11-1

11

MODELS FOR PANEL DATA

11.1 INTRODUCTION

Data sets that combine time series and cross sections are common in economics. The published statistics of the OECD contain numerous series of economic aggregates observed yearly for many countries. The Penn World Tables [CIC (2010)] is a data bank that contains national income data on 188 countries for over 50 years. Recently constructed **longitudinal data sets** contain observations on thousands of individuals or families, each observed at several points in time. Other empirical studies have examined time series data on sets of firms, states, countries or industries simultaneously. These data sets provide rich sources of information about the economy. The analysis of panel data allows the model builder to learn about economic processes while accounting for both heterogeneity across individuals, firms, countries, etc. and for dynamic effects that are not visible in cross sections. Modeling in this context often calls for complex stochastic specifications. In this chapter, we will survey the most commonly used techniques for time series – cross section (e.g., cross country) and panel (e.g., longitudinal) data. The methods considered here provide extensions to most of the models we have examined in the preceding chapters. Section 11.2 describes the specific features of panel data. Most of this analysis is focused on individual data, rather than cross country aggregates. We will examine some aspects of aggregate modeling in Section 11.11. Sections 11.3, 11.4 and 11.5 consider in turn the three main approaches to regression analysis with panel data, pooled regression, the fixed effects model and the random effects model. Section 11.6 considers robust estimation of covariance matrices for the panel data estimators, including a general treatment of "cluster" effects. Sections 11.7 – 11.11 examine some specific applications and extensions of panel data methods. Spatial autocorrelation is discussed in Section 11.7. In Section 11.8, we consider sources of endogeneity in the random effects model, including a model of the sort considered in Chapter 8 with an endogenous right hand side variable, then two approaches to dynamic models. Section 11.9 builds the fixed and random effects models into nonlinear regression models. In Section 11.10, the random effects model is extended to the multiple equation systems developed in Chapter 10. Finally, Section 11.11 examines random parameter models. The random parameters approach is an extension of the fixed and random effects model in which the heterogeneity that the FE and RE models build into the constant terms ~~of the models~~ is extended to other parameters as well.

and so on

AU: OK to spell out "etc."?

data

and

Panel data methods are used throughout the remainder of this book. We will develop several extensions of the fixed and random effects models in Chapter 14 on maximum likelihood methods, and in Chapter 15 where we will continue the development of random parameter models that is begun in Section 11.11. Chapter 14 will also present methods for handling discrete distributions of random parameters under the heading of latent class models. In Chapter 23, we will return to the models of nonstationary panel data that are suggested in Section 11.8.4. The fixed and random effects approaches will be used throughout the applications of discrete and limited dependent variables models in microeconometrics in Chapters 17, 18 and 19.

11-2

~~9~~ 11

MODELS FOR PANEL DATA

9.1 INTRODUCTION

Data sets that combine time series and cross sections are common in economics. For example, the published statistics of the OECD contain numerous series of economic aggregates observed yearly for many countries. Recently constructed **longitudinal data sets** contain observations on thousands of individuals or families, each observed at several points in time. Other empirical studies have analyzed time-series data on sets of firms, states, countries, or industries simultaneously. These data sets provide rich sources of information about the economy. Modeling in this setting, however, calls for some complex stochastic specifications. In this chapter, we will survey the most commonly used techniques for time-series cross-section data analyses in single equation models.

9.2 PANEL DATA MODELS

Many recent studies have analyzed **panel,** or longitudinal, data sets. Two very famous ones are the National Longitudinal Survey of Labor Market Experience (NLS, http://www.bls.gov/nls/nlsdoc.htm) and the Michigan Panel Study of Income Dynamics (PSID, http://psidonline.isr.umich.edu/). In these data sets, very large cross sections, consisting of thousands of microunits, are followed through time, but the number of periods is often quite small. The PSID, for example, is a study of roughly 6,000 families and 15,000 individuals who have been interviewed periodically from 1968 to the present. An ongoing study in the United Kingdom is the British Household Panel Survey (BHPS, http://www.iser.essex.ac.uk/ulsc/bhps/) which was begun in 1991 and is now in its 15th wave. The survey follows several thousand households (currently over 5,000) for several years. Many very rich data sets have recently been developed in the area of health care and health economics, including the German Socioeconomic Panel (GSOEP, http://dpls.dacc.wisc.edu/apdu/GSOEP/gsoep_cd_data.html) and the Medical Expenditure Panel Survey (MEPS, http://www.meps.ahrq.gov/). Constructing long, evenly spaced time series in contexts such as these would be prohibitively expensive, but for the purposes for which these data are typically used, it is unnecessary. Time effects are often viewed as "transitions" or discrete changes of state. The Current Population Survey (CPS, http://www.census.gov/cps/), for example, is a monthly survey of about 50,000 households that interviews households monthly for four months, waits for eight months, then reinterviews. This two-wave, **rotating panel** format allows analysis of short-term changes as well as a more general analysis of the U.S. national labor market. They are typically modeled as specific to the period in which they occur and are not carried across periods within a cross-sectional unit.[1] Panel data sets are more oriented toward cross-section analyses; they are wide but typically short. Heterogeneity across units is an integral part—indeed, often the central focus—of the analysis.

The analysis of panel or longitudinal data is the subject of one of the most active and innovative bodies of literature in econometrics,[2] partly because panel data provide such a rich environment for the development of estimation techniques and theoretical results. In more practical terms, however, researchers have been able to use time-series cross-sectional data to examine issues that could not be studied in either cross-sectional or time-series settings alone. Two examples are as follows.

1. In a widely cited study of labor supply, Ben-Porath (1973) observes that at a certain point in time, in a cohort of women, 50 percent may appear to be working. It is ambiguous whether this finding implies that, in this cohort, one-half of the women on average will be working or that the same one-half will be working in every period. These have very different implications for policy and for the interpretation of any statistical results. Cross-sectional data alone will not shed any light on the question.

2. A long-standing problem in the analysis of production functions has been the inability to separate economies of scale and technological change.[3] Cross-sectional data provide information only about the former, whereas time-series data muddle the two effects, with no prospect of separation. It is common, for example, to assume constant returns to scale so as to reveal the technical change.[4] Of course, this practice assumes away the problem. A panel of data on costs or output for a number of firms each observed over several years can provide estimates of both the rate of technological change (as time progresses) and economies of scale (for the sample of different sized firms at each point in time).

Recent applications have allowed researchers to study the impact of health policy changes [e.g., Riphahn et al.'s (2003) analysis of reforms in German public health insurance regulations] and more generally the dynamics of labor market behavior. In principle, the methods of Chapters 6 and ~~20~~ 21 can be applied to longitudinal data sets. In the typical panel, however, there are a large number of cross-sectional units and

[1]Formal time-series modeling for panel data is briefly examined in Section ~~22~~ 23.5.

[2]The panel data literature rivals the received research on unit roots and cointegration in econometrics in its rate of growth. A compendium of the earliest literature is Maddala (1993). Book-length surveys on the econometrics of panel data include Hsiao (2003), Dielman (1989), Matyas and Sevestre (1996), Raj and Baltagi (1992), Nerlove (2002), Arellano (2003), and Baltagi (2001, 2005). There are also lengthy surveys devoted to specific topics, such as limited dependent variable models [Hsiao, Lahiri, Lee, and Pesaran (1999)] and semiparametric methods [Lee (1998)]. An extensive bibliography is given in Baltagi (2005).

[3]The distinction between these two effects figured prominently in the policy question of whether it was appropriate to break up the AT&T Corporation in the 1980s and, ultimately, to allow competition in the provision of long-distance telephone service.

[4]In a classic study of this issue, Solow (1957) states: "From time series of $\Delta Q/Q$, w_K, $\Delta K/K$, w_L and $\Delta L/L$ or their discrete year-to-year analogues, we could estimate $\Delta A/A$ and thence $A(t)$ itself. Actually an amusing thing happens here. Nothing has been said so far about returns to scale. But if all factor inputs are classified either as K or L, then the available figures always show w_K and w_L adding up to one. Since we have assumed that factors are paid their marginal products, this amounts to assuming the hypothesis of Euler's theorem. The calculus being what it is, we might just as well assume the conclusion, namely, the F is homogeneous of degree one."

11-4

only a few periods. Thus, the time-series methods discussed there may be somewhat problematic. Recent work has generally concentrated on models better suited to these short and wide data sets. The techniques are focused on cross-sectional variation, or heterogeneity. In this chapter, we shall examine in detail the most widely used models and look briefly at some extensions.

11 ~~9~~.2.1 GENERAL MODELING FRAMEWORK FOR ANALYZING PANEL DATA

The fundamental advantage of a panel data set over a cross section is that it will allow the researcher great flexibility in modeling differences in behavior across individuals. The basic framework for this discussion is a regression model of the form

$$
\begin{aligned}
y_{it} &= \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{z}_i'\boldsymbol{\alpha} + \varepsilon_{it} \\
&= \mathbf{x}_{it}'\boldsymbol{\beta} + c_i + \varepsilon_{it}.
\end{aligned}
\tag{11 ~~9~~-1}
$$

There are K regressors in $\mathbf{x}_{it}$, *not including a constant term*. KT The **heterogeneity**, or KT **individual effect** is $\mathbf{z}_i'\boldsymbol{\alpha}$ where $\mathbf{z}_i$ contains a constant term and a set of individual or group-specific variables, which may be observed, such as race, sex, location, etc, or unobserved, and so on such as family specific characteristics, individual heterogeneity in skill or preferences, and so on, all of which are taken to be constant over time t. As it stands, this model is a classical regression model. If $\mathbf{z}_i$ is observed for all individuals, then the entire model can be treated as an ordinary linear model and fit by least squares. The complications arise when c_i is unobserved, which will be the case in most applications. Consider, for example, analyses of the effect of education and experience on earnings from which "ability" will always be a missing and unobservable variable. In health care studies, for example, of usage of the health care system, "health" and "health care" will be unobservable factors in the analysis.

AU: OK to spell out "etc." in text?

KT The main objective of the analysis will be consistent and efficient estimation of the **partial effects,**

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

KT Whether this is possible depends on the assumptions about the unobserved effects. We begin with a **strict exogeneity** assumption for the independent variables,

$$
E[\varepsilon_{it}\,|\,\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots,] = 0.
$$

That is, the current disturbance is uncorrelated with the independent variables in every period, past, present and future. The crucial aspect of the model concerns the heterogeneity. A particularly convenient assumption would be **mean independence,** KT

$$
E[c_i\,|\,\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots] = \alpha.
$$

If the missing variable(s) are uncorrelated with the included variables, then, as we shall see, they may be included in the disturbance of the model. This is the assumption that underlies the random effects model, as we will explore below. It is, however, a particularly strong assumption—it would be unlikely in the labor market and health later

11-5

care examples mentioned previously above. The alternative would be

$$E[c_i \mid \mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots,] = h(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots)$$
$$= h(\mathbf{X}_i).$$

This formulation is more general, but at the same time, considerably more complicated, the more so since it may require yet further assumptions about the nature of the function.

9.2.2 MODEL STRUCTURES

We will examine a variety of different models for panel data. Broadly, they can be arranged as follows:

1. **Pooled Regression:** If $\mathbf{z}_i$ contains only a constant term, then ordinary least squares provides consistent and efficient estimates of the common α and the slope vector $\boldsymbol{\beta}$.

2. **Fixed Effects:** If $\mathbf{z}_i$ is unobserved, but correlated with $\mathbf{x}_{it}$, then the least squares estimator of $\boldsymbol{\beta}$ is biased and inconsistent as a consequence of an omitted variable. However, in this instance, the model

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

KT where $\alpha_i = \mathbf{z}_i'\boldsymbol{\alpha}$, embodies all the observable effects and specifies an estimable conditional mean. This **fixed effects** approach takes α_i to be a group-specific constant term in the regression model. It should be noted that the term "fixed" as used here signifies the correlation of c_i and $\mathbf{x}_{it}$, not that c_i is nonstochastic.

3. **Random Effects:** If the unobserved individual heterogeneity, however formulated, can be assumed to be uncorrelated with the included variables, then the model may be formulated as

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + E[\mathbf{z}_i'\boldsymbol{\alpha}] + \{\mathbf{z}_i'\boldsymbol{\alpha} - E[\mathbf{z}_i'\boldsymbol{\alpha}]\} + \varepsilon_{it}$$
$$= \mathbf{x}_{it}'\boldsymbol{\beta} + \alpha + u_i + \varepsilon_{it},$$

that is, as a linear regression model with a compound disturbance that may be consistently, albeit inefficiently, estimated by least squares. This **random effects** approach specifies that u_i is a group-specific random element, similar to ε_{it} except that for each group, there is but a single draw that enters the regression identically in each period. Again, the crucial distinction between fixed and random effects is whether the unobserved individual effect embodies elements that are correlated with the regressors in the model, not whether these effects are stochastic or not. We will examine this basic formulation, then consider an extension to a dynamic model. KT

4. **Random Parameters:** The random effects model can be viewed as a regression model with a random constant term. With a sufficiently rich data set, we may extend this idea to a model in which the other coefficients vary randomly across individuals as well. The extension of the model might appear as

$$y_{it} = \mathbf{x}_{it}'(\boldsymbol{\beta} + \mathbf{h}_i) + (\alpha + u_i) + \varepsilon_{it},$$

where $\mathbf{h}_i$ is a random vector that induces the variation of the parameters across individuals. This random parameters model was proposed quite early in this literature, but has only fairly recently enjoyed widespread attention in several fields. It represents a natural extension in which researchers broaden the amount of heterogeneity across individuals while retaining some commonalities—the parameter vectors still share a common mean. Some recent applications have extended this yet another step by allowing the mean value of the parameter distribution to be person-specific, as in

$$y_{it} = \mathbf{x}_{it}'(\boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \mathbf{h}_i) + (\alpha + u_i) + \varepsilon_{it},$$

where $\mathbf{z}_i$ is a set of observable, person specific variables, and $\boldsymbol{\Delta}$ is a matrix of parameters to be estimated. As we will examine ~~later~~, this **hierarchical model** is extremely versatile. KT

in Chapter 17

11 ~~9~~.2.3 EXTENSIONS

The short list of model types provided earlier only begins to suggest the variety of applications of panel data methods in econometrics. We will begin in this chapter to study some of the formulations and uses of linear models. The random and fixed effects models and random parameters models have also been widely used in models of censoring, binary, and other discrete choices, and models for event counts. We will examine all of these in the chapters to follow. In some cases, such as the models for count data in Chapter ~~25~~ 19 the extension of random and fixed effects models is straightforward, if somewhat more complicated computationally. In others, such as in binary choice models in Chapter ~~23~~ 17 and censoring models in Chapter ~~24~~ 18, these panel data models have been used, but not before overcoming some significant methodological and computational obstacles.

11 ~~9~~.2.4 BALANCED AND UNBALANCED PANELS

discussion

By way of preface to the analysis to follow, we note an important aspect of panel data analysis. As suggested by the preceding, a "panel" data set will consist of n sets of observations on individuals to be denoted $i = 1, \ldots, n$. If each individual in the data set is observed the same number of times, usually denoted T, the data set is a **balanced panel.** An **unbalanced panel** data set is one in which individuals may be observed different numbers of times. We will denote this T_i. A **fixed panel** is one in which the same set of individuals is observed for the duration of the study. The data sets we will examine in this chapter, while not all balanced, are fixed. A rotating panel is one in which the cast of individuals changes from one period to the next. For example, Gonzalez and Maloney (1999) examined self-employment decisions in Mexico using the National Urban Employment Survey. This is a quarterly data set drawn from 1987 to 1993 in which individuals are interviewed five times. Each quarter, one-fifth of the individuals is rotated out of the data set. We will not treat rotating panels in this text. Some discussion and numerous references may be found in Baltagi (2005).

The development to follow is structured so that the distinction between balanced and unbalanced panels will entail nothing more than a trivial change in notation—where for convenience we write T suggesting a balanced panel, merely changing T to T_i generalizes the results. We will note specifically when this is not the case, such as in Breusch and Pagan's (1980) LM statistic.

11-7

In Chapter 16, we will show that if the disturbances are normally distributed, then the least squares estimator is also the **maximum likelihood estimator.** Maximum likelihood estimators are asymptotically efficient among consistent and asymptotically normally distributed estimators. This gives us a partial result, albeit a somewhat narrow one since to claim it, we must assume normally distributed disturbances. If some other distribution is specified for ε and it emerges that $\mathbf{b}$ is not the maximum likelihood estimator, then least squares may not be efficient.

Example 4.9 The Gamma Regression Model

Greene (1980a) considers estimation in a regression model with an asymmetrically distributed disturbance,

$$y = (\alpha + \sigma\sqrt{P}) + \mathbf{x}'\beta + (\varepsilon - \sigma\sqrt{P}) = \alpha^* + \mathbf{x}'\beta + \varepsilon^*,$$

where ε has the gamma distribution in Section B.4.5 [see (B-39)] and $\sigma = \sqrt{P}/\lambda$ is the standard deviation of the disturbance. In this model, the covariance matrix of the least squares estimator of the slope coefficients (not including the constant term) is,

$$\text{Asy. Var}[\mathbf{b} \mid \mathbf{X}] = \sigma^2(\mathbf{X}'\mathbf{M}^0\mathbf{X})^{-1},$$

whereas for the maximum likelihood estimator (which is not the least squares estimator),[14]

$$\text{Asy. Var}[\hat{\beta}_{ML}] \approx [1 - (2/P)]\sigma^2(\mathbf{X}'\mathbf{M}^0\mathbf{X})^{-1}.$$

But for the asymmetry parameter, this result would be the same as for the least squares estimator. We conclude that the estimator that accounts for the asymmetric disturbance distribution is more efficient asymptotically.

11.2.5 ~~4.9.6~~ Well Behaved Panel Data ~~MORE GENERAL DATA GENERATING PROCESSES~~

The asymptotic properties of the estimators in the classical regression model were established in Section 4.4 under the following assumptions:

- **A1.** ***Linearity:*** $y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{iK}\beta_K + \varepsilon_i$.
- **A2.** ***Full rank:*** The $n \times K$ sample data matrix, $\mathbf{X}$ has full column rank.
- **A3.** ***Exogeneity of the independent variables:*** $E[\varepsilon_i \mid x_{j1}, x_{j2}, \ldots, x_{jK}] = 0$, $i, j = 1, \ldots, n$.
- **A4.** ***Homoscedasticity and nonautocorrelation.***
- **A5.** ***Data generating mechanism-independent observations.***

The following are the crucial results needed: For consistency of $\mathbf{b}$, we need ~~(4-21) and (4-24)~~:

$$\text{plim}(1/n)\mathbf{X}'\mathbf{X} = \text{plim}\,\overline{\mathbf{Q}}_n = \mathbf{Q}, \quad \text{a positive definite matrix,}$$

$$\text{plim}(1/n)\mathbf{X}'\varepsilon = \text{plim}\,\overline{\mathbf{w}}_n = E[\overline{\mathbf{w}}_n] = \mathbf{0}.$$

[14] The matrix $\mathbf{M}^0$ produces data in the form of deviations from sample means. (See Section A.2.8.) In Greene's model, P must be greater than 2.

11-8

(For consistency of s^2, we added a fairly weak assumption about the moments of the disturbances.) To establish asymptotic normality, we ~~will require~~ required consistency and ~~(4-32)~~ ~~which is~~

$$\sqrt{n}\,\overline{\mathbf{w}}_n \xrightarrow{d} N[0, \sigma^2\mathbf{Q}].$$

With these in place, the desired characteristics are then established by the methods of Sections ~~4.9.1~~ 4.4.1 and ~~4.9.2~~ 4.4.2. ~~To analyze other cases, we can merely focus on these results. It is not necessary to reestablish the consistency or asymptotic normality themselves, since they follow as a consequence.~~

Au: "panel data" already a KT in chap. Here also?

Exceptions to the assumptions ~~made here~~ are likely to arise in ~~two settings. In~~ a **panel data** set. The sample will consist of multiple observations on each of many observational units. For example, a study might consist of a set of observations made at different points in time on a large number of families. In this case, the $\mathbf{x}$'s will surely be correlated across observations, at least within observational units. They might even be the same for all the observations on a single family. ~~They are also likely to be a mixture of random variables, such as family income, and nonstochastic regressors, such as a fixed "family effect" represented by a dummy variable. The second case would be a time-series model in which lagged values of the dependent variable appear on the right-hand side of the model.~~

The panel data set could be treated as follows. Assume for the moment that the data consist of a fixed number of observations, say T, on a set of N families, so that the total number of rows in $\mathbf{X}$ is $n = NT$. The matrix

$$\overline{\mathbf{Q}}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbf{Q}_i$$

in which n is all the observations in the sample, could be viewed as

$$\overline{\mathbf{Q}}_n = \frac{1}{N}\sum_i \frac{1}{T}\sum_{\substack{\text{observations}\\ \text{for family } i}} \mathbf{Q}_{it} = \frac{1}{N}\sum_{i=1}^{N}\overline{\mathbf{Q}}_i,$$

where $\overline{\mathbf{Q}}_i$ = average $\mathbf{Q}_{it}$ for family i. We might then view the set of observations on the ith unit as if they were a single observation and apply our convergence arguments to the number of families increasing without bound. The point is that the conditions that are needed to establish convergence will apply with respect to the number of observational units. The number of observations taken for each observation unit might be fixed and could be quite small.

~~The second difficult case arises when there are lagged dependent variables among the variables on the right-hand side or, more generally, in time-series settings in which the observations are no longer independent or even uncorrelated. Suppose that the model may be written~~

$$\cancel{y_t = \mathbf{z}_t'\theta + \gamma_1 y_{t-1} + \cdots + \gamma_p y_{t-p} + \varepsilon_t.} \quad \text{(4-37)}$$

~~(Since this model is a time-series setting, we use t instead of i to index the observations.) We continue to assume that the disturbances are uncorrelated across observations. Since y_{t-1} is dependent on y_{t-2} and so on, it is clear that although the disturbances are uncorrelated across observations, the regressor vectors, including the lagged y's surely~~

This chapter will contain relatively little development of the properties of estimators as was done in Chapter 4. We will rely on earlier results in Chapters 4, 8, and 9 and focus instead on a variety of models and specifications.

9.3 THE POOLED REGRESSION MODEL

We begin the analysis by assuming the simplest version of the model, the **pooled model,**

$$
\begin{aligned}
y_{it} &= \alpha + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it},\ i = 1,\ldots,n,\ t = 1,\ldots,T_i,\\
E[\varepsilon_{it}\,|\,\mathbf{x}_{i1},\mathbf{x}_{i2},\ldots,\mathbf{x}_{iT_i}] &= 0,\\
Var[\varepsilon_{it}\,|\,\mathbf{x}_{i1},\mathbf{x}_{i2},\ldots,\mathbf{x}_{iT_i}] &= \sigma_\varepsilon^2,\\
Cov[\varepsilon_{it},\varepsilon_{js}\,|\,\mathbf{x}_{i1},\mathbf{x}_{i2},\ldots,\mathbf{x}_{iT_i}] &= 0 \text{ if } i \neq j \text{ or } t \neq s.
\end{aligned}
\tag{9-2}
$$

(In the panel data context, this is also called the **population averaged model** under the assumption that any latent heterogeneity has been averaged out.) In this form, if the remaining assumptions of the classical model are met (zero conditional mean of ε_{it}, homoscedasticity, independence across observations, i, and strict exogeneity of $\mathbf{x}_{it}$), then no further analysis beyond the results of Chapter 4 is needed. Ordinary least squares is the efficient estimator and inference can reliably proceed along the lines developed in Chapter 5.

9.3.1 LEAST SQUARES ESTIMATION OF THE POOLED MODEL

The crux of the panel data analysis in this chapter is that the assumptions underlying ordinary least squares estimation of the pooled model are unlikely to be met. The question, then, is what can be expected of the estimator when the heterogeneity does differ across individuals? The fixed effects case is obvious. As we will examine later, omitting (or ignoring) the heterogeneity when the fixed effects model is appropriate renders the least squares estimator inconsistent—sometimes wildly so. In the random effects case, in which the true model is

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

where $E[c_i\,|\,\mathbf{X}_i] = \alpha$, we can write the model

$$
\begin{aligned}
y_{it} &= \alpha + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it} + (c_i - E[c_i\,|\,\mathbf{X}_i])\\
&= \alpha + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it} + u_i\\
&= \alpha + \mathbf{x}_{it}'\boldsymbol{\beta} + w_{it}.
\end{aligned}
$$

In this form, we can see that the unobserved heterogeneity induces **autocorrelation;** $E[w_{it}w_{is}] = \sigma_u^2$ when $t \neq s$. As we explored in Chapter 8—we will revisit it in Chapter 19—the ordinary least squares estimator in the generalized regression model may be consistent, but the conventional estimator of its asymptotic variance is likely to underestimate the true variance of the estimator.

9.3.2 ROBUST COVARIANCE MATRIX ESTIMATION

Suppose we consider the model more generally than this. Stack the T_i observations for individual i in a single equation,

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

where $\boldsymbol{\beta}$ now includes the constant term. In this setting, there may be heteroscedasticity across individuals. However, in a panel data set, the more substantive problem is cross-observation correlation, or autocorrelation. In a longitudinal data set, the group of observations may all pertain to the same individual, so any latent effects left out of the model will carry across all periods. Suppose, then, we assume that the disturbance vector consists of ε_{it} plus these omitted components. Then,

$$\begin{aligned}\operatorname{Var}[\mathbf{w}_i \mid \mathbf{X}_i] &= \sigma_\varepsilon^2 \mathbf{I}_{Ti} + \boldsymbol{\Sigma}_i \\ &= \boldsymbol{\Omega}_i.\end{aligned}$$

The ordinary least squares estimator of $\boldsymbol{\beta}$ is

$$\begin{aligned}\mathbf{b} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \left[\sum_{i=1}^{n} \mathbf{X}_i'\mathbf{X}_i\right]^{-1} \sum_{i=1}^{n} \mathbf{X}_i'\mathbf{y}_i \\ &= \left[\sum_{i=1}^{n} \mathbf{X}_i'\mathbf{X}_i\right]^{-1} \sum_{i=1}^{n} \mathbf{X}_i'(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{w}_i) \\ &= \boldsymbol{\beta} + \left[\sum_{i=1}^{n} \mathbf{X}_i'\mathbf{X}_i\right]^{-1} \sum_{i=1}^{n} \mathbf{X}_i'\mathbf{w}_i.\end{aligned}$$

Consistency can be established along the lines developed in Chapter 4. The true asymptotic covariance matrix would take the form we saw for the generalized regression model in (8-10),

$$\begin{aligned}\text{Asy. Var}[\mathbf{b}] &= \frac{1}{n}\operatorname{plim}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{X}_i\right]^{-1} \operatorname{plim}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{w}_i\mathbf{w}_i'\mathbf{X}_i\right] \operatorname{plim}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{X}_i\right]^{-1} \\ &= \frac{1}{n}\operatorname{plim}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{X}_i\right]^{-1} \operatorname{plim}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i'\boldsymbol{\Omega}_i\mathbf{X}_i\right] \operatorname{plim}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{X}_i\right]^{-1}.\end{aligned}$$

This result provides the counterpart to (8-28). As before, the center matrix must be estimated. In the same spirit as the White estimator, we can estimate this matrix with

$$\text{Est. Asy. Var}[\mathbf{b}] = \frac{1}{n}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{X}_i\right]^{-1} \left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i'\hat{\mathbf{w}}_i\hat{\mathbf{w}}_i'\mathbf{X}_i\right] \left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{X}_i\right]^{-1}, \tag{9-3}$$

where $\hat{\mathbf{w}}'$ is the vector of T_i residuals for individual i. In fact, the logic of the White estimator *does* carry over to this estimator. Note, however, this is not quite the same as (8-27). It is quite likely that the more important issue for appropriate estimation of the asymptotic covariance matrix is the correlation across observations, not heteroscedasticity. As such, it is quite likely that the White estimator in (8-27) is not the solution to the inference problem here. Example 9.1 shows this effect at work.

Example 9.1 Wage Equation

Cornwell and Rupert (1988) analyzed the returns to schooling in a (balanced) panel of 595 observations on heads of households. The sample data are drawn from years 1976–1982 from the "Non-Survey of Economic Opportunity" from the Panel Study of Income Dynamics.

The estimating equation is

$$\begin{aligned}\ln Wage_{it} = {} & \beta_1 + \beta_2\, Exp_{it} + \beta_3\, Exp_{it}^2 + \beta_4\, Wks_{it} + \beta_5\, Occ_{it} \\ & + \beta_6\, Ind_{it} + \beta_7\, South_{it} + \beta_8\, SMSA_{it} + \beta_9\, MS_{it} \\ & + \beta_{10}\, Union_{it} + \beta_{11}\, Ed_i + \beta_{12}\, Fem_i + \beta_{13}\, Blk_i + \varepsilon_{it}\end{aligned}$$

where the variables are

Exp = years of full time work experience, 0 if not,
Wks = weeks worked, 0 if not,
Occ = 1 if blue collar occupation, 0 if not,
Ind = 1 if the individual works in a manufacturing industry, 0 if not,
South = 1 if the individual resides in the south, 0 if not,
SMSA = 1 if the individual resides in an SMSA, 0 if not,
MS = 1 if the individual is married, 0 if not,
Union = 1 if the individual wage is set by a union contract, 0 if not,
Ed = years of education,
Fem = 1 if the individual is female, 0 if not,
Blk = 1 if the individual is black, 0 if not.

TB 11.1

Note that *Ed*, *Fem*, and *Blk* are **time-invariant.** See Appendix Table F~~9~~11.1 for the data source. The main interest of the study, beyond comparing various estimation methods, is β_{11}, the return to education. Table ~~9~~11.1 reports the least squares estimates based on the full sample of 4,165 observations. [The authors do not report OLS estimates. However, they do report linear least squares estimates of the fixed effects model, which are simple least squares using deviations from individual means. (See Section ~~9~~11.4.) It was not possible to match their reported results for these or any of their other reported results. Because our purpose is to compare the various estimators to each other, we have not attempted to resolve the discrepancy.] The conventional OLS standard errors are given in the second column of results. The third column gives the robust standard errors computed using (~~9~~11-3). For these data, the computation is

$$\text{Est. Asy. Var}[\mathbf{b}] = \left[\sum_{i=1}^{595} \mathbf{X}_i'\mathbf{X}_i\right]^{-1} \left[\sum_{i=1}^{595} \left(\sum_{t=1}^{7} \mathbf{x}_{it}e_{it}\right)\left(\sum_{t=1}^{7} \mathbf{x}_{it}e_{it}\right)'\right] \left[\sum_{i=1}^{595} \mathbf{X}_i'\mathbf{X}_i\right]^{-1}.$$

TABLE ~~9~~11.1 Wage Equation Estimated by OLS

Coefficient	*Estimated Coefficient*	*OLS Standard Error*	*Panel Robust Standard Error*	*White Hetero. Consistent Std. Error*
β_1: *Constant*	5.2511	0.07129	0.1233	0.07435
β_2: *Exp*	0.04010	0.002159	0.004067	0.002158
β_3: *Exp*^{2}	−0.0006734	0.00004744	0.00009111	0.00004789
β_4: *Wks*	0.004216	0.001081	0.001538	0.001143
β_5: *Occ*	−0.1400	0.01466	0.02718	0.01494
β_6: *Ind*	0.04679	0.01179	0.02361	0.01199
β_7: *South*	−0.05564	0.01253	0.02610	0.01274
β_8: *SMSA*	0.1517	0.01207	0.02405	0.01208
β_9: *MS*	0.04845	0.02057	0.04085	0.02049
β_{10}: *Union*	0.09263	0.01280	0.02362	0.01233
β_{11}: *Ed*	0.05670	0.002613	0.005552	0.002726
β_{12}: *Fem*	−0.3678	0.02510	0.04547	0.02310
β_{13}: *Blk*	−0.1669	0.02204	0.04423	0.02075

The robust standard errors are generally about twice the uncorrected ones. In contrast, the White robust standard errors are almost the same as the uncorrected ones. This suggests that for this model, ignoring the within group correlations does, indeed, substantially affect the inferences one would draw.

11.3.3 CLUSTERING AND STRATIFICATION

Many recent studies have analyzed survey data sets, such as the Current Population Survey (CPS). Survey data are often drawn in "clusters," partly to reduce costs. For example, interviewers might visit all the families in a particular block. In other cases, effects that resemble the common random effects in panel data treatments might arise naturally in the sampling setting. Consider, for example, a study of student test scores across several states. Common effects could arise at many levels in such a data set. Education curriculum or funding policies in a state could cause a "state effect;" there could be school district effects, school effects within districts, and even teacher effects within a particular school. Each of these is likely to induce correlation across observations that resembles the random (or fixed) effects we have identified above. One might be reluctant to assume that a tightly structured model such as the simple random effects specification is at work. But, as we saw in Example 11.1, ignoring common effects can lead to serious inference errors. The robust estimator suggested in Section 11.3.2 provides a useful approach.

Au: OK to spell out i.e.?

For a two-level model, such as might arise in a sample of firms that are grouped by industry, or students who share teachers in particular schools, a natural approach to this "clustering" would be the robust common effects approach shown earlier. The resemblance of the now standard **cluster estimator** for a one level model to the common effects panel model considered above is more than coincidental. However, there is a difference in the data generating mechanism in that in this setting, the individuals in the group are generally observed once, and their association, i.e., common effect, is likely to be less clearly defined than in a panel such as the one analyzed in Example 11.1. A refinement to (11-3) is often employed to account for small sample effects when the number of clusters is likely to be a significant proportion of a finite total, such as the number of school districts in a state. A degrees of freedom correction as shown in (11-4) is often employed for this purpose. The robust covariance matrix estimator would be

KT

earlier

that is

$$Est.Asy.Var[\mathbf{b}] = \left[\sum_{g=1}^{G}\mathbf{X}'_g\mathbf{X}_g\right]^{-1}\left[\frac{G}{G-1}\sum_{g=1}^{G}\left(\sum_{i=1}^{n_g}\mathbf{x}_{ig}\hat{w}_{ig}\right)\left(\sum_{i=1}^{n_g}\mathbf{x}_{ig}\hat{w}_{ig}\right)'\right]\left[\sum_{g=1}^{G}\mathbf{X}'_g\mathbf{X}_g\right]^{-1}$$
$$= \left[\sum_{g=1}^{G}\mathbf{X}'_g\mathbf{X}_g\right]^{-1}\left[\frac{G}{G-1}\sum_{g=1}^{G}\left(\mathbf{X}'_g\hat{\mathbf{w}}_g\right)\left(\hat{\mathbf{w}}'_g\mathbf{X}_g\right)\right]\left[\sum_{g=1}^{G}\mathbf{X}'_g\mathbf{X}_g\right]^{-1}, \tag{11-4}$$

Au: OK to set "Est. Asy. Var" italic?

minus

where G is the number of clusters in the sample and each cluster consists of n_g, $g = 1,\ldots,G$ observations. [Note that this matrix is simply $G/(G-1)$ times the matrix in (11-3).] A further correction (without obvious formal motivation) sometimes employed is a "degrees of freedom correction," $\Sigma_g n_g / [(\Sigma_g n_g) - K]$.

Many further refinements for more complex samples — consider the test scores example — have been suggested. For a detailed analysis, see Cameron and Trivedi (2005, Chapter 24). Several aspects of the computation are discussed in Wooldridge (2003) as well. An important question arises concerning the use of asymptotic distributional results in cases in which the number of clusters might be relatively small. Angrist and Lavy (2002) find that the clustering correction after pooled OLS, as we have done in Example 9.1, is not as helpful as might be hoped for. (Though our correction with 595 clusters each of size 7 would be "safe" by these standards.) But, the difficulty might arise, at least in part, from the use of OLS in the presence of the common effects. Kezde (2001) and Bertrand, Dufflo and Mullainathan (2002) find more encouraging results when the correction is applied after estimation of the fixed effects regression. Yet another complication arises when the groups are very large and the number of groups is relatively small, for example when the panel consists of many large samples from a subset (or even all) of the U.S. states. Since the asymptotic theory we have used to this point assumes the opposite, the results will be less reliable in this case. Donald and Lang (2007) find that this case gravitates toward analysis of group means, rather than the individual data. Wooldridge (2003) provides results that help explain this finding. Finally, there is a natural question as to whether the correction is even called for if one has used a random effects, generalized least squares procedure (see Section 11.5 following) to do the estimation at the first step. If the data generating mechanism were strictly consistent with the random effects model, the answer would clearly be negative. Under the view that the random effects specification is only an approximation to the correlation across observations in a cluster, then there would remain "residual correlation" that would be accommodated by the correction in (11-4) (or some GLS counterpart). (This would call the specific random effects correction in Section 11.5 into question, however.) A similar argument would motivate the correction after fitting the fixed effects model as well. We will pursue these possibilities in Section 11.6.4 after we develop the fixed and random effects estimator in detail.

Example 11.2 Repeat Sales of Monet Paintings

We examined in Examples 4.5, 4.10 and 6.2 the relationship between the sale price and the surface area of a sample of 430 sales of Monet paintings. In fact, these were not sales of 430 paintings. Many of them were repeat sales of the same painting at different points in time. The sample actually contains 376 paintings. The numbers of sales per painting were one, 333, two, 34, three, 7 and four, 2. If the sale price of the painting is motivated at least partly by intrinsic features of the painting, then this would motivate a correction of the least squares standard errors as suggested in (11-4). Table 11.2 displays the OLS regression results with the coventional and with the corrected standard errors. Even with the quite modest amount of grouping in the data, the impact of the correction, in the expected direction of larger standard errors, is evident.

TB 11.2

TABLE 11.2 Sale Price Equation

Variable	*Estimated Coefficient*	*OLS Standard Error*	*Corrected Standard Error*
Constant	−9.7068	0.5661	0.6791
ln *Area*	1.3473	0.0822	0.1030
Signature	1.3614	0.1251	0.1281
ln *Aspect Ratio*	−0.0225	0.1479	0.1661

Note minus signs

11.3.4 ROBUST ESTIMATION USING GROUP MEANS

The pooled regression model can be estimated using the sample means of the data. The implied regression model is obtained by premultiplying each group by $(1/T)\mathbf{i}'$ where $\mathbf{i}'$ is a row vector of ones;

$$(1/T)\mathbf{i}'\mathbf{y}_i = (1/T)\mathbf{i}'\mathbf{X}_i\boldsymbol{\beta} + (1/T)\mathbf{i}'\mathbf{w}_i$$

or

$$\bar{y}_{i.} = \bar{\mathbf{x}}_{i.}'\boldsymbol{\beta} + \bar{w}_{i.}$$

In the transformed linear regression, the disturbances continue to have zero conditional means but heteroscedastic variances $\sigma_i^2 = (1/T^2)\mathbf{i}'\boldsymbol{\Omega}_i\mathbf{i}$. With $\boldsymbol{\Omega}_i$ unspecified, this is a heteroscedastic regression for which we would use the White estimator for appropriate inference. Why might one want to use this estimator when the full data set is available? If the classical assumptions are met, then it is straightforward to show that the asymptotic covariance matrix for the group means estimator is unambiguously larger, and the answer would be that there is no benefit. But, failure of the classical assumptions is what brought us to this point, and then the issue is less clear-cut. In the presence of unstructured cluster effects the efficiency of least squares can be considerably diminished, as we saw in the preceding example. The loss of information that occurs through the averaging might be relatively small, though in principle, the disaggregated data should still be better.

KT

We emphasize, using **group means** does not solve the problem that is addressed by the fixed effects estimator. Consider the general model,

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + c_i\mathbf{i} + \mathbf{w}_i,$$

where as before, c_i is the latent effect. If the mean independence assumption, $E[c_i \mid \mathbf{X}_i] = \alpha$, is not met, then, the effect will be transmitted to the group means as well. In this case, $E[c_i \mid \mathbf{X}_i] = h(\mathbf{X}_i)$. A common specification is Mundlak's (1978),

Au: Is x-ref to Section 9.5.5 OK?

$$E[c_i \mid \mathbf{X}_i] = \bar{\mathbf{x}}_{i.}'\boldsymbol{\gamma}.$$

(We will revisit this specification in Section 9.5.5.) Then,

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}'\boldsymbol{\beta} + c_i + \varepsilon_{it} \\ &= \mathbf{x}_{it}'\boldsymbol{\beta} + \bar{\mathbf{x}}_{i.}'\boldsymbol{\gamma} + [\varepsilon_{it} + c_i - E[c_i \mid \mathbf{X}_i]] \\ &= \mathbf{x}_{it}'\boldsymbol{\beta} + \bar{\mathbf{x}}_{i.}'\boldsymbol{\gamma} + u_{it} \end{aligned}$$

where by construction, $E[u_{it} \mid \mathbf{X}_i] = 0$. Taking means as before,

$$\begin{aligned} \bar{y}_{i.} &= \bar{\mathbf{x}}_{i.}'\boldsymbol{\beta} + \bar{\mathbf{x}}_{i.}'\boldsymbol{\gamma} + \bar{u}_{i.} \\ &= \bar{\mathbf{x}}_{i.}'(\boldsymbol{\beta} + \boldsymbol{\gamma}) + \bar{u}_{i.}. \end{aligned}$$

The implication is that the group means estimator estimates not $\boldsymbol{\beta}$, but $\boldsymbol{\beta} + \boldsymbol{\gamma}$. Averaging the observations in the group collects the entire set of effects, observed and latent, in the group means.

KT

One consideration that remains, which, unfortunately, we cannot resolve analytically, is the possibility of **measurement error.** If the regressors are measured with error, then, as we ~~will explore~~ examined in Section ~~12.5~~ 8.5 the least squares estimator is inconsistent and, as a consequence, efficiency is a moot point. In the panel data setting, if the measurement error is random, then using group means would work in the direction of averaging it out—indeed, in this instance, assuming the benchmark case $x_{itk} = x_{itk}^* + u_{itk}$, one could show that the group means estimator would be consistent as $T \to \infty$ while the OLS estimator would not.

TABLE ~~9.2~~ 11.3 Wage Equation Estimated by OLS

Coefficient	OLS Estimated Coefficient	Panel Robust Standard Error	Group Means Estimates	White Robust Standard Error
β_1: *Constant*	5.2511	0.1233	5.1214	0.2078
β_2: *Exp*	0.04010	0.004067	0.03190	0.004597
β_3: *Exp*^{2}	−0.0006734	0.00009111	−0.0005656	0.0001020
β_4: *Wks*	0.004216	0.001538	0.009189	0.003578
β_5: *Occ*	−0.1400	0.02718	−0.1676	0.03338
β_6: *Ind*	0.04679	0.02361	0.05792	0.02636
β_7: *South*	−0.05564	0.02610	−0.05705	0.02660
β_8: *SMSA*	0.1517	0.02405	0.1758	0.02541
β_9: *MS*	0.04845	0.04085	0.1148	0.04989
β_{10}: *Union*	0.09263	0.02362	0.1091	0.02830
β_{11}: *Ed*	0.05670	0.005552	0.05144	0.005862
β_{12}: *Fem*	−0.3678	0.04547	−0.3171	0.05105
β_{13}: *Blk*	−0.1669	0.04423	−0.1578	0.04352

AU: Is x-ref to Exm 9.1 OK?

Example ~~9.2~~ 11.3 Robust Estimators of the Wage Equation

TB 11.3 11.3

Table ~~9.2~~ shows the group means estimator of the wage equation shown in Example 9.1 with the original least squares estimates. In both cases, a robust estimator is used for the covariance matrix of the estimator. It appears that similar results are obtained with the means.

11 ~~9~~.3.5 ESTIMATION WITH FIRST DIFFERENCES

First differencing is another approach to estimation. Here, the intent would explicitly be to transform latent heterogeneity out of the model. The base case would be

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

which implies the first differences equation

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

or

$$\begin{aligned}\Delta y_{it} &= (\Delta \mathbf{x}_{it})'\boldsymbol{\beta} + \varepsilon_{it} - \varepsilon_{i,t-1}\\ &= (\Delta \mathbf{x}_{it})'\boldsymbol{\beta} + u_{it}.\end{aligned}$$

KT

The advantage of the **first difference** approach is that it removes the latent heterogeneity from the model whether the fixed or random effects model is appropriate. The disadvantage is that the differencing also removes any time-invariant variables from the model. In our example, we had three, *Ed*, *Fem*, and *Blk*. If the time-invariant variables in the model are of no interest, then this is a robust approach that can estimate the parameters of the time-varying variables consistently. Of course, this is not helpful for the application in the example, because the impact of *Ed* on ln *Wage* was the primary object of the analysis. Note, as well, that the differencing procedure trades the cross-observation correlation in c_i for a moving average (MA) disturbance, $u_{i,t} = \varepsilon_{i,t} - \varepsilon_{i,t-1}$. The new disturbance, $u_{i,t}$ is autocorrelated, though across only one period. Procedures are available for using two-step feasible GLS for an MA disturbance (see Chapter 19). Alternatively, this model is a natural candidate for OLS with the Newey–West robust covariance estimator, since the right number of lags (one) is known. (See Section ~~19~~ 20.5.2.)

11-16

As a general observation, with a variety of approaches available, the first difference estimator does not have much to recommend it, save for one very important application. Many studies involve two period "panels," a before and after treatment. In these cases, as often as not, the phenomenon of interest may well specifically be the change in the outcome variable—the "treatment effect." Consider the model

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

where $t = 1, 2$ and $S_{it} = 0$ in period 1 and 1 in period 2; S_{it} indicates a "treatment" that takes place between the two observations. The "treatment effect" would be

$$\mathrm{E}[\Delta y_i \mid (\Delta \mathbf{x}_i = 0)] = \theta,$$

which is precisely the constant term in the first difference regression,

$$\Delta y_i = \theta + (\Delta \mathbf{x}_i)'\boldsymbol{\beta} + u_i.$$

We will examine cases like these in detail in Section ~~24.5.~~ 18.5

11 ~~9~~.3.6 THE WITHIN- AND BETWEEN-GROUPS ESTIMATORS

We can formulate the pooled regression model in three ways. First, the original formulation is

$$y_{it} = \alpha + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}. \qquad (\cancel{9}\text{-4a})\ 11\text{-5a}$$

In terms of the group means,

$$\bar{y}_{i.} = \alpha + \bar{\mathbf{x}}_{i.}'\boldsymbol{\beta} + \bar{\varepsilon}_{i.}, \qquad (\cancel{9}\text{-4b})\ 11\text{-5b}$$

while in terms of deviations from the group means,

$$y_{it} - \bar{y}_{i.} = (\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.})'\boldsymbol{\beta} + \varepsilon_{it} - \bar{\varepsilon}_{i.}. \qquad (\cancel{9}\text{-4c})\ 11\text{-5c}$$

[We are assuming there are no time-invariant variables, such as *Ed* in Example 9.1, in $\mathbf{x}_{it}$. These would become all zeros in (~~9-4c~~ 11-5c).] All three are classical regression models, and in principle, all three could be estimated, at least consistently if not efficiently, by ordinary least squares. [Note that (~~9~~ 11-4b) defines only n observations, the group means.] Consider then the matrices of sums of squares and cross products that would be used in each case, where we focus only on estimation of $\boldsymbol{\beta}$. In (~~9-4a~~ 11-5a), the moments would accumulate variation about the overall means, $\bar{\bar{y}}$ and $\bar{\bar{\mathbf{x}}}$, and we would use the total sums of squares and cross products,

$$\mathbf{S}_{xx}^{total} = \sum_{i=1}^{n}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\bar{\mathbf{x}}})(\mathbf{x}_{it} - \bar{\bar{\mathbf{x}}})' \quad \text{and} \quad \mathbf{S}_{xy}^{total} = \sum_{i=1}^{n}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\bar{\mathbf{x}}})(y_{it} - \bar{\bar{y}}). \qquad (\cancel{9}\text{-5})\ 11\text{-6}$$

For (~~9-4c~~ 11-5c), because the data are in deviations already, the means of $(y_{it} - \bar{y}_{i.})$ and $(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.})$ are zero. The moment matrices are **within-groups** (i.e., variation around group means) sums of squares and cross products,

OK

$$\mathbf{S}_{xx}^{within} = \sum_{i=1}^{n}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.})(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.})' \quad \text{and} \quad \mathbf{S}_{xy}^{within} = \sum_{i=1}^{n}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.})(y_{it} - \bar{y}_{i.}).$$

Au: Confirm renumbering of equations from here on. EQ 11-4 is on msp 11-12

Au: Is X-ref to Exm 9.1 OK?

Finally, for (9-4b), the mean of group means is the overall mean. The moment matrices are the **between-groups** sums of squares and cross products—that is, the variation of the group means around the overall means;

$$\mathbf{S}_{xx}^{between} = \sum_{i=1}^{n} T(\bar{\mathbf{x}}_{i.} - \bar{\bar{\mathbf{x}}})(\bar{\mathbf{x}}_{i.} - \bar{\bar{\mathbf{x}}})' \quad \text{and} \quad \mathbf{S}_{xy}^{between} = \sum_{i=1}^{n} T(\bar{\mathbf{x}}_{i.} - \bar{\bar{\mathbf{x}}})(\bar{y}_{i.} - \bar{\bar{y}}).$$

It is easy to verify that

$$\mathbf{S}_{xx}^{total} = \mathbf{S}_{xx}^{within} + \mathbf{S}_{xx}^{between} \quad \text{and} \quad \mathbf{S}_{xy}^{total} = \mathbf{S}_{xy}^{within} + \mathbf{S}_{xy}^{between}.$$

Therefore, there are three possible least squares estimators of $\boldsymbol{\beta}$ corresponding to the decomposition. The least squares estimator is

$$\mathbf{b}^{total} = \left[\mathbf{S}_{xx}^{total}\right]^{-1}\mathbf{S}_{xy}^{total} = \left[\mathbf{S}_{xx}^{within} + \mathbf{S}_{xx}^{between}\right]^{-1}\left[\mathbf{S}_{xy}^{within} + \mathbf{S}_{xy}^{between}\right]. \tag{9-6}$$

The **within-groups estimator** is

$$\mathbf{b}^{within} = \left[\mathbf{S}_{xx}^{within}\right]^{-1}\mathbf{S}_{xy}^{within}. \tag{9-7}$$

This is the dummy variable estimator developed in Section 9.4. An alternative estimator would be the **between-groups estimator,**

$$\mathbf{b}^{between} = \left[\mathbf{S}_{xx}^{between}\right]^{-1}\mathbf{S}_{xy}^{between}. \tag{9-8}$$

This is the **group means estimator.** This least squares estimator of (9-4b) is based on the n sets of groups means. (Note that we are assuming that n is at least as large as K.) From the preceding expressions (and familiar previous results),

$$\mathbf{S}_{xy}^{within} = \mathbf{S}_{xx}^{within}\mathbf{b}^{within} \quad \text{and} \quad \mathbf{S}_{xy}^{between} = \mathbf{S}_{xx}^{between}\mathbf{b}^{between}.$$

Inserting these in (9-6), we see that the least squares estimator is a **matrix weighted average** of the within- and between-groups estimators:

$$\mathbf{b}^{total} = \mathbf{F}^{within}\mathbf{b}^{within} + \mathbf{F}^{between}\mathbf{b}^{between}, \tag{9-9}$$

where

$$\mathbf{F}^{within} = \left[\mathbf{S}_{xx}^{within} + \mathbf{S}_{xx}^{between}\right]^{-1}\mathbf{S}_{xx}^{within} = \mathbf{I} - \mathbf{F}^{between}.$$

The form of this result resembles the Bayesian estimator in the classical model discussed in Chapter 18. The resemblance is more than passing; it can be shown [see, e.g., Judge et al. (1985)] that

$$\mathbf{F}^{within} = \{[\text{Asy. Var}(\mathbf{b}^{within})]^{-1} + [\text{Asy. Var}(\mathbf{b}^{between})]^{-1}\}^{-1}[\text{Asy. Var}(\mathbf{b}^{within})]^{-1},$$

which is essentially the same mixing result we have for the Bayesian estimator. In the weighted average, the estimator with the smaller variance receives the greater weight.

Example ~~9.8~~ 11.4 Analysis of Covariance and the World Health Organization Data

The decomposition of the total variation in Section ~~9~~ 11.3.6 extends to the linear regression model the familiar "analysis of variance," or ANOVA, that is often used to decompose the variation in a variable in a clustered or stratified sample, or in a panel data set. One of the useful features of panel data analysis as we are doing here is the ability to analyze the

AU: Term "within-groups estimator" not in chap. list

AU: X-ref to Sect. 9.4 OK?

AU: Term "between-groups estimator" not in chap. list

TABLE ~~9.3~~ 11.4 Analysis of Variance for WHO Data on Health Care Attainment

Variable	*Within-Groups Variation*	*Between-Groups Variation*
DALE	5.645%	94.355%
COMP	0.150%	99.850%
Expenditure	0.635%	99.365%
Education	0.178%	99.822%

between-groups variation (heterogeneity) to learn about the main regression relationships and the within-groups variation to learn about dynamic effects. 6.10

The World Health Organization data used in Example 6.6 is an unbalanced panel data set—we used only one year of the data in Example 6.6. Of the 191 countries in the sample, 140 are observed in the full five years, one is observed four times, and 50 are observed only once. The original WHO studies (2000a, 2000b) analyzed these data using the fixed effects model developed in the next section. The estimator is that in ~~(9-4c)~~. It is easy to see that groups with one observation will fall out of the computation, because if $T_i = 1$, then the observation equals the group mean. These data have been used by many researchers in similar panel data analyses. [See, e.g., Greene (2004c) and several references.] Gravelle et al. (2002a) have strongly criticized these analyses, arguing that the WHO data are much more like a cross section than a panel data set. 6.10 11-5c

From Example 6.~~6~~10, the model used by the researchers at WHO was

$$\ln DALE_{it} = \alpha_i + \beta_1 \ln Health\ Expenditure_{it} + \beta_2 \ln Education_{it} + \beta_3 \ln^2 Education_{it} + \varepsilon_{it}.$$

Additional models were estimated using WHO's composite measure of health care attainment, *COMP*. The analysis of variance for a variable x_{it} is based on the decomposition

$$\sum_{i=1}^{n}\sum_{t=1}^{T_i}(x_{it} - \overline{x})^2 = \sum_{i=1}^{n}\sum_{t=1}^{T_i}(x_{it} - \overline{x}_{i.})^2 + \sum_{t=1}^{n} T_i(\overline{x}_{i.} - \overline{x})^2.$$

Dividing both sides of the equation by the left-hand side produces the decomposition:

$$1 = \textit{Within-groups proportion} + \textit{Between-groups proportion}.$$

The first term on the right-hand side is the within-group variation that differentiates a panel data set from a cross section (or simply multiple observations on the same variable). Table ~~9.3~~ 11.4 lists the decomposition of the variation in the variables used in the WHO studies.

TB 11.4

The results suggest the reasons for the authors' concern about the data. For all but ~~DALE~~ COMP, virtually all the variation in the data is between groups—that is cross-sectional variation. As the authors argue, these data are only slightly different from a cross section.

~~9~~11.4 THE FIXED EFFECTS MODEL

The fixed effects model arises from the assumption that the omitted effects, c_i, in the general model,

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

are correlated with the included variables. In a general form,

$$E[c_i \mid \mathbf{X}_i] = h(\mathbf{X}_i). \qquad \text{(~~9-10~~ 11-11)}$$

11-19

Because the conditional mean is the same in every period, we can write the model as

$$
\begin{aligned}
y_{it} &= \mathbf{x}_{it}'\boldsymbol{\beta} + h(\mathbf{X}_i) + \varepsilon_{it} + [c_i - h(\mathbf{X}_i)] \\
&= \mathbf{x}_{it}'\boldsymbol{\beta} + \alpha_i \qquad + \varepsilon_{it} + [c_i - h(\mathbf{X}_i)].
\end{aligned}
$$

By construction, the bracketed term is uncorrelated with $\mathbf{X}_i$, so we may absorb it in the disturbance, and write the model as

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \alpha_i + \varepsilon_{it}. \qquad (\cancel{9}\text{-}11)\ 11\text{-}12$$

11-12 11-11

A further assumption (usually unstated) is that $\mathrm{Var}[c_i \mid \mathbf{X}_i]$ is constant. With this assumption, (~~9-11~~) becomes a classical linear regression model. (We will reconsider the homoscedasticity assumption shortly.) We emphasize, it is (~~9-10~~) that signifies the "fixed effects" model, not that any variable is "fixed" in this context and random elsewhere. The fixed effects formulation implies that differences across groups can be captured in differences in the constant term.[5] Each α_i is treated as an unknown parameter to be estimated.

FN 5

AU: Term "time-invariant" already a KT in chapter. Here also?

Before proceeding, we note once again a major shortcoming of the fixed effects approach. Any **time invariant** variables in $\mathbf{x}_{it}$ will mimic the individual specific constant term. Consider the application of Examples ~~9.1~~ 11.1 and ~~9.2~~ 11.3. We could write the fixed effects formulation as

$$\ln Wage_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + [\beta_{10}Ed_i + \beta_{11}Fem_i + \beta_{12}Blk_i + c_i] + \varepsilon_{it}.$$

The fixed effects formulation of the model will absorb the last four terms in the regression in α_i. The coefficients on the time-invariant variables cannot be estimated. This lack of identification is the price of the robustness of the specification to unmeasured correlation between the common effect and the exogenous variables.

~~9~~ 11.4.1 LEAST SQUARES ESTIMATION

FN 6

Let $\mathbf{y}_i$ and $\mathbf{X}_i$ be the T observations for the ith unit, $\mathbf{i}$ be a $T \times 1$ column of ones, and let $\boldsymbol{\varepsilon}_i$ be the associated $T \times 1$ vector of disturbances.[6] Then,

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{i}\alpha_i + \boldsymbol{\varepsilon}_i.$$

Collecting these terms gives

$$
\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} =
\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{bmatrix} \boldsymbol{\beta} +
\begin{bmatrix} \mathbf{i} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{i} & \cdots & \mathbf{0} \\ & & \vdots & \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{i} \end{bmatrix}
\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} +
\begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_n \end{bmatrix}
$$

or

$$\mathbf{y} = [\mathbf{X} \quad \mathbf{d}_1 \quad \mathbf{d}_2 \ldots \mathbf{d}_n] \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{bmatrix} + \boldsymbol{\varepsilon}, \qquad (\cancel{9}\text{-}12)\ 11\text{-}13$$

[5] It is also possible to allow the slopes to vary across i, but this method introduces some new methodological issues, as well as considerable complexity in the calculations. A study on the topic is Cornwell and Schmidt (1984).

[6] The assumption of a fixed group size, T, at this point is purely for convenience. As noted in Section ~~9~~ 11.2.4, the unbalanced case is a minor variation.

11-20

where $\mathbf{d}_i$ is a dummy variable indicating the ith unit. Let the $nT \times n$ matrix $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_n]$. Then, assembling all nT rows gives

AU: This KT is not in chapter list

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

KT

This model is usually referred to as the **least squares dummy variable (LSDV) model** (although the "least squares" part of the name refers to the technique usually used to estimate it, not to the model itself).

This model is a classical regression model, so no new results are needed to analyze it. If n is small enough, then the model can be estimated by ordinary least squares with K regressors in $\mathbf{X}$ and n columns in $\mathbf{D}$, as a multiple regression with $K + n$ parameters. Of course, if n is thousands, as is typical, then this model is likely to exceed the storage capacity of any computer. But, by using familiar results for a partitioned regression, we can reduce the size of the computation.[7] We write the least squares estimator of $\boldsymbol{\beta}$ as

FN 7

$$\mathbf{b} = [\mathbf{X}'\mathbf{M}_\mathbf{D}\mathbf{X}]^{-1}[\mathbf{X}'\mathbf{M}_\mathbf{D}\mathbf{y}] = \mathbf{b}^{within}, \qquad (\cancel{9}\text{-13})$$

11-14

where

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

This amounts to a least squares regression using the transformed data $\mathbf{X}_* = \mathbf{M}_\mathbf{D}\mathbf{X}$ and $\mathbf{y}_* = \mathbf{M}_\mathbf{D}\mathbf{y}$. The structure of $\mathbf{D}$ is particularly convenient; its columns are orthogonal, so

$$\mathbf{M}_\mathbf{D} = \begin{bmatrix} \mathbf{M}^0 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{M}^0 & \mathbf{0} & \cdots & \mathbf{0} \\ & & \cdots & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{M}^0 \end{bmatrix}.$$

Each matrix on the diagonal is

11-15

$$\mathbf{M}^0 = \mathbf{I}_T - \frac{1}{T}\mathbf{i}\mathbf{i}'. \qquad (\cancel{9}\text{-14})$$

Premultiplying any $T \times 1$ vector $\mathbf{z}_i$ by $\mathbf{M}^0$ creates $\mathbf{M}^0\mathbf{z}_i = \mathbf{z}_i - \bar{z}\mathbf{i}$. (Note that the mean is taken over only the T observations for unit i.) Therefore, the least squares regression of $\mathbf{M}_\mathbf{D}\mathbf{y}$ on $\mathbf{M}_\mathbf{D}\mathbf{X}$ is equivalent to a regression of $[y_{it} - \bar{y}_{i.}]$ on $[\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.}]$, where $\bar{y}_{i.}$ and $\bar{\mathbf{x}}_{i.}$ are the scalar and $K \times 1$ vector of means of y_{it} and $\mathbf{x}_{it}$ over the T observations for group i.[8] The dummy variable coefficients can be recovered from the other normal equation in the partitioned regression:

FN 8

$$\mathbf{D}'\mathbf{D}\mathbf{a} + \mathbf{D}'\mathbf{X}\mathbf{b} = \mathbf{D}'\mathbf{y}$$

or

$$\mathbf{a} = [\mathbf{D}'\mathbf{D}]^{-1}\mathbf{D}'(\mathbf{y} - \mathbf{X}\mathbf{b}).$$

This implies that for each i,

11-16

$$a_i = \bar{y}_{i.} - \bar{\mathbf{x}}_{i.}'\mathbf{b}. \qquad (\cancel{9}\text{-15})$$

[7] See Theorem 3.3.

[8] An interesting special case arises if $T = 2$. In the two-period case, you can show—we leave it as an exercise—that this least squares regression is done with $nT/2$ first difference observations, by regressing observation $(y_{i2} - y_{i1})$ (and its negative) on $(\mathbf{x}_{i2} - \mathbf{x}_{i1})$ (and its negative).

The appropriate estimator of the asymptotic covariance matrix for $\mathbf{b}$ is

$$\text{Est. Asy. Var}[\mathbf{b}] = s^2[\mathbf{X}'\mathbf{M_D}\mathbf{X}]^{-1} = s^2\left[\mathbf{S}_{xx}^{within}\right]^{-1}, \qquad (\cancel{9}\text{-}16)\ \ 11\text{-}17$$

which uses the second moment matrix with $\mathbf{x}$'s now expressed as deviations from their respective group means. The disturbance variance estimator is

$$s^2 = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it} - \mathbf{x}_{it}'\mathbf{b} - a_i)^2}{nT - n - K} = \frac{(\mathbf{M_D}\mathbf{y} - \mathbf{M_D}\mathbf{X}\mathbf{b})'(\mathbf{M_D}\mathbf{y} - \mathbf{M_D}\mathbf{X}\mathbf{b})}{nT - n - K}. \qquad (\cancel{9}\text{-}17)\ \ 11\text{-}18$$

The itth residual used in this computation is

$$e_{it} = y_{it} - \mathbf{x}_{it}'\mathbf{b} - a_i = y_{it} - \mathbf{x}_{it}'\mathbf{b} - (\bar{y}_{i.} - \bar{\mathbf{x}}_{i.}'\mathbf{b}) = (y_{it} - \bar{y}_{i.}) - (\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.})'\mathbf{b}.$$

Thus, the numerator in s^2 is exactly the sum of squared residuals using the least squares slopes and the data in group mean deviation form. But, done in this fashion, one might then use $nT - K$ instead of $nT - n - K$ for the denominator in computing s^2, so a correction would be necessary. For the individual effects,

$$\text{Asy. Var}[a_i] = \frac{\sigma_\varepsilon^2}{T} + \bar{\mathbf{x}}_{i.}'\{\text{Asy. Var}[\mathbf{b}]\}\bar{\mathbf{x}}_{i.}, \qquad (\cancel{9}\text{-}18)\ \ 11\text{-}19$$

so a simple estimator based on s^2 can be computed.

11 ~~9~~.4.2 SMALL T ASYMPTOTICS

From (~~9~~-16), we find

$$
\begin{aligned}
11\text{-}17 \qquad \text{Asy. Var}[\mathbf{b}] &= \sigma_\varepsilon^2[\mathbf{X}'\mathbf{M_D}\mathbf{X}]^{-1} \\
&= \frac{\sigma_\varepsilon^2}{n}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{M}^0\mathbf{X}_i\right]^{-1} \\
&= \frac{\sigma_\varepsilon^2}{n}\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.})(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.})'\right]^{-1} \qquad (\cancel{9}\text{-}19)\ \ 11\text{-}20 \\
&= \frac{\sigma_\varepsilon^2}{n}\left[T\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.})(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.})'\right]^{-1} \\
&= \frac{\sigma_\varepsilon^2}{n}\left[T\bar{\mathbf{S}}_{xx,i}\right]^{-1}.
\end{aligned}
$$

Since least squares is unbiased in this model, the question of (mean square) consistency turns on the covariance matrix. Does the matrix above converge to zero? It is necessary to be specific about what is meant by convergence. In this setting, increasing sample size refers to increasing n, that is, increasing the number of groups. The group size, T, is assumed fixed. The leading scalar clearly vanishes with increasing n. The matrix in the square brackets is T times the average over the n groups of the within groups covariance matrices of the variables in $\mathbf{X}_i$. So long as the data are well behaved, we can assume that the bracketed matrix does not converge to a zero matrix (or a matrix with zeros on the diagonal). On this basis, we can expect consistency of the least squares estimator. In practical terms, this requires within-groups variation of the data. Notice that the result falls apart if there are time invariant variables in $\mathbf{X}_i$, because then there are zeros on the diagonals of the bracketed matrix. This result also suggests the nature of the problem of the WHO data in Example 9.3 as analyzed by Gravelle et al. (2002). 11.4

Now, consider the result in (9-18) for the asymptotic variance of a_i. Assume that $\mathbf{b}$ is consistent, as shown above. Then, with increasing n, the asymptotic variance of a_i declines to a lower bound of σ_ε^2/T *which does not converge to zero*. The constant term estimators in the fixed effects model are not consistent estimators of α_i. They are not inconsistent because they gravitate toward the wrong parameter. They are so because their asymptotic variances do not converge to zero, even as the sample size grows. It is easy to see why this is the case. From (9-15), we see that each a_i is estimated using only T observations—assume n were infinite, so that β were known. Because T is not assumed to be increasing, we have the surprising result. The constant terms are inconsistent unless $T \to \infty$, which is not part of the model.

previously / 11-19 / 11-16

9.4.3 TESTING THE SIGNIFICANCE OF THE GROUP EFFECTS

The t ratio for a_i can be used for a test of the hypothesis that α_i equals zero. This hypothesis about one specific group, however, is typically not useful for testing in this regression context. If we are interested in differences across groups, then we can test the hypothesis that the constant terms are all equal with an F test. Under the null hypothesis of equality, the efficient estimator is pooled least squares. The F ratio used for this test is

$$F(n-1, nT-n-K) = \frac{(R^2_{LSDV} - R^2_{Pooled})/(n-1)}{(1-R^2_{LSDV})/(nT-n-K)}, \tag{9-20}$$

11-21

where *LSDV* indicates the dummy variable model and *Pooled* indicates the pooled or restricted model with only a single overall constant term. Alternatively, the model may have been estimated with an overall constant and $n-1$ dummy variables instead. All other results (i.e., the least squares slopes, s^2, R^2) will be unchanged, but rather than estimate α_i, each dummy variable coefficient will now be an estimate of $\alpha_i - \alpha_1$ where group "1" is the omitted group. The F test that the coefficients on these $n-1$ dummy variables are zero is identical to the one above. It is important to keep in mind, however, that although the statistical results are the same, the interpretation of the dummy variable coefficients in the two formulations is different.[9]

FN 9

9.4.4 FIXED TIME AND GROUP EFFECTS

The least squares dummy variable approach can be extended to include a time-specific effect as well. One way to formulate the extended model is simply to add the time effect, as in

$$y_{it} = \mathbf{x}_{it}'\beta + \alpha_i + \delta_t + \varepsilon_{it}. \tag{9-21}$$

11-22

This model is obtained from the preceding one by the inclusion of an additional $T-1$ dummy variables. (One of the time effects must be dropped to avoid perfect collinearity—the group effects and time effects both sum to one.) If the number of variables is too large to handle by ordinary regression, then this model can also be

[9] For a discussion of the differences, see Suits (1984).

11-23

FN 10

KT

estimated by using the partitioned regression.[10] There is an asymmetry in this formulation, however, since each of the group effects is a group-specific intercept, whereas the time effects are **contrasts**—that is, comparisons to a base period (the one that is excluded). A symmetric form of the model is

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \mu + \alpha_i + \delta_t + \varepsilon_{it}, \qquad \text{(\sout{9}-21')}\ 11\text{-}23$$

where a full n and T effects are included, but the restrictions

$$\sum_i \alpha_i = \sum_t \delta_t = 0$$

are imposed. Least squares estimates of the slopes in this model are obtained by regression of

$$y_{*it} = y_{it} - \bar{y}_{i.} - \bar{y}_{.t} + \bar{\bar{y}} \qquad \text{(\sout{9}-22)}\ 11\text{-}24$$

on

$$\mathbf{x}_{*it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_{i.} - \bar{\mathbf{x}}_{.t} + \bar{\bar{\mathbf{x}}},$$

where the period-specific and overall means are

$$\bar{y}_{.t} = \frac{1}{n}\sum_{i=1}^{n} y_{it} \quad \text{and} \quad \bar{\bar{y}} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} y_{it},$$

and likewise for $\bar{\mathbf{x}}_{.t}$ and $\bar{\bar{\mathbf{x}}}$. The overall constant and the dummy variable coefficients can then be recovered from the normal equations as

$$\begin{aligned}
\hat{\mu} &= m = \bar{\bar{y}} - \bar{\bar{\mathbf{x}}}'\mathbf{b},\\
\hat{\alpha}_i &= a_i = (\bar{y}_{i.} - \bar{\bar{y}}) - (\bar{\mathbf{x}}_{i.} - \bar{\bar{\mathbf{x}}})'\mathbf{b}, \qquad \text{(\sout{9}-23)}\ 11\text{-}25\\
\hat{\delta}_t &= d_t = (\bar{y}_{.t} - \bar{\bar{y}}) - (\bar{\mathbf{x}}_{.t} - \bar{\bar{\mathbf{x}}})'\mathbf{b}.
\end{aligned}$$

The estimated asymptotic covariance matrix for $\mathbf{b}$ is computed using the sums of squares and cross products of $\mathbf{x}_{*it}$ computed in (~~9~~-22) 11 and

$$s^2 = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it} - \mathbf{x}'_{it}\mathbf{b} - m - a_i - d_t)^2}{nT - (n-1) - (T-1) - K - 1} \qquad \text{(\sout{9}-24)}\ 11\text{-}26$$

If one of n or T is small and the other is large, then it may be simpler just to treat the smaller set as an ordinary set of variables and apply the previous results to the one-way fixed effects model defined by the larger set. Although more general, this model is infrequently used in practice. There are two reasons. First, the cost in terms of degrees of freedom is often not justified. Second, in those instances in which a model of the timewise evolution of the disturbance is desired, a more general model than this simple dummy variable formulation is usually used.

[10] The matrix algebra and the theoretical development of two-way effects in panel data models are complex. See, for example, Baltagi (2005). Fortunately, the practical application is much simpler. The number of periods analyzed in most panel data sets is rarely more than a handful. Because modern computer programs uniformly allow dozens (or even hundreds) of regressors, almost any application involving a second fixed effect can be handled just by literally including the second effect as a set of actual dummy variables.