

PANEL DATA

RANDOM AND FIXED EFFECTS MODEL

Professor Menelaos Karanasos

December 2011

PANEL DATA

Notation

y_{it} is the value of the dependent variable for cross-section unit i at time t where $i = 1, \dots, n$ and $t = 1, \dots, T$

X_{it}^j is the value of the j th explanatory variable for unit i at time t where $j = 1, \dots, K$. That is, there are K explanatory variables
Thus the upperscript j denotes an explanatory variable.

BALANCED PANELS

We will restrict our discussion to estimation with balanced panels.

That is we have the same number of observations on each cross-section unit

so that the total number of observations is nT

When $n = 1$ and T is large we have the familiar time-series data case

Likewise, when $T = 1$ and n is large we have cross-section data

Panel data estimation refers to cases where $n > 1$ and $T > 1$

POOLED ESTIMATOR

The most common way of organizing the data is by decision units. Thus, let

$$\underbrace{\mathbf{y}_i}_{T \times 1} = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix}, \quad \underbrace{\boldsymbol{\varepsilon}_i}_{T \times 1} = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix},$$

and

$$\underbrace{\mathbf{X}_i}_{T \times K} = \begin{bmatrix} X_{i1}^1 & X_{i1}^2 & \cdots & X_{i1}^K \\ X_{i2}^1 & X_{i2}^2 & \cdots & X_{i2}^K \\ \vdots & \vdots & \ddots & \vdots \\ X_{iT}^1 & X_{iT}^2 & \cdots & X_{iT}^K \end{bmatrix},$$

That is \mathbf{X}_i is a $T \times K$ matrix

Often the data are stacked to form

$$\underbrace{\mathbf{y}}_{nT \times 1} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{bmatrix}, \quad \underbrace{\mathbf{X}}_{nT \times K} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{bmatrix}, \quad \underbrace{\boldsymbol{\varepsilon}}_{nT \times 1} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_n \end{bmatrix},$$

where \mathbf{y} and $\boldsymbol{\varepsilon}$ are $nT \times 1$ vectors, and \mathbf{X} is an $nT \times K$ matrix

MATRIX FORM

The standard model can be expressed as

$$\underbrace{\mathbf{y}}_{nT \times 1} = \underbrace{\mathbf{X}}_{nT \times K} \underbrace{\mathbf{b}}_{K \times 1} + \underbrace{\boldsymbol{\varepsilon}}_{nT \times 1} \quad (1)$$

where

$$\underbrace{\mathbf{b}}_{K \times 1} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_K \end{bmatrix}$$

The models we will discuss next are all variants of the standard linear model given by equation (1)

The simplest estimation method is to assume that $\varepsilon_{it} \sim iid(0, \sigma^2)$ for all i and t

Estimation of this model is straightforward using OLS

By assuming each observation is *iid*, however, we have essentially ignored the panel structure of the data

TWO EXTENSIONS

Our starting point is the following model:

$$y_{it} = \underbrace{\mathbf{x}'_{it}}_{1 \times K} \underbrace{\mathbf{b}}_{K \times 1} + \varepsilon_{it},$$

That is, \mathbf{x}_{it} is the t th row of the \mathbf{X}_i matrix or the it th row of the \mathbf{X} matrix
We go one step further and specify the following error structure for the disturbance term

$$\varepsilon_{it} = a_i + \eta_{it}, \quad (2)$$

where we assume that η_{it} is uncorrelated with \mathbf{x}_{it}

The first term of the decomposition, a_i , is called the individual effect

ERROR STRUCTURE

In this formulation our ignorance has two parts:

The first part varies across individuals or the cross-section unit but is constant across time

This part may or may not be correlated with the explanatory variables

The second part varies unsystematically (i.e., independently) across time and individuals

RANDOM AND FIXED EFFECTS

This formulation is the simplest way of capturing the notion that two observations from the same individual(unit) will be more like each other than observations from two different individuals

A large proportion of empirical applications involve one of the following assumptions about the individual effect:

1. Random effects model: a_i is uncorrelated with \mathbf{x}_{it}
2. Fixed effects model: a_i is correlated with \mathbf{x}_{it}

THE RANDOM EFFECTS MODEL

It is important to stress that the substantive assumption that distinguishes this model from the fixed effects model is that the time-invariant person-specific effect a_i is uncorrelated with \mathbf{x}_{it} .

This orthogonality condition, along with our assumption about η_{it} , is sufficient for OLS to be asymptotically unbiased

However, there are two problems:

1. OLS will produce consistent estimates of \mathbf{b} but the standard errors will be understated.
2. OLS is not efficient compared to a feasible GLS procedure

THE ERROR

It will be helpful to be a bit more explicit about the precise nature of the error.

Recall that

$$\underbrace{\boldsymbol{\varepsilon}_i}_{T \times 1} = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix}, \quad \underbrace{\boldsymbol{\varepsilon}}_{nT \times 1} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_n \end{bmatrix},$$

and

$$\varepsilon_{it} = a_i + \eta_{it},$$

Thus, we can write

$$\underbrace{\boldsymbol{\eta}_i}_{T \times 1} = \begin{bmatrix} \eta_{i1} \\ \eta_{i2} \\ \vdots \\ \eta_{iT} \end{bmatrix}, \quad \underbrace{\boldsymbol{\eta}}_{nT \times 1} = \begin{bmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \vdots \\ \boldsymbol{\eta}_n \end{bmatrix}.$$

Moreover, we have

$$\begin{aligned} E(\boldsymbol{\eta}_i) &= 0, & E(\boldsymbol{\eta}_i \boldsymbol{\eta}_i') &= \sigma_\eta^2 \mathbf{1}_T, \\ E(\boldsymbol{\eta}) &= 0, & E(\boldsymbol{\eta} \boldsymbol{\eta}') &= \sigma_\eta^2 \mathbf{1}_{nT}. \end{aligned}$$

THE \mathbf{a}_i

Further,

$$\underbrace{\mathbf{a}_i}_{T \times 1} = \begin{bmatrix} a_i \\ a_i \\ \vdots \\ a_i \end{bmatrix} = a_i \mathbf{i}, \quad \underbrace{\mathbf{a}}_{nT \times 1} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n \end{bmatrix},$$

where \mathbf{i} is a $T \times 1$ vector of ones, and

$$\begin{aligned} E(a_i) &= 0, & E(a_i a_j) &= 0, \text{ for } i \neq j, \\ E(a_i^2) &= \sigma_a^2, & E(a_i \eta_{jt}) &= 0. \end{aligned}$$

THE Σ MATRIX

Given these assumptions we can write the error covariance matrix of the disturbance term of each individual cross-section unit:

$$\begin{aligned} \underbrace{\Sigma}_{T \times T} &= \underbrace{E(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i')}_{T \times T} = E(\mathbf{a}_i \mathbf{a}_i') + E(\boldsymbol{\eta}_i \boldsymbol{\eta}_i') & (3) \\ &= \underbrace{\sigma_a^2 \mathbf{1} \mathbf{1}'}_{T \times T} + \sigma_\eta^2 \mathbf{I}_T = \\ &= \sigma_a^2 \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix} + \sigma_\eta^2 \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_a^2 + \sigma_\eta^2 & \sigma_a^2 & \cdots & \sigma_a^2 \\ \sigma_a^2 & \sigma_a^2 + \sigma_\eta^2 & \cdots & \sigma_a^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_a^2 & \sigma_a^2 & \cdots & \sigma_a^2 + \sigma_\eta^2 \end{bmatrix}. \end{aligned}$$

THE Ω MATRIX

When the data are organized as in equation (1) the covariance of the error term for all the observations in the stacked model can be written as

$$\underbrace{\Omega}_{nT \times nT} = \mathbf{I}_n \otimes \underbrace{\Sigma}_{T \times T} = \mathbf{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \begin{bmatrix} \Sigma & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \Sigma & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & & \Sigma \end{bmatrix}.$$

The block diagonality of Ω makes finding an inverse simpler and we can focus on finding the inverse of Σ . It is straightforward but tedious to show that

$$\Sigma^{-1/2} = \frac{1}{\sigma_\eta} \left[\mathbf{I}_T - \frac{(1-\theta)}{T} \underbrace{\mathbf{ii}'}_{T \times T} \right],$$

$$\text{where } \theta = \sqrt{\frac{\sigma_\eta^2}{T\sigma_a^2 + \sigma_\eta^2}} \quad (4)$$

is the unknown quantity that must be estimated.

Feasible GLS estimators require that we get estimates of the variances σ_a^2 and σ_η^2 in θ .

RANDOM EFFECTS AS A COMBINATION OF WITHIN AND BETWEEN ESTIMATORS

We consider two estimators that are consistent but not efficient relative to GLS

The first one is quite intuitive: convert all the data into specific averages and perform OLS on this 'collapsed' data set.

In particular,

$$\underbrace{\mathbf{y}_i}_{T \times 1} = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix} \Rightarrow \underbrace{\bar{y}_i}_{1 \times 1} = \sum_{t=1}^T y_{it} / T,$$

so that we have one observation for each section i . Notice that

$$\bar{\mathbf{y}}_i = \bar{y}_i \times \mathbf{i} = \begin{bmatrix} \bar{y}_i \\ \bar{y}_i \\ \vdots \\ \bar{y}_i \end{bmatrix}.$$

Moreover,

$$\underbrace{\mathbf{X}_i}_{T \times K} = \begin{bmatrix} X_{i1}^1 & X_{i1}^2 & \cdots & X_{i1}^K \\ X_{i2}^1 & X_{i2}^2 & \cdots & X_{i2}^K \\ \vdots & \vdots & \ddots & \vdots \\ X_{iT}^1 & X_{iT}^2 & \cdots & X_{iT}^K \end{bmatrix} \Rightarrow$$
$$\underbrace{\bar{\mathbf{x}}_i'}_{1 \times K} = \left[\bar{X}_i^1 \quad \bar{X}_i^2 \quad \cdots \quad \bar{X}_i^K \right],$$

where $\bar{X}_i^j = \sum_{t=1}^T X_{it}^j / T$. Recall that we also denote the t th row of the \mathbf{X}_i matrix by $\mathbf{x}'_{it} = [X_{it}^1 \quad X_{it}^2 \quad \cdots \quad X_{it}^K]$. Specifically we perform OLS on the following equation

$$\underbrace{\bar{y}_i}_{1 \times 1} = \underbrace{\bar{\mathbf{x}}_i'}_{1 \times K} \underbrace{\mathbf{b}}_{K \times 1} + \text{error}, \quad i = 1, \dots, n \quad (5)$$

Thus, the between-group estimator can be expressed as

$$\beta_B = \left[\sum_{i=1}^n (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})' \right]^{-1} \sum_{i=1}^n (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{y}_i - \bar{y}),$$

where $\bar{\mathbf{x}} = \frac{1}{nT} \sum_{i=1}^n \sum_{j=1}^T \mathbf{x}_{it}$ and $\bar{y} = \frac{1}{nT} \sum_{i=1}^n \sum_{j=1}^T y_{it}$.

The estimator β_B is called the between-group estimator because it ignores variation within the group.

DUMMY VARIABLES

To put this expression in matrix terms, stack the data as before and define a new $nT \times n$ matrix \mathbf{D} , which is merely the matrix of n dummy variables corresponding to each cross-section unit.

That is

$$\underbrace{\mathbf{D}}_{nT \times n} = \left[\underbrace{\mathbf{D}_1}_{nT \times 1} \quad \underbrace{\mathbf{D}_2}_{nT \times 1} \quad \cdots \quad \underbrace{\mathbf{D}_n}_{nT \times 1} \right],$$

where $\mathbf{D}_i = 1$ for observation $(i-1)T + 1, \dots, iT$, and zero otherwise

Next define $\underbrace{\mathbf{P}_D}_{nT \times nT} = \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$ a symmetric and idempotent matrix

that is: $\mathbf{P}'_D = \mathbf{P}_D$ and $\mathbf{P}^2_D = \mathbf{P}_D$

Premultiplying by this matrix transforms the data into the means described in equation (5)

PREDICTED VALUES (MEANS)

This is because the predicted value of $\underbrace{\mathbf{y}}_{nT \times 1}$ from a regression on nothing

but the individual dummies is merely $\mathbf{P}_D \mathbf{y} = \begin{bmatrix} \bar{y}_1 \times \mathbf{i} \\ \vdots \\ \bar{y}_n \times \mathbf{i} \end{bmatrix} = \hat{\mathbf{y}}$

where \mathbf{i} is the $T \times 1$ unit vector.

(See Matrices Appendix C)

In other words, we run the following OLS regression:

$$\mathbf{P}_D \mathbf{y} = \mathbf{P}_D \mathbf{X} \mathbf{b} + \text{error}$$

The β estimated this way is called the between estimator and is given by

$$\begin{aligned} \beta_B &= (\mathbf{X}' \mathbf{P}'_D \mathbf{P}_D \mathbf{X})^{-1} \mathbf{X}' \mathbf{P}'_D \mathbf{P}_D \mathbf{y} \\ &= (\mathbf{X}' \mathbf{P}_D^2 \mathbf{X})^{-1} \mathbf{X}' \mathbf{P}_D^2 \mathbf{y} \\ &= (\mathbf{X}' \mathbf{P}_D \mathbf{X})^{-1} \mathbf{X}' \mathbf{P}_D \mathbf{y} \end{aligned}$$

since $\mathbf{P}'_D = \mathbf{P}_D$ and $\mathbf{P}_D^2 = \mathbf{P}_D$.

2SLS ESTIMATION

The between estimator is consistent (though not efficient) when OLS on the pooled sample is consistent.

This estimator corresponds to a 2SLS, using the person dummies as instruments.

That is regress \mathbf{X} on \mathbf{D} and get the predicted values:

$$\underbrace{\widehat{\mathbf{X}}}_{nT \times K} = \underbrace{\mathbf{D}}_{nT \times n} \underbrace{\boldsymbol{\gamma}}_{n \times K} = \mathbf{D} \underbrace{(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{X}}_{\boldsymbol{\gamma}} = \mathbf{P}_D \mathbf{X}.$$

(see Matrices Appendix C).

Then regress \mathbf{y} on $\widehat{\mathbf{X}}$:

$$\begin{aligned}\beta_B &= (\widehat{\mathbf{X}}'\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}'\mathbf{y} \\ &= [(\mathbf{P}_D\mathbf{X})'\mathbf{P}_D\mathbf{X}]^{-1}(\mathbf{P}_D\mathbf{X})'\mathbf{y} \\ &= (\mathbf{X}'\mathbf{P}'_D\mathbf{P}_D\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}'_D\mathbf{y} \\ &= (\mathbf{X}'\mathbf{P}_D\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_D\mathbf{y}.\end{aligned}$$

WITHIN ESTIMATOR

We can also use the information 'thrown away' by the between estimator.

Define $\mathbf{M}_D = \mathbf{I}_{nT} - \overbrace{\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'}^{\mathbf{P}_D}$, which is also a symmetric idempotent matrix.

If we premultiply the data by \mathbf{M}_D and compute OLS on the transformed data we can derive the following within estimator

$$\begin{aligned}\beta_W &= [(\mathbf{M}_D\mathbf{X})'_D\mathbf{M}\mathbf{X}]^{-1}(\mathbf{M}_D\mathbf{X})'\mathbf{M}_D\mathbf{y} \\ &= (\mathbf{X}'\mathbf{M}'_D\mathbf{M}_D\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}'_D\mathbf{M}_D\mathbf{y} \\ &= (\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}_D\mathbf{y}.\end{aligned}\tag{6}$$

A. The matrix \mathbf{M}_D can be interpreted as a residual-maker matrix.

Premultiplying by this matrix transforms the data into residuals from auxiliary regressions of all the variables on a complete set of individual specific constants

(see also Matrices Appendix B):

RESIDUALS (DEVIATIONS FROM PERSON-SPECIFIC MEANS)

i) We regress \mathbf{y} and \mathbf{X} on \mathbf{D} and we save the residuals $\tilde{\mathbf{y}} = \mathbf{M}_D \mathbf{y}$ and $\tilde{\mathbf{X}} = \mathbf{M}_D \mathbf{X}$:

$$\begin{aligned}\tilde{\mathbf{y}} &= \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{D}\boldsymbol{\gamma} \\ &= \mathbf{y} - \underbrace{\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{y}}_{\boldsymbol{\gamma}} \\ &= [\mathbf{I}_{nT} - \underbrace{\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'}_{\mathbf{P}_D}]\mathbf{y} \\ &= \mathbf{M}_D \mathbf{y}.\end{aligned}$$

ii) We regress $\tilde{\mathbf{y}} = \mathbf{M}_D \mathbf{y}$ on $\tilde{\mathbf{X}} = \mathbf{M}_D \mathbf{X}$.

Since the predicted value from regressing \mathbf{y} on \mathbf{D} , that is $\hat{\mathbf{y}} = \mathbf{D}\boldsymbol{\gamma} = \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{y}$, is merely the individual specific mean (see the analysis above), the residuals $\tilde{\mathbf{y}} = \mathbf{y} - \hat{\mathbf{y}}$ are merely deviations from person-specific means.

Similarly, the residuals

$\mathbf{X} - \widehat{\mathbf{X}} = \mathbf{X} - \mathbf{P}_D \mathbf{X} = [\mathbf{I}_{nT} - \overbrace{\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'}^{\mathbf{P}_D}] \mathbf{X} = \widetilde{\mathbf{X}} = \mathbf{M}_D \mathbf{X}$ are merely deviations from person-specific means.

B. Specifically, equation (6) is equivalent to performing OLS on the following equation:

$$y_{it} - \bar{y}_i = \underbrace{(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'}_{1 \times K} \underbrace{\mathbf{b}}_{K \times 1} + \text{error}, \quad i = 1, \dots, N, \quad t = 1, \dots, T.$$

If the assumptions underlying the random effects model are correct, the within estimator is also a consistent estimator, but it is not efficient. This defect is clear since we have included n unnecessary extra variables.

Since

$$\bar{y}_i = \bar{\mathbf{x}}_i' \mathbf{b} + a_i,$$

we can estimate a_i as

$$\alpha_i = \bar{y}_i - \bar{\mathbf{x}}_i' \beta.$$

C. The within estimator can also be expressed as

$$\beta_W = \left[\sum_{i=1}^n \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \right]^{-1} \sum_{i=1}^n \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(y_{it} - \bar{y}_i).$$

Its variance-covariance matrix is given by

$$V(\beta_W) = \sigma_\eta^2 \left[\sum_{i=1}^n \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \right]^{-1}.$$

D. The within estimator is merely an estimator that would result from running OLS on the data including a set of dummy variables

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{D}\mathbf{a} + \boldsymbol{\eta},$$

where $\mathbf{a}' = [a_1 \ a_2 \ \cdots \ a_n]$. Then

$$\boldsymbol{\beta}_W = (\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}_D\mathbf{y}.$$

(see Matrices Appendix B).

It is called the within estimator because it uses only the variation within each cross-section unit.

Its variance covariance matrix is given by

$$V(\boldsymbol{\beta}_W) = \sigma_{\eta}^2(\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}$$

E. Alternatively, we can write

$$\underbrace{\mathbf{y}_i}_{T \times 1} = \underbrace{\mathbf{X}_i}_{T \times K} \mathbf{b} + a_i \underbrace{\mathbf{i}}_{T \times 1} + \boldsymbol{\eta}_i, \quad i = 1, \dots, n$$

We premultiply the above i th equation by a $T \times T$ idempotent transformation matrix

$$\mathbf{Q} = \mathbf{I}_T - \frac{1}{T} \mathbf{i} \mathbf{i}',$$

to 'swipe out' the individual effect a_i so that individual observations are measured as deviations from individual means (over time):

$$\begin{aligned} \mathbf{Qy}_i &= \mathbf{QX}_i \mathbf{b} + \mathbf{Q} \mathbf{i} a_i + \mathbf{Q} \boldsymbol{\eta}_i \\ &= \mathbf{QX}_i \mathbf{b} + \mathbf{Q} \boldsymbol{\eta}_i, \end{aligned}$$

Applying the OLS procedure to the above equation, we have

$$\beta_W = \left[\sum_{i=1}^N \mathbf{x}'_i \mathbf{Q} \mathbf{x}_i \right]^{-1} \sum_{i=1}^N \mathbf{x}_i \mathbf{Q} \mathbf{y}_i.$$

Its variance-covariance matrix is given by

$$V(\beta_W) = \sigma_{\eta}^2 \left[\sum_{i=1}^N \mathbf{x}'_i \mathbf{Q} \mathbf{x}_i \right]^{-1}.$$

Notice that the pooled OLS estimate is just a weighted sum of the between and within estimators:

$$\beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

Since $\mathbf{M}_D = \mathbf{I}_{nT} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}' = \mathbf{I}_{nT} - \mathbf{P}_D \Rightarrow \mathbf{I}_{nT} = \mathbf{M}_D + \mathbf{P}_D$, we have

$$\begin{aligned}\beta &= (\mathbf{X}'\mathbf{X})^{-1} \left\{ \mathbf{X}' \overbrace{(\mathbf{M}_D + \mathbf{P}_D)\mathbf{y}}^{\mathbf{I}_{nT}} \right\} \\ &= (\mathbf{X}'\mathbf{X})^{-1} \{ \mathbf{X}'\mathbf{M}_D\mathbf{y} + \mathbf{X}'\mathbf{P}_D\mathbf{y} \}.\end{aligned}$$

Next premultiplying $\mathbf{X}'\mathbf{M}_D\mathbf{y}$ by $\mathbf{I}_K = (\mathbf{X}'\mathbf{M}_D\mathbf{X})(\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}$ and $\mathbf{X}'\mathbf{P}_D\mathbf{y}$ by $\mathbf{I}_K = (\mathbf{X}'\mathbf{P}_D\mathbf{X})(\mathbf{X}'\mathbf{P}_D\mathbf{X})^{-1}$ gives

$$\begin{aligned}\beta &= (\mathbf{X}'\mathbf{X})^{-1}\left\{(\mathbf{X}'\mathbf{M}_D\mathbf{X})\overbrace{(\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}_D\mathbf{y}}^{\beta_W}\right. \\ &\quad \left. + \underbrace{(\mathbf{X}'\mathbf{P}_D\mathbf{X})(\mathbf{X}'\mathbf{P}_D\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_D\mathbf{y}}_{\beta_B}\right\} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\{(\mathbf{X}'\mathbf{M}_D\mathbf{X})\beta_W + (\mathbf{X}'\mathbf{P}_D\mathbf{X})\beta_B\}.\end{aligned}$$

Notice that the sum of the two weights is equal to \mathbf{I}_K :

$$\begin{aligned}& (\mathbf{X}'\mathbf{X})^{-1}\{(\mathbf{X}'\mathbf{M}_D\mathbf{X}) + (\mathbf{X}'\mathbf{P}_D\mathbf{X})\} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\{\mathbf{X}'(\mathbf{M}_D + \mathbf{P}_D)\mathbf{X}\} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{I}_K.\end{aligned}$$

In the random effects case, OLS on the pooled data fails to use information about the heteroscedasticity that results from using repeated observations of the same cross-section units.

The problem with the pooled estimator is that it weights all observations equally.

This treatment is not generally optimal because an additional observation on a person already in the data set is unlikely to add as much information as an additional observation from a new (independent) individual.

Next we have to compute the necessary quantities for a feasible GLS. Standard ANOVA suggests the following estimators

$$\hat{\sigma}_\eta^2 = \frac{\hat{\boldsymbol{\varepsilon}}_W' \hat{\boldsymbol{\varepsilon}}_W}{nT - n - K},$$

$$\hat{\sigma}_B^2 = \frac{\hat{\boldsymbol{\varepsilon}}_B' \hat{\boldsymbol{\varepsilon}}_B}{n - K},$$

$$\hat{\sigma}_a^2 = \hat{\sigma}_B^2 - \frac{\hat{\sigma}_\eta^2}{T},$$

where $\hat{\boldsymbol{\varepsilon}}_W$ are the residuals from the within regression and $\hat{\boldsymbol{\varepsilon}}_B$ are the residuals from the between regression. These can then be used to construct $\hat{\boldsymbol{\theta}}$.

These estimators are asymptotically unbiased estimates of the relevant variances.

THE RANDOM EFFECT ESTIMATOR

A simple procedure to compute the random effect estimator is as follows:

1. Compute the between and within estimators
2. Use the residuals to calculate the appropriate variance terms
3. Calculate $\hat{\theta}$
4. Run OLS on the following transformed variables \mathbf{y}^* and \mathbf{X}^*

$$y_{it}^* = y_{it} - (1 - \hat{\theta})\bar{y}_i, \quad (7)$$

$$\mathbf{x}_{it}^* = \mathbf{x}_{it} - (1 - \hat{\theta})\bar{\mathbf{x}}_i. \quad (8)$$

The transformation is intuitively appealing.

When there is no uncorrelated person-specific component of variance ($\sigma_a^2 = 0$), $\theta = 1$, and the random effects estimator reduces to the pooled OLS estimator.

When $\hat{\theta} = 0$ the random effects estimator reduces to the within estimator.

To get efficient estimates of the parameters we have to use the GLS method

$$\beta_{GLS} = \left[\sum_{i=1}^N \mathbf{x}'_i \Sigma^{-1} \mathbf{x}_i \right]^{-1} \sum_{i=1}^N \mathbf{x}_i \Sigma^{-1} \mathbf{y}_i.$$

where Σ is given by (3).

We can also obtain the following

$$\begin{aligned}\beta_{GLS} &= \left[\frac{1}{T} \sum_{i=1}^N \mathbf{x}'_i \mathbf{Q} \mathbf{x}_i + \theta^2 \sum_{i=1}^n (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})' \right]^{-1} \times \\ &\quad \left[\frac{1}{T} \sum_{i=1}^N \mathbf{x}'_i \mathbf{Q} \mathbf{y}_i + \theta^2 \sum_{i=1}^n (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{y}_i - \bar{y})' \right] \\ &= \Delta \beta_B + (\mathbf{I}_K - \Delta) \beta_W,\end{aligned}\tag{9}$$

where θ^2 has been defined in (4) and

$$\Delta = \theta^2 T \left[\sum_{i=1}^N \mathbf{x}'_i \mathbf{Q} \mathbf{x}_i + \theta^2 T \sum_{i=1}^n (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})' \right]^{-1} \times \left[\sum_{i=1}^n (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})' \right].$$

The GLS estimator (9) is a weighted average of the between-group and the within-group estimators.

If $\theta \rightarrow 0$, then $\Delta \rightarrow \mathbf{0}$, and the GLS becomes the within estimator.

If $\theta \rightarrow 1$, then the GLS becomes the pooled OLS estimator.

Moreover, the variance covariance matrix of β_{GLS} is

$$V(\beta_{GLS}) = \sigma_{\eta}^2 \left[\sum_{i=1}^N \mathbf{x}'_i \mathbf{Q} \mathbf{x}_i + T \theta^2 \sum_{i=1}^n (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})' \right]^{-1}.$$

THE FIXED EFFECTS MODEL

The fixed effects model starts with the presumption that $Cov(X_{it}, a_i) \neq 0$. Thus we must estimate the model conditionally on the presence of the fixed effects.

That is, we must rewrite the model as

$$y_{it} = \mathbf{x}_{it}\mathbf{b} + a_i + \eta_i, \quad i = 1, \dots, n \quad (10)$$

where the a_i are treated as unknown parameters to be estimated. In the typical case, T is small and n is large.

Although we can not estimate a_i consistently, we can estimate the remaining parameters consistently.

To do so we need only run the regression

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{D}\mathbf{a} + \boldsymbol{\eta} \quad (11)$$

where $\mathbf{a}' = [a_1 \ a_2 \ \cdots \ a_n]$, $\mathbf{D} = \mathbf{I}_n \otimes \mathbf{i}_T$ as before, is a set of n dummy variables.

The above equation is just the same as running a regression of each of our variables y and X on this set of dummies and then running the regression of the y residuals on the X residuals.

The matrix that produces such residuals is the familiar

$$\mathbf{M}_D = \mathbf{I}_{nT} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}' = \mathbf{I}_{nT} - \mathbf{P}_D.$$

We can run OLS on the transformed variables $\tilde{\mathbf{y}} = \mathbf{M}_D\mathbf{y}$ and $\tilde{\mathbf{X}} = \mathbf{M}_D\mathbf{X}$ to get

$$\beta_W = (\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}_D\mathbf{y}.$$

This is merely the within estimator we derived before. The within estimator is only one possible fixed effects estimator.

Any transformation that rids us of the fixed effect will produce a fixed effect estimator.

Deviations from means purges the data of the fixed effects by removing means of these variables across individual cross-section units.

That is, the predicted value of y , which belongs to group i , is just the mean of that group:

$$\bar{y}_i = \bar{\mathbf{x}}_i \mathbf{b} + a_i + \bar{\eta}_i \quad (12)$$

We can difference Eqs (10) and (12) to yield

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) \mathbf{b} + (\eta_i - \bar{\eta}_i)$$

In many applications, the easiest way to implement a fixed effects estimator with conventional software is to include a different dummy variable for each individual unit.

This method is called the least-squares dummy variable (LSDV) method as in Eq. (11).

If n is very large, however, it may be computationally prohibitive to compute coefficients for each cross-section unit.

In that case another way to implement a fixed effect estimator is as follows:

- Transform all the variables by subtracting person-specific means
- Run OLS on the transformed variables

Moreover, the standard errors need to be corrected. The correct standard errors are

$$V(\boldsymbol{\beta}_W) = \sigma_{\eta}^2 (\mathbf{X}' \mathbf{M}_D \mathbf{X})^{-1}.$$

This result is almost exactly the same output one would get from the two-step procedure defined earlier.

The correct way to estimate σ_{η}^2 is

$$\hat{\sigma}_{\eta}^2 = \frac{\hat{\varepsilon}'_W \hat{\varepsilon}_W}{nT - n - K},$$

The fact that the fixed effects estimator can be interpreted as a simple OLS regression of means-difference variables explains why this estimator is often called a within group estimator. That is, it uses only the variation within an individual's set of observations.

A WU-HAUSMAN TEST

The random and fixed effects estimators have different properties depending on the correlation between a_i and the regressors. Specifically,

1. If the effects are uncorrelated with the explanatory variables, the random effects estimator is consistent and efficient.

The fixed effects estimator is consistent but not efficient.

2. If the effects are correlated with the explanatory variables, the fixed effects estimator is consistent and efficient but the random effects estimator is now inconsistent.

One can perform a Hausman test, defined as

$$H = (\beta_{RE} - \beta_{FE})' [V(\beta_{RE}) - V(\beta_{FE})]^{-1} (\beta_{RE} - \beta_{FE}).$$

The Hausman test statistic will be distributed asymptotically as χ^2 with K degrees of freedom under the null hypothesis that the random effects estimator is correct.

An alternative method is to perform a simple auxiliary regression.

Let y_{it}^* and \mathbf{x}_{it}^* be the data transformed for the random effects model as in Eqs. (7) and (8). Recall that the $\tilde{\mathbf{X}}$ variables are transformed for fixed-effects regression.

The Hausman test can be computed by means of a simple F test on \mathbf{d} in the following auxiliary regression:

$$\mathbf{y}^* = \mathbf{X}^* \mathbf{b} + \tilde{\mathbf{X}} \mathbf{d} + \text{error}.$$

The hypothesis being tested is whether the omission of fixed effects in the random effects model has any effect on the consistency of the random effects estimates.

MATRICES APPENDIX A

A matrix may be partitioned into as set of submatrices by indicating subgroups of rows and/or columns. For example

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

The inverse of a partitioned matrix may also be expressed in partitioned form:

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{B}_{11} & -\mathbf{B}_{11}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{B}_{11} & \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{B}_{11}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \end{bmatrix}, \quad (13)$$

where $\mathbf{B}_{11} = (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}$ or, alternatively,

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{22}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{22} \\ -\mathbf{B}_{22}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{B}_{22} \end{bmatrix},$$

where $\mathbf{B}_{22} = (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}$.

Consider partition the matrix X as $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{bmatrix}$.

Then

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} \mathbf{X}'_1 \\ \mathbf{X}'_2 \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} \underbrace{\mathbf{X}'_1\mathbf{X}_1}_{\mathbf{A}_{11}} & \underbrace{\mathbf{X}'_1\mathbf{X}_2}_{\mathbf{A}_{12}} \\ \underbrace{\mathbf{X}'_2\mathbf{X}_1}_{\mathbf{A}_{22}} & \underbrace{\mathbf{X}'_2\mathbf{X}_2}_{\mathbf{A}_{22}} \end{bmatrix}.$$

The inverse of the above matrix is given by equation (13) where

$$\begin{aligned} \mathbf{B}_{11} &= (\mathbf{X}'_1\mathbf{X}_1 - \mathbf{X}'_1\mathbf{X}_2(\mathbf{X}'_2\mathbf{X}_2)^{-1}\mathbf{X}'_2\mathbf{X}_1)^{-1} \\ &= (\mathbf{X}'_1\mathbf{I}\mathbf{X}_1 - \mathbf{X}'_1\mathbf{X}_2(\mathbf{X}'_2\mathbf{X}_2)^{-1}\mathbf{X}'_2\mathbf{X}_1)^{-1} \\ &= (\mathbf{X}'_1\mathbf{M}_2\mathbf{X}_1)^{-1}, \end{aligned}$$

where

$$\mathbf{M}_2 = \mathbf{I} - \mathbf{X}_2(\mathbf{X}'_2\mathbf{X}_2)^{-1}\mathbf{X}'_2,$$

that is, M_2 is a symmetric idempotent matrix.

Similarly,

$$\mathbf{B}_{22} = (\mathbf{X}'_2\mathbf{M}_1\mathbf{X}_2)^{-1}, \text{ where } \mathbf{M}_1 = \mathbf{I} - \mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1.$$

Moreover, $-\mathbf{B}_{11} \underbrace{\mathbf{A}_{12}}_{\mathbf{X}'_1 \mathbf{X}_2} \underbrace{\mathbf{A}_{22}^{-1}}_{(\mathbf{X}'_2 \mathbf{X}_2)^{-1}}$ is given by

$$-\mathbf{B}_{11} \mathbf{X}'_1 \mathbf{X}_2 (\mathbf{X}'_2 \mathbf{X}_2)^{-1}.$$

In addition, $\mathbf{X}'\mathbf{y}$ in partitioned form is given by

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} \mathbf{X}'_1 \\ \mathbf{X}'_2 \end{bmatrix} \mathbf{y} = \begin{bmatrix} \mathbf{X}'_1 \mathbf{y} \\ \mathbf{X}'_2 \mathbf{y} \end{bmatrix}$$

Thus the OLS estimates from the regression of \mathbf{y} on \mathbf{X} in partitioned form are given by

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} = \begin{bmatrix} \underbrace{\mathbf{A}_{11}}_{\mathbf{X}'_1\mathbf{X}_1} & \underbrace{\mathbf{A}_{12}}_{\mathbf{X}'_1\mathbf{X}_2} \\ \underbrace{\mathbf{A}_{22}}_{\mathbf{X}'_2\mathbf{X}_2} & \underbrace{\mathbf{A}_{22}}_{\mathbf{X}'_2\mathbf{X}_2} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'_1\mathbf{y} \\ \mathbf{X}'_2\mathbf{y} \end{bmatrix}.$$

Taking the first row we have

$$\begin{aligned} \boldsymbol{\beta}_1 &= \mathbf{B}_{11}\mathbf{X}'_1\mathbf{y} - \mathbf{B}_{11} \underbrace{\mathbf{A}_{12}}_{\mathbf{x}'_1\mathbf{x}_2} \underbrace{\mathbf{A}_{22}^{-1}}_{(\mathbf{x}'_2\mathbf{x}_2)^{-1}} \mathbf{X}'_2\mathbf{y} \\ &= \mathbf{B}_{11}\mathbf{X}'_1\mathbf{y} - \mathbf{B}_{11}\mathbf{X}'_1\mathbf{X}_2(\mathbf{X}'_2\mathbf{X}_2)^{-1}\mathbf{X}'_2\mathbf{y} \\ &= \underbrace{\mathbf{B}_{11}}_{(\mathbf{X}'_1\mathbf{M}_2\mathbf{X}_1)^{-1}} \underbrace{\mathbf{X}'_1(\mathbf{I} - \mathbf{X}_2(\mathbf{X}'_2\mathbf{X}_2)^{-1}\mathbf{X}'_2)}_{\mathbf{M}_2}\mathbf{y} \\ &= (\mathbf{X}'_1\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{M}_2\mathbf{y}. \end{aligned}$$

Similarly, we can show that

$$\boldsymbol{\beta}_2 = (\mathbf{X}'_2\mathbf{M}_1\mathbf{X}_2)^{-1}\mathbf{X}'_2\mathbf{M}_1\mathbf{y}$$

Rewrite

$$\beta_1 = (\mathbf{X}'_1 \mathbf{M}_2 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \mathbf{M}_2 \mathbf{y}.$$

Regressing \mathbf{y} and \mathbf{X}_1 on \mathbf{X}_2 yields a vector of residuals, $\mathbf{M}_2 \mathbf{y}$, and a matrix of residuals, $\mathbf{M}_2 \mathbf{X}_1$.

Regressing the former on the latter gives the β_1 coefficient vector.

$$\mathbf{M}_i = \mathbf{I} - \mathbf{X}_i(\mathbf{X}'_i\mathbf{X}_i)^{-1}\mathbf{X}'_i, \quad i = 1, 2.$$

Premultiplication of any vector by M_i gives the residuals from the regression of that vector on X_i .

Consider the following regression

$$\mathbf{y} = \mathbf{X}_i\mathbf{b} + \varepsilon.$$

The estimator of b is given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'_i\mathbf{X}_i)^{-1}\mathbf{X}'_i\mathbf{y} = \mathbf{y}.$$

The predicted value of y is given by

$$\hat{\mathbf{y}} = \mathbf{X}_i\hat{\boldsymbol{\beta}} = \mathbf{X}_i(\mathbf{X}'_i\mathbf{X}_i)^{-1}\mathbf{X}'_i\mathbf{y} = \mathbf{P}_i\mathbf{y}.$$

The vector of the residuals is given by

$$\begin{aligned} \varepsilon &= \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}_i(\mathbf{X}'_i\mathbf{X}_i)^{-1}\mathbf{X}'_i\mathbf{y} \\ &= [\mathbf{I} - \mathbf{X}_i(\mathbf{X}'_i\mathbf{X}_i)^{-1}\mathbf{X}'_i]\mathbf{y} = [\mathbf{I} - \mathbf{P}_i]\mathbf{y} = \mathbf{M}_i\mathbf{y}. \end{aligned}$$

MATRICES APPENDIX C

Assume for simplicity and without loss of generality that we have only two sections. That is $n = 2$.

Define the $2T \times 2$ matrix \mathbf{D} , which is merely the matrix of 2 dummy variables corresponding to each cross-section unit.

In a partition form we have

$$\underbrace{2T \times 2}_{\mathbf{D}} = \begin{bmatrix} \underbrace{T \times 1} & \underbrace{T \times 1} \\ \mathbf{i} & \mathbf{0} \\ \underbrace{0} & \underbrace{\mathbf{i}} \\ \underbrace{T \times 1} & \underbrace{T \times 1} \end{bmatrix} \quad \text{or}$$
$$\underbrace{2 \times 2T}_{\mathbf{D}'} = \begin{bmatrix} \underbrace{1 \times T} & \underbrace{1 \times T} \\ \mathbf{i}' & \mathbf{0}' \\ \underbrace{0'} & \underbrace{\mathbf{i}'} \\ \underbrace{1 \times T} & \underbrace{1 \times T} \end{bmatrix}$$

where \mathbf{i} is a $T \times 1$ unit vector.

Then

$$\begin{aligned} \overbrace{\mathbf{D}'\mathbf{D}}^{2 \times 2} &= \begin{bmatrix} \mathbf{i}'\mathbf{i} & \mathbf{0} \\ \mathbf{0} & \mathbf{i}'\mathbf{i} \end{bmatrix} = \begin{bmatrix} T & 0 \\ 0 & T \end{bmatrix} = TI \Rightarrow \\ &= (\mathbf{D}'\mathbf{D})^{-1} = (1/T)I. \end{aligned}$$

Moreover,

$$\overbrace{\mathbf{D}\mathbf{D}'}^{2T \times 2T} = \begin{bmatrix} \overbrace{T \times T}^{\mathbf{i}\mathbf{i}'} & \overbrace{T \times T}^{\mathbf{0}} \\ \mathbf{0} & \mathbf{i}\mathbf{i}' \end{bmatrix},$$

where $\mathbf{i}\mathbf{i}'$ is a $T \times T$ matrix with unit elements.

Thus

$$\begin{aligned} \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{y} &= \frac{1}{T}\mathbf{D}\mathbf{D}'\mathbf{y} \\ &= \frac{1}{T} \begin{bmatrix} \overbrace{\mathbf{i}\mathbf{i}'}^{T \times T} & \overbrace{\mathbf{0}}^{T \times T} \\ \mathbf{0} & \mathbf{i}\mathbf{i}' \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \\ &= \frac{1}{T} \begin{bmatrix} \sum_{t=1}^T y_{1t} \times \mathbf{i} \\ \sum_{t=1}^T y_{2t} \times \mathbf{i} \end{bmatrix} = \begin{bmatrix} \bar{y}_1 \times \mathbf{i} \\ \bar{y}_2 \times \mathbf{i} \end{bmatrix} \end{aligned}$$

In general if we have n cross-sections then \mathbf{D} is an $nT \times n$ matrix

$$\underbrace{\mathbf{D}}_{nT \times n} = \begin{bmatrix} \overbrace{\mathbf{i}}^{T \times 1} & \mathbf{0} & \dots & \overbrace{\mathbf{0}}^{T \times 1} \\ \mathbf{0} & \mathbf{i} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{i} \end{bmatrix} .$$

It can be shown that

$$\underbrace{\mathbf{D}}_{nT \times n} (\mathbf{D}'\mathbf{D})^{-1} \mathbf{D}' \underbrace{\mathbf{y}}_{nT \times 1} = \begin{bmatrix} \bar{y}_1 \times \mathbf{i} \\ \bar{y}_2 \times \mathbf{i} \\ \vdots \\ \bar{y}_n \times \mathbf{i} \end{bmatrix} = \hat{\mathbf{y}} .$$

Similarly,

$$\underbrace{\mathbf{D}}_{nT \times n} (\mathbf{D}'\mathbf{D})^{-1} \mathbf{D}' \underbrace{\mathbf{X}}_{nT \times k} = \begin{bmatrix} \overline{\mathbf{X}}_1 \\ \overline{\mathbf{X}}_2 \\ \vdots \\ \overline{\mathbf{X}}_n \end{bmatrix} = \widehat{\mathbf{X}},$$

where

$$\overline{\mathbf{X}}_i = \left[\overline{X}_i^1 \times \mathbf{i} \quad \overline{X}_i^2 \times \mathbf{i} \quad \dots \quad \overline{X}_i^K \times \mathbf{i} \right],$$

with $\overline{X}_i^j = \sum_{t=1}^T X_{it}^j / T$.