

Ch. 11 Panel Data Model

(January 9, 2018)



1 Introduction

Data sets that combine time series and cross sections are common in econometrics. For example, the published statistics of the OECD contain numerous series of economic aggregate observed yearly for many countries. The Panel Study of Income Dynamics (PSID) is a study of roughly 6000 families and 15000 individuals who have been interviewed periodically from 1968 to the present. Data of this type are often referred to as *panel data*.

Estimation of relationships that combine time series and cross-sectional data is a problem frequently encountered in econometrics. Typically, one may possess several years of data on a number of firms, households, geographical areas, or biological units. The problem, when using these data to estimate a relationship, is to specify a model that will adequately allow for difference in behavior *over cross-sectional* units as well as any difference in behavior *over time* for a given cross-sectional units. Once a model has been specified, there are the additional problems of the most efficient estimation procedure and how to test hypothesis about the parameters.

For simplicity, we restrict our attention to the linear regression model

$$Y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + u_{it}, \quad i = 1, 2, \dots, N, \quad t = 1, 2, \dots, T, \quad (11-1)$$

where \mathbf{x}_{it} is a $k \times 1$ vector of observations on explanatory variables. There are assumed to be N cross-sectional units and T time periods, for a total of NT observations. If each u_{it} has expectation zero conditional on its corresponding \mathbf{x}_{it} , we can estimate (11-1) by OLS. But the OLS is not efficient if the u_{it} is not *i.i.d.*, and the *i.i.d.* assumption is rarely realistic with panel data.

If certain shocks affect the same cross-sectional unit at all point in time, the error terms u_{it} and u_{is} must be correlated (serially correlated) for all $t \neq s$. Similarly, if certain shocks affect all cross-sectional units at the same point in time, the error term u_{it} and u_{jt} must be correlated (cross-dependent) for all $i \neq j$. In consequence, if we use OLS, not only do we obtain inefficient parameter estimates, but we could possibly obtain inconsistent parameter estimate. This happen, for example, when \mathbf{x}_{it} contains lagged dependent variables and the u_{it} is serially correlated.

2 Error-Components Model

The two most popular approaches for dealing with panel data are both based on what are called *error-components models*. This idea is to specify the error term u_{it} in (11-1) as consisting of two or three separate shocks, each of which is assumed to be independent of the other. A fairly general specification is

$$u_{it} = \lambda_t + v_i + \varepsilon_{it}. \quad (11-2)$$

Here λ_t affects all observations for time period t , v_i affects all observations for cross-sectional unit i , and ε_{it} affects only observation it . It is generally assumed that λ_t are independent across t ,¹ the v_i are independent across i , and ε_{it} are independent across i and t .

In order to estimate an error-components model, λ_t and v_i can be regarded as being either **fixed** or **random**. If the λ_t and v_i are thought of as fixed effects or *are correlated with the explanatory variables \mathbf{x}_{it}* , then they are treated as parameters to be estimate. It turns out that they can then be estimated by OLS using dummy variables. If they are thought of as random effects, *as disturbances that are uncorrelated with \mathbf{x}_{it}* , then we must figure out the covariance matrix of u_{it} , and use the feasible GLS. Each of these approaches can be appropriate in some circumstances but may be inappropriate in others.

In what follows of this chapter, we simplify the error-components specification (11-2) by eliminating the λ_t . Thus we assume that there are shocks specific to each cross-sectional unit, or group, but no time-specific shocks. This assumption is often made in empirical work,² and it considerably simplifies the algebra.

2.1 Pooled, Fixed and Random Effects Models

With the assumption that there are shocks specific to each cross-sectional unit or group, the fundamental advantage of a panel data set over a cross section is that it will allow the researcher greater flexibility in model difference in behavior **across individuals**.

¹It is to be noted that although λ_t is independent across t , u_{it} is still correlated. It is because $E(u_{it}u_{jt}) \neq 0$.

²That is, the assumptions are oriented toward cross-section analysis; they are wide but typically short (relatively). Heterogeneity across units is an integral part of this analysis.

To consider whether an overall constant or a individual-specific constant in the model, we consider the following basic framework of the form

$$Y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_i\boldsymbol{\alpha} + \varepsilon_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T. \quad (11-3)$$

There are k regressors in \mathbf{x}_{it} , **not including a constant term**. The heterogeneity, or individual effect is $\mathbf{z}'_i\boldsymbol{\alpha}$ where \mathbf{z}'_i contains a constant term and a set of individual or group specific variables, which may be observed, such as race, sex, location, and so on or unobserved, such as family specific characteristics, individual heterogeneity in skill or preference, and so on, all of which are taken to be constant over time t . The various cases we will consider are:

(a). Pooled Regression:

If \mathbf{z}'_i contains only a constant term, then there is no individual specific characteristics in this model. All we need is pooling the data,

$$Y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \alpha + \varepsilon_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T. \quad (11-4)$$

and OLS provides consistent and efficient estimates of the common $\boldsymbol{\beta}$ and α .

(b). Fixed Effects:

If $\mathbf{z}'_i\boldsymbol{\alpha} = \alpha_i$, then it is the fixed effect approach to take α_i as a group-specific variable in the regression model,

$$Y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \alpha_i + \varepsilon_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T. \quad (11-5)$$

It should be noted that the term “fixed” as used here signifies the correlation of α_i and \mathbf{x}_{it} , not that α_i is non-stochastic. The α_i could be in fact be random,³ the essential thing is that they must be independent of the error term ε_{it} . They may, however, be correlated with the explanatory variables \mathbf{x}_{it} .

(c). Random Effects:

If the individual heterogeneity is an unobserved random variable and can be assumed to be uncorrelated with the included explanatory variables, then the model may be formulated as

$$\begin{aligned} Y_{it} &= \mathbf{x}'_{it}\boldsymbol{\beta} + E(\mathbf{z}'_i\boldsymbol{\alpha}) + [\mathbf{z}'_i\boldsymbol{\alpha} - E(\mathbf{z}'_i\boldsymbol{\alpha})] + \varepsilon_{it} \\ &= \mathbf{x}'_{it}\boldsymbol{\beta} + \alpha + v_i + \varepsilon_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T. \end{aligned} \quad (11-6)$$

³We can regard α_i as a explanatory variable like \mathbf{x}_{it} .

The random effects approach specifies that v_i is a group specific random element, similar to ε_{it} except that for each group, there is but a single draw that enters the regression identically in each period. \square

3 Pooled Regression

We begin the analysis by assuming the simplest version of the model, i.e., there is no heterogeneity among the groups, the pooled model,

$$Y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \alpha + \varepsilon_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T.$$

Stack the T observations for individual i in a single equation,⁴

$$\mathbf{y}_i = \mathbf{X}_i\ddot{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}_i, \quad i = 1, 2, \dots, N,$$

or⁵

$$\mathbf{y} = \mathbf{X}\ddot{\boldsymbol{\beta}} + \boldsymbol{\varepsilon},$$

where $\ddot{\boldsymbol{\beta}} = [\alpha, \boldsymbol{\beta}']'$ now includes the constant term. If it is assumed that $\boldsymbol{\varepsilon}_i$ is well behaved:

$$\begin{aligned} E(\varepsilon_{it}|\mathbf{X}_i) &= 0, \quad t = 1, 2, \dots, T. \\ E(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}'_i|\mathbf{X}_i) &= \sigma^2\mathbf{I}_T; \\ E(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}'_j|\mathbf{X}_i) &= \mathbf{0} \quad \text{if } i \neq j, \end{aligned}$$

then the most efficient estimator of $\ddot{\boldsymbol{\beta}}$ is

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

with estimated variance

$$\widehat{Var}(\hat{\ddot{\boldsymbol{\beta}}}) = s^2(\mathbf{X}'\mathbf{X})^{-1}, \quad \text{in which}$$

⁴Here,

$$\mathbf{y}_i = \begin{bmatrix} y_{i,1} \\ y_{i,2} \\ \vdots \\ y_{i,T} \end{bmatrix}, \quad \mathbf{X}_i = \begin{bmatrix} \mathbf{x}'_{i,1} \\ \mathbf{x}'_{i,2} \\ \vdots \\ \mathbf{x}'_{i,T} \end{bmatrix}.$$

⁵Here,

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_N \end{bmatrix}.$$

$$s^2 = \frac{\mathbf{e}'\mathbf{e}}{NT - k}. \text{ Here, } \mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}.$$

It is quite likely that there is correlation across observations such that $E(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i' | \mathbf{X}_i) = \boldsymbol{\Omega}_i$, then in the same spirit as the White estimator, we can estimate the appropriate asymptotic covariance matrix as

$$\widehat{Asy.Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1} \left(\sum_{i=1}^N \mathbf{X}_i' \mathbf{e}_i \mathbf{e}_i' \mathbf{X}_i \right) (\mathbf{X}'\mathbf{X})^{-1}.$$

Exercise 1.

Reproduce the results at Example 9.1 (Table 9.1) on p. 187 of Greene 6th edition. ■

4 Fixed-Effects Model

This formulation of the model assumes that differences across units can be captured in difference in the constant term. Each α_i is the unobserved time-invariant individual effect:⁶

$$Y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \alpha_i + \varepsilon_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T.$$

Unlike the random effects model where the unobserved α_i is independent of \mathbf{x}_{it} for all $t = 1, \dots, T$, the fixed effects (FE) model allows α_i to be correlated with the regressor matrix \mathbf{x}_{it} . Strict exogeneity with respect to the idiosyncratic error term ε_{it} , however, is still required.

Let \mathbf{y}_i and \mathbf{X}_i be the T observations the i th unit, \mathbf{i} be a $T \times 1$ column of ones, and let $\boldsymbol{\varepsilon}_i$ be associated $T \times 1$ vector of disturbances. Then the matrix form of (11-4) for an individual i can be written as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{i}\alpha_i + \boldsymbol{\varepsilon}_i, \quad i = 1, 2, \dots, N. \quad (11-7)$$

It is also assumed that the disturbance terms are well behaved, that is

$$E(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}'_i|\mathbf{X}_i) = \sigma^2\mathbf{I}_T; \quad (11-8)$$

$$E(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}'_j|\mathbf{X}_i) = \mathbf{0} \quad \text{if } i \neq j,$$

$$E(\varepsilon_{it}|\mathbf{X}_i, \alpha_i) = 0, \quad t = 1, 2, \dots, T. \quad (11-9)$$

4.1 Fixed Effects Transformation (Within Transformation)

4.1.1 Estimation of $\boldsymbol{\beta}$

Collect all the cross-sectional unit in (11-7) to be matrix form we have

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_N \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{i} & \mathbf{0} & \cdot & \cdot & \cdot & \mathbf{0} \\ \mathbf{0} & \mathbf{i} & \mathbf{0} & \cdot & \cdot & \mathbf{0} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \mathbf{0} & \cdot & \cdot & \cdot & \mathbf{0} & \mathbf{i} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \alpha_N \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{\varepsilon}_N \end{bmatrix},$$

⁶Unlike \mathbf{x}_{it} , α_i cannot be directly observed.

or in more compact form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\alpha} + \boldsymbol{\varepsilon},$$

where \mathbf{y} and $\boldsymbol{\varepsilon}$ are $NT \times 1$, \mathbf{X} is $NT \times k$, $\boldsymbol{\beta}$ is $k \times 1$, $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_N]'$ and

$$\mathbf{D} = \begin{bmatrix} \mathbf{i} & \mathbf{0} & . & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{i} & \mathbf{0} & . & . & \mathbf{0} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \mathbf{0} & . & . & . & \mathbf{0} & \mathbf{i} \end{bmatrix} = [\mathbf{d}_1 \ \mathbf{d}_2 \ \dots \mathbf{d}_N]$$

is $NT \times N$ with \mathbf{d}_i is a dummy variables indicating the i th unit. This model is usually also referred to as the *least squares dummy variable (LSDV) model*.

Since this model satisfies the ideal conditions, OLS estimator is BLUE. By using the familiar partitioned regression of Ch. 6, the slope estimator would be⁷

$$\hat{\boldsymbol{\beta}}_{FE} = (\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}_D\mathbf{y}, \quad (11-10)$$

where

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

Lemma. (Group-Demeaned Matrix)

$$\mathbf{M}_D = \mathbf{I}_{NT} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}' = \begin{bmatrix} \mathbf{M}_i & \mathbf{0} & . & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_i & \mathbf{0} & . & . & \mathbf{0} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \mathbf{0} & . & . & . & \mathbf{0} & \mathbf{M}_i \end{bmatrix},$$

where $\mathbf{M}_i = \mathbf{I}_T - \mathbf{i}(\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}' = \mathbf{I}_T - 1/T(\mathbf{ii}')$ is the demean-matrix.

⁷Since α_i is not observable, it cannot be directly controlled for. The FE model eliminates α_i by demeaning the variables using the within transformation.

Proof.

By definition,

$$\begin{aligned}
 \mathbf{D}'\mathbf{D} &= \begin{bmatrix} \mathbf{i}' & \mathbf{0} & . & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{i}' & \mathbf{0} & . & . & \mathbf{0} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \mathbf{0} & . & . & . & \mathbf{0} & \mathbf{i}' \end{bmatrix} \begin{bmatrix} \mathbf{i} & \mathbf{0} & . & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{i} & \mathbf{0} & . & . & \mathbf{0} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \mathbf{0} & . & . & . & \mathbf{0} & \mathbf{i} \end{bmatrix} \\
 &= \begin{bmatrix} \mathbf{i}'\mathbf{i} & 0 & . & . & . & 0 \\ 0 & \mathbf{i}'\mathbf{i} & 0 & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & . & . & . & 0 & \mathbf{i}'\mathbf{i} \end{bmatrix} = \begin{bmatrix} T & 0 & . & . & . & 0 \\ 0 & T & 0 & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & . & . & . & 0 & T \end{bmatrix}_{N \times N},
 \end{aligned}$$

and therefore

$$\begin{aligned}
 \mathbf{M}_D &= \mathbf{I}_{NT} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}' \\
 &= \begin{bmatrix} \mathbf{I}_T & \mathbf{0} & . & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_T & \mathbf{0} & . & . & \mathbf{0} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \mathbf{0} & . & . & . & \mathbf{0} & \mathbf{I}_T \end{bmatrix} - \begin{bmatrix} \mathbf{i} & \mathbf{0} & . & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{i} & \mathbf{0} & . & . & \mathbf{0} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \mathbf{0} & . & . & . & \mathbf{0} & \mathbf{i} \end{bmatrix} \\
 &\quad \times \begin{bmatrix} T & 0 & . & . & . & 0 \\ 0 & T & 0 & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & . & . & . & 0 & T \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{i}' & \mathbf{0} & . & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{i}' & \mathbf{0} & . & . & \mathbf{0} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \mathbf{0} & . & . & . & \mathbf{0} & \mathbf{i}' \end{bmatrix}
 \end{aligned}$$

$$\begin{aligned}
&= \begin{bmatrix} \mathbf{I}_T - \frac{1}{T}\mathbf{i}\mathbf{i}' & \mathbf{0} & . & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_T - \frac{1}{T}\mathbf{i}\mathbf{i}' & \mathbf{0} & . & . & \mathbf{0} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \mathbf{0} & . & . & . & \mathbf{0} & \mathbf{I}_T - \frac{1}{T}\mathbf{i}\mathbf{i}' \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{M}_i & \mathbf{0} & . & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_i & \mathbf{0} & . & . & \mathbf{0} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \mathbf{0} & . & . & . & \mathbf{0} & \mathbf{M}_i \end{bmatrix}.
\end{aligned}$$

It is easy to see that the matrix \mathbf{M}_D is idempotent and that

$$\mathbf{M}_D \mathbf{y} = \begin{bmatrix} \mathbf{M}_i & \mathbf{0} & . & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_i & \mathbf{0} & . & . & \mathbf{0} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \mathbf{0} & . & . & . & \mathbf{0} & \mathbf{M}_i \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ . \\ . \\ . \\ \mathbf{y}_N \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1 - \bar{Y}_1 \mathbf{i} \\ \mathbf{y}_2 - \bar{Y}_2 \mathbf{i} \\ . \\ . \\ . \\ \mathbf{y}_N - \bar{Y}_N \mathbf{i} \end{bmatrix}$$

and

$$\mathbf{M}_D \mathbf{X} = \begin{bmatrix} \mathbf{M}_i & \mathbf{0} & . & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_i & \mathbf{0} & . & . & \mathbf{0} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \mathbf{0} & . & . & . & \mathbf{0} & \mathbf{M}_i \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ . \\ . \\ . \\ \mathbf{X}_N \end{bmatrix} = \begin{bmatrix} \mathbf{M}_i \mathbf{X}_1 \\ \mathbf{M}_i \mathbf{X}_2 \\ . \\ . \\ . \\ \mathbf{M}_i \mathbf{X}_N \end{bmatrix},$$

where the scalar $\bar{Y}_i = 1/T \sum_{t=1}^T Y_{it}$, $i = 1, 2, \dots, N$, and let $\mathbf{X}_i = [\mathbf{x}_{i1} \ \mathbf{x}_{i2} \ \dots \ \mathbf{x}_{ik}]$,⁸ then

⁸That is,

$$\mathbf{X}_i = \begin{bmatrix} \mathbf{x}'_{i1} \\ \mathbf{x}'_{i2} \\ . \\ . \\ . \\ \mathbf{x}'_{iT} \end{bmatrix} = [\mathbf{x}_{i1} \ \mathbf{x}_{i2} \ \dots \ \mathbf{x}_{ik}] = \begin{bmatrix} X_{i11} & X_{i12} & . & . & . & X_{i1k} \\ X_{i21} & X_{i22} & . & . & . & X_{i2k} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ X_{iT1} & X_{iT2} & . & . & . & X_{iT k} \end{bmatrix}.$$

$\mathbf{M}_i \mathbf{X}_i = [\mathbf{M}_i \mathbf{x}_{i1} \ \mathbf{M}_i \mathbf{x}_{i2} \ \dots \ \mathbf{M}_i \mathbf{x}_{ik}]$. Therefore $\mathbf{M}_i \mathbf{x}_{ij} = \mathbf{x}_{ij} - \bar{X}_{ij} \mathbf{i}$, $j = 1, 2, \dots, k$ with $\bar{X}_{ij} = 1/T \sum_{t=1}^T X_{itj}$. Denote $\bar{\mathbf{x}}_i = [\bar{X}_{i1} \ \bar{X}_{i2} \ \dots \ \bar{X}_{ik}]'$, the least squares regression estimation of fixed effects slope estimator $\hat{\beta}_{FE}$ in (11-10), i.e. $\mathbf{M}_D \mathbf{y}$ on $\mathbf{M}_D \mathbf{X}$ is equivalent to regression of $[Y_{it} - \bar{Y}_i]$ on $[\mathbf{x}_{it} - \bar{\mathbf{x}}_i]$.⁹

The idea for estimating β is to transform the equations to eliminate the fixed effect α_i . The above fixed effects transformation are called *within transformation*. To see this, the FE transformation is obtained by first averaging (11-5) over $t = 1, \dots, T$ to get the cross-sectional equation

$$\bar{Y}_i = \bar{\mathbf{x}}_i' \beta + \alpha_i + \bar{\varepsilon}_i, \quad (11-11)$$

where $\bar{\mathbf{x}}_i = T^{-1} \sum_{t=1}^T \mathbf{x}_{it}$ and $\bar{\varepsilon}_i = T^{-1} \sum_{t=1}^T \varepsilon_{it}$. Subtracting (11-11) from (11-5) for each t gives the FE transformation equation,

$$Y_{it} - \bar{Y}_i = (\mathbf{x}_{it}' - \bar{\mathbf{x}}_i') \beta + \varepsilon_{it} - \bar{\varepsilon}_i. \quad (11-12)$$

The time demeaning of the original equation has removed the individual specific effect α_i .

Exercise 2.

Show that $(\varepsilon_{it} - \bar{\varepsilon}_i)$ in Eq. (11-12) satisfies the ideal conditions. ■

4.1.2 Estimation of α

Denote the LSDV residuals to be

$$\mathbf{e} = \mathbf{y} - \mathbf{X} \hat{\beta}_{FE} - \mathbf{D} \hat{\alpha}, \quad (11-13)$$

⁹To see this, recall the notation that

$$\mathbf{M}_i \mathbf{X}_i = \begin{bmatrix} X_{i11} - \bar{X}_{i1} & X_{i12} - \bar{X}_{i2} & \cdot & \cdot & \cdot & X_{i1k} - \bar{X}_{ik} \\ X_{i21} - \bar{X}_{i1} & X_{i22} - \bar{X}_{i2} & \cdot & \cdot & \cdot & X_{i2k} - \bar{X}_{ik} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ X_{iT1} - \bar{X}_{i1} & X_{iT2} - \bar{X}_{i2} & \cdot & \cdot & \cdot & X_{iT k} - \bar{X}_{ik} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{i1}' - \bar{\mathbf{x}}_i' \\ \mathbf{x}_{i2}' - \bar{\mathbf{x}}_i' \\ \cdot \\ \cdot \\ \mathbf{x}_{iT}' - \bar{\mathbf{x}}_i' \end{bmatrix}.$$

then the dummy variables coefficient estimators can be recovered from

$$\mathbf{D}'\mathbf{y} = \mathbf{D}'\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{D}'\mathbf{D}\hat{\boldsymbol{\alpha}} + \mathbf{D}'\mathbf{e},$$

or

$$\hat{\boldsymbol{\alpha}} = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{FE}), \quad (11-14)$$

because $\mathbf{D}'\mathbf{e} = \mathbf{0}$. This implies that

$$\begin{aligned} \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{\alpha}_N \end{bmatrix} &= \frac{1}{T} \begin{bmatrix} \mathbf{i}' & \mathbf{0} & \cdot & \cdot & \cdot & \mathbf{0} \\ \mathbf{0} & \mathbf{i}' & \mathbf{0} & \cdot & \cdot & \mathbf{0} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \mathbf{0} & \cdot & \cdot & \cdot & \mathbf{0} & \mathbf{i}' \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 - \mathbf{X}_1\hat{\boldsymbol{\beta}}_{FE} \\ \mathbf{y}_2 - \mathbf{X}_2\hat{\boldsymbol{\beta}}_{FE} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{y}_N - \mathbf{X}_N\hat{\boldsymbol{\beta}}_{FE} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{T} \sum_{t=1}^T y_{1t} \\ \frac{1}{T} \sum_{t=1}^T y_{2t} \\ \cdot \\ \cdot \\ \cdot \\ \frac{1}{T} \sum_{t=1}^T y_{Nt} \end{bmatrix} - \begin{bmatrix} \frac{1}{T} \sum_{t=1}^T [X_{1t1} & X_{1t2} & \cdots & X_{1tk}] \\ \frac{1}{T} \sum_{t=1}^T [X_{2t1} & X_{2t2} & \cdots & X_{2tk}] \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \frac{1}{T} \sum_{t=1}^T [X_{Nt1} & X_{Nt2} & \cdots & X_{Ntk}] \end{bmatrix} \hat{\boldsymbol{\beta}}_{FE} \\ &= \begin{bmatrix} \bar{Y}_1 - \bar{\mathbf{x}}_1'\hat{\boldsymbol{\beta}}_{FE} \\ \bar{Y}_2 - \bar{\mathbf{x}}_2'\hat{\boldsymbol{\beta}}_{FE} \\ \cdot \\ \cdot \\ \cdot \\ \bar{Y}_N - \bar{\mathbf{x}}_N'\hat{\boldsymbol{\beta}}_{FE} \end{bmatrix}. \end{aligned}$$

4.1.3 Variance of the Fixed Effects Estimators

The variance of the Fixed-Effect estimators $\hat{\boldsymbol{\beta}}_{FE}$ and $\hat{\boldsymbol{\alpha}}$ are in the following.

Theorem. ($\text{Var}(\hat{\beta}_{FE})$)

Let the fixed effects model be partitioned as

$$\mathbf{y} = \mathbf{X}\hat{\beta}_{FE} + \mathbf{D}\hat{\alpha} + \mathbf{e},$$

the variance of $\hat{\beta}_{FE}$ is $\text{Var}(\hat{\beta}_{FE}) = \sigma^2(\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}$.

Proof.

$$\begin{aligned}\hat{\beta} &= (\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}_D\mathbf{y} \\ &= \beta + (\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}_D\epsilon,\end{aligned}$$

therefore,

$$\begin{aligned}\text{Var}(\hat{\beta}) &= E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] \\ &= E[(\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}_D\epsilon][(\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}_D\epsilon'] \\ &= E[(\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}_D\epsilon\epsilon'\mathbf{M}_D\mathbf{X}(\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}] \\ &= \sigma^2[(\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}_D\mathbf{I}_{NT}\mathbf{M}_D\mathbf{X}(\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}] \\ &= \sigma^2[(\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}_D\mathbf{X}(\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}] \\ &= \sigma^2(\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}.\end{aligned}\tag{11-15}$$

$$\tag{11-16}$$

■

With the above results, the appropriate estimator of $\text{Var}(\hat{\beta})$ is therefore

$$\text{Est}(\text{Var}(\hat{\beta})) = s^2(\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1},$$

where the disturbance variance estimator is s^2

$$s^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\beta} - \mathbf{D}\hat{\alpha})'(\mathbf{y} - \mathbf{X}\hat{\beta} - \mathbf{D}\hat{\alpha})}{NT - N - k} = \frac{\sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \mathbf{x}'_{it}\hat{\beta} - \hat{\alpha}_i)^2}{NT - N - k}.\tag{11-17}$$

Theorem. ($\text{Var}(\hat{\alpha})$)

The variance of individual effect estimator is

$$\text{Var}(\hat{\alpha}_i) = \frac{\sigma^2}{T} + \bar{\mathbf{x}}_i' \text{Var}(\hat{\beta}) \bar{\mathbf{x}}_i.$$

Proof.

$$\begin{aligned}
\hat{\alpha}_i &= \bar{y}_i - \bar{\mathbf{x}}_i' \hat{\boldsymbol{\beta}} \\
&= \frac{1}{T} \sum_{t=1}^T (\alpha_i + \mathbf{x}_{it}' \boldsymbol{\beta} + \varepsilon_{it}) - \bar{\mathbf{x}}_i' \hat{\boldsymbol{\beta}}.
\end{aligned}$$

The expected value of $\hat{\alpha}_i$ therefore is

$$\begin{aligned}
E(\hat{\alpha}_i) &= \frac{1}{T} \sum_{t=1}^T (\alpha_i + \mathbf{x}_{it}' \boldsymbol{\beta}) - \bar{\mathbf{x}}_i' \boldsymbol{\beta} \\
&= \alpha_i.
\end{aligned}$$

The variance of $\hat{\alpha}_i$ therefore is

$$\begin{aligned}
Var(\hat{\alpha}_i) &= E(\alpha_i - \hat{\alpha}_i)^2 \\
&= E \left[\frac{1}{T} \sum_{t=1}^T \varepsilon_{it} - \bar{\mathbf{x}}_i' (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \right]^2 \\
&= \frac{1}{T^2} E \left(\sum_{t=1}^T \varepsilon_{it} \right)^2 + E(\bar{\mathbf{x}}_i' (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}))^2 \\
&= \frac{\sigma^2}{T} + E[\bar{\mathbf{x}}_i' (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \bar{\mathbf{x}}_i] \\
&= \frac{\sigma^2}{T} + \bar{\mathbf{x}}_i' Var(\hat{\boldsymbol{\beta}}) \bar{\mathbf{x}}_i.
\end{aligned}$$

■

4.1.4 Testing the Significance of the Group Effects

Consider the null hypothesis that $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_N = \alpha$. Under this null hypothesis, the efficient estimator is the pooled least squares. The F ration used for the test would be

$$F_{N-1, NT-N-k} = \frac{(R_{LSDV}^2 - R_{Pooled}^2)/(N-1)}{(1 - R_{LSDV}^2)/(NT - N - k)}, \quad (11-18)$$

where R_{LSDV}^2 indicates the R^2 from the dummy variables model and R_{Pooled}^2 indicates the R^2 from the pooled or restricted model with only a single overall constant.

Example.

See Example 9.4 (p.199) of Greene 6th edition. ■

4.1.5 Nonspherical Disturbance and Robust Variance Matrix Estimator

Without assumption (11-8), express (11-16) gives an improper variance matrix estimator. While heteroscedasticity in ε_{it} is always a potential problem, the more substantive problem is cross-observation correlation, or autocorrelation. In a longitudinal data set, the group of observations may all pertain to the same individual, so any latent effects left out of the model will carry across all periods. Suppose, we assume that the disturbance is (instead of (11-8))

$$E(\varepsilon_i \varepsilon_i') = \Omega_i, \quad (11-19)$$

following (11-15), then the robust variance matrix estimator of $\hat{\beta}_{FE}$ is

$$Asy.Var(\hat{\beta}) = (\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1} \left(\sum_{i=1}^N \ddot{\mathbf{X}}_i' \mathbf{e}_i \mathbf{e}_i' \ddot{\mathbf{X}}_i \right) (\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}, \quad (11-20)$$

where $\mathbf{M}_D\mathbf{X} = \ddot{\mathbf{X}} = [\ddot{\mathbf{X}}_1' \ddot{\mathbf{X}}_2' \dots, \ddot{\mathbf{X}}_N']'$ and \mathbf{e}_i is the vector of T residuals for the i th individual (see (11-13)). This is suggested by Arellano (1987) and follows the general results of White (1984). The robust variance matrix estimator is valid in the presence of any heteroscedasticity and serial correlation in ε_{it} , provided that T is small relative to N .

4.2 First Differencing Transformation

We can also use differencing to eliminate the unobserved effect α_i . Lagging the model (11-5) one period and subtracting gives

$$\Delta Y_{it} = \Delta \mathbf{x}_{it}' \boldsymbol{\beta} + \Delta \varepsilon_{it}, \quad i = 1, 2, \dots, N; \quad t = 2, \dots, T, \quad (11-21)$$

where $\Delta Y_{it} = Y_{it} - Y_{i,t-1}$, $\Delta \mathbf{x}_{it} = \mathbf{x}_{it} - \mathbf{x}_{i,t-1}$ and $\Delta \varepsilon_{it} = \varepsilon_{it} - \varepsilon_{i,t-1}$. As with the fixed effects transformation, this *first-differencing transformation* eliminates the unobserved effect α_i . In differencing we lose the first time period for each cross section.

The first-difference (FD) estimator, $\hat{\beta}_{FD}$, the pooled OLS estimator from the regression

$$\Delta Y_{it} \text{ on } \Delta \mathbf{x}_{it}, \quad t = 2, \dots, T; \quad i = 1, 2, \dots, N. \quad (11-22)$$

Denote $v_{it} = \Delta \varepsilon_{it}$. When we assume that

$$E(\mathbf{v}_i \mathbf{v}_i') = \sigma_v^2 \mathbf{I}_{T-1}, \quad (11-23)$$

then the variance of $\hat{\beta}_{FD}$ can be estimated by

$$\widehat{Var}(\hat{\beta}_{FD}) = \hat{\sigma}_v^2 (\Delta \mathbf{X}' \Delta \mathbf{X})^{-1},$$

where $\hat{\sigma}_v^2$ is a consistent estimator of σ_v^2 . The simplest estimator is obtained by computing the OLS residuals

$$\hat{v}_{it} = \Delta Y_{it} - \Delta \mathbf{x}_{it}' \hat{\beta}_{FD}$$

from the pooled regression (11-22).

A consistent estimate of σ_v^2 is

$$\hat{\sigma}_v^2 = \frac{\sum_{i=1}^N \sum_{t=2}^T \hat{v}_{it}^2}{N(T-1) - k}.$$

If the assumption (11-23) is violated, then, as usual, we can compute a robust variance matrix which is

$$asy \widehat{Var}(\hat{\beta}_{FD}) = (\Delta \mathbf{X}' \Delta \mathbf{X})^{-1} \left(\sum_{i=1}^N \Delta \mathbf{x}_i' \mathbf{v}_i \mathbf{v}_i' \Delta \mathbf{x}_i \right) (\Delta \mathbf{X}' \Delta \mathbf{X})^{-1},$$

where $\Delta \mathbf{X}$ denote the $N(T-1) \times k$ matrix of stacked first difference of \mathbf{x}_{it} .

4.3 The Within and Between Groups Estimators

We could formulate a pooled regression model in three ways.

(a). First, the original formulation is

$$Y_{it} = \alpha + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T. \quad (11-24)$$

(b). In term of deviations from the group means,

$$Y_{it} - \bar{Y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\boldsymbol{\beta} + \varepsilon_{it} - \bar{\varepsilon}_i, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T, \quad (11-25)$$

(c.) and in terms of the group means,

$$\bar{Y}_i = \alpha + \bar{\mathbf{x}}_i'\boldsymbol{\beta} + \bar{\varepsilon}_i, \quad i = 1, 2, \dots, N. \quad (11-26)$$

□

To estimate $\boldsymbol{\beta}$ by pooled OLS in (11-24) (partial out the effect of α) we would use the total sum of squares and cross products,

$$\mathbf{S}_{xx}^{total} = \sum_{i=1}^N \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}})(\mathbf{x}_{it} - \bar{\mathbf{x}})'$$

and

$$\mathbf{S}_{xy}^{total} = \sum_{i=1}^N \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}})(Y_{it} - \bar{Y}),$$

where $\bar{\mathbf{x}} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \mathbf{x}_{it}$ and $\bar{Y} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T Y_{it}$.

In (11-25), the moments matrices we use are **within-group's** (i.e., deviations from the group means) sums of squares and cross products,

$$\mathbf{S}_{xx}^{Within} = \sum_{i=1}^N \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \quad (11-27)$$

and

$$\mathbf{S}_{xy}^{Within} = \sum_{i=1}^N \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(Y_{it} - \bar{Y}_i), \quad (11-28)$$

Finally, for (11-26), the means of group mean are the overall mean (i.e., $1/N \sum_{i=1}^N \bar{Y}_i = \bar{\bar{Y}}$). Therefore the moment matrices are the **between-groups** sums of squares and cross products,

$$\mathbf{S}_{xx}^{Between} = \sum_{i=1}^N T(\bar{\mathbf{x}}_i - \bar{\bar{\mathbf{x}}})(\bar{\mathbf{x}}_i - \bar{\bar{\mathbf{x}}})'$$

and

$$\mathbf{S}_{xy}^{Between} = \sum_{i=1}^N T(\bar{\mathbf{x}}_i - \bar{\bar{\mathbf{x}}})(\bar{Y}_i - \bar{\bar{Y}}).$$

It is easy to verify that

$$\mathbf{S}_{xx}^{Total} = \mathbf{S}_{xx}^{Within} + \mathbf{S}_{xx}^{Between},$$

and

$$\mathbf{S}_{xy}^{Total} = \mathbf{S}_{xy}^{Within} + \mathbf{S}_{xy}^{Between}.$$

Therefore, there are three possible least squares estimator of β corresponding to theses decompositions.

(a). The least squares estimator in the pooling regression (11-24) is

$$\begin{aligned} \hat{\beta}^{Total} &= (\mathbf{S}_{xx}^{Total})^{-1} \mathbf{S}_{xy}^{Total} \\ &= (\mathbf{S}_{xx}^{Within} + \mathbf{S}_{xx}^{Between})^{-1} (\mathbf{S}_{xy}^{Within} + \mathbf{S}_{xy}^{Between}). \end{aligned}$$

(b). The **within-groups** estimator from (11-25) is

$$\begin{aligned} \hat{\beta}^{Within} &= (\mathbf{S}_{xx}^{Within})^{-1} \mathbf{S}_{xy}^{Within} \\ &= \left(\sum_{i=1}^N \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \right)^{-1} \left(\sum_{i=1}^N \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(Y_{it} - \bar{Y}_i) \right). \end{aligned}$$

This is the LSDV estimator computed earlier.

(c). An alternative estimator would be the **between-groups** estimator from (11-26)

$$\hat{\beta}^{Between} = (\mathbf{S}_{xx}^{Between})^{-1} \mathbf{S}_{xy}^{Between}.$$

This is the least squares estimator based on the N sets of group means. \square

From the preceding expression we know that

$$\mathbf{S}_{xy}^{Within} = \mathbf{S}_{xx}^{Within} \cdot \hat{\beta}^{Within}$$

and

$$\mathbf{S}_{xy}^{Between} = \mathbf{S}_{xx}^{Between} \cdot \hat{\beta}^{Between},$$

hence

$$\begin{aligned} \hat{\beta}^{Total} &= (\mathbf{S}_{xx}^{Within} + \mathbf{S}_{xx}^{Between})^{-1} (\mathbf{S}_{xx}^{Within} \hat{\beta}^{Within} + \mathbf{S}_{xx}^{Between} \hat{\beta}^{Between}) \\ &= (\mathbf{S}_{xx}^{Within} + \mathbf{S}_{xx}^{Between})^{-1} \cdot \mathbf{S}_{xx}^{Within} \hat{\beta}^{Within} \\ &\quad + (\mathbf{S}_{xx}^{Within} + \mathbf{S}_{xx}^{Between})^{-1} \mathbf{S}_{xx}^{Between} \hat{\beta}^{Between} \\ &= \mathbf{F}^{Within} \hat{\beta}^{Within} + \mathbf{F}^{Between} \hat{\beta}^{Between}, \end{aligned}$$

where

$$\begin{aligned} \mathbf{F}^{Within} &= (\mathbf{S}_{xx}^{Within} + \mathbf{S}_{xx}^{Between})^{-1} \mathbf{S}_{xx}^{Within} \text{ and,} \\ \mathbf{F}^{Within} + \mathbf{F}^{Between} &= (\mathbf{S}_{xx}^{Within} + \mathbf{S}_{xx}^{Between})^{-1} (\mathbf{S}_{xx}^{Within} + \mathbf{S}_{xx}^{Between}) \\ &= \mathbf{I}. \end{aligned}$$

That is the pooling OLS estimator is a matrix weighted average of the within- and between-groups estimator.

4.4 Fixed Effects with Unbalanced Panels

In our treatment of panel data models we have assumed that a balanced panel is available—each cross section unit has the same time periods available. Often, some time periods are missing for some units in the population of interest, and we are left with an *unbalanced panel*. The preceding analysis assumed equal group sizes and relied

on the assumption at several points. A modification to allow unequal group sizes is quite simple. First, the full sample size is $\sum_{i=1}^N T_i$ instead of NT , which calls for minor modifications in the computations of s^2 , $Var(\hat{\beta})$ and the F statistics. Second, group means must be based on T_i , which varies across group.

The regressors moment matrix shown in (11-27),

$$\mathbf{S}_{xx}^{Within} = \sum_{i=1}^N \sum_{t=1}^{T_i} (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'$$

now is

$$\mathbf{S}_{xx,unbalance}^{Within} = \sum_{i=1}^N \left(\sum_{t=1}^{T_i} (\mathbf{x}_{it} - \bar{\mathbf{x}}_{i,un})(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i,un})' \right), \quad (11-29)$$

where

$$\bar{\mathbf{x}}_{i,un} = \frac{1}{T_i} \sum_{t=1}^{T_i} \mathbf{x}_{it}.$$

The other moment $\mathbf{S}_{xy,unbalance}^{Within}$ and $\mathbf{S}_{yy,unbalance}^{Within}$ are computed likewise. No other changes are necessary for the one factor LSDV estimators.

5 Random Effects

The fixed effects model allows the unobserved individual effects to be correlated with the included variables. We then modeled the differences between units strictly as parametric shifts of the regression function. This model might be viewed as applying only to the cross-sectional units in the study, not to additional ones outside the sample. For example, an intercountry comparison may well include the full set of countries for which it is reasonable to assume that the model is constant.

If the individual effects are strictly uncorrelated with the regressors, then it might be appropriate to model the individual specific constant terms as randomly distributed across cross-sectional units. This view would be appropriate if we believed that sampled cross-sectional units were drawn from a large population. The payoff to this form is that it greatly reduces the number of parameters to be estimated. The cost is the possibility of inconsistent estimates, should the assumption turn out to be inappropriate.

Consider the model

$$Y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \alpha + v_i + \varepsilon_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T. \quad (11-30)$$

where there are k regressors including a constant and now the single constant term α is the mean of the unobserved heterogeneity, $E(\mathbf{z}'_i\boldsymbol{\alpha})$ in (11-6). The component v_i is the random heterogeneity specific to the i th observation and is constant through time.¹⁰ We assume further for $t = 1, 2, \dots, T$,

$$\begin{aligned} E(\varepsilon_{it}|\mathbf{X}_i) &= E(v_i|\mathbf{X}_i) = 0; \\ E(\varepsilon_{it}^2|\mathbf{X}_i) &= \sigma_\varepsilon^2; \\ E(v_i^2|\mathbf{X}_i) &= \sigma_v^2; \\ E(\varepsilon_{it}v_j|\mathbf{X}_i) &= 0 \text{ for all } i, t, \text{ and } j; \\ E(\varepsilon_{it}\varepsilon_{js}|\mathbf{X}_i) &= 0 \text{ if } t \neq s \text{ or } i \neq j; \\ E(v_iv_j|\mathbf{X}_i) &= 0 \text{ if } i \neq j. \end{aligned}$$

Denote

$$u_{it} = \varepsilon_{it} + v_i,$$

¹⁰In this stage, we assume that $E(v_i^2) = \sigma_v^2$ below which mean that v_i capture all the random heterogeneity specific to the i th observation, though they have **same** variance.

and let \mathbf{y}_i and \mathbf{X}_i (including the constant term) be the T observations of the i th unit, and let

$$\mathbf{u}_i = [u_{i1}, u_{i2}, \dots, u_{iT}]',$$

then collecting the T observations of i th unit we have

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{u}_i, \quad i = 1, 2, \dots, N,$$

and the variance of the disturbances would be

$$\begin{aligned} \boldsymbol{\Sigma} = E(\mathbf{u}_i \mathbf{u}_i') &= E \begin{bmatrix} \varepsilon_{i1} + v_i \\ \varepsilon_{i2} + v_i \\ \cdot \\ \cdot \\ \varepsilon_{iT} + v_i \end{bmatrix} \begin{bmatrix} \varepsilon_{i1} + v_i & \varepsilon_{i2} + v_i & \cdot & \cdot & \cdot & \varepsilon_{iT} + v_i \end{bmatrix} \\ &= \begin{bmatrix} \sigma_\varepsilon^2 + \sigma_v^2 & \sigma_v^2 & \sigma_v^2 & \cdot & \cdot & \cdot & \sigma_v^2 \\ \sigma_v^2 & \sigma_\varepsilon^2 + \sigma_v^2 & \sigma_v^2 & \cdot & \cdot & \cdot & \sigma_v^2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_v^2 & \sigma_v^2 & \sigma_v^2 & \cdot & \cdot & \cdot & \sigma_\varepsilon^2 + \sigma_v^2 \end{bmatrix} = \sigma_\varepsilon^2 \mathbf{I}_T + \sigma_v^2 \mathbf{i}_T \mathbf{i}_T', \end{aligned}$$

where \mathbf{i}_T is a $T \times 1$ column of ones.

Observations of all the cross-sectional unit can be collected and be rewritten as

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{y}_N \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{X}_N \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{u}_N \end{bmatrix},$$

or in more compact form

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{u},$$

where \mathbf{y} and \mathbf{u} are $NT \times 1$, \mathbf{X} is $NT \times k$, $\boldsymbol{\beta}$ is $k \times 1$ and the variance-covariances of

the stacked model are

$$\mathbf{\Omega} = E(\mathbf{uu}') = \begin{bmatrix} \mathbf{\Sigma} & \mathbf{0} & \cdot & \cdot & \cdot & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma} & \mathbf{0} & \cdot & \cdot & \mathbf{0} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \mathbf{0} & \cdot & \cdot & \cdot & \mathbf{0} & \mathbf{\Sigma} \end{bmatrix} = \mathbf{I}_N \otimes \mathbf{\Sigma}. \quad (11-31)$$

5.1 Generalized Least Squares Estimation

The generalized least squares estimator of the coefficients would be

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y} = \left(\sum_{i=1}^N \mathbf{X}_i' \mathbf{\Sigma}^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^N \mathbf{X}_i' \mathbf{\Sigma}^{-1} \mathbf{y}_i \right), \quad (11-32)$$

if $\mathbf{\Omega}$ are known. As with many generalized least squares problem it is convenient to find a transform matrix

$$\mathbf{\Omega}^{-1/2} = [\mathbf{I}_N \otimes \mathbf{\Sigma}]^{-1/2}$$

so that the OLS can be applied to the transformed model.

Fuller and Battese (1973) show that in this case

$$\mathbf{\Sigma}^{-1/2} = \frac{1}{\sigma_\varepsilon} \left[\mathbf{I}_T - \frac{\theta}{T} \mathbf{i}_T \mathbf{i}_T' \right], \quad (11-33)$$

where

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

The transformation of \mathbf{y}_i and \mathbf{X}_i for GLS is therefore

$$\mathbf{\Sigma}^{-1/2} \mathbf{y}_i = \frac{1}{\sigma_\varepsilon} \begin{bmatrix} Y_{i1} - \theta \bar{Y}_i \\ Y_{i2} - \theta \bar{Y}_i \\ \cdot \\ \cdot \\ \cdot \\ Y_{iT} - \theta \bar{Y}_i \end{bmatrix} \quad (11-34)$$

and likewise for the rows of \mathbf{X}_i (including constant term). Eq. (11-34) implies that the random-effect GLS estimator is identical to the pooled OLS (11-4) estimator when $\theta = 0$, which happens when $\sigma_v^2 = 0$,¹¹ and equals to the within-groups, or fixed-effects, estimator when $\theta = 1$, which happens when $\sigma_\varepsilon^2 = 0$. (One would interpret θ as the effect that would remain if $\sigma_\varepsilon^2 = 0$, because the only effect then would be v_i . In this case, the fixed and the random effects model would be indistinguishable, so this results make sense.)

5.2 FGLS When Σ is Unknown

If the variance component $(\sigma_\varepsilon^2, \sigma_v^2)$ are known, generalized least squares estimation (11-32) can be computed as shown earlier. Of course, this is unlikely, so as usual, we must first estimate the disturbance variance and then use an FGLS procedure. A heuristic approach to estimate the variance components is as follows.¹²

5.2.1 Estimation of σ_ε^2

The random effects model is

$$Y_{it} = \alpha + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it} + v_i, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T,$$

and in term of group means,

$$\bar{Y}_i = \alpha + \bar{\mathbf{x}}'_i\boldsymbol{\beta} + \bar{\varepsilon}_i + v_i, \quad i = 1, 2, \dots, N.$$

Therefore, taking deviation from the group means removes the heterogeneity:

$$Y_{it} - \bar{Y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\boldsymbol{\beta} + \varepsilon_{it} - \bar{\varepsilon}_i, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T, \quad (11-35)$$

¹¹In that case there is no individual's specific random effect

¹²Because random effects model is a FGLS, all that we need are consistent estimators of σ_ε^2 and σ_v^2 , there are alternative consistent estimators, see for example, Hsiao (1986, Section 3.3) and Wooldridge (2002, section 10.4, p.260).

Because¹³

$$E \left(\sum_{t=1}^T (\varepsilon_{it} - \bar{\varepsilon}_i)^2 \right) = (T-1)\sigma_\varepsilon^2,$$

if β is observable (and therefore ε is observed), then an unbiased and consistent estimator of σ_ε^2 based on T observations in group i would be (from laws of large number)

$$\hat{\sigma}_\varepsilon^2(i) = \frac{\sum_{t=1}^T (\varepsilon_{it} - \bar{\varepsilon}_i)^2}{T-1}.$$

Since β (without constant term) must be estimated, we may use the residuals from the LSDV estimator (which is consistent and unbiased in general) and correct the degree of freedom to form the estimator:

$$s_e^2(i) = \frac{\sum_{t=1}^T (e_{it} - \bar{e}_i)^2}{T-k-1}. \quad (11-36)$$

We have N such estimators, so we average them to obtain

$$\bar{s}_e^2 = \frac{1}{N} \sum_{i=1}^N s_e^2(i) = \frac{1}{N} \left[\frac{\sum_{i=1}^N \sum_{t=1}^T (e_{it} - \bar{e}_i)^2}{T-k-1} \right] = \frac{\sum_{i=1}^N \sum_{t=1}^T (e_{it} - \bar{e}_i)^2}{NT - Nk - N}.$$

The degree of freedom correction in \bar{s}_e^2 is excessive because it assumes that α and β are re-estimated for each i . The unbiased estimator would be

$$\hat{\sigma}_\varepsilon^2 = s_{LSDV}^2 = \frac{\sum_{i=1}^N \sum_{t=1}^T (e_{it} - \bar{e}_i)^2}{NT - N - k}. \quad (11-37)$$

where \bar{e}_i should be 0.¹⁴ This is the same variance estimator as (11-17).¹⁵

¹³This can be proved by denoting that $\varepsilon = (\varepsilon_{i1} \ \varepsilon_{i2} \ \dots \ \varepsilon_{iT})'$, Then it is by assumption that $\varepsilon \sim N(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_T)$, and therefore $\frac{\sum_{t=1}^T (\varepsilon_{it} - \bar{\varepsilon}_i)^2}{\sigma_\varepsilon^2} \equiv \frac{\varepsilon' \mathbf{M}_0 \varepsilon}{\sigma_\varepsilon^2} \sim \chi_{T \text{ trace of } M_0 (=T-1)}^2$, then $E \left(\sum_{t=1}^T (\varepsilon_{it} - \bar{\varepsilon}_i)^2 \right) = (T-1)\sigma_\varepsilon^2$.

¹⁴To see this results, the fixed-effect model in terms of deviation from the group means can be derived as follows:

$$Y_{it} = \alpha_i + \mathbf{x}'_{it} \beta + \varepsilon_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T.$$

and in term of group means,

$$\bar{Y}_i = \alpha_i + \bar{\mathbf{x}}'_i \beta + \bar{\varepsilon}_i, \quad i = 1, 2, \dots, N.$$

Therefore, taking deviation from the group means removes the heterogeneity:

$$Y_{it} - \bar{Y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \beta + \varepsilon_{it} - \bar{\varepsilon}_i, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T. \quad (11-38)$$

Eq. (11-38)) is exactly the same with (11-35). Therefore the variance, σ_ε^2 of a random effects model can be consistently estimated from a fixed-effect model.

¹⁵It should be noted here that it seems the e_{it} is from (11-35) where no constant term is added. But from the example below, Greene indeed use the LSDV residuals to estimate σ_ε^2 .

5.2.2 Estimation of σ_v^2

It remains to estimate σ_v^2 . Back to the original model:

$$Y_{it} = \alpha + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it} + u_i, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T. \quad (11-39)$$

In spite of the correlation across observations, this is a classical regression model in which the OLS estimation slope and variance estimator are both consistent and, in general, unbiased. Therefore, using the OLS (pooling) residuals from the model with only a single overall constant, we have

$$\text{plim } s_{Pooled}^2 = \text{plim } \frac{\mathbf{e}'\mathbf{e}}{NT - k - 1} = \sigma_\varepsilon^2 + \sigma_v^2.$$

This provides the two estimators needed for the variance components; the second would be

$$\hat{\sigma}_u^2 = s_{Pooled}^2 - s_{LSDV}^2. \quad (11-40)$$

Example.

See Example 9.6 (Table 9.5) on p. 207 of Greene 6th edition. ■

5.3 Testing for Presence of Random Effects

The absence of an unobserved random effect in the random-effect model

$$Y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \alpha + v_i + \varepsilon_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T.$$

is statistical equivalent to the null hypothesis that $H_0 : \sigma_v^2 = 0$.

To test $H_0 : \sigma_v^2 = 0$, we can use the simple test for serial correlation because the error $u_{it} = \varepsilon_{it} + v_i$ are serially uncorrelated under the $H_0 : \sigma_v^2 = 0$.¹⁶

Under the null hypothesis, pooled OLS is efficient and all associated pooled OLS statistics are asymptotically valid. Breusch and Pagan (1980) have devised a Lagrange

¹⁶If $\sigma_v^2 \neq 0$, then $E(u_{it}u_{is}) = E[(\varepsilon_{it}v_i)(\varepsilon_{is}v_i)] = \sigma_v^2 \neq 0$.

multiplier test for the random effects model based on the OLS (pooling) residuals for¹⁷.

$$H_0 : \sigma_v^2 = 0, \quad (\text{or } \text{Cov}(u_{it}u_{is}) = 0),$$

$$H_1 : \sigma_v^2 \neq 0,$$

the test statistics is

$$LM = \frac{NT}{2(T-1)} \left[\frac{\sum_{i=1}^N (T\bar{e}_i)^2}{\sum_{i=1}^N \sum_{t=1}^T e_{it}^2} - 1 \right]^2 = \frac{NT}{2(T-1)} \left[\frac{\mathbf{e}'\mathbf{D}\mathbf{D}'\mathbf{e}}{\mathbf{e}'\mathbf{e}} - 1 \right]^2,$$

where \mathbf{e} is the OLS (pooling) residual vector and \mathbf{D} is matrix of dummy variables as defined in fixed effects model. Under the null hypothesis, LM is distributed as χ^2 with one degree of freedom.

Example.

See Example 13.3 on p. 299 of Greene 5th edition. ■

5.4 Hausman's Specification Test for the Random Effects Model

Since the key consideration in choosing between a random effects and a fixed effects approach is whether α_i (or v_i) and \mathbf{x}_{it} are correlated, it is important to have a method for testing this assumption. From a purely practical standpoint, the dummy variables approach is costly in terms of degree of freedom lost, and in a wide, longitudinal data set, the random effects model has some intuitive appeals. On the other hand, the fixed effects approach has one considerable virtue. There is no justification for treating the individual effects as uncorrelated with other regressors, as assumed in the random effects model. The random effects treatment, therefore, may suffer from the inconsistency due to omitted variables.

The specification test developed by Hausman (1978) is used to test for orthogonality of the random effects and the regressors. Under the null hypothesis of no correlation (between α_i (or v_i) and \mathbf{x}_{it}), both OLS $\hat{\beta}$, in the LSDV model and GLS, $\tilde{\beta}$, in the random effects model are consistent, but OLS is inefficient,¹⁸ whereas under the alternative (of correlation between α_i (or u_i) and \mathbf{x}_{it}), OLS from LSDV is consistent,

¹⁷Remember that a LM test statistics is a test statistics derived under the null hypothesis.

¹⁸Referring to the GLS matrix weighted average given earlier, we see that the efficient weight uses θ , whereas OLS sets θ only to 1.

but GLS is not. Therefore, under the null hypothesis, the two estimates should not differ systematically, and a test can be based on the difference.

The essential ingredient for the test is the covariance matrix of the difference vector, $[\hat{\beta} - \tilde{\beta}]$:

$$Var[\hat{\beta} - \tilde{\beta}] = Var[\hat{\beta}] + Var[\tilde{\beta}] - Cov[\hat{\beta}, \tilde{\beta}] - Cov[\hat{\beta}, \tilde{\beta}]'. \quad (11-41)$$

Hausman's essential result is that the covariance of an efficient estimator with its difference from an inefficient estimator is zero, which implies

$$Cov[(\hat{\beta} - \tilde{\beta}), \tilde{\beta}] = Cov[\hat{\beta}, \tilde{\beta}] - Var[\tilde{\beta}] = \mathbf{0}$$

or that

$$Cov[\hat{\beta}, \tilde{\beta}] = Var[\tilde{\beta}].$$

Inserting this result to (11-40) produces the required covariance matrix for the test

$$Var[\hat{\beta} - \tilde{\beta}] = Var[\hat{\beta}] - Var[\tilde{\beta}] = \Xi.$$

The chi-squared test is based on the Wald criterion:

$$W = [\hat{\beta} - \tilde{\beta}]' \hat{\Xi}^{-1} [\hat{\beta} - \tilde{\beta}] \sim \chi^2_{[k]}.$$

For $\hat{\Xi}$, we use the estimated covariance matrices of the slope estimator in the LSDV model and estimated covariance matrix in the random effects model, excluding the constant term. Under the null hypothesis, W is asymptotically distributed as chi-squared with k degrees of freedom.

Exercise 3.

Reproduce the results of Example 9.7 on p. 209 of Greene 6th edition. ■

5.5 Unbalanced Panels and Random Effects

Unbalanced panels add a layer of difficulty in the random effects model. The first problem can be seen in (11-31). The matrix Ω is no longer $\mathbf{I}_N \otimes \Sigma$ because the

diagonal blocks in Σ_i are of different sizes. There is also groupwise heteroscedasticity in (11-34), because the i th diagonal block in $\Omega^{-1/2}$ is

$$\Sigma_i^{-1/2} = \frac{1}{\sigma_\varepsilon} \left[\mathbf{I}_{T_i} - \frac{\theta}{T_i} \mathbf{i}_{T_i} \mathbf{i}_{T_i}' \right], \quad (11-42)$$

and

$$\theta_i = 1 - \frac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2 + T_i \sigma_v^2}}.$$

The transformation of \mathbf{y}_i and \mathbf{X}_i for GLS in unbalanced panel is therefore

$$\Sigma_i^{-1/2} \mathbf{y}_i = \frac{1}{\sigma_\varepsilon} \begin{bmatrix} Y_{i1} - \theta_i \bar{Y}_i \\ Y_{i2} - \theta_i \bar{Y}_i \\ \vdots \\ Y_{iT_i} - \theta_i \bar{Y}_i \end{bmatrix},$$

and likewise for the rows of \mathbf{X}_i (including constant term).

5.6 Nonspherical Disturbances and Robust Covariance Estimation

Since the random effects model is already a generalized regression model with a known structure, robust estimation of the covariance matrix for the OLS estimator in this context is not the best use of the data.¹⁹

¹⁹Can we use OLS to estimate the random effects model with robust covariance estimator ?

6 Dynamic Models

The analysis so far has involved static models and relatively straightforward estimation problems. Dynamic effects in the panel data model, with or without heterogeneity, raise complex new issues in estimation and inference. This section will explore a few of the specifications and show that the familiar estimation techniques (OLS, FGLS, etc.) are not effective in these cases.

6.1 Random Effects Model

Consider a homogeneous dynamic panel data model

$$Y_{it} = \phi Y_{i,t-1} + \mathbf{x}_{it}'\boldsymbol{\beta} + v_i + \varepsilon_{it}, \quad (11-43)$$

where $|\phi| < 1$ and v_i is, as in the preceding sections of this chapter, individual unmeasured heterogeneity, that may or may not be correlated with \mathbf{x}_{it} . We consider methods of estimation for this model when T is fixed and relatively small, and N may be large and increasing.

The disturbance $v_i + \varepsilon_{it}$ may be correlated with \mathbf{x}_{it} , but either way, it is surely correlated with $Y_{i,t-1}$. By substitution,

$$Cov[Y_{i,t-1}, (v_i + \varepsilon_{it})] = \sigma_v^2 + \phi Cov[Y_{i,t-2}, (v_i + \varepsilon_{it})],$$

and so on. By repeated substitution, it can be seen that for moderately large T ,

$$Cov[Y_{i,t-1}, (v_i + \varepsilon_{it})] \approx \frac{\sigma_v^2}{1 - \phi}.$$

Consequently, OLS (pooled) and GLS are inconsistent.

6.2 Fixed Effects Model

The fixed effects approach does not solve the problem either. Let

$$Y_{it} = \alpha_i + \phi Y_{i,t-1} + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it},$$

taking deviations from individual means, we have

$$Y_{it} - \bar{Y}_i = (\mathbf{x}'_{it} - \bar{\mathbf{x}}'_i)\boldsymbol{\beta} + \phi(Y_{i,t-1} - \bar{Y}_i) + \varepsilon_{it} - \bar{\varepsilon}_i.$$

Anderson and Hsiao (1982) showed that²⁰

$$\begin{aligned} & Cov((Y_{i,t-1} - \bar{Y}_i), (\varepsilon_{it} - \bar{\varepsilon}_i)) \\ &= E \left[\left(Y_{i,t-1} - \frac{Y_{i,1} + Y_{i,2} + \cdots + Y_{i,T}}{T} \right) \left(\varepsilon_{it} - \frac{\varepsilon_{i,1} + \varepsilon_{i,2} + \cdots + \varepsilon_{i,T}}{T} \right) \right] \\ &\approx \frac{-\sigma_\varepsilon^2}{T(1-\phi)^2} \left[\frac{(T-1) - T\phi + \phi^T}{T} \right] \\ &= \frac{-\sigma_\varepsilon^2}{T(1-\phi)^2} \left[(1-\phi) - \frac{1-\phi^T}{T} \right]. \end{aligned} \quad (11-44)$$

This does converge to zero as T increases, but, again, we are considering cases in which T is small or moderate, say 5 to 15, in which case, the bias in the OLS estimator could be 15 percent to 60 percent. The implication is that the “within” transformation does not produce a consistent estimator.

6.3 Instrumental Variables Estimation

The general approach, which has been developed in several stages in the literature, relies on instrumental variables estimators and most recently on a GMM estimator. In either the fixed or the random effects cases, the heterogeneity can be swept from the model by taking first differences, which produces

$$Y_{it} - Y_{i,t-1} = (\mathbf{x}_{it} - \mathbf{x}_{i,t-1})'\boldsymbol{\beta} + \phi(Y_{i,t-1} - Y_{i,t-2}) + (\varepsilon_{it} - \varepsilon_{i,t-1}).$$

This model is still complicated by correlation between the lagged dependent variable and the disturbance,

$$Cov[(Y_{i,t-1} - Y_{i,t-2}), (\varepsilon_{it} - \varepsilon_{i,t-1})] = E(Y_{i,t-1} \cdot \varepsilon_{i,t-1}) \neq 0. \quad (11-45)$$

But without the group effects, there is a simple instrumental variables estimator available. Assuming that the time series is long enough, one could use the differences, $(Y_{i,t-2} - Y_{i,t-3})$, or the lagged levels, $Y_{i,t-2}$ and $Y_{i,t-3}$, as one or two instrumental variables for $(Y_{i,t-1} - Y_{i,t-2})$. By this construction, then, the treatment of this model is

²⁰See also the famous Nickell (1981)’s “Nickell bias”.

a standard application of the instrumental variables technique that we developed in Section 3.2 of Chapter 7. There is a question as to whether one should use differences or levels as instruments. Arellano (1989) gives evidence that the latter is preferable.

7 Appendix: Proofs of Hausman Test

The theory underlying the proposed specification tests rests on one fundamental idea. Under the null hypothesis of no misspecification, there will exist a *consistent*, asymptotically normal and asymptotically *efficient* estimator, where efficiency means attaining the asymptotic Cramer-Rao bound. Under the alternative hypothesis of misspecification, however, this estimator (random-effected estimator) will be biased and *inconsistent*.

To construct a test of misspecification, it is necessary to find another estimator which is not adversely affected by the misspecification; but this estimator will not be asymptotically efficient under the null hypothesis. A consideration of the difference between the two estimators, $\mathbf{q} = \hat{\beta}_1 - \tilde{\beta}_0$, where $\tilde{\beta}_0$ is the efficient estimator under H_0 and $\hat{\beta}_1$ is a consistent estimator under H_1 , will then lead to a specification test. If no misspecification is present, the probability limit of \mathbf{q} is zero. With misspecification $\text{plim } \mathbf{q}$ will differ from zero.

In constructing tests based on \mathbf{q} , an immediate problem comes to mind. To develop test not only is the probability limit of \mathbf{q} required, but the variance of the asymptotic distribution of $\sqrt{T}\mathbf{q}$, i.e. $\mathbf{V}(\mathbf{q})$, must also be determined. Luckily, the calculation is easy and in fact, $\mathbf{V}(\mathbf{q}) = \mathbf{V}(\hat{\beta}_1) - \mathbf{V}(\tilde{\beta}_0) \equiv \mathbf{V}_1 - \mathbf{V}_0$ under the null hypothesis of no misspecification. Thus, the construction of specification error tests is simplified, since the estimators may be considered separately because the variance of the difference $\sqrt{T}\mathbf{q} = \sqrt{T}(\hat{\beta}_1 - \tilde{\beta}_0)$ is the difference of the respective variance. To prove the result formally, consider the following lemma.

Lemma.

Consider two estimators $\tilde{\beta}_0$ and $\hat{\beta}_1$ which are both consistent and asymptotically normally distributed with $\tilde{\beta}_0$ attaining the asymptotic Cramer-Rao bound so $\sqrt{T}(\tilde{\beta}_0 - \beta) \sim N(\mathbf{0}, \mathbf{V}_0)$ and $\sqrt{T}(\hat{\beta}_1 - \beta) \sim N(\mathbf{0}, \mathbf{V}_1)$, where \mathbf{V}_0 is the inverse of Fisher's information matrix. Let the difference between the two estimators be $\mathbf{q} = \hat{\beta}_1 - \tilde{\beta}_0$. The limiting distribution of the efficient estimator, $\sqrt{T}(\tilde{\beta}_0 - \beta)$, and $\sqrt{T}\mathbf{q}$ have zero covariance, $\text{Cov}(\tilde{\beta}_0, \mathbf{q}) = \text{Cov}(\tilde{\beta}_0, \hat{\beta}_1 - \tilde{\beta}_0) = \mathbf{0}$, a zero matrix.

Proof.

Suppose $\tilde{\beta}_0$ and \mathbf{q} are not orthogonal. Since $\text{plim } \mathbf{q} = \mathbf{0}$ (both estimators are consistent), we define a new estimator $\tilde{\beta}_2 = \tilde{\beta}_0 + r\mathbf{A}\mathbf{q}$ where r is a scalar and \mathbf{A} is an

arbitrary $k \times k$ matrix to be chosen. The new estimator is consistent and asymptotic normal with asymptotic variance

$$\mathbf{V}(\tilde{\beta}_2) = \mathbf{V}(\tilde{\beta}_0) + r\mathbf{A} \cdot \text{Cov}(\tilde{\beta}_0, \mathbf{q}) + r \cdot \text{Cov}(\tilde{\beta}_0, \mathbf{q})' \mathbf{A}' + r^2 \mathbf{A} \mathbf{V}(\mathbf{q}) \mathbf{A}'.$$

Now consider the difference between the asymptotic variance of the new estimator and the old asymptotically efficient estimator

$$\mathbf{F}(r) = \mathbf{V}(\tilde{\beta}_2) - \mathbf{V}(\tilde{\beta}_0) = r\mathbf{A} \cdot \text{Cov}(\tilde{\beta}_0, \mathbf{q}) + r \cdot \text{Cov}(\tilde{\beta}_0, \mathbf{q})' \mathbf{A}' + r^2 \mathbf{A} \mathbf{V}(\mathbf{q}) \mathbf{A}'.$$

Taking derivative with respect to r yields

$$\mathbf{F}'(r) = \mathbf{A} \cdot \text{Cov}(\tilde{\beta}_0, \mathbf{q}) + \text{Cov}(\tilde{\beta}_0, \mathbf{q})' \mathbf{A}' + 2r \mathbf{A} \mathbf{V}(\mathbf{q}) \mathbf{A}'.$$

Then choose $\mathbf{A} = -\text{Cov}(\tilde{\beta}_0, \mathbf{q})'$ and note that $\text{Cov}(\tilde{\beta}_0, \mathbf{q})$ is symmetric, which leads to

$$\mathbf{F}'(r) = -2\text{Cov}(\tilde{\beta}_0, \mathbf{q})' \text{Cov}(\tilde{\beta}_0, \mathbf{q}) + 2r \text{Cov}(\tilde{\beta}_0, \mathbf{q})' \mathbf{V}(\mathbf{q}) \text{Cov}(\tilde{\beta}_0, \mathbf{q}).$$

Therefore at $r = 0$, $\mathbf{F}'(0) = -2\text{Cov}(\tilde{\beta}_0, \mathbf{q})' \text{Cov}(\tilde{\beta}_0, \mathbf{q})$ is the sense of being nonpositive definite, which means that for r small²¹

$$\begin{aligned} F'_j(0) &= \lim_{r \rightarrow 0^+} \frac{F_j(r) - F_j(0)}{r - 0} \\ &= \lim_{r \rightarrow 0^+} \frac{F_j(r) - 0}{r} \leq 0, \quad \forall j = 1, 2, \dots, k, \end{aligned}$$

implying $F_j(r) < 0$, $\forall j$.

There is a contradiction unless $\text{Cov}(\tilde{\beta}_0, \mathbf{q}) = \mathbf{0}$ since $\tilde{\beta}_0$ is asymptotically efficient already implies $\mathbf{F}(r)$ is positive definite or $F_j(r) \geq 0$, $\forall j$. ■

Once it has been shown that the efficient estimator is uncorrelated with \mathbf{q} , the asymptotic variance of \mathbf{q} is easily calculated.

Corollary.

$\mathbf{V}(\mathbf{q}) = \mathbf{V}(\hat{\beta}_1) - \mathbf{V}(\tilde{\beta}_0) \equiv \mathbf{V}_1 - \mathbf{V}_0$ is nonnegative definite.

Proof.

Since $\mathbf{q} + \tilde{\beta}_0 = \hat{\beta}_1$, then

$$\mathbf{V}(\mathbf{q}) + \mathbf{V}(\tilde{\beta}_0) + 2\text{Cov}(\tilde{\beta}_0, \mathbf{q}) = \mathbf{V}(\hat{\beta}_1).$$

²¹ $\mathbf{F}(0) = \mathbf{0}$.

Because it has been shown that $Cov(\tilde{\beta}_0, \mathbf{q}) = \mathbf{0}$, thus

$\mathbf{V}(\mathbf{q}) = \mathbf{V}(\hat{\beta}_1) - \mathbf{V}(\tilde{\beta}_0)$ is apparent.²² ■

Given the above result a general misspecification test can be specified by considering the statistics

$$W = T\mathbf{q}'\hat{\mathbf{V}}(\mathbf{q})^{-1}\mathbf{q}$$

where $\hat{\mathbf{V}}(\mathbf{q})$ is a consistent estimator of $\mathbf{V}(\mathbf{q})$. This statistic will shown to be distributed asymptotically as central χ_k^2 under the null hypothesis where k is the number of unknown parameters in β when no misspecification is present.



The Seashore-Entrance to NSYSU.

End of this Chapter

²²It is to be noted that here $\mathbf{V}(\mathbf{q})$ is the variance of $\sqrt{T}\mathbf{q}$.