Econ 582
Fixed Effects Estimation of Panel Data

Eric Zivot

May 28, 2012

Panel Data Framework

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}, \\ i &= 1, \ldots, n \text{ (individuals); } t = 1, \ldots, T \text{ (time periods)} \\ \underset{T\times 1}{\mathbf{y}_i} &= \underset{(T\times L)}{\mathbf{X}_i}\underset{(L\times 1)}{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}_i, \end{aligned}$$

Main question: Is $\mathbf{x}_{it}$ uncorrelated with ε_{it}?

1. If yes, then we have a SUR type model with common coefficients.

2. If no, then we have a multi-equation system with common coefficients and endogenous regressors. We need to use an estimation procedure to deal with the endogeneity. In the panel set-up, under certain assumptions, we can deal with the endogeneity without using instruments using the so-called fixed effects (FE) estimator.

Error Components Assumption

$$\begin{aligned}
\varepsilon_{it} &= \alpha_i + \eta_{it} \\
\alpha_i &= \text{unobserved fixed effect} \\
E[\mathbf{x}_{it}\eta_{it}] &= 0 \\
E[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'] &= \boldsymbol{\Sigma},
\end{aligned}$$

The fixed effect component α_i (which is actually an unobserved random variable) captures unobserved heterogeneity across individuals that is fixed over time.

With the error components assumption, the RE and FE models are defined as follows:

$$\begin{aligned}
&\text{RE model:} \quad E[\mathbf{x}_{it}\alpha_i] = 0 \\
&\text{FE model:} \quad E[\mathbf{x}_{it}\alpha_i] \neq 0
\end{aligned}$$

Example: Panel wage equation

$$\begin{aligned}
LW69_i &= \phi + \beta S69_i + \gamma IQ_i + \pi EXPR69_i + \alpha_i + \eta_{i,69} \\
LW80_i &= \phi + \beta S80_i + \gamma IQ_i + \pi EXPR80_i + \alpha_i + \eta_{i,80} \\
E[IQ_i\alpha_i] &\neq 0
\end{aligned}$$

Here α_i captures unobserved ability that is correlated with IQ_i. It is assumed that ability does not vary over time.

Note: The existence of α_i guarantees across time error correlation:

$$\begin{aligned}
E[\varepsilon_{i,69}\varepsilon_{i,80}] &= E[\left(\alpha_i + \eta_{i,69}\right)\left(\alpha_i + \eta_{i,80}\right)] \\
&= E[\alpha_i^2] + E[\alpha_i\eta_{i,80}] + E[\alpha_i\eta_{i,69}] + E[\eta_{i,69}\eta_{i,80}] \\
&\neq 0
\end{aligned}$$

Remark: With panel data, as we saw in the last lecture, the endogeneity due to unobserved heterogeneity (i.e., $E[\mathbf{x}_{it}\alpha_i] \neq 0$) can be eliminated without the use of instruments. To see this, consider the difference in log-wages over time:

$$
\begin{gathered}
LW80_i - LW69_i = (\phi - \phi) + \beta(S80_i - S69_i) \\
+\pi(EXPR80_i - EXPR69_i) + (\alpha_i - \alpha_i) + \left(\eta_{i,80} - \eta_{i,69}\right) \\
= \beta(S80_i - S69_i) + \pi(EXPR80_i - EXPR69_i) + \left(\eta_{i,80} - \eta_{i,69}\right)
\end{gathered}
$$

Hence, we can consistently estimate β and π by using the first differenced data!

Fixed Effects Estimation

Key insight: With panel data, $\boldsymbol{\beta}$ can be consistently estimated without using instruments.

There are 3 equivalent approaches

1. Within group estimator

2. Least squares dummy variable estimator

3. First difference estimator

Within group estimator

To illustrate the within group estimator consider the simplified panel regression with a single regressor

$$\begin{aligned} y_{it} &= \beta x_{it} + \alpha_i + \eta_{it} \\ E[x_{it}\alpha_i] &\neq 0 \\ E[x_{it}\eta_{it}] &= 0 \end{aligned}$$

Trick to remove fixed effect α_i : First, for each i average over time t

$$\begin{aligned} \bar{y}_i &= \beta \bar{x}_i + \alpha_i + \bar{\eta}_i \\ \bar{y}_i &= \frac{1}{T}\sum_{t=1}^{T} y_{it},\ \bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}, \\ \alpha_i &= \frac{1}{T}\sum_{t=1}^{T} \alpha_i \end{aligned}$$

Second, form the transformed regression

$$y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + (\alpha_i - \alpha_i) + \eta_{it} - \bar{\eta}_i$$

or

$$\tilde{y}_{it} = \beta \tilde{x}_{it} + \tilde{\eta}_{it}$$

Now, stack by observation for $t = 1, \ldots, T$ giving the giant regression

$$\begin{bmatrix} \tilde{\mathbf{y}}_1 \\ \vdots \\ \tilde{\mathbf{y}}_n \end{bmatrix} = \beta \begin{bmatrix} \tilde{\mathbf{x}}_1 \\ \vdots \\ \tilde{\mathbf{x}}_n \end{bmatrix} + \begin{bmatrix} \tilde{\boldsymbol{\eta}}_1 \\ \vdots \\ \tilde{\boldsymbol{\eta}}_n \end{bmatrix}$$

or

$$\underset{Tn\times 1}{\tilde{\mathbf{y}}} = \beta \underset{Tn\times 1}{\tilde{\mathbf{x}}} + \underset{Tn\times 1}{\tilde{\boldsymbol{\eta}}}$$

The within-group FE estimator is pooled OLS on the transformed regression (stacked by observation)

$$\begin{aligned}\hat{\beta}_{FE} &= (\tilde{\mathbf{x}}'\tilde{\mathbf{x}})^{-1}\tilde{\mathbf{x}}'\tilde{\mathbf{y}} \\ &= \left(\sum_{i=1}^{n} \tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_i\right)^{-1} \sum_{i=1}^{n} \tilde{\mathbf{x}}_i'\tilde{\mathbf{y}}_i\end{aligned}$$

Remarks

1. If $\mathbf{x}_{it}$ does not vary with t (e.g. $\mathbf{x}_{it} = \mathbf{x}_i$) then $\tilde{\mathbf{x}}_{it} = \mathbf{0}$ and we cannot estimate $\boldsymbol{\beta}$.

2. Must be careful computing the degrees of freedom for the FE estimator. There are Tn total observations and L parameters in $\boldsymbol{\beta}$, so it appreas that there are $Tn - L$ degrees of freedom. However, you lose 1 degree of freedom for each fixed effect eliminated. So the actual degrees of freedom are $Tn - L - n = n(T-1) - L$.

Matrix Algebra Derivation of Within Group Fixed Effects Estimator

Consider the general model (assume all variables vary with i and t)

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

Stack the observations for $t = 1, \ldots, T$ giving

$$\underset{T\times 1}{\mathbf{y}_i} = \underset{(T\times L)(L\times 1)}{\mathbf{X}_i \quad \boldsymbol{\beta}} + \underset{(1\times 1)(T\times 1)}{\alpha_i \quad \mathbf{1}_T} + \underset{T\times 1}{\boldsymbol{\eta}_i}$$

Define

$$\begin{aligned}
\underset{T\times T}{\mathbf{Q}_T} &= \mathbf{I}_T - \mathbf{1}_T(\mathbf{1}_T'\mathbf{1}_T)^{-1}\mathbf{1}_T' \\
&= \mathbf{I}_T - \mathbf{P}_T \\
\mathbf{P}_T &= \mathbf{1}_T(\mathbf{1}_T'\mathbf{1}_T)^{-1}\mathbf{1}_T' = T^{-1}\mathbf{1}_T\mathbf{1}_T'
\end{aligned}$$

Note

$$\begin{aligned}
\mathbf{P}_T\mathbf{1}_T &= \mathbf{1}_T,\ \mathbf{Q}_T\mathbf{1}_T = \mathbf{0} \\
\mathbf{P}_T\mathbf{y}_i &= \mathbf{1}_T(\mathbf{1}_T'\mathbf{1}_T)^{-1}\mathbf{1}_T'\mathbf{y}_i \\
&= \mathbf{1}_T\bar{y}_i \\
\mathbf{Q}_T\mathbf{y}_i &= (\mathbf{I}_T - \mathbf{P}_T)\mathbf{y}_i \\
&= \mathbf{y}_i - \mathbf{1}_T\bar{y}_i \\
&= \tilde{\mathbf{y}}_i
\end{aligned}$$

The transformed error components model is then

$$\begin{aligned}\mathbf{Q}_T\mathbf{y}_i &= \mathbf{Q}_T\mathbf{X}_i\boldsymbol{\beta} + \alpha_i\mathbf{Q}_T\mathbf{1}_T + \mathbf{Q}_T\boldsymbol{\eta}_i \\ &\Rightarrow \tilde{\mathbf{y}}_i = \tilde{\mathbf{X}}_i\boldsymbol{\beta} + \tilde{\boldsymbol{\eta}}_i\end{aligned}$$

The giant regression (stacked by observation) is

$$\begin{bmatrix}\tilde{\mathbf{y}}_1 \\ \vdots \\ \tilde{\mathbf{y}}_n\end{bmatrix} = \begin{bmatrix}\tilde{\mathbf{X}}_1 \\ \vdots \\ \tilde{\mathbf{X}}_n\end{bmatrix}\boldsymbol{\beta} + \begin{bmatrix}\tilde{\boldsymbol{\eta}}_1 \\ \vdots \\ \tilde{\boldsymbol{\eta}}_n\end{bmatrix}$$

or

$$\underset{Tn\times 1}{\tilde{\mathbf{y}}} = \underset{(Tn\times L)}{\tilde{\mathbf{X}}}\underset{(L\times 1)}{\boldsymbol{\beta}} + \underset{Tn\times 1}{\tilde{\boldsymbol{\eta}}}$$

Note: Unless $E[\boldsymbol{\eta}\boldsymbol{\eta}'] = \boldsymbol{\sigma}_\eta^2\mathbf{I}_{Tn}$ $\hat{\boldsymbol{\beta}}_{OLS} = \hat{\boldsymbol{\beta}}_{FE}$ is not efficient.

The FE estimator is again pooled OLS on the transformed system

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{FE} &= \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}} = \left(\sum_{i=1}^n \tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\right)^{-1}\sum_{i=1}^n \tilde{\mathbf{X}}_i'\tilde{\mathbf{y}}_i \\ &= \left(\sum_{i=1}^n (\mathbf{Q}_T\mathbf{X}_i)'\mathbf{Q}_T\mathbf{X}_i\right)^{-1}\sum_{i=1}^n (\mathbf{Q}_T\mathbf{X}_i)'\mathbf{Q}_T\mathbf{y}_i \\ &= \left(\sum_{i=1}^n \mathbf{X}_i'\mathbf{Q}_T\mathbf{X}_i\right)^{-1}\sum_{i=1}^n \mathbf{X}_i'\mathbf{Q}_T\mathbf{y}_i\end{aligned}$$

since $\mathbf{Q}_T$ is idempotent.

Least Squares Dummy Variable Model

Consider the general model

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

Stack the observations over t giving

$$\underset{T\times 1}{\mathbf{y}_i} = \mathbf{X}_i\boldsymbol{\beta} + \alpha_i \mathbf{1}_T + \boldsymbol{\eta}_i$$

Now create the giant regression

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_n \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{1}_T & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathbf{1}_T \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} \boldsymbol{\eta}_1 \\ \vdots \\ \boldsymbol{\eta}_n \end{bmatrix}$$

or

$$\begin{aligned} \underset{Tn\times 1}{\mathbf{y}} &= \underset{(Tn\times L)}{\mathbf{X}} \underset{(L\times 1)}{\boldsymbol{\beta}} + \underset{(Tn\times n)}{\mathbf{D}} \underset{(n\times 1)}{\boldsymbol{\alpha}} + \underset{(Tn\times 1)}{\boldsymbol{\eta}} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{I}_n \otimes \mathbf{1}_T)\, \boldsymbol{\alpha} + \boldsymbol{\eta} \\ \mathbf{D} &= \mathbf{I}_n \otimes \mathbf{1}_T \end{aligned}$$

Aside: Partitioned Regression

Consider the partitioned regression equation

$$\underset{n\times 1}{\mathbf{y}} = \underset{n\times k_1}{\mathbf{X}_1} \underset{k_1\times 1}{\boldsymbol{\beta}_1} + \underset{n\times k_2}{\mathbf{X}_2} \underset{k_2\times 2}{\boldsymbol{\beta}_2} + \boldsymbol{\varepsilon}$$

The LS estimators for $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ can be expressed as

$$\begin{aligned} \hat{\boldsymbol{\beta}}_1 &= (\mathbf{X}'_1\mathbf{Q}_2\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{Q}_2\mathbf{y},\ \mathbf{Q}_2 = \mathbf{I}_n - \mathbf{P}_{\mathbf{X}_2} \\ \hat{\boldsymbol{\beta}}_2 &= (\mathbf{X}'_2\mathbf{Q}_1\mathbf{X}_2)^{-1}\mathbf{X}'_2\mathbf{Q}_1\mathbf{y},\ \mathbf{Q}_1 = \mathbf{I}_n - \mathbf{P}_{\mathbf{X}_1} \end{aligned}$$

where

$$\mathbf{P}_{\mathbf{X}_1} = \mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1,\ \mathbf{P}_{\mathbf{X}_2} = \mathbf{X}_2(\mathbf{X}'_2\mathbf{X}_2)^{-1}\mathbf{X}'_2$$

Result: The FE estimator is the partitioned OLS estimator of $\boldsymbol{\beta}$ in the giant regression

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{FE} &= (\mathbf{X}'\mathbf{Q}_D\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}_D\mathbf{y} \\ \mathbf{Q}_D &= \mathbf{I}_{Tn} - \mathbf{P}_D,\ \mathbf{P}_D = \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}',\ \mathbf{D} = \mathbf{I}_n \otimes \mathbf{1}_T\end{aligned}$$

Now,

$$\begin{aligned}\mathbf{P}_D &= (\mathbf{I}_n \otimes \mathbf{1}_T)\left[(\mathbf{I}_n \otimes \mathbf{1}_T)'(\mathbf{I}_n \otimes \mathbf{1}_T)\right]^{-1}(\mathbf{I}_n \otimes \mathbf{1}_T)' \\ &= (\mathbf{I}_n \otimes \mathbf{1}_T)\left[\mathbf{I}_n \otimes \mathbf{1}_T'\mathbf{1}_T\right]^{-1}(\mathbf{I}_n \otimes \mathbf{1}_T)' \\ &= (\mathbf{I}_n \otimes \mathbf{1}_T)[\mathbf{I}_n \otimes \left(\mathbf{1}_T'\mathbf{1}_T\right)^{-1}]\left(\mathbf{I}_n \otimes \mathbf{1}_T'\right) \\ &= \mathbf{I}_n \otimes \mathbf{P}_T\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbf{Q}_D &= \mathbf{I}_{Tn} - \mathbf{P}_D \\ &= \mathbf{I}_{Tn} - (\mathbf{I}_n \otimes \mathbf{P}_T) \\ &= \mathbf{I}_n \otimes \mathbf{Q}_T\end{aligned}$$

As a result

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{FE} &= (\mathbf{X}'\mathbf{Q}_D\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}_D\mathbf{y} \\ &= (\mathbf{X}'(\mathbf{I}_n \otimes \mathbf{Q}_T)\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I}_n \otimes \mathbf{Q}_T)\mathbf{y} \\ &= \left(\sum_{i=1}^{n}\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\right)^{-1}\sum_{i=1}^{n}\tilde{\mathbf{X}}_i'\tilde{\mathbf{y}}_i\end{aligned}$$

which is exactly the result we got when we transformed the model by subtracting off group means. That is, the LSDV estimator of $\boldsymbol{\beta}$ in the FE model is numerically identical to the Within estimator of β.

Recovering Estimates of α

With the LSDV approach, it is straightforward to deduce estimates for α_i. By examining the normal equations for $\boldsymbol{\alpha}$ in the Giant Regression, one can deduce that

$$\begin{aligned}\hat{\alpha}_i &= \bar{y}_i - \bar{\mathbf{x}}_i' \hat{\boldsymbol{\beta}}_{FE} \\ \bar{y}_i &= \frac{1}{T}\sum_{t=1}^{T} y_{it},\ \bar{\mathbf{x}}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}\end{aligned}$$

Comments:

- For short panels (small T) $\hat{\alpha}_i$ is inconsistent (T fixed and $n \to \infty$)

FE as a First Difference Estimator

Results:

- When $T = 2$, pooled OLS on the first differenced model is numerically identical to the LSDV and Within estimators of $\boldsymbol{\beta}$.
- When $T > 2$, pooled OLS on the first differenced model is not numerically the same as the LSDV and Within estimators of $\boldsymbol{\beta}$. It is consistent, but generally less efficient that the LSDV and Within estimators.
- When $T > 2$ and $E[\boldsymbol{\eta}_i \boldsymbol{\eta}_i'] = \sigma_\eta^2 \mathbf{I}_T$ (no serial correlation), then pooled GLS on the first differenced model is numerically the same as the LSDV and Within Estimators.

Equivalence of First Differenced (FD) Estimator and LSDV and Within Estimators

Consider the error components model

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}'\boldsymbol{\beta} + \alpha_i + \eta_{it} \\ E[\boldsymbol{\eta}_i\boldsymbol{\eta}_i'] &= \sigma_\eta^2\mathbf{I}_T \end{aligned}$$

Note: The spherical error assumption is made so that GLS estimation is possible.

Take 1st differences over t:

$$\Delta y_{i2} = y_{i2} - y_{i1} = \Delta\mathbf{x}_{i2}'\boldsymbol{\beta} + \Delta\eta_{i2}$$
$$\vdots$$
$$\Delta y_{iT} = y_{iT} - y_{i(T-1)} = \Delta\mathbf{x}_{iT}'\boldsymbol{\beta} + \Delta\eta_{iT}$$

Notice that differencing removes the fixed effect.

In matrix form the model is

$$\underset{(T-1)\times 1}{\mathbf{C}'\mathbf{y}_i} = \mathbf{C}'\mathbf{X}_i\boldsymbol{\beta} + \mathbf{C}'\boldsymbol{\eta}_i$$

or

$$\hat{\mathbf{y}}_i = \hat{\mathbf{X}}_i\boldsymbol{\beta} + \hat{\boldsymbol{\eta}}_i$$

where

$$\underset{(T-1)\times T}{\mathbf{C}'} = \begin{pmatrix} -1 & 1 & 0 & & 0 \\ 0 & -1 & 1 & & 0 \\ & & & & \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}$$

Note that

$$E[\hat{\boldsymbol{\eta}}_i\hat{\boldsymbol{\eta}}_i'] = E[\mathbf{C}'\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{C}] = \mathbf{C}'E[\boldsymbol{\eta}_i\boldsymbol{\eta}_i']\mathbf{C} = \sigma_\eta^2\mathbf{C}'\mathbf{C}$$

The transformed Giant Regression is

$$\begin{bmatrix} \hat{\mathbf{y}}_1 \\ \vdots \\ \hat{\mathbf{y}}_n \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{X}}_1 \\ \vdots \\ \hat{\mathbf{X}}_n \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \hat{\boldsymbol{\eta}}_1 \\ \vdots \\ \hat{\boldsymbol{\eta}}_n \end{bmatrix}$$

or

$$\underset{(T-1)n\times 1}{\hat{\mathbf{y}}} = \underset{((T-1)n\times L)}{\hat{\mathbf{X}}} \underset{(L\times 1)}{\boldsymbol{\beta}} + \underset{(T-1)n\times 1}{\hat{\boldsymbol{\eta}}}$$

Notice that

$$\underset{(T-1)\times(T-1)}{E[\hat{\boldsymbol{\eta}}\hat{\boldsymbol{\eta}}']} = \begin{pmatrix} \sigma_\eta^2 \mathbf{C}'\mathbf{C} & 0 & \cdots & 0 \\ 0 & \sigma_\eta^2 \mathbf{C}'\mathbf{C} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_\eta^2 \mathbf{C}'\mathbf{C} \end{pmatrix} = \sigma_\eta^2 \left[\mathbf{I}_n \otimes (\mathbf{C}'\mathbf{C})\right]$$

Pooled GLS on the transformed regression gives

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{FE} &= \left[\hat{\mathbf{X}}' \left(\mathbf{I}_n \otimes (\mathbf{C}'\mathbf{C})^{-1}\right) \hat{\mathbf{X}}\right]^{-1} \hat{\mathbf{X}}' \left(\mathbf{I}_n \otimes (\mathbf{C}'\mathbf{C})^{-1}\right) \hat{\mathbf{y}} \\
&= \left(\sum_{i=1}^{n} \mathbf{X}_i{}'\mathbf{C}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{X}_i\right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i{}'\mathbf{C}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{y}_i \\
&= \left(\sum_{i=1}^{n} \mathbf{X}_i{}'\mathbf{P}_C\mathbf{X}_i\right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i{}'\mathbf{P}_C\mathbf{y}_i \\
&= \left(\sum_{i=1}^{n} \mathbf{X}_i{}'\mathbf{Q}_T\mathbf{X}_i\right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i{}'\mathbf{Q}_T\mathbf{y}_i \\
&= \left(\sum_{i=1}^{n} \tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\right)^{-1} \sum_{i=1}^{n} \tilde{\mathbf{X}}_i'\tilde{\mathbf{y}}_i
\end{aligned}$$

Now:

$$\begin{aligned}
& \hat{\mathbf{X}}' \left(\mathbf{I}_n \otimes (\mathbf{C}'\mathbf{C})^{-1}\right) \hat{\mathbf{X}} \\
= & \left[\hat{\mathbf{X}}_1' \cdots \hat{\mathbf{X}}_n'\right] \begin{pmatrix} (\mathbf{C}'\mathbf{C})^{-1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & (\mathbf{C}'\mathbf{C})^{-1} \end{pmatrix} \begin{bmatrix} \hat{\mathbf{X}}_1 \\ \vdots \\ \hat{\mathbf{X}}_n \end{bmatrix} \\
= & \sum_{i=1}^{n} \hat{\mathbf{X}}_i'(\mathbf{C}'\mathbf{C})^{-1}\hat{\mathbf{X}}_i \\
= & \sum_{i=1}^{n} (\mathbf{C}'\mathbf{X}_i)'(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{X}_i \\
= & \sum_{i=1}^{n} \mathbf{X}_i{}'\mathbf{C}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{X}_i = \sum_{i=1}^{n} \mathbf{X}_i{}'\mathbf{P}_C\mathbf{X}_i, \ \mathbf{P}_C = \mathbf{C}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'
\end{aligned}$$

The result

$$\mathbf{P}_C = \mathbf{C}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}' = \mathbf{Q}_T = \mathbf{I}_T - \mathbf{1}_T(\mathbf{1}_T'\mathbf{1}_T)^{-1}\mathbf{1}_T'$$

follows from the result that

$$\mathbf{C}'\mathbf{1}_T = \begin{pmatrix} -1 & 1 & 0 & & 0 \\ 0 & -1 & 1 & & 0 \\ & & & & \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

That is, $\mathbf{P}_C = \mathbf{C}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'$ is an idempotent matrix satisfying $\mathbf{P}_C\mathbf{1}_T = \mathbf{0}$. It projects onto the space orthogonal to $\mathbf{1}_T$, and this is exactly what $\mathbf{Q}_T$ does:

$$\mathbf{Q}_T\mathbf{1}_T = \mathbf{0}$$

Hence, when $E[\boldsymbol{\eta}_i\boldsymbol{\eta}_i'] = \sigma_\eta^2\mathbf{I}_T$ pooled GLS on the FD model is numerically equivalent to the LSDV and Within estimators of $\boldsymbol{\beta}$.

Panel Robust Statistical Inference

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}'\boldsymbol{\beta} + \alpha_i + \eta_{it}, \\ i &= 1, \ldots, n; t = 1, \ldots, T \end{aligned}$$

Assumptions:

1. $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$, $i = 1, \ldots, n$
2. $\boldsymbol{\varepsilon}_i = \alpha_i \mathbf{1}_T + \boldsymbol{\eta}_i$ (error components)
3. $\{\mathbf{y}_i, \mathbf{X}_i\}$ is i.i.d. (over i but not t)
4. $E[x_{it}\alpha_i] \neq 0$ (Endogeneity)
5. $E[x_{it}\eta_{ih}] = 0$ for $t, h = 1, 2 \ldots, T$

Comments

- It is reasonable to assume independence over i (i.e., cross-sectional independence due to random sampling at given t)
- Errors are generally serially correlated over t for a given i (η_{it} is autocorrelated) and heteroskedastic over i (cross-sectional heteroskedasticity)
- OLS standard errors are typically downward biased due to serial correlation
- Panel robust standard errors correct for serial correlation and heteroskedasticity.

FE Panel Data Estimators in Common Notation

The Within (LSDV) and FD estimators of $\boldsymbol{\beta}$ in

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}'\boldsymbol{\beta} + \alpha_i + \eta_{it}, \\ i &= 1, \ldots, n; t = 1, \ldots, T \end{aligned}$$

are pooled OLS estimators in a transformed model

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

- Within Estimator: $\tilde{y}_{it} = y_{it} - \bar{y}_i$, $\tilde{\mathbf{x}}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_i$, $\eta_{it} - \bar{\eta}_i$
- FD Estimator: $\tilde{y}_{it} = y_{it} - y_{it-1}$, $\tilde{\mathbf{x}}_{it} = \mathbf{x}_{it} - \mathbf{x}_{it-1}$, $\eta_{it} - \eta_{it-1}$

In matrix notation, we have

$$\underset{T\times 1}{\tilde{\mathbf{y}}_i} = \tilde{\mathbf{X}}_i\boldsymbol{\beta} + \tilde{\boldsymbol{\eta}}_i, \ i = 1, \ldots, n$$

and the giant regression is

$$\underset{nT\times 1}{\tilde{\mathbf{y}}} = \underset{nT\times L}{\tilde{\mathbf{X}}}\underset{L\times 1}{\boldsymbol{\beta}} + \tilde{\boldsymbol{\eta}}$$

Pooled OLS on the giant regression is

$$\hat{\boldsymbol{\beta}}_{FE} = \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}} = \left(\sum_{i=1}^{n}\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\right)^{-1}\sum_{i=1}^{n}\tilde{\mathbf{X}}_i'\tilde{\mathbf{y}}_i, \ FE = Within, FD$$

Asymptotic Distribution Theory (Advanced)

Using $\mathbf{y}_i = \tilde{\mathbf{X}}_i\boldsymbol{\beta} + \tilde{\boldsymbol{\eta}}_i$, we have

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{FE} &= \left(\sum_{i=1}^n \tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\right)^{-1} \sum_{i=1}^n \tilde{\mathbf{X}}_i' \left(\tilde{\mathbf{X}}_i\boldsymbol{\beta} + \tilde{\boldsymbol{\eta}}_i\right) \\
&= \left(\sum_{i=1}^n \tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\right)^{-1} \left(\sum_{i=1}^n \tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\right)\boldsymbol{\beta} + \left(\sum_{i=1}^n \tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\right)^{-1} \sum_{i=1}^n \tilde{\mathbf{X}}_i'\tilde{\boldsymbol{\eta}}_i \\
&= \boldsymbol{\beta} + \left(\sum_{i=1}^n \tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\right)^{-1} \sum_{i=1}^n \tilde{\mathbf{X}}_i'\tilde{\boldsymbol{\eta}}_i
\end{aligned}$$

It follows that

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}\right) = \left(\frac{1}{n}\sum_{i=1}^n \tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{\mathbf{X}}_i'\tilde{\boldsymbol{\eta}}_i$$

The Law of Large Numbers (LLN) gives

$$\frac{1}{n}\sum_{i=1}^n \tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i \to E[\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i]^{-1}$$

and the Central Limit Theorem (CLT) gives

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{\mathbf{X}}_i'\tilde{\boldsymbol{\eta}}_i \to N(\mathbf{0}, E[\tilde{\mathbf{X}}_i'\tilde{\boldsymbol{\eta}}_i\tilde{\boldsymbol{\eta}}_i'\tilde{\mathbf{X}}_i])$$

Hence, by Slutsky's theorem

$$\begin{aligned}\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{FE}-\boldsymbol{\beta}\right) &= \left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\mathbf{X}}_i'\tilde{\boldsymbol{\eta}}_i \\ &\xrightarrow{d} E[\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i]^{-1}N(\mathbf{0},E[\tilde{\mathbf{X}}_i'\tilde{\boldsymbol{\eta}}_i\tilde{\boldsymbol{\eta}}_i'\tilde{\mathbf{X}}_i]) \\ &\equiv N(\mathbf{0},\text{avar}(\hat{\boldsymbol{\beta}}_{FE}))\end{aligned}$$

where

$$\text{avar}(\hat{\boldsymbol{\beta}}_{FE}) = E[\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i]^{-1}E[\tilde{\mathbf{X}}_i'\tilde{\boldsymbol{\eta}}_i\tilde{\boldsymbol{\eta}}_i'\tilde{\mathbf{X}}_i]E[\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i]^{-1}$$

Therefore,

$$\hat{\boldsymbol{\beta}}_{FE} \overset{A}{\sim} N(\boldsymbol{\beta}, n^{-1}\widehat{\text{avar}}(\hat{\boldsymbol{\beta}}_{FE}))$$

Remark: The sandwich form of $\text{avar}(\hat{\boldsymbol{\beta}}_{FE})$ suggests that it is an inefficient estimator in general.

Panel Robust Inference

Under serial correlation and heteroskedasticity, it can be shown (Econ 583)

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i \xrightarrow{p} E[\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i],\ \frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{X}}_i'\breve{\boldsymbol{\eta}}_i\breve{\boldsymbol{\eta}}_i'\tilde{\mathbf{X}}_i \xrightarrow{p} E[\tilde{\mathbf{X}}_i'\tilde{\boldsymbol{\eta}}_i\tilde{\boldsymbol{\eta}}_i'\tilde{\mathbf{X}}_i]$$

$$\breve{\boldsymbol{\eta}}_i = \tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\hat{\boldsymbol{\beta}}_{FE}$$

Then a consistent estimate for $\text{avar}(\hat{\boldsymbol{\beta}}_{FE})$ is

$$\begin{aligned}\widehat{\text{avar}}(\hat{\boldsymbol{\beta}}_{FE}) &= \left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\right)^{-1}\times\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{X}}_i'\breve{\boldsymbol{\eta}}_i\breve{\boldsymbol{\eta}}_i'\tilde{\mathbf{X}}_i\right)^{-1} \\ &\quad\times\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\right)^{-1}\end{aligned}$$

Remarks

- The formula for $\widehat{\text{avar}}(\hat{\boldsymbol{\beta}}_{FE})$ is the panel robust covariance estimate. It is not the same as the usual White correction for heteroskedasticity in a pooled OLS regression. The White correction does not account for serial correlation. It is also not the Newey-West correction for heteroskedasticity and autocorrelation.

- The formula for $\widehat{\text{avar}}(\hat{\boldsymbol{\beta}}_{FE})$ is also known as the cluster robust covariance estimate when the clustering variable is i (each individual is a cluster)

Special Case: When $\boldsymbol{\eta}_i$ is Spherical

$$E[\boldsymbol{\eta}_i\boldsymbol{\eta}_i'] = \sigma_\eta^2\mathbf{I}_T$$

For the Within estimator, we have $\tilde{\boldsymbol{\eta}}_i = \mathbf{Q}_T\boldsymbol{\eta}_i$ so that

$$E[\tilde{\boldsymbol{\eta}}_i\tilde{\boldsymbol{\eta}}_i'] = E[\mathbf{Q}_T\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{Q}_T] = \sigma_\eta^2\mathbf{Q}_T$$

This implies that there is no serial correlation (correlation across time) and no cross-sectional heteroskedasticity.

$$E[\tilde{\mathbf{X}}_i'\tilde{\boldsymbol{\eta}}_i\tilde{\boldsymbol{\eta}}_i'\tilde{\mathbf{X}}_i] = E[\tilde{\mathbf{X}}_i'E[\tilde{\boldsymbol{\eta}}_i\tilde{\boldsymbol{\eta}}_i']\tilde{\mathbf{X}}_i] = \sigma_\eta^2E[\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i]$$

and

$$\text{avar}(\hat{\boldsymbol{\beta}}_{FE}) = \sigma_\eta^2E[\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i]^{-1}$$

Then a consistent estimate for $\mathsf{avar}(\hat{\boldsymbol{\beta}}_{FE})$ is

$$\begin{aligned}
\widehat{\mathsf{avar}}(\hat{\boldsymbol{\beta}}_{FE}) &= \hat{\sigma}_\eta^2 \left(\frac{1}{n} \sum_{i=1}^{n} \tilde{\mathbf{X}}_i' \tilde{\mathbf{X}}_i \right)^{-1} \\
\hat{\sigma}_\eta^2 &= \frac{1}{Tn - n - L} \sum_{i=1}^{n} \breve{\boldsymbol{\eta}}_i' \breve{\boldsymbol{\eta}}_i \\
\breve{\boldsymbol{\eta}}_i &= (\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i \hat{\boldsymbol{\beta}}_{FE})
\end{aligned}$$

Remark: In this case the Within estimator is an efficient estimator.