bar{\boldsymbol{x}}_i)']^{-1}$$

2.3.3 Between effect model and estimation of σ_α^2

$$\begin{aligned}
\bar{y}_i &= \bar{\boldsymbol{x}}_i'\boldsymbol{\beta} + \overbrace{\alpha_i + \bar{\varepsilon}_i}^{v_i} \\
\sigma_B^2 = Var(v_i) &= Var(\alpha_i + \bar{\varepsilon}_i) \\
&= Var(\alpha_i) + Var(\bar{\varepsilon}_i) \\
&= Var(\alpha_i) + T^{-1}Var(\varepsilon_{it}) && \text{as } \varepsilon_{it} \text{ is serially uncorrelated}
\end{aligned}$$

$$\underbrace{Var(\alpha_i)}_{\sigma_\alpha^2} = \underbrace{Var(v_i)}_{\sigma_B^2} - T^{-1}\underbrace{Var(\varepsilon_{it})}_{\sigma_\varepsilon^2}$$

3 GMM Estimation of Linear Panel Model

3.1 Linear panel model

$$\begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}'_{i1} \\ \vdots \\ \boldsymbol{x}'_{iT} \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{pmatrix}$$
$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{u}_i$$

3.2 Exogeneity assumption

$$\mathbb{E}(\boldsymbol{Z}'_i \boldsymbol{u}_i) = \boldsymbol{0}$$

$\boldsymbol{Z}_i$ is a T x r matrix. r is the number of exogeneous variables in $\boldsymbol{X}_i$ plus the number of instrumental variables for endogeneous variables in $\boldsymbol{X}_i$. In GMM context, r is also the number of moment conditions.

K is the number of parameters.

$r \geq K$. If $r = K$, the model is just-identified, GMM is the same as MM; if $r > K$, the model is over-identified.

3.2.1 Summation assumption

The weakest exogeneity assumption

$$\boldsymbol{Z}_i = \begin{pmatrix} \boldsymbol{z}'_{i1} \\ \vdots \\ \boldsymbol{z}'_{iT} \end{pmatrix}$$

$$\mathbb{E}(\boldsymbol{Z}'_i \boldsymbol{u}_i) = \mathbb{E}(\begin{pmatrix} \boldsymbol{z}'_{i1} \\ \vdots \\ \boldsymbol{z}'_{iT} \end{pmatrix}' \begin{pmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{pmatrix}) = \mathbb{E}(\begin{pmatrix} \boldsymbol{z}_{i1} & \cdots & \boldsymbol{z}_{iT} \end{pmatrix} \begin{pmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{pmatrix}) = \mathbb{E}(\sum_{t=1}^{T} \boldsymbol{z}_{it} u_{it}) = \boldsymbol{0}$$

3.2.2 Contemporaneous exogeneity assumption

Stronger

$$\boldsymbol{Z}_i = \begin{pmatrix} \boldsymbol{z}'_{i1} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{z}'_{iT} \end{pmatrix}$$

$$\begin{aligned} \mathbb{E}(\boldsymbol{Z}'_i \boldsymbol{u}_i) &= \mathbb{E}(\begin{pmatrix} \boldsymbol{z}'_{i1} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{z}'_{iT} \end{pmatrix}' \begin{pmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{pmatrix}) \\ &= \mathbb{E}(\begin{pmatrix} \boldsymbol{z}_{i1} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{z}_{iT} \end{pmatrix} \begin{pmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{pmatrix}) \\ &= \mathbb{E}(\begin{pmatrix} \boldsymbol{z}_{i1} u_{i1} \\ \vdots \\ \boldsymbol{z}_{iT} u_{iT} \end{pmatrix}) \\ &= \begin{pmatrix} \mathbb{E}(\boldsymbol{z}_{i1} u_{i1}) \\ \vdots \\ \mathbb{E}(\boldsymbol{z}_{iT} u_{iT}) \end{pmatrix} = \boldsymbol{0} \end{aligned}$$

3.2.3 Weak/sequential exogeneity assumption

Stronger

$$\boldsymbol{Z}_i = \begin{pmatrix} \boldsymbol{z}_{i1}' & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \begin{pmatrix} \boldsymbol{z}_{i1}' & \boldsymbol{z}_{i2}' \end{pmatrix} & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \begin{pmatrix} \boldsymbol{z}_{i1}' & \cdots & \boldsymbol{z}_{iT}' \end{pmatrix} \end{pmatrix}$$

$$\begin{aligned}
\mathbb{E}(\boldsymbol{Z}_i'\boldsymbol{u}_i) &= \mathbb{E}(\begin{pmatrix} \boldsymbol{z}_{i1}' & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \begin{pmatrix} \boldsymbol{z}_{i1}' & \boldsymbol{z}_{i2}' \end{pmatrix} & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \begin{pmatrix} \boldsymbol{z}_{i1}' & \cdots & \boldsymbol{z}_{iT}' \end{pmatrix} \end{pmatrix}' \begin{pmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{pmatrix}) \\
&= \mathbb{E}(\begin{pmatrix} \boldsymbol{z}_{i1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \begin{pmatrix} \boldsymbol{z}_{i1} \\ \boldsymbol{z}_{i2} \end{pmatrix} & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \begin{pmatrix} \boldsymbol{z}_{i1} \\ \vdots \\ \boldsymbol{z}_{iT} \end{pmatrix} \end{pmatrix} \begin{pmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{pmatrix}) \\
&= \mathbb{E}(\begin{pmatrix} \boldsymbol{z}_{i1}u_{i1} \\ \begin{pmatrix} \boldsymbol{z}_{i1}u_{i2} \\ \boldsymbol{z}_{i2}u_{i2} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \boldsymbol{z}_{i1}u_{iT} \\ \vdots \\ \boldsymbol{z}_{iT}u_{iT} \end{pmatrix} \end{pmatrix}) = \begin{pmatrix} \mathbb{E}(\boldsymbol{z}_{i1}u_{i1}) \\ \begin{pmatrix} \mathbb{E}(\boldsymbol{z}_{i1}u_{i2}) \\ \mathbb{E}(\boldsymbol{z}_{i2}u_{i2}) \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \mathbb{E}(\boldsymbol{z}_{i1}u_{iT}) \\ \vdots \\ \mathbb{E}(\boldsymbol{z}_{iT}u_{iT}) \end{pmatrix} \end{pmatrix} = \boldsymbol{0}
\end{aligned}$$

which is equivalent as $\mathbb{E}(\boldsymbol{z}_{is}u_{it}) = \boldsymbol{0}$ for $s \leq t$.

Strong form of sequential exogeneity $\mathbb{E}(u_{it}|\boldsymbol{z}_{it}, \cdots, \boldsymbol{z}_{i1}) = 0$ implies weak form of sequential exogeneity $\mathbb{E}(\boldsymbol{z}_{is}u_{it}) = \boldsymbol{0}$ for $s \leq t$ as $\mathbb{E}(\boldsymbol{z}_{is}u_{it}) = \mathbb{E}(\mathbb{E}(\boldsymbol{z}_{is}u_{it}|\boldsymbol{z}_{it}, \cdots, \boldsymbol{z}_{i1})) = \mathbb{E}(\boldsymbol{z}_{is}\underbrace{\mathbb{E}(u_{it}|\boldsymbol{z}_{it}, \cdots, \boldsymbol{z}_{i1})}_{0}) = \boldsymbol{0}$ for $s \leq t$.

It also implies $Cov(\boldsymbol{z}_{is}, u_{it}) = \boldsymbol{0}$ for $s \leq t$ as $Cov(\boldsymbol{z}_{is}, u_{it}) = \underbrace{\mathbb{E}(\boldsymbol{z}_{is}u_{it})}_{\boldsymbol{0}} - \mathbb{E}(\boldsymbol{z}_{is})\mathbb{E}(u_{it}) = -\mathbb{E}(\boldsymbol{z}_{is})\mathbb{E}(\underbrace{\mathbb{E}(u_{it}|\boldsymbol{z}_{it}, \cdots, \boldsymbol{z}_{i1})}_{0}) = \boldsymbol{0}$ for $s \leq t$.

3.2.4 Strong/strict exogeneity assumption

The strongest exogeneity assumption

$$\boldsymbol{Z}_i = \begin{pmatrix} \begin{pmatrix} \boldsymbol{z}_{i1}' & \cdots & \boldsymbol{z}_{iT}' \end{pmatrix} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \begin{pmatrix} \boldsymbol{z}_{i1}' & \cdots & \boldsymbol{z}_{iT}' \end{pmatrix} & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \begin{pmatrix} \boldsymbol{z}_{i1}' & \cdots & \boldsymbol{z}_{iT}' \end{pmatrix} \end{pmatrix}$$

$$\mathbb{E}(\boldsymbol{Z}_i'\boldsymbol{u}_i) = \mathbb{E}(\begin{pmatrix} \begin{pmatrix}\boldsymbol{z}_{i1}' & \cdots & \boldsymbol{z}_{iT}'\end{pmatrix} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \begin{pmatrix}\boldsymbol{z}_{i1}' & \cdots & \boldsymbol{z}_{iT}'\end{pmatrix} & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \begin{pmatrix}\boldsymbol{z}_{i1}' & \cdots & \boldsymbol{z}_{iT}'\end{pmatrix} \end{pmatrix}' \begin{pmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{pmatrix})$$

$$= \mathbb{E}(\begin{pmatrix} \begin{pmatrix}\boldsymbol{z}_{i1} \\ \vdots \\ \boldsymbol{z}_{iT}\end{pmatrix} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \begin{pmatrix}\boldsymbol{z}_{i1} \\ \vdots \\ \boldsymbol{z}_{iT}\end{pmatrix} & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \begin{pmatrix}\boldsymbol{z}_{i1} \\ \vdots \\ \boldsymbol{z}_{iT}\end{pmatrix} \end{pmatrix} \begin{pmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{pmatrix})$$

$$= \mathbb{E}(\begin{pmatrix} \begin{pmatrix}\boldsymbol{z}_{i1}u_{i1} \\ \vdots \\ \boldsymbol{z}_{iT}u_{i1}\end{pmatrix} \\ \begin{pmatrix}\boldsymbol{z}_{i1}u_{i2} \\ \vdots \\ \boldsymbol{z}_{iT}u_{i2}\end{pmatrix} \\ \vdots \\ \begin{pmatrix}\boldsymbol{z}_{i1}u_{iT} \\ \vdots \\ \boldsymbol{z}_{iT}u_{iT}\end{pmatrix} \end{pmatrix})$$

$$= \begin{pmatrix} \begin{pmatrix}\mathbb{E}(\boldsymbol{z}_{i1}u_{i1}) \\ \vdots \\ \mathbb{E}(\boldsymbol{z}_{iT}u_{i1})\end{pmatrix} \\ \begin{pmatrix}\mathbb{E}(\boldsymbol{z}_{i1}u_{i2}) \\ \vdots \\ \mathbb{E}(\boldsymbol{z}_{iT}u_{i2})\end{pmatrix} \\ \vdots \\ \begin{pmatrix}\mathbb{E}(\boldsymbol{z}_{i1}u_{iT}) \\ \vdots \\ \mathbb{E}(\boldsymbol{z}_{iT}u_{iT})\end{pmatrix} \end{pmatrix} = \boldsymbol{0}$$

which is equivalent as $\mathbb{E}(\boldsymbol{z}_{is}u_{it}) = \boldsymbol{0}$ for $s = 1, \cdots, T$

Strong form of strict exogeneity $\mathbb{E}(u_{it}|\boldsymbol{z}_{i1}, \cdots, \boldsymbol{z}_{iT}) = 0$ implies weak form of strict exogeneity $\mathbb{E}(\boldsymbol{z}_{is}u_{it}) = \boldsymbol{0}$ for $s = 1, \cdots, T$. Since for $s = 1, \cdots, T$,

$$\begin{aligned} \mathbb{E}(\boldsymbol{z}_{is}u_{it}) &= \mathbb{E}(\mathbb{E}(\boldsymbol{z}_{is}u_{it}|\boldsymbol{z}_{i1}, \cdots, \boldsymbol{z}_{iT})) \\ &= \mathbb{E}(\boldsymbol{z}_{is}\underbrace{\mathbb{E}(u_{it}|\boldsymbol{z}_{i1}, \cdots, \boldsymbol{z}_{iT})}_{0}) \\ &= \boldsymbol{0} \end{aligned}$$

3.3 GMM Estimator of Linear Panel Model

3.3.1 Unconditional moment condition

$$\mathbb{E}(\boldsymbol{Z}_i'\boldsymbol{u}_i) = \mathbb{E}(\boldsymbol{Z}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_0)) = \boldsymbol{0}$$

where $\boldsymbol{\beta}_0$ is the true population parameter. So, $g(\boldsymbol{d}_i; \boldsymbol{\theta}_0) = \boldsymbol{Z}_i'\boldsymbol{u}_i = \boldsymbol{Z}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_0)$

3.3.2 Objective/loss function

We want to find $\boldsymbol{\beta}$ from the parameter space such that the squared distance between $\sum_{i=1}^{N} \boldsymbol{Z}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})/N$ and $\mathbb{E}(\boldsymbol{Z}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_0))$ i.e.,

$$
\begin{aligned}
&[\rho(\sum_{i=1}^{N} \boldsymbol{Z}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})/N, \mathbb{E}(\boldsymbol{Z}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_0)))]^2 && \text{where } \rho(.) \text{ is a metric function} \\
&= ||\sum_{i=1}^{N} \boldsymbol{Z}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})/N - \mathbb{E}(\boldsymbol{Z}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_0))||^2 \\
&= (\sum_{i=1}^{N} \boldsymbol{Z}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})/N - \underbrace{\mathbb{E}(\boldsymbol{Z}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_0))}_{\mathbf{0}})'\boldsymbol{W}_N(\sum_{i=1}^{N} \boldsymbol{Z}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})/N - \underbrace{\mathbb{E}(\boldsymbol{Z}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_0))}_{\mathbf{0}}) \\
&= (\sum_{i=1}^{N} \boldsymbol{Z}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})/N)'\boldsymbol{W}_N(\sum_{i=1}^{N} \boldsymbol{Z}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})/N) \geq 0 && \text{as distance cannot be negative}
\end{aligned}
$$

is as close to the zero as possible. The distance is a function of $\boldsymbol{\beta}$ i.e.,

$$
Q_N(\boldsymbol{\beta}) := (\sum_{i=1}^{N} \boldsymbol{Z}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})/N)'\boldsymbol{W}_N(\sum_{i=1}^{N} \boldsymbol{Z}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})/N) \geq 0
$$

If $\boldsymbol{W}_N$ is symmetric and positive definite, then $Q_N(\boldsymbol{\beta})$ is strictly convex. So, first order condition becomes sufficient and there is an unique minimizer.

3.3.3 Gradient vector

$$
\begin{aligned}
\triangledown Q_N(\boldsymbol{\beta}) = \frac{\partial Q_N(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \frac{\partial(\sum_{i=1}^{N} \boldsymbol{Z}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})/N)'\boldsymbol{W}_N(\sum_{i=1}^{N} \boldsymbol{Z}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})/N)}{\partial \boldsymbol{\beta}} \\
&= 2(\frac{\partial(\sum_{i=1}^{N} \boldsymbol{Z}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})/N)}{\partial \boldsymbol{\beta}'})'\boldsymbol{W}_N(\sum_{i=1}^{N} \boldsymbol{Z}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})/N) \\
&= 2[\sum_{i=1}^{N}(\frac{\partial \boldsymbol{Z}_i'\boldsymbol{y}_i}{\partial \boldsymbol{\beta}'} - \frac{\partial \boldsymbol{Z}_i'\boldsymbol{X}_i\boldsymbol{\beta}}{\partial \boldsymbol{\beta}'})/N]'\boldsymbol{W}_N(\sum_{i=1}^{N} \boldsymbol{Z}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})/N) \\
&= 2[\sum_{i=1}^{N} -\frac{\partial \boldsymbol{Z}_i'\boldsymbol{X}_i\boldsymbol{\beta}}{\partial \boldsymbol{\beta}'}/N]'\boldsymbol{W}_N(\sum_{i=1}^{N} \boldsymbol{Z}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})/N) \\
&= -2(1/N^2)\sum_{i=1}^{N}(\boldsymbol{Z}_i'\boldsymbol{X}_i)'\boldsymbol{W}_N\sum_{i=1}^{N}(\boldsymbol{Z}_i'\boldsymbol{y}_i - \boldsymbol{Z}_i'\boldsymbol{X}_i\boldsymbol{\beta}) \\
&= -2(1/N^2)\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Z}_i''\boldsymbol{W}_N(\sum_{i=1}^{N}\boldsymbol{Z}_i'\boldsymbol{y}_i - \sum_{i=1}^{N}\boldsymbol{Z}_i'\boldsymbol{X}_i\boldsymbol{\beta}) \\
&= -2(1/N^2)[(\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Z}_i)\boldsymbol{W}_N(\sum_{i=1}^{N}\boldsymbol{Z}_i'\boldsymbol{y}_i) - (\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{Z}_i)\boldsymbol{W}_N(\sum_{i=1}^{N}\boldsymbol{Z}_i'\boldsymbol{X}_i)\boldsymbol{\beta}]
\end{aligned}
$$

If $r = K$, both $(\frac{\partial(\sum_{i=1}^{N} \boldsymbol{Z}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})/N)}{\partial \boldsymbol{\beta}'})'$ and $\boldsymbol{W}_N$ are square matrixes and invertible. In this case, FOC is $\triangledown Q_N(\widehat{\boldsymbol{\beta}}_{pmm}) = \sum_{i=1}^{N} \boldsymbol{Z}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}_{pmm})/N = \mathbf{0}$ which is MM estimation.

3.3.4 First order condition

$$
\begin{aligned}
-2(1/N^2)[(\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{Z}_i)\boldsymbol{W}_N(\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{y}_i) - (\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{Z}_i)\boldsymbol{W}_N(\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{X}_i)\widehat{\boldsymbol{\beta}}_{pgmm}] &= \boldsymbol{0} \\
(\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{Z}_i)\boldsymbol{W}_N(\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{y}_i) - (\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{Z}_i)\boldsymbol{W}_N(\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{X}_i)\widehat{\boldsymbol{\beta}}_{pgmm} &= \boldsymbol{0} \\
(\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{Z}_i)\boldsymbol{W}_N(\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{y}_i) &= (\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{Z}_i)\boldsymbol{W}_N(\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{X}_i)\widehat{\boldsymbol{\beta}}_{pgmm} \\
[(\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{Z}_i)\boldsymbol{W}_N(\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{X}_i)]^{-1}(\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{Z}_i)\boldsymbol{W}_N(\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{y}_i) &= \widehat{\boldsymbol{\beta}}_{pgmm}
\end{aligned}
$$

Special case: if $\boldsymbol{W}_N = (\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{Z}_i)^{-1}$,

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}_{pgmm} &= [(\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{Z}_i)\overbrace{(\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{Z}_i)^{-1}(\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{X}_i)}^{\widehat{\boldsymbol{\Gamma}}_{2SLS}}]^{-1}\overbrace{(\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{Z}_i)(\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{Z}_i)^{-1}}^{\widehat{\boldsymbol{\Gamma}}_{2SLS}'}(\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{y}_i) \\
&= [\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{Z}_i\widehat{\boldsymbol{\Gamma}}_{2SLS}]^{-1}\widehat{\boldsymbol{\Gamma}}_{2SLS}'\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{y}_i \\
&= [\sum_{i=1}^N \boldsymbol{X}_i'\underbrace{\boldsymbol{Z}_i\widehat{\boldsymbol{\Gamma}}_{2SLS}}_{\widehat{\boldsymbol{X}}_i}]^{-1}\sum_{i=1}^N \underbrace{(\boldsymbol{Z}_i\widehat{\boldsymbol{\Gamma}}_{2SLS})'}_{\widehat{\boldsymbol{X}}_i'}\boldsymbol{y}_i = \widehat{\boldsymbol{\beta}}_{p2SLS}
\end{aligned}
$$

Special case: if $r = K$, the model is just-identified, GMM is the same as MM,

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}_{pmm} = \widehat{\boldsymbol{\beta}}_{pgmm} &= [(\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{Z}_i)\boldsymbol{W}_N(\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{X}_i)]^{-1}(\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{Z}_i)\boldsymbol{W}_N(\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{y}_i) \\
&= (\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{W}_N^{-1}(\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{Z}_i)^{-1}(\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{Z}_i)\boldsymbol{W}_N(\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{y}_i) \\
&= (\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{X}_i)^{-1}(\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{y}_i) = \widehat{\boldsymbol{\beta}}_{piv}
\end{aligned}
$$

Special case: if all regressors are exogeneous: $\boldsymbol{Z}_i = \boldsymbol{X}_i$ (which implies $r = K$),

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}_{pgmm} &= \widehat{\boldsymbol{\beta}}_{piv} \\
&= (\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}(\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{y}_i) = \widehat{\boldsymbol{\beta}}_{pols}
\end{aligned}
$$

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}_{pgmm} &= [(\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{Z}_i)\boldsymbol{W}_N(\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{X}_i)]^{-1}(\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{Z}_i)\boldsymbol{W}_N(\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{y}_i) \\
&= [\begin{pmatrix}\boldsymbol{X}_1' & \cdots & \boldsymbol{X}_N'\end{pmatrix}\begin{pmatrix}\boldsymbol{Z}_1 \\ \vdots \\ \boldsymbol{Z}_N\end{pmatrix}\boldsymbol{W}_N\begin{pmatrix}\boldsymbol{Z}_1' & \cdots & \boldsymbol{Z}_N'\end{pmatrix}\begin{pmatrix}\boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_N\end{pmatrix}]^{-1}\begin{pmatrix}\boldsymbol{X}_1' & \cdots & \boldsymbol{X}_N'\end{pmatrix}\begin{pmatrix}\boldsymbol{Z}_1 \\ \vdots \\ \boldsymbol{Z}_N\end{pmatrix}\boldsymbol{W}_N\begin{pmatrix}\boldsymbol{Z}_1' & \cdots & \boldsymbol{Z}_N'\end{pmatrix}\begin{pmatrix}\boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_N\end{pmatrix} \\
&= [\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_N\boldsymbol{Z}'\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_N\boldsymbol{Z}'\boldsymbol{y}
\end{aligned}
$$

3.4 Conditional variance of $\widehat{\boldsymbol{\beta}}_{pgmm}$

$$\begin{aligned}
Var(\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_N\boldsymbol{Z}'\boldsymbol{y}|\boldsymbol{X},\boldsymbol{Z}) &= \boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_N\boldsymbol{Z}'Var(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{Z})(\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_N\boldsymbol{Z}')' \\
&= \boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_N\boldsymbol{Z}'Var(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{u}|\boldsymbol{X},\boldsymbol{Z})(\boldsymbol{Z}''\boldsymbol{W}_N'\boldsymbol{Z}'\boldsymbol{X}'') \\
&= \boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_N\boldsymbol{Z}'Var(\boldsymbol{u}|\boldsymbol{X},\boldsymbol{Z})(\boldsymbol{Z}\boldsymbol{W}_N\boldsymbol{Z}'\boldsymbol{X}) \\
&= \boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_N Var(\boldsymbol{Z}'\boldsymbol{u}|\boldsymbol{X},\boldsymbol{Z})\boldsymbol{W}_N\boldsymbol{Z}'\boldsymbol{X} \\
&= \boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_N\mathbb{E}((\boldsymbol{Z}'\boldsymbol{u}-\mathbb{E}(\boldsymbol{Z}'\boldsymbol{u}|\boldsymbol{X},\boldsymbol{Z}))(\boldsymbol{Z}'\boldsymbol{u}-\mathbb{E}(\boldsymbol{Z}'\boldsymbol{u}|\boldsymbol{X},\boldsymbol{Z}))'|\boldsymbol{X},\boldsymbol{Z})\boldsymbol{W}_N\boldsymbol{Z}'\boldsymbol{X} \\
&= \boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_N\mathbb{E}((\boldsymbol{Z}'\boldsymbol{u})(\boldsymbol{Z}'\boldsymbol{u})'|\boldsymbol{X},\boldsymbol{Z})\boldsymbol{W}_N\boldsymbol{Z}'\boldsymbol{X} \\
&= \boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_N\mathbb{E}(\boldsymbol{Z}'\boldsymbol{u}\boldsymbol{u}'\boldsymbol{Z}''|\boldsymbol{X},\boldsymbol{Z})\boldsymbol{W}_N\boldsymbol{Z}'\boldsymbol{X}
\end{aligned}$$

$$\begin{aligned}
[\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_N\boldsymbol{Z}'\boldsymbol{X}]^{-1'} &= [\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_N\boldsymbol{Z}'\boldsymbol{X}]'^{-1} \\
&= [\boldsymbol{X}'\boldsymbol{Z}''\boldsymbol{W}_N'\boldsymbol{Z}'\boldsymbol{X}'']^{-1} \\
&= [\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_N\boldsymbol{Z}'\boldsymbol{X}]^{-1}
\end{aligned}$$

$$\begin{aligned}
Var(\widehat{\boldsymbol{\beta}}_{pgmm}|\boldsymbol{X},\boldsymbol{Z}) &= Var([\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_N\boldsymbol{Z}'\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_N\boldsymbol{Z}'\boldsymbol{y}|\boldsymbol{X},\boldsymbol{Z}) \\
&= [\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_N\boldsymbol{Z}'\boldsymbol{X}]^{-1}Var(\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_N\boldsymbol{Z}'\boldsymbol{y}|\boldsymbol{X},\boldsymbol{Z})[\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_N\boldsymbol{Z}'\boldsymbol{X}]^{-1'} \\
&= [\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_N\boldsymbol{Z}'\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_N\mathbb{E}(\boldsymbol{Z}'\boldsymbol{u}\boldsymbol{u}'Z|\boldsymbol{X},\boldsymbol{Z})\boldsymbol{W}_N\boldsymbol{Z}'\boldsymbol{X}[\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_N\boldsymbol{Z}'\boldsymbol{X}]^{-1}
\end{aligned}$$

4 GMM Estimation of Fixed Effect Model

$$
\begin{aligned}
y_{it} &= \boldsymbol{x}'_{it}\boldsymbol{\beta} + \alpha_i + \varepsilon_{it} \\
\boldsymbol{y}_i &= \boldsymbol{X}_i\boldsymbol{\beta} + \underbrace{(\boldsymbol{e}\alpha_i + \boldsymbol{\varepsilon}_i)}_{\boldsymbol{u}_i} \\
\boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I}_N \otimes \boldsymbol{e})\boldsymbol{\alpha} + \boldsymbol{\varepsilon}
\end{aligned}
$$

4.1 Assumption

α_i is potentially correlated with $\boldsymbol{X}_i$, so $\boldsymbol{u}_i$ is potentially correlated with $\boldsymbol{X}_i$

$\boldsymbol{\varepsilon}_i$ is also potentially correlated with $\boldsymbol{X}_i$, so $\boldsymbol{u}_i$ is potentially correlated with $\boldsymbol{X}_i$

Even after eliminating α_i by using any arbitrary operators $\boldsymbol{T}$, $\tilde{\boldsymbol{u}}_i := \boldsymbol{T}\boldsymbol{u}_i$ is still potentially correlated with $\tilde{\boldsymbol{X}}_i := \boldsymbol{T}\boldsymbol{X}_i$ because of the potential correlation between $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{X}_i$. Thus, $\tilde{\boldsymbol{X}}_i$ is potentially endogeneous.

If $\tilde{\boldsymbol{X}}_i$ is endogeneous, OLS estimation is inconsistent and biased. We should use IV estimation (for just-identified case) and 2SLS estimation (for over-identified case). IV and 2SLS estimation are special cases of GMM estimation.

4.2 GMM estimator of fixed effect model

There exists a $\boldsymbol{T}$ such that $\boldsymbol{T}\boldsymbol{e} = \boldsymbol{0}$.

4.2.1 Transformed model

$$
\begin{aligned}
\tilde{\boldsymbol{y}}_i &:= \boldsymbol{T}\boldsymbol{y}_i = \boldsymbol{T}(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{u}_i) = \boldsymbol{T}\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{T}\boldsymbol{u}_i := \tilde{\boldsymbol{X}}_i\boldsymbol{\beta} + \tilde{\boldsymbol{u}}_i \\
\tilde{\boldsymbol{u}}_i &:= \boldsymbol{T}\boldsymbol{u}_i = \boldsymbol{T}(\boldsymbol{e}\alpha_i + \boldsymbol{\varepsilon}_i) = \boldsymbol{T}\boldsymbol{e}\alpha_i + \boldsymbol{T}\boldsymbol{\varepsilon}_i = \boldsymbol{0} + \boldsymbol{T}\boldsymbol{\varepsilon}_i = \boldsymbol{T}\boldsymbol{\varepsilon}_i = \tilde{\boldsymbol{\varepsilon}}_i
\end{aligned}
$$

It is obvious that $\tilde{\boldsymbol{u}}_i = \boldsymbol{T}\boldsymbol{\varepsilon}_i$ is correlated with $\tilde{\boldsymbol{X}}_i := \boldsymbol{T}\boldsymbol{X}_i$ if $\boldsymbol{\varepsilon}_i$ is correlated with $\boldsymbol{X}_i$.

If $\boldsymbol{T} = \boldsymbol{Q} = \boldsymbol{I}_T - T^{-1}\boldsymbol{e}\boldsymbol{e}'$,

$$
\begin{aligned}
\tilde{\boldsymbol{y}}_i &= \tilde{\boldsymbol{X}}_i\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}_i \\
(\boldsymbol{y}_i - \boldsymbol{e}\bar{y}_i) &= (\boldsymbol{X}_i - \boldsymbol{e}\bar{\boldsymbol{x}}'_i)\boldsymbol{\beta} + (\boldsymbol{\varepsilon}_i - \boldsymbol{e}\bar{\varepsilon}_i) \\
(y_{it} - \bar{y}_i) &= (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)'\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i)
\end{aligned}
$$

Under weak form of weak/sequential exogeneity assumption $\mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{it}) = \boldsymbol{0}$ for $s \leq t$.

For $s \leq t$, we have

$$
\begin{aligned}
\mathbb{E}(\boldsymbol{z}_{is}(\varepsilon_{it} - \bar{\varepsilon}_i)) &= \mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{it}) - \mathbb{E}(\boldsymbol{z}_{is}\bar{\varepsilon}_i) \\
&= \boldsymbol{0} - \mathbb{E}(\boldsymbol{z}_{is}\sum_{t=1}^{T}\varepsilon_{it}/T) \\
&= -\frac{1}{T}\mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{i1} + \cdots + \boldsymbol{z}_{is}\varepsilon_{i,s-1} + \boldsymbol{z}_{is}\varepsilon_{is} + \cdots + \boldsymbol{z}_{is}\varepsilon_{iT}) \\
&= -\frac{1}{T}(\mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{i1}) + \cdots + \mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{i,s-1}) + \mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{is}) + \cdots + \mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{iT})) \\
&= -\frac{1}{T}(\mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{i1}) + \cdots + \mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{i,s-1}) + \boldsymbol{0} + \cdots + \boldsymbol{0}) \\
&= -\frac{1}{T}(\mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{i1}) + \cdots + \mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{i,s-1}))
\end{aligned}
$$

So $\mathbb{E}(\boldsymbol{z}_{it}(\varepsilon_{it} - \bar{\varepsilon}_i))$ is not necessarily equal to zero under weak form of weak/sequential exogeneity assumption. If weak form of strong/strict exogeneity is assumed $\mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{it}) = \boldsymbol{0}\ \forall s$, then $\mathbb{E}(\boldsymbol{z}_{is}(\varepsilon_{it} - \bar{\varepsilon}_i)) = \boldsymbol{0}\ \forall s$. So, $\boldsymbol{z}_{is}$, $s = 1, \cdots, T$ satisfy the exclusion restriction (exogeneity) requirement of valid instrument since $Cov(\boldsymbol{z}_{is}, \varepsilon_{it} - \bar{\varepsilon}_i) = \underbrace{\mathbb{E}(\boldsymbol{z}_{is}(\varepsilon_{it} - \bar{\varepsilon}_i))}_{\boldsymbol{0}} - \mathbb{E}(\boldsymbol{z}_{is})\mathbb{E}(\varepsilon_{it} - \bar{\varepsilon}_i) =$

$-\mathbb{E}(\boldsymbol{z}_{is})(\underbrace{\mathbb{E}(\varepsilon_{it})}_{0} - T^{-1}\sum_{t=1}^{T}\underbrace{\mathbb{E}(\varepsilon_{it})}_{0}) = \mathbf{0}\ \forall s$ (additionally assume $\mathbb{E}(\varepsilon_{it}) = 0$). So, we have

$$\begin{aligned} &\mathbb{E}(\boldsymbol{z}_{is}(\varepsilon_{it} - \bar{\varepsilon}_i)) = \mathbf{0} && \text{for } \forall s \\ \iff &\mathbb{E}(\boldsymbol{Z}_i'(\boldsymbol{\varepsilon}_i - \boldsymbol{e}\bar{\varepsilon}_i)) = \mathbf{0} \\ \iff &\mathbb{E}(\boldsymbol{Z}_i'\tilde{\boldsymbol{\varepsilon}}_i) = \mathbf{0} \end{aligned}$$

We can then apply IV estimation in GMM framework.

If $\boldsymbol{T} = \boldsymbol{\Delta}$

$$\begin{aligned} \tilde{\boldsymbol{y}}_i &= \tilde{\boldsymbol{X}}_i\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}_i \\ \boldsymbol{\Delta y}_i &= \boldsymbol{\Delta X}_i\boldsymbol{\beta} + \boldsymbol{\Delta\varepsilon}_i \\ (y_{it} - y_{i,t-1}) &= (\boldsymbol{x}_{it} - \boldsymbol{x}_{i,t-1})'\boldsymbol{\beta} + (\varepsilon_{it} - \varepsilon_{i,t-1}) \end{aligned}$$

Under weak form of weak/sequential exogeneity assumption $\mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{it}) = \mathbf{0}$ for $s \leq t$.

For $s < t$, we have

$$\begin{aligned} \mathbb{E}(\boldsymbol{z}_{is}(\varepsilon_{it} - \varepsilon_{i,t-1})) &= \mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{it}) - \mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{i,t-1}) \\ &= \mathbf{0} - \mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{i,t-1}) && \text{as } s < t \implies s \leq t \\ &= \mathbf{0} && \text{as } s < t \iff s \leq t-1 \end{aligned}$$

So, $\boldsymbol{z}_{is}$ for $s < t$ satisfy the exclusion restriction (exogeneity) requirement of valid instrument since $Cov(\boldsymbol{z}_{is}, \varepsilon_{it} - \varepsilon_{i,t-1}) = 0$ for $s < t$ (additionally assume $\mathbb{E}(\varepsilon_{it}) = 0$). Equivalently,

$$\boldsymbol{Z}_i = \begin{pmatrix} t=2; \boldsymbol{z}_{i1}' & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & t=3; \begin{pmatrix} \boldsymbol{z}_{i1}' & \boldsymbol{z}_{i2}' \end{pmatrix} & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & t=T; \begin{pmatrix} \boldsymbol{z}_{i1}' & \cdots & \boldsymbol{z}_{i,T-1}' \end{pmatrix} \end{pmatrix}$$

So, we have

$$\begin{aligned} &\mathbb{E}(\boldsymbol{z}_{is}(\varepsilon_{it} - \varepsilon_{i,t-1})) = \mathbf{0} && \text{for } s < t \\ \iff &\mathbb{E}(\boldsymbol{Z}_i'\boldsymbol{\Delta\varepsilon}_i) = \mathbf{0} \\ \iff &\mathbb{E}(\boldsymbol{Z}_i'\tilde{\boldsymbol{\varepsilon}}_i) = \mathbf{0} \end{aligned}$$

We can then apply IV estimation in GMM framework.

5 GMM Estimation of Random Effect Model

$$
\begin{aligned}
y_{it} &= \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i + \varepsilon_{it} \\
\boldsymbol{y}_i &= \boldsymbol{X}_i\boldsymbol{\beta} + \underbrace{(\boldsymbol{e}\alpha_i + \boldsymbol{\varepsilon}_i)}_{\boldsymbol{u}_i} \\
\boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I}_N \otimes \boldsymbol{e})\boldsymbol{\alpha} + \boldsymbol{\varepsilon}
\end{aligned}
$$

5.1 Assumption

α_i is not correlated with $\boldsymbol{X}_i$.

$\boldsymbol{\varepsilon}_i$ is potentially correlated with $\boldsymbol{X}_i$, so $\boldsymbol{u}_i$ is potentially correlated with $\boldsymbol{X}_i$ Thus, $\boldsymbol{X}_i$ is potentially endogeneous.

If $\boldsymbol{X}_i$ is endogeneous, OLS estimation is inconsistent and biased. We should use IV estimation (for just-identified case) and 2SLS estimation (for over-identified case). IV and 2SLS estimations are special cases of GMM estimation.

Assume

$$
\mathbb{E}(\boldsymbol{u}_i|\boldsymbol{Z}_i) = \boldsymbol{0}
$$

Which is stronger than $\mathbb{E}(\boldsymbol{Z}_i'\boldsymbol{u}_i) = \boldsymbol{0}$ as $\mathbb{E}(\boldsymbol{u}_i|\boldsymbol{Z}_i) = \boldsymbol{0}$ implies $\mathbb{E}(\boldsymbol{Z}_i'\boldsymbol{u}_i) = \boldsymbol{0}$

And assume

$$
Var(\boldsymbol{u}_i|\boldsymbol{Z}_i) = \boldsymbol{\Omega}_i = \begin{pmatrix} \sigma_\alpha^2 + \sigma_\varepsilon^2 & \cdots & \sigma_\alpha^2 \\ \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 + \sigma_\varepsilon^2 \end{pmatrix}
$$

5.1.1 Optimal moment condition

$$
\begin{aligned}
\boldsymbol{D}_i &= \mathbb{E}(\frac{\partial \boldsymbol{u}_i'}{\partial \boldsymbol{\beta}}|\boldsymbol{Z}_i)Var(\boldsymbol{u}_i|\boldsymbol{Z}_i)^{-1} \\
&= \mathbb{E}(\frac{\partial (\boldsymbol{Z}_i\boldsymbol{\beta})'}{\partial \boldsymbol{\beta}}|\boldsymbol{Z}_i)\boldsymbol{\Omega}_i^{-1} \\
&= \mathbb{E}(\boldsymbol{Z}_i'|\boldsymbol{Z}_i)\boldsymbol{\Omega}_i^{-1} \\
&= \boldsymbol{Z}_i'\boldsymbol{\Omega}_i^{-1}
\end{aligned}
$$

Optimal unconditional moment is

$$
\begin{aligned}
\mathbb{E}(\boldsymbol{D}_i\boldsymbol{u}_i) &= \boldsymbol{0} \\
\mathbb{E}(\boldsymbol{Z}_i'\boldsymbol{\Omega}_i^{-1}\boldsymbol{u}_i) &= \boldsymbol{0} \\
\mathbb{E}(\boldsymbol{Z}_i'\boldsymbol{\Omega}_i^{-1/2}\boldsymbol{\Omega}_i^{-1/2}\boldsymbol{u}_i) &= \boldsymbol{0} \\
\mathbb{E}(\boldsymbol{Z}_i'\boldsymbol{\Omega}_i'^{-1/2}\boldsymbol{\Omega}_i^{-1/2}\boldsymbol{u}_i) &= \boldsymbol{0} \\
\mathbb{E}(\boldsymbol{Z}_i'\boldsymbol{\Omega}_i^{-1/2'}\boldsymbol{\Omega}_i^{-1/2}\boldsymbol{u}_i) &= \boldsymbol{0} \\
\mathbb{E}((\boldsymbol{\Omega}_i^{-1/2}\boldsymbol{Z}_i)'\boldsymbol{\Omega}_i^{-1/2}\boldsymbol{u}_i) &= \boldsymbol{0} \\
\sigma_\varepsilon^2\mathbb{E}((\boldsymbol{\Omega}_i^{-1/2}\boldsymbol{Z}_i)'\boldsymbol{\Omega}_i^{-1/2}\boldsymbol{u}_i) &= \sigma_\varepsilon^2\boldsymbol{0} \\
\mathbb{E}((\sigma_\varepsilon\boldsymbol{\Omega}_i^{-1/2}\boldsymbol{Z}_i)'\sigma_\varepsilon\boldsymbol{\Omega}_i^{-1/2}\boldsymbol{u}_i) &= \boldsymbol{0}
\end{aligned}
$$

This implies that the model should be transformed by $\sigma_\varepsilon\boldsymbol{\Omega}_i^{-1/2}$

5.2 GMM Estimator of Random Effect Model

5.2.1 Transformed model

$$
\begin{aligned}
\sigma_\varepsilon\boldsymbol{\Omega}^{-1/2}\boldsymbol{y}_i &= \sigma_\varepsilon\boldsymbol{\Omega}^{-1/2}(\boldsymbol{X}_i\boldsymbol{\beta} + (\boldsymbol{e}\alpha_i + \boldsymbol{\varepsilon}_i)) = \sigma_\varepsilon\boldsymbol{\Omega}^{-1/2}(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{u}_i) = \sigma_\varepsilon\boldsymbol{\Omega}^{-1/2}\boldsymbol{X}_i\boldsymbol{\beta} + \sigma_\varepsilon\boldsymbol{\Omega}^{-1/2}\boldsymbol{u}_i \\
(\boldsymbol{y}_i - \lambda\boldsymbol{e}\bar{y}_i) &= (\boldsymbol{X}_i - \lambda\boldsymbol{e}\bar{\boldsymbol{x}}_i')\boldsymbol{\beta} + [(1-\lambda)\boldsymbol{e}\alpha_i + (\boldsymbol{\varepsilon}_i - \lambda\boldsymbol{e}\bar{\varepsilon}_i)] \qquad \lambda = 1 - \psi = 1 - \tfrac{\sigma_\varepsilon}{\sqrt{T\sigma_\alpha^2 + \sigma_\varepsilon^2}} \\
(y_{it} - \lambda\bar{y}_i) &= (\boldsymbol{x}_{it} - \lambda\bar{\boldsymbol{x}}_i)'\boldsymbol{\beta} + [(1-\lambda)\alpha_i + (\varepsilon_{it} - \lambda\bar{\varepsilon}_i)]
\end{aligned}
$$

Under weak form of weak/sequential exogeneity assumption $\mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{it}) = \mathbf{0}$ for $s \leq t$.

For $s \leq t$, we have

$$\begin{aligned}
\mathbb{E}(\boldsymbol{z}_{is}[(1-\lambda)\alpha_i + (\varepsilon_{it} - \lambda\bar{\varepsilon}_i)]) &= \mathbb{E}(\boldsymbol{z}_{is}(1-\lambda)\alpha_i + \boldsymbol{z}_{is}(\varepsilon_{it} - \lambda\bar{\varepsilon}_i)) \\
&= (1-\lambda)\mathbb{E}(\boldsymbol{z}_{is}\alpha_i) + \mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{it}) - \lambda\mathbb{E}(\boldsymbol{z}_{is}\bar{\varepsilon}_i) \\
&= (1-\lambda)\mathbf{0} + \mathbf{0} - \lambda E(\boldsymbol{z}_{is}\sum_{t=1}^{T}\varepsilon_{it}/T) \\
&= -\frac{\lambda}{T}\mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{i1} + \cdots + \boldsymbol{z}_{is}\varepsilon_{i,s-1} + \boldsymbol{z}_{is}\varepsilon_{is} + \cdots + \boldsymbol{z}_{is}\varepsilon_{iT}) \\
&= -\frac{\lambda}{T}(\mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{i1}) + \cdots + \mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{i,s-1}) + \mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{is}) + \cdots + \mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{iT})) \\
&= -\frac{\lambda}{T}(\mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{i1}) + \cdots + \mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{i,s-1}) + \mathbf{0} + \cdots + \mathbf{0}) \\
&= -\frac{\lambda}{T}(\mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{i1}) + \cdots + \mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{i,s-1}))
\end{aligned}$$

So $\mathbb{E}(\boldsymbol{z}_{it}(\varepsilon_{it} - \bar{\varepsilon}_i))$ is not necessarily equal to zero under weak form of weak/sequential exogeneity assumption.

If weak form of strong/strict exogeneity assumption is assumed $\mathbb{E}(\boldsymbol{z}_{is}\varepsilon_{it}) = \mathbf{0}\ \forall s$, then $\mathbb{E}(\boldsymbol{z}_{is}(\varepsilon_{it} - \bar{\varepsilon}_i)) = \mathbf{0}\ \forall s$
So, $\boldsymbol{z}_{is}$, $s = 1, \cdots, T$ satisfy the exclusion restriction (exogeneity) requirement of valid instrument.

So, we have

$$\begin{aligned}
&\mathbb{E}(\boldsymbol{z}_{is}[(1-\lambda)\alpha_i + (\varepsilon_{it} - \lambda\bar{\varepsilon}_i)]) = \mathbf{0} \qquad \text{for } \forall s \\
\iff\ &\mathbb{E}(\boldsymbol{Z}_i'[(1-\lambda)\boldsymbol{e}\alpha_i + (\boldsymbol{\varepsilon}_i - \lambda\boldsymbol{e}\bar{\varepsilon}_i)]) = \mathbf{0}
\end{aligned}$$

We can then apply IV estimation in GMM framework.

6 Dynamic Linear Panel Model

6.1 Assumption

6.1.1 Weak/sequential Exogeneity

For $t = 2, \cdots, T$

$$\mathbb{E}(\varepsilon_{it}|y_{i,t-1}, \cdots y_{i1}, \alpha_i) = 0$$

This implies

$$\mathbb{E}(y_{is}\varepsilon_{it}) = 0,\ \mathbb{E}(\varepsilon_{it}) = 0 \quad and \quad \mathbb{E}(\alpha_i\varepsilon_{it}) = 0 \qquad \text{for } s < t$$

And

$$Cov(y_{is}, \varepsilon_{it}) = 0 \quad and \quad Cov(\alpha_i, \varepsilon_{it}) = 0 \qquad \text{for } s < t$$

It is because

$$\begin{aligned} Cov(y_{is}, \varepsilon_{it}) &= \mathbb{E}(y_{is}\varepsilon_{it}) - \mathbb{E}(y_{is})\mathbb{E}(\varepsilon_{it}) \\ &= \mathbb{E}(\mathbb{E}(y_{is}\varepsilon_{it}|y_{i,t-1}, \cdots y_{i1}, \alpha_i)) - \mathbb{E}(y_{is})\mathbb{E}(\mathbb{E}(\varepsilon_{it}|y_{i,t-1}, \cdots y_{i1}, \alpha_i)) \\ &= \mathbb{E}(y_{is}\underbrace{\mathbb{E}(\varepsilon_{it}|y_{i,t-1}, \cdots y_{i1}, \alpha_i)}_{0}) - \mathbb{E}(y_{is})\mathbb{E}(\underbrace{\mathbb{E}(\varepsilon_{it}|y_{i,t-1}, \cdots y_{i1}, \alpha_i)}_{0}) \qquad \text{as } s < t \\ &= 0 \end{aligned}$$

6.2 Model

6.2.1 No transformation

$$y_{it} = \gamma y_{i,t-1} + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \underbrace{(\alpha_i + \varepsilon_{it})}_{u_{it}}$$

$$\begin{aligned} Cov(y_{i,t-1}, \alpha_i) &= Cov(\gamma y_{i,t-2} + \boldsymbol{x}'_{i,t-1}\boldsymbol{\beta} + \alpha_i + \varepsilon_{i,t-1}, \alpha_i) \\ &= \gamma Cov(y_{i,t-2}, \alpha_i) + Cov(\boldsymbol{x}'_{i,t-1}\boldsymbol{\beta}, \alpha_i) + Var(\alpha_i) + \underbrace{Cov(\varepsilon_{i,t-1}, \alpha_i)}_{0} \\ &= \gamma Cov(y_{i,t-2}, \alpha_i) + \boldsymbol{\beta}' Cov(\boldsymbol{x}_{i,t-1}, \alpha_i) + Var(\alpha_i) \\ &\neq 0 \qquad \text{assume } Cov(\boldsymbol{x}_{i,t-1}, \alpha_i) \neq 0 \text{ and } Var(\alpha_i) > 0 \end{aligned}$$

so that

$$\begin{aligned} Cov(y_{i,t-1}, u_{it}) &= Cov(y_{i,t-1}, \alpha_i + \varepsilon_{it}) \\ &= \underbrace{Cov(y_{i,t-1}, \alpha_i)}_{\neq 0} + \underbrace{Cov(y_{i,t-1}, \varepsilon_{it})}_{0} \\ &\neq 0 \end{aligned}$$

The necessary condition for OLS estimator to be unbiased is $\mathbb{E}(u_{it}|y_{i,t-1}, \boldsymbol{x}_{it}) = 0$. As $\mathbb{E}(u_{it}|y_{i,t-1}, \boldsymbol{x}_{it}) = 0 \implies Cov(y_{i,t-1}, u_{it}) = 0$. As a result, $Cov(y_{i,t-1}, u_{it}) \neq 0 \implies \mathbb{E}(u_{it}|y_{i,t-1}, \boldsymbol{x}_{it}) \neq 0$. Thus, OLS estimator is biased.

6.2.2 Special case: no $\boldsymbol{x}_{it}$

$$y_{it} = \gamma y_{i,t-1} + \underbrace{(\alpha_i + \varepsilon_{it})}_{u_{it}}$$

The necessary condition for OLS estimator to be consistent is $\mathbb{E}(y_{i,t-1}u_{it}) = 0$. However,

$$\begin{aligned}\mathbb{E}(y_{i,t-1}u_{it}) &= \mathbb{E}(y_{i,t-1}(\alpha_i + \varepsilon_{it})) \\ &= \mathbb{E}(y_{i,t-1}\alpha_i) + \underbrace{\mathbb{E}(y_{i,t-1}\varepsilon_{it})}_{0} > 0\end{aligned}$$

$$\begin{aligned}\mathbb{E}(y_{i,t-1}\alpha_i) &= \mathbb{E}((\gamma y_{i,t-2} + \alpha_i + \varepsilon_{i,t-1})\alpha_i) \\ &= \gamma\mathbb{E}(y_{i,t-2}\alpha_i) + \mathbb{E}(\alpha_i^2) + \mathbb{E}(\varepsilon_{i,t-1}\alpha_i) \\ &= \gamma\mathbb{E}((\gamma y_{i,t-3} + \alpha_i + \varepsilon_{i,t-2})\alpha_i) + \mathbb{E}(\alpha_i^2) + \mathbb{E}(\mathbb{E}(\varepsilon_{i,t-1}\alpha_i | y_{i,t-2}, \cdots, y_{i1}, \alpha_i)) \\ &= \gamma^2\mathbb{E}(y_{i,t-3}\alpha_i) + \gamma\mathbb{E}(\alpha_i^2) + \gamma\mathbb{E}(\varepsilon_{i,t-2}\alpha_i) + \mathbb{E}(\alpha_i^2) + \mathbb{E}(\alpha_i \underbrace{\mathbb{E}(\varepsilon_{i,t-1} | y_{i,t-2}, \cdots, y_{i1}, \alpha_i)}_{0}) \\ &= \gamma^2\mathbb{E}(y_{i,t-3}\alpha_i) + \gamma\mathbb{E}(\alpha_i^2) + \mathbb{E}(\alpha_i^2) \\ &\cdots \\ &= \gamma^{t-2}\mathbb{E}(y_{i,t-(t-2+1)}) + \gamma^{t-2-1}\mathbb{E}(\alpha_i^2) + \cdots + \mathbb{E}(\alpha_i^2) \\ &= \gamma^{t-2}\mathbb{E}(y_{i1}) + \gamma^{t-3}\mathbb{E}(\alpha_i^2) + \cdots + \mathbb{E}(\alpha_i^2) \\ &= \gamma^{t-2}y_{i1} + \gamma^{t-3}Var(\alpha_i) + \cdots + Var(\alpha_i) && y_{i1} \text{ is initial value and assume } \mathbb{E}(\alpha_i) = 0 \\ &> 0 && \text{assume } Var(\alpha_i) > 0,\ y_{i1} > 0 \text{ and } 0 < \gamma < 1\end{aligned}$$

Thus, OLS estimator is inconsistent. The necessary condition for OLS estimator to be unbiased is $\mathbb{E}(u_{it}|y_{i,t-1}) = 0$. As $\mathbb{E}(u_{it}|y_{i,t-1}) = 0 \implies \mathbb{E}(y_{i,t-1}u_{it}) = 0$, $\mathbb{E}(y_{i,t-1}u_{it}) \neq 0 \implies \mathbb{E}(u_{it}|y_{i,t-1}) \neq 0$. Thus, OLS estimator is biased. It can also be seen by OVB formula.

$$\begin{aligned}\gamma_{short} &= \frac{Cov(y_{it}, y_{i,t-1})}{Var(y_{i,t-1})} \\ &= \frac{Cov(\gamma_{long}y_{i,t-1} + \alpha_i + \varepsilon_{it}, y_{i,t-1})}{Var(y_{i,t-1})} \\ &= \gamma_{long} + \frac{Cov(\alpha_i, y_{i,t-1})}{Var(y_{i,t-1})} + \frac{\overbrace{Cov(\varepsilon_{it}, y_{i,t-1})}^{0}}{Var(y_{i,t-1})} \\ &= \gamma_{long} + \frac{Cov(\alpha_i, y_{i,t-1})}{Var(y_{i,t-1})}\end{aligned}$$

$$\gamma_{short} - \gamma_{long} = \frac{Cov(\alpha_i, y_{i,t-1})}{Var(y_{i,t-1})} > 0 \qquad \text{if } Var(y_{i,t-1}) > 0$$

$$Cov(\alpha_i, y_{i,t-1}) = \mathbb{E}(\alpha_i y_{i,t-1}) - \mathbb{E}(\alpha_i)\mathbb{E}(y_{i,t-1}) > 0 \qquad \text{see above for } \mathbb{E}(\alpha_i y_{i,t-1}) > 0 \text{ and assume } \mathbb{E}(\alpha_i) = 0$$

Thus, OLS estimator is biased upward / over-estimate.

6.2.3 Within transformation

$$y_{it} - \bar{y}_i = \gamma(y_{i,t-1} - \bar{y}_{i,-1}) + (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)'\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

$$\begin{aligned}Cov(y_{i,t-1}, \bar{\varepsilon}_i) &= Cov(\gamma y_{i,t-2} + \boldsymbol{x}'_{i,t-1}\boldsymbol{\beta} + \alpha_i + \varepsilon_{i,t-1}, T^{-1}\sum_{t=1}^{T}\varepsilon_{it}) \\ &\neq 0 && \text{since } \varepsilon_{i,t-1} \text{ is correlated with } T^{-1}\textstyle\sum_{t=1}^{T}\varepsilon_{it}\end{aligned}$$

so that

$$Cov(y_{i,t-1} - \bar{y}_{i,-1}, \varepsilon_{it} - \bar{\varepsilon}_i) \neq 0$$

The necessary condition for FE estimator to be unbiased is $\mathbb{E}(\varepsilon_{it} - \bar{\varepsilon}_i | y_{i,t-1} - \bar{y}_{i,-1}, \boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i) = 0$. As $\mathbb{E}(\varepsilon_{it} - \bar{\varepsilon}_i | y_{i,t-1} - \bar{y}_{i,-1}, \boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i) = 0 \implies Cov(y_{i,t-1} - \bar{y}_{i,-1}, \varepsilon_{it} - \bar{\varepsilon}_i) = 0$. As a result, $Cov(y_{i,t-1} - \bar{y}_{i,-1}, \varepsilon_{it} - \bar{\varepsilon}_i) \neq 0 \implies \mathbb{E}(\varepsilon_{it} - \bar{\varepsilon}_i | y_{i,t-1} - \bar{y}_{i,-1}, \boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i) \neq 0$. Thus, FE estimator is biased.

6.2.4 Special case: no $\boldsymbol{x}_{it}$

$$y_{it} - \bar{y}_i = \gamma(y_{i,t-1} - \bar{y}_{i,-1}) + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

The bias is called Nickell (1981) bias / dynamic panel bias. If $\gamma > 0$, the bias must be negative. The bias converges to zero when $T \to \infty$.

6.2.5 First difference transformation

$$\begin{aligned}
\tilde{\boldsymbol{y}}_i &= \tilde{\boldsymbol{X}}_i \boldsymbol{\delta} + \tilde{\boldsymbol{\varepsilon}}_i \\
\begin{pmatrix} y_{i3} - y_{i2} \\ \vdots \\ y_{iT} - y_{i,T-1} \end{pmatrix} &= \begin{pmatrix} y_{i2} - y_{i1} & (\boldsymbol{x}_{i3} - \boldsymbol{x}_{i2})' \\ \vdots & \\ y_{i,T-1} - y_{i,T-2} & (\boldsymbol{x}_{iT} - \boldsymbol{x}_{i,T-1})' \end{pmatrix} \begin{pmatrix} \gamma \\ \boldsymbol{\beta} \end{pmatrix} + \begin{pmatrix} \varepsilon_{i3} - \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iT} - \varepsilon_{i,T-1} \end{pmatrix} \\
y_{it} - y_{i,t-1} &= \gamma(y_{i,t-1} - y_{i,t-2}) + (\boldsymbol{x}_{it} - \boldsymbol{x}_{i,t-1})'\boldsymbol{\beta} + (\varepsilon_{it} - \varepsilon_{i,t-1}) && t \geq 3
\end{aligned}$$

$$\begin{aligned}
Cov(y_{i,t-1} - y_{i,t-2}, \varepsilon_{it} - \varepsilon_{i,t-1}) &= Cov(y_{i,t-1}, \varepsilon_{it}) - Cov(y_{i,t-1}, \varepsilon_{i,t-1}) - Cov(y_{i,t-2}, \varepsilon_{it}) + Cov(y_{i,t-2}, \varepsilon_{i,t-1}) \\
&= 0 - Cov(y_{i,t-1}, \varepsilon_{i,t-1}) - 0 + 0 && \text{as } Cov(y_{is}, \varepsilon_{it}) = 0 \text{ for } s < t \\
&= -Cov(\gamma y_{i,t-2} + \boldsymbol{x}'_{i,t-1}\boldsymbol{\beta} + \alpha_i + \varepsilon_{i,t-1}, \varepsilon_{i,t-1}) \\
&= -\gamma \underbrace{Cov(y_{i,t-2}, \varepsilon_{i,t-1})}_{0} - \boldsymbol{\beta}' \underbrace{Cov(\boldsymbol{x}_{i,t-1}, \varepsilon_{i,t-1})}_{0} - \underbrace{Cov(\alpha_i, \varepsilon_{i,t-1})}_{0} - Var(\varepsilon_{i,t-1}) \\
&< 0 && \text{assume } Var(\varepsilon_{i,t-1}) > 0
\end{aligned}$$

The necessary condition for FD estimator to be unbiased is $\mathbb{E}(\varepsilon_{it} - \varepsilon_{i,t-1} | y_{i,t-1} - y_{i,t-2}, \boldsymbol{x}_{it} - \boldsymbol{x}_{i,t-1}) = 0$. As $\mathbb{E}(\varepsilon_{it} - \varepsilon_{i,t-1} | y_{i,t-1} - y_{i,t-2}, \boldsymbol{x}_{it} - \boldsymbol{x}_{i,t-1}) = 0 \implies Cov(y_{i,t-1} - y_{i,t-2}, \varepsilon_{it} - \varepsilon_{i,t-1}) = 0$. As a result, $Cov(y_{i,t-1} - y_{i,t-2}, \varepsilon_{it} - \varepsilon_{i,t-1}) \neq 0 \implies \mathbb{E}(\varepsilon_{it} - \varepsilon_{i,t-1} | y_{i,t-1} - y_{i,t-2}, \boldsymbol{x}_{it} - \boldsymbol{x}_{i,t-1}) \neq 0$. Thus, FD estimator is biased.

6.2.6 Special case: no $\boldsymbol{x}_{it}$

$$y_{it} - y_{i,t-1} = \gamma(y_{i,t-1} - y_{i,t-2}) + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

The necessary condition for FD estimator to be consistent is $\mathbb{E}((y_{i,t-1} - y_{i,t-2})(\varepsilon_{it} - \varepsilon_{i,t-1})) = 0$. However,

$$\begin{aligned}
\mathbb{E}((y_{i,t-1} - y_{i,t-2})(\varepsilon_{it} - \varepsilon_{i,t-1})) &= \mathbb{E}(y_{i,t-1}\varepsilon_{it}) - \mathbb{E}(y_{i,t-1}\varepsilon_{i,t-1}) - \mathbb{E}(y_{i,t-2}\varepsilon_{it}) + \mathbb{E}(y_{i,t-2}\varepsilon_{i,t-1}) \\
&= 0 - \mathbb{E}(y_{i,t-1}\varepsilon_{i,t-1}) - 0 + 0 && \text{as } \mathbb{E}(y_{is}\varepsilon_{it}) = 0 \text{ for } s < t \\
&= -\mathbb{E}((\gamma y_{i,t-2} + \alpha_i + \varepsilon_{i,t-1})\varepsilon_{i,t-1}) \\
&= -\gamma \underbrace{\mathbb{E}(y_{i,t-2}\varepsilon_{i,t-1})}_{0} - \underbrace{\mathbb{E}(\alpha_i \varepsilon_{i,t-1})}_{0} - \mathbb{E}(\varepsilon_{i,t-1}^2) \\
&= -Var(\varepsilon_{i,t-1}) && \text{as } \mathbb{E}(\varepsilon_{i,t-1}) = 0 \\
&< 0 && \text{assume } Var(\varepsilon_{i,t-1}) > 0
\end{aligned}$$

Thus, FD estimator is inconsistent. The necessary condition for FD estimator to be unbiased is $\mathbb{E}(\varepsilon_{it} - \varepsilon_{i,t-1} | y_{i,t-1} - y_{i,t-2}) = 0$. As $\mathbb{E}(\varepsilon_{it} - \varepsilon_{i,t-1} | y_{i,t-1} - y_{i,t-2}) = 0 \implies \mathbb{E}((y_{i,t-1} - y_{i,t-2})(\varepsilon_{it} - \varepsilon_{i,t-1})) = 0$, $\mathbb{E}((y_{i,t-1} - y_{i,t-2})(\varepsilon_{it} - \varepsilon_{i,t-1})) \neq 0 \implies \mathbb{E}(\varepsilon_{it} - \varepsilon_{i,t-1} | y_{i,t-1} - y_{i,t-2}) \neq 0$. Thus, FD estimator is biased.

Thus, IV estimation (for just-identified case) or 2SLS estimation (for over-identified case) is applied. IV and 2SLS estimations are special cases of GMM estimation.

Under weak/sequential exogeneity, $Cov(y_{is}, \varepsilon_{it}) = 0$ for $s < t$. This implies for $s < t - 1 \iff s \leq t - 2$

$$\begin{aligned}
Cov(y_{is}, \varepsilon_{it} - \varepsilon_{it-1}) &= Cov(y_{is}, \varepsilon_{it}) - Cov(y_{is}, \varepsilon_{i,t-1}) \\
&= 0 - Cov(y_{is}, \varepsilon_{i,t-1}) && \text{as } s < t - 1 \implies s < t \\
&= 0 && \text{as } s < t - 1
\end{aligned}$$

Note that $Cov(y_{is}, \varepsilon_{it} - \varepsilon_{i,t-1}) = 0 \implies \mathbb{E}(y_{is}(\varepsilon_{it} - \varepsilon_{i,t-1})) = 0$ as $\mathbb{E}(\varepsilon_{it} - \varepsilon_{i,t-1}) = 0$ under weak/sequential exogeneity. So, y_{is} for $s \leq t-2$ satisfy the exclusion restriction (exogeneity) requirement of valid instrument. i.e.,

$$\begin{aligned}
\tilde{\boldsymbol{z}}'_{i3} =& (y_{i1}, \Delta\boldsymbol{x}'_{i3}) && \text{at } t=3\\
\tilde{\boldsymbol{z}}'_{i4} =& (y_{i1}, y_{i2}, \Delta\boldsymbol{x}'_{i4}) && \text{at } t=4\\
& \cdots & \\
\tilde{\boldsymbol{z}}'_{iT} =& (y_{i1}, \cdots, y_{i,T-2}, \Delta\boldsymbol{x}'_{iT}) && \text{at } t=T
\end{aligned}$$

That is, $\tilde{\boldsymbol{z}}'_{it} = [y_{i1}, \cdots, y_{i,t-2}, \Delta\boldsymbol{x}'_{it}]$. $\boldsymbol{Z}_i = \begin{pmatrix} \tilde{\boldsymbol{z}}'_{i3} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \tilde{\boldsymbol{z}}'_{i4} & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \tilde{\boldsymbol{z}}'_{iT} \end{pmatrix}$

So, we have

$$\begin{aligned}
&\mathbb{E}(\tilde{\boldsymbol{z}}_{it}(\varepsilon_{it} - \varepsilon_{i,t-1})) = \mathbf{0}\\
\iff &\mathbb{E}(\boldsymbol{Z}'_i \Delta\boldsymbol{\varepsilon}_i) = \mathbf{0}
\end{aligned}$$

We can then apply 2SLS estimation in GMM framework. This is the same as Arellano-Bond estimator with 2SLS weight.

6.2.7 Anderson-Hsiao estimator

Anderson & Hsiao (1981) considers a special case y_{is} for $s = t-2$ i.e., $y_{i,t-2}$ as the instrument since they not only satisfy the exclusion restriction (exogeneity) requirement but also satisfy the relevancy requirement of valid instrument i.e., correlates with $y_{i,t-1} - y_{i,t-2}$. Thus, $\tilde{\boldsymbol{z}}'_{it} = [y_{i,t-2}, \Delta\boldsymbol{x}'_{it}]$

$$\boldsymbol{Z}_i = \begin{pmatrix} \begin{pmatrix} y_{i1} & \Delta\boldsymbol{x}'_{i3} \end{pmatrix} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \begin{pmatrix} y_{i2} & \Delta\boldsymbol{x}'_{i4} \end{pmatrix} & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \begin{pmatrix} y_{i,T-2} & \Delta\boldsymbol{x}'_{iT} \end{pmatrix} \end{pmatrix}$$

and

$$\tilde{\boldsymbol{z}}'_{it} = [\underbrace{\Delta y_{i,t-2}}_{y_{i,t-2} - y_{i,t-3}}, \Delta\boldsymbol{x}'_{it}]$$

$$\boldsymbol{Z}_i = \begin{pmatrix} \begin{pmatrix} \Delta y_{i2} & \Delta\boldsymbol{x}'_{i4} \end{pmatrix} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \begin{pmatrix} \Delta y_{i3} & \Delta\boldsymbol{x}'_{i5} \end{pmatrix} & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \begin{pmatrix} \Delta y_{i,T-2} & \Delta\boldsymbol{x}'_{iT} \end{pmatrix} \end{pmatrix}$$

As only one instrument is used at each t, the number of moments is equal to the number of parameters i.e., $r = K$. In such case, GMM estimation = MM estimation = IV estimation.

$$\widehat{\boldsymbol{\delta}}^{pgmm}_{AH} = [\sum_{i=1}^{N} \boldsymbol{Z}'_i \tilde{\boldsymbol{X}}_i]^{-1} \sum_{i=1}^{N} \boldsymbol{Z}'_i \tilde{\boldsymbol{y}}_i = \widehat{\boldsymbol{\delta}}^{piv}_{AH}$$

6.2.8 Arellano-Bond estimator

Arellano & Bond (1991) considers all the possible cases i.e., y_{is} for $s \leq t-2$. Except $t = 3$, more than one instruments are used, number of moments is larger than the number of parameters i.e., $r > K$. GMM estimation is 2SLS estimation if $\boldsymbol{W}_N = (\sum_{i=1}^{N} \boldsymbol{Z}'_i \boldsymbol{Z}_i)^{-1}$.

$$\tilde{\boldsymbol{z}}'_{it} = [y_{i1}, \cdots, y_{i,t-2}, \Delta\boldsymbol{x}'_{it}]$$

$$\boldsymbol{Z}_i = \begin{pmatrix} \begin{pmatrix} y_{i1} & \Delta\boldsymbol{x}'_{i3} \end{pmatrix} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \begin{pmatrix} y_{i1} & y_{i2} & \Delta\boldsymbol{x}'_{i4} \end{pmatrix} & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \begin{pmatrix} y_{i1} & \cdots & y_{i,T-2} & \Delta\boldsymbol{x}'_{iT} \end{pmatrix} \end{pmatrix}$$

$$\widehat{\boldsymbol{\delta}}_{AB}^{pgmm} = [(\sum_{i=1}^{N} \tilde{\boldsymbol{X}}_i' \boldsymbol{Z}_i) \boldsymbol{W}_N (\sum_{i=1}^{N} \boldsymbol{Z}_i' \tilde{\boldsymbol{X}}_i)]^{-1} (\sum_{i=1}^{N} \tilde{\boldsymbol{X}}_i' \boldsymbol{Z}_i) \boldsymbol{W}_N (\sum_{i=1}^{N} \boldsymbol{Z}_i' \tilde{\boldsymbol{y}}_i)$$

If $\boldsymbol{W}_N = (\sum_{i=1}^{N} \boldsymbol{Z}_i' \boldsymbol{Z}_i)^{-1}$, $\widehat{\boldsymbol{\delta}}_{AB}^{pgmm} = \widehat{\boldsymbol{\delta}}_{AB}^{2SLS}$

If $\boldsymbol{W}_N = \widehat{\boldsymbol{S}}^{-1}$, $\widehat{\boldsymbol{\delta}}_{AB}^{pgmm} = \widehat{\boldsymbol{\delta}}_{AB}^{opgmm}$

7 Pooled Model and Clustered Standard Error

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i + \varepsilon_{it} \qquad \text{Level 1}$$

$$\begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_{i1}' \\ \vdots \\ \boldsymbol{x}_{iT}' \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \alpha_i + \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iT} \end{pmatrix}$$

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \underbrace{(\boldsymbol{e}\alpha_i + \boldsymbol{\varepsilon}_i)}_{\boldsymbol{u}_i} \qquad \text{Level 2}$$

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_N \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_N \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{e} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{e} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_N \end{pmatrix}$$

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I}_N \otimes \boldsymbol{e})\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \qquad \text{Level 3}$$

where α_i is unobserved heterogeneity, $\boldsymbol{\varepsilon}_i$ is idiosyncratic error, $\boldsymbol{u}_i$ is composite error.

If $\mathbb{E}(\alpha_i|\boldsymbol{X}_i) = 0$, OLS estimator is likely to be unbiased and consistent. $\mathbb{E}(\alpha_i|\boldsymbol{X}_i) = 0 \implies \mathbb{E}(\boldsymbol{u}_i|\boldsymbol{X}_i) = \boldsymbol{0}$ as $\mathbb{E}(\boldsymbol{u}_i|\boldsymbol{X}_i) = \mathbb{E}(\boldsymbol{X}_i'\boldsymbol{e}\alpha_i + \boldsymbol{\varepsilon}_i|\boldsymbol{X}_i) = \boldsymbol{X}_i'\boldsymbol{e}\mathbb{E}(\alpha_i|\boldsymbol{X}) + \mathbb{E}(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i) = \boldsymbol{0}$. Thus, the necessary condition for OLS estimator to be unbiased is satisfied if $\mathbb{E}(\alpha_i|\boldsymbol{X}_i) = 0$.

$\mathbb{E}(\boldsymbol{u}_i|\boldsymbol{X}_i) = \boldsymbol{0} \implies \mathbb{E}(\boldsymbol{X}_i'\boldsymbol{u}_i) = \boldsymbol{0}$ as $\mathbb{E}(\boldsymbol{X}_i'\boldsymbol{u}_i) = \mathbb{E}(\mathbb{E}(\boldsymbol{X}_i'\boldsymbol{u}_i|\boldsymbol{X}_i)) = \mathbb{E}(\boldsymbol{X}_i'\mathbb{E}(\boldsymbol{u}_i|\boldsymbol{X}_i)) = \mathbb{E}(\boldsymbol{X}_i'\boldsymbol{0}) = \boldsymbol{0}$. Thus, the necessary condition for OLS estimator to be consistent is satisfied if $\mathbb{E}(\alpha_i|\boldsymbol{X}_i) = 0$.

$\mathbb{E}(\alpha_i|\boldsymbol{X}_i) = 0 \implies \mathbb{E}(\alpha_i\boldsymbol{X}_i) = \boldsymbol{0}$ and $\mathbb{E}(\alpha_i) = 0$. Thus, $\mathbb{E}(\alpha_i|\boldsymbol{X}_i) = 0 \implies Cov(\alpha_i, \boldsymbol{X}_i) = \boldsymbol{0}$ as $Cov(\alpha_i, \boldsymbol{X}_i) = \mathbb{E}(\alpha_i\boldsymbol{X}_i) - \mathbb{E}(\alpha_i)\mathbb{E}(\boldsymbol{X}_i) = \boldsymbol{0}$.

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{pooled}^{ols} &= [\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{X}_i]^{-1} \sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{y}_i && \text{Level 2} \\ &= [\sum_{i=1}^N \sum_{t=1}^T \boldsymbol{x}_{it}\boldsymbol{x}_{it}']^{-1} \sum_{i=1}^N \sum_{t=1}^T \boldsymbol{x}_{it}y_{it} && \text{Level 1} \end{aligned}$$

From now on in this section, we assume $\alpha_i = 0$ for $\forall i$. Therefore, $\boldsymbol{u}_i = \boldsymbol{\varepsilon}_i$.

$$Var(\widehat{\boldsymbol{\beta}}_{pooled}^{ols}|\boldsymbol{X}_i) = [\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{X}_i]^{-1} \sum_{i=1}^N \boldsymbol{X}_i' Var(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i)\boldsymbol{X}_i [\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{X}_i]^{-1}$$

If ε_{it} is homoskedasticity and serially uncorrelated across t i.e., $Var(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i) = \sigma_\varepsilon^2\boldsymbol{I}_T$ (further assume independence of i and strict exogeneity), we have $\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i \sim iid\ [\boldsymbol{0}, \sigma_\varepsilon^2\boldsymbol{I}_T]$

$$\begin{aligned} &= \sigma_\varepsilon^2[\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{X}_i]^{-1} \\ &= \sigma_\varepsilon^2[\sum_{i=1}^N \sum_{t=1}^T \boldsymbol{x}_{it}\boldsymbol{x}_{it}']^{-1} \end{aligned}$$

If $Var(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i) = \boldsymbol{\Omega}_i$, we have $\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i \sim inid\ [\boldsymbol{0}, \boldsymbol{\Omega}_i]$

$$\begin{aligned} &= [\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{X}_i]^{-1} \sum_{i=1}^N \boldsymbol{X}_i' \overbrace{\mathbb{E}[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'|\boldsymbol{X}_i]}^{\boldsymbol{\Omega}_i} \boldsymbol{X}_i [\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{X}_i]^{-1} \\ &= [\sum_{i=1}^N \sum_{t=1}^T \boldsymbol{x}_{it}\boldsymbol{x}_{it}']^{-1} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T \boldsymbol{x}_{it}\mathbb{E}[\varepsilon_{it}\varepsilon_{is}|\boldsymbol{X}_i]\boldsymbol{x}_{is}' [\sum_{i=1}^N \sum_{t=1}^T \boldsymbol{x}_{it}\boldsymbol{x}_{it}']^{-1} \end{aligned}$$

7.1 Block bootstrapping

Suggested by MacKinnon, Nielsen, & Webb (2022).

7.2 Clustered standard error with independence of i

To be more precise, clustered covariance matrix is discussed here.

7.2.1 Liang & Zeger (1986) and Arellano (1987)

Clustered standard error can handle both heteroscedasticity and serial correlation within a cluster/group.

$$\widehat{Var}(\widehat{\boldsymbol{\beta}}^{ols}_{pooled}|\boldsymbol{X}_i) = [\sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{X}_i]^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i'\widehat{\boldsymbol{\Omega}}_i\boldsymbol{X}_i[\sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{X}_i]^{-1}$$

$$\widehat{\boldsymbol{\Omega}}_i = \widehat{\boldsymbol{\varepsilon}}_i\widehat{\boldsymbol{\varepsilon}}_i' = \begin{pmatrix} \hat{\varepsilon}_{i1}^2 & \hat{\varepsilon}_{i1}\hat{\varepsilon}_{i2} & \cdots & \hat{\varepsilon}_{i1}\hat{\varepsilon}_{iT} \\ \vdots & \hat{\varepsilon}_{i2}^2 & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\varepsilon}_{iT}\hat{\varepsilon}_{i1} & \cdots & \hat{\varepsilon}_{iT}\hat{\varepsilon}_{i,T-1} & \hat{\varepsilon}_{iT}^2 \end{pmatrix}$$

$$\widehat{\boldsymbol{\Omega}} = \begin{pmatrix} \widehat{\boldsymbol{\Omega}}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \widehat{\boldsymbol{\Omega}}_2 & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \widehat{\boldsymbol{\Omega}}_N \end{pmatrix}$$

$$\begin{aligned} \widehat{Var}(\widehat{\boldsymbol{\beta}}^{ols}_{pooled}|\boldsymbol{X}_i) &= [\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{x}_{it}\boldsymbol{x}_{it}']^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} \boldsymbol{x}_{it}\hat{\varepsilon}_{it}\hat{\varepsilon}_{is}\boldsymbol{x}_{is}'[\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{x}_{it}\boldsymbol{x}_{it}']^{-1} \\ &= [\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{x}_{it}\boldsymbol{x}_{it}']^{-1}(\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{x}_{it}\hat{\varepsilon}_{it}^2\boldsymbol{x}_{it}' + \\ &\qquad \sum_{l=1}^{T}[\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{x}_{it}\hat{\varepsilon}_{it}\hat{\varepsilon}_{i,t-l}\boldsymbol{x}_{i,t-l}' + \sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}\hat{\varepsilon}_{it}\hat{\varepsilon}_{i,t-l}\boldsymbol{x}_{i,t-l}')'])[\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{x}_{it}\boldsymbol{x}_{it}']^{-1} \end{aligned}$$

It is the panel generalization of Eicker-Huber-White estimator (White, 1980). If there is no serial correlation within the cluster/group, clustered standard error reduces to the exact form of Eicker-Huber-White estimator i.e.,

$$\begin{aligned} \widehat{Var}(\widehat{\boldsymbol{\beta}}^{ols}_{pooled}|\boldsymbol{X}_i) &= [\sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{X}_i]^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i' diag(\hat{\varepsilon}_{it}^2)\boldsymbol{X}_i[\sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{X}_i]^{-1} \\ &= [\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{x}_{it}\boldsymbol{x}_{it}']^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{x}_{it}\hat{\varepsilon}_{it}^2\boldsymbol{x}_{it}'[\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{x}_{it}\boldsymbol{x}_{it}']^{-1} \end{aligned}$$

7.2.2 Panel Newey-West (Petersen, 2009)

A weight can also be added to clustered standard error, this generalizes the Newey-West estimator (Newey & West, 1987).

$$\widehat{Var}(\widehat{\boldsymbol{\beta}}^{ols}_{pooled}|\boldsymbol{X}_i) = [\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{x}_{it}\boldsymbol{x}_{it}']^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} \boldsymbol{x}_{it}w_{t,s}\hat{\varepsilon}_{it}\hat{\varepsilon}_{is}\boldsymbol{x}_{is}'[\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{x}_{it}\boldsymbol{x}_{it}']^{-1}$$

where

$$w_{t,s} = \begin{cases} 1 - \frac{|s-t|}{L+1} & \text{if } |s-t| \leq L \\ 0 & \text{otherwise} \end{cases}$$

This can also be written as

$$\begin{aligned}\widehat{Var}(\widehat{\boldsymbol{\beta}}^{ols}_{pooled}|\boldsymbol{X}_i) = [\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}'_{it}]^{-1}(\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\hat{\varepsilon}^2_{it}\boldsymbol{x}'_{it}+\\ \sum_{l=1}^{L}w_l[\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\hat{\varepsilon}_{it}\hat{\varepsilon}_{i,t-l}\boldsymbol{x}'_{i,t-l}+\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}\hat{\varepsilon}_{it}\hat{\varepsilon}_{i,t-l}\boldsymbol{x}'_{i,t-l})'])[\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}'_{it}]^{-1}\end{aligned}$$

where $w_l = 1 - \frac{l}{L+1}$. Petersen (2009) finds that this adjustment is worser than the one without weight.

7.2.3 Generalization of HC1, HC2, and HC3 in MacKinnon & White (1985)

Finite sample adjustment e.g., $\frac{N}{N-1}\frac{NT-1}{NT-K}$ is multiplied in Stata (generalization of HC1 in MacKinnon & White (1985)).

If N (the number of cluster) is small e.g., less than 50 for state-year panel (Cameron & Miller, 2015), clustered standard error is inconsistent because law of large number cannot be applied (even $T \to \infty$). However, we can adjust it by Bell & McCaffrey (2002)'s Bias-Reduced Linearization (BRL) adjustment (generalization of HC2) and use t-distribution with $N-K$ degree of freedom, instead of standard normal distribution.

In BRL adjustment, we replace $\widehat{\boldsymbol{\varepsilon}}_i$ by

$$\widetilde{\boldsymbol{\varepsilon}}_i = \boldsymbol{A}_i\widehat{\boldsymbol{\varepsilon}}_i$$

where $\boldsymbol{A}'_i\boldsymbol{A}_i = (\boldsymbol{I}_T - \boldsymbol{H}_i)^{-1}$ where $\boldsymbol{H}_i = \boldsymbol{X}_i(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'_i$ the projection/hat matrix.

There are many possible $\boldsymbol{A}_i$, Bell & McCaffrey (2002) uses eigen-decomposition of the inverse of the residual marker $\boldsymbol{I}_T - \boldsymbol{H}_i$ i.e.,

$$\begin{aligned}(\boldsymbol{I}_T - \boldsymbol{H}_i)^{-1} &= \boldsymbol{P\Lambda P'}\\ &= \boldsymbol{P\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2}\boldsymbol{P'}\\ &= \boldsymbol{P\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2'}\boldsymbol{P'}\\ &= \boldsymbol{P\Lambda}^{1/2}(\boldsymbol{P\Lambda}^{1/2})'\\ &= \boldsymbol{A'A''}\end{aligned}$$

where $\boldsymbol{P}$ is a matrix in which vectors are eigenvectors and $\boldsymbol{\Lambda}$ is a diagonal matrix with eigenvalues items. Similar to HC2, BRL adjusted clustered standard error is unbiased when there is homoskedasticity i.e., $Var(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i) = \sigma^2_\varepsilon\boldsymbol{I}_T$.

Bell & McCaffrey (2002) also considers

$$\widetilde{\boldsymbol{\varepsilon}}_i = \sqrt{\frac{N-1}{N}}(\boldsymbol{I}_T - \boldsymbol{H}_i)^{-1}\widehat{\boldsymbol{\varepsilon}}_i$$

It is the generalization of HC3, which is less popular compared to HC2 generalization (Cameron & Miller, 2015).

7.3 Clustered standard error with dependence of i

7.3.1 Spatial Correlation Consistent (SCC) estimator (Driscoll & Kraay, 1998)

$$\widehat{Var}(\widehat{\boldsymbol{\beta}}^{ols}_{pooled}|\boldsymbol{X}_i) = [\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}'_{it}]^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\boldsymbol{x}_{it}w_{t,s}\hat{\varepsilon}_{it}\hat{\varepsilon}_{js}\boldsymbol{x}'_{js}[\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}'_{it}]^{-1}$$

where

$$w_{t,s} = \begin{cases}1 - \frac{|s-t|}{L+1} & \text{if } |s-t| \le L\\ 0 & \text{otherwise}\end{cases}$$

This can also be written as

$$\widehat{Var}(\widehat{\boldsymbol{\beta}}^{ols}_{pooled}|\boldsymbol{X}_t) = [\sum_{t=1}^{T}\boldsymbol{X}'_t\boldsymbol{X}_t]^{-1}(\sum_{t=1}^{T}\boldsymbol{X}'_t\widehat{\boldsymbol{\varepsilon}}_t\widehat{\boldsymbol{\varepsilon}}'_t\boldsymbol{X}_t + \sum_{l=1}^{L}w_l[\sum_{t=1}^{T}\boldsymbol{X}'_t\widehat{\boldsymbol{\varepsilon}}_t\widehat{\boldsymbol{\varepsilon}}'_{t-l}\boldsymbol{X}_{t-l} + \sum_{t=1}^{T}(\boldsymbol{X}'_t\widehat{\boldsymbol{\varepsilon}}_t\widehat{\boldsymbol{\varepsilon}}'_{t-l}\boldsymbol{X}_{t-l})'])[\sum_{t=1}^{T}\boldsymbol{X}'_t\boldsymbol{X}_t]^{-1}$$

where $w_l = 1 - \frac{l}{L+1}$. It requires large T while L is up to you.

7.4 Fama-Macbeth estimation

Fama-Macbeth estimation (Fama & Macbeth, 1973) was invented before the development of linear panel model in Econometrics. It is still widely applied in the areas of empirical asset pricing. Petersen (2009) finds that FM standard error performs well when there is only time fixed effect and no OVB in his simulation setting. Its large sample properties are derived in Jagannathan & Wang (1998). The derivation depends on the linear beta pricing model in Finance which implies the data generating process of the return $\boldsymbol{y}_t$.

$$y_{ti} = \boldsymbol{x}'_{ti}\boldsymbol{\beta} + \varepsilon_{ti} \qquad \text{Level 1}$$

$$\begin{pmatrix} y_{t1} \\ \vdots \\ y_{tN} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}'_{t1} \\ \vdots \\ \boldsymbol{x}'_{tN} \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \varepsilon_{t1} \\ \vdots \\ \varepsilon_{tN} \end{pmatrix}$$

$$\boldsymbol{y}_t = \boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t \qquad \text{Level 2}$$

Fama-Macbeth estimator is

$$\widehat{\boldsymbol{\beta}}_{FM} = \frac{1}{T}\sum_{t=1}^{T}[(\boldsymbol{X}'_t\boldsymbol{X}_t)^{-1}\boldsymbol{X}'_t\boldsymbol{y}_t]$$

Its covariance matrix is

$$\begin{aligned}
Var(\widehat{\boldsymbol{\beta}}_{FM}|\boldsymbol{X}_t) &= Var(\frac{1}{T}\sum_{t=1}^{T}[(\boldsymbol{X}'_t\boldsymbol{X}_t)^{-1}\boldsymbol{X}'_t\boldsymbol{y}_t]|\boldsymbol{X}_t) \\
&= \frac{1}{T^2}Var(\sum_{t=1}^{T}[(\boldsymbol{X}'_t\boldsymbol{X}_t)^{-1}\boldsymbol{X}'_t\boldsymbol{y}_t]|\boldsymbol{X}_t) \\
&= \frac{1}{T^2}\sum_{t=1}^{T}Var([(\boldsymbol{X}'_t\boldsymbol{X}_t)^{-1}\boldsymbol{X}'_t\boldsymbol{y}_t]|\boldsymbol{X}_t) && \text{due to independence} \\
&= \frac{1}{T^2}\sum_{t=1}^{T}(\boldsymbol{X}'_t\boldsymbol{X}_t)^{-1}\boldsymbol{X}'_t Var(\boldsymbol{y}_t|\boldsymbol{X}_t)\boldsymbol{X}_t(\boldsymbol{X}'_t\boldsymbol{X}_t)^{-1} \\
&= \frac{1}{T^2}\sum_{t=1}^{T}(\boldsymbol{X}'_t\boldsymbol{X}_t)^{-1}\boldsymbol{X}'_t \underbrace{Var(\boldsymbol{\varepsilon}_t|\boldsymbol{X}_t)}_{\boldsymbol{\Omega}_t}\boldsymbol{X}_t(\boldsymbol{X}'_t\boldsymbol{X}_t)^{-1}
\end{aligned}$$

Cochrane (2005) demonstrates that Fama-Macbeth estimator is equivalent to pooled OLS estimator if $\boldsymbol{X}_t = \boldsymbol{X}$ i.e., not time changing. FM variance is same as clustered standard error if $\boldsymbol{\Omega}_t = \boldsymbol{\Sigma}$ in addition to the assumption just mentioned.

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}^{ols}_{pooled} &= (\sum_{t=1}^{T}\boldsymbol{X}'\boldsymbol{X})^{-1}\sum_{t=1}^{T}\boldsymbol{X}'\boldsymbol{y}_t \\
&= (T\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\sum_{t=1}^{T}\boldsymbol{y}_t \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{y}_t = \widehat{\boldsymbol{\beta}}_{FM}
\end{aligned}$$

$$\begin{aligned}
Var(\widehat{\boldsymbol{\beta}}^{ols}_{pooled}|\boldsymbol{X}) &= [\sum_{t=1}^{T}\boldsymbol{X}'\boldsymbol{X}]^{-1}\sum_{t=1}^{T}\boldsymbol{X}'\boldsymbol{\Sigma}\boldsymbol{X}[\sum_{t=1}^{T}\boldsymbol{X}'\boldsymbol{X}]^{-1} \\
&= [T\boldsymbol{X}'\boldsymbol{X}]^{-1}T\boldsymbol{X}'\boldsymbol{\Sigma}\boldsymbol{X}[T\boldsymbol{X}'\boldsymbol{X}]^{-1} \\
&= \frac{1}{T}[\boldsymbol{X}'\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}\boldsymbol{X}[\boldsymbol{X}'\boldsymbol{X}]^{-1}
\end{aligned}$$

$$Var(\widehat{\boldsymbol{\beta}}_{FM}|\boldsymbol{X}) = \frac{1}{T^2}T(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} = Var(\widehat{\boldsymbol{\beta}}^{ols}_{pooled}|\boldsymbol{X})$$

7.5 Petersen (2009) Simulation Result

7.5.1 Only individual fixed effect

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

If there is only α_i (individual fixed effect) and α_i is not correlated with $\boldsymbol{x}_{it}$ (so no OVB), OLS estimator is unbiased and clustered standard error clustered by individual i.e.,

$$[\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}']^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\boldsymbol{x}_{it}\hat{\varepsilon}_{it}\hat{\varepsilon}_{is}\boldsymbol{x}_{is}'[\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}']^{-1}$$

is unbiased. In contrast, conventional standard error i.e.,

$$\widehat{\sigma}_{\varepsilon}^2[\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}']^{-1}$$

Eicker-Huber-White standard error i.e.,

$$[\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}']^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\hat{\varepsilon}_{it}^2\boldsymbol{x}_{it}'[\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}']^{-1}$$

Newey-West standard error i.e.,

$$[\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}']^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\boldsymbol{x}_{it}w_{t,s}\hat{\varepsilon}_{it}\hat{\varepsilon}_{is}\boldsymbol{x}_{is}'[\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}']^{-1}$$

Fama-Macbeth standard error are biased downward (over-rejection).

7.5.2 Only time fixed effect

$$y_{ti} = \boldsymbol{x}_{ti}'\boldsymbol{\beta} + \gamma_t + \varepsilon_{ti}$$

If there is only γ_t (time fixed effect) and γ_t is not correlated with $\boldsymbol{x}_{it}$ (so no OVB), OLS estimator is unbiased and Fama-Macbeth standard error i.e.,

$$\frac{1}{T^2}\sum_{t=1}^{T}(\boldsymbol{X}_t'\boldsymbol{X}_t)^{-1}\boldsymbol{X}_t'\boldsymbol{\Omega}_t\boldsymbol{X}_t(\boldsymbol{X}_t'\boldsymbol{X}_t)^{-1}$$

and clustered standard error clustered by time (only if T is large) is i.e.,

$$[\sum_{t=1}^{T}\sum_{i=1}^{N}\boldsymbol{x}_{ti}\boldsymbol{x}_{ti}']^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{x}_{ti}\hat{\varepsilon}_{ti}\hat{\varepsilon}_{tj}\boldsymbol{x}_{ti}'[\sum_{t=1}^{T}\sum_{i=1}^{N}\boldsymbol{x}_{ti}\boldsymbol{x}_{ti}']^{-1}$$

unbiased. In contrast, conventional standard error are biased downward (over-rejection).

7.5.3 Both individual and time fixed effect

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i + \gamma_t + \varepsilon_{it}$$

Cameron, Gelbach & Miller (2011), and Thompson (2011) suggests Clustered standard error clustered by individual + Clustered standard error clustered by time - Eicker-Huber-White standard error. i.e.,

$$\begin{aligned}
&[\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}']^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\boldsymbol{x}_{it}\hat{\varepsilon}_{it}\hat{\varepsilon}_{is}\boldsymbol{x}_{is}'[\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}']^{-1} + [\sum_{t=1}^{T}\sum_{i=1}^{N}\boldsymbol{x}_{ti}\boldsymbol{x}_{ti}']^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{x}_{ti}\hat{\varepsilon}_{ti}\hat{\varepsilon}_{tj}\boldsymbol{x}_{ti}'[\sum_{t=1}^{T}\sum_{i=1}^{N}\boldsymbol{x}_{ti}\boldsymbol{x}_{ti}']^{-1} - \\
&\quad [\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}']^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\hat{\varepsilon}_{it}^2\boldsymbol{x}_{it}'[\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}']^{-1} \\
&= [\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}']^{-1}(\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\boldsymbol{x}_{it}\hat{\varepsilon}_{it}\hat{\varepsilon}_{is}\boldsymbol{x}_{is}' + \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{x}_{ti}\hat{\varepsilon}_{ti}\hat{\varepsilon}_{tj}\boldsymbol{x}_{ti}' - \sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\hat{\varepsilon}_{it}^2\boldsymbol{x}_{it}')[\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}']^{-1}
\end{aligned}$$

The last term is subtracted in order to prevent double-counting of diagonal items. Simulation shows it works.

8 References

Anderson, T. W., & Hsiao, C. (1981). Estimation of Dynamic Models with Error Components. Journal of the American Statistical Association, 76(375), 598–606.

Arellano, M. (1987). PRACTITIONERS' CORNER: Computing Robust Standard Errors for Within-groups Estimators. Oxford Bulletin of Economics and Statistics, 49, 431–434.

Arellano, M., & Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. The Review of Economic Studies, 58(2), 277-297.

Bell, R. M., & McCaffrey, D. F. (2002). Bias Reduction in Standard Errors for Linear Regression with Multi-Stage Samples. Survey Methodology, 28(2), 169–179.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust Inference With Multiway Clustering. Journal of Business & Economic Statistics, 29(2), 238–249.

Cameron, A. C., & Miller, D. L. (2015). A Practitioner's Guide to Cluster-Robust Inference. Journal of Human Resources,, 50 (2) 317–372.

Cameron, A. C., & Trivedi, P. K. (2005). Microeconometrics: Methods and Applications.

Cochrane, J. H. (2005). Asset Pricing: Revised Edition.

Driscoll, J. C., & Kraay, A. C. (1998). Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. The Review of Economics and Statistics, 80(4), 549–560.

Fama, E., & MacBeth, J. (1973). Risk, Return, and Equilibrium Empirical Tests. The Journal of Political Economy, 81, 607-636

Hansen, B. E. (2022). Econometrics.

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal Data Analysis Using Generalized Linear Models. Biometrika, 73(1), 13–22.

MacKinnon, J. G., Nielsen, M. Ø., & Webb, M. D. (2022). Cluster-robust inference: A guide to empirical practice. Journal of Econometrics, 232(2), 272–299.

MacKinnon, J. G., & White, H. (1985). Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties. Journal of Econometrics, 29(3), 305–325.

Millo, G. (2017). Robust Standard Error Estimators for Panel Models: A Unifying Approach. Journal of Statistical Software, 82(3), 1–27.

Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica, 55(3), 703–708.

Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. Econometrica, 49(6), 1417–1426.

Jagannathan, R., & Wang, Z. (1998). An Asymptotic Theory for Estimating Beta-Pricing Models Using Cross-Sectional Regression. The Journal of Finance, 53(4), 1285–1309.

Petersen, M. A. (2009). Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. The Review of Financial Studies, 22(1), 435–480.

Thompson, S. B. (2011). Simple formulas for standard errors that cluster by both firm and time. Journal of Financial Economics, 99(1), 1–10.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. Econometrica, 48(4), 817–838.

9 Appendix - R Code

```r
library(lmtest)
library(plm)
library(fastDummies)
library(tidyverse)

set.seed(15)

gen_long_data <- function (N, T, var_gamma_i, static = TRUE) {
  df <- tibble(.rows = N) %>% mutate(i = seq_len(N))
  df["gamma_i"] <- rnorm(N, mean = 0, sd = sqrt(var_gamma_i))

  for (t in seq_len(T)) {
    df[str_c("x_i", t)] <-
      df[["gamma_i"]] + rnorm(N, mean = 0, sd = sqrt(1 - var_gamma_i))

    if (t == 1 | static) {
      df[str_c("y_i", t)] <-
        2 + df[[str_c("x_i", t)]] * 2 + df[["gamma_i"]] + rnorm(N, mean = 0, sd = sqrt(4 - var_gamma_i))
    } else {
      df[str_c("ly_i", t)] <- df[[str_c("y_i", t - 1)]]
      df[str_c("y_i", t)] <-
        2 + df[[str_c("y_i", t - 1)]] * 0.5 + df[[str_c("x_i", t)]] * 2 + df[["gamma_i"]] +
          rnorm(N, mean = 0, sd = sqrt(4 - var_gamma_i))
    }
  }

  if (static) {
    df %>%
      select(-gamma_i) %>%
      pivot_longer(
        cols = starts_with(c("y_i", "x_i")),
        names_to = c(".value", "t"),
        names_sep = "_i"
      )
  } else {
    df %>%
      select(-gamma_i) %>%
      pivot_longer(
        cols = starts_with(c("y_i", "ly_i", "x_i")),
        names_to = c(".value", "t"),
        names_sep = "_i"
      )
  }
}

long_data_static<-
  gen_long_data(100, 5, 0.1, static = TRUE) %>%
  mutate(t = as.numeric(t)) %>%
  dummy_cols(select_columns = "i")

long_data_static_plm <- long_data_static %>% pdata.frame(index = c("i", "t"))

# within / FE estimator
plm(y ~ x, model = "within", effect = "individual", data = long_data_static_plm) %>%
  coeftest(., vcov = plm::vcovHC(., type = "HC1", cluster = "group"))

# first difference estimator
plm(y ~ x, model = "fd", effect = "individual", data = long_data_static_plm) %>%
  coeftest(., vcov = plm::vcovHC(., type = "HC1", cluster = "group"))

# LSDV estimator
plm(y ~ . + 0, model = "pooling", data = long_data_static_plm %>% select(-i, -t)) %>%
  coeftest(., vcov = plm::vcovHC(., type = "HC1", cluster = "group"))

# GLS / RE estimator
plm(y ~ x, model = "random", effect = "individual", data = long_data_static_plm) %>% coeftest()

# Pooled OLS with clustered standard error
plm(y ~ x, model = "pooling", data = long_data_static_plm) %>%
  coeftest(., vcov = plm::vcovHC(., type = "HC1", cluster = "group"))

# Pooled OLS with BRL adjusted clustered standard error
plm(y ~ x, model = "pooling", data = long_data_static_plm) %>%
```

```r
  coeftest(., vcov = plm::vcovHC(., type = "HC2", cluster = "group"))

long_data_dynamic<-
  gen_long_data(100, 5, 0.1, static = FALSE) %>%
  mutate(t = as.numeric(t))

long_data_dynamic_plm <- long_data_dynamic %>% pdata.frame(index = c("i", "t"))

# Anderson-Hsiao estimator
pgmm(
  y ~ lag(y, 1) + x | lag(y, 2) | x,
  effect = "individual",
  model = "onestep",
  transformation = "d",
  data = long_data_dynamic_plm
) %>% summary(robust = TRUE)

# Arellano-Bond estimator
pgmm(
  y ~ lag(y, 1) + x | lag(y, 2:4) | x,
  effect = "individual",
  model = "twosteps",
  transformation = "d",
  data = long_data_dynamic_plm
) %>% summary(robust = TRUE)
```