CHAPTER 21

Linear Panel Models: Basics

21.1. Introduction

Panel data are repeated observations on the same cross section, typically of individuals or firms in microeconomics applications, observed for several time periods. Other terms used for such data include **longitudinal data** and **repeated measures**. The focus is on data from a **short panel**, meaning a large cross section of individuals observed for a few time periods, rather than a long panel such as a small cross section of countries observed for many time periods.

A major advantage of panel data is increased precision in estimation. This is the result of an increase in the number of observations owing to combining or **pooling** several time periods of data for each individual. However, for valid statistical inference one needs to control for likely correlation of regression model errors over time for a given individual. In particular, the usual formula for OLS standard errors in a pooled OLS regression typically overstates the precision gains, leading to underestimated standard errors and t-statistics that can be greatly inflated.

A second attraction of panel data is the possibility of consistent estimation of the **fixed effects** model, which allows for unobserved individual heterogeneity that may be correlated with regressors. Such unobserved heterogeneity leads to **omitted variables bias** that could in principle be corrected by instrumental variables methods using only a single cross section, but in practice it can be difficult to obtain a valid instrument. Data from a short panel, with as few as two periods, offers an alternative way to proceed if the unobserved individual-specific effects are assumed to be additive and time-invariant.

Most disciplines in applied statistics other than microeconometrics treat any unobserved individual heterogeneity as being distributed independently of the regressors. Then the effects are called **random effects**, though a better term is *purely* random effects. Compared to fixed effects models this stronger assumption has the advantage of permitting consistent estimation of all parameters, including coefficients of time-invariant regressors. However, random effects and pooled estimators are inconsistent if the true model is one with fixed effects. Economists often view the assumptions for the random effects model as being unsupported by the data.

A third attraction of panel data is the possibility of learning more about the **dynamics** of individual behavior than is possible from a single cross section. Thus a cross section may yield a poverty rate of 20% but we need panel data to determine whether the same 20% are in poverty each year. As a related example, panel data may determine whether high serial correlation of individual earnings or unemployment spell length is due to an individual specific tendency to have high earnings or a long unemployment spell, or whether it is a consequence of having past high earnings or unemployment. This topic is deferred to Chapter 22.

The linear panel data models and associated estimators are conceptually simple, aside from the fundamental issue of whether or not fixed effects are necessary. The considerable algebra used to derive the properties of panel data estimators should not distract one from an understanding of the basics: The statistical properties of panel data estimators vary with the assumed model and its treatment of unobserved effects. Furthermore, much of the algebra does not generalize to nonlinear panel models.

The current chapter presents the basic estimators for various linear panel data models. A lengthy introduction in Sections 21.2 and 21.3 provides, respectively, the commonly used models and estimators and an application to the relationship between annual hours worked and wages. The important distinction between fixed and random effects models is studied in Section 21.4. Sections 21.5–21.7 present additional detail on estimation for, respectively, pooled models, individual-specific fixed effects models, and individual-specific random effects models. Section 21.8 considers other basic aspects such as inference and prediction in linear panel data models.

21.2. Overview of Models and Estimators

Panel data provide information on individual behavior both across time and across individuals.

Even for linear regression, standard panel data analysis uses a much wider range of models and estimators than is the case with cross-section data. Several standard models are presented in Section 21.2.1, followed by several estimators presented in Section 21.2.2. Table 21.1 gives a summary that also indicates that several of the estimators are inconsistent if the dgp is the **individual-specific fixed effects model**.

Obtaining correct standard errors of estimators is also more complicated than in the cross-section case. One needs to control for correlation over time in errors for a given individual, in addition to possible heteroskedasticity. This topic is covered in Section 21.2.3.

21.2.1. Panel Data Models

A very general linear model for panel data permits the intercept and slope coefficients to vary over both individual and time, with

$$y_{it} = \alpha_{it} + \mathbf{x}_{it}'\boldsymbol{\beta}_{it} + u_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T,$$

Table 21.1. *Linear Panel Model: Common Estimators and Models*[a]

	Assumed Model		
Estimator of β	**Pooled (21.1)**	**Random Effects (21.3) and (21.5)**	**Fixed Effects (21.3) Only**
Pooled OLS (21.1)	Consistent	Consistent	Inconsistent
Between (21.7)	Consistent	Consistent	Inconsistent
Within (or Fixed Effects) (21.8)	Consistent	Consistent	Consistent
First Differences (21.9)	Consistent	Consistent	Consistent
Random Effects (21.10)	Consistent	Consistent	Inconsistent

[a] This table considers only consistency of estimators of β. For correct computation of standard errors see Section 21.2.3.

where y_{it} is a scalar dependent variable, $\mathbf{x}_{it}$ is a $K \times 1$ vector of independent variables, u_{it} is a scalar disturbance term, i indexes individual (or firm or country) in a cross section, and t indexes time.

This model is too general and is not estimable as there are more parameters to estimate than observations. Further restrictions need to be placed on the extent to which α_{it} and β_{it} vary with i and t, and on the behavior of the error u_{it}.

Pooled Model

The most restrictive model is a **pooled model** that specifies **constant coefficients**, the usual assumption for cross-section analysis, so that

$$y_{it} = \alpha + \mathbf{x}_{it}'\beta + u_{it}. \tag{21.1}$$

If this model is correctly specified and regressors are uncorrelated with the error then it can be consistently estimated using **pooled OLS**. The error term is likely to be correlated over time for a given individual, however, in which case the usual reported standard errors should not be used as they can be greatly downward biased. Furthermore, the pooled OLS estimator is inconsistent if the fixed effects model, defined in the following, is appropriate.

Individual and Time Dummies

A simple variant of the model (21.1) permits intercepts to vary across individuals and over time while slope parameters do not. Then $y_{it} = \alpha_i + \gamma_t + \mathbf{x}_{it}'\beta + u_{it}$, or

$$y_{it} = \sum_{j=1}^{N} \alpha_j d_{j,it} + \sum_{s=2}^{T} \gamma_s d_{s,it} + \mathbf{x}_{it}'\beta, \tag{21.2}$$

where the N **individual dummies** $d_{j,it}$ equal one if $i = j$ and equal zero otherwise, the $(T-1)$ **time dummies** $d_{s,it}$ equal one if $t = s$ and equal zero otherwise, and it is assumed that $\mathbf{x}_{it}$ does not include an intercept. (If an intercept is included then one of the N individual dummies must be dropped).

This model has $N + (T - 1) + \dim[\mathbf{x}]$ parameters that can be consistently estimated if both $N \to \infty$ and $T \to \infty$. We focus on **short panels** where $N \to \infty$ but T does not. Then the γ_s can be consistently estimated, so the $(T - 1)$ time dummies are simply incorporated into the regressors $\mathbf{x}_{it}$. The challenge then lies in estimating the parameters $\boldsymbol{\beta}$ controlling for the N individual intercepts α_i. One possibility is to instead have dummies for groups of observations, such as grouping by region, in which case the clustering methods of Chapter 24 are relevant. Here instead we specify a full set of N individual intercepts, which causes problems as $N \to \infty$.

Fixed Effects and Random Effects Models

The **individual-specific effects model** allows each cross-sectional unit to have a different intercept term though all slopes are the same, so that

$$y_{it} = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}, \tag{21.3}$$

where ε_{it} is iid over i and t. This is a more parsimonious way to express model (21.2), with any time dummies included in the regressors $\mathbf{x}_{it}$. The α_i are random variables that capture **unobserved heterogeneity**, already studied in Sections 18.2–18.5 and 20.4.

Throughout this chapter we make the assumption of **strong exogeneity** or **strict exogeneity**

$$\mathrm{E}[\varepsilon_{it}|\alpha_i, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}] = 0, \quad t = 1, \ldots, T, \tag{21.4}$$

so that the error term is assumed to have mean zero conditional on past, current, and future values of the regressors. Chamberlain (1980) gives a detailed discussion of exogeneity assumptions and tests for exogeneity for panel data. Strong exogeneity rules out models with lagged dependent variables or with endogenous variables as regressors; these models are deferred to Chapter 22.

One variant of the model (21.3) treats α_i as an unobserved random variable that is potentially correlated with the observed regressors $\mathbf{x}_{it}$. This variant is called the **fixed effects (FE) model** as early treatments modeled these effects as parameters $\alpha_1, \ldots, \alpha_N$ to be estimated. If fixed effects are present and correlated with $\mathbf{x}_{it}$ then many estimators such as pooled OLS are inconsistent. Instead, alternative estimation methods that eliminate the α_i are needed to ensure consistent estimation of $\boldsymbol{\beta}$ in a short panel.

The other variant of the model (21.3) assumes that the unobservable individual effects α_i are random variables that are distributed independently of the regressors. This model is called the **random effects (RE) model**, which usually makes the additional assumptions that

$$\begin{aligned} \alpha_i &\sim \left[\alpha, \sigma_\alpha^2\right], \\ \varepsilon_{it} &\sim \left[0, \sigma_\varepsilon^2\right], \end{aligned} \tag{21.5}$$

so that both the random effects and the error term in (21.3) are assumed to be iid. Note that no specific distributions have been specified in (21.5). A more precise term for this model is the **one-way individual-specific random effects model,** or more simply the **random intercept model**, to distinguish the model with more general random effects models such as the **mixed linear models** presented in Section 22.8. Yet another name is the **random components model**.

The term fixed effect is potentially misleading and the term random effect is more precisely a purely random effect. To avoid such confusion, M-J. Lee (2002) calls a fixed effect a "related effect" and a random effect an "unrelated effect." We use the traditional notation and terminology, but it should be clear that α_i is a random variable in both fixed and random effects models.

Equicorrelated Model

The RE model can be viewed as a specialization of the pooled model, as the α_i can be subsumed into the error term. Then (21.3) can be viewed as regression of y_{it} on $\mathbf{x}_{it}$ with composite error term $u_{it} = \alpha_i + \varepsilon_{it}$, and (21.5) implies that

$$\mathrm{Cov}[(\alpha_i + \varepsilon_{it}), (\alpha_i + \varepsilon_{is})] = \begin{cases} \sigma_\alpha^2, & t \neq s, \\ \sigma_\alpha^2 + \sigma_\varepsilon^2, & t = s. \end{cases} \tag{21.6}$$

The RE model therefore imposes the constraint that the composite error u_{it} is **equicorrelated**, since $\mathrm{Cor}[u_{it}, u_{is}] = \sigma_\alpha^2/[\sigma_\alpha^2 + \sigma_\varepsilon^2]$ for $t \neq s$ does not vary with the time difference $t - s$. Clearly, pooled OLS will be consistent but inefficient under the RE model. The random effects model is also called the **equicorrelated model** or **exchangeable errors model.**

Fixed versus Random Effects Models

The fundamental distinction is between models with and without fixed effects. The modern econometrics literature emphasizes fixed effects, but we also provide details for the random effects model.

Some authors, including Chamberlain (1980, 1984) and Wooldridge (2002), use the notation

$$y_{it} = c_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}$$

in (21.3) to make it very clear that the individual effect is a random variable in both fixed and random effects models. Both models assume that

$$\mathrm{E}[y_{it}|c_i, \mathbf{x}_{it}] = c_i + \mathbf{x}_{it}'\boldsymbol{\beta}.$$

The individual-specific effect c_i is unknown and in short panels cannot be consistently estimated, so we cannot estimate $\mathrm{E}[y_{it}|c_i, \mathbf{x}_{it}]$. Instead, we can eliminate c_i by taking the expectation with respect to c_i, leading to

$$\mathrm{E}[y_{it}|\mathbf{x}_{it}] = \mathrm{E}[c_i|\mathbf{x}_{it}] + \mathbf{x}_{it}'\boldsymbol{\beta}.$$

For the RE model it is assumed that $\mathrm{E}[c_i|\mathbf{x}_{it}] = \alpha$, so $\mathrm{E}[y_{it}|\mathbf{x}_{it}] = \alpha + \mathbf{x}_{it}'\boldsymbol{\beta}$ and hence it is possible to identify $\mathrm{E}[y_{it}|\mathbf{x}_{it}]$. In the FE model, however, $\mathrm{E}[c_i|\mathbf{x}_{it}]$ varies with $\mathbf{x}_{it}$ and it is not known how it varies, so we cannot identify $\mathrm{E}[y_{it}|\mathbf{x}_{it}]$. It is nonetheless possible to consistently estimate $\boldsymbol{\beta}$ in the FE model with short panels (as will be discussed in the following). Thus it is possible in the FE model to identify the marginal effect

$$\boldsymbol{\beta} = \partial \mathrm{E}[y_{it}|c_i, \mathbf{x}_{it}]/\partial \mathbf{x}_{it},$$

even though the conditional mean is not identified. For example, it is possible to identify the effect on earnings of an additional year of schooling, controlling for individual effects, even though the individual effects and the conditional mean are not identified.

In short panels the FE model permits only **identification** of the **marginal effect** $\partial \mathrm{E}[y_{it}|c_i, \mathbf{x}_{it}]/\partial \mathbf{x}_{it}$, and even then **only for time-varying regressors**, so the marginal effect of race or gender, for example, is not identified. The RE model permits identification of all components of $\boldsymbol{\beta}$ and of $\mathrm{E}[y_{it}|\mathbf{x}_{it}]$, but the key RE assumption that $\mathrm{E}[c_i|\mathbf{x}_{it}]$ is constant is viewed as untenable in many microeconometrics applications.

21.2.2. Panel Data Estimators

We now introduce several commonly used panel data estimators of $\boldsymbol{\beta}$, with further detail provided in Sections 21.5–21.7. The estimators differ in the extent to which cross-section and time-series variation in the data are used, and their properties vary according to whether or not the fixed effects model is the appropriate model.

A regressor x_{it} may be either **time-invariant**, with $x_{it} = x_i$ for $t = 1, \ldots, T$, or **time-varying**. For some estimators, notably the within and first differences estimators defined in the following, only the coefficients of time-varying regressors are identified.

Pooled OLS

The **pooled OLS estimator** is obtained by stacking the data over i and t into one long regression with NT observations, and estimating by OLS

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

If $\mathrm{Cov}[u_{it}, \mathbf{x}_{it}] = \mathbf{0}$ then either $N \to \infty$ or $T \to \infty$ is sufficient for consistency.

The pooled OLS estimator is clearly consistent if the pooled model (21.1) is appropriate and regressors are uncorrelated with the error term. The usual OLS variance matrix based on iid errors, however, is not appropriate here as the errors for a given individual i are almost certainly positively correlated over t. The *NT* correlated observations have less information than *NT* independent observations.

To understand this correlation, note that for a given individual we expect considerable correlation in y over time, so that $\mathrm{Cor}[y_{it}, y_{is}]$ is high. Even after inclusion of regressors $\mathrm{Cor}[u_{it}, u_{is}]$ may remain nonzero, and it often can still be quite high. For example, if a model overpredicts individual earnings in one year it may also overpredict earnings for the same individual in other years. The RE model accommodates this correlation, with $\mathrm{Cor}[u_{it}, u_{is}] = \sigma_\alpha^2/[\sigma_\alpha^2 + \sigma_\varepsilon^2]$ for $t \neq s$ from (21.6).

The usual OLS output treats each of the T years as independent pieces of information, but the information content is less than this given the positive error correlation. This leads to overstatement of estimator precision that can be very large, as illustrated in Section 21.3.2 and formally demonstrated in Section 21.5.4. One therefore needs to use panel-corrected standard errors (see Section 21.2.3) whenever OLS is applied in a panel setting. Many corrections are possible, depending on the correlation and heteroskedasticity structure assumed for the errors and whether the panel is short or long (see Section 21.5).

The pooled OLS estimator is inconsistent if the true model is the fixed effects model. To see this, rewrite the model (21.3) as

$$y_{it} = \alpha + \mathbf{x}_{it}'\boldsymbol{\beta} + (\alpha_i - \alpha + \varepsilon_{it}).$$

Then pooled OLS regression of y_{it} on $\mathbf{x}_{it}$ and an intercept leads to an inconsistent estimator of $\boldsymbol{\beta}$ if the individual effect α_i is correlated with the regressors $\mathbf{x}_{it}$, since such correlation implies that the combined error term $(\alpha_i - \alpha + \varepsilon_{it})$ is correlated with the regressors.

In summary, pooled OLS is appropriate if the constant-coefficients or random effects models are appropriate, but panel-corrected standard errors and t-statistics must be used for statistical inference. Pooled OLS is inconsistent if the fixed effects model is appropriate.

Between Estimator

The pooled OLS estimator uses variation over both time and cross-sectional units to estimate $\boldsymbol{\beta}$.

The between estimator in short panels instead uses just the cross-sectional variation. Begin with the individual-specific effects model (21.3). Averaging over all years yields $\bar{y}_i = \alpha_i + \bar{\mathbf{x}}_i'\boldsymbol{\beta} + \bar{\varepsilon}_i$, which can be rewritten as the **between model**

$$\bar{y}_i = \alpha + \bar{\mathbf{x}}_i'\boldsymbol{\beta} + (\alpha_i - \alpha + \bar{\varepsilon}_i), \quad i = 1, \ldots, N, \tag{21.7}$$

where $\bar{y}_i = T^{-1}\sum_t y_{it}$, $\bar{\varepsilon}_i = T^{-1}\sum_t \varepsilon_{it}$, and $\bar{\mathbf{x}}_i = T^{-1}\sum_t \mathbf{x}_{it}$.

The **between estimator** is the OLS estimator from regression of $\bar{y}_i$ on an intercept and $\bar{\mathbf{x}}_i$. It uses variation between different individuals and is the analogue of cross-section regression, which is the special case $T = 1$.

The between estimator is consistent if the regressors $\bar{\mathbf{x}}_i$ are independent of the composite error $(\alpha_i - \alpha + \bar{\varepsilon}_i)$ in (21.7). This will be the case for the constant-coefficients model and the random effects model. In contrast, for the fixed effects model the between estimator is inconsistent as α_i is then assumed to be correlated with $\mathbf{x}_{it}$ and hence $\bar{\mathbf{x}}_i$.

Within Estimator or Fixed Effects Estimator

The within estimator is an estimator that, unlike the pooled OLS or between estimators, exploits the special features of panel data. In a short panel it measures the association between individual-specific deviations of regressors from their time-averaged values and individual-specific deviations of the dependent variable from its time-averaged value. This is done using the variation in the data over time.

Specifically, begin with the individual-specific effects model (21.3), which nests (21.1) as the special case $\alpha_i = \alpha$. Then taking the average over time yields $\bar{y}_i = \alpha_i + \bar{\mathbf{x}}_i'\boldsymbol{\beta} + \bar{\varepsilon}_i$. Subtracting this from y_{it} in (21.3) yields the **within model**

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i), \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{21.8}$$

as the α_i terms cancel.

The **within estimator** is the OLS estimator in (21.8). A special feature of this estimator is that it yields consistent estimates of $\boldsymbol{\beta}$ in the fixed effects model, whereas the pooled OLS and between estimators do not.

From Section 21.6 the within estimator has several interpretations. It is called the **fixed effects estimator** as it is the efficient estimator of $\boldsymbol{\beta}$ in the model (21.3) if α_i are fixed effects and the error ε_{it} is iid. This chapter focuses on a literature that treats fixed effects as **nuisance parameters** that can be ignored since interest lies solely in estimation of $\boldsymbol{\beta}$. If instead the fixed effects are of interest they can also be estimated. In short panels these estimates of the individual α_i are inconsistent, though their distribution or their variation with a key variable may be informative. If N is not too large an alternative and simpler way to compute the within estimator is by **least-squares dummy variable estimation**. This directly estimates (21.2) by OLS regression of y_{it} on $\mathbf{x}_{it}$ and the N individual dummy variables and yields the within estimator for $\boldsymbol{\beta}$, along with estimates of the N fixed effects (see Section 21.6.4). Yet another interpretation of the within estimator is the covariance estimator. Finally, taking deviations from individual-specific averages is equivalent to taking residuals from auxiliary regression of y_{it} and $\mathbf{x}_{it}$ on individual dummies and then working with the residuals.

A major limitation of within estimation is that the coefficients of time-invariant regressors are not identified in the within model, since if $x_{it} = x_i$ then $\bar{x}_i = x_i$ so $(x_{it} - \bar{x}_i) = 0$. Many studies seek to estimate the effect of time-invariant regressors. For example, in panel wage regressions we may be interested in the effect of gender or race. For this reason many practitioners prefer not to use the within estimator. Pooled OLS or random effects estimators permit estimation of coefficients of time-invariant regressors, but these estimators are inconsistent if the fixed effects model is the correct model.

First-Differences Estimator

The first-differences estimator also exploits the special features of panel data. In a short panel it measures the association between individual-specific one-period changes in regressors and individual-specific one-period changes in the dependent variable.

Specifically, begin with the individual-specific effects model (21.3). Then lagging one period yields $y_{i,t-1} = \alpha_i + \mathbf{x}_{i,t-1}'\boldsymbol{\beta} + \varepsilon_{i,t-1}$. Subtracting this from y_{it} in (21.3) yields the **first-differences model**

$$y_{it} - y_{i,t-1} = (\mathbf{x}_{it} - \mathbf{x}_{i,t-1})'\boldsymbol{\beta} + (\varepsilon_{it} - \varepsilon_{i,t-1}), \quad i = 1, \ldots, N, \quad t = 2, \ldots, T, \tag{21.9}$$

as the α_i terms cancel.

The **first-differences estimator** is the OLS estimator in (21.9). Like the within estimator, this estimator yields consistent estimates of $\boldsymbol{\beta}$ in the fixed effects model, though the coefficients of time-invariant regressors are not identified. The first-differences estimator is less efficient than the within estimator for $T > 2$ if ε_{it} is iid.

Random Effects Estimator

The random effects estimator is an estimator that also exploits the special features of panel data.

Begin with the individual-specific effects model (21.3), but assume a random effects model where α_i and ε_{it} are iid as in (21.5). Pooled OLS is consistent but pooled GLS will be more efficient. The **feasible GLS estimator** (see Section 4.5.1) of the RE model, called the **random effects estimator**, can be calculated from OLS estimation of the transformed model

$$y_{it} - \widehat{\lambda}\bar{y}_i = (1-\widehat{\lambda})\mu + (\mathbf{x}_{it} - \widehat{\lambda}\bar{\mathbf{x}}_i)'\boldsymbol{\beta} + v_{it}, \tag{21.10}$$

where $v_{it} = (1-\widehat{\lambda})\alpha_i + (\varepsilon_{it} - \widehat{\lambda}\bar{\varepsilon}_i)$ is asymptotically iid, and $\widehat{\lambda}$ is consistent for

$$\lambda = 1 - \frac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2 + T\sigma_\alpha^2}}. \tag{21.11}$$

Section 21.7 provides a derivation of (21.10) and ways to estimate σ_α^2 and σ_ε^2 and hence to estimate λ. Note that $\widehat{\lambda} = 0$ corresponds to pooled OLS, $\widehat{\lambda} = 1$ corresponds to within estimation, and $\widehat{\lambda} \to 1$ as $T \to \infty$. This is a two-step estimator of $\boldsymbol{\beta}$.

The RE estimator is fully efficient under the RE model, though the efficiency gain compared to pooled OLS need not be great. It is inconsistent, however, if the fixed effects model is the correct model.

21.2.3. Panel-Robust Statistical Inference

The various panel models include error terms denoted u_{it}, ε_{it}, and α_i. In many microeconometrics applications it is reasonable to assume independence over i. However, the errors are potentially (1) **serially correlated** (i.e., correlated over t for given i) and/or (2) **heteroskedastic**. Valid statistical inference requires controlling for both of these factors.

The White heteroskedastic consistent estimator of Section 4.4.5 is easily extended to short panels since for the ith observation the error variance matrix is of finite dimension $T \times T$ while $N \to \infty$. Thus panel-robust standard errors can be obtained without assuming specific functional forms for either within-individual error correlation or heteroskedasticity. More efficient estimators using GMM are deferred to Section 22.2.3.

It is crucial to note that frequently the panel commands in many computer packages calculate default standard errors assuming iid model errors, leading to erroneous inference. In particular, for pooled OLS regression of y_{it} on $\mathbf{x}_{it}$ without any control for individual effects it is very likely that $\mathrm{Cov}[u_{it}, u_{is}] > 0$ for $t \neq s$. Ignoring this serial correlation can lead to greatly underestimated standard errors and over-estimated t-statistics, as demonstrated in the Section 21.3 data example and shown algebraically in Section 21.5.4. Once fixed or random individual-specific effects are included the serial correlation in errors can be greatly reduced, but it may not be completely eliminated.

Additionally, one may need to control for potential heteroskedasticity as is routinely done for cross-section data.

Panel-Robust Sandwich Standard Errors

The panel estimators of Section 21.2.2 can be obtained by OLS estimation of $\boldsymbol{\theta}$ in the pooled regression

$$\widetilde{y}_{it} = \widetilde{\mathbf{w}}_{it}'\boldsymbol{\theta} + \widetilde{u}_{it}, \tag{21.12}$$

where different panel estimators correspond to different transformations $\widetilde{y}_{it}$, $\widetilde{\mathbf{w}}_{it}$, and $\widetilde{u}_{it}$ of y_{it}, $\mathbf{w}_{it}' = [1\ \mathbf{x}_{it}']$, and u_{it}. The key is that $\widetilde{y}_{it}$ is a known function of only $y_{i1}, \ldots, y_{iT}$, and similarly for $\widetilde{\mathbf{w}}_{it}$ and $\widetilde{u}_{it}$.

In the simplest case of pooled OLS, no transformation is necessary and $\boldsymbol{\theta} = [\alpha\ \boldsymbol{\beta}']'$. For the within estimator $\widetilde{y}_{it} = y_{it} - \bar{y}_i$, $\widetilde{\mathbf{w}}_{it} = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$, where only time-varying regressors appear, and $\boldsymbol{\theta}$ equals the coefficients of the time-varying regressors. For first-differences estimation $\widetilde{y}_{it} = y_{it} - y_{i,t-1}$, $\widetilde{\mathbf{w}}_{it} = (\mathbf{x}_{it} - \mathbf{x}_{i,t-1})$ and again only coefffi-cients of time-varying regressors are identified. For random effects $\widetilde{y}_{it} = y_{it} - \widehat{\lambda}\bar{y}_i$ and $\widetilde{\mathbf{w}}_{it}' = (\mathbf{w}_{it} - \widehat{\lambda}\bar{\mathbf{w}}_i)$ and $\boldsymbol{\theta} = [\alpha\ \boldsymbol{\beta}']'$. Such transformations can induce serial correlation even if underlying errors are uncorrelated.

It is convenient to stack observations over time periods for a given individual, leading to

$$\widetilde{\mathbf{y}}_i = \widetilde{\mathbf{W}}_i\boldsymbol{\theta} + \widetilde{\mathbf{u}}_i,$$

where $\widetilde{\mathbf{y}}_i$ is a $T \times 1$ vector in the preceding examples, except for the first-differences model where it is $(T-1) \times 1$, and $\widetilde{\mathbf{W}}_i$ is a $T \times q$ matrix or, for the first-differences model, a $(T-1) \times q$ matrix. Further stacking over the N individuals yields

$$\widetilde{\mathbf{y}} = \widetilde{\mathbf{W}}\boldsymbol{\theta} + \widetilde{\mathbf{u}}.$$

Three representations of the OLS estimator are therefore

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{\text{OLS}} &= [\widetilde{\mathbf{W}}'\widetilde{\mathbf{W}}]^{-1}\widetilde{\mathbf{W}}'\widetilde{\mathbf{y}} \\ &= \left[\sum_{i=1}^{N}\widetilde{\mathbf{W}}_i{}'\widetilde{\mathbf{W}}_i\right]^{-1}\sum_{i}\widetilde{\mathbf{W}}_i{}'\widetilde{\mathbf{y}}_i \\ &= \left[\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{\mathbf{w}}_{it}\widetilde{\mathbf{w}}_{it}'\right]^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{\mathbf{w}}_{it}\widetilde{\mathbf{y}}_{it},\end{aligned}$$

where in the third expression the sum is from $t = 2$ to T in the case of the first-differences estimator. The most convenient representation to use varies with the context.

To consider consistency, note that if the model is correctly specified then the usual algebra yields $\widehat{\boldsymbol{\theta}}_{\text{OLS}} = \boldsymbol{\theta} + [\widetilde{\mathbf{W}}'\widetilde{\mathbf{W}}]^{-1}\widetilde{\mathbf{W}}'\widetilde{\mathbf{u}}$ or

$$\widehat{\boldsymbol{\theta}}_{\text{OLS}} = \boldsymbol{\theta} + \left[\sum_{i=1}^{N}\widetilde{\mathbf{W}}_i{}'\widetilde{\mathbf{W}}_i\right]^{-1}\sum_{i=1}^{N}\widetilde{\mathbf{W}}_i{}'\widetilde{\mathbf{u}}_i.$$

Given independence over i the essential condition for consistency is $\mathrm{E}[\widetilde{\mathbf{W}}_i{}'\widetilde{\mathbf{u}}_i] = \mathbf{0}$. This generally requires a stronger assumption than $\mathrm{E}[u_{it}|\mathbf{w}_{it}] = 0$. A sufficient assumption is that of strong exogeneity given in (21.4). See Chapter 22 for estimation under assumptions weaker than strong exogeneity that permit, for example, lagged dependent variables as regressors.

The asymptotic variance of $\widehat{\boldsymbol{\theta}}_{\mathrm{OLS}}$ is then

$$\mathrm{V}[\widehat{\boldsymbol{\theta}}_{\mathrm{OLS}}] = \left[\sum_{i=1}^{N} \widetilde{\mathbf{W}}_i{}'\widetilde{\mathbf{W}}_i\right]^{-1} \sum_{i=1}^{N} \widetilde{\mathbf{W}}_i{}'\mathrm{E}[\widetilde{\mathbf{u}}_i\widetilde{\mathbf{u}}_i'|\widetilde{\mathbf{W}}_i]\widetilde{\mathbf{W}}_i \left[\sum_{i=1}^{N} \widetilde{\mathbf{W}}_i{}'\widetilde{\mathbf{W}}_i\right]^{-1},$$

given independence of errors over i. Consistent estimation of $\mathrm{V}[\widehat{\boldsymbol{\theta}}_{\mathrm{OLS}}]$ in this panel setting is analogous to the cross-section problem of obtaining a consistent estimate of $\mathrm{V}[\widehat{\boldsymbol{\theta}}_{\mathrm{OLS}}]$ that is robust to heteroskedasticity of unknown form. The only complication is the appearance of a vector $\mathbf{u}_i$ rather than a scalar u_i, which poses no problem if the panel is short as then the dimension of $\mathbf{u}_i$ is finite.

This leads to a **panel-robust estimate** of the asymptotic variance matrix of the pooled OLS estimator, one that controls for both serial correlation and heteroskedasticity, given by

$$\widehat{\mathrm{V}}[\widehat{\boldsymbol{\theta}}_{\mathrm{OLS}}] = \left[\sum_{i=1}^{N} \widetilde{\mathbf{W}}_i{}'\widetilde{\mathbf{W}}_i\right]^{-1} \sum_{i=1}^{N} \widetilde{\mathbf{W}}_i{}'\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\widetilde{\mathbf{W}}_i \left[\sum_{i=1}^{N} \widetilde{\mathbf{W}}_i{}'\widetilde{\mathbf{W}}_i\right]^{-1}, \tag{21.13}$$

where $\widehat{\mathbf{u}}_i = \widehat{\widetilde{\mathbf{u}}}_i = \widetilde{\mathbf{y}}_i - \widetilde{\mathbf{W}}_i\widehat{\boldsymbol{\theta}}$. The estimator in (21.13) assumes independence over i and $N \to \infty$, the case for short panels, but otherwise permits $\mathrm{V}[u_{it}]$ and $\mathrm{Cov}[u_{it}, u_{is}]$ to vary with i, t, and s. An equivalent expression is

$$\widehat{\mathrm{V}}[\widehat{\boldsymbol{\theta}}_{\mathrm{OLS}}] = \left[\sum_{i=1}^{N}\sum_{t=1}^{T} \widetilde{\mathbf{w}}_{it}\widetilde{\mathbf{w}}_{it}'\right]^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} \widetilde{\mathbf{w}}_{it}\widetilde{\mathbf{w}}_{is}'\widehat{u}_{it}\widehat{u}_{is} \left[\sum_{i=1}^{N}\sum_{t=1}^{T} \widetilde{\mathbf{w}}_{it}\widetilde{\mathbf{w}}_{it}'\right]^{-1},$$

where $\widehat{u}_{it} = \widetilde{y}_{it} - \widetilde{\mathbf{w}}_{it}'\widehat{\boldsymbol{\theta}}$. This estimator was proposed by Arellano (1987) for the fixed effects estimator.

Panel-robust standard errors based on (21.13) can be computed by use of a regular OLS command, if the command has a **cluster-robust** standard error option (see Section 24.5.2). Since the clustering here is on the individual one selects the identifier for individual i as the **cluster variable**. This method was used to obtain the panel-robust standard errors given in Table 24.1.

The term "robust" standard error can cause confusion. A common error made in pooled regression is to estimate the OLS regression (21.12) using the standard robust standard error option (see Section 4.4.5). However, this only adjusts for heteroskedasticity, and in practice in a panel setting it is much more important to correct for the correlation in individual errors. Another common error, though one that has smaller impact, is to use cluster-robust standard errors that assume homoskedasticity so that $\mathrm{E}[\mathbf{u}_i\mathbf{u}_i']$ is constant over i.

Panel Bootstrap Standard Errors

The **bootstrap method** provides an alternative way to obtain panel-robust standard errors. The key assumption is that observations are independent over i, so one can do a bootstrap pairs procedure that resamples **with replacement over** i and uses all observed time periods for a given individual. For data $\{(\mathbf{y}_i, \mathbf{X}_i), i = 1, \ldots, N\}$ this yields B pseudo-samples and for each **pseudo-sample** one performs OLS regression of [illegible] on $\widetilde{\mathbf{w}}_{it}$, yielding B estimates $\widehat{\theta}_b, b = 1, \ldots, B$.

The **panel bootstrap estimate** of the variance matrix is then

$$\widehat{\mathrm{V}}_{\mathrm{Boot}}[\widehat{\theta}] = \frac{1}{B-1} \sum_{b=1}^{B} \left(\widehat{\theta}_b - \overline{\widehat{\theta}}\right)\left(\widehat{\theta}_b - \overline{\widehat{\theta}}\right)', \qquad (21.1[illegible])$$

where $\overline{\widehat{\theta}} = B^{-1} \sum_b \widehat{\theta}_b$. This bootstrap provides **no asymptotic refinement** (see Section 11.2.2). Given independence over i the estimate is consistent as $N \to \infty$. It is asymptotically equivalent to the estimate (21.13), just as in the cross-section case bootstrap pairs are asymptotically equivalent to White's heteroskedastic consistent estimate. This bootstrap does not offer an asymptotic refinement though bootstrap with asymptotic refinement is possible (see Section 11.6.2).

This bootstrap method can be applied to any panel estimator that relies on independence over i and $N \to \infty$, including the pooled feasible GLS estimators of Section 21.5.2 for short panels. The key is to resample over i only, and not over both i and t.

Discussion

The importance of correcting standard errors for serial correlation in errors at the individual level cannot be overemphasized. Computer packages currently do not automatically do this. Bertrand, Duflo, and Mullainathan (2004) illustrate the resulting downward bias in standard error computation, in the context of difference-in-differences estimation (see Section 22.6). They find that the panel-robust and panel bootstrap methods work well, even though in their application with state-year data N (the number of states) is relatively small whereas the asymptotic theory uses $N \to \infty$.

The following example (see Table 21.2) also shows the importance of correcting standard errors for any error serial correlation and autocorrelation.

21.3. Linear Panel Example: Hours and Wages

An important issue in labor economics is the responsiveness of labor supply to wages. The standard textbook model of labor supply suggests that for people already working the effect of a wage increase on labor supply is ambiguous, with an income effect pushing in the direction of less work offsetting a substitution effect in the direction of more work.

Cross-section analysis for adult males finds a relatively small positive response to hours worked. However, it is possible that this association is spurious, merely reflecting a greater unobserved desire to work being positively associated with higher wages.

Panel data analysis can control for this, under the assumption that the unobserved desire to work is time-invariant. For example, the within estimator does so by measuring the extent to which an individual works above-average (or below-average) hours in periods with above-average (or below-average) wages.

The data on 532 males for each of the 10 years from 1979 to 1988 come from Ziliak (1997). The variable of interest is lnhrs, the natural logarithm of annual hours worked. The single explanatory variable is lnwg, the natural logarithm of hourly wage. We consider the regression model

$$\mathrm{lnhrs}_{it} = \alpha_i + \beta \mathrm{lnwg}_{it} + \varepsilon_{it},$$

where the individual-specific effect α_i is simplified to α in some models and β measures the wage elasticity of labor supply. The error term ε_{it} is assumed to be independent over i, but it may be correlated over t for given i. As noted we expect β, the labor supply elasticity, to be small and positive.

Ziliak (1997) additionally included a quadratic in age, number of children, and an indicator variable for bad health. These regressors and year dummies make relatively small difference to the estimate of β and its standard error, and for simplicity they are omitted here. In Chapter 22 we consider more general models that permit lnwg to be endogenous and permit lags of lnhrs to appear as a regressor.

21.3.1. Data Summary

For the 5,320 observations, the sample means of lnhrs and lnwg are respectively 7.66 and 2.61, implying geometric means of 2,120 hours and \$13.60 per hour. The sample standard deviations are respectively 0.29 and 0.43, indicating considerably greater variability in percentage terms in wages rather than hours.

For panel data it is useful to know whether variability is mostly across individuals or across time. The total variation of a series x_{it} around its grand mean $\bar{x}$ can be decomposed as

$$\begin{aligned}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x})^2 &= \sum_{i=1}^{N}\sum_{t=1}^{T}[(x_{it}-\bar{x}_i)+(\bar{x}_i-\bar{x})]^2 \\ &= \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2+\sum_{i=1}^{N}\sum_{t=1}^{T}(\bar{x}_i-\bar{x})^2,\end{aligned}$$

as the cross-product term sums to zero. In words, the total sum of squares equals the **within sum of squares** plus the **between sum of squares**. This leads to **within standard deviation** s_{W} and **between standard deviation** s_{B}, where

$$s_{\mathrm{W}}^2 = \frac{1}{NT-N}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2$$

and

$$s_{\mathrm{B}}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\bar{x}_i-\bar{x})^2.$$

Table 21.2. *Hours and Wages: Standard Linear Panel Model Estimators*[a]

	POLS	Between	Within	First Diff	RE–GLS	RE–MLE
α	7.442	7.483	7.220	.001	7.346	7.346
β	.083	.067	.168	.109	.119	.120
Robust se[b]	(.030)	(.024)	(.085)	(.084)	(.051)	(.052)
Boot se	[.030]	[.019]	[.084]	[.083]	[.056]	[.058]
Default se	{.009}	{.020}	{.019}	{.021}	{.014}	{.014}
R^2	.015	.021	.016	.008	.014	.014
RMSE	.283	.177	.233	.296	.233	.233
RSS	427.225	0.363	259.398	417.944	288.860	288.612
TSS	433.831	17.015	263.677	420.223	293.023	292.773
σ_α	.000		.181		.161	.162
σ_ε	.283		.232		.233	.233
λ	0.000	–	1.000	–	.585	.586
N	5320	532	5320	4788	5320	5320

[a] Shown are pooled OLS (POLS), between, within, first-differences, random effects (RE) GLS and MLE linear panel regression of lnhrs on lnwg. Standard errors for the slope coefficients are panel robust in parentheses, panel bootstrap in square brackets, and default estimates that assume iid errors in curly braces. The R^2, root mean square error (RMSE), residual sum of squares (RSS), total sum of squares (TSS), and sample size come from the appropriate regression given in Section 21.2. The parameter λ is defined after (21.11).

[b] se, standard error.

The within and between sample standard deviations are, respectively, 0.22 and 0.18 for lnhrs and 0.19 and 0.39 for lnwg. The larger total variation in wages compared to hours is therefore due to between individual variation being much higher for wages. Within individuals the variation is actually somewhat smaller for wages than it is for hours.

21.3.2. Comparison of Panel Data Estimators

Table 21.2 summarizes results from application of the standard panel estimators defined in Section 21.2.2 to these data, along with three different estimates of the standard errors. As detailed in the following, statistical inference should use either the panel-robust standard error or the panel bootstrap standard error.

Slope Parameter Estimates

The estimate of the slope parameter β differs across the different estimation methods. The between estimate that uses only cross-section variation is less than the pooled OLS estimate. The within or fixed effects estimate of 0.168 is much higher than the pooled OLS estimate of 0.083 and is borderline statistically significant using a two-tailed test at 5% and standard error estimate of 0.084 or 0.085. The first-differences estimate of 0.109 is also higher than that of pooled OLS but is considerably less than the within estimate, which also uses only time-series variation. The RE estimates of 0.119 or 0.120 lie between the between and within estimates. This is expected, as RE estimates can be shown to be a **weighted average of between and within estimates**. The two RE estimates are very close to each other as here the estimates of the variances σ_α^2 and σ_ε^2 are similar, leading to similar values $\widehat{\lambda} = 0.585$ and $\widehat{\lambda} = 0.586$ in the regression (21.10). The RE estimates are surprisingly less efficient than the pooled OLS estimates, a sign that the RE model fails to model the error correlation well.

Which estimates are preferred? The within and first-difference estimators are consistent under all models (pooled, RE, and FE) whereas the other estimators are inconsistent under the fixed effects model. The most robust estimates are therefore the within or first-differences estimates of 0.168 or 0.109.

There is, however, an efficiency loss in using these more robust estimators, with standard errors of 0.83 to 0.85 that are much larger than those from pooled OLS and RE estimates. A formal Hausman test (see Section 21.4.3 for details and discussion) can be used to test whether or not the individual effects are fixed. Given the relative imprecision of estimation in this example, the Hausman test does not reject the null hypothesis of random effects, despite the large difference between FE and RE estimates. So the more efficient random effects estimates could be used here. Another advantage of random effects estimation is that it permits estimation of the coefficients of time-invariant estimators.

Standard Error Estimation

We now turn to comparison of the standard error estimates. From Section 21.2.3, inference should be based on panel-robust standard errors that permit errors to be correlated over time for a given individual and to have variances and covariances that differ across individuals. Also, as detailed in later sections, the standard errors for estimators based on deviations from means, such as (21.8) and (21.10), need to account for loss of $N + K$ rather than K degrees of freedom.

The first standard error estimate is computed by the panel-robust method given in (21.13), and the second is computed by the panel bootstrap given in (21.14) with 500 replications. For brevity these estimates are called panel robust, though they are additionally robust to heteroskedasticity. The two estimates are very close, aside from the random effects models where the panel-robust standard errors are underestimated because they are computed for the regression (21.10), which ignores estimation error in $\widehat{\lambda}$.

The third standard error estimate is the standard default computer output that is based on the assumption of iid errors. In this example the correctly estimated standard errors are a remarkable three to four times as large as the default standard errors. The one exception is the between estimator, an estimator with standard errors that need only correction for heteroskedasticity since it uses only cross-section variation.

For example, for the pooled OLS estimator of β the default standard error is 0.09, leading to incorrect t-statistic of 9.07. The panel-robust standard error is a much larger 0.30, leading to correct t-statistic of a much smaller 2.83. Default standard errors assume independence of model errors over t for given i when in practice they are likely to be positively correlated. This erroneous assumption overestimates the benefit of additional time periods, leading to downward bias in standard errors (see Section 21.5.4). Additionally, ignoring heteroskedasticity in errors also leads to bias, though this bias could be in either direction. For these data a failure to control for heteroskedasticity also imparts a large downward bias: The standard error of $\widehat{\beta}_{\text{POLS}}$ controlling for heteroskedasticity, but not for correlation over t for given i, is 0.020. For other data, correction for heteroskedasticity is usually much less important than the correction for panel correlation.

For the within and between estimators inclusion of the term α_i should control for some of the correlation in the error across time for a given individual. For these data, however, the differences between panel-robust and nonrobust standard errors remain large, in part because of failure to additionally control for heteroskedasticity.

Clearly panel-robust standard errors should be used.

21.3.3. Graphical Analysis

It is insightful to perform a graphical comparison of overall, between, and fixed effects (within or first-differences) regressions. Such plots are rarely performed in panel data regression, but they are easily applied here as there is only one regressor.

All plots include a nonparametric regression curve using the Lowess smoother (see Section 9.6.2) and a linear regression curve that corresponds to the estimates given in Table 21.2.

Figure 21.1 plots lnhrs against lnwg for all firms in all years (5,320 observations). The plot suggests a positive relationship, roughly linear except at the extreme ends, and from Table 21.2 the line has slope 0.083 with a low $R^2 = 0.015$.

The between estimator (21.7) regresses $\bar{y}_i$ on $\bar{x}_i$. The corresponding plot for the lnhrs–lnwg data is given in Figure 21.2 and again shows a positive relationship.

The within or fixed effects estimator (21.8) regresses $(y_{it} - \bar{y}_i)$ on $(x_{it} - \bar{x}_i)$. Figure 21.3 gives the related plot of $(y_{it} - \bar{y}_i + \bar{\bar{y}})$ on $(x_{it} - \bar{x}_i + \bar{\bar{x}})$, where $\bar{\bar{y}} = N^{-1}\sum_i \bar{y}_i$ and $\bar{\bar{x}} = N^{-1}\sum_i \bar{x}_i$ are the grand means of y and x. Comparison with Figure 21.1 shows that differencing the individual mean leads to a considerable decrease in the range of variability in lnwg, with less of a decrease in the variability of lnhrs.

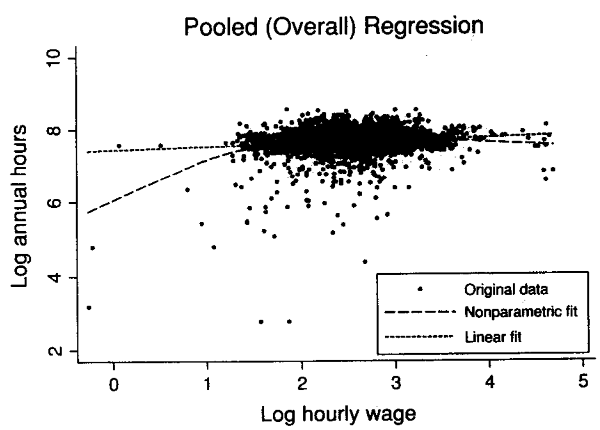

Figure 21.1: Hours and wages: pooled (overall) regression. Natural logarithm of annual hours worked plotted against natural logarithm of hourly wage. Data for 532 U.S. males for each of the ten years 1979–88.

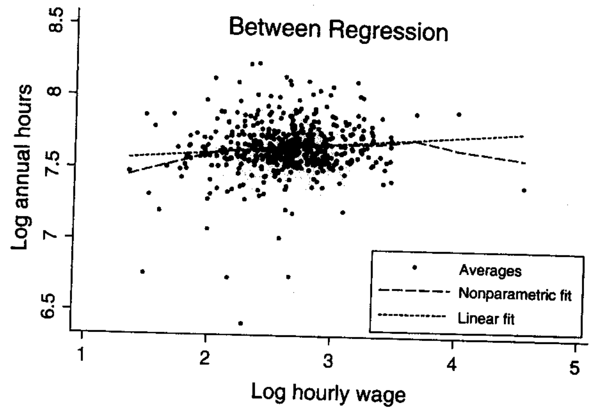

Figure 21.2: Hours and wages: between regression. Ten-year average of log hours plotted against ten-year average of log wage for 532 men. Same sample as Figure 21.1.

The slope does appear steeper than that for pooled OLS, and from Table 21.2 the slope increased from 0.083 to 0.168.

The first-differences estimator (21.9) regresses $(y_{it} - y_{i,t-1})$ on $(x_{it} - x_{i,t-1})$. The corresponding plot for the lnhrs – lnwg data is given in Figure 21.4. The figure is qualitatively similar to Figure 21.3.

The conclusion of the preceding analysis is that there is greater response to wage changes using time-series variation than using cross-section variation.

21.3.4. Residual Analysis

It is instructive to consider the autocorrelation patterns of the data and of residuals. For example, for residuals $\widehat{u}_{it} = y_{it} - \widehat{y}_{it}$ the autocorrelation between period s and period t is calculated as $\widehat{\rho}_{st} = c_{st}/\sqrt{c_{ss}c_{tt}}$, $s, t = 1, \ldots, T$, where the covariance estimate $c_{st} = (N-1)^{-1}\sum_i(\widehat{u}_{it} - \overline{\widehat{u}}_t)(\widehat{u}_{is} - \overline{\widehat{u}}_s)$ and $\overline{\widehat{u}}_t = N^{-1}\sum_i \widehat{u}_{it}$.

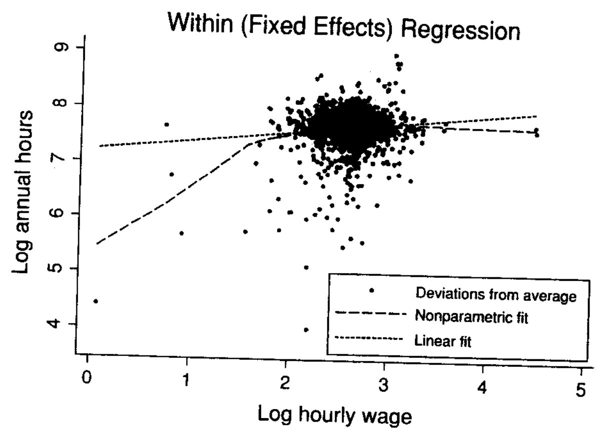

Figure 21.3: Hours and wages: within (fixed effects) regression. Deviation of log hours from ten-year average plotted against deviation of log wage from ten-year average using ten years of data for 532 men. Same sample as Figure 21.1.

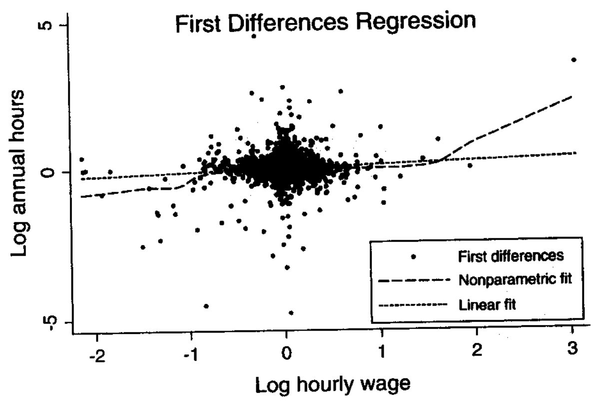

Figure 21.4: Hours and wages: first differences regression. First difference of log hours plotted against first difference of log wage using ten years of data for 532 men. Same sample as Figure 21.1.

Table 21.3 gives the residual autocorrelations after pooled OLS regression of lnhrs on lnwg. The autocorrelations generally lie between 0.2 and 0.4 for data two to nine periods apart. The decay rate is very slow, and the autocorrelations appear closer to a random effects model that assumes that $\text{Cor}[u_{it}, u_{is}]$ is constant for $t \neq s$ than to an AR(1) model that has exponential decay.

The autocorrelations for lnhrs before regression are very similar to those given in Table 21.3, since $\widehat{u}_{it} \simeq y_{it}$ as evident from the poor explanatory power of pooled OLS with $R^2 = 0.015$. The autocorrelations for the regressor lnwg, not tabulated here, are much higher, ranging from approximately 0.9 at one lag, to 0.7 at nine lags.

The correlations of the residuals from the within regression are given in Table 21.4. If the original errors ε_{it} in (21.3) are iid then it can be shown that the transformed errors $\varepsilon_{it} - \bar{\varepsilon}_i$ have autocorrelations at all lags equal to $-1/(T-1) = -0.11$. There is some departure from this here, particularly for the first lag, which is always positive.

Table 21.3. *Hours and Wages: Autocorrelations of Pooled OLS Residuals*[a]

	u79	u80	u81	u82	u83	u84	u85	u86	u87	u88
upols79	1.00									
upols80	.33	1.00								
upols81	.44	.40	1.00							
upols82	.30	.31	.57	1.00						
upols83	.21	.23	.37	.47	1.00					
upols84	.20	.23	.32	.34	.64	1.00				
upols85	.24	.32	.41	.35	.39	.58	1.00			
upols86	.20	.19	.28	.25	.31	.35	.40	1.00		
upols87	.20	.32	.33	.29	.31	.34	.39	.35	1.00	
upols88	.16	.25	.30	.26	.21	.25	.34	.55	.53	1.00

[a] Note: Autocorrelations of residuals are from pooled OLS regression of lnhrs on lnwg for 532 men in 10 years. The autocorrelations die slowly.

Table 21.4. *Hours and Wages: Autocorrelations of Within Regression Residuals*[a]

	u79	u80	u81	u82	u83	u84	u85	u86	u87	u88
ufe79	1.00									
ufe80	.10	1.00								
ufe81	.21	.08	1.00							
ufe82	.00	−.04	.26	1.00						
ufe83	−.26	−.27	−.21	.01	1.00					
ufe84	−.26	−.27	−.30	−.20	.32	1.00				
ufe85	−.18	−.10	−.11	−.17	−.16	.17	1.00			
ufe86	−.19	−.25	−.26	−.27	−.17	−.14	−.08	1.00		
ufe87	−.15	−.05	−.16	−.20	−.24	−.21	−.09	−.09	1.00	
ufe88	−.17	−.11	−.14	−.18	−.38	−.31	.13	.24	.24	1.00

[a] Autocorrelations of residuals are from within (fixed effects) regression of lnhrs on lnwg for 532 men in 10 years.

The correlations of the residuals from random effects regression are quite similar to those for fixed effects given in Table 21.4. The correlations of residuals from first-differences regression are qualitatively similar to the theoretical result that if the original errors ε_{it} in (21.3) are iid then the transformed errors $\varepsilon_{it} - \varepsilon_{it-1}$ have autocorrelations of 0.5 at lag one and 0 at other lags.

21.4. Fixed Effects versus Random Effects Models

The fixed effects model has the attraction of allowing one to use panel data to establish causation under weaker assumptions (presented in Section 21.4.1) than those needed to establish causation with cross-section data or with panel data models without fixed effects, such as pooled models and random effects models.

In some studies causation is clear, so random effects may be appropriate. For example, in a controlled experiment such as crop yield from different amounts of fertilizers applied to different fields the causation is clear. In other cases it may be sufficient to use a random effects analysis to measure the extent of correlation, with determination of causation left to further research taking other approaches. The effect of smoking on lung cancer is an example. Economists are unusual in instead preferring a fixed effects approach, however, because of a desire to measure **causation** in spite of reliance on observational data.

The fixed effects model has several practical weaknesses. Estimation of the coefficient of any time-invariant regressor, such as an indicator variable for gender, is not possible as it is absorbed into the individual-specific effect. Coefficients of time-varying regressors are estimable, but these estimates may be very imprecise if most of the variation in a regressor is cross sectional rather than over time. Prediction of the conditional mean is not possible. Instead, only changes in the conditional mean caused by changes in time-varying regressors can be predicted. Even coefficients of time-varying regressors may be difficult or theoretically impossible to identify in nonlinear models with fixed effects (see Chapter 23). For these reasons economists also use random effects models, even if causal interpretation may then be unwarranted.

21.4.1. Fixed Effects Example

Consider the effect of computer use on wage. Several cross-sectional studies, most notably those by Krueger (1993) and DiNardo and Pischke (1997), find that computer use in a job is associated with substantially higher wages, even after controlling for many determinants of the wage such as education, age, gender, industry, and occupation. As emphasized by DiNardo and Pischke (1997) this does not necessarily imply causation, if regressors are correlated with the error term owing to endogeneity or omitted variables.

Specifically, we suppose that in the cross section

$$y_i = \mathbf{x}_i'\beta + \alpha_i + \varepsilon_i,$$

where y is the natural logarithm of wage, $\mathbf{x}$ is a vector of individual characteristics including an indicator variable for computer use at work, and ε is an error that is assumed to be independent of $\mathbf{x}$. The complication is the addition of the unobserved variable α, which is assumed to be correlated with computer use at work, and hence with the observed regressors $\mathbf{x}$, even though the components of $\mathbf{x}$ other than computer use, such as occupation and education, may partly control for computer use at work. Regression of y on $\mathbf{x}$ leads to **omitted variables bias** leading to inconsistent estimates of β as the combined error $(\alpha + \varepsilon)$ is correlated with $\mathbf{x}$.

Panel data offer a way around this problem, if we assume that the unobserved variable α_i is time-invariant. Then

$$y_{it} = \mathbf{x}_{it}'\beta + \alpha_i + \varepsilon_{it},$$

where again ε is uncorrelated with $\mathbf{x}$ and α is correlated with $\mathbf{x}$. Differencing eliminates α_i (see Section 21.2.2), permitting consistent estimation of β. For the computer use example, the causative effect of computer use on wages is then measured by the association between individual changes in wages and individual movements to or from a job with a computer. Haisken-DeNew and Schmidt (1999) found no effect using German panel data.

This fixed effects panel approach permits determination of causation under weaker assumptions than those of cross-section analysis, but it still requires assumptions. The key assumption is that the unobservables α_i are time-invariant, rather than being of more general form α_{it}. In the computer use example it is being assumed that an individual's propensity to have a job with a computer may be endogenous, but the unobserved component of the effect of this propensity α_i on wage is constant over time once we control for observables $\mathbf{x}_{it}$. Essentially the particular time periods in which an individual's job does or does not involve use of a computer are assumed to be purely random, once we control for time-invariant unobservable α_i and observable $\mathbf{x}_{it}$.

A random effects or pooled panel approach does not have similar properties. It instead assumes away the original concern that α is correlated with $\mathbf{x}$, since it assumes that α is iid $[0, \sigma^2]$ and hence is uncorrelated with $\mathbf{x}$. This leads to inconsistent parameter estimates if in fact α is correlated with $\mathbf{x}$, whereas the fixed effects estimator is consistent if α is correlated with $\mathbf{x}$, provided α is time-invariant.

21.4.2. Conditional versus Marginal Analysis

Fixed effects estimation is a **conditional analysis**, measuring the effect of $\mathbf{x}_{it}$ on y_{it} controlling for the individual effect α_i. Prediction is possible only for individuals in the particular sample being used, and even then it is only possible if the panel is long enough so that α_i can be consistently estimated. Random effects estimation is instead an example of **marginal analysis** or **population-averaged analysis**, as the individual effects are integrated out as iid random variables. The random effects estimators can be applied outside the sample.

If the true model is a random effects model, then whether to perform a conditional or marginal analysis will vary with the application. If analysis is for a random sample of countries then one uses random effects, but if one is intrinsically interested in the particular countries in the sample then one does fixed effects estimation even though this can entail a loss of efficiency.

If the true model instead has individual-specific effects correlated with regressors, however, then a random effects analysis is no longer meaningful as the random effects estimator is inconsistent. Instead, alternative estimators such as the fixed effects and first-differences estimators are necessary. Because of the desire to determine causation microeconomic applications emphasize these latter estimators.

21.4.3. Hausman Test

If individual effects are fixed the within estimator $\widehat{\beta}_{\mathrm{W}}$ is consistent whereas the random effects estimator $\widetilde{\beta}_{\mathrm{RE}}$ is inconsistent. Here β refers to the vector of coefficients of just the time-varying regressors. One can therefore test whether fixed effects are present by using a Hausman test of whether there is a statistically significant difference between these estimators. Alternatively, any other pair of estimators with similar properties, such as first differences versus pooled OLS, can be used.

A large value of the Hausman test statistic leads to rejection of the null hypothesis that the individual-specific effects are uncorrelated with regressors and to the conclusion that fixed effects are present. It may still be possible to avoid using a fixed effects model. If regressors are correlated with individual-specific effects caused by omitted variables, then one can add further regressors, either time varying or time-invariant, and again perform a Hausman test in this larger model to see whether fixed effects are still necessary. Even if such correlation persists it may be possible to estimate a random effects model using instrumental variables methods (see Sections 22.4.3–22.4.4).

Computation When RE Is Fully Efficient

We begin by assuming that the true model is the random effects model (21.3) with α_i iid $[0, \sigma_\alpha^2]$ uncorrelated with regressors and error ε_{it} iid $[0, \sigma_\varepsilon^2]$.

Then the estimator $\widetilde{\boldsymbol{\beta}}_{\text{RE}}$ is fully efficient, so from Section 8.3 the **Hausman test** statistic simplifies to

$$\text{H} = (\widetilde{\boldsymbol{\beta}}_{1,\text{RE}} - \widehat{\boldsymbol{\beta}}_{1,\text{W}})' \left[\widehat{\text{V}}[\widehat{\boldsymbol{\beta}}_{1,\text{W}}] - \widehat{\text{V}}[\widetilde{\boldsymbol{\beta}}_{1,\text{RE}}]\right]^{-1} (\widetilde{\boldsymbol{\beta}}_{1,\text{RE}} - \widehat{\boldsymbol{\beta}}_{1,\text{W}}),$$

where $\boldsymbol{\beta}_1$ denotes the subcomponent of $\boldsymbol{\beta}$ corresponding to time-varying regressors since only that component can be estimated by the within estimator. This test stastistic is asymptotically $\chi^2(\dim[\boldsymbol{\beta}_1])$ distributed under the null hypothesis.

Hausman (1978) showed that an asymptotically equivalent version of this test is to perform a Wald test of $\boldsymbol{\gamma} = \mathbf{0}$ in the auxiliary OLS regression,

$$y_{it} - \widehat{\lambda}\bar{y}_i = (1 - \widehat{\lambda})\mu + (\mathbf{x}_{1it} - \widehat{\lambda}\bar{\mathbf{x}}_{1i})'\boldsymbol{\beta}_1 + (\mathbf{x}_{1it} - \bar{\mathbf{x}}_{1i})'\boldsymbol{\gamma} + v_{it}, \qquad (21.15)$$

where $\mathbf{x}_{1it}$ denotes the time-varying regressors and $\widehat{\lambda}$ is defined in (21.11) and only the time-varying regressors are used. This algebraic result can be interpreted as follows. The individual-specific effects model (21.10) implies that $v_{it} = (1 - \widehat{\lambda})\alpha_i + (\varepsilon_{it} - \widehat{\lambda}\bar{\varepsilon}_i)$. The random effects estimator is actually obtained by OLS estimation of (21.15) with $\boldsymbol{\gamma} = \mathbf{0}$ (see (21.10)). If instead the fixed effects specification is valid then the error v_{it} will be correlated with the regressors, via correlation of α_i with regressors. This correlation leads to additional functions of the regressors, such as $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$, being statistically significant variables in (21.15).

Computation When RE Is Not Fully Efficient

The simple form of the Hausman test is invalid if α_i or ε_{it} are not iid, which is more than likely given heteroskedasticity inherent in much microeconometrics data. Then the RE estimator is not fully efficient under the null hypothesis so the expression $\widehat{\text{V}}[\widehat{\boldsymbol{\beta}}_{\text{W}}] - \widehat{\text{V}}[\widehat{\boldsymbol{\beta}}_{\text{RE}}]$ in the formula for H needs to be replaced by the more general $\widehat{\text{V}}[\widetilde{\boldsymbol{\beta}}_{\text{RE}} - \widehat{\boldsymbol{\beta}}_{\text{W}}]$ (see Section 8.3).

For short panels this variance matrix can be consistently estimated by bootstrap resampling over i (see Section 21.2.3). Then a panel-robust Hausman test statistic is

$$\text{H}_{\text{Robust}} = (\widetilde{\boldsymbol{\beta}}_{1,\text{RE}} - \widehat{\boldsymbol{\beta}}_{1,\text{W}})' \left[\widehat{\text{V}}_{\text{Boot}}[\widetilde{\boldsymbol{\beta}}_{1,\text{RE}} - \widehat{\boldsymbol{\beta}}_{1,\text{W}}]\right]^{-1} (\widetilde{\boldsymbol{\beta}}_{1,\text{RE}} - \widehat{\boldsymbol{\beta}}_{1,\text{W}}), \qquad (21.16)$$

where

$$\widehat{\text{V}}_{\text{Boot}}[\widetilde{\boldsymbol{\beta}}_{1,\text{RE}} - \widehat{\boldsymbol{\beta}}_{1,\text{W}}] = \frac{1}{B-1}\sum_{b=1}^{B}\left(\widehat{\boldsymbol{\delta}}_b - \overline{\widehat{\boldsymbol{\delta}}}\right)\left(\widehat{\boldsymbol{\delta}}_b - \overline{\widehat{\boldsymbol{\delta}}}\right)',$$

b denotes the bth of B bootstrap replications (see Section 21.2.3), and $\widehat{\boldsymbol{\delta}} = \widetilde{\boldsymbol{\beta}}_{1,\text{RE}} - \widehat{\boldsymbol{\beta}}_{1,\text{W}}$. This test statistic can be applied to subcomponents of $\boldsymbol{\beta}_1$ and can use alternative estimators such as $\widetilde{\boldsymbol{\beta}}_{1,\text{POLS}}$ in place of $\widetilde{\boldsymbol{\beta}}_{1,\text{RE}}$ and $\widehat{\boldsymbol{\beta}}_{1,\text{FD}}$ in place of $\widehat{\boldsymbol{\beta}}_{1,\text{W}}$.

Alternatively, Wooldridge (2002) suggests estimating the auxiliary OLS regression (21.15) and testing $\boldsymbol{\gamma} = \mathbf{0}$ using panel-robust standard errors. If the effects are random, though not necessarily such that α_i and ε_{it} are iid, then $v_{it} = (1 - \widehat{\lambda})\alpha_i + (\varepsilon_{it} - \widehat{\lambda}\bar{\varepsilon}_i)$ is still uncorrelated with regressors though v_{it} is no longer asymptotically iid, so **cluster-robust standard errors** need to be used. If the effects are fixed then the error v_{it} is correlated with the regressors, leading to significance of additional functions of the regressors such as $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$. This robust version of the auxiliary regression for the Hausman test is preferred to one that assumes v_{it} is asymptotically iid, on the usual grounds of minimizing distributional assumptions. However, it is not clear whether this test actually coincides with the Hausman test when RE is inefficient.

Hausman Test Example

For the lnhrs–lnwg example estimates given in Table 21.2, a comparison of FE and RE estimates using the default standard errors yields $\text{H} \simeq (0.168 - 0.119)^2/(0.019^2 - 0.014^2)$. This leads to $\text{H} = 14 > \chi^2_{.05}(1) = 3.84$, so the random effects model is rejected.

This test is not appropriate, however. The statistic H is inflated because the usual standard errors in this example are greatly downward biased (see Section 21.3.2). Furthermore, this bias is a signal that the RE estimator is not fully efficient under H_0, so that the more general form of the Hausman test needs to be used.

The auxiliary regression (21.15) yields a panel-robust t-statistic for $\widehat{\gamma}$ of 1.28 and hence $\text{H}^* = 1.28^2 = 1.65$, leading to nonrejection of the random effects model at 5%. Even though the wage elasticity estimates differ by 0.049, the estimates are sufficiently imprecise that the difference is not statistically significant. Note that if the nonrobust t-statistic for $\widehat{\gamma}$ is used instead, then $t^2 = 13.69$, close to the previous incorrect Hausman test statistic.

21.4.4. Richer Models for Random Effects

The random effects model specifies that the random effect α_i is distributed independently of regressors. Richer models, closer in spirit to fixed effects models, relax this assumption.

Mundlak (1978) allowed individual effects in the panel model (21.3) to be determined by **time averages** of the regressors, so that $\alpha_i = \bar{\mathbf{x}}_i'\boldsymbol{\pi} + w_i$, where w_i is iid. Then efficient GLS estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\pi}$ in this expanded model leads to an estimator of $\boldsymbol{\beta}$ that equals the fixed effects estimator in model (21.3). By contrast the usual random effects estimator of $\boldsymbol{\beta}$ in model (21.3) that erroneously specifies iid random effects will be inconsistent.

Chamberlain (1982, 1984) considered an even richer model for the random effects, with $\alpha_i = \mathbf{x}_{1i}'\boldsymbol{\pi}_1 + \cdots + \mathbf{x}_{Ti}'\boldsymbol{\pi}_T + w_i$, a **weighted sum** of the regressors. He proposed estimation by minimum distance methods (see Section 22.2.7 for details), leading to an estimator of $\boldsymbol{\beta}$ that equals the fixed effects estimator.

More generally, mixed linear models and hierarchical linear models of Section 24.6 permit quite general models for random intercepts and also random slope parameters. Bayesian analysis of panel data also uses this framework. See Section 22.8 for details.

In linear models the fixed effects approach is used if the unobserved individual effect is correlated with regressors. In more complicated models, such as nonlinear models, fixed effects models are not always estimable and richer random effects models provide an alternative approach.

21.5. Pooled Models

The **pooled cross-section time-series model** or **constant-coefficients model** is

$$y_{it} = \alpha + \mathbf{x}_{it}{}'\beta + u_{it}. \tag{21.17}$$

In the statistics literature the model is called a **population-averaged model**, as there is no explicit model of y_{it} conditional on individual effects. Instead, any individual effects have implicitly been averaged out. The random effects model is a special case where the error u_{it} is equicorrelated over t for given i (see Section 21.2.1).

The main complication for statistical inference, assuming no fixed effects, is that the distribution of least-squares estimators of this model varies with the assumed distribution of u_{it}. In short panels, panel-robust standard errors can be obtained using (21.13).

Here we instead focus on GLS estimation using many of the different specifications, including equicorrelation, for the covariance structure of u_{it} over time and individuals that have been proposed in the literature.

Although we focus on pooled GLS estimation of (21.17), a model without individual-specific fixed effects, the methods of this section can be applied more generally to pooled GLS estimation of the transformed model (21.12) of Section 21.2.3.

21.5.1. Pooled OLS, FGLS, and WLS Estimators

It is convenient to use matrix notation. Combining observations over time for a given individual, define

$$\mathbf{y}_i = \mathbf{W}_i\delta + \mathbf{u}_i, \tag{21.18}$$

where $\delta = [\alpha \quad \beta']'$ is a $(K+1)\times 1$ parameter vector, $\mathbf{y}_i$ and $\mathbf{u}_i$ are $T \times 1$ vectors with tth entries y_{it} and u_{it}, respectively, and $\mathbf{W}_i$ is a $T \times (K+1)$ matrix with tth row $\mathbf{w}_{it}' = [1 \quad \mathbf{x}_{it}]'$. Stacking all individuals yields

$$\mathbf{y} = \mathbf{W}\delta + \mathbf{u}, \tag{21.19}$$

where $\mathbf{y}$ and $\mathbf{u}$ are $NT \times 1$ vectors, for example $\mathbf{y} = [\mathbf{y}_1' \ldots \mathbf{y}_N']'$, and $\mathbf{W}$ is an $NT \times (K+1)$ regressor matrix whose first column is a vector of ones. We assume that $\mathrm{E}[\mathbf{u}|\mathbf{W}] = \mathbf{0}$, so errors are strictly exogenous, and define $\boldsymbol{\Omega} = \mathrm{E}[\mathbf{uu}'|\mathbf{W}]$.

There are several possible least-squares estimators of this model, summarized in Table 21.5.

First, **pooled OLS** is consistent and asymptotically normal. However, in a panel setting it is unlikely that $\boldsymbol{\Omega} = \sigma^2\mathbf{I}_{NT}$, so OLS is inefficient except in some special cases such as when all regressors are time-invariant. More importantly, the usual OLS variance estimate of $\sigma^2(\mathbf{W}'\mathbf{W})^{-1}$ should not be used and a panel-robust estimate such as that in (21.13) needs to be used.

Second, **pooled feasible GLS** (PFGLS) is consistent and fully efficient if $\boldsymbol{\Omega}$ is correctly specified and $\widehat{\boldsymbol{\Omega}}$ is consistent for $\boldsymbol{\Omega}$. Some of the very large range of structures on u_{it} and hence $\boldsymbol{\Omega}$ that have been proposed in the panel literature and incorporated

Table 21.5. *Pooled Least-Squares Estimators and Their Asymptotic Variances*

Estimator	Formula[a]	Variance Matrix[b]
Pooled OLS: $\widehat{\delta}_{\text{POLS}}$	$(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y}$	$(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\widehat{\boldsymbol{\Omega}}\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}$
Pooled FGLS: $\widehat{\delta}_{\text{PFGLS}}$	$(\mathbf{W}'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{W})^{-1}\mathbf{W}'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{y}$	$(\mathbf{W}'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{W})^{-1}$
Pooled WLS: $\widehat{\delta}_{\text{PWLS}}$	$(\mathbf{W}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{W})^{-1}\mathbf{W}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{y}$	$(\mathbf{W}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{W})^{-1}\mathbf{W}'\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\Omega}}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{W}$ $\times(\mathbf{W}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{W})^{-1}$

[a] The formulas are for the model $\mathbf{y} = \mathbf{W}\delta + \mathbf{u}$ defined in (21.19) and error matrix Ω.
[b] For computation of $\widehat{\Omega}$ for the variance matrices of POLS and PWLS see the text; in those cases $\widehat{\Omega}$ need not be consistent for Ω. For pooled FGLS it is assumed that $\widehat{\Omega}$ is consistent for Ω.

into regression packages are given in Sections 21.5.2 and 21.5.3 for, respectively, short and long panels.

Third, the **pooled weighted LS** (PWLS) estimator guards against misspecification of $\boldsymbol{\Omega}$. It posits a **working matrix** $\boldsymbol{\Sigma}$ for the error variance matrix $\boldsymbol{\Omega}$ but then performs inference that is valid even if $\boldsymbol{\Sigma} \neq \boldsymbol{\Omega}$. Ordinary least squares is an example, with $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}_{NT}$, but other choices of $\boldsymbol{\Sigma}$ may improve efficiency.

Estimation of the variance matrix of the pooled OLS estimator requires an $\widehat{\boldsymbol{\Omega}}$ such that $(NT)^{-1}\mathbf{W}'\widehat{\boldsymbol{\Omega}}\mathbf{W}$ consistently estimates $(NT)^{-1}\mathbf{W}'\boldsymbol{\Omega}\mathbf{W}$.

For short panels this is possible by direct application of the results of Section 21.2.3. Estimation of the variance matrix of the pooled WLS estimator requires an $\widehat{\boldsymbol{\Omega}}$ such that $(NT)^{-1}\mathbf{W}'\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\Omega}}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{W}$ consistently estimates $(NT)^{-1}\mathbf{W}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega}\boldsymbol{\Sigma}^{-1}\mathbf{W}$. The panel-robust estimate for OLS given in (21.13) can be adapted to pooled WLS by replacing $\mathbf{W}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega}\boldsymbol{\Sigma}^{-1}\mathbf{W}$, or equivalently $\sum_i \mathbf{W}_i'\boldsymbol{\Sigma}_i^{-1}\mathrm{E}[\mathbf{u}_i\mathbf{u}_i'|\mathbf{W}_i]\boldsymbol{\Sigma}_i^{-1}\mathbf{W}_i$ given independence over i, by the quantity $\sum_i \mathbf{W}_i'\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\widehat{\boldsymbol{\Sigma}}_i^{-1}\mathbf{W}_i$, where $\widehat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{W}_i\widehat{\delta}$. Alternatively, a panel bootstrap can be used.

21.5.2. Error Variance Matrix for Short Panels

In short panels there are few time periods but many individuals, usually people or firms. It is assumed that errors are independent over individuals, so that $\mathrm{Cov}[u_{it}, u_{js}] = 0$, $i \neq j$. In such cases it is convenient to revert to summation notation. For example, the PFGLS estimator given in Table 21.5 becomes

$$\widehat{\beta}_{\text{PFGLS}} = \left[\sum_{i=1}^{N} \mathbf{W}_i'\widehat{\boldsymbol{\Omega}}_i^{-1}\mathbf{W}_i\right]^{-1} \sum_{i=1}^{N} \mathbf{W}_i'\widehat{\boldsymbol{\Omega}}_i^{-1}\mathbf{y}_i, \tag{21.20}$$

where $\widehat{\boldsymbol{\Omega}}_i$ is consistent for

$$\boldsymbol{\Omega}_i = \mathrm{E}[\mathbf{u}_i\mathbf{u}_i'|\mathbf{W}_i], \tag{21.21}$$

and $\boldsymbol{\Omega}_i$ is nondiagonal as errors for a given individual are likely to be correlated over time. Note that $\widehat{\boldsymbol{\Omega}}_i$ needs to come from estimation of a specified model for $\boldsymbol{\Omega}_i$, and we cannot use $\widehat{\boldsymbol{\Omega}}_i = \widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'$ (see the related discussion after equation (5.88)).

Equicorrelated Errors

The most commonly used error structure is the random effects model presented in Section 21.2.1. Then from (21.6) $\mathbf{\Omega}_i$ has common diagonal entries $\sigma_\alpha^2 + \sigma_\varepsilon^2$ and common off-diagonal entries σ_α^2. Equivalently, the errors are **equicorrelated**, with $\mathbf{\Omega}_i$ having common diagonal entries σ^2 and common off-diagonal entries $\rho\sigma^2$. Implementation of FGLS requires only estimation of σ_α^2 and σ_ε^2, or of σ^2 and ρ (see Sections 21.2.2 and 21.7).

ARMA Errors

An alternative error structure is to assume an ARMA error model. For example, an AR(1) error model specifies that $u_{it} = \rho u_{i,t-1} + \varepsilon_{it}$, where ε_{it} are iid. Then $\mathrm{Cov}[u_{it}, u_{is}] = \rho^{|t-s|}\sigma^2$. In this case the covariance between errors falls as the number of time periods between the errors increases. The RE model and an AR(1) error model are compared in Section 21.5.4.

Baltagi and Li (1991) combine the two error models to consider a random effects model with AR(1) errors. This can be easily generalized to the AR(p) case, and methods for moving average and **ARMA errors** (see Section 5.8.7) in random effects models have also been developed more recently. A summary is given in Baltagi (2001, Chapter 5).

Homoskedastic Errors with Unstructured Autocorrelation

For FGLS estimation in short panels there is actually no need to impose as much structure as that imposed by an RE model or an AR(1) error model, if the assumption is made that the $T \times T$ matrix $\mathbf{\Omega}_i$ is constant over i. Then there are "only" $T(T+1)/2$ covariance parameters to estimate. A consistent estimate of $\mathbf{\Omega}_i$ is then $\widehat{\mathbf{\Omega}}_i$ with (t, s)th entry $\widehat{\sigma}_{ts} = N^{-1}\sum_{i=1}^{N}\widehat{u}_{it}\widehat{u}_{is}$. The preceding models also assume homoskedasticity, but place additional structure on $\mathbf{\Omega}_i$.

Robust Inference

All of the preceding specifications assume that error covariances are the same across individuals, which rules out heteroskedasticity. Provided the panel is short one can nonetheless use the preceding restrictive error variance matrix models as the basis for pooled WLS estimation, but then obtain robust standard errors as discussed after Table 21.5. Alternatively, richer mixed models, presented in Chapter 22, can be estimated.

The assumption of independence over i is maintained throughout Chapters 21–23, though it can be relaxed even for small T provided structure can be placed on the correlation. An example is an explicit model for spatial correlation for panel data on regions such as states or countries, with correlations declining as physical distance between individual observations increases.

21.5.3. Error Variance Matrix for Long Panels

In **long panels** there are many time periods but relatively few individuals. Such data can arise in microeconometrics analysis if the individual observational unit is one of only a few regions, such as a state or country, or firms, but these are observed over enough time periods to base inference on the assumption that $T \to \infty$.

Correlation across time for a given individual can be introduced using an ARMA model for the errors, with the parameters of the ARMA model permitted to differ across individuals as now N is fixed and $T \to \infty$. For example, consider an AR(1) error with $u_{it} = \rho_i u_{i,t-1} + \varepsilon_{it}$, where $\varepsilon_{it} \sim [0, \sigma_i^2]$ is heteroskedastic and ρ_i also differs across individuals. Separate regressions of y_{it} on $\mathbf{w}_{it}$ with AR(1) errors for each individual using T time periods yields consistent estimates $\widehat{\rho}_i$ and $\widehat{\sigma}_i^2$, since $T \to \infty$. These can then be used for feasible GLS estimation of $\boldsymbol{\delta}$ using all NT observations. For details see Kmenta (1986). This model permits both heteroskedasticity across individuals and correlation over time for a given individual. Pesaran (2004) proposes a considerably richer model that is estimated by GLS.

For long panels it is possible to introduce correlation across individuals, so that $\mathrm{Cov}[u_{it}, u_{jt}] \neq 0$ for $i \neq j$, since N is fixed and asymptotic results rely on $T \to \infty$. In particular, one can perform pooled GLS estimation as done earlier, with the assumption of independence across individuals, but then calculate standard errors using the method of Newy and West (1987b), mentioned briefly in Section 6.4.4, that permits arbitrary cross-sectional dependence and serial dependence, provided the serial dependence dies away sufficiently fast. For details see Arellano (2003, p. 19).

Time-series considerations for panel data are discussed in more detail in Section 22.5 for models with lagged dependent variables as regressors.

21.5.4. The Impact of Autocorrelated Errors

Panel data regression models have errors that are usually autocorrelated over time for a given individual. If fixed effects are absent then pooled OLS regression gives consistent parameter estimates. However, the **error correlation** can lead to **large bias** in standard errors for pooled OLS if autocorrelation is ignored and to relatively small efficiency gains as the length of a panel is increased.

The analysis is particularly simple for estimation of the mean of y based on T observations for one individual (so $N = 1$) with equicorrelation. Then $y_t = \beta + u_t$, and the OLS estimator is the sample mean, so $\widehat{\beta} = \bar{y} = T^{-1}\sum_t y_t$. The OLS estimator has true variance $\mathrm{V}[\widehat{\beta}] = \mathrm{V}[\bar{y}] = T^{-2}\sum_t\sum_s \mathrm{Cov}[u_t, u_s]$. Assuming equicorrelation the double sum has T variances equal to σ^2 and $T(T-1)$ covariances all equal to $\rho\sigma^2$. Hence $\mathrm{V}[\bar{y}] = T^{-1}\sigma^2(1 + (T-1)\rho)$. Thus the iid result that $\mathrm{V}[\bar{y}] = T^{-1}\sigma^2$ needs to be modified by inflation by a multiple $(1 + \rho(T-1))$. In particular $\mathrm{V}[\bar{y}]$ approaches σ^2 as $\rho \to 1$.

Table 21.6 presents the impact of correlation on the variance of $\bar{y}$ for different values of T and ρ, where for simplicity we normalize $\sigma^2 = 1$. The precision of estimation falls considerably as ρ increases, and the estimate of $\mathrm{V}[\bar{y}]$ under the assumption of independence given in the first column (assuming σ^2 is known for simplicity) can

Table 21.6. *Variances of Pooled OLS Estimator with Equicorrelated Errors*[a]

T	$\rho = 0.0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$	$\rho = 1.0$
1	1.00	1.00	1.00	1.00	1.00	1.00
2	0.50	0.60	0.70	0.80	0.90	1.00
5	0.20	0.36	0.52	0.68	0.84	1.00
10	0.10	0.28	0.46	0.64	0.82	1.00

[a] Given are the variances of the pooled OLS estimator as the correlation ρ of equicorrelated errors increases, for an intercept-only model with error variance normalized to one assuming errors are correlated though homoskedastic.

greatly understate the true variance. Furthermore, for $\rho > 0$ the gain in precision due to increase in the number of time periods is much smaller than with independent data where a doubling of the number of time periods will halve estimator variance. For example, if $\rho = 0.4$ then with five time periods the estimator variance is only 0.52 times that with one period, instead of the much lower multiple of 0.2 with independent data. Moreover, a doubling from 5 to 10 time periods leads to only a small reduction in estimator variance from 0.52 to 0.46.

This result holds more generally for balanced panel regression with equicorrelated errors and regressors that are time-invariant, where the true variance of the OLS estimator is $(1 + \rho(T - 1))$ times that assuming independent errors (see Kloek, 1981). In practice time-varying regressors are also included and clear analytical results are more difficult to obtain. For regression with intercept and single time-varying regressor, Scott and Holt (1982) show that the variance of the slope coefficient is inflated by the multiple $(1 + \widehat{\rho}_x \rho(T - 1))$, where $\widehat{\rho}_x$ can be viewed as an estimate of the individual-specific autocorrelation in x. For panel data $\widehat{\rho}_x$ is often high so that there is still considerable inflation. These results also apply to other forms of clustered data and are presented in more detail in Section 24.5.2.

The preceding analysis assumes equicorrelated errors, a property of the RE model. If instead errors are AR(1) there is greater benefit from increasing panel length. Then $\mathrm{Cov}[u_t, u_s] = \rho^{|t-s|}\sigma^2$, so $\mathrm{V}[\bar{y}] = T^{-2}\sigma^2[T + 2\sum_{s=1}^{T-1}(T - s)\rho^s]$. For example, if $\rho = 0.8$ then $\mathrm{V}[\bar{y}] = 0.72\sigma^2$ for $T = 5$ and $0.54\sigma^2$ for $T = 10$, lower than the corresponding values from Table 21.6 of $0.84\sigma^2$ and $0.82\sigma^2$ for equicorrelation with $\rho = 0.8$, but still much higher than values of $0.2\sigma^2$ and $0.1\sigma^2$ for $\rho = 0.0$.

Microeconometricians gravitate to the RE model or equicorrelated error models for short panels as an outgrowth of the literature on clustered data presented in Chapter 24. For example, consider data on different siblings in a family for many families. Then it is natural to assume that correlations of unobservables across siblings in the same family are the same for different siblings pairs. For example, the correlation between the first and second siblings equals that between the first and third siblings. Those using long panel data instead often have a time-series background and naturally assume that correlation declines over time, leading to models such as an AR(1) error.

Determining which model of time-series correlation is more reasonable really depends on the data. Many short panels used in microeconomics applications yield pooled OLS residual autocorrelations that are qualitatively similar to those given in Table 21.3. These are closer to an RE model than an AR(1) model, though an ARMA(1,1) model may do well. Better still may be an RE model with AR(1) error. In all cases error correlation leads to a loss of information and the usual OLS standard errors understate the true standard errors. For short panels one can base inference on panel robust standard errors (see Section 21.2.3) that do not require specifying a model for the error correlation.

21.5.5. Hours and Wages Pooled GLS Example

A variety of pooled GLS estimates and associated default and robust standard errors of the model $y_{it} = \alpha_i + \beta x_{it} + u_{it}$ for the lnhrs on lnwg regression are given in Table 21.7. All assume the error u_{it} is independent over i and identically distributed over i, and then have different assumptions on correlation in u_{it} over t.

The first column of Table 21.7, for the pooled OLS estimator, repeats the first column of Table 21.2.

Pooled GLS estimates assuming equicorrelated errors are given in the second column of Table 21.7. These coincide with the RE–GLS column in Table 21.2, since the random effects model implies equicorrelated errors (see (21.6)).

Pooled GLS estimates assuming AR(1) errors, so that $u_{it} = \rho u_{it-1} + \varepsilon_{it}$ where ε_{it} is iid, are given in the third column of Table 21.7. The slope coefficient estimate is close to the pooled OLS estimate.

Pooled GLS estimates with no structure placed on error correlation aside from homoskedasticity, so that $\mathrm{Cov}[u_{it}, u_{is}] = \sigma_{ts}$, are given in the fourth column of Table 21.7. Then σ_{ts} is consistently estimated given small T by $\widehat{\sigma}_{ts} = N^{-1}\sum_{i=1}^{N}\widehat{u}_{it}\widehat{u}_{is}$ for all t and s. These are again close to the pooled OLS estimate.

It is clear from Table 21.7 that panel-robust standard errors should be used rather than the default standard errors, which here assume homoskedasticity and correctly-specified model for serial correlation.

Table 21.7. *Hours and Wages: Pooled OLS and GLS Estimates*[a]

Estimator	POLS	PFGLS		
Error correlation	None	Equi	AR1	General
α	7.442	7.346	7.440	7.426
β	.083	.120	.084	.091
Robust se	(.029)	(.052)	(.037)	(.050)
Boot se	[.032]	[.060]	[.050]	[-]
Default se	{.009}	{.014}	{.012}	{.014}

[a] Pooled OLS and GLS linear panel regression of lnhrs on lnwg for a short panel assuming independence and identical distribution over i and no fixed effects. Pooled GLS estimators assume equicorrelated or random effects errors (equi), AR(1) errors (AR1), or no structure on the correlations (general). Standard errors for the slope coefficients are panel robust in parentheses, panel bootstrap in square brackets, and usual default estimates that assume iid errors in curly braces.

21.6. Fixed Effects Model

The **fixed effects model** specifies

$$y_{it} = \alpha_i + \mathbf{x}_{it}'\beta + \varepsilon_{it}, \qquad (21.22)$$

where the individual-specific effects $\alpha_1, \ldots, \alpha_N$ measure unobserved heterogeneity that is possibly correlated with the regressors, $\mathbf{x}_{it}$ and β are $K \times 1$ vectors, and to begin with the errors ε_{it} are iid $[0, \sigma^2]$.

The challenge for estimation is the presence of the N individual-specific effects that increase in number as $N \to \infty$. For practical purposes we are most interested in the K slope parameters β, which give the marginal effect of change in regressors since $\partial \mathrm{E}[y_{it}]/\partial \mathbf{x}_{it} = \beta$. The N parameters $\alpha_1, \ldots, \alpha_N$ are **nuisance parameters** or **incidental parameters** that are not of intrinsic interest. Nevertheless, their presence potentially prevents estimation of the parameters β that are of interest.

Remarkably, for the linear model there are several ways to consistently estimate β despite the presence of these nuisance parameters. These include (1) OLS in the within model (21.8); (2) direct OLS estimation of the model (21.2) with indicator variables for each of the N fixed effects; (3) GLS in the within model (21.8); (4) ML estimation conditional on the individual means $\bar{y}_i$, $i = 1, \ldots, N$; and (5) OLS in the first-differences model (21.9).

The first two methods always lead to the same estimator for β. So too does the third if additionally the ε_{it} in (21.22) are iid and the fourth if $\varepsilon_{it} \sim \mathcal{N}[0, \sigma^2]$. The last estimator differs from the others for $T > 2$. Such equivalences generally do not hold in nonlinear models, which are considered in Chapter 23.

The essential results for the within estimator are given in the next Section. The first-differences estimator, presented in Section 21.6.2, is extensively used in Chapter 22 when regressors are no longer strongly exogenous. The other estimators are presented in the remainder of Section 21.6, which some readers may wish to skip.

21.6.1. Within or Fixed Effects Estimator

The within model is obtained by subtraction of the time-averaged model $\bar{y}_i = \alpha_i + \bar{\mathbf{x}}_i{}'\beta + \bar{\varepsilon}_i$ from the original model. Then

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\beta + (\varepsilon_{it} - \bar{\varepsilon}_i), \qquad (21.23)$$

so the fixed effect α_i is eliminated, along with time-invariant regressors since $\mathbf{x}_{it} - \bar{\mathbf{x}}_i = \mathbf{0}$ if $\mathbf{x}_{it} = \mathbf{x}_i$ for all t.

Using OLS estimation yields the **within estimator** or **fixed effects estimator** $\widehat{\beta}_{\mathrm{W}}$, where

$$\widehat{\beta}_{\mathrm{W}} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\right]^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(y_{it} - \bar{y}_i). \qquad (21.24)$$

The individual fixed effects α_i can then be estimated by

$$\widehat{\alpha}_i = \bar{y}_i - \bar{\mathbf{x}}_i'\widehat{\beta}_{\mathrm{W}}, \quad i = 1, \ldots, N. \qquad (21.25)$$

The estimate $\widehat{\alpha}_i$ is unbiased for α_i, and it is consistent provided $T \to \infty$ since $\widehat{\alpha}_i$ averages T observations. In short panels the estimates $\widehat{\alpha}_i$ are inconsistent, but $\widehat{\beta}_W$ is nonetheless consistent for $\boldsymbol{\beta}$. The α_i are viewed as **nuisance parameters** or **ancillary parameters** that fortunately do not need to be consistently estimated to obtain consistent estimates of the more important slope parameters $\boldsymbol{\beta}$. This remarkable result need not carry over to more complicated fixed effects models such as nonlinear models.

Consistency of the Within Estimator

The within estimator of $\boldsymbol{\beta}$ is consistent if $\text{plim}(NT)^{-1}\sum_i\sum_t(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\varepsilon_{it} - \bar{\varepsilon}_i) = \mathbf{0}$. This should happen if either $N \to \infty$ or $T \to \infty$ and

$$\mathrm{E}[\varepsilon_{it} - \bar{\varepsilon}_i | \mathbf{x}_{it} - \bar{\mathbf{x}}_i] = 0. \qquad (21.26)$$

Owing to the presence of the averages $\bar{\mathbf{x}}_i = T^{-1}\sum_i \mathbf{x}_{it}$ and $\bar{\varepsilon}_i$ this condition is stronger than $\mathrm{E}[\varepsilon_{it}|\mathbf{x}_{it}] = 0$. A sufficient condition for (21.26) is the strong exogeneity condition that $\mathrm{E}[\varepsilon_{it}|\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}] = 0$. This precludes within estimation with lagged endogenous variables as regressors (see Section 22.5).

Asymptotic Distribution of the Within Estimator

The distribution of $\widehat{\beta}_W$ appears potentially complicated because the error $(\varepsilon_{it} - \bar{\varepsilon}_i)$ in the within model (21.8) is correlated over t for given i. It is shown in the following that the usual OLS results nonetheless apply. Under the strong assumption that ε_{it} is iid,

$$\mathrm{V}[\widehat{\beta}_W] = \sigma_\varepsilon^2 \left[\sum_{i=1}^N \sum_{t=1}^T \ddot{\mathbf{x}}_{it}\ddot{\mathbf{x}}_{it}'\right]^{-1}, \qquad (21.27)$$

where $\ddot{\mathbf{x}}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_i$. A consistent and unbiased estimate of σ_ε^2 is $\widehat{\sigma}_\varepsilon^2 = [N(T-1) - K]^{-1}\sum_i\sum_t \widehat{\ddot{\varepsilon}}_{it}^2$, where the degrees of freedom equal the sample size NT less the number of model parameters K and the N individual effects. Note that if the regression (21.23) is estimated using a standard least-squares package then we need to inflate the reported variances by $[N(T-1) - K]^{-1}[NT - K]$.

For short panels (21.13) yields the robust estimate of the asymptotic variance

$$\mathrm{V}[\widehat{\beta}_W] = \left[\sum_{i=1}^N \sum_{t=1}^T \ddot{\mathbf{x}}_{it}\ddot{\mathbf{x}}_{it}'\right]^{-1} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T \ddot{\mathbf{x}}_{it}\ddot{\mathbf{x}}_{is}'\widehat{\ddot{\varepsilon}}_{it}\widehat{\ddot{\varepsilon}}_{is} \left[\sum_{i=1}^N \sum_{t=1}^T \ddot{\mathbf{x}}_{it}\ddot{\mathbf{x}}_{it}'\right]^{-1}, \qquad (21.28)$$

where $\ddot{\varepsilon}_{it} = \varepsilon_{it} - \bar{\varepsilon}_i$. This preferred estimate permits arbitrary autocorrelations for the ε_{it} and arbitrary heteroskedasticity.

Derivation of the Variance of the Within Estimator

We now derive the estimates of the variance of the within estimator given in (21.27) and (21.28), using matrix algebra. We begin with the model for the itth observation

$$y_{it} = \alpha_i + \mathbf{x}_{it}'\beta + \varepsilon_{it},$$

where $\mathbf{x}_{it}$ and β are $K \times 1$ vectors. For the ith individual, stack all T observations, so

$$\begin{bmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \alpha_i + \begin{bmatrix} \mathbf{x}_{i1}' \\ \vdots \\ \mathbf{x}_{iT}' \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix}, \quad i = 1, \ldots, N,$$

or

$$\mathbf{y}_i = \mathbf{e}\alpha_i + \mathbf{X}_i\beta + \varepsilon_i, \quad i = 1, \ldots, N, \tag{21.29}$$

where $\mathbf{e} = (1, 1, \ldots, 1)'$ is a $T \times 1$ vector of ones, $\mathbf{X}_i$ is a $T \times K$ matrix, and $\mathbf{y}_i$ and ε_i are $T \times 1$ vectors.

To transform model (21.29) to the within model, which subtracts the individual-specific mean, introduce the $T \times T$ matrix

$$\mathbf{Q} = \mathbf{I}_T - T^{-1}\mathbf{e}\mathbf{e}'. \tag{21.30}$$

Premultiplication by the matrix $\mathbf{Q}$ creates deviations from the mean, since

$$\mathbf{Q}\mathbf{W}_i = \mathbf{W}_i - \mathbf{e}\bar{\mathbf{w}}_i', \tag{21.31}$$

where $\mathbf{W}_i$ is a $T \times m$ matrix with tth row $\mathbf{w}_{it}'$ and $\bar{\mathbf{w}}_i = T^{-1}\sum_{t=1}^{T}\mathbf{w}_{it}$ is a $m \times 1$ vector of averages. The result (21.31) is obtained using $\mathbf{e}'\mathbf{W}_i = T\bar{\mathbf{w}}_i'$. Note also that $\mathbf{Q}\mathbf{Q}' = \mathbf{Q}$, using $\mathbf{e}'\mathbf{e} = T$ and $\mathbf{Q}\mathbf{e} = \mathbf{0}$, so $\mathbf{Q}$ is idempotent.

Premultiplying the fixed effects model (21.29) for the ith individual by $\mathbf{Q}$ yields

$$\mathbf{Q}\mathbf{y}_i = \mathbf{Q}\mathbf{X}_i\beta + \mathbf{Q}\varepsilon_i, \quad i = 1, \ldots, N, \tag{21.32}$$

using $\mathbf{Q}\mathbf{e} = \mathbf{0}$. This is the within model (21.23), since equivalently $\mathbf{y}_i - \mathbf{e}\bar{y}_i' = (\mathbf{X}_i - \mathbf{e}\bar{\mathbf{x}}_i')\beta + (\varepsilon_i - \mathbf{e}\bar{\varepsilon}_i)$ using (21.31). Thus premultiplication by $\mathbf{Q}$ yields the within model. An OLS estimation of (21.32) yields $\widehat{\beta}_{\mathrm{W}}$ with variance matrix, assuming independence over i, equal to

$$\mathrm{V}[\widehat{\beta}_{\mathrm{W}}] = \left[\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{Q}'\mathbf{Q}\mathbf{X}_i\right]^{-1} \sum_{i=1}^{N}\mathbf{X}_i'\mathbf{Q}'\mathrm{V}[\mathbf{Q}\varepsilon_i|\mathbf{X}_i]\mathbf{Q}\mathbf{X}_i \left[\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{Q}'\mathbf{Q}\mathbf{X}_i\right]^{-1}. \tag{21.33}$$

Begin with the strong the assumption that ε_{it} are iid $[0, \sigma_\varepsilon^2]$, so that ε_i are iid $[\mathbf{0}, \sigma_\varepsilon^2\mathbf{I}]$. The $T \times 1$ error $\mathbf{Q}\varepsilon_i$ is then independent over i with mean zero and variance $\mathrm{V}[\mathbf{Q}\varepsilon_i] = \mathbf{Q}\mathrm{V}[\varepsilon_i]\mathbf{Q}' = \sigma_\varepsilon^2\mathbf{Q}\mathbf{Q}' = \sigma_\varepsilon^2\mathbf{Q}$. Then

$$\begin{aligned}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{Q}'\mathrm{V}[\mathbf{Q}\varepsilon_i|\mathbf{X}_i]\mathbf{Q}\mathbf{X}_i &= \sum_{i=1}^{N}\mathbf{X}_i'\mathbf{Q}'\sigma_\varepsilon^2\mathbf{Q}\mathbf{Q}\mathbf{X}_i \\ &= \sigma_\varepsilon^2\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{Q}'\mathbf{Q}\mathbf{X}_i,\end{aligned}$$

so that (21.33) simplifies to the estimate given in (21.27), using

$$(\mathbf{QX}_i)'(\mathbf{QX}_i) = \sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'.$$

At the time of writing many packages use (21.27) but alternative estimators may be better. In particular, the assumption of serially uncorrelated error ε_{it} is easily relaxed. If $\boldsymbol{\varepsilon}_i$ are iid $[\mathbf{0}, \boldsymbol{\Sigma}_i]$ we use the more general form of the variance matrix (21.33) with $\mathrm{Cov}[\mathbf{Q}\boldsymbol{\varepsilon}_i, \mathbf{Q}\boldsymbol{\varepsilon}_j] = \mathbf{0}$, for $i \neq j$, and $\mathrm{V}[\mathbf{Q}\boldsymbol{\varepsilon}_i]$ replaced by $(\mathbf{Q}\widehat{\boldsymbol{\varepsilon}}_i)(\mathbf{Q}\widehat{\boldsymbol{\varepsilon}}_i)'$, where $\widehat{\boldsymbol{\varepsilon}}_i = \mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}_{\mathrm{W}}$. This yields the estimate given in (21.28).

From the derivation it should be clear that $\widehat{\boldsymbol{\beta}}_{\mathrm{W}}$ is also consistent in the random effects model, though as shown in Section 21.7 it is less efficient than the random effects estimator if the random effects model is appropriate.

GLS Estimation of the Within Model

The within model (21.32) can also be estimated by feasible GLS.

If in fact ε_{it} are iid $[0, \sigma_\varepsilon^2]$, however, then there are no gains to doing GLS. To see this, note that then $\mathbf{Q}\boldsymbol{\varepsilon}_i$ is independent of $\mathbf{Q}\boldsymbol{\varepsilon}_j$, $i \neq j$, with $\mathrm{V}[\mathbf{Q}\boldsymbol{\varepsilon}_i] = \sigma_\varepsilon^2\mathbf{Q}$, so the **GLS estimator** is

$$\widehat{\boldsymbol{\beta}}_{\mathrm{W,GLS}} = \left[\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{Q}'\mathbf{Q}^{-}\mathbf{Q}\mathbf{X}_i\right]^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{Q}'\mathbf{Q}^{-}\mathbf{Q}\mathbf{y}_i,$$

where the generalized inverse $\mathbf{Q}^{-}$ is used as $\mathbf{Q}$ is not of full rank. However, $\mathbf{Q}'\mathbf{Q}^{-}\mathbf{Q} = \mathbf{Q}'\mathbf{Q}$ since $\mathbf{Q}'\mathbf{Q}^{-}\mathbf{Q} = \mathbf{Q}$, for a generalized inverse, and $\mathbf{Q} = \mathbf{Q}\mathbf{Q}'$ as $\mathbf{Q}$ here is idempotent. Replacing $\mathbf{Q}'\mathbf{Q}^{-}\mathbf{Q}$ by $\mathbf{Q}'\mathbf{Q}$ in the formula for $\widehat{\boldsymbol{\beta}}_{\mathrm{W,GLS}}$ yields the OLS estimator in (21.32).

There can be gains to GLS if other models for ε_{it} are assumed. The approach is essentially the same as that in Section 21.5.2 for pooled GLS without fixed effects, except that first the fixed effect must be eliminated. This leads to error $\mathbf{Q}\boldsymbol{\varepsilon}_i$ that is less than full rank, so we first drop one time period and apply pooled GLS to only $(T-1)$ time periods. It is easier, and often not much less efficient, to instead just use the usual within FE estimator and then obtain panel-robust standard errors using (21.28).

MaCurdy (1982b) gives a Box–Jenkins-type analysis for identification and estimation of ARMA processes for ε_{it} in a fixed effects model for a short panel. For short panels it is not necessary to assume an ARMA process for ε_{it} or even stationarity, since for $N \to \infty$ we can always consistently estimate $\mathrm{Cov}[u_{it}, u_{is}]$ by $N^{-1}\sum_i \widehat{u}_{it}\widehat{u}_{is}$. Nonetheless, there may be interest in determining the ARMA process for the errors.

21.6.2. First-Differences Estimator

The within model is obtained by subtraction of the time-averaged model $\bar{y}_i = \alpha_i + \bar{\mathbf{x}}_i'\boldsymbol{\beta} + \bar{\varepsilon}_i$ from the original model. Alternatively, one can subtract the model lagged one period, $y_{i,t-1} = \alpha_i + \mathbf{x}_{i,t-1}'\beta + \varepsilon_{i,t-1}$. Then

$$(y_{it} - y_{i,t-1}) = (\mathbf{x}_{it} - \mathbf{x}_{i,t-1})'\beta + (\varepsilon_{it} - \varepsilon_{i,t-1}), \quad t = 2, \ldots, T, \tag{21.34}$$

so the fixed effect α_i is eliminated. An OLS estimation yields the **first-differences estimator**

$$\widehat{\beta}_{\mathrm{FD}} = \left[\sum_{i=1}^{N}\sum_{t=2}^{T}(\mathbf{x}_{it} - \mathbf{x}_{i,t-1})(\mathbf{x}_{it} - \mathbf{x}_{i,t-1})'\right]^{-1} \sum_{i=1}^{N}\sum_{t=2}^{T}(\mathbf{x}_{it} - \mathbf{x}_{i,t-1})(y_{it} - y_{i,t-1}). \tag{21.35}$$

Note that there only $N(T-1)$ observations in this regression. An easy error to make in implementation is to stack all NT observations and then subtract the first lag. Then only the (1, 1) observation is dropped, whereas all T first-period observations $(i, 1)$, $i = 1, \ldots, N$, must be dropped after differencing.

Consistency of the First-Differences Estimator

Consistency of the first differences estimator requires that $\mathrm{E}[\varepsilon_{it} - \varepsilon_{i,t-1}|\mathbf{x}_{it} - \mathbf{x}_{i,t-1}]$. This is a stronger condition than $\mathrm{E}[\varepsilon_{it}|\mathbf{x}_{it}] = 0$ but a weaker condition than the strong exogeneity condition needed for consistency of the within estimator.

Asymptotic Distribution of the First-Differences Estimator

Statistical inference requires adjusting the usual OLS standard errors to account for the correlation over time in the error term $\varepsilon_{it} - \varepsilon_{i,t-1}$. To obtain the asymptotic variance of $\widehat{\beta}_{\mathrm{FD}}$, stack the model for the ith individual as

$$\Delta\mathbf{y}_i = \Delta\mathbf{X}_i'\beta + \Delta\varepsilon_i,$$

where $\Delta\mathbf{y}_i$ is a $(T-1)\times 1$ vector with entries $(y_{i2} - y_{i1}), \ldots, (y_{iT} - y_{i,T-1})$, and $\Delta\mathbf{X}_i$ is a $(T-1)\times K$ vector with rows $(\mathbf{x}_{i2} - \mathbf{x}_{i1})', \ldots, (\mathbf{x}_{iT} - \mathbf{x}_{i,T-1})'$. Then

$$\widehat{\beta}_{\mathrm{FD}} = \left[\sum_{i=1}^{N}(\Delta\mathbf{X}_i)'(\Delta\mathbf{X}_i)\right]^{-1}\sum_{i=1}^{N}(\Delta\mathbf{X}_i)'(\Delta\mathbf{y}_i) \tag{21.36}$$

has variance matrix, assuming independence over i, of

$$\mathrm{V}[\widehat{\beta}_{\mathrm{FD}}] = \left[\sum_{i=1}^{N}(\Delta\mathbf{X}_i)'(\Delta\mathbf{X}_i)\right]^{-1}\left[\sum_{i=1}^{N}(\Delta\mathbf{X}_i)'\mathrm{V}[\Delta\varepsilon_i|\Delta\mathbf{X}_i](\Delta\mathbf{X}_i)\right]\left[\sum_{i=1}^{N}(\Delta\mathbf{X}_i)'(\Delta\mathbf{X}_i)\right]^{-1}. \tag{21.37}$$

The simplest assumption is that ε_{it} are iid $[0, \sigma_\varepsilon^2]$. Then the error $(\varepsilon_{it} - \varepsilon_{i,t-1})$ is now an MA(1) error, with variance $2\sigma_\varepsilon^2$ and one-period apart autocovariance σ_ε^2 for individual i. It follows that $\mathrm{V}[\Delta\varepsilon_i]$ equals σ_ε^2 times a $(T-1)\times(T-1)$ matrix with entries of 2 on the diagonal, entries of 1 on the immediate off-diagonals, and 0s elsewhere.

A more realistic assumption is that ε_{it} is correlated over time for given i, so that $\mathrm{Cov}[\varepsilon_{it}, \varepsilon_{is}] \neq 0$ for $t \neq s$, but is still independent over i. From (21.13), for short panels an estimator that is robust to general forms of autocorrelation and heteroskedasticity is (21.37) with $V[\Delta\varepsilon_i]$ replaced by $(\widehat{\Delta\varepsilon_i})'(\widehat{\Delta\varepsilon_i})$. One should never use the usual OLS standard errors from OLS regression of the first-differences model (21.37), as these are only correct in the unlikely event that ε_{it} is a random walk, so that $(\varepsilon_{it} - \varepsilon_{i,t-1})$ are iid.

For $T = 2$ the first-differences and within estimators are equal since $\bar{y} = (y_1 + y_2)/2$ so $(y_1 - \bar{y}) = (y_1 - y_2)/2$ and $(y_2 - \bar{y}) = -(y_1 - y_2)/2$, and similarly for $\mathbf{x}$. For $T > 2$ the two estimators differ. Under the simplest assumption that ε_{it} are iid, it can be shown that the GLS estimator of the first-difference model (21.34) equals the within estimator. The estimator $\widehat{\beta}_{\text{FD}}$ instead estimates (21.34) by OLS and is less efficient than $\widehat{\beta}_{\text{W}}$. For this reason the first-difference estimator is not mentioned much in introductory courses. However, it is used extensively once lagged dependent variables are introduced (see Chapter 22). Then the within estimator is inconsistent. The first-differences estimator is also inconsistent, but relies on weaker exogeneity assumptions that permit consistent IV estimation.

21.6.3. Conditional ML Estimator

The conditional MLE maximizes the joint likelihood of $y_{11}, \ldots, y_{NT}$ conditional on the individual averages $\bar{y}_1, \ldots, \bar{y}_T$. This method has the attraction that, for the linear panel model under normality, the fixed effects α_i are eliminated, so maximization is with respect to $\boldsymbol{\beta}$ alone.

Assume that y_{it} conditional on regressors $\mathbf{x}_{it}$ and parameters α_i, $\boldsymbol{\beta}$, and σ^2 are iid with normal distribution $\mathcal{N}[\alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta}, \sigma^2]$. Then the **conditional likelihood function** is

$$
\begin{aligned}
\mathrm{L}_{\text{COND}}(\boldsymbol{\beta},\sigma^2,\boldsymbol{\alpha}) &= \prod_{i=1}^{N} f(y_{i1},\ldots,y_{iT}|\bar{y}_i) \\
&= \prod_{i=1}^{N} \frac{f(y_{i1},\ldots,y_{iT},\bar{y}_i)}{f(\bar{y}_i)} \\
&= \prod_{i=1}^{N} \frac{(2\pi\sigma^2)^{-T/2}}{(2\pi\sigma^2/T)^{-1/2}} \exp\left\{\sum_{t=1}^{T} -[(y_{it}-\mathbf{x}_{it}'\boldsymbol{\beta})^2 + (\bar{y}_i - \bar{\mathbf{x}}_i'\boldsymbol{\beta})^2]/2\sigma^2\right\}.
\end{aligned}
\tag{21.38}
$$

The first equality defines the conditional likelihood assuming independence over i. The second equality always holds since, suppressing subscript i, $f(y_1,\ldots,y_T|\bar{y}) = f(y_1,\ldots,y_T,\bar{y})/f(\bar{y})$ and $f(y_1,\ldots,y_T,\bar{y}) = f(y_1,\ldots,y_T)$ as knowledge of $\bar{y} = T^{-1}\sum_i y_i$ adds nothing given knowledge of $y_1,\ldots,y_T$. The third equality under normality comes after considerable algebra that is left as an exercise.

The key result is that the fixed effects $\boldsymbol{\alpha}$ do not appear in the final equality in (21.38), so $\mathrm{L}_{\text{COND}}(\boldsymbol{\beta},\sigma^2,\boldsymbol{\alpha})$ is in fact $\mathrm{L}_{\text{COND}}(\boldsymbol{\beta},\sigma^2)$, and we need to maximize the conditional log-likelihood function (21.38) with respect to $\boldsymbol{\beta}$ and σ^2 only. The resulting **conditional ML estimator** $\widehat{\boldsymbol{\beta}}_{\text{CML}}$ solves the first-order conditions

$$
\frac{1}{\sigma^2}\sum_{t=1}^{T}\sum_{i=1}^{N}[(y_{it}-\mathbf{x}_{it}'\boldsymbol{\beta})\mathbf{x}_{it} - (\bar{y}_i - \bar{\mathbf{x}}_i'\boldsymbol{\beta})\bar{\mathbf{x}}_i] = \mathbf{0},
$$

or equivalently

$$\sum_{t=1}^{T}\sum_{i=1}^{N}[(y_{it}-\bar{y}_i)-(\mathbf{x}_{it}-\bar{\mathbf{x}}_i)'\boldsymbol{\beta})](\mathbf{x}_{it}-\bar{\mathbf{x}}_i)=\mathbf{0}.$$

However, these are just the first-order conditions from OLS regression of $(y_{it}-\bar{y}_i)$ on $(\mathbf{x}_{it}-\bar{\mathbf{x}}_i)$.

The conditional MLE $\widehat{\boldsymbol{\beta}}_{\text{CML}}$ therefore equals the within estimator $\widehat{\boldsymbol{\beta}}_{\text{W}}$.

Intuitively, the method yields a consistent estimator because conditioning on $\bar{y}_i$ in (21.38) eliminated the fixed effects. More formally, $\bar{y}_i$ is a sufficient statistic for α_i and conditioning on the sufficient statistic enables consistent estimation of $\boldsymbol{\beta}$ (see Section 23.2.2).

21.6.4. Least-Squares Dummy Variable Estimator

Consider the original fixed effects model (21.22) before any differencing. An OLS analysis can be applied directly to the model, simultaneously estimating $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

In principle no special software is needed. One simply estimates the OLS regression of y_{it} on $\mathbf{x}_{it}$ and a set of N indicator variables $d_{1,it},\ldots,d_{N,it}$, where $d_{j,it}$ equals one if $j=i$ and equals zero otherwise. However, as N gets large there are too many regressors to permit inversion of the $(N+K)\times(N+K)$ regressor matrix. Some matrix algebra, however, reduces the problem to inversion of a $K\times K$ matrix.

The resulting estimator of $\boldsymbol{\beta}$ turns out to equal the within estimator. This is a special case of the so-called Frisch-Waugh Theorem for a subset regression. If dummy variables are partialled out by regression of all the variables on the dummies, and if the residuals from these regressions are used in a second stage regression, then we get the same estimates as in the full regression. But these residuals here are simply deviations from their respective means, i.e. the within regression. For completeness we now present the relevant matrix algebra.

Stack the $T\times 1$ vectors in (21.29) over all N individuals to yield the fixed effects **dummy variable model**

$$\begin{bmatrix}\mathbf{y}_1\\ \vdots\\ \mathbf{y}_N\end{bmatrix}=\begin{bmatrix}\mathbf{e} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \ddots & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{e}\end{bmatrix}\begin{bmatrix}\alpha_1\\ \vdots\\ \alpha_N\end{bmatrix}+\begin{bmatrix}\mathbf{X}_1\\ \vdots\\ \mathbf{X}_N\end{bmatrix}\boldsymbol{\beta}+\begin{bmatrix}\boldsymbol{\varepsilon}_1\\ \vdots\\ \boldsymbol{\varepsilon}_N\end{bmatrix},$$

or

$$\mathbf{y}=[(\mathbf{I}_N\otimes\mathbf{e})\quad \mathbf{X}]\begin{bmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\beta}\end{bmatrix}+\boldsymbol{\varepsilon}, \qquad (21.39)$$

where $\mathbf{y}$ is an $NT\times 1$ vector, the Kronecker product $(\mathbf{I}_N\otimes\mathbf{e})$ is an $NT\times N$ block-diagonal matrix, and $\mathbf{X}$ is the $NT\times K$ matrix of nonconstant regressors.

An OLS estimation of this model yields the **least-squares dummy variable (LSDV) estimator**

$$\begin{bmatrix} \widehat{\alpha}_{\text{LSDV}} \\ \widehat{\beta}_{\text{LSDV}} \end{bmatrix} = \begin{bmatrix} (\mathbf{I}_N \otimes \mathbf{e})'(\mathbf{I}_N \otimes \mathbf{e}) & (\mathbf{I}_N \otimes \mathbf{e})'\mathbf{X} \\ \mathbf{X}'(\mathbf{I}_N \otimes \mathbf{e}) & \mathbf{X}'\mathbf{X} \end{bmatrix}^{-1} \times \begin{bmatrix} (\mathbf{I}_N \otimes \mathbf{e})'\mathbf{y} \\ \mathbf{X}'\mathbf{y} \end{bmatrix}$$

$$= \begin{bmatrix} T\mathbf{I}_N & T\bar{\mathbf{X}} \\ T\bar{\mathbf{X}}' & \mathbf{X}'\mathbf{X} \end{bmatrix}^{-1} \times \begin{bmatrix} \bar{\mathbf{y}} \\ \mathbf{X}'\mathbf{y} \end{bmatrix},$$

where the matrix of sample means $\bar{\mathbf{X}} = [\bar{\mathbf{x}}_1' \cdots \bar{\mathbf{x}}_N']'$, $\bar{\mathbf{x}}_i = T^{-1}\sum_{t=1}^{T} \mathbf{x}_{it}$, $\bar{\mathbf{y}} = [\bar{y}_1 \cdots \bar{y}_N]'$, and $\bar{y}_i = T^{-1}\sum_{t=1}^{T} y_{it}$. Using the formula for partitioned inverse and performing further algebra leads to

$$\begin{bmatrix} \widehat{\alpha}_{\text{LSDV}} \\ \widehat{\beta}_{\text{LSDV}} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{y}} - \bar{\mathbf{X}}\widehat{\beta}_{\text{W}} \\ [\mathbf{X}'\mathbf{X} - \bar{\mathbf{X}}'\bar{\mathbf{X}}]^{-1}(\mathbf{X}'\mathbf{y} - \bar{\mathbf{X}}'\bar{\mathbf{y}}) \end{bmatrix}. \tag{21.40}$$

Reexpressing this in summation notation, we have $\widehat{\beta}_{\text{LSDV}} = \widehat{\beta}_{\text{W}}$ defined in (21.24) and $\widehat{\alpha}_{\text{LSDV}} = \widehat{\alpha}_{\text{FE}}$ defined in (21.25), so the LSDV estimators equal the within or fixed effects estimator

For short panels an obvious potential problem is that consistent estimation of β and α is not guaranteed as there are $N + K$ parameters to estimate and $N \to \infty$. Remarkably, consistent estimation of β is possible, even though α is inconsistently estimated, unless additionally $T \to \infty$.

This estimator is second-moment efficient if ε_{it} are iid $[0, \sigma^2]$. It follows that the within estimator of β is more efficient than alternative differencing estimators that also eliminate α_i, such as subtracting the first observation or the previous period's observation. If additionally the errors are normally distributed, the LSDV estimator equals the MLE by the usual equivalence of OLS and MLE in the linear model with spherical normal errors.

21.6.5. Covariance Estimator

Suppose data belong to one of N classes, with y_{it} denoting the tth observation in the ith class. The **analysis of variance** decomposes the total variation of y_{it} around the grand mean $\bar{y}$, $\sum_i \sum_t (y_{it} - \bar{y})^2$, into **within-group** variation $\sum_i \sum_t (y_{it} - \bar{y}_i + \bar{y})^2$ and **between-group** variation $\sum_i (\bar{y}_i - \bar{y})^2$, where $\bar{y}_i$ is the mean in the ith group. Group membership becomes more important as between-group variation increases. The **analysis of covariance** extends this approach to introduce regressors, in which case the residual sum of squares is similarly decomposed. This framework is widely used in applied statistics.

For short panels each individual is viewed as a class, observed for several time periods. The model (21.3) is called the **analysis-of-covariance model**, as it permits the mean residual in the ith class to differ over classes. The estimator of this model, the within estimator, is accordingly also called the **covariance estimator**.

21.7. Random Effects Model

The **random effects model** (21.3) can be rewritten as

$$y_{it} = \mu + \mathbf{x}_{it}'\beta + \alpha_i + \varepsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{21.41}$$

or

$$y_{it} = \mathbf{w}_{it}'\delta + \alpha_i + \varepsilon_{it}, \tag{21.42}$$

where $\mathbf{w}_{it} = [1\ \mathbf{x}_{it}]$ and $\delta = [\mu \quad \beta']'$. The individual-specific effects α_i are assumed to be realizations of iid random variables with distribution $[0, \sigma_\alpha^2]$ and the error ε_{it} is iid $[0, \sigma_\varepsilon^2]$. The nonrandom scalar intercept μ is added so that, unlike in (21.5), the random effects can be normalized to have zero mean.

The model can alternatively be viewed as a special case of a **random coefficient** or **varying coefficient model**, where only the intercept coefficient is random. The model can be re-expressed as $y_{it} = \mu + \mathbf{x}_{it}'\beta + u_{it}$, where the error term u_{it} has two components $u_{it} = \alpha_i + \varepsilon_{it}$. For this reason the random effects model is also called the **error components model**. Even clearer terminology may be the **random intercept model**. Richer mixed models also permit random slopes, see Chapter 22.

There are many consistent estimators of the random effects model, including (1) GLS estimation in the model (21.42); (2) ML estimation in the model (21.42) assuming α_i and ε_{it} are normally distributed; (3) OLS estimation in the model (21.42); and (4) fixed effects model estimators such as the within and first-differences estimators, though these only estimate the coefficients of time-varying regressors. The first two estimators are asymptotically equivalent but can vary in finite samples depending on the specific estimates used for σ_α^2 and σ_ε^2. The remaining estimators are consistent, though they are inefficient if in fact α_i and ε_{it} are iid.

21.7.1. GLS Estimator

The **random effects estimator** of μ and β is the feasible GLS estimator of the model (21.42), and it is shown later in this section that it can be implemented by OLS regression of the transformed equation

$$y_{it} - \widehat{\lambda}\bar{y}_i = (1 - \widehat{\lambda})\mu + (\mathbf{x}_{it} - \widehat{\lambda}\bar{\mathbf{x}}_i)'\beta + v_{it}, \tag{21.43}$$

where $v_{it} = (1 - \widehat{\lambda})\alpha_i + (\varepsilon_{it} - \widehat{\lambda}\bar{\varepsilon}_i)$ and $\widehat{\lambda}$ is consistent for

$$\lambda = 1 - \sigma_\varepsilon/(T\sigma_\alpha^2 + \sigma_\varepsilon^2)^{1/2}. \tag{21.44}$$

Equivalently,

$$\widehat{\delta}_{\text{RE}} = \begin{bmatrix} \widehat{\mu}_{\text{RE}} \\ \widehat{\beta}_{\text{RE}} \end{bmatrix} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{w}_{it} - \widehat{\lambda}\bar{\mathbf{w}}_i)(\mathbf{w}_{it} - \widehat{\lambda}\bar{\mathbf{w}}_i)'\right]^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{w}_{it} - \widehat{\lambda}\bar{\mathbf{w}}_i)(y_{it} - \widehat{\lambda}\bar{y}_i), \tag{21.45}$$

where $\mathbf{w}_{it} = [1\ \mathbf{x}_{it}]$ and $\bar{\mathbf{w}}_i = [1\ \bar{\mathbf{x}}_i]$. Consistency requires $NT \to \infty$, through either $N \to \infty$ or $T \to \infty$ or both.

Assuming that ε_{it} and α_i are iid, the usual OLS output from OLS regression of (21.43) can be used to obtain the variance matrix estimate, so that

$$\mathrm{V}\begin{bmatrix}\widehat{\mu}_{\mathrm{RE}}\\ \widehat{\beta}_{\mathrm{RE}}\end{bmatrix} = \sigma_\varepsilon^2\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{w}_{it}-\widehat{\lambda}\bar{\mathbf{w}}_i)(\mathbf{w}_{it}-\widehat{\lambda}\bar{\mathbf{w}}_i)'\right]^{-1}. \tag{21.46}$$

Alternatively, for short panels a robust variance estimate that permits quite general behavior for $\alpha_i + \varepsilon_{it}$ can be obtained using (21.13). This yields

$$\mathrm{V}\begin{bmatrix}\widehat{\mu}_{\mathrm{RE}}\\ \widehat{\beta}_{\mathrm{RE}}\end{bmatrix} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{\mathbf{w}}_{it}\widetilde{\mathbf{w}}_{it}'\right]^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{\mathbf{w}}_{it}\widetilde{\mathbf{w}}_{is}'\widehat{\widetilde{\varepsilon}}_{it}\widehat{\widetilde{\varepsilon}}_{is}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{\mathbf{w}}_{it}\widetilde{\mathbf{w}}_{it}'\right]^{-1}, \tag{21.47}$$

where $\widetilde{\mathbf{w}}_{it} = \mathbf{w}_{it} - \widehat{\lambda}\bar{\mathbf{w}}_i$ and $\widetilde{\varepsilon}_{it} = \widehat{\varepsilon}_{it} - \widehat{\lambda\bar{\varepsilon}}_i$ where $\widehat{\varepsilon}_{it}$ is the RE residual. This estimate permits arbitrary autocorrelations for the ε_{it} and arbitrary heteroskedasticity.

Equation (21.46) requires consistent estimates of the **variance components** σ_ε^2 and σ_α^2. From the within or fixed effects regression of $(y_{it} - \bar{y}_i)$ on $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$ we obtain

$$\widehat{\sigma}_\varepsilon^2 = \frac{1}{N(T-1)-K}\sum_i\sum_t((y_{it}-\bar{y}_i)-(\mathbf{x}_{it}-\bar{\mathbf{x}}_i)'\widehat{\beta}_{\mathrm{W}})^2. \tag{21.48}$$

From the between regression of $\bar{y}_i$ on an intercept and $\bar{\mathbf{x}}_i$, an equation that has error term with variance $\sigma_\alpha^2 + \sigma_\varepsilon^2/T$, we obtain

$$\widehat{\sigma}_\alpha^2 = \frac{1}{N-(K+1)}\sum_i(\bar{y}_i - \widehat{\mu}_{\mathrm{B}} - \bar{\mathbf{x}}_i'\widehat{\beta}_{\mathrm{B}})^2 - \frac{1}{T}\widehat{\sigma}_\varepsilon^2. \tag{21.49}$$

More efficient estimators of the variance components σ_ε^2 and σ_α^2 are possible (see, for example, Amemiya, 1985), but these will not necessarily increase the efficiency of $\widehat{\beta}_{\mathrm{RE}}$. A wide range of estimators are possible. The variance estimator (21.49) can be negative, in which case programs often set $\widehat{\sigma}_\alpha^2 = 0$, so $\widehat{\lambda} = 0$ and estimation is then by pooled OLS.

To verify that the feasible GLS estimator simplifies to OLS estimation of (21.43), stack (21.42) by observations from all T time periods for given i in the same way as for the fixed effects model. Then

$$\mathbf{y}_i = \mathbf{W}_i\delta + (\mathbf{e}\alpha_i + \varepsilon_i), \tag{21.50}$$

where $\mathbf{y}_i$, $\mathbf{e}$, ε_i, and $\mathbf{X}_i$ are defined after (21.29), and $\mathbf{W}_i' = [\mathbf{e}\quad \mathbf{X}_i']$. To estimate by GLS we need to obtain the variance matrix $\boldsymbol{\Omega}$ of the $T \times 1$ vector error $(\mathbf{e}\alpha_i + \varepsilon_i)$. Given independence of α_i and ε_{it} we have $\mathrm{E}[(\mathbf{e}\alpha_i + \varepsilon_i)(\mathbf{e}\alpha_i + \varepsilon_i)'] = \mathrm{E}[\varepsilon_i\varepsilon_i'] + \mathrm{E}[\alpha_i^2]\mathbf{e}\mathbf{e}'$. Since ε_{it} are iid $[0, \sigma_\varepsilon^2]$ and α_i are iid $[0, \sigma_\alpha^2]$ we obtain

$$\boldsymbol{\Omega} = \sigma_\varepsilon^2\mathbf{I}_T + \sigma_\alpha^2\mathbf{e}\mathbf{e}' = \sigma_\varepsilon^2\left[\mathbf{Q} + \frac{1}{\psi^2}(\mathbf{I}_T - \mathbf{Q})\right],$$

where $\mathbf{Q} = \mathbf{I}_T - T^{-1}\mathbf{e}\mathbf{e}'$ was introduced in (21.30) and $\psi^2 = \sigma_\varepsilon^2/[\sigma_\varepsilon^2 + T\sigma_\alpha^2]$. Using $\mathbf{Q}\mathbf{Q}' = \mathbf{Q}$ we can easily verify that $\boldsymbol{\Omega}^{-1} = \sigma_\varepsilon^{-2}[\mathbf{Q} + \psi^2(\mathbf{I}_T - \mathbf{Q})]$ and

$$\boldsymbol{\Omega}^{-1/2} = \frac{1}{\sigma_\varepsilon}\left[\mathbf{Q} + \psi(\mathbf{I}_T - \mathbf{Q})\right]. \tag{21.51}$$

The GLS estimator is obtained by premultiplication of (21.50) by any scalar multiple of $\boldsymbol{\Omega}^{-1/2}$. Now

$$\begin{aligned}\left[\mathbf{Q}+\psi(\mathbf{I}_T-\mathbf{Q})\right]\mathbf{y}_i &= \mathbf{y}_i - \mathbf{e}\bar{y}_i' + \psi(\mathbf{y}_i - (\mathbf{y}_i - \mathbf{e}\bar{y}_i')) \\ &= \mathbf{y}_i - \lambda\mathbf{e}\bar{y}_i',\end{aligned}$$

where $\lambda = (1-\psi)$. Performing similar algebra for $\mathbf{W}_i$, $\mathbf{e}\alpha_i$, and $\boldsymbol{\varepsilon}_i$ in (21.50) yields the following model:

$$\mathbf{y}_i - \lambda\mathbf{e}\bar{y}_i' = (\mathbf{W}_i - \lambda\bar{\mathbf{W}})'\boldsymbol{\delta} + (1-\lambda)\alpha_i + (\boldsymbol{\varepsilon}_i - \lambda\mathbf{e}\bar{\varepsilon}_i'), \tag{21.52}$$

where the transformed error in (21.52) has variance matrix $\sigma_\varepsilon^2\mathbf{I}_T$. The GLS estimator is the OLS estimator of (21.52), but (21.52) is just a stacked version of (21.43) with the scalar λ replaced by a consistent estimate.

The random effects estimator $\widehat{\boldsymbol{\beta}}_{\text{RE}}$ of the slope parameters converges to the within estimator as $T \to \infty$ since then $\lambda \to 1$. Otherwise, $\widehat{\boldsymbol{\beta}}_{\text{RE}}$ can be shown, after some algebra, to equal a **matrix-weighted combination** of the within estimator and the between estimator. If the random effects model is appropriate, this weighted average works better than using the within estimator alone. However, if the fixed effects model is appropriate then this weighted average is inconsistent, as the between estimator is then inconsistent. The estimator of the intercept can be shown to simplify to $\widehat{\mu}_{\text{RE}} = \bar{y} - \bar{\mathbf{X}}\widehat{\boldsymbol{\beta}}_{\text{RE}}$. For more details see, for example, Hsiao (2003, p. 36) or Greene (2003).

21.7.2. ML Estimator

In the derivation in the previous section, normality of the errors is not assumed. If they are in fact **normal**, we can maximize the log-likelihood function with respect to $\boldsymbol{\beta}$, μ, σ_ε^2, and σ_α^2. For given σ_ε^2 and σ_α^2 the MLE for $\boldsymbol{\beta}$ and μ is the same as the GLS estimator, but the MLE gives estimators $\widetilde{\sigma}_\varepsilon^2$ and $\widetilde{\sigma}_\alpha^2$ that differ from those given in (21.48) and (21.49).

Thus the MLE for $\boldsymbol{\beta}$ and μ is given by (21.45) with $\widehat{\lambda}$ replaced by the alternative consistent estimate $\widetilde{\lambda} = 1 - \widetilde{\sigma}_\varepsilon/(T\widetilde{\sigma}_\alpha^2 + \widetilde{\sigma}_\varepsilon^2)^{1/2}$. Asymptotically, the MLE and GLS estimators of the random effects model are equivalent, but the two will differ in finite samples.

For the MLE there may be two local maxima rather than one of the likelihood for $0 < \psi^2 \leq 1$, so care is needed to ensure a global maximum.

21.7.3. Other Estimators

Many different estimators of $\boldsymbol{\beta}$ are consistent if the random effects model is the correct model. In particular, the pooled OLS, within, first-differences, and between estimators are all consistent. However they are inefficient if α_i and ε_{it} are iid, and the within and first-differences estimators can only estimate the coefficients of time-varying regressors.

21.8. Modeling Issues

In this section we consider some practical issues that arise in linear panel data models, even in the absence of complications such as endogeneity and lagged dependent variables, topics that are deferred to Chapter 22.

21.8.1. Tests for Pooling

The random effects model restricts all regression parameters to be the same in different cross sections and time periods, whereas the fixed effects models imposes parameter constancy except for the intercept, which may vary across individuals. **Tests of poolability** test the appropriateness of these constraints.

These tests are usually done using a Chow test (see Greene, 2003, p. 130) based on the tests for equality of regressors in two linear regressions assuming a common variance. Depending on the assumptions about errors, the Chow test may be applied to models estimated by OLS or by GLS. Baltagi (2001, Chapter 4) and Hsiao (2003, Chapter 2) provide detailed coverage.

For short panels it is not possible to allow the slope parameters to differ across individuals, as then the number of parameters goes to infinity. However, parameters can be permitted to vary over time. The model $y_{it} = \gamma + \mathbf{x}_{it}'\boldsymbol{\beta} + u_{it}$ is then tested against $y_{it} = \gamma_t + \mathbf{x}_{it}'\boldsymbol{\beta}_t + u_{it}$. The most obvious method is to assume random effects with $u_{it} = \varepsilon_{it} + \alpha_i$, estimate the restricted model ($\gamma_t = \gamma$ and $\boldsymbol{\beta}_t = \boldsymbol{\beta}$) using the random effects GLS estimator, and compare the restricted and unrestricted residual sums of squares in the transformed models. If more robust inference is preferred then panel-robust standard errors should be obtained and a Wald test performed. For short panels it is common to specify models with slope parameters $\boldsymbol{\beta}$ constant, though the intercept γ_t may be permitted to vary over time by inclusion of time dummies as additional regressors.

21.8.2. Tests for Individual-Specific Effects

Breusch and Pagan (1980) derived Lagrange-multiplier tests for the presence of individual-specific random effects against the null hypothesis assumption of iid errors. These have the advantage of being easily implemented by an auxiliary regression that requires only residuals from pooled OLS estimates. Alternatively, one can assume normality and do a likelihood ratio test of the random effects MLE against the MLE of the constant-coefficients model, or a Wald test of $\sigma_\alpha = 0$ in the random effects model.

In practice one often rejects the null hypothesis that the errors in the constant-coefficients model are iid. It is easiest to immediately estimate by pooled OLS with panel-robust standard errors or by random effects GLS.

For a short panel formal tests for the presence of individual-specific fixed effects are not possible because of the incidental parameters problem. It is not possible to test whether N parameters are zero when there are only NT observations and T is small. Instead, the Hausman test of Section 21.4.3 is used to test the null hypothesis of random effects against the alternative of fixed effects.

21.8.3. Prediction

Prediction in models without individual effects is straightforward: Use $\widehat{y}_{js} = \mathbf{x}'_{js}\widehat{\beta}$. This is a prediction of the population average $\mathrm{E}[y_{js}|\mathbf{x}_{js}]$.

Prediction for a given individual conditional on the individual-specific effect is more difficult. This is prediction of $\mathrm{E}[y_{js}|\mathbf{x}_{js}, \alpha_i]$. We consider out-of-sample forecasts for the ith individual using the random effects model (21.42). Then $y_{i,t+s} = \mathbf{w}'_{it}\delta + u_{i,t+s}$, where $u_{i,t+s} = \alpha_i + \varepsilon_{i,t+s}$. The obvious predictor replaces δ by $\widehat{\delta}_{\mathrm{RE}}$ and $u_{i,t+s}$ by either 0 or $\widehat{\bar{u}}_i$, where $\widehat{\bar{u}}_i = \bar{y}_i - \bar{\mathbf{w}}'_i\widehat{\delta}_{\mathrm{RE}}$ is the average within-sample residual for the ith individual. However, this is inefficient as it ignores the correlation between $u_{i,t+s}$ and in-sample errors induced by the individual-specific random effect α_i. The problem is an example of the more general problem of prediction within a GLS rather than an OLS framework. For this special case the best linear unbiased predictor (see Section 22.8.3) is $\widehat{y}_{i,t+s} = \mathbf{x}'_{it}\widehat{\delta}_{\mathrm{RE}} + (T\sigma^2_\alpha/(T\sigma^2_\alpha + \sigma^2_\varepsilon))\widehat{\bar{u}}_i$. For the fixed effects model the obvious predictor is $\widehat{y}_{i,t+s} = \mathbf{x}'_{it}\widehat{\beta}_{\mathrm{W}} + \widehat{\alpha}_{i,\mathrm{FE}}$, but again this is inconsistent in short panels.

21.8.4. Two-Way Effects Models

The analysis to date has focused on the one-way model, which is (21.1) with $u_{it} = \alpha_i + \varepsilon_{it}$. A more general model is the **two-way effects model**, with $u_{it} = \alpha_i + \gamma_t + \varepsilon_{it}$, which additionally allows for time-specific effects. Then

$$y_{it} = \alpha_i + \gamma_t + \mathbf{x}'_{it}\beta + \varepsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T. \tag{21.53}$$

This model was presented originally in (21.2).

As already noted, for short panels the usual approach is to treat the time-specific effects as fixed and estimate them as the coefficients of time dummies that are included in the regressors, with analysis then differing according to whether the individual-specific effects are treated as fixed or random.

If both α_i and γ_t are fixed then the OLS estimator of β in (21.53) is equivalent to regression of $y_{it} - \bar{y}_i - \bar{y}_t + \bar{\bar{y}}$ on $\mathbf{x}_{it} - \bar{\mathbf{x}}_i - \bar{\mathbf{x}}_t + \bar{\bar{\mathbf{x}}}$, where $\bar{y}_i = T^{-1}\sum_{t=1}^{T} y_{it}$, $\bar{y}_t = N^{-1}\sum_{i=1}^{N} y_{it}$, and $\bar{\bar{y}} = (NT)^{-1}\sum_{i=1}^{N}\sum_{i=1}^{T} y_{it}$, with similar definitions for $\bar{\mathbf{x}}_i$, $\bar{\mathbf{x}}_t$, and $\bar{\bar{\mathbf{x}}}$. This method of estimation is convenient if T is large.

If instead both α_i and γ_t are random then the error term will have a component γ_t that induces error correlation across individuals, whereas we have focused on independence over i. It can be shown that the GLS estimator can be computed by OLS estimation of y^*_{it} on a constant and $\mathbf{x}^*_{it}$,

$$y^*_{it} = y_{it} - \lambda_1\bar{y}_i - \lambda_2\bar{y}_t + \lambda_3\bar{\bar{y}},$$

where $\bar{y}_i$, $\bar{y}_t$, and $\bar{\bar{y}}$ have already been defined and $\mathbf{x}^*_{it}$ is defined analogously to y^*_{it}. For this and other results for the two-way effects model see Hsiao (2003) or Baltagi (2001).

21.8.5. Unbalanced Panel Data

The discussion thus far has assumed the panel is **balanced**, meaning that data are available for every individual in every year. For panel data on different regions this is often the case. In contrast, for panel surveys of individuals there is usually a drop off or **attrition** over time in the proportion of individuals still answering the survey. Moreover, some individuals may miss one or more periods but return later, in some cases by design as in the case of **rotating panels** such as the CPS, where households are surveyed for four consecutive months, not surveyed for eight months, and then surveyed for another four months. Such panels where different individuals appear in different years are called **unbalanced panels** or **incomplete panels.**

Let d_{it} be an indicator variable equal to one if the itth observation is observed and equal to zero otherwise. Then for the individual-specific effects model (21.3) the FE estimator is consistent if the strong exogeneity assumption (21.4) becomes

$$\mathrm{E}[u_{it}|\alpha_i, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}, d_{i1}, \ldots, d_{iT}] = 0, \tag{21.54}$$

and the RE estimator is consistent if additionally α_i is independent of the other conditioning variables. The fixed and random effects estimators can then be applied to unbalanced data with relatively little adjustment. This should be clear from the initial presentation of the estimators as OLS estimators in various models given in Section 21.2.2. For example, for the random effects model replace $\widehat{\lambda}$ in (21.10) by $\widehat{\lambda}_i = 1 - \sigma_\varepsilon/(T_i\sigma_\alpha^2 + \sigma_\varepsilon^2)^{1/2}$, where T_i is the number of observations for individual i (see Baltagi, 1985, and Wansbeek and Kapteyn, 1989). Davis (2002) considers multi-way random effects models. For the fixed effects model an individual observation must be observed at least twice in the sample and degrees of freedom must be appropriately adjusted. Baltagi (2001) gives a lengthy treatment of unbalanced panels. Econometrics packages that estimate the more standard of the panel models presented in Chapters 21–23 usually automatically handle missing observations.

At times it may be convenient to convert an unbalanced panel into a balanced panel, by including in the sample only those individuals with data available in all years. This obviously can greatly reduce efficiency because of the loss of many observations. Furthermore, if data are not randomly missing this can exacerbate potential problems of a nonrepresentative sample.

One reason for missing data can be that although most variables are observed, at least one variable is not. For example, the **nonresponse rate** to income questions can be quite high. Rather than drop an entire observation because data for one regressor, income, is missing there may be efficiency gains to using the imputation methods presented in Chapter 27.

Unbalanced panels require special methods if the reason for individuals dropping out of the sample is correlated with the error term, so that (21.54) does not hold. For example, those individuals with unusually low wages (after controlling for observed characteristics) may be more likely to drop out of a panel sample. The result is an unrepresentative panel that will lead to **attrition bias** if wage is the dependent variable. Consistent estimation requires use of sample selection methods extended to panel data (see Section 23.5.2).

21.8.6. Measurement Error

Measurement error in regressors leads to inconsistent parameter estimates in cross-section regression models. If panel data methods are used that involve differencing of the data, the result may be a large increase in the inconsistency caused by measurement error depending on the assumptions made about the dgp. This is pursued in Chapter 26.

21.9. Practical Considerations

The various estimators presented in this chapter are easily implemented. The most foolproof method is to use the panel commands available in econometric packages such as LIMDEP, STATA, and TSP, all of which have the added advantage of usually handling unbalanced panels. Most estimators can alternatively be estimated using an appropriate pooled OLS regression on transformed data that requires only a cross-section package, though standard errors may then differ from panel package standard errors because the latter may ignore autocorrelation induced by transformation and may use different degrees of freedom.

A weakness of panel commands in packages is that they currently compute standard errors based on restrictive distributional assumptions such as iid errors in the fixed effects models, and iid individual effect and iid errors in the random effects model. To compute the more robust standard error estimates presented in this chapter may require panel estimation with a panel bootstrap or estimation of an appropriate pooled OLS regression using an option to compute cluster-robust standard errors.

In microeconometric analysis there is a fundamental distinction between models with and models without fixed effects. If a model without fixed effects is preferred it should be justified by passing a Hausman test. If this test rejects the random effects model then it may still be possible to consistently estimate coefficients of time-invariant regressors using the instrumental variables methods presented in the next chapter.

21.10. Bibliographic Notes

Most textbooks, such as Greene's (2003), include at least a chapter on panel data models. Wooldridge (2002) has several chapters that cover both linear and nonlinear panel models. Econometrics monographs on panel data include those by Hsiao (1986, 2003), Baltagi (1995, 2001), Matyas and Sevestre (1995), M-J. Lee (2002), and Arellano (2003). The last three books place greater emphasis on the methods presented in Chapter 22 and 23. Diggle, Liang, and Zeger (1994, 2002) is a standard statistics reference.

21.4 Mundlak (1978) wrote a classic article on fixed versus random effects models. Hausman (1978) used tests between these two models to illustrate his testing approach.

21.6 Kuh (1959) and Hoch (1962) provide two early panel data applications to estimation of investment functions and of Cobb–Douglas production functions. These studies contrast use of within estimates using time-series variation and between estimates using cross-section variation.

Exercises

21–1 (Adapted from Baltagi, 1999) Consider the panel model $y_{it} = \alpha + \beta x_{it} + u_{it}$, where α and β are scalars.

(a) Show by appropriate subtraction that this model implies

$$y_{it} - \bar{y} = \beta(x_{it} - \bar{x}_i) + \beta(\bar{x}_i - \bar{x}) + (u_{it} - \overline{u}),$$

where $\bar{y} = (NT)^{-1} \sum_{i,t} y_{it}$, $\bar{x} = (NT)^{-1} \sum_{i,t} x_{it}$ and $\bar{x}_i = T^{-1} \sum_t x_{it}$.

(b) For the corresponding unrestricted least-squares regression

$$y_{it} - \bar{y} = \beta_1(x_{it} - \bar{x}_i) + \beta_2(\bar{x}_i - \bar{x}) + (u_{it} - \overline{u}),$$

show that the least-squares estimator of β_1 is the within estimator and that of β_2 is the between estimator.

(c) Show that if $u_{it} = \mu_i + v_{it}$, where μ_i ~iid$[0, \sigma_\mu^2]$ and v_{it} ~iid$[0, \sigma_v^2]$, and the two are mutually independent across both i and t, the OLS and the GLS estimators are equivalent.

21–2 Consider estimation of the fixed effects linear regression model $y_{it} = \alpha_i + \mathbf{x}_{it}'\beta + \varepsilon_{it}$, where α_i are fixed effects possibly correlated with $\mathbf{x}_{it}$. Stacking all T observations for individual i yields $\mathbf{y}_i = \alpha_i \mathbf{e} + \mathbf{X}_i\beta + \varepsilon_i$ (see (21.29) for definitions). Consider the estimator $\widehat{\beta} = [\sum_{i=1}^N \mathbf{X}_i'\mathbf{J}'\mathbf{J}\mathbf{X}_i]^{-1} \times \sum_{i=1}^N \mathbf{X}_i'\mathbf{J}'\mathbf{J}\mathbf{y}_i$, where $\mathbf{J}$ is a $T \times T$ matrix of known constants such that $\mathbf{Je} = \mathbf{0}$. [Note that an example of $\mathbf{J}$ is $\mathbf{Q} = \mathbf{I}_T - T^{-1}\mathbf{ee}'$.]

(a) Provide a motivation for the estimator $\widehat{\beta}$.

(b) Find $\mathrm{E}[\widehat{\beta}]$. For simplicity assume that $\mathbf{X}_i$ are fixed regressors and that ε_{it} are iid $[0, \sigma^2]$. Is $\widehat{\beta}$ unbiased for β?

(c) Find $\mathrm{V}[\widehat{\beta}]$. For simplicity assume that $\mathbf{X}_i$ are fixed regressors and that ε_{it} are iid $[0, \sigma^2]$.

(d) Now suppose ε_{it} are independent over i but correlated over t with $\mathrm{V}[\varepsilon_i] = \Omega_i$. Give $\mathrm{V}[\widehat{\beta}]$.

(e) Suppose that the effects α_i are random $(0, \sigma_\alpha^2)$ rather than fixed. Would the estimator in this exercise be consistent?

21–3 (Adapted from Baltagi, 1998) Consider the fixed effects, two-way error component panel data model

$$y_{it} = \alpha + \mathbf{x}_{it}'\beta + \mu_i + \lambda_t + \varepsilon_{it},$$

where α is a scalar, $\mathbf{x}_{it}$ is a $k \times 1$ vector of exogenous regressors, β is a $K \times 1$ vector, μ and λ denote fixed individual and time effects, respectively, and $\varepsilon_{it} \sim$ iid$[0, \sigma^2]$.

(a) Show that the within estimator of β, which is best linear unbiased, can be obtained by applying two within (one-way) transformations on this model. The first is the within transformation ignoring the time effects followed by the within transformation ignoring the individual effects.

(b) Show that the order of these two within (one-way) transformations is unimportant. Give an intuitive explanation for this result.

21–4 Use a 50% random subsample of the wage–hours data in Section 21.3

(a) Can β be directly interpreted as a labor supply elasticity? Explain.

(b) For the following estimators: (1) pooled OLS, (2) between, (3) within, (4) first differences, (5) random effects GLS, (6) random effects MLE give (i) $\widehat{\beta}$ (estimated coefficient of lnwg), (ii) default standard error, and (iii) panel bootstrap standard error with 200 replications.

(c) Are the estimates of β similar?

(d) Is there a systematic difference between default standard errors and panel-robust standard errors?

(e) Will the pooled OLS estimator in part (b) be consistent for β in a fixed effects model? Will the pooled OLS estimator be consistent for β in a random effects model?

(f) Perform a Hausman test of the difference between the fixed and random effects (GLS) estimates of β in this model. Do this manually using the earlier regression output with the default standard errors. What do you conclude and which model is favored?

(g) Given the preceding evidence, do you believe that the labor supply curve is upward sloping? Explain.

Less consideration has been given to nonstationary data in **short panels**. Harris and Tzavalis (1999) consider the unit root tests of Levin and Lin (1992) in short panels. Let $\widehat{\gamma}$ denote the within estimate of γ in the AR(1) fixed effects model $y_{it} = \alpha_i + \gamma y_{i,t-1} + \varepsilon_{it}$, where $\varepsilon_{it} \sim \mathcal{N}[0, \sigma^2]$. We consider the null hypothesis of a unit root, so $\gamma = 1$, and no intercept $\alpha_i = 0$, which corresponds to the pure time series case 2 in Hamilton (1994, p. 490). Under this null hypothesis the unit root test statistic

$$\frac{\sqrt{N}(\widehat{\gamma} - 1 + 3/(T+1))}{[3(17T^2 - 20T + 17)]/[5(T-1)(T+1)^3]} \xrightarrow{d} \mathcal{N}[0, 1]$$

as $N \to \infty$ for fixed T. Large negative values of this statistic lead to rejection of the unit root hypothesis. Levin and Lin (1992) provide additional tests, such as for models with individual time trends.

Binder, Hsiao, and Pesaran (2003) consider short panel estimation of fixed effect dynamic panel models with unit roots and cointegration. With unit roots the Arellano–Bond estimator is inconsistent, though the extensions due to Ahn and Schmidt (1995) and others discussed at the end of Section 22.5.3 yield consistent estimates. Binder et al. (2003) propose quasi-ML estimators that perform better in finite samples when unit roots are present.

22.6. Difference-in-Differences Estimator

The evaluation literature presented in Chapter 25 focuses on measuring the **treatment effect**, in the simplest case the impact or marginal effect of a single binary regressor that equals one if treatment occurs and equals zero if treatment does not occur. For example, interest may lie in measuring the effect on earnings of a policy change (the binary treatment) that alters tax rates or welfare eligibility or access to training for some individuals but not for others.

In this section we relate one of the methods of Chapter 25 to panel methods. Specifically the treatment effect can be measured using standard panel data methods if panel data are available before and after the treatment and if not all individuals receive the treatment. Then the first-differences estimator for the fixed effects model reduces to a simple estimator called the differences-in-differences estimator, introduced in Section 3.4.2 and also studied in Section 25.5. The latter estimator has the advantage that it can also be used when repeated cross-section data rather than panel data are available. However, it does rely on model assumptions that are often not made explicit. The treatment here follows Blundell and MaCurdy (2000).

22.6.1. Fixed Effects with Binary Treatment

Let the binary regressor of interest be

$$D_{it} = \begin{cases} 1 \text{ if individual } i \text{ receives treatment in period } t, \\ 0 \text{ otherwise.} \end{cases} \tag{22.39}$$

Assume a fixed effects model for y_{it} with

$$y_{it} = \phi D_{it} + \delta_t + \alpha_i + \varepsilon_{it}, \tag{22.40}$$

where δ_t is a time-specific fixed effect and α_i is an individual-specific fixed effect. As noted in Section 21.2.1 this is equivalent to regression of y_{it} on D_{it} and a full set of time dummies with the complication of individual-specific fixed effects. For simplicity there are no other regressors.

The individual effects α_i can be eliminated by first differencing. Then

$$\Delta y_{it} = \phi \Delta D_{it} + (\delta_t - \delta_{t-1}) + \Delta \varepsilon_{it}. \tag{22.41}$$

The treatment effect ϕ can be consistently estimated by pooled OLS regression of Δy_{it} on ΔD_{it} and a full set of time dummies.

22.6.2. Differences in Differences

Now consider specialization to only two time periods. Furthermore, suppose treatment occurs only in period 2, so that in period 1 $D_{i1} = 0$ for all individuals and in period 2 $D_{i2} = 1$ for the treated and $D_{i2} = 0$ for the nontreated. Then the subscript t can be dropped from (22.41) and

$$\Delta y_i = \phi D_i + \delta + v_i, \tag{22.42}$$

where D_i is a binary treatment variable indicating whether or not the individual received treatment.

The treatment effect can be estimated by OLS regression of Δy on an intercept and the binary regressor D. Define $\Delta \bar{y}^{\text{tr}}$ to denote the sample average of Δy_i for the treated $(D_i = 1)$ and $\Delta \bar{y}^{\text{nt}}$ to denote the sample average of Δy_i for the nontreated $(D_i = 0)$. Then the OLS estimator reduces to

$$\widehat{\phi} = \Delta \bar{y}^{\text{tr}} - \Delta \bar{y}^{\text{nt}}. \tag{22.43}$$

This estimator is called the **differences-in-differences (DID) estimator**, since one estimates the time difference for the treated and untreated groups and then takes the difference in the time differences.

The estimator is appealing for its intuitive simplicity. Additionally, it can be extended from panel data to the case where separate cross sections are available in the two periods. In the second period compute the averages $\bar{y}_2^{\text{tr}}$ and $\bar{y}_2^{\text{nt}}$ for the treated and untreated groups. Compute similar averages $\bar{y}_1^{\text{tr}}$ and $\bar{y}_1^{\text{nt}}$ in the first pretreatment period. This assumes that it is possible to identify in the first period whether or not an individual is eligible for treatment. This is easy if, for example, the treatment applies only to women and data on gender are available. Then compute

$$\widehat{\phi} = (\bar{y}_2^{\text{tr}} - \bar{y}_1^{\text{tr}}) - (\bar{y}_2^{\text{nt}} - \bar{y}_1^{\text{nt}}). \tag{22.44}$$

As an example, if average annual earnings for the group eligible for treatment equals 10,000 before treatment and 13,000 after treatment then $\bar{y}_2^{\text{tr}} - \bar{y}_1^{\text{tr}} = 3{,}000$. Similarly, if average annual earnings for the group not eligible for treatment equals 15,000 before treatment and 17,000 after treatment then $\bar{y}_2^{\text{nt}} - \bar{y}_1^{\text{nt}} = 2{,}000$. The DID estimate of the treatment effect $\widehat{\phi}$ is then $3{,}000 - 2{,}000 = 1{,}000$.

22.6.3. Assumptions Underlying Differences in Differences

The preceding formulation of the DID estimator makes explicit the underlying assumptions for consistent estimation of ϕ.

First, it is assumed that the time effects δ_t are common across treated and untreated individuals. For example, time trends may differ by gender, in which case identifying ϕ is problematic if treatment depends on gender. The common trends assumption is needed if either panel or cross-section data are used.

Second, if cross-section data are used then the composition of the treated and untreated groups is assumed to be stable before and after the change. With panel data differencing eliminates the fixed effects α_i. With repeated cross-section data the original model (22.40) implies that $\bar{y}_t^{\text{tr}} = \phi + \delta_t + \bar{\alpha}_t^{\text{tr}} + \bar{\varepsilon}_t^{\text{tr}}$ and $\bar{y}_t^{\text{nt}} = \delta_t + \bar{\alpha}_t^{\text{nt}} + \bar{\varepsilon}_t^{\text{nt}}$. Given that treatment only occurs in the second period it follows that

$$\phi = (\bar{y}_2^{\text{tr}} - \bar{y}_1^{\text{tr}}) - (\bar{y}_2^{\text{nt}} - \bar{y}_1^{\text{nt}}) + (\bar{\alpha}_2^{\text{tr}} - \bar{\alpha}_1^{\text{tr}}) - (\bar{\alpha}_2^{\text{nt}} - \bar{\alpha}_1^{\text{nt}}) + v,$$

where $v = (\bar{\varepsilon}_2^{\text{nt}} - \bar{\varepsilon}_1^{\text{nt}}) - (\bar{\varepsilon}_2^{\text{nt}} - \bar{\varepsilon}_1^{\text{nt}})$. Consistency of $\widehat{\phi}$ in (22.44) occurs if $\text{plim}(\bar{\alpha}_2^{\text{tr}} - \bar{\alpha}_1^{\text{tr}}) = 0$ and $\text{plim}(\bar{\alpha}_2^{\text{nt}} - \bar{\alpha}_1^{\text{nt}}) = 0$. This will happen if assignment to treatment is random. However, often this is not the case.

22.6.4. Richer Models

In practice richer models are used. An obvious extension is to include regressors other than the treatment indicator and time dummies. By grouping data the individual-specific effects can at least be permitted to differ on average across groups. The general procedure is to estimate

$$y_{igt} = \phi D_{igt} + \delta_t + \alpha_i + \varepsilon_{it},$$

where g denotes the gth group.

In a classic example of DID estimation, Card (1990) studied the effect on unemployment of low-wage workers in Miami of a sudden influx of immigrants from Cuba. This example is also reviewed in Angrist and Krueger (1999). Athey and Imbens (2002) present extension to nonlinear models.

22.7. Repeated Cross Sections and Pseudo Panels

The key potential advantages of panel data arise from being able to observe subjects over time. This makes it possible to control for unobserved individual heterogeneity, differences in initial conditions, and dynamic dependence of outcomes. In many cases, however, genuine panel data are unavailable.

22.7.1. Repeated Cross Sections

We consider analysis when data are for several **repeated cross sections**, derived from responses to a series of independent sample surveys, where independence means that