Econometric Methods for Panel Data

Based on the books by BALTAGI: *Econometric Analysis of Panel Data* and by HSIAO: *Analysis of Panel Data*

Robert M. Kunst
robert.kunst@univie.ac.at

University of Vienna
and
Institute for Advanced Studies Vienna

May 4, 2010

Outline

The definition of panel data

The word *panel* has a Dutch origin and it essentially means a board. Data for a variable on a board is two-dimensional, the variable X has two subscripts. One dimension is an *individual* index (i), and the other dimension is *time* (t):

$$\begin{bmatrix} X_{11} & \dots & X_{1T} \\ \vdots & X_{it} & \vdots \\ X_{N1} & \dots & X_{NT} \end{bmatrix}$$

Long and broad boards

$$\begin{bmatrix} X_{11} & \dots & \dots & \dots & X_{1T} \\ \vdots & \dots & X_{it} & \dots & \vdots \\ X_{N1} & \dots & \dots & \dots & X_{NT} \end{bmatrix} \quad \begin{bmatrix} X_{11} & \dots & X_{1T} \\ \vdots & \vdots & \vdots \\ \vdots & X_{it} & \vdots \\ \vdots & \vdots & \vdots \\ X_{N1} & \dots & X_{NT} \end{bmatrix}$$

If $T \gg N$, the panel is a *time-series panel*, as it is often encountered in macroeconomics. If $N \gg T$, it is a *cross-section panel*, as it is common in microeconomics.

Not quite panels

- *longitudinal data* sets look like panels but the time index may not be common across individuals. For examples, growing plants may be measured according to individual time;
- in *pseudo-panels*, individuals may change between time points: X_{it} and $X_{i,t+1}$ may relate to different persons;
- in *unbalanced panels*, T differs among individuals and is replaced by T_i: no more matrix or 'board' shape.

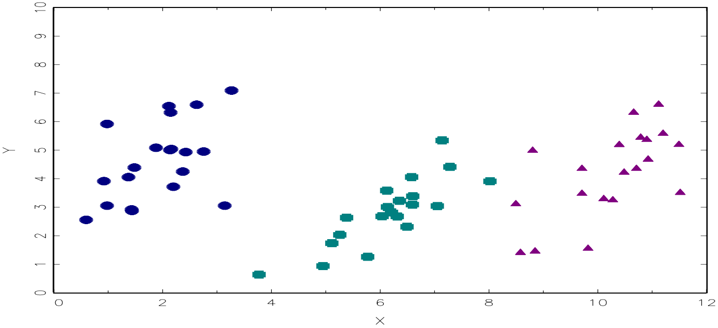
The two dimensions have different quality

- In the time dimension t, the panel behaves like a time series: natural ordering, systematic dependence over time, asymptotics depend on stationarity, ergodicity etc.
- In the cross-section dimension i, there is no natural ordering, cross-section dependence may play a role ('second generation'), otherwise asymptotics may also be simple assuming independence ('first generation'). Sometimes, e.g. in *spatial panels*, i may have structure and natural ordering.

Advantages of panel data analysis

- Panels are more informative than simple time series of aggregates, as they allow tracking individual histories. A 10% unemployment rate is less informative than a panel of individuals with all of them unemployed 10% of the time or one with 10% always unemployed;
- Panels are more informative than cross-sections, as they reflect dynamics and Granger causality across variables.

The heterogeneity problem



$N = 3$, $T = 20$. For each i, there is an identical, positive slope in a linear relationship between Y and X. For the whole sample, the relationship is slightly falling and nonlinear. If interest focuses on the former model, naive estimation over the whole sample results in a *heterogeneity bias*.

Books on panels

BALTAGI, B. *Econometric Analysis of Panel Data*, Wiley. (textbook)

HSIAO, C. *Analysis of Panel Data*, Cambridge University Press. (textbook)

ARELLANO, M. *Panel Data Econometrics*, Oxford University Press. (very formal state of the art)

DIGGLE, P., HEAGERTY, P., LIANG, K.Y., AND S. ZEGER *Analysis of Longitudinal Data*, Oxford University Press. (non-economic monograph)

The pooled regression model

Consider the model

$$y_{it} = \alpha + \beta' X_{it} + u_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T.$$

Assume there are K regressors (covariates), such that $\dim(\beta) = K$.

Panel models mainly differ in their assumptions on u. u independent across i and t, $\mathrm{E}u = 0$, and $\mathrm{var}\, u = \sigma^2$ define the (usually unrealistic) *pooled regression model*. It is efficiently estimated by least squares (OLS).

Sometimes, one may consider digressing from the homogeneity assumption $\beta_i \equiv \beta$. This entails that most advantages of panel modelling are lost.

The fixed-effects regression

The model

$$\begin{aligned} y_{it} &= \alpha + \beta' X_{it} + u_{it}, \\ u_{it} &= \mu_i + \nu_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \end{aligned}$$

with the identifying condition $\sum_{i=1}^{N} \mu_i = 0$, and the *individual effects* μ_i assumed as unobserved constants (parameters), is the *fixed-effects (FE) regression* model. (ν_{it}) fulfills the usual conditions on errors: independent, $\mathrm{E}\nu = 0$, $\mathrm{var}\nu = \sigma_\nu^2$.

Incidental parameters

Most authors call a parameter *incidental* when its dimension increases with the sample size. The nuisance due to such a parameter is worse than for a typical nuisance parameter whose dimension may be constant.

As $N \to \infty$, the number of fixed effects μ_i increases. Thus, the fixed effects are incidental parameters.

Incidental parameters cannot be consistently estimated, and they may cause inconsistency in the ML estimation of the remaining parameters.

OLS in the FE regression model

In the FE model, the regression equation

$$y_{it} = \alpha + \beta' X_{it} + u_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T,$$

does not fulfill the conditions of the Gauss-Markov Theorem, as $\mathrm{E} u_{it} = \mu_i \neq 0$. OLS may be biased, inconsistent, and even if it is unbiased, it is usually inefficient.

Least-squares dummy variables

Let $Z_{\mu,it}^{(j)}$ denote a dummy variable that is 0 for all observations *it* with $i \neq j$ and 1 for $i = j$. Then, convening $Z_{\mu,it} = (Z_{\mu,it}^{(1)}, \ldots, Z_{\mu,it}^{(N)})'$ and $\mu = (\mu_1, \ldots, \mu_N)'$, the regression model

$$y_{it} = \alpha + \beta' X_{it} + \mu' Z_{\mu,it} + \nu_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T,$$

fulfills all conditions of the Gauss-Markov Theorem. OLS for this regression is called LSDV (least-squares dummy variables), the *within*, or the FE estimator. Assuming X as non-stochastic, LSDV is unbiased, consistent, and linear efficient (BLUE).

The dummy-variable trap in LSDV

Note that $\sum_{j=1}^{N} Z_{\mu,it}^{(j)} = 1$. Inhomogeneous LSDV regression would be multicollinear. Two (equivalent) solutions:

1. restricted OLS imposing $\sum_{i=1}^{N} \mu_i = 0$;
2. homogeneous regression with free μ_i^* coefficients. Then, $\hat{\alpha}$ is recovered from $\sum_{i=1}^{N} \hat{\mu}_i^*$, and $\hat{\mu}_i = \hat{\mu}_i^* - \hat{\alpha}$.

In any case, parameter dimension is $K + N$.

Within and between

By conditioning on individual ('group') dummies, the within or within-groups estimator concentrates exclusively on variation within the individuals.

By contrast, the *between estimator* results from a regression among N individual time averages:

$$\bar{y}_{i.} = \alpha + \beta' \bar{X}_{i.} + \bar{u}_{i.}, \quad i = 1 \ldots, N,$$

with $\bar{y}_{i.} = T^{-1} \sum_{t=1}^{T} y_{it}$ etc. Because of $\mathrm{E}\bar{u}_{i.} = \mu_i \neq 0$, it violates the Gauss-Markov conditions and is more of theoretical interest.

LSDV regression in matrix form

Stacking all NT observations yields the compact form

$$y = \alpha\iota_{NT} + \mathbf{X}\beta + \mathbf{Z}_\mu\mu + \nu,$$

where ι_{NT}, y, and ν are NT–vectors. Generally, ι_m stands for an m–vector of ones. Convention is that i is the 'slow' index and t the 'fast' index, such that the first T observations belong to $i = 1$. $\mathbf{X}$ is an $NT \times K$–matrix, β is a K–vector, μ is an N–vector. $\mathbf{Z}_\mu$ is an $NT \times N$–matrix.

The matrix $\mathbf{Z}_\mu$

The $NT \times N$–matrix for the dummy regressors looks like

$$\mathbf{Z}_\mu = \begin{bmatrix} 1 & 0 & \cdots & & 0 \\ \vdots & \vdots & & & \vdots \\ 1 & 0 & & & \\ 0 & 1 & & & \\ \vdots & \vdots & & & \\ 0 & 1 & & & \\ & & \ddots & & \\ & & & 0 & 1 \\ \vdots & & & & \vdots \\ 0 & \cdots & & 0 & 1 \end{bmatrix}.$$

This matrix can be written in Kronecker notation as $\mathbf{I}_N \otimes \iota_T$. $\mathbf{I}_N$ is the $N \times N$ identity matrix.

Review: Kronecker products

The Kronecker product $\mathbf{A} \otimes \mathbf{B}$ of two matrices $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{kl}]$ of dimensions $n \times m$ and $n_1 \times m_1$ is defined by

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \ldots & a_{1m}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \ldots & a_{2m}\mathbf{B} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1}\mathbf{B} & a_{n2}\mathbf{B} & \ldots & a_{nm}\mathbf{B} \end{bmatrix},$$

which gives a large $(nn_1) \times (mm_1)$–matrix. Left factor determines 'crude' form and right factor determines 'fine' form. (Note: some authors use different definitions, t.ex. David Hendry)

$\mathbf{I} \otimes \mathbf{B}$ is a block-diagonal matrix.

Three calculation rules for Kronecker products

$$
\begin{aligned}
(\mathbf{A} \otimes \mathbf{B})' &= \mathbf{A}' \otimes \mathbf{B}' \\
(\mathbf{A} \otimes \mathbf{B})^{-1} &= \mathbf{A}^{-1} \otimes \mathbf{B}^{-1} \\
(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) &= (\mathbf{AC} \otimes \mathbf{BD})
\end{aligned}
$$

for non-singular matrices (second rule) and fitting dimensions (third rule).

The matrix $\mathbf{Z}_\mu$ is very big

Direct regression of y on the full $NT \times (N+K)$–matrix is feasible (unless N is too large) but inconvenient:

- This straightforward regression does not exploit the simple structure of the matrix $\mathbf{Z}_\mu$;
- it involves inversion of an $(N+K) \times (N+K)$–matrix, an obstacle especially for cross-section panels.

Therefore, the regression is run in two steps.

The Frisch-Waugh theorem

Theorem

The OLS estimator on the partitioned regression

$$y = X\beta + Z\gamma + u$$

can be obtained by first regressing y and all X variables on Z, which yields residuals y_Z and X_Z, and then regressing y_Z on X_Z in

$$y_Z = X_Z\beta + v.$$

The resulting estimate $\hat{\beta}$ is identical to the one obtained from the original regression.

Remarks on Frisch-Waugh

- Note that the outlined sequence does not yield an OLS estimate of γ. If that is really needed, one may run the same sequence with X and Z exchanged.
- Frisch-Waugh emphasizes that OLS coefficients in any multiple regression are really effects conditional on all remaining regressors.
- Frisch-Waugh permits running a multiple regression on two regressors if only simple regression is available.
- Proof by direct evaluation of the regressions, using inversion of partitioned matrices.

Interpreting regression on Z_μ

If any variable y is regressed on Z_μ, the residuals are

$$\begin{aligned}
& y - \mathbf{Z}_\mu \left(\mathbf{Z}_\mu' \mathbf{Z}_\mu\right)^{-1} \mathbf{Z}_\mu' y \\
= \; & y - (\mathbf{I}_N \otimes \iota_T)\{(\mathbf{I}_N \otimes \iota_T)'(\mathbf{I}_N \otimes \iota_T)\}^{-1}(\mathbf{I}_N \otimes \iota_T)' y \\
= \; & y - T^{-1}(\mathbf{I}_N \otimes \iota_T \iota_T') y \\
= \; & y - T^{-1}\left(\sum_{t=1}^{T} y_{1t}, \ldots, \sum_{t=1}^{T} y_{1t}, \sum_{t=1}^{T} y_{2t}, \ldots, \sum_{t=1}^{T} y_{Nt}\right)' \\
= \; & y - (\bar{y}_{1.}, \ldots, \bar{y}_{N.})',
\end{aligned}$$

such that all observations are adjusted for their individual time averages. This is done for all variables, y and the covariates X.

Applying Frisch-Waugh to LSDV

1. In the first step, y and all regressors in $\mathbf{X}$ are regressed on $\mathbf{Z}_\mu$ and the residuals are formed, i.e. the observations corrected for their individual time averages;
2. in the second step, the 'purged' y observations are regressed on the purged covariates. $\hat{\beta}$ is obtained.

The procedure fails for time-constant variables. An OLS estimate $\hat{\mu}$ follows from regressing the residuals $y - \mathbf{X}\hat{\beta}$ on dummies.

The purging matrix $\mathbf{Q}$

The matrix that defines the purging (or sweep-out) operation in the first step is $\mathbf{Q} = \mathbf{I}_{NT} - T^{-1}(\mathbf{I}_N \otimes \mathbf{J}_T)$ with $\mathbf{J}_T = \iota_T \iota_T'$ a $T \times T$–matrix filled with ones:

$$\begin{aligned} \tilde{y} &= y - (\bar{y}_{1.}, \ldots, \bar{y}_{N.})' \otimes \iota_T \\ &= \mathbf{Q}y \end{aligned}$$

Using this matrix allows to write the FE estimator compactly:

$$\begin{aligned} \hat{\beta}_{FE} &= (\mathbf{X}'\mathbf{Q}'\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}'\mathbf{Q}y \\ &= (\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}y, \end{aligned}$$

as $\mathbf{Q}' = \mathbf{Q}$ and $\mathbf{Q}^2 = \mathbf{Q}$.

The variance matrix of the LSDV estimator

The LSDV estimator looks like a GLS estimator, and the algebra for its variance follows the GLS variance:

$$\mathrm{var}\hat{\beta}_{FE} = \sigma_\nu^2(\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}$$

Note: This differs from usual GLS algebra, as $\mathbf{Q}$ is singular. There is no GLS model that motivates FE, though the result is analogous.

N and T asymptotics of the LSDV estimator

If $\mathrm{E}\hat{\beta}_{FE} = \beta$ and $\mathrm{var}\hat{\beta}_{FE} \to 0$, $\hat{\beta}_{FE}$ will be consistent.

1. If $T \to \infty$, $(\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}$ converges to 0, assuming that all covariates show constant or increasing variation around their individual means, which is plausible (excludes time-constant covariates).
2. If $N \to \infty$, $(\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}$ converges to 0, assuming that covariates vary across individuals (excludes covariates indicative of time points).

$T \to \infty$ allows to retrieve consistent estimates of all 'effects' μ_i. For $N \to \infty$, this is not possible.

t–statistics on coefficients

1. Plugging in a sensible estimator $\hat{\sigma}_\nu^2$ for σ_ν^2 in

$$\Sigma_{FE} = \sigma_\nu^2(\mathbf{X}'\mathbf{QX})^{-1}$$

   yields $\hat{\Sigma}_{FE} = [\hat{\sigma}_{FE,jk}]_{j,k=1,\ldots,K}$ and allows to construct t–values for the vector β:

$$t_\beta = \mathrm{diag}(\hat{\sigma}_{FE,11}^{-1/2}, \ldots, \hat{\sigma}_{FE,NN}^{-1/2})\hat{\beta}_{FE}.$$

   A suggestion for $\hat{\sigma}_\nu^2$ is $(NT - N - K)^{-1} \sum_{i=1}^N \sum_{t=1}^T \hat{\nu}_{it}^2$.

2. t–statistics on μ_i and α are obtained by evaluating the full model in an analogous way.

These t–values are, under their null, asymptotically $N(0,1)$. Under Gauss-Markov assumptions, they are t distributed in small samples. Statistics for effects need $T \to \infty$.

The bias of pooled regression in the FE model

The ‘pooled’ OLS estimate in the FE model is

$$\hat{\beta}_{\#} = (\mathbf{X}'_{\#}\mathbf{X}_{\#})^{-1}\mathbf{X}'_{\#}y,$$

where $X_{\#}$ etc. denotes the X matrix extended by the intercept column. As in usual derivation of ‘omitted-variable bias’:

$$\mathrm{E}\left(\hat{\beta}_{\#}\right) = \beta_{\#} + \left(\mathbf{X}'_{\#}\mathbf{X}_{\#}\right)^{-1}\mathbf{X}'_{\#}\mathbf{Z}_{\mu}\mu,$$

so the bias depends on the correlation of X and Z_{μ}, on T, and on μ. For $T \to \infty$, the bias should disappear.

Idea of the random-effects model

If N is large, one may view the 'effects' as unobserved random variables and not as incidental parameters: *random effects* (RE). The model becomes more 'parsimonious', as it has less parameters.

- It should be plausible that all effects are drawn from the same probability distribution. Strong heterogeneity across individuals invalidates the RE model.
- The concept is unattractive for small N.
- The concept is unattractive for large T.
- The concept presumes that covariates and effects are approximately independent. This assumption may often be implausible a priori.

The random-effects model

The basic one-way RE model (or variance-components model) reads:

$$\begin{aligned} y_{it} &= \alpha + \beta' X_{it} + u_{it}, \\ u_{it} &= \mu_i + \nu_{it}, \\ \mu_i &\sim i.i.d. \left(0, \sigma_\mu^2\right), \\ \nu_{it} &\sim i.i.d. \left(0, \sigma_\nu^2\right), \end{aligned}$$

for $i = 1, \ldots, N$, and $t = 1, \ldots, T$. μ and ν are mutually independent.

GLS estimation for the RE model

The RE model corresponds to a usual GLS regression model

$$
\begin{aligned}
y &= \alpha\iota_{NT} + \mathbf{X}'\beta + u, \\
\operatorname{var} u &= \sigma_\mu^2(\mathbf{I}_N \otimes \mathbf{J}_T) + \sigma_\nu^2 \mathbf{I}_{NT}.
\end{aligned}
$$

Denoting the error variance matrix by $\mathrm{E}uu' = \boldsymbol{\Omega}$, the GLS estimator is

$$\hat{\beta}_{RE} = \left\{\mathbf{X}_{\#}'\boldsymbol{\Omega}^{-1}\mathbf{X}_{\#}\right\}^{-1}\mathbf{X}_{\#}'\boldsymbol{\Omega}^{-1}y.$$

This estimator is BLUE for the RE model.

The matrix $\mathbf{\Omega}$ is big

GLS estimation involves the inversion of the $(NT \times NT)$–matrix $\mathbf{\Omega}$, which may be inconvenient. It is easily inverted, however, from a 'spectral' decomposition into orthogonal parts. $\bar{\mathbf{J}}_T$ is a $T \times T$–matrix of elements T^{-1}.

$$\begin{aligned}
\mathbf{\Omega} &= T\sigma_\mu^2 \left(\mathbf{I}_N \otimes \bar{\mathbf{J}}_T\right) + \sigma_\nu^2 \mathbf{I}_{NT} \\
&= \left(T\sigma_\mu^2 + \sigma_\nu^2\right)\left(\mathbf{I}_N \otimes \bar{\mathbf{J}}_T\right) + \sigma_\nu^2 \left(\mathbf{I}_{NT} - \mathbf{I}_N \otimes \bar{\mathbf{J}}_T\right) \\
&= \left(T\sigma_\mu^2 + \sigma_\nu^2\right)\mathbf{P} + \sigma_\nu^2 \mathbf{Q},
\end{aligned}$$

where $\mathbf{P}$ and $\mathbf{Q}$ have the properties

$$\mathbf{PQ} = \mathbf{QP} = \mathbf{0}, \quad \mathbf{P}^2 = \mathbf{P}, \quad \mathbf{Q}^2 = \mathbf{Q}, \quad \mathbf{P} + \mathbf{Q} = \mathbf{I}.$$

Inversion of $\boldsymbol{\Omega}$

Because of the special properties of the spectral decomposition, $\boldsymbol{\Omega}$ can be inverted componentwise:

$$\boldsymbol{\Omega}^{-1} = \left(T\sigma_\mu^2 + \sigma_\nu^2\right)^{-1}\left(\mathbf{I}_N \otimes \bar{\mathbf{J}}_T\right) + \sigma_\nu^{-2}\left(\mathbf{I}_{NT} - \mathbf{I}_N \otimes \bar{\mathbf{J}}_T\right).$$

Thus, the GLS estimator has the simpler closed form

$$\begin{aligned}\hat{\beta}_{\#,RE} &= \left[\mathbf{X}_\#' \left\{\left(T\sigma_\mu^2 + \sigma_\nu^2\right)^{-1}\mathbf{P} + \sigma_\nu^{-2}\mathbf{Q}\right\}\mathbf{X}_\#\right]^{-1} \times \\ &\quad \times \mathbf{X}_\#' \left\{\left(T\sigma_\mu^2 + \sigma_\nu^2\right)^{-1}\mathbf{P} + \sigma_\nu^{-2}\mathbf{Q}\right\} y,\end{aligned}$$

which only requires inversion of a $(K+1) \times (K+1)$–matrix.

Splitting $\boldsymbol{\Omega}$

Any GLS estimator $(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$ can also be represented in the form $\{(MX)'MX\}^{-1}(MX)'My$, as OLS on data transformed by the square root of Ω^{-1}. Because of $\mathbf{P}^2 = \mathbf{P}$ etc., this form is very simple here:

$$\hat{\beta}_{\#,RE} = \left\{ \left(\tilde{\mathbf{P}} \mathbf{X}_{\#} \right)' \left(\tilde{\mathbf{P}} \mathbf{X}_{\#} \right) \right\}^{-1} \left(\tilde{\mathbf{P}} \mathbf{X}_{\#} \right)' \tilde{\mathbf{P}} y,$$

with

$$\tilde{\mathbf{P}} = \left(\sqrt{T\sigma_\mu^2 + \sigma_\nu^2} \right)^{-1} \mathbf{P} + \sigma_\nu^{-1} \mathbf{Q}.$$

The weight parameter θ

The GLS estimator is invariant to any re-scaling of the transformation matrix (cancels out). For example,

$$\hat{\beta}_{\#,RE} = \left\{ \left(\tilde{\mathbf{P}}_1 \mathbf{X}_\# \right)' \left(\tilde{\mathbf{P}}_1 \mathbf{X}_\# \right) \right\}^{-1} \left(\tilde{\mathbf{P}}_1 \mathbf{X}_\# \right)' \tilde{\mathbf{P}}_1 y$$

with

$$\begin{aligned} \tilde{\mathbf{P}}_1 &= \frac{\sigma_\nu}{\sqrt{T\sigma_\mu^2 + \sigma_\nu^2}} \mathbf{P} + \mathbf{Q} \\ &= (1-\theta)\,\mathbf{P} + \mathbf{Q}, \end{aligned}$$

and

$$\theta = 1 - \frac{\sigma_\nu}{\sqrt{T\sigma_\mu^2 + \sigma_\nu^2}}.$$

The value of θ determines the RE estimate

Note the following cases:

1. $\theta = 0$, the transformation is $\mathbf{I}$, RE-GLS becomes OLS. This happens if $\sigma_\mu^2 = 0$, no 'effects';
2. $\theta = 1$, the transformation is $\mathbf{Q}$, RE-GLS becomes FE-LSDV. This happens if T is very large or $\sigma_\nu^2 = 0$ or σ_μ^2 is large;
3. The transformation can never become just $\mathbf{P}$, which would yield the *between* estimator.

θ must be estimated

In practice, RE-GLS requires the unknown weight parameter

$$\theta = 1 - \frac{\sigma_\nu}{\sqrt{T\sigma_\mu^2 + \sigma_\nu^2}}$$

to be estimated.

Most algorithms use a consistent (inefficient, non-GLS) estimator in a first step, in order to estimate variances from residuals. Some algorithms iterate this idea. Some use joint estimation of all parameters by nonlinear ML-type optimization.

OLS and LSDV are consistent in the RE model

The RE model is a traditional GLS model. OLS is consistent and (for fixed regressors) unbiased though not efficient. LSDV is also consistent and unbiased for β. It may be more efficient, as typically θ is closer to one, where LSDV and RE coincide.

Thus, LSDV and OLS are appropriate first-step routines for estimating θ.

Estimating numerator and denominator

A customary method is to estimate $\sigma_1^2 = T\sigma_\mu^2 + \sigma_\nu^2$ by

$$\hat{\sigma}_1^2 = \frac{T}{N} \sum_{i=1}^{N} \left(\frac{1}{T} \sum_{t=1}^{T} \hat{u}_{it} \right)^2$$

and σ_ν^2 by

$$\hat{\sigma}_\nu^2 = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \left(\hat{u}_{it} - T^{-1} \sum_{s=1}^{T} \hat{u}_{is} \right)^2}{N\,(T-1)}.$$

$\hat{u}$ may be OLS or LSDV residuals. A drawback is that the implied estimate $\hat{\sigma}_\mu^2 = T^{-1}\left(\hat{\sigma}_1^2 - \hat{\sigma}_\nu^2\right)$ can be negative and inadmissible.

Estimating σ_μ^2 and σ_ν^2

The NERLOVE method estimates σ_μ^2 directly from the effect estimates in a first-step LSDV regression, and forms $\hat{\theta}$ accordingly.

Estimating σ_1^2, however, is no coincidence. That term appears in the evaluation of the likelihood. It is implied by an eigenvalue decomposition of the matrix $\mathbf{\Omega}$.

Usually, discrepancies across θ estimation methods are minor.

RE-GLS is linear efficient for the correct model

By construction, the RE estimator is linear efficient (BLUE) for the RE model. It has the GLS variance

$$\mathrm{var}\left(\hat{\beta}_{\#,RE}\right) = \left\{\mathbf{X}'_{\#}\mathbf{\Omega}^{-1}\mathbf{X}_{\#}\right\}^{-1}.$$

For the fGLS estimator, this variance is attained asymptotically, for $N \to \infty$ and $T \to \infty$.

For $T \to \infty$, the RE estimator becomes the FE estimator. Thus, FE and RE have similar properties for time-series panels.

Two-way panels: the idea

In a two-way panel, there are individual-specific unobserved constants (individual effects) as well as time-specific constants (*time effects*):

$$
\begin{aligned}
y_{it} &= \alpha + \beta' X_{it} + u_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \\
u_{it} &= \mu_i + \lambda_t + \nu_{it}.
\end{aligned}
$$

The model is important in economic applications but it is less ubiquitous than the one-way model. Some software routines have not implemented the two-way RE model, for example.

A compact form for the two-way FE regression

Assume all effects are constants (incidental parameters). Again, the model can be written compactly as

$$y = \alpha + \mathbf{X}\beta + \mathbf{Z}_\mu \mu + \mathbf{Z}_\lambda \lambda + \nu,$$

with the $(TN \times T)$–matrix $\mathbf{Z}_\lambda$ for time dummies, and $\lambda' = (\lambda_1, \ldots, \lambda_T)$. The model assumes the identifying restriction $\sum_{i=1}^N \mu_i = \sum_{t=1}^T \lambda_t = 0$. This may be achieved by just using $N-1$ individual dummies, $T-1$ time dummies, and an 'inhomogeneous' intercept.

Note that $\mathbf{Z}_\lambda = \iota_N \otimes \mathbf{I}_T$.

Two-way LSDV estimation

By construction, OLS regression on X and all time and individual dummies yields the BLUE for the model. Direct regression requires the inversion of a $(K+N+T-1)\times(K+N+T-1)$–matrix, which is inconveniently large. Thus, application of the Frisch-Waugh theorem may be advisable.

The two-way purging matrix

Simple algebra shows that residuals of the first-step regression on all dummies is equivalent to applying the sweep-out (or purging) matrix

$$\mathbf{Q}_2 = \mathbf{I}_N \otimes \mathbf{I}_T - \mathbf{I}_N \otimes \bar{\mathbf{J}}_T - \bar{\mathbf{J}}_N \otimes \mathbf{I}_T + \bar{\mathbf{J}}_N \otimes \bar{\mathbf{J}}_T.$$

The third matrix has N^2 blocks of diagonal matrices with elements N^{-1} on their diagonals; the fourth matrix is filled with the constant value $(NT)^{-1}$.

The second matrix subtracts individual time average, the third matrix subtracts time-point averages across individuals, which subtracts means twice, so the last matrix adds a total average.

The GLS-type form of the two-way FE estimator

Again, the FE estimator has the form

$$\hat{\beta}_{FE,2-way} = \left(\mathbf{X}'\mathbf{Q_2}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Q}_2 y,$$

with the just defined two-way $\mathbf{Q}_2$. Its variance is

$$\mathrm{var}\hat{\beta} = \sigma_\nu^2\left(\mathbf{X}'\mathbf{Q_2}\mathbf{X}\right)^{-1},$$

which can be estimated by plugging in some variance estimate $\hat{\sigma}_\nu^2$ based on the LSDV residuals.

Some properties of the two-way LSDV estimator

- The estimator $\hat{\beta}_{FE,2-way}$ is BLUE for non-stochastic or exogenous X;
- $\hat{\beta}_{FE,2-way}$ is consistent for $T \to \infty$ and for $N \to \infty$;
- $\hat{\mu}$ is consistent for $T \to \infty$;
- $\hat{\lambda}$ is consistent for $N \to \infty$.

The two-way random-effects model

The basic two-way RE model (or variance-components model) reads:

$$\begin{aligned} y_{it} &= \alpha + \beta' X_{it} + u_{it}, \\ u_{it} &= \mu_i + \lambda_t + \nu_{it}, \\ \mu_i &\sim i.i.d.\left(0, \sigma_\mu^2\right), \\ \lambda_t &\sim i.i.d.\left(0, \sigma_\lambda^2\right), \\ \nu_{it} &\sim i.i.d.\left(0, \sigma_\nu^2\right), \end{aligned}$$

for $i = 1, \ldots, N$, and $t = 1, \ldots, T$. μ, λ_t, and ν are mutually independent.

Estimation in the two-way random-effects model

This is a GLS regression model, with error variance matrix

$$\begin{aligned}\mathbf{\Omega} &= \mathrm{E}\left(uu'\right) \\ &= \sigma_\mu^2\left(\mathbf{I}_N \otimes \mathbf{J}_T\right) + \sigma_\lambda^2\left(\mathbf{J}_N \otimes \mathbf{I}_T\right) + \sigma_\nu^2 \mathbf{I}_{NT}\end{aligned}$$

Thus, β is estimated consistently though inefficiently by OLS and (linear) efficiently by GLS.

The GLS estimator for the two-way model

Calculation of the estimator

$$\hat{\beta}_{RE,2-way} = \{\mathbf{X}'_{\#}\mathbf{\Omega}^{-1}\mathbf{X}_{\#}\}^{-1}\mathbf{X}'_{\#}\mathbf{\Omega}^{-1}y$$

requires the inversion of an $NT \times NT$–matrix, which is inconvenient. Unfortunately, the spectral matrix decomposition is much more involved than for the one-way model.

The decomposition of the two-way $\mathbf{\Omega}$

Some matrix algebra yields

$$\mathbf{\Omega}^{-1} = a_1 \left(\mathbf{I}_N \otimes \mathbf{J}_T \right) + a_2 \left(\mathbf{J}_N \otimes \mathbf{I}_T \right) + a_3 \mathbf{I}_{NT} + a_4 \mathbf{J}_{NT}$$

with the coefficients

$$\begin{aligned}
a_1 &= -\frac{\sigma_\mu^2}{\left(\sigma_\nu^2 + T\sigma_\mu^2\right)\sigma_\nu^2}, \\
a_2 &= -\frac{\sigma_\lambda^2}{\left(\sigma_\nu^2 + N\sigma_\lambda^2\right)\sigma_\nu^2}, \\
a_3 &= \frac{1}{\sigma_\nu^2}, \\
a_4 &= \frac{\sigma_\mu^2 \sigma_\lambda^2}{\sigma_\nu^2 \left(\sigma_\nu^2 + T\sigma_\mu^2\right)\left(\sigma_\nu^2 + N\sigma_\lambda^2\right)} \frac{2\sigma_\nu^2 + T\sigma_\mu^2 + N\sigma_\lambda^2}{\sigma_\nu^2 + T\sigma_\mu^2 + N\sigma_\lambda^2}.
\end{aligned}$$

Properties of the two-way GLS estimator

The GLS estimator is linear efficient with variance matrix

$$\{\mathbf{X}'_{\#}\mathbf{\Omega}^{-1}\mathbf{X}_{\#}\}^{-1}.$$

For $T \to \infty$, $N \to \infty$, $N/T \to c \neq 0$, o.c.s. that $\mathbf{\Omega}^{-1} \to \mathbf{Q}_2$, and FE and RE become equivalent.

Implementation of feasible GLS in the two-way RE model

The square roots of the weights can be used to transform all variables y and X in a first step, then inhomogeneous OLS regression of transformed y on transformed X yields the GLS estimate.

The variance parameters σ_ν^2, σ_μ^2, σ_λ^2 are unknown, and so are the a_j terms and the $\sqrt{a_j}$ for the transformation. If the variance parameters are estimated from a first-stage LSDV, the resulting estimator can be shown to be asymptotically efficient. First-stage OLS yields similar results but entails some slight asymptotic inefficiency.