Multiple Equation GMM with Common Coefficients: Panel Data

Eric Zivot

Winter 2013

Multi-equation GMM with common coefficients

Example (panel wage equation)

$$\begin{aligned} LW69_i &= \phi + \beta S69_i + \gamma IQ_i + \pi EXPR69_i + \varepsilon_{i1} \\ LW80_i &= \phi + \beta S80_i + \gamma IQ_i + \pi EXPR80_i + \varepsilon_{i2} \end{aligned}$$

Note: common coefficients across two equations and $cov(\varepsilon_{i1}, \varepsilon_{i2}) \neq 0$

The general multi-equation model with common coefficients $\boldsymbol{\delta}$ is

$$\begin{aligned} y_{im} &= \underset{1\times L}{\mathbf{z}'_{im}} \underset{L\times 1}{\boldsymbol{\delta}} + \varepsilon_{im},\ i = 1, \ldots, n; m = 1, \ldots, M \\ E[\mathbf{z}_{im}\varepsilon_{im}] &\neq \mathbf{0} \text{ for some element of } z_{im} \end{aligned}$$

The model for the m^{th} equation is

$$\underset{n\times 1}{\mathbf{y}_m} = \underset{n\times L}{\mathbf{Z}_m} \underset{L\times 1}{\boldsymbol{\delta}} + \underset{n\times 1}{\boldsymbol{\varepsilon}_m},\ m = 1, \ldots, M$$

Note

$$\boldsymbol{\delta}_1 = \cdots = \boldsymbol{\delta}_M = \boldsymbol{\delta}$$

The stacked system (across equations) is

$$\left.\begin{array}{c}\mathbf{y}_1 = \mathbf{Z}_1\boldsymbol{\delta} + \boldsymbol{\varepsilon}_1 \\ \vdots \\ \mathbf{y}_M = \mathbf{Z}_M\boldsymbol{\delta} + \boldsymbol{\varepsilon}_M\end{array}\right\} \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_M \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_M \end{bmatrix} \boldsymbol{\delta} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_M \end{bmatrix}$$

This is in the form of a giant regression (with common coefficients)

$$\underset{Mn\times 1}{\mathbf{y}} = \underset{(Mn\times L)}{\mathbf{Z}}\underset{(L\times 1)}{\boldsymbol{\delta}} + \underset{Mn\times 1}{\boldsymbol{\varepsilon}}$$

Assume each equation m has k_m instruments $\mathbf{x}_{im}$ satisfying

$$E[\mathbf{x}_{im}\varepsilon_{im}] = 0$$

Very often the same instruments $\mathbf{x}_i$ are used for each equation.

The moment conditions are

$$\underset{K\times 1}{\mathbf{g}_i(\boldsymbol{\delta})} = \begin{bmatrix} \mathbf{x}_{i1}\varepsilon_{i1} \\ \vdots \\ \mathbf{x}_{iM}\varepsilon_{iM} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{i1}(y_{i1} - \mathbf{z}'_{i1}\boldsymbol{\delta}) \\ \vdots \\ \mathbf{x}_{iM}(y_{iM} - \mathbf{z}'_{iM}\boldsymbol{\delta}) \end{bmatrix}$$

There are a total of $K = \sum_{m=1}^{M} k_m$ moment conditions.

Identification

$$\begin{aligned} E[\mathbf{g}_i(\boldsymbol{\delta})] &= \mathbf{0} \text{ for } \boldsymbol{\delta} = \boldsymbol{\delta}_0 \\ &\neq \mathbf{0} \text{ for } \boldsymbol{\delta} \neq \boldsymbol{\delta}_0 \end{aligned}$$

Identification implies that

$$\begin{bmatrix} E[\mathbf{x}_{i1} y_{i1}] \\ \vdots \\ E[\mathbf{x}_{iM} y_{iM}] \end{bmatrix} - \begin{bmatrix} E[\mathbf{x}_{i1} \mathbf{z}'_{i1}] \\ \vdots \\ E[\mathbf{x}_{iM} \mathbf{z}'_{iM}] \end{bmatrix} \boldsymbol{\delta}_0 = \mathbf{0}$$

or

$$\underset{K\times 1}{\boldsymbol{\sigma}_{xy}} - \underset{(K\times L)}{\boldsymbol{\Sigma}_{xz}} \underset{(L\times 1)}{\boldsymbol{\delta}_0} = \underset{K\times 1}{\mathbf{0}}$$

For $\boldsymbol{\delta}_0$ to be the unique solution requires the rank condition

$$\text{rank}(\boldsymbol{\Sigma}_{xz}) = L$$

Remarks:

- In the model with common coefficients, some equations may be individually unidentified but identified within the system.

- A sufficient condition for identification is that $E[\mathbf{x}_{im}\mathbf{z}'_{im}]$ have full rank L for some equation m.

The moments $\mathbf{g}_i(\boldsymbol{\delta})$ are a MDS with

$$E[\mathbf{g}_i(\boldsymbol{\delta}_0)\mathbf{g}_i(\boldsymbol{\delta}_0)'] = \mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & \cdots & \mathbf{S}_{1M} \\ \vdots & \ddots & \vdots \\ \mathbf{S}_{M1} & \cdots & \mathbf{S}_{MM} \end{pmatrix}, \ \mathbf{S}_{mh} = E[\mathbf{x}_{im}\mathbf{x}_{ih}'\varepsilon_{im}\varepsilon_{ih}]$$

Efficient GMM estimation

$$\begin{aligned} \min_{\delta} \ J(\boldsymbol{\delta}, \hat{\mathrm{S}}^{-1}) &= n\mathbf{g}_n(\boldsymbol{\delta})'\hat{\mathrm{S}}^{-1}\mathbf{g}_n(\boldsymbol{\delta}) \\ \mathbf{g}_n(\boldsymbol{\delta}) &= \mathbf{S}_{xy} - \mathbf{S}_{xz}\boldsymbol{\delta} \\ \mathbf{S}_{xy} &= \begin{bmatrix} \mathbf{S}_{x_1y_1} \\ \vdots \\ \mathbf{S}_{x_My_M} \end{bmatrix}, \ \mathbf{S}_{xz} = \begin{bmatrix} \mathbf{S}_{x_1z_1} \\ \vdots \\ \mathbf{S}_{x_Mz_M} \end{bmatrix} \end{aligned}$$

Straightforward algebra gives

$$\boldsymbol{\delta}(\hat{\mathrm{S}}^{-1}) = (\mathbf{S}_{xz}'\hat{\mathrm{S}}^{-1}\mathbf{S}_{xz})^{-1}\mathbf{S}_{xz}'\hat{\mathrm{S}}^{-1}\mathbf{S}_{xy}$$

Special cases

1. 3SLS with common coefficients

$$\begin{aligned} & \text{conditional homoskedasticity} \\ \mathbf{x}_{i1} &= \mathbf{x}_{i2} = \cdots = \mathbf{x}_{iM} = \mathbf{x}_i \\ \mathbf{S} &= \mathbf{S}_{3SLS} = \mathbf{\Sigma} \otimes E[\mathbf{x}_i \mathbf{x}_i'] \end{aligned}$$

2. SUR with common coefficients (Random effects)

$$\begin{aligned} & \text{conditional homoskedasticity} \\ \mathbf{x}_{i1} &= \mathbf{x}_{i2} = \cdots = \mathbf{x}_{iM} = \mathbf{x}_i \\ \mathbf{x}_i &= \text{union}(\mathbf{z}_{i1}, \ldots, \mathbf{z}_{iM}) \\ \mathbf{S} &= \mathbf{S}_{SUR} = \mathbf{\Sigma} \otimes E[\mathbf{z}_i \mathbf{z}_i'] \end{aligned}$$

Random Effects Estimator (Efficient GMM with common coefficients)

The SUR model with common coefficients is the giant regression

$$
\begin{aligned}
\underset{Mn\times 1}{\mathbf{y}} &= \underset{(Mn\times L)}{\mathbf{Z}}\underset{(L\times 1)}{\boldsymbol{\delta}} + \boldsymbol{\varepsilon},\ \mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_M \end{bmatrix} \\
E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] &= \boldsymbol{\Sigma} \otimes \mathbf{I}_n \\
\boldsymbol{\Sigma} &= \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1M} \\ \vdots & \ddots & \vdots \\ \sigma_{M1} & \cdots & \sigma_{MM} \end{pmatrix}
\end{aligned}
$$

The FGLS estimator is the random effects (RE) estimator and has the form

$$
\hat{\boldsymbol{\delta}}_{RE} = (\mathbf{Z}'(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{I}_n)\mathbf{Z})^{-1}\mathbf{Z}'(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{I}_n)\mathbf{y}
$$

The elements of $\hat{\mathbf{\Sigma}}$ are usually estimated using the pooled OLS estimator of the giant regression

$$\hat{\boldsymbol{\delta}}_{OLS} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$$

and forming

$$\hat{\sigma}_{mh} = (\mathbf{y}_m - \mathbf{Z}_m\hat{\boldsymbol{\delta}}_{OLS})'(\mathbf{y}_h - \mathbf{Z}_h\hat{\boldsymbol{\delta}}_{OLS})/n$$

Note: Hayashi uses a somewhat non-standard version of the RE estimator. He assumes that $\mathbf{\Sigma}$ is unstructured. Most other treatments of the RE estimator assume a special structure for $\mathbf{\Sigma}$ based on an error components representation for ε_{im}. In this case, the FGLS estimator uses a different estimator for $\mathbf{\Sigma}$.

Simplifying the RE estimator

It turns out that the RE estimator may also be derived from an alternative representation of the giant regression.

In the alternative representation, the system is stacked by observations instead of by equations. We can do this because $\boldsymbol{\delta}$ is the same across equations. Define

$$\underset{M\times 1}{\mathbf{y}_i} = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{iM} \end{bmatrix}, \quad \underset{M\times L}{\mathbf{Z}_i} = \begin{bmatrix} \mathbf{z}'_{i1} \\ \vdots \\ \mathbf{z}'_{iM} \end{bmatrix}, \quad \underset{M\times 1}{\boldsymbol{\varepsilon}_i} = \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iM} \end{bmatrix}$$

Then, the model for the i^{th} individual across all equations is

$$\begin{aligned} \underset{M\times 1}{\mathbf{y}_i} &= \underset{(M\times L)}{\mathbf{Z}_i} \underset{(L\times 1)}{\boldsymbol{\delta}} + \underset{M\times 1}{\boldsymbol{\varepsilon}_i}, \; i = 1, \ldots, n \\ E[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i'] &= \underset{M\times M}{\boldsymbol{\Sigma}} \\ E[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_j'] &= 0 \end{aligned}$$

Example: 2 period panel wage equation

$$\begin{aligned}
\mathbf{y}_i &= \begin{pmatrix} LW69_i \\ LW80_i \end{pmatrix}, \mathbf{Z}_i = \begin{pmatrix} 1 & S69_i & IQ_i & EXPR69_i \\ 1 & S80_i & IQ_i & EXPR80_i \end{pmatrix} \\
\boldsymbol{\delta} &= (\phi, \beta, \gamma, \pi)', \ \boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2})' \\
\mathbf{\Sigma} &= \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}
\end{aligned}$$

The giant regression stacked by observations has the form

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_n \end{bmatrix} \boldsymbol{\delta} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_n \end{bmatrix}$$

or

$$\begin{aligned}
\underset{Mn\times 1}{\mathbf{y}} &= \underset{(Mn\times L)}{\mathbf{Z}}\underset{(L\times 1)}{\boldsymbol{\delta}} + \underset{Mn\times 1}{\boldsymbol{\varepsilon}} \\
\underset{Mn\times Mn}{E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}']} &= \begin{pmatrix} E[\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1'] & \cdots & E[\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_M'] \\ \vdots & \ddots & \vdots \\ E[\boldsymbol{\varepsilon}_M\boldsymbol{\varepsilon}_1'] & \cdots & E[\boldsymbol{\varepsilon}_M\boldsymbol{\varepsilon}_M'] \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{\Sigma} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \boldsymbol{\Sigma} \end{pmatrix} = \mathbf{I}_n \otimes \boldsymbol{\Sigma}
\end{aligned}$$

The FGLS or RE estimator is

$$\hat{\boldsymbol{\delta}}_{RE} = (\mathbf{Z}'(\mathbf{I}_n \otimes \hat{\boldsymbol{\Sigma}}^{-1})\mathbf{Z})^{-1}(\mathbf{Z}'(\mathbf{I}_n \otimes \hat{\boldsymbol{\Sigma}}^{-1})\mathbf{y}$$

Now

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

and

$$
\begin{aligned}
& \mathbf{Z}'(\mathbf{I}_n \otimes \hat{\boldsymbol{\Sigma}}^{-1})\mathbf{y} \\
= \ & \left[\mathbf{Z}_1' \cdots \mathbf{Z}_n'\right] \begin{pmatrix} \hat{\boldsymbol{\Sigma}}^{-1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \hat{\boldsymbol{\Sigma}}^{-1} \end{pmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} \\
= \ & \sum_{i=1}^{n} \mathbf{Z}_i' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{y}_i
\end{aligned}
$$

Therefore,

$$
\hat{\boldsymbol{\delta}}_{RE} = \left(\sum_{i=1}^{n} \mathbf{Z}_i' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{Z}_i \right)^{-1} \sum_{i=1}^{n} \mathbf{Z}_i' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{y}_i
$$

Asymptotic Properties of the RE estimator

Consistency of the RE estimator follows from the usual manipulations

$$\hat{\boldsymbol{\delta}}_{RE} - \boldsymbol{\delta} = \left(n^{-1} \sum_{i=1}^{n} \mathbf{Z}_i' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{Z}_i \right)^{-1} n^{-1} \sum_{i=1}^{n} \mathbf{Z}_i' \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\varepsilon}_i$$

as long as

$$n^{-1} \sum_{i=1}^{n} \mathbf{Z}_i' \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\varepsilon}_i \xrightarrow{p} \mathbf{0}$$

the RE estimator will be consistent. This will occur, for example, if $\mathbf{Z}_i$ is uncorrelated with ε_i (no endogeneity).

Asymptotic normality also follows from the usual manipulations:

$$\sqrt{n}(\hat{\boldsymbol{\delta}}_{RE} - \boldsymbol{\delta}) = \left(n^{-1} \sum_{i=1}^{n} \mathbf{Z}_i' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{Z}_i \right)^{-1} n^{-1/2} \sum_{i=1}^{n} \mathbf{Z}_i' \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\varepsilon}_i$$

Under the SUR model assumptions

$$n^{-1} \sum_{i=1}^{n} \mathbf{Z}_i' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{Z}_i \xrightarrow{p} E[\mathbf{Z}_i' \boldsymbol{\Sigma}^{-1} \mathbf{Z}_i]$$

$$n^{-1/2} \sum_{i=1}^{n} \mathbf{Z}_i' \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\varepsilon}_i \xrightarrow{d} N(0, E[\mathbf{Z}_i' \boldsymbol{\Sigma}^{-1} \mathbf{Z}_i])$$

so that

$$\sqrt{n}(\hat{\boldsymbol{\delta}}_{RE} - \boldsymbol{\delta}) \xrightarrow{d} N(0, E[\mathbf{Z}_i' \boldsymbol{\Sigma}^{-1} \mathbf{Z}_i]^{-1})$$

Pooled OLS again

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_n \end{bmatrix} \boldsymbol{\delta} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_n \end{bmatrix}$$

or

$$\underset{Mn\times 1}{\mathbf{y}} = \underset{(Mn\times L)}{\mathbf{Z}} \underset{(L\times 1)}{\boldsymbol{\delta}} + \underset{Mn\times 1}{\boldsymbol{\varepsilon}}$$

The pooled OLS estimator may also be computed using the giant regression stacked by observations:

$$\begin{aligned} \hat{\boldsymbol{\delta}}_{OLS} &= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} \\ &= \left(\sum_{i=1}^{n} \mathbf{Z}_i'\mathbf{Z}_i \right)^{-1} \sum_{i=1}^{n} \mathbf{Z}_i'\mathbf{y}_i \end{aligned}$$

The asymptotic properties of the OLS estimator are straightforward to derive.

For asymptotic normality,

$$\sqrt{n}(\hat{\boldsymbol{\delta}}_{OLS} - \boldsymbol{\delta}) = \left(n^{-1}\sum_{i=1}^{n}\mathbf{Z}_i'\mathbf{Z}_i\right)^{-1} n^{-1/2}\sum_{i=1}^{n}\mathbf{Z}_i'\boldsymbol{\varepsilon}_i$$

$$n^{-1}\sum_{i=1}^{n}\mathbf{Z}_i'\mathbf{Z}_i \xrightarrow{p} E[\mathbf{Z}_i'\mathbf{Z}_i]$$

$$n^{-1/2}\sum_{i=1}^{n}\mathbf{Z}_i'\boldsymbol{\varepsilon}_i \xrightarrow{d} N(\mathbf{0}, E[\mathbf{Z}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{Z}_i])$$

Note, by iterated expectations and conditional homoskedasticity

$$\begin{aligned} E[\mathbf{Z}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{Z}_i] &= E[E[\mathbf{Z}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{Z}_i|\mathbf{Z}_i] \\ &= E[\mathbf{Z}_i'E[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i']\mathbf{Z}_i]|\mathbf{Z}_i] \\ &= E[\mathbf{Z}_i'\boldsymbol{\Sigma}\mathbf{Z}_i] \end{aligned}$$

Therefore,

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\delta}}_{OLS} - \boldsymbol{\delta}) &\xrightarrow{d} E[\mathbf{Z}_i'\mathbf{Z}_i]^{-1} \times N(\mathbf{0}, E[\mathbf{Z}_i'\boldsymbol{\Sigma}\mathbf{Z}_i]) \\ &\equiv N(\mathbf{0}, E[\mathbf{Z}_i'\mathbf{Z}_i]^{-1}E[\mathbf{Z}_i'\boldsymbol{\Sigma}\mathbf{Z}_i]E[\mathbf{Z}_i'\mathbf{Z}_i]^{-1})\end{aligned}$$

Equivalently,

$$\begin{aligned}\hat{\boldsymbol{\delta}}_{OLS} &\overset{A}{\sim} N(\boldsymbol{\delta}, n^{-1}\widehat{\mathsf{avar}}(\hat{\boldsymbol{\delta}}_{OLS})) \\ \widehat{\mathsf{avar}}(\hat{\boldsymbol{\delta}}_{OLS}) &= \left(n^{-1}\sum_{i=1}^{n}\mathbf{Z}_i'\mathbf{Z}_i\right)^{-1}\left(n^{-1}\sum_{i=1}^{n}\mathbf{Z}_i'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{Z}_i\right) \\ &\quad \times \left(n^{-1}\sum_{i=1}^{n}\mathbf{Z}_i'\mathbf{Z}_i\right)^{-1}\end{aligned}$$

Question: When is pooled OLS efficient?

Introduction to Panel Data

$$\begin{aligned} y_{im} &= \mathbf{z}'_{im}\boldsymbol{\delta} + \varepsilon_{im}, \\ i &= 1, \ldots, n \text{ (individuals); } m = 1, \ldots, M \text{ (time periods)} \\ \underset{M\times 1}{\mathbf{y}_i} &= \underset{(M\times L)}{\mathbf{Z}_i}\underset{(L\times 1)}{\boldsymbol{\delta}} + \boldsymbol{\varepsilon}_i, \ \{\mathbf{y}_i, \mathbf{Z}_i\} \text{ is i.i.d.} \end{aligned}$$

Main question: Is $\mathbf{z}_{im}$ uncorrelated with ε_{im}?

1. If yes, then we have a SUR type model with common coefficients. Under conditional homoskedasticity, the appropriate estimator is the RE estimator.

2. If no, then we have a multi-equation system with 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

$$\begin{aligned} \varepsilon_{im} &= \alpha_i + \eta_{im} \\ \alpha_i &= \text{unobserved fixed effect} \\ E[\mathbf{z}_{im}\eta_{im}] &= 0 \\ E[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'|\mathbf{x}_i] &= \boldsymbol{\Sigma},\ \mathbf{x}_i = \text{union}(\mathbf{z}_{i1}, \ldots, \mathbf{z}_{iM}) \end{aligned}$$

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:

$$\begin{aligned} \text{RE model:} &\qquad E[\mathbf{z}_{im}\alpha_i] = 0 \\ \text{FE model:} &\qquad E[\mathbf{z}_{im}\alpha_i] \neq 0 \end{aligned}$$

Note: Hayashi's treatment of error components model is slightly non-standard.

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 equation 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, the endogeneity due to unobserved endogeneity (i.e., $E[\mathbf{z}_{im}\alpha_i] \neq 0$) 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, $\boldsymbol{\delta}$ 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

$$\begin{aligned} y_{im} &= \delta z_{im} + \alpha_i + \eta_{im} \\ E[z_{im}\alpha_i] &\neq 0 \\ E[z_{im}\eta_{im}] &= 0 \end{aligned}$$

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

$$\begin{aligned} \bar{y}_i &= \delta \bar{z}_i + \alpha_i + \bar{\eta}_i \\ \bar{y}_i &= \frac{1}{M}\sum_{m=1}^{M} y_{im}, \ \bar{z}_i = \frac{1}{M}\sum_{m=1}^{M} z_{im}, \\ \alpha_i &= \frac{1}{M}\sum_{m=1}^{M} \alpha_i \end{aligned}$$

Second, form the transformed regression

$$y_{im} - \bar{y}_i = \delta(z_{im} - \bar{z}_i) + (\alpha_i - \alpha_i) + \eta_{im} - \bar{\eta}_i$$

or

$$\tilde{y}_{im} = \delta \tilde{z}_{im} + \tilde{\eta}_{im}$$

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

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

Remarks

1. If $\mathbf{z}_{im}$ does not vary with m (e.g. $\mathbf{z}_{im} = \mathbf{z}_i$) then $\tilde{\mathbf{z}}_{im} = \mathbf{0}$ and we cannot estimate $\boldsymbol{\delta}$.

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

Hayashi Notation for Within Estimator

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

$$y_{im} = \mathbf{z}'_{im}\boldsymbol{\delta} + \alpha_i + \eta_{im}$$

Stack the observations over m giving

$$\underset{M\times 1}{\mathbf{y}_i} = \underset{(M\times L)}{\mathbf{Z}_i}\underset{(L\times 1)}{\boldsymbol{\delta}} + \underset{(1\times 1)}{\alpha_i}\underset{(M\times 1)}{\mathbf{1}_M} + \underset{M\times 1}{\boldsymbol{\eta}_i}$$

Define

$$\begin{aligned}
\underset{M\times M}{\mathbf{Q}_M} &= \mathbf{I}_M - \mathbf{1}_M(\mathbf{1}'_M\mathbf{1}_M)^{-1}\mathbf{1}'_M \\
&= \mathbf{I}_M - \mathbf{P}_M \\
\mathbf{P}_M &= \mathbf{1}_M(\mathbf{1}'_M\mathbf{1}_M)^{-1}\mathbf{1}'_M = M^{-1}\mathbf{1}_M\mathbf{1}'_M
\end{aligned}$$

Note

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

The transformed error components model is then

$$\begin{aligned}\mathbf{Q}_M\mathbf{y}_i &= \mathbf{Q}_M\mathbf{Z}_i\boldsymbol{\delta} + \alpha_i\mathbf{Q}_M\mathbf{1}_M + \mathbf{Q}_M\boldsymbol{\eta}_i \\ &\Rightarrow \tilde{\mathbf{y}}_i = \tilde{\mathbf{Z}}_i\boldsymbol{\delta} + \tilde{\boldsymbol{\eta}}_i\end{aligned}$$

The giant regression (stacked by observation) is

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

or

$$\underset{Mn\times 1}{\tilde{\mathbf{y}}} = \underset{(Mn\times L)}{\tilde{\mathbf{Z}}}\underset{(L\times 1)}{\boldsymbol{\delta}} + \underset{Mn\times 1}{\tilde{\boldsymbol{\eta}}}$$

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

The FE estimator is again pooled OLS on the transformed system

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

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

Least Squared Dummy Variable Model

Consider the general model

$$y_{im} = \mathbf{z}'_{im}\boldsymbol{\delta} + \alpha_i + \eta_{im}$$

Stack the observations over m giving

$$\underset{M\times 1}{\mathbf{y}_i} = \mathbf{Z}_i\boldsymbol{\delta} + \alpha_i\mathbf{1}_M + \boldsymbol{\eta}_i$$

Now create the giant regression

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_n \end{bmatrix}\boldsymbol{\delta} + \begin{bmatrix} \mathbf{1}_M & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathbf{1}_M \end{bmatrix}\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} \boldsymbol{\eta}_1 \\ \vdots \\ \boldsymbol{\eta}_n \end{bmatrix}$$

or

$$\begin{aligned} \underset{Mn\times 1}{\mathbf{y}} &= \underset{(Mn\times L)}{\mathbf{Z}}\underset{(L\times 1)}{\boldsymbol{\delta}} + \underset{(Mn\times n)}{\mathbf{D}}\underset{(n\times 1)}{\boldsymbol{\alpha}} + \underset{(Mn\times 1)}{\boldsymbol{\eta}} = \mathbf{Z}\boldsymbol{\delta} + (\mathbf{I}_n \otimes \mathbf{1}_M)\,\boldsymbol{\alpha} + \boldsymbol{\eta} \\ \mathbf{D} &= \mathbf{I}_n \otimes \mathbf{1}_M \end{aligned}$$

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 OLS estimator of $\boldsymbol{\delta}$ in the giant regression

$$\begin{aligned}
\hat{\boldsymbol{\delta}}_{FE} &= (\mathbf{Z}'\mathbf{Q}_D\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Q}_D\mathbf{y} \\
\mathbf{Q}_D &= \mathbf{I}_{Mn} - \mathbf{P}_D, \ \mathbf{P}_D = \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}', \ \mathbf{D} = \mathbf{I}_n \otimes \mathbf{1}_M
\end{aligned}$$

Now,

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

Therefore,

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

As a result

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

which is exactly the result we got when we transformed the model by subtracting off group means.

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

$$\begin{aligned}
\hat{\alpha}_i &= \bar{y}_i - \bar{\mathbf{z}}_i' \hat{\boldsymbol{\delta}}_{FE} \\
\bar{y}_i &= \frac{1}{M} \sum_{m=1}^{M} y_{im} \\
\bar{\mathbf{z}}_i &= \frac{1}{M} \sum_{m=1}^{M} \mathbf{z}_{im}
\end{aligned}$$

FE as a First Difference Estimator

Consider the error components model

$$\begin{aligned} y_{im} &= \mathbf{z}'_{im}\boldsymbol{\delta} + \alpha_i + \eta_{im} \\ E[\boldsymbol{\eta}_i\boldsymbol{\eta}'_i] &= \sigma^2_\eta \mathbf{I}_M \end{aligned}$$

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

Take 1st differences over m:

$$\Delta y_{i2} = y_{i2} - y_{i1} = \Delta\mathbf{z}'_{i2}\boldsymbol{\delta} + \Delta\eta_{i2}$$

$$\vdots$$

$$\Delta y_{iM} = y_{iM} - y_{i(M-1)} = \Delta\mathbf{z}'_{iM}\boldsymbol{\delta} + \Delta\eta_{iM}$$

Notice that differencing removes the fixed effect.

In matrix form the model is

$$\underset{(M-1)\times 1}{\mathbf{C}'\mathbf{y}_i} = \mathbf{C}'\mathbf{Z}_i\boldsymbol{\delta} + \mathbf{C}'\boldsymbol{\eta}_i$$

or (in Hayashi notation)

$$\hat{\mathbf{y}}_i = \hat{\mathbf{Z}}_i\boldsymbol{\delta} + \hat{\boldsymbol{\eta}}_i$$

where $\hat{\mathbf{y}}_i = \mathbf{C}'\mathbf{y}_i, \hat{\mathbf{Z}}_i = \mathbf{C}'\mathbf{Z}_i, \hat{\boldsymbol{\eta}}_i = \mathbf{C}'\boldsymbol{\eta}_i$ and

$$\underset{(M-1)\times M}{\mathbf{C}'} = \begin{pmatrix} -1 & 1 & 0 & & 0 \\ 0 & -1 & 1 & & 0 \\ & & & & \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}$$

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

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

or

$$\underset{(M-1)n\times 1}{\hat{\mathbf{y}}} = \underset{((M-1)n\times L)(L\times 1)}{\hat{\mathbf{Z}} \quad \boldsymbol{\delta}} + \underset{(M-1)n\times 1}{\hat{\boldsymbol{\eta}}}$$

Notice that

$$\underset{(M-1)\times(M-1)}{E[\hat{\boldsymbol{\eta}}\hat{\boldsymbol{\eta}}']} = \begin{pmatrix} \sigma_\eta^2 \mathbf{C}'\mathbf{C} & 0 & \cdots & 0 \\ 0 & \sigma_\eta^2 \mathbf{C}'\mathbf{C} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_\eta^2 \mathbf{C}'\mathbf{C} \end{pmatrix} = \sigma_\eta^2 \left[\mathbf{I}_n \otimes (\mathbf{C}'\mathbf{C}) \right]$$

Pooled GLS on the transformed regression gives

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

Here

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

The result

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

follows from the result that

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

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

The FE estimator is a GMM estimator

Using the first-difference transformation of the error components model

$$\hat{\mathbf{y}}_i = \hat{\mathbf{Z}}_i \boldsymbol{\delta} + \hat{\boldsymbol{\eta}}_i,$$

the FE estimator can be viewed as a GMM estimator of the form

$$\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) = \left(\underline{\mathsf{S}}_{xz}' \hat{\mathbf{W}} \underline{\mathsf{S}}_{xz} \right)^{-1} \underline{\mathsf{S}}_{xz}' \hat{\mathbf{W}} \underline{\mathsf{S}}_{xy}$$

where

$$\begin{aligned}
\underline{\mathsf{S}}_{xz} &= \frac{1}{n} \sum_{i=1}^{n} \hat{\mathbf{Z}}_i \otimes \mathbf{x}_i, \;\; \mathbf{x}_i = \text{union of } (z_{i1}, \ldots, z_{iM}) \\
\underline{\mathsf{S}}_{xy} &= \frac{1}{n} \sum_{i=1}^{n} \hat{\mathbf{y}}_i \otimes \mathbf{x}_i \\
\hat{\mathbf{W}} &= (\mathbf{C}'\mathbf{C})^{-1} \otimes \mathbf{S}_{xx}^{-1}
\end{aligned}$$

Asymptotic Distribution of FE estimator

Assumptions:

1. $\mathbf{y}_i = \mathbf{Z}_i\boldsymbol{\delta} + \boldsymbol{\varepsilon}_i$, $i = 1, \ldots, n$
2. $\boldsymbol{\varepsilon}_i = \alpha_i \mathbf{1}_M + \boldsymbol{\eta}_i$ (error components)
3. $\{\mathbf{y}_i, \mathbf{Z}_i\}$ is i.i.d.
4. $E[z_{im}\alpha_i] \neq 0$ (Endogeneity)
5. $E[z_{im}\eta_{ih}] = 0$ for $m, h = 1, 2 \ldots, M$

It follows that

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

where

$$\mathsf{avar}(\hat{\boldsymbol{\delta}}_{FE}) = E[\tilde{\mathbf{Z}}_i'\tilde{\mathbf{Z}}_i]^{-1}E[\tilde{\mathbf{Z}}_i'\tilde{\boldsymbol{\eta}}_i\tilde{\boldsymbol{\eta}}_i'\tilde{\mathbf{Z}}_i]E[\tilde{\mathbf{Z}}_i'\tilde{\mathbf{Z}}_i]^{-1}$$

Therefore,

$$\hat{\boldsymbol{\delta}}_{FE} \sim N(\boldsymbol{\delta}, n^{-1}\widehat{\mathsf{avar}}(\hat{\boldsymbol{\delta}}_{FE}))$$

Remark: The sandwich form of $\mathsf{avar}(\hat{\boldsymbol{\delta}}_{FE})$ suggests that it is an inefficient GMM estimator.

General Case: Conditional Heteroskedasticity

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

where $\breve{\boldsymbol{\eta}}_i = \tilde{\mathbf{y}}_i - \tilde{\mathbf{Z}}_i\hat{\boldsymbol{\delta}}_{FE}$. Then a consistent estimate for avar$(\hat{\boldsymbol{\delta}}_{FE})$ is

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

Remarks

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

Special Case: Conditional Homoskedasticity

Assume: $E[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i' | \mathbf{x}_i] = \boldsymbol{\Sigma}$, $\mathbf{x}_i =$ union$(\mathbf{z}_{i1}, \ldots, \mathbf{z}_{iM})$

Then it can be shown that (Hayashi page 332)

$$E[\tilde{\mathbf{Z}}_i' \tilde{\boldsymbol{\eta}}_i \tilde{\boldsymbol{\eta}}_i' \tilde{\mathbf{Z}}_i] = E[\tilde{\mathbf{Z}}_i' E[\tilde{\boldsymbol{\eta}}_i \tilde{\boldsymbol{\eta}}_i'] \tilde{\mathbf{Z}}_i] = E[\tilde{\mathbf{Z}}_i' \mathbf{V} \tilde{\mathbf{Z}}_i], \ \mathbf{V} = E[\tilde{\boldsymbol{\eta}}_i \tilde{\boldsymbol{\eta}}_i']$$

and

$$\hat{\mathbf{V}} = \frac{1}{n} \sum_{i=1}^{n} \breve{\boldsymbol{\eta}}_i \breve{\boldsymbol{\eta}}_i' \xrightarrow{p} E[\tilde{\boldsymbol{\eta}}_i \tilde{\boldsymbol{\eta}}_i'], \text{ where } \breve{\boldsymbol{\eta}}_i = (\tilde{\mathbf{y}}_i - \tilde{\mathbf{Z}}_i \hat{\boldsymbol{\delta}}_{FE})$$

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

$$\widehat{\mathsf{avar}}(\hat{\boldsymbol{\delta}}_{FE}) = \left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{Z}}_i'\tilde{\mathbf{Z}}_i\right)^{-1} \times \left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{Z}}_i'\hat{\mathbf{V}}\tilde{\mathbf{Z}}_i\right)^{-1} \times \left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{Z}}_i'\tilde{\mathbf{Z}}_i\right)^{-1}$$

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

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

This implies that there is no serial correlation (correlation across time)

Then

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

and

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

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

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

Remark: In this case the FE estimator is an efficient GMM estimator!

Hausman Specification Test for FE vs. RE

The hypotheses to be tested are

$$\begin{aligned} H_0 &: E[\mathbf{z}_{im}\alpha_i] = 0 \text{ (RE estimation)} \\ H_1 &: E[\mathbf{z}_{im}\alpha_i] \neq 0 \text{ (FE estimation)} \end{aligned}$$

Hausman and Taylor (1981, Ecta) considered a test statistic based on

$$\hat{\mathbf{q}} = \hat{\boldsymbol{\delta}}_{FE} - \hat{\boldsymbol{\delta}}_{RE}$$

in the context of maximum likelihood estimation. Newey (1985, see also Hayashi Chapter 3, Analytic Exercise 9) studied the test in the context of GMM.

Intuition: Under H_0, both $\hat{\boldsymbol{\delta}}_{FE}$ and $\hat{\boldsymbol{\delta}}_{RE}$ are consistent (but $\hat{\boldsymbol{\delta}}_{RE}$ is efficient) so that

$$\hat{\mathbf{q}} \xrightarrow{p} \mathbf{0}$$

Under H_1, $\hat{\boldsymbol{\delta}}_{RE}$ is not consistent but $\hat{\boldsymbol{\delta}}_{FE}$ is consistent so that

$$\hat{\mathbf{q}} \overset{p}{\nrightarrow} \mathbf{0}$$

Therefore, consider the test statistic

$$H = n\hat{\mathbf{q}} \left(\widehat{\mathsf{avar}}(\hat{\mathbf{q}})\right)^{-1} \hat{\mathbf{q}}$$

If H is big then reject H_0; otherwise do not reject H_0.

Q1: What is $\widehat{\mathsf{avar}}(\hat{\mathbf{q}})$?

Q2: What is the asymptotic distribution of H?

To answer these questions consider the general linear GMM model

$$\begin{aligned} y_i &= \mathbf{z}_i'\boldsymbol{\delta} + \varepsilon_i,\ i = \mathbf{1}, \ldots, n \\ E[\mathbf{x}_i \varepsilon_i] &= \mathbf{0},\ E[\mathbf{x}_i \mathbf{x}_i' \varepsilon_i^2] = \mathbf{S} \end{aligned}$$

Let $\hat{\boldsymbol{\delta}}_1(\hat{\mathbf{W}}_1)$ denote an inefficient GMM estimator with weight matrix $\hat{\mathbf{W}}_1 \xrightarrow{p} \mathbf{W}_1 \neq \mathbf{S}^{-1}$.

Let $\hat{\boldsymbol{\delta}}_2(\hat{\mathbf{W}}_2)$ denote another inefficient GMM estimator with weight matrix $\hat{\mathbf{W}}_2 \xrightarrow{p} \mathbf{W}_2 \neq \mathbf{W}_1 \neq \mathbf{S}^{-1}$.

The following results are based on Newey 1985 (See also Hayashi, Chapter 3, Analytic Exercise 9)

Result 1

$$\begin{bmatrix} \sqrt{n}\left(\hat{\boldsymbol{\delta}}_1(\hat{\mathbf{W}}_1) - \boldsymbol{\delta}\right) \\ \sqrt{n}\left(\hat{\boldsymbol{\delta}}_2(\hat{\mathbf{W}}_2) - \boldsymbol{\delta}\right) \end{bmatrix} \xrightarrow{d} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{12}' & \mathbf{A}_{22} \end{bmatrix}\right]$$

where

$$\begin{aligned} \mathbf{A}_{11} &= \text{avar}(\hat{\boldsymbol{\delta}}_1(\hat{\mathbf{W}}_1)) \\ \mathbf{A}_{22} &= \text{avar}(\hat{\boldsymbol{\delta}}_2(\hat{\mathbf{W}}_2)) \\ \mathbf{A}_{12} &= \text{acov}(\hat{\boldsymbol{\delta}}_1(\hat{\mathbf{W}}_1), \hat{\boldsymbol{\delta}}_2(\hat{\mathbf{W}}_2)) \end{aligned}$$

Define

$$\hat{\mathbf{q}} = \hat{\boldsymbol{\delta}}_1(\hat{\mathbf{W}}_1) - \hat{\boldsymbol{\delta}}_2(\hat{\mathbf{W}}_2)$$

Result 2:

$$\sqrt{n}\hat{\mathbf{q}} \xrightarrow{d} N(\mathbf{0}, \mathsf{avar}(\hat{\mathbf{q}}))$$

$$\begin{aligned}
\mathsf{avar}(\hat{\mathbf{q}}) &= \mathbf{A}_{11} + \mathbf{A}_{22} - \mathbf{A}_{12} - \mathbf{A}'_{12} \\
&= \mathsf{avar}(\hat{\boldsymbol{\delta}}_1(\hat{\mathbf{W}}_1)) + \mathsf{avar}(\hat{\boldsymbol{\delta}}_2(\hat{\mathbf{W}}_2)) \\
&\quad - \mathsf{acov}(\hat{\boldsymbol{\delta}}_1(\hat{\mathbf{W}}_1), \hat{\boldsymbol{\delta}}_2(\hat{\mathbf{W}}_2)) - \mathsf{acov}(\hat{\boldsymbol{\delta}}_1(\hat{\mathbf{W}}_1), \hat{\boldsymbol{\delta}}_2(\hat{\mathbf{W}}_2))'
\end{aligned}$$

Result 3: Set $\hat{\mathbf{W}}_2 = \hat{\mathbf{S}}^{-1}$ where $\hat{\mathbf{S}} \xrightarrow{p} \mathbf{S}$. Then

$$\mathbf{A}_{12} = \mathbf{A}_{12}' = \mathbf{A}_{22} = \mathsf{avar}(\hat{\boldsymbol{\delta}}_2(\hat{\mathbf{S}}^{-1}))$$

and

$$\begin{aligned} \mathsf{avar}(\hat{\mathbf{q}}) &= \mathbf{A}_{11} + \mathbf{A}_{22} - 2\mathbf{A}_{22} = \mathbf{A}_{11} - \mathbf{A}_{22} \\ &= \mathsf{avar}(\hat{\boldsymbol{\delta}}_1(\hat{\mathbf{W}}_1)) - \mathsf{avar}(\hat{\boldsymbol{\delta}}_2(\hat{\mathbf{S}}^{-1})) \end{aligned}$$

Remark: Because $\hat{\boldsymbol{\delta}}_2(\hat{\mathbf{S}}^{-1})$ is the efficient GMM estimator $\mathsf{avar}(\hat{\mathbf{q}}) \geq 0$.

Applying Newey's result to

$$\hat{\mathbf{q}} = \hat{\boldsymbol{\delta}}_{FE} - \hat{\boldsymbol{\delta}}_{RE}$$

gives

$$\text{avar}(\hat{\mathbf{q}}) = \text{avar}(\hat{\boldsymbol{\delta}}_{FE}) - \text{avar}(\hat{\boldsymbol{\delta}}_{RE})$$

and

$$H = n\hat{\mathbf{q}}\left(\widehat{\text{avar}}(\hat{\mathbf{q}})\right)^{-1}\hat{\mathbf{q}} \overset{A}{\sim} \chi^2(L)$$

Therefore, we reject $H_0 : E[\mathbf{z}_{im}\alpha_i] = 0$ at the $\alpha \times (100)\%$ level if

$$H > \chi^2_{1-\alpha}(L)$$

Remark: Hayashi shows (Appendix to chapter 5) that $\text{avar}(\hat{\mathbf{q}}) > 0$ so that $\text{avar}(\hat{\mathbf{q}})^{-1}$ is well defined.

Remarks

The form of avar($\hat{\mathrm{q}}$) $=$ avar($\hat{\boldsymbol{\delta}}_{FE}$) $-$ avar($\hat{\boldsymbol{\delta}}_{RE}$) depends on assumptions about heteroskedasticity.

Typical case: Assume conditional homoskedasticity

$$\sqrt{n}(\hat{\boldsymbol{\delta}}_{RE} - \boldsymbol{\delta}) \xrightarrow{d} N(0, E[\mathbf{Z}_i' \boldsymbol{\Sigma}^{-1} \mathbf{Z}_i]^{-1})$$

$$\sqrt{n}\left(\hat{\boldsymbol{\delta}}_{FE} - \boldsymbol{\delta}\right) \xrightarrow{d}$$

$$N(0, E[\tilde{\mathbf{Z}}_i' \tilde{\mathbf{Z}}_i]^{-1} E[\tilde{\mathbf{Z}}_i' E[\tilde{\boldsymbol{\eta}}_i \tilde{\boldsymbol{\eta}}_i'] \tilde{\mathbf{Z}}_i] E[\tilde{\mathbf{Z}}_i' \tilde{\mathbf{Z}}_i]^{-1})$$

Then

$$\begin{aligned}
\widehat{\mathsf{avar}}(\hat{\boldsymbol{\delta}}_{FE}) &= \left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{Z}}_i'\tilde{\mathbf{Z}}_i\right)^{-1} \times \left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{Z}}_i'\hat{\mathbf{V}}\tilde{\mathbf{Z}}_i\right)^{-1} \\
&\quad \times \left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{Z}}_i'\tilde{\mathbf{Z}}_i\right)^{-1} \\
\widehat{\mathsf{avar}}(\hat{\boldsymbol{\delta}}_{RE}) &= \left(n^{-1}\sum_{i=1}^{n}\mathbf{Z}_i'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{Z}_i\right)
\end{aligned}$$