CHAPTER 3

Simple Regression with Variable Intercepts

3.1 INTRODUCTION

When the overall homogeneity hypothesis is rejected by the panel data while the specification of a model appears proper, a simple way to take account of the unobserved heterogeneity across individuals and/or through time is to use the variable-intercept models (1.3.1) and (1.3.2). The basic assumption of such models is that, conditional on the observed explanatory variables, the effects of all omitted (or excluded) variables are driven by three types of variables: individual time-invariant, period individual-invariant, and individual time-varying variables.[1] The individual time-invariant variables are variables that are the same for a given cross-sectional unit through time but that vary across cross-sectional units. Examples of these are attributes of individual firm management, ability, sex, and socioeconomic background variables. The period individual-invariant variables are variables that are the same for all cross-sectional units at a given point in time but that vary through time. Examples of these variable are prices, interest rates, and widespread optimism or pessimism. The individual time-varying variables are variables that vary across cross-sectional units at a given point in time and also exhibit variations through time. Examples of these variables are firm profits, sales, and capital stock.

The variable-intercept models assume that the effects of the numerous omitted individual time-varying variables are each individually unimportant but are collectively significant and possess the property of a random variable that is uncorrelated with (or independent of) all other included and excluded variables. On the other hand, because the effects of remaining omitted variables either stay constant through time for a given cross-sectional unit or are the same for all cross-sectional units at a given point in time, or a combination of both, they can be absorbed into the intercept term of a regression model as a means to allow explicitly for the individual and/or time heterogeneity contained in the

[1] These three different sorts of variations apply, of course, to both included and excluded variables. Throughout this monograph we concentrate on relations between excluded variables and included variables.

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

temporal cross-sectional data. Moreover, when the individual- or time-specific effects are absorbed into the intercept term, there is no need to assume that the individual- or time-specific effects are uncorrelated with **x**, although sometimes they are.

The variable-intercept models can provide a fairly useful specification for fitting regression models using panel data. For example, consider fitting a Cobb–Douglas production function

$$y_{it} = \mu + \beta_1 x_{1it} + \cdots + \beta_K x_{Kit} + v_{it}, \quad \begin{array}{l} i = 1, \ldots, N, \\ t = 1, \ldots, T, \end{array} \tag{3.1.1}$$

where y is the logarithm of output and $x_1, \ldots, x_K$ are the logarithms of respective inputs. The classic procedure is to assume that the effects of omitted variables are independent of **x** and are independently identically distributed. Thus, conditioning on **x** all observations are random variations of a representative firm. However, (3.1.1) has often been criticized for ignoring variables reflecting managerial and other technical differences between firms or variables that reflect general conditions affecting the productivity of all firms but that are fluctuating over time (such as weather factors in agriculture production) (e.g., Hoch 1962; Mundlak 1961; Nerlove 1965). Ideally, such firm- and time-effects variables, say M_i and P_t, should be introduced explicitly into (3.1.1). Thus, v_{it} can be written as

$$v_{it} = \alpha M_i + \lambda P_t + u_{it}, \tag{3.1.2}$$

with u_{it} representing the effects of all remaining omitted variables. However, if there are no observations on M_i and P_t, it is impossible to estimate α and λ directly. A natural alternative would then be to consider the effects of the product, $\alpha_i = \alpha M_i$ and $\lambda_t = \lambda P_t$, which then leads to a variable-intercept model: (1.3.1) or (1.3.2).

Such a procedure was used by Hoch (1962) to estimate parameters of a Cobb–Douglas production function based on annual data for 63 Minnesota farms from 1946 to 1951. He treated output, y, as a function of labor, x_1; real estate, x_2; machinery, x_3; and feed, fertilizer, and related expenses, x_4. However, because of the difficulties of measuring real estate and machinery variables, he also tried an alternative specification that treated y as a function of x_1, x_4, a current-expenditures item, x_5, and fixed capital, x_6. Regression results for both specifications rejected the overall homogeneity hypothesis at the 5 percent significance level. The least-squares estimates under three assumptions ($\alpha_i = \lambda_t = 0$; $\alpha_i = 0$, $\lambda_t \neq 0$; and $\alpha_i \neq 0$, $\lambda_t \neq 0$) are summarized in Table 3.1. They exhibit an increase in the adjusted R^2 from 0.75 to about 0.88 when α_i and λ_t are introduced. There are also some important changes in parameter estimates when we move from the assumption of identical α_i's to the assumption that both α_i and λ_t differ from zero. There is a significant drop in the sum of the elasticities, with the drop concentrated mainly in the labor variable. If one interprets α_i as the firm scale effect, then this indicates that efficiency increases

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

Table 3.1. *Least-squares estimates of elasticity of Minnesota farm production function based on alternative assumptions*

	Assumption		
Estimate of Elasticity: β_k	α_i and λ_t are identically zero for all i and t	α_i only is identically zero for all i	α_i and λ_t different from zero
Variable set 1[a]			
$\hat{\beta}_1$, labor	0.256	0.166	0.043
$\hat{\beta}_2$, real estate	0.135	0.230	0.199
$\hat{\beta}_3$, machinery	0.163	0.261	0.194
$\hat{\beta}_4$, feed & fertilizer	0.349	0.311	0.289
Sum of $\hat{\beta}$'s	0.904	0.967	0.726
Adjusted R^2	0.721	0.813	0.884
Variable set 2			
$\hat{\beta}_1$, labor	0.241	0.218	0.057
$\hat{\beta}_5$, current expenses	0.121	0.185	0.170
$\hat{\beta}_6$, fixed capital	0.278	0.304	0.317
$\hat{\beta}_4$, feed & fertilizer	0.315	0.285	0.288
Sum of $\hat{\beta}$'s	0.954	0.991	0.832
Adjusted R^2	0.752	0.823	0.879

[a] All output and input variables are in service units, measured in dollars.

Source: Hoch (1962).

with scale. As demonstrated in Figure 1.1, when the production hyperplane of larger firms lies above the average production plane and the production plane of smaller firm below the average plane, the pooled estimates, neglecting firm differences, will have greater slope than the average plane. Some confirmation of this argument was provided by Hoch (1962). Table 3.2 lists the characteristics of firms grouped on the basis of firm-specific effects α_i. The table suggests a fairly pronounced association between scale and efficiency.

This example demonstrates that by introducing the unit- and/or time-specific variables into the specification for panel data, it is possible to reduce or avoid the omitted-variable bias. In this chapter we focus on the estimation and hypothesis testing of models (1.3.1) and (1.3.2) under the assumption that all explanatory variables, $\mathbf{x}_{kit}$, are nonstochastic (or exogenous). For ease of seeing the relations between fixed and random effects inference, we shall assume there are no time-specific effects in Sections 3.2–3.5. In Section 3.2 we discuss estimation methods when the specific effects are treated as fixed constants (FE). Section 3.3 discusses estimation methods when they are treated as random variables (effects) (RE). Section 3.4 discusses the pros and cons of treating the specific effects as fixed or random. Tests for misspecification are discussed in Section 3.5. Section 3.6 discusses models with both individual- and time-specific effects and models with specific variables. Section 3.7 discusses

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

Table 3.2. *Characteristics of firms grouped on the basis of the firm constant*

		Firms classified by value of $\exp(\alpha_i)^a$				
Characteristics	All firms	<0.85	0.85–0.95	0.95–1.05	1.05–1.15	>1.15
Numbers of firms in group	63	6	17	19	14	7
Average value of:						
e^{α_i}, firm constant	1.00	0.81	0.92	1.00	1.11	1.26
Output (dollars)	15,602	10,000	15,570	14,690	16,500	24,140
Labor (dollars)	3,468	2,662	3,570	3,346	3,538	4,280
Feed & fertilizer (dollars)	3,217	2,457	3,681	3,064	2,621	5,014
Current expenses (dollars)	2,425	1,538	2,704	2,359	2,533	2,715
Fixed capital (dollars)	3,398	2,852	3,712	3,067	3,484	3,996
Profit (dollars)	3,094	491	1,903	2,854	4,324	8,135
Profit/output	0.20	0.05	0.12	0.19	0.26	0.33

[a] The mean of firm effects, α_i, is zero is invoked.
Source: Hoch (1962).

heteroscedasticity and autocorrelation adjustment. In Section 3.8 we use a multivariate setup of a single-equation model to provide a synthesis of the issues involved and to provide a link between the single equation model and the linear simultaneous equations model (see Chapter 5).

3.2 FIXED-EFFECTS MODELS: LEAST-SQUARES DUMMY VARIABLE APPROACH

The obvious generalization of the constant-intercept-and-slope model for panel data is to introduce dummy variables to account for the effects of those omitted variables that are specific to individual cross-sectional units but stay constant over time, and the effects that are specific to each time period but are the same for all cross-sectional units. For ease of highlighting the difference between the FE and RE specifications in this section and the next three sections we assume no time-specific effects and focus only on individual-specific effects. Thus, the value of the dependent variable for the ith unit at time t, y_{it}, depends on K exogenous variables, $(x_{1it}, \ldots, x_{Kit}) = \mathbf{x}'_{it}$, that differ among individuals in a cross section at a given point in time and also exhibit variation through time, as well as on variables that are specific to the ith unit and that stay (more or less) constant over time. This is model (1.3.1), which we can rewrite as

$$y_{it} = \alpha_i^* + \underset{1\times K}{\mathbf{x}'_{it}} \underset{K\times 1}{\boldsymbol{\beta}} + u_{it}, \quad \begin{array}{l} i = 1, \ldots, N, \\ t = 1, \ldots, T, \end{array} \tag{3.2.1}$$

where $\boldsymbol{\beta}$ is a $K \times 1$ vector of constants and α_i^* is a 1×1 scalar constant representing the effects of those variables peculiar to the ith individual in more or less the same fashion over time. The error term, u_{it}, represents the effects of the omitted variables that are peculiar to both the individual units and time periods. We assume that u_{it} is uncorrelated with $(\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT})$ and can be characterized by an independently identically distributed random variable with mean 0 and variance σ_u^2.

The model (3.2.1) is also called the ANCOVA model. Without attempting to make the boundaries between regression analysis, ANOVA, and ANCOVA precise, we can say that regression model assumes that the expected value of y is a function of exogenous factors, $\mathbf{x}$, while the conventional ANOVA model stipulates that the expected value of y_{it} depends only on the class, i, to which the observation considered belongs and that the value of the measured quantity, y, assumes the relation that $y_{it} = \alpha_i^* + u_{it}$, where the effects of all other characteristics, u_{it}, are random and are in no way dependent on the individual-specific effects, α_i^*. But if y is also affected by other variables that we are not able to control and standardize within classes, the simple within-class sum of squares will be an overestimate of the stochastic component in y, and the differences between class means will reflect not only any class effect but also the effects of any differences in the values assumed by the uncontrolled variables in different classes. It was for this kind of problem that the ANCOVA model of the form (3.2.1) was first developed. The models are of a mixed character, involving genuine exogenous variables, $\mathbf{x}_{it}$, as do regression models, and at the same time allowing the true relation for each individual to depend on the class to which the individual belongs, α_i^*, as do the usual ANOVA models. The regression model enables us to assess the effects of quantitative factors and the ANOVA model those of qualitative factors; the ANCOVA model covers both quantitative and qualitative factors.

Stacking all NT observations of y_{it} ((3.2.1)) in vector form, we have

$$Y = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{bmatrix} = \begin{bmatrix} \mathbf{e} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \alpha_1^* + \begin{bmatrix} \mathbf{0} \\ \mathbf{e} \\ \vdots \\ \mathbf{0} \end{bmatrix} \alpha_2^* + \cdots + \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{e} \end{bmatrix} \alpha_N^* + \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_N \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_N \end{bmatrix}, \tag{3.2.2}$$

where

$$\underset{T\times 1}{\mathbf{y}_i} = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix}, \qquad \underset{T\times K}{\mathbf{X}_i} = \begin{bmatrix} x_{1i1} & x_{2i1} & \cdots & x_{Ki1} \\ x_{1i2} & x_{2i2} & \cdots & x_{Ki2} \\ \vdots & \vdots & & \vdots \\ x_{1iT} & x_{2iT} & & x_{KiT} \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_N \end{bmatrix}$$

$$\underset{1\times T}{\mathbf{e}'} = (1, 1, \dots, 1), \qquad \underset{1\times T}{\mathbf{u}_i'} = (u_{i1}, \dots, u_{iT}),$$

$$E\mathbf{u}_i = \mathbf{0}, \qquad E\mathbf{u}_i\mathbf{u}_i' = \sigma_u^2 I_T, \qquad E\mathbf{u}_i\mathbf{u}_j' = \mathbf{0} \quad \text{if } i \neq j,$$

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

I_T denotes the $T \times T$ identity matrix. Let $\tilde{X} = (\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_N, X)$, where $\mathbf{d}_i$ is an $NT \times 1$ vector dummy variable with the first $(i-1) \times T$ elements equal to 0, $(i-1)T+1$ to iT elements equal to 1, and 0 from $iT+1, \ldots, NT$, $i = 1, \ldots, N$. Then $\mathbf{y} = \tilde{X}\boldsymbol{\theta} + \mathbf{u}$, where $\boldsymbol{\theta} = (\alpha_1^*, \ldots, \alpha_N^*, \boldsymbol{\beta}')'$.

Given the assumed properties of u_{it}, we know that the ordinary least-squares (OLS) estimator of (3.2.2) is the best linear unbiased estimator (BLUE). The OLS estimators of α_i^* and $\boldsymbol{\beta}$ are obtained by minimizing

$$\begin{aligned} S &= (\mathbf{y} - \tilde{X}\boldsymbol{\theta})'(\mathbf{y} - \tilde{X}\boldsymbol{\theta}) = \sum_{i=1}^{N} \mathbf{u}_i'\mathbf{u}_i \\ &= \sum_{i=1}^{N} (\mathbf{y}_i - \mathbf{e}\alpha_i^* - X_i\boldsymbol{\beta})'(\mathbf{y}_i - \mathbf{e}\alpha_i^* - X_i\boldsymbol{\beta}). \qquad (3.2.3) \end{aligned}$$

Taking partial derivatives of S with respect to α_i^* and setting them equal to 0, we have

$$\hat{\alpha}_i^* = \bar{y}_i - \bar{\mathbf{x}}_i'\boldsymbol{\beta}, \quad i = 1, \ldots, N, \qquad (3.2.4)$$

where

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \quad \bar{\mathbf{x}}_i = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_{it}.$$

Substituting (3.2.4) into (3.2.3) and taking the partial derivative of S with respect to $\boldsymbol{\beta}$, we have[2]

$$\hat{\boldsymbol{\beta}}_{cv} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T} (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\right]^{-1} \left[\sum_{i=1}^{N}\sum_{t=1}^{T} (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(y_{it} - \bar{y}_i)\right]. \qquad (3.2.5)$$

The OLS estimator (3.2.5) is called the least-squares dummy variable (LSDV) estimator because the observed values to the coefficients α_i^* takes the form of dummy variables. However, the computational procedure for estimating the slope parameters in this model does not require the dummy variables for the individual (and/or time) effects actually be included in the matrix of explanatory variables. We need only find the means of time series observations separately for each cross-sectional unit, transform the observed variables by subtracting out the appropriate time series means, and then apply the least-squares method to the transformed data. Hence, we need only invert a matrix of order $K \times K$.

[2] Although the notations are different, (3.2.5) is identical with (2.2.10).

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

The foregoing procedure is equivalent to premultiplying the ith equation

$$\mathbf{y}_i = \mathbf{e}\alpha_i^* + X_i\boldsymbol{\beta} + \mathbf{u}_i$$

by a $T \times T$ idempotent (covariance) transformation matrix

$$Q = I_T - \frac{1}{T}\mathbf{e}\mathbf{e}' \tag{3.2.6}$$

to "sweep out" the individual effect α_i^* so that individual observations are measured as deviations from individual means (over time):

$$\begin{aligned} Q\mathbf{y}_i &= Q\mathbf{e}\alpha_i^* + QX_i\boldsymbol{\beta} + Q\mathbf{u}_i \\ &= QX_i\boldsymbol{\beta} + Q\mathbf{u}_i, \quad i = 1, \ldots, N. \end{aligned} \tag{3.2.7}$$

Applying the OLS procedure to (3.2.7) we have[3]

$$\hat{\boldsymbol{\beta}}_{cv} = \left[\sum_{i=1}^N X_i'QX_i\right]^{-1}\left[\sum_{i=1}^N X_i'Q\mathbf{y}_i\right], \tag{3.2.8}$$

which is identical to (3.2.5). Because (3.2.2) is called the ANCOVA model, the LSDV estimator of $\boldsymbol{\beta}$ is sometimes called the covariance (CV) estimator. It is also called the within-group estimator, because only the variation within each group is utilized in forming this estimator.[4]

The CV estimator of $\boldsymbol{\beta}$ can also be derived as a method of moment estimator. The strict exogeneity of $\mathbf{x}_{it}$ implies that

$$E(\mathbf{u}_i \mid X_i, \alpha_i^*) = E(\mathbf{u}_i \mid X_i) = \mathbf{0}. \tag{3.2.9}$$

It follows that

$$E[(\mathbf{u}_i - \mathbf{e}\bar{u}_i) = (\mathbf{y}_i - \mathbf{e}\bar{y}_i) - (X_i - \mathbf{e}\bar{\mathbf{x}}_i')\boldsymbol{\beta} \mid X_i] = \mathbf{0}. \tag{3.2.10}$$

[3] Equation (3.2.7) can be viewed as a linear-regression model with singular-disturbance covariance matrix $\sigma_u^2 Q$. A generalization of Aitken's theorem leads to the generalized least-squares estimator

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{cv} &= \left(\sum_{i=1}^N X_i'Q'Q^-QX_i\right)^{-1}\left(\sum_{i=1}^N X_i'Q'Q^-Q\mathbf{y}_i\right) \\ &= \left[\sum_{i=1}^N X_i'QX_i\right]^{-1}\left[\sum_{i=1}^N X_t'Q\mathbf{y}_i\right], \end{aligned}$$

where Q^- is the generalized inverse of Q satisfying the conditions $QQ^-Q = Q$ (Theil (1971), their Sections 6.6 and 6.7).

[4] Because the slope coefficients are assumed the same for all i and t, for simplicity we shall not distinguish the individual mean corrected estimator and the within-group estimator as we did in Chapter 2. We shall simply refer to (3.2.8) or its equivalent as the within-group estimator.

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

Approximating the moment conditions (3.2.10) by their sample moments yields

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N} X_i'[(\mathbf{y}_i - \mathbf{e}\bar{y}_i) - (X_i - \mathbf{e}\bar{\mathbf{x}}_i')\hat{\boldsymbol{\beta}}]\\ = \frac{1}{N}\sum_{i=1}^{N} X_i'[Q\mathbf{y}_i - QX_i\hat{\boldsymbol{\beta}}] = \mathbf{0}.\end{aligned} \tag{3.2.10$'$}$$

Solving (3.2.10′) yields the CV estimator (3.2.8).

The CV estimator $\hat{\beta}_{cv}$ is unbiased. It is also consistent when either N or T or both tend to infinity. Its variance–covariance matrix is

$$\text{Var}\,(\hat{\boldsymbol{\beta}}_{cv}) = \sigma_u^2 \left[\sum_{t=1}^{N} X_i' Q X_i\right]^{-1}. \tag{3.2.11}$$

However, the estimator for the intercept, (3.2.4), although unbiased, is consistent only when $T \to \infty$.

It should be noted that an alternative and equivalent formulation of (3.2.1) is to introduce a "mean intercept," μ, so that

$$y_{it} = \mu + \mathbf{x}_{it}'\boldsymbol{\beta} + \alpha_i + u_{it}. \tag{3.2.12}$$

Because both μ and α_i are fixed constants, without additional restriction, they are not separately identifiable or estimable. One way to identify μ and α_i is to introduce the restriction that $\sum_{i=1}^{N}\alpha_i = 0$. Then the individual effect α_i represents the deviation of the ith individual from the common mean μ.

Equations (3.2.12) and (3.2.1) lead to the same least-squares estimator for $\boldsymbol{\beta}$ [equation (3.2.5)]. This easily can be seen by noting that the BLUEs for μ, α_i, and $\boldsymbol{\beta}$ are obtained by minimizing

$$\sum_{i=1}^{N}\mathbf{u}_i'\mathbf{u}_i = \sum_{i=1}^{N}\sum_{t=1}^{T} u_{it}^2$$

subject to the restriction $\sum_{i=1}^{N}\alpha_i = 0$. Utilizing the restriction $\sum_{i=1}^{N}\alpha_i = 0$ in solving the marginal conditions, we have

$$\begin{aligned}\hat{\mu} = \bar{y} - \bar{\mathbf{x}}'\boldsymbol{\beta}, \quad \text{where } \bar{y} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it},\\ \bar{\mathbf{x}} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it},\end{aligned} \tag{3.2.13}$$

$$\hat{\alpha}_i = \bar{y}_i - \hat{\mu} - \bar{\mathbf{x}}_i'\boldsymbol{\beta}. \tag{3.2.14}$$

Substituting (3.2.13) and (3.2.14) into (3.2.12) and solving the marginal condition for $\boldsymbol{\beta}$, we obtain (3.2.5).

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

When var $(u_{it}) = \sigma_i^2$, the LSDV estimator is no longer BLUE. However, it remains consistent. An efficient estimator is to apply the weighted least-squares estimator where each $(y_{it}, \mathbf{x}'_{it}, 1)$ is weighted by the inverse of σ_i before applying the LSDV estimator. An initial estimator of σ_i can be obtained from

$$\hat{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T}(y_{it} - \hat{\alpha}_i^* - \mathbf{x}'_{it}\hat{\boldsymbol{\beta}}_{cv})^2. \tag{3.2.15}$$

3.3 RANDOM EFFECTS MODELS: ESTIMATION OF VARIANCE-COMPONENTS MODELS

In Section 3.2 we discussed the estimation of linear regression models when the effects of omitted individual-specific variables (α_i) are treated as fixed constants over time. In this section we treat the individual-specific effects, α_i, like u_{it}, as random variables.

It is a standard practice in the regression analysis to assume that the large number of factors that affect the value of the dependent variable, but that have not been explicitly included as explanatory variables, can be appropriately summarized by a random disturbance. When numerous individual units are observed over time, it is sometimes assumed that some of the omitted variables will represent factors peculiar to both the individual units and time periods for which observations are obtained, whereas other variables will reflect individual differences that tend to affect the observations for a given individual in more or less the same fashion over time. Still other variables may reflect factors peculiar to specific time periods, but affecting individual units more or less equally. Thus, the residual, v_{it}, is often assumed to consist of three components:[5]

$$v_{it} = \alpha_i + \lambda_t + u_{it}, \tag{3.3.1}$$

However, the sample provides information only about the joint density of $(y_{it}, \mathbf{x}'_{it})$, $f(\mathbf{y}_i, \mathbf{x}_i)$, not the joint density of $f(y_i, \mathbf{x}_i, \alpha_i, \boldsymbol{\lambda})$, where $\mathbf{x}_i$ denotes the $TK \times 1$ observed $\mathbf{x}_{it}$, and $\boldsymbol{\lambda}$ denotes the $T \times 1$ vector $(\lambda_1, \dots, \lambda_T)$. Since

$$\begin{aligned} f(\mathbf{y}_i, \mathbf{x}_i) &= f(\mathbf{y}_i \mid \mathbf{x}_i) f(\mathbf{x}_i) \\ &= \left[\int f(\mathbf{y}_i \mid \mathbf{x}_i, \alpha_i, \boldsymbol{\lambda}) f(\alpha_i, \boldsymbol{\lambda} \mid \mathbf{x}_i) d\alpha_i d\boldsymbol{\lambda}\right] \cdot f(\mathbf{x}_i), \end{aligned} \tag{3.3.2}$$

we need to know $f(\alpha_i, \boldsymbol{\lambda}_t \mid \mathbf{x}_i)$ to derive the random-effects estimator. However, α_i and λ_t are unobserved. A common assumption for the random-effects model

[5] Note that we follow the formulation of (3.2.10) by treating α_i and λ_t as deviations from the population mean. For ease of exposition we also restrict our attention to the homoscedastic variances of α_i and λ_t. For the heteroscedasticity generalization of the error-component model, see Chapter 3, Section 3.7 or Mazodier and Trognon (1978) and Wansbeek and Kapteyn (1982). For a test of individual heteroscedasticity, see Holly and Gardiol (2000).

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

is to assume

$$f(\alpha_i, \lambda_t \mid \mathbf{x}_i) = f(\alpha_i, \lambda_t) = f(\alpha_i)f(\lambda_t). \tag{3.3.3}$$

In other words, we assume that

$$\begin{gathered} E\alpha_i = E\lambda_t = Eu_{it} = 0, \quad E\alpha_i\lambda_t = E\alpha_i u_{it} = E\lambda_t u_{it} = 0, \\ E\alpha_i\alpha_j = \begin{cases} \sigma_\alpha^2 & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases} \\ E\lambda_t\lambda_s = \begin{cases} \sigma_\lambda^2 & \text{if } t = s, \\ 0 & \text{if } t \neq s, \end{cases} \\ Eu_{it}u_{js} = \begin{cases} \sigma_u^2 & \text{if } i = j,\ t = s, \\ 0 & \quad \text{otherwise,} \end{cases} \end{gathered} \tag{3.3.4}$$

and

$$E\alpha_i\mathbf{x}'_{it} = E\lambda_t\mathbf{x}'_{it} = Eu_{it}\mathbf{x}'_{it} = \mathbf{0}'.$$

The variance of y_{it} conditional on $\mathbf{x}_{it}$ is, from (3.3.1) and (3.3.4), $\sigma_y^2 = \sigma_\alpha^2 + \sigma_\lambda^2 + \sigma_u^2$. The variances $\sigma_\alpha^2, \sigma_\lambda^2$, and σ_u^2 are accordingly called variance components; each is a variance in its own right and is a component of σ_y^2. Therefore, this kind of model is sometimes referred to as a variance-components (or error-components) model.

For ease of exposition we assume $\lambda_t = 0$ for all t in this and the following three sections. That is, we concentrate on models of the form (3.2.12).

Rewriting (3.2.12) in vector form, we have

$$\underset{T\times 1}{\mathbf{y}_i} = \underset{T\times(K+1)}{\tilde{X}_i}\ \underset{(K+1)\times 1}{\boldsymbol{\delta}} + \underset{T\times 1}{\mathbf{v}_i}, \quad i = 1, 2, \ldots, N, \tag{3.3.5}$$

where $\tilde{X}_i = (\mathbf{e}, X_i)$, $\boldsymbol{\delta}' = (\mu, \boldsymbol{\beta}')$, $\mathbf{v}'_i = (v_{i1}, \ldots, v_{iT})$, and $v_{it} = \alpha_i + u_{it}$. The presence of α_i creates correlations of v_{it} over time for a given individual, although v_{it} remains uncorrelated across individuals. The variance–covariance matrix of $\mathbf{v}_i$ takes the form,

$$E\mathbf{v}_i\mathbf{v}'_i = \sigma_u^2 I_T + \sigma_\alpha^2 \mathbf{e}\mathbf{e}' = V. \tag{3.3.6}$$

Its inverse is (see Graybill 1969; Nerlove 1971b; Wallace and Hussain 1969)

$$V^{-1} = \frac{1}{\sigma_u^2}\left[I_T - \frac{\sigma_\alpha^2}{\sigma_u^2 + T\sigma_\alpha^2}\mathbf{e}\mathbf{e}'\right]. \tag{3.3.7}$$

3.3.1 Covariance Estimation

Regardless of whether the α_i's are treated as fixed or as random, the individual-specific effects for a given sample can be swept out by the idempotent (covariance) transformation matrix Q [equation (3.2.6)], because $Q\mathbf{e} = \mathbf{0}$, and hence

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

$Q\mathbf{v}_i = Q\mathbf{u}_i$. Thus, premultiplying (3.3.5) by Q, we have

$$\begin{aligned} Q\mathbf{y}_i &= Q\mathbf{e}\mu + QX_i\boldsymbol{\beta} + Q\mathbf{e}\alpha_i + Q\mathbf{u}_i \\ &= QX_i\boldsymbol{\beta} + Q\mathbf{u}_i. \end{aligned} \tag{3.3.8}$$

Applying the least-squares method to (3.3.8), we obtain the CV estimator (3.2.8) of $\boldsymbol{\beta}$. We estimate μ by $\hat{\mu} = \bar{y} - \bar{\mathbf{x}}'\hat{\boldsymbol{\beta}}_{cv}$.

Whether α_i are treated as fixed or random, the CV estimator of $\boldsymbol{\beta}$ is unbiased and consistent either N or T or both tend to infinity. However, whereas the CV estimator is the BLUE under the assumption that α_i are fixed constants, the CV estimator is not the BLUE in finite samples when α_i are assumed random. The BLUE in the latter case is the generalized least-squares (GLS) estimator.[6] Moreover, if the explanatory variables contain some time-invariant variables, $\mathbf{z}_i$, then $\mathbf{e}\mathbf{z}_i'$ and $\mathbf{e}$ are perfectly correlated. Their coefficients cannot be estimated by CV because the CV transformation eliminates $\mathbf{z}_i$ from (3.3.8).

3.3.2 Generalized Least-Squares (GLS) Estimation

Under (3.3.4), $E(\mathbf{v}_i \mid \mathbf{x}_i) = 0$. The least-squares method can be applied. However, because v_{it} and v_{is} both contain α_i, the residuals of (3.3.5) are serially correlated. To get efficient estimates of $\boldsymbol{\delta}' = (\mu, \boldsymbol{\beta}')$ we have to use the GLS method. The normal equations for the GLS estimators are

$$\left[\sum_{i=1}^{N} \tilde{X}_i' V^{-1} \tilde{X}_i\right] \hat{\boldsymbol{\delta}}_{\text{GLS}} = \left[\sum_{i=1}^{N} \tilde{X}_i' V^{-1} \mathbf{y}_i\right]. \tag{3.3.9}$$

Following Maddala (1971a), we write V^{-1} [equation (3.3.7)] as

$$V^{-1} = \frac{1}{\sigma_u^2}\left[\left(I_T - \frac{1}{T}\mathbf{e}\mathbf{e}'\right) + \psi \cdot \frac{1}{T}\mathbf{e}\mathbf{e}'\right] = \frac{1}{\sigma_u^2}\left[Q + \psi \cdot \frac{1}{T}\mathbf{e}\mathbf{e}'\right], \tag{3.3.10}$$

where

$$\psi = \frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}. \tag{3.3.11}$$

Hence, (3.3.9) can conveniently be written as

$$[W_{\tilde{x}\tilde{x}} + \psi B_{\tilde{x}\tilde{x}}] \begin{bmatrix} \hat{\mu} \\ \hat{\boldsymbol{\beta}} \end{bmatrix}_{\text{GLS}} = [W_{\tilde{x}y} + \psi B_{\tilde{x}y}], \tag{3.3.12}$$

[6] For details, see Section 3.3.2.

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

where

$$T_{\tilde{x}\tilde{x}} = \sum_{i=1}^{N} \tilde{X}_i' \tilde{X}_i, \qquad T_{\tilde{x}y} = \sum_{i=1}^{N} \tilde{X}_i' \mathbf{y}_i,$$

$$B_{\tilde{x}\tilde{x}} = \frac{1}{T} \sum_{i=1}^{N} (\tilde{X}_i' \mathbf{e}\mathbf{e}' \tilde{X}_i), \quad B_{\tilde{x}y} = \frac{1}{T} \sum_{i=1}^{N} (\tilde{X}_i' \mathbf{e}\mathbf{e}' y_i),$$

$$W_{\tilde{x}\tilde{x}} = T_{\tilde{x}\tilde{x}} - B_{\tilde{x}\tilde{x}}, \qquad W_{\tilde{x}y} = T_{\tilde{x}y} - B_{\tilde{x}y}.$$

The matrices $B_{\tilde{x}\tilde{x}}$ and $B_{\tilde{x}y}$ contain the sums of squares and sums of cross products between groups, $W_{\tilde{x}\tilde{x}}$ and $W_{\tilde{x}y}$ are the corresponding matrices within groups, and $T_{\tilde{x}\tilde{x}}$ and $T_{\tilde{x}y}$ are the corresponding matrices for total variation.

Solving (3.3.12), we have

$$\begin{bmatrix} \psi NT & \psi T \sum_{i=1}^{N} \bar{\mathbf{x}}_i' \\ \psi T \sum_{i=1}^{N} \bar{\mathbf{x}}_i & \sum_{i=1}^{N} X_i' Q X_i + \psi T \sum_{i=1}^{N} \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i' \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\boldsymbol{\beta}} \end{bmatrix}_{\text{GLS}} . \tag{3.3.13}$$

$$= \begin{bmatrix} \psi NT \bar{y} \\ \sum_{i=1}^{N} X_i' Q \mathbf{y}_i + \psi T \sum_{i=1}^{N} \bar{\mathbf{x}}_i \bar{y}_i \end{bmatrix}$$

Using the formula of the partitioned inverse, we obtain

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{\text{GLS}} &= \left[\frac{1}{T} \sum_{i=1}^{N} X_i' Q X_i + \psi \sum_{i=1}^{N} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})' \right]^{-1} \\ &\quad \cdot \left[\frac{1}{T} \sum_{i=1}^{N} X_i' Q \mathbf{y}_i + \psi \sum_{i=1}^{N} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{y}_i - \bar{y}) \right] \qquad (3.3.14) \\ &= \Delta \hat{\boldsymbol{\beta}}_b + (I_K - \Delta) \hat{\boldsymbol{\beta}}_{cv}, \\ \hat{\mu}_{\text{GLS}} &= \bar{y} - \bar{\mathbf{x}}' \hat{\boldsymbol{\beta}}_{\text{GLS}}, \end{aligned}$$

where

$$\begin{aligned} \Delta = \psi T &\left[\sum_{i=1}^{N} X_i' Q X_i + \psi T \sum_{i=1}^{N} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})' \right]^{-1} \\ &\cdot \left[\sum_{i=1}^{N} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})' \right], \end{aligned}$$

$$\hat{\boldsymbol{\beta}}_b = \left[\sum_{i=1}^{N} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})' \right]^{-1} \left[\sum_{i=1}^{N} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{y}_i - \bar{y}) \right].$$

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

The estimator $\hat{\boldsymbol{\beta}}_b$ is called the between-group estimator because it ignores variation within the group.

The GLS estimator (3.3.14) is a weighted average of the between-group and within-group estimators. If $\psi \rightarrow 1$, $\hat{\boldsymbol{\delta}}_{\text{GLS}}$ converges to the OLS estimator $T_{\tilde{x}\tilde{x}}^{-1} T_{\tilde{x}y}$. If $\psi \rightarrow 0$, the GLS estimator for $\boldsymbol{\beta}$ becomes the CV estimator (LSDV) [equation (3.2.5)]. In essence, ψ measures the weight given to the between-group variation. In the LSDV (or fixed-effects model) procedure, this source of variation is completely ignored. The OLS procedure corresponds to $\psi = 1$. The between-group and within-group variations are just added up. Thus, one can view the OLS and LSDV as somewhat all-or-nothing ways of utilizing the between-group variation. The procedure of treating α_i as random provides a solution intermediate between treating them all as different and treating them all as equal, as implied by the GLS estimator given in (3.3.14).

If $[W_{\tilde{x}\tilde{x}} + \psi B_{\tilde{x}\tilde{x}}]$ is nonsingular, the covariance matrix of GLS estimators of $\boldsymbol{\delta}$ can be written as

$$\text{Var}\begin{bmatrix} \hat{\mu} \\ \hat{\boldsymbol{\beta}} \end{bmatrix}_{\text{GLS}} = \sigma_u^2 [W_{\tilde{x}\tilde{x}} + \psi B_{\tilde{x}\tilde{x}}]^{-1} \tag{3.3.15}$$

$$= \sigma_u^2 \left[\begin{pmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & \sum_{i=1}^N X_i' Q X_i \end{pmatrix} + T\psi \begin{pmatrix} N & \sum_{i=1}^N \bar{\mathbf{x}}_i' \\ \sum_{i=1}^N \bar{\mathbf{x}}_i & \sum_{i=1}^N \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i' \end{pmatrix} \right]^{-1}.$$

Using the formula for partitioned inversion (e.g., Rao 1973, Chapter 2; Theil 1971, Chapter 1), we obtain

$$\text{Var}\,(\hat{\boldsymbol{\beta}}_{\text{GLS}}) = \sigma_u^2 \left[\sum_{i=1}^N X_i' Q X_i + T\psi \sum_{i=1}^N (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})' \right]^{-1}. \tag{3.3.16}$$

Because $\psi > 0$, we see immediately that the difference between the covariance matrices of $\hat{\boldsymbol{\beta}}_{cv}$ and $\hat{\boldsymbol{\beta}}_{\text{GLS}}$ is a positive semidefinite matrix. However, for fixed N, as $T \rightarrow \infty$, $\psi \rightarrow 0$. Thus, under the assumption that $(1/NT)\Sigma_{i=1}^N X_i' X_i$ and $(1/NT)\Sigma_{i=1}^N X_i' Q X_i$, converge to finite positive definitive matrices, when $T \rightarrow \infty$, we have $\hat{\beta}_{\text{GLS}} \rightarrow \hat{\beta}_{cv}$ and $\text{Var}(\sqrt{T}\hat{\boldsymbol{\beta}}_{\text{GLS}}) \rightarrow \text{Var}\,(\sqrt{T}\hat{\boldsymbol{\beta}}_{cv})$. This is because when $T \rightarrow \infty$, we have an infinite number of observations for each i. Therefore, we can consider each α_i as a random variable that has been drawn once and forever so that for each i we can pretend that they are just like fixed parameters.

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

Computation of the GLS estimator can be simplified by noting the special form of V^{-1} (3.3.10). Let $P = [I_T - (1-\psi^{1/2})(1/T)\mathbf{e}\mathbf{e}']$; we have $V^{-1} = \frac{1}{\sigma_u^2}P'P$. Premultiplying (3.3.5) by the transformation matrix, P, we obtain the GLS estimator (3.3.12) by applying the least-squares method to the transformed model (Theil 1971, Chapter 6). This is equivalent to first transforming the data by subtracting a fraction $(1-\psi^{1/2})$ of individual means $\bar{y}_i$, and $\bar{\mathbf{x}}_i$ from their corresponding y_{it} and $\mathbf{x}_{it}$, then regressing $[y_{it} - (1-\psi^{1/2})\bar{y}_i]$ on a constant and $[\mathbf{x}_{it} - (1-\psi^{1/2})\bar{\mathbf{x}}_i]$. Since $\psi^{1/2} \neq 0$, $\mathbf{x}_{it} - (1-\psi^{1/2})\bar{\mathbf{x}}_i$ is different from 0 even $\mathbf{x}_{it}$ is time-invariant. In other words, the random-effects model allows one to estimate the coefficients of both time-varying and time-invariant variables while the fixed-effects model only allows us to estimate the coefficients of time-varying explanatory variables.

The GLS requires that of σ_u^2 and σ_α^2 be known. If the variance components, σ_u^2 and σ_α^2, are unknown, we can use two-step GLS estimation (feasible GLS, FGLS). In the first step we estimate the variance components using some consistent estimators. In the second step we substitute their estimated values into (3.3.10) or its equivalent form. Noting that $\bar{y}_i = \mu + \boldsymbol{\beta}'\bar{\mathbf{x}}_i + \alpha_i + \bar{u}_i$ and $(y_{it} - \bar{y}_i) = \boldsymbol{\beta}'(\mathbf{x}_{it} - \bar{\mathbf{x}}_i) + (u_{it} - \bar{u}_i)$, we can use the within- and between-group residuals to estimate σ_u^2 and σ_α^2 respectively, by[7]

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^N \sum_{t=1}^T [(y_{it} - \bar{y}_i) - \hat{\boldsymbol{\beta}}_{cv}'(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)]^2}{N(T-1) - K}, \tag{3.3.17}$$

and

$$\hat{\sigma}_\alpha^2 = \frac{\sum_{i=1}^N (\bar{y}_i - \tilde{\mu} - \tilde{\boldsymbol{\beta}}'\bar{\mathbf{x}}_i)^2}{N - (K+1)} - \frac{1}{T}\hat{\sigma}_u^2, \tag{3.3.18}$$

where $(\tilde{\mu}, \tilde{\boldsymbol{\beta}}')' = B_{\tilde{x}\tilde{x}}^{-1} B_{\tilde{x}\tilde{y}}$. When the sample size is large (in the sense of $N \to \infty$, $T \to \infty$), the two-step GLS estimator will have the same asymptotic efficiency as the GLS procedure with known variance components (Fuller and Battese 1974). Even for moderate sample size [for $T \geq 3$, $N - (K+1) \geq 9$; for $T = 2$, $N - (K+1) \geq 10$], the two-step procedure is still more efficient than the CV (or within-group) estimator in the sense that the difference between the covariance matrices of the CV estimator and the two-step estimator is non-negative definite (Taylor 1980).

Amemiya (1971) has discussed efficient estimation of the variance components. However, substituting more efficiently estimated variance components into (3.3.12) need not lead to more efficient estimates of μ and $\boldsymbol{\beta}$ (Maddala and Mount 1973; Taylor 1980).

[7] Equation (3.3.18) may yield a negative estimate of σ_α^2. For additional discussion on this issue, see Section 3.3.3.

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

3.3.3 Maximum-Likelihood Estimation

When α_i and u_{it} are random and normally distributed, the logarithm of the likelihood function is

$$\begin{aligned}\log L &= -\frac{NT}{2}\log 2\pi - \frac{N}{2}\log \mid V \mid \\ &\quad -\frac{1}{2}\sum_{i=1}^{N}(\mathbf{y}_i - \mathbf{e}\mu - X_i\boldsymbol{\beta})'V^{-1}(\mathbf{y}_i - \mathbf{e}\mu - X_i\boldsymbol{\beta}) \\ &= -\frac{NT}{2}\log 2\pi - \frac{N(T-1)}{2}\log \sigma_u^2 - \frac{N}{2}\log\left(\sigma_u^2 + T\sigma_\alpha^2\right) \\ &\quad -\frac{1}{2\sigma_u^2}\sum_{i=1}^{N}(\mathbf{y}_i - \mathbf{e}\mu - X_i\boldsymbol{\beta})'Q(\mathbf{y}_i - \mathbf{e}\mu - X_i\boldsymbol{\beta}) \\ &\quad -\frac{T}{2(\sigma_u^2 + T\sigma_\alpha^2)}\sum_{i=1}^{N}(\bar{y}_i - \mu - \boldsymbol{\beta}'\bar{\mathbf{x}}_i)^2, \end{aligned} \tag{3.3.19}$$

where the second equality follows from (3.3.10) and

$$\mid V \mid = \sigma_u^{2(T-1)}\left(\sigma_u^2 + T\sigma_\alpha^2\right). \tag{3.3.20}$$

The maximum-likelihood estimator (MLE) of $(\mu, \boldsymbol{\beta}', \sigma_u^2, \sigma_\alpha^2) = \tilde{\boldsymbol{\theta}}'$ is obtained by solving the following first-order conditions simultaneously:

$$\frac{\partial \log L}{\partial \mu} = \frac{T}{(\sigma_u^2 + T\sigma_\alpha^2)}\sum_{i=1}^{N}\left(\bar{y}_i - \mu - \bar{\mathbf{x}}_i'\boldsymbol{\beta}\right) = 0, \tag{3.3.21}$$

$$\begin{aligned}\frac{\partial \log L}{\partial \boldsymbol{\beta}'} &= \frac{1}{\sigma_u^2}\Bigg[\sum_{i=1}^{N}(\mathbf{y}_i - \mathbf{e}\mu - X_i\boldsymbol{\beta})'QX_i \\ &\quad + \frac{T\sigma_u^2}{\left(\sigma_u^2 + T\sigma_\alpha^2\right)}\sum_{i=1}^{N}(\bar{y}_i - \mu - \bar{\mathbf{x}}_i'\boldsymbol{\beta})\bar{\mathbf{x}}_i'\Bigg] = \mathbf{0}', \end{aligned} \tag{3.3.22}$$

$$\begin{aligned}\frac{\partial \log L}{\partial \sigma_u^2} &= -\frac{N(T-1)}{2\sigma_u^2} - \frac{N}{2(\sigma_u^2 + T\sigma_\alpha^2)} + \frac{1}{2\sigma_u^4}\sum_{i=1}^{N}(\mathbf{y}_i - \mathbf{e}\mu \\ &\quad - X_i\boldsymbol{\beta})'Q(\mathbf{y}_i - \mathbf{e}\mu - X_i\boldsymbol{\beta}) \\ &\quad + \frac{T}{2\left(\sigma_u^2 + T\sigma_\alpha^2\right)^2}\sum_{i=1}^{N}(\bar{y}_i - \mu - \bar{\mathbf{x}}_i'\boldsymbol{\beta})^2 = 0, \end{aligned} \tag{3.3.23}$$

$$\frac{\partial \log L}{\partial \sigma_\alpha^2} = -\frac{NT}{2\left(\sigma_u^2 + T\sigma_\alpha^2\right)} + \frac{T^2}{2\left(\sigma_u^2 + T\sigma_\alpha^2\right)^2}\sum_{i=1}^{N}(\bar{y}_i - \mu \ - \bar{\mathbf{x}}_i'\boldsymbol{\beta})^2 = 0. \tag{3.3.24}$$

Simultaneous solution of (3.3.21)–(3.3.24) is complicated. The Newton–Raphson iterative procedure can be used to solve for the MLE. The procedure uses an initial trial value of $\tilde{\boldsymbol{\theta}}$, $\hat{\tilde{\boldsymbol{\theta}}}^{(1)}$, to start the iteration by substituting it into the formula

$$\hat{\tilde{\boldsymbol{\theta}}}^{(j)} = \hat{\tilde{\boldsymbol{\theta}}}^{(j-1)} - \left[\frac{\partial^2 \log L}{\partial \tilde{\boldsymbol{\theta}} \partial \tilde{\boldsymbol{\theta}}'}\right]^{-1}_{\tilde{\boldsymbol{\theta}}=\hat{\tilde{\boldsymbol{\theta}}}^{(j-1)}} \left. \frac{\partial \log L}{\partial \tilde{\boldsymbol{\theta}}} \right|_{\tilde{\boldsymbol{\theta}}=\hat{\tilde{\boldsymbol{\theta}}}^{(j-1)}} \tag{3.3.25}$$

to obtain a revised estimate of $\tilde{\boldsymbol{\theta}}$, $\hat{\tilde{\boldsymbol{\theta}}}^{(2)}$. The process is repeated until the jth iterative solution $\hat{\tilde{\boldsymbol{\theta}}}^{(j)}$ is close to the $(j-1)$th iterative solution $\hat{\tilde{\boldsymbol{\theta}}}^{(j-1)}$.

Alternatively, we can use a sequential iterative procedure to obtain the MLE. We note that from (3.3.21) and (3.3.22) we have

$$\begin{aligned} \begin{bmatrix} \hat{\mu} \\ \hat{\boldsymbol{\beta}} \end{bmatrix} &= \left[\sum_{i=1}^{N} \tilde{X}_i' V^{-1} \tilde{X}_i\right]^{-1} \left[\sum_{i=1}^{N} \tilde{X}_i' V^{-1} \mathbf{y}_i\right] \\ &= \left\{ \sum_{i=1}^{N} \begin{bmatrix} \mathbf{e}' \\ X_i' \end{bmatrix} \left[I_T - \frac{\sigma_\alpha^2}{\sigma_u^2 + T\sigma_\alpha^2} \mathbf{e}\mathbf{e}' \right] (\mathbf{e}, X_i) \right\}^{-1} \\ &\quad \cdot \left\{ \sum_{i=1}^{N} \begin{bmatrix} \mathbf{e}' \\ X_i' \end{bmatrix} \left[I_T - \frac{\sigma_\alpha^2}{\sigma_u^2 + T\sigma_\alpha^2} \mathbf{e}\mathbf{e}' \right] \mathbf{y}_i \right\}. \end{aligned} \tag{3.3.26}$$

Substituting (3.3.24) into (3.3.23), we have

$$\hat{\sigma}_u^2 = \frac{1}{N(T-1)} \sum_{i=1}^{N} (\mathbf{y}_i - \mathbf{e}\mu - X_i \boldsymbol{\beta})' Q (\mathbf{y}_i - \mathbf{e}\mu - X_i \boldsymbol{\beta}). \tag{3.3.27}$$

From (3.3.24) we have

$$\hat{\sigma}_\alpha^2 = \frac{1}{N} \sum_{i=1}^{N} (\bar{y}_i - \hat{\mu} - \bar{\mathbf{x}}_i' \hat{\boldsymbol{\beta}})^2 - \frac{1}{T} \hat{\sigma}_u^2. \tag{3.3.28}$$

Thus, we can obtain the MLE by first substituting an initial trial value of $\sigma_\alpha^2/(\sigma_u^2 + T\sigma_\alpha^2)$ into (3.3.26) to estimate μ and $\boldsymbol{\beta}$, and then estimate σ_u^2 by (3.3.27) using the solution of (3.3.26). Substituting the solutions of (3.3.26) and (3.3.27) into (3.3.28), we obtain an estimate of σ_α^2. Then we repeat the process by substituting the new values of σ_u^2 and σ_α^2 into (3.3.26) to obtain new estimates of μ and $\boldsymbol{\beta}$, and so on until the solution converges.

When T is fixed and N goes to infinity, the MLE is consistent and asymptotically normally distributed with variance–covariance matrix

$$\operatorname{Var}\left(\sqrt{N}\hat{\tilde{\boldsymbol{\theta}}}_{\text{MLE}}\right) = NE\left[-\frac{\partial^2 \log L}{\partial\tilde{\boldsymbol{\theta}}\partial\tilde{\boldsymbol{\theta}}'}\right]^{-1}$$

$$= \begin{bmatrix} \frac{T}{\sigma^2} & \frac{T}{\sigma^2}\frac{1}{N}\sum_{i=1}^{N}\bar{\mathbf{x}}_i' & 0 & 0 \\ & \frac{1}{\sigma_u^2}\frac{1}{N}\sum_{i=1}^{N} X_i'\left(I_T - \frac{\sigma_\alpha^2}{\sigma^2}\mathbf{e}\mathbf{e}'\right)X_i & \mathbf{0} & \mathbf{0} \\ & & \frac{T-1}{2\sigma_u^2} + \frac{1}{2\sigma^4} & \frac{T}{2\sigma^4} \\ & & & \frac{T^2}{2\sigma^4} \end{bmatrix}^{-1} \qquad (3.3.29)$$

where $\sigma^2 = \sigma_u^2 + T\sigma_\alpha^2$. When N is fixed and T tends to infinity, the MLEs of μ, $\boldsymbol{\beta}$ and σ_u^2 converge to the CV estimator, and are consistent, but the MLE of σ_α^2 is inconsistent. This is because when N is fixed, there is not sufficient variation in α_i no matter how large T is; for details, see Anderson and Hsiao (1981, 1982).

Although the MLE is asymptotically efficient, sometimes simultaneous solution of (3.3.21)–(3.3.24) yields an estimated value of σ_α^2 that is negative.[8] When there is a unique solution to the partial derivative equations (3.3.21)–(3.3.24), with $\sigma_u^2 > 0, \sigma_\alpha^2 > 0$, the solution is the MLE. However, when we constrain $\sigma_u^2 \geq 0$ and $\sigma_\alpha^2 \geq 0$, a boundary solution may occur. The solution, then, no longer satisfies all the derivative equations (3.3.21)–(3.3.24). Maddala (1971a) has shown that the boundary solution of $\sigma_u^2 = 0$ cannot occur, but the boundary solution of $\sigma_\alpha^2 = 0$ will occur when $T_{yy} - T_{\tilde{x}y}' T_{\tilde{x}\tilde{x}}^{-1} T_{\tilde{x}y} > T[B_{yy} - 2T_{\tilde{x}y}' T_{\tilde{x}\tilde{x}}^{-1} T_{\tilde{x}y} + T_{\tilde{x}y}' T_{\tilde{x}\tilde{x}}^{-1} B_{\tilde{x}\tilde{x}} T_{\tilde{x}\tilde{x}}^{-1} T_{\tilde{x}y}]$. However, the probability of a boundary solution tends to 0 when either T or N tends to infinity.

3.4 FIXED EFFECTS OR RANDOM EFFECTS

3.4.1 An Example

In previous sections we discussed the estimation of a linear regression model (3.2.1) when the effects, α_i, are treated either as fixed or as random. Whether to treat the effects as fixed or random makes no difference when N is fixed and

[8] The negative-variance-components problem also arises in the two-step GLS method. As one can see from (3.3.17) and (3.3.18) that there is no guarantee that (3.3.18) necessarily yields a positive estimate of σ_α^2. A practical guide in this situation is to replace a negative estimated variance component by its boundary value, zero. See Baltagi (1981b) and Maddala and Mount (1973) for a Monte Carlo studies of the desirable results of using this procedure in terms of the mean square error of the estimate. For additional discussion of the MLE of random effects model, see Breusch (1987).

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

T is large because both the LSDV estimator (3.2.8) and the generalized least-squares estimator (3.3.14) become the same estimator. When T is finite and N is large, whether to treat the effects as fixed or random is not an easy question to answer. It can make a surprising amount of difference in the estimates of the parameters. In fact, when only a few observations are available for different individuals over time, it is exceptionally important to make the best use of the lesser amount of information over time for the efficient estimation of the common behavioral relationship.

For example, Hausman (1978) found that using a fixed-effects specification produced significantly different results from a random-effects specification when estimating a wage equation using a sample of 629 high school graduates followed over six years by the Michigan income dynamics study. The explanatory variables in the Hausman wage equation include a piecewise-linear representation of age, the presence of unemployment or poor health in the previous year, and dummy variables for self-employment, living in the South, or living in a rural area. The fixed-effects specification was estimated using (3.2.5).[9] The random-effects specification was estimated using (3.3.14). The results are reproduced in Table 3.3. In comparing these two estimates, it is apparent that the effects of unemployment, self-employment, and geographical location differ widely (relative to their standard errors) in the two models.

3.4.2 Conditional Inference or Unconditional (Marginal) Inference

If the effects of omitted variables can be appropriately summarized by a random variable and the individual (or time) effects represent the ignorance of the investigator, it does not seem reasonable to treat one source of ignorance (α_i) as fixed and the other source of ignorance (u_{it}) as random. It appears that one way to unify the fixed-effects and random-effects models is to assume from the outset that the effects are random. The fixed-effects model is viewed as one in which investigators make inferences conditional on the effects that are in the sample. The random-effects model is viewed as one in which investigators make unconditional or marginal inferences with respect to the population of all effects. There is really no distinction in the "nature (of the effect)." It is up to the investigator to decide whether to make inference with respect to the population characteristics or only with respect to the effects that are in the sample.

In general, whether one wishes to consider the conditional likelihood function or the marginal likelihood function depends on the context of the data, the manner in which they were gathered, and the environment from which they came. For instance, consider an example in which several technicians provide maintenance for machines. The effects of technicians can be assumed random if the technicians are all randomly drawn from a common population. However, if the situation were one of analyzing just a few individuals, say five or six,

[9] We note that the fixed-effects estimator, although not efficient, is consistent under the random-effects formulation (Chapter 3, Section 3.3.1).

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

Table 3.3. *Wage equations (dependent variable: log wage[a])*

Variable	Fixed effects	Random effects
1. Age 1 (20–35)	0.0557	0.0393
	(0.0042)	(0.0033)
2. Age 2 (35–45)	0.0351	0.0092
	(0.0051)	(0.0036)
3. Age 3 (45–55)	0.0209	−0.0007
	(0.0055)	(0.0042)
4. Age 4 (55–65)	0.0209	−0.0097
	(0.0078)	(0.0060)
5. Age 5 (65–)	−0.0171	−0.0423
	(0.0155)	(0.0121)
6. Unemployed previous year	−0.0042	−0.0277
	(0.0153)	(0.0151)
7. Poor health previous year	−0.0204	−0.0250
	(0.0221)	(0.0215)
8. Self-employment	−0.2190	−0.2670
	(0.0297)	(0.0263)
9. South	−0.1569	−0.0324
	(0.0656)	(0.0333)
10. Rural	−0.0101	−0.1215
	(0.0317)	(0.0237)
11. Constant	—	0.8499
	—	(0.0433)
s^2	0.0567	0.0694
Degrees of freedom	3,135	3,763

[a] 3,774 observations; standard errors are in parentheses.
Source: Hausman (1978).

and the sole interest lay in just these individuals, and if we want to assess differences between those specific technicians, then the fixed-effects model is more appropriate. On the other hand, if an experiment involves hundreds of individuals who are considered a random sample from some larger population, random effects would be more appropriate. The situation to which a model applies and the inferences based on it are the deciding factors in determining whether we should treat effects as random or fixed. When inferences are going to be confined to the effects in the model, the effects are more appropriately considered fixed. When inferences will be made about a population of effects from which those in the data are considered to be a random sample, then the effects should be considered random.[10]

If one accepts this view, then why do the fixed-effects and random-effects approaches sometimes yield vastly different estimates of the common slope coefficients that are not supposed to vary across individuals? It appears that

[10] In this sense, if N becomes large, one would not be interested in the specific effect of each individual but rather in the characteristics of the population. A random-effects framework would be more appropriate.

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

in addition to the efficiency issue discussed earlier, there is also a different but important issue of whether or not the model is properly specified, that is, whether the differences in individual effects can be attributed to the chance mechanism.

In the random effects framework of (3.3.3)–(3.3.5), there are two fundamental assumptions. One is that the unobserved individual effects, α_i, are random draws from a common population. The other is that the explanatory variables are strictly exogenous. That is, the error terms are uncorrelated with (or orthogonal to) the past, current, and future values of the regressors,

$$\begin{aligned} E(u_{it} \mid \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}) &= E(\alpha_i \mid \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}) \\ &= E(v_{it} \mid \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}) = 0 \quad \text{for } t = 1, \ldots, T. \end{aligned} \tag{3.4.1}$$

In the aforementioned example if there are fundamental differences in the technicians, for instance, in the ability, age, years of experiences, etc., then the difference in technician cannot be attributed to a pure chance mechanism. It is more appropriate to view the technicians as drawn from heterogeneous populations and the individual effects $\alpha_i^* = \alpha_i + \mu$ representing the fundamental difference among the heterogeneous populations. If the difference in technicians, captured by α_i^* is ignored, the least-squares estimator of (3.3.5) yields

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{LS} &= \left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}})(\mathbf{x}_{it} - \bar{\mathbf{x}})'\right]^{-1}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}})(y_{it} - \bar{y})\right] \\ &= \boldsymbol{\beta} + \left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}})(\mathbf{x}_{it} - \bar{\mathbf{x}})'\right]^{-1}\left\{T\sum_{i=1}^{N}(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\alpha_i^* - \bar{\alpha})\right\} + \mathrm{o}(1) \end{aligned} \tag{3.4.2}$$

where $\bar{\alpha} = \frac{1}{N}\sum_{i=1}^{N}\alpha_i^*$. However, if the fundamental characteristics that drive α_i^*, say, ability, age, and years of experience in the example of technicians, are correlated with $\mathbf{x}_i$, then it is clear that $\frac{1}{N}\sum_{i=1}^{N}(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\alpha_i^* - \bar{\alpha})$ will not converge to 0 as $N \to \infty$. The least-squares estimator of $\boldsymbol{\beta}$ is inconsistent. The bias of $\hat{\boldsymbol{\beta}}_{LS}$ depends on the correlation between $\mathbf{x}_{it}$ and α_i^*.

On the other hand, if α_i^* (or α_i) are treated as fixed constants, then the regressors for y_{it} are $(\mathbf{x}_{it}', 1)$. As long as $(\mathbf{x}_{it}', 1)$ are uncorrelated with u_{it}, the least-squares estimators for $\boldsymbol{\beta}$ and α_i^* (or α_i) are unbiased. The issue of whether α_i^* are correlated with $\mathbf{x}_{it}$ is no longer relevant under the fixed-effects formulation. Thus, unless the distribution of α_i^* conditional on $\mathbf{x}_i$ can be appropriately formulated, it would be more appropriate to treat α_i^* as fixed and different (Hsiao and Sun 2000).

3.4.2.1 Mundlak's Formulation

Mundlak (1978a) criticized the random-effects formulation (3.3.4) on the grounds that it neglects the correlation that may exist between the effects,

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

α_i, and the explanatory variables, $\mathbf{x}_{it}$. There are reasons to believe that in many circumstances α_i and $\mathbf{x}_{it}$ are indeed correlated. For instance, consider the estimation of production function using firm data. The output of each firm, y_{it}, may be affected by unobservable managerial ability, α_i. Firms with more efficient management tend to produce more and use more inputs, X_i. Less efficient firms tend to produce less and use fewer inputs. In this situation, α_i and X_i cannot be independent. Ignoring this correlation can lead to biased estimation.

The properties of various estimators we have discussed thus far depend on the existence and extent of the relations between the X's and the effects. Therefore, we have to consider the joint distribution of these variables. However, α_i are unobservable. Mundlak (1978a) suggested that we approximate $E(\alpha_i \mid X_i)$ by a linear function. He introduced the auxiliary regression

$$\alpha_i = \sum_t \mathbf{x}'_{it}\mathbf{a}_t + \omega_i, \quad \omega_i \sim N\left(0, \sigma_\omega^2\right). \tag{3.4.3a}$$

A simple approximation to (3.4.3a) is to let

$$\alpha_i = \bar{\mathbf{x}}'_i\mathbf{a} + \omega_i, \quad \omega_i \sim N\left(0, \sigma_\omega^2\right). \tag{3.4.3b}$$

Clearly, $\mathbf{a}$ will be equal to 0 (and $\sigma_\omega^2 = \sigma_\alpha^2$) if (and only if) the explanatory variables are uncorrelated with the effects.

Substituting (3.4.3b) into (3.3.5), and stacking equations over t and i, we have

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{bmatrix} = \begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \\ \vdots \\ \tilde{X}_N \end{bmatrix} \boldsymbol{\delta} + \begin{bmatrix} \mathbf{e}\bar{\mathbf{x}}'_1 \\ \mathbf{e}\bar{\mathbf{x}}'_2 \\ \vdots \\ \mathbf{e}\bar{\mathbf{x}}'_N \end{bmatrix} \mathbf{a} + \begin{bmatrix} \mathbf{e} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \omega_1$$
$$+ \begin{bmatrix} \mathbf{0} \\ \mathbf{e} \\ \vdots \\ \mathbf{0} \end{bmatrix} \omega_2 + \cdots + \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{e}_N \end{bmatrix} \omega_N + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_N \end{bmatrix}, \tag{3.4.4}$$

where

$$E(\mathbf{u}_i + \mathbf{e}\omega_i) = \mathbf{0},$$

$$E(\mathbf{u}_i + \mathbf{e}\omega_i)(\mathbf{u}_j + \mathbf{e}\omega_j)' = \begin{cases} \sigma_u^2 I_T + \sigma_\omega^2 \mathbf{e}\mathbf{e}' = \tilde{V}, & \text{if } i = j, \\ \mathbf{0}, & \text{if } i \neq j, \end{cases}$$

$$\tilde{V}^{-1} = \frac{1}{\sigma_u^2}\left[I_T - \frac{\sigma_\omega^2}{\sigma_u^2 + T\sigma_\omega^2}\mathbf{e}\mathbf{e}'\right].$$

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

Utilizing the expression for the inverse of a partitioned matrix (Theil 1971, Chapter 1), we obtain the GLS of $(\mu, \boldsymbol{\beta}', \mathbf{a}')$ as

$$\hat{\mu}^*_{\text{GLS}} = \bar{y} - \bar{\mathbf{x}}'\hat{\boldsymbol{\beta}}_b, \tag{3.4.5}$$

$$\hat{\boldsymbol{\beta}}^*_{\text{GLS}} = \hat{\boldsymbol{\beta}}_{cv}, \tag{3.4.6}$$

$$\hat{\mathbf{a}}^*_{\text{GLS}} = \hat{\boldsymbol{\beta}}_b - \hat{\boldsymbol{\beta}}_{cv}. \tag{3.4.7}$$

Thus, in the present framework, the BLUE of $\boldsymbol{\beta}$ is the CV estimator of (3.2.1) or (3.2.10′). It does not depend on knowledge of the variance components. Therefore, Mundlak (1978a) maintained that the imaginary difference between the fixed-effects and random-effects approaches is based on an incorrect specification. In fact, applying GLS to (3.2.12) yields a biased estimator. This can be seen by noting that the GLS estimate of $\boldsymbol{\beta}$ for (3.3.5), that is, (3.3.12), can be viewed as the GLS estimate of (3.4.4) after imposing the restriction $\mathbf{a} = \mathbf{0}$. As shown in (3.3.12),

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = \Delta\hat{\boldsymbol{\beta}}_b + (I_K - \Delta)\hat{\boldsymbol{\beta}}_{CV}. \tag{3.4.8}$$

If (3.4.4) is the correct specification, $E\hat{\boldsymbol{\beta}}_b$ is equal to $\boldsymbol{\beta} + \mathbf{a}$, and $E\hat{\boldsymbol{\beta}}_{cv} = \boldsymbol{\beta}$, so that

$$E\hat{\boldsymbol{\beta}}_{\text{GLS}} = \boldsymbol{\beta} + \Delta\mathbf{a}. \tag{3.4.9}$$

This is a biased estimator if $\mathbf{a} \neq \mathbf{0}$. However, when T tends to infinity, Δ tends to 0, and $\hat{\boldsymbol{\beta}}_{\text{GLS}}$ tends to $\hat{\boldsymbol{\beta}}_{cv}$ and is asymptotically unbiased. But in the more relevant situation in which T is fixed and N tends to infinity, $\text{plim}_{N\to\infty}\hat{\boldsymbol{\beta}}_{\text{GLS}} \neq \boldsymbol{\beta}$ in Mundlak's formulation.

Though it is important to recognize the possible correlation between the effects and the explanatory variables, Mundlak's (1978a) claim that there is only one estimator and that efficiency is not a consideration in distinguishing between the random-effects and fixed-effects approaches is perhaps a bit strong. Mandlak derived (3.4.6) from the assumption that $f(\alpha_i \mid \mathbf{x}_i)$ has mean $\bar{\mathbf{x}}'\mathbf{a}$ and variance σ^2_ω for the linear model (3.2.12) only. In the dynamic, random-coefficient, and discrete-choice models to be discussed later, one can show that the two approaches do not lead to the same estimator even when one allows for the correlation between α_i and X_i following the formulation of Mundlak (1978a). Moreover, in the linear static model, if $\mathbf{a} = \mathbf{0}$, the efficient estimator is (3.3.14), not the CV estimator (3.2.8).

3.4.2.2 Conditional and Unconditional Inferences in the Presence or Absence of Correlation between Individual Effects and Attributes

To gain further intuitive notions about the differences between models (3.3.5) and (3.4.4) within the conditional and unconditional inference frameworks, we consider the following two experiments. Let a population be made up of a certain composition of red and black balls. The first experiment consists of N individuals, each picking a fixed number of balls randomly from this population

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

to form his person-specific jar. Each individual then makes T independent trials of drawing a ball from his specific jar and putting it back. The second experiment assumes that individuals have different preferences for the compositions of red and black balls for their specific jars and allows personal attributes to affect the compositions. Specifically, before making T independent trials with replacement from their respective jars, individuals are allowed to take any number of balls from the population until their compositions reach the desired proportions.

If one is interested in making inferences regarding an individual jar's composition of red and black balls, a fixed-effects model should be used, whether the sample comes from the first or the second experiment. On the other hand, if one is interested in the population composition, a marginal or unconditional inference should be used. However, the marginal distributions are different for these two cases. In the first experiment, differences in individual jars are outcomes of random sampling. The subscript i is purely a labeling device, with no substantive content. A conventional random-effects model assuming independence between α_i and $\mathbf{x}_{it}$ would be appropriate. In the second experiment, the differences in individual jars reflect differences in personal attributes. A proper marginal inference has to allow for these nonrandom effects. In other words, individuals are not random draws from a common population, but from heterogeneous populations. In Mundlek's formulation, this heterogeneity is captured by the observed attributes $\mathbf{x}_i$. For the Mundlak's formulation a marginal inference that properly allows for the correlation between individual effects (α_i) and the attributes $(\mathbf{x}_i)$ in the data-generating process gives rise to the same estimator as when the individual effects are treated as fixed. It is not that in making inferences about population characteristics, we should assume a fixed-effects model.

Formally, let u_{it} and α_i be independent normal processes that are mutually independent. In the case of the first experiment, α_i are independently distributed and independent of individual attributes, $\mathbf{x}_i$, so the distribution of α_i must be expressible as random sampling from a univerate distribution (Box and Tiao 1968; Chamberlain 1980). Thus, the conditional distribution of $\{(u_i + e\alpha_i)', \alpha_i \mid X_i\}$ is identical with the marginal distribution of $\{(\mathbf{u}_i + \mathbf{e}\alpha_i)', \alpha_i\}$,

$$\begin{bmatrix} u_{i1} + \alpha_i \\ \vdots \\ u_{iT} + \alpha_i \\ \dots \\ \alpha_i \end{bmatrix} = \left[\begin{array}{c|c} u_{i1} + \alpha_i & \\ \vdots & \\ u_{iT} + \alpha_i & X_i \\ \dots & \\ \alpha_i & \end{array}\right] \tag{3.4.10a}$$

$$\sim N\left[\begin{bmatrix} \mathbf{0} \\ \dots \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 I_T + \sigma_\alpha^2 \mathbf{e}\mathbf{e}' & \vdots\, \sigma_\alpha^2 \mathbf{e} \\ \dots & \vdots\, \dots \\ \sigma_\alpha^2 \mathbf{e}' & \vdots\, \sigma_\alpha^2 \end{bmatrix}\right].$$

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

In the second experiment, α_i may be viewed as a random draw from a heterogeneous population with mean a_i^* and variance $\sigma_{\omega_i}^2$ (Mundlak's (1978a) formulation may be viewed as a special case of this in which $E(\alpha_i \mid X_i) = a_i^* = \mathbf{a}'\bar{\mathbf{x}}_i$ and $\sigma_{\omega_i}^2 = \sigma_\omega^2$ for all i). Then the conditional distribution of $\{(\mathbf{u}_i + \mathbf{e}\alpha_i)' \vdots \alpha_i \mid X_i\}$ is

$$\begin{bmatrix} u_{i1}+\alpha_i & | & \\ \vdots & | & \\ u_{iT}+\alpha_i & | & X_i \\ \cdots & | & \\ \alpha_i & | & \end{bmatrix} \sim N\left[\begin{bmatrix} \mathbf{e}a_i^* \\ \cdots \\ a_i^* \end{bmatrix}, \begin{bmatrix} \sigma_u^2 I_T + \sigma_{\omega i}^2 \mathbf{e}\mathbf{e}' & \vdots\, \sigma_{\omega i}^2 \mathbf{e} \\ \cdots & \\ \sigma_{\omega i}^2 \mathbf{e}' & \vdots\, \sigma_{\omega i}^2 \end{bmatrix}\right]. \qquad (3.4.10b)$$

In both cases, the conditional density of $\mathbf{u}_i + \mathbf{e}\alpha_i$, given α_i, is[11]

$$\left(2\pi\sigma_u^2\right)^{T/2} \exp\left\{-\frac{1}{2\sigma_u^2}\mathbf{u}_i'\mathbf{u}_i\right\}. \qquad (3.4.11)$$

But the marginal densities of $\mathbf{u}_i + \mathbf{e}\alpha_i$, given X_i, are different [(3.4.10a) and (3.4.10b), respectively]. Under the independence assumption, $\{\mathbf{u}_i + \mathbf{e}\alpha_i \mid X_i\}$ has a common mean of 0 for $i = 1, \ldots, N$. Under the assumption that α_i and X_i are correlated or α_i is a draw from a heterogeneous population, $\{\mathbf{u}_i + \mathbf{e}\alpha_i \mid X_i\}$ has a different mean $\mathbf{e}a_i^*$ for different i.

In the linear regression model, conditional on α_i the Jacobian of transformation from $\mathbf{u}_i + \mathbf{e}\alpha_i$ to $\mathbf{y}_i$ is 1. Maximizing the conditional likelihood function of $(\mathbf{y}_1 \mid \alpha_1, X_1), \ldots, (\mathbf{y}_N \mid \alpha_N, X_N)$, treating α_i as unknown parameters, yields the CV (or within-group) estimators for both cases. Maximizing the marginal likelihood function of $(y_1, \ldots, y_N \mid X_1, \ldots, X_N)$ yields the GLS estimator for model (3.3.12) under (3.4.10a) if σ_u^2 and σ_α^2 are known, and it happens to yield the CV estimator for model (3.2.12) under (3.4.10b). In other words, there is no loss of information using a conditional approach for the case of (3.4.10b). However, there is a loss in efficiency in maximizing the conditional likelihood function for the former case [i.e., (3.4.10a)] because of the loss of degrees of freedom in estimating additional $(\alpha_1, \ldots, \alpha_N)$ unknown parameters, which leads to ignoring the information contained in the between-group variation.

The advantage of the unconditional inference is that the likelihood function may depend on only a finite number of parameters, and hence can often lead to efficient inference. The disadvantage is that the correct specification of the

[11] If $(Y^{(1)'}, Y^{(2)'})'$ is normally distributed with mean $(\boldsymbol{\mu}^{(1)'}, \boldsymbol{\mu}^{(2)'})'$ and variance–covariance matrix

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

the conditional distribution of $Y^{(1)}$ given $Y^{(2)} = \mathbf{y}^{(2)}$ is normal, with mean $\boldsymbol{\mu}^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{y}^{(2)} - \boldsymbol{\mu}^{(2)})$ and covariance matrix $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ (e.g., Anderson 1985, Section 2.5).

conditional density of $\mathbf{y}_i$ given $\mathbf{X}_i$,

$$f(\mathbf{y}_i \mid X_i) = \int f(\mathbf{y}_i \mid X_i, \alpha_i) f(\alpha_i \mid X_i)\, d\alpha_i \tag{3.4.12}$$

depends on the correct specification of $f(\alpha_i \mid X_i)$. A misspecified $f(\alpha_i \mid X_i)$ can lead to a misspecified $f(\mathbf{y}_i \mid X_i)$. Maximizing the wrong $f(\mathbf{y}_i \mid X_i)$ can lead to biased and inconsistent estimators. The bias of the GLS estimator (3.3.12) in the case that $\alpha_i \sim N(a_i^*, \sigma_{\omega i}^2)$ is not due to any fallacy of the unconditional inference, but due to the misspecification of $f(\alpha_i \mid X_i)$.

The advantage of the conditional inference is that there is no need to specify $f(\alpha_i \mid X_i)$. Therefore, if the distribution of effects cannot be represented by a simple parametric functional form (say bimodal), or one is not sure of the correlation pattern between the effects and X_i, there may be an advantage to base one's inference conditionally. For instance, in the situation that there are fundamental differences between the effects, if there are fundamental differences in the ability, years of experiences, etc. as in the previous example of technicians, then it is more appropriate to treat the technicians' effects as fixed.

The disadvantage of the conditional inference is that not only there is a loss of efficiency due to the loss of degrees of freedom of estimating the effects, but there is also an issue of incidental parameters if T is finite (Neyman–Scott 1948). A typical panel contains a large number of individuals observed over a short time period, and the number of individual effects parameters (α_i^*) increases with the number of cross-sectional dimension, N. Because an increase in N provides no information on a particular α_i^* apart from those already contained in $\mathbf{y}_i$, α_i^* cannot be consistently estimated with finite T. The condition that

$$E(u_{it} \mid \mathbf{x}_{it}) = 0 \tag{3.4.13}$$

is not informative about the common parameters, $\boldsymbol{\beta}$, in the absence of any knowledge about α_i^*. If the estimation of the incidental parameters, α_i^*, is not asymptotically independent of the estimation of the common parameters (called structural parameters in statistical literature), the conditional inference of the common parameter, $\boldsymbol{\beta}$, conditional on the inconsistently estimated α_i^*, in general, will be inconsistent.

In the case of linear static model (3.2.1) or (3.2.12), the strict exogeneity of $\mathbf{x}_{it}$ to u_{it},

$$E(u_{it} \mid \mathbf{x}_i) = 0, \quad t = 1, 2, \ldots, T, \tag{3.4.14}$$

where $\mathbf{x}_i' = (\mathbf{x}_{i1}', \ldots, \mathbf{x}_{iT}')$, implies that

$$E(u_{it} - \bar{u}_i \mid \mathbf{x}_i) = 0, \quad \begin{array}{c} t = 1, 2, \ldots, T, \\ i = 1, \ldots, N. \end{array} \tag{3.4.15}$$

Since $\boldsymbol{\beta}$ can be identified from the moment conditions of the form (3.4.15) in the linear static model and (3.4.15) no longer involves α_i^*, consistent estimators of $\boldsymbol{\beta}$ can be proposed by making use of these moment conditions (e.g., (3.2.8)).

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

Unfortunately, for nonlinear panel data models, it is in general not possible to find moment conditions that are independent of α_i^* to provide consistent estimators of common parameters.

The advantage of fixed-effects inference is that there is no need to assume that the effects are independent of $\mathbf{x}_i$. The disadvantage is that it introduces the issue of incidental parameters. Moreover, in the case of linear regression model, condition (3.4.15) implies that the coefficients of time-invariant variables cannot be estimated and the estimation of $\boldsymbol{\beta}$ makes use of only within-group variation, which is usually much smaller than between-group variation. The advantage of random-effects inference is that the number of parameters is fixed when sample size increases. It also allows the derivation of efficient estimators that make use of both within- and between-group variation. The impact of time-invariant variables can also be estimated. The disadvantage is that one has to make specific assumption about the pattern of correlation (or no correlation) between the effects and the included explanatory variables. A common assumption is that $f(\alpha_i \mid \mathbf{x}_i)$ is identical to the marginal density $f(\alpha_i)$. However, if the effects are correlated with $\mathbf{x}_{it}$ or if there is a fundamental difference among individual units, that is, conditional on $\mathbf{x}_{it}$, y_{it} cannot be viewed as a random draw from a common distribution, common random-effects model (3.3.4) is misspecified and the resulting estimator is biased. In short, the advantages of random-effects specification are the disadvantages of fixed-effects specification and the disadvantages of random-effects specification are the advantages of fixed-effects specification. Unfortunately, there is no universally accepted way to make explicit assumptions about the way in which observables and unobservables interact in all contexts.

Finally, it should be noted that the assumption of randomness does not carry with it the assumption of normality. Often this assumption is made for random effects, but it is a separate assumption made subsequent to the randomness assumption. Most estimation procedures do not require normality, although if distributional properties of the resulting estimators are to be investigated, then normality is often assumed.

3.5 TESTS FOR MISSPECIFICATION

As discussed in Section 3.4, the fundamental issue is not whether α_i should be treated fixed or random. The issue is whether or not $f(\alpha_i \mid \mathbf{x}_i) \equiv f(\alpha_i)$, or whether α_i can be viewed as random draws from a common population. In the linear regression framework, treating α_i as fixed in (3.2.12) leads to the identical estimator of $\boldsymbol{\beta}$ whether α_i is correlated with $\mathbf{x}_i$ as in (3.4.3a) or is from a heterogeneous population. Hence, for ease of reference, when α_i is correlated with $\mathbf{x}_i$, we shall follow the convention and call (3.2.12) a fixed-effects model, and when α_i is uncorrelated with $\mathbf{x}_i$, we shall call it a random-effects model.

Thus, one way to decide whether to use a fixed-effects or random-effects model is to test for misspecification of (3.3.4), where α_i is assumed random and uncorrelated with $\mathbf{x}_i$. Using Mundlak's formulation, (3.4.3a) or (3.4.3b),

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

this test can be reduced to a test of

$$H_0 : \mathbf{a} = \mathbf{0},$$

against

$$H_1 : \mathbf{a} \neq \mathbf{0}.$$

If the alternative hypothesis, H_1, holds, we use the fixed-effects model (3.2.1). If the null hypothesis, H_0, holds, we use the random-effects model (3.3.4). The ratio

$$F = \frac{\left[\Sigma_{i=1}^{N}(\mathbf{y}_i - \tilde{X}_i\hat{\boldsymbol{\delta}}_{\text{GLS}})'V^{*-1}(\mathbf{y}_i - \tilde{X}_i\hat{\boldsymbol{\delta}}_{\text{GLS}}) - \Sigma_{i=1}^{N}(\mathbf{y}_i - \tilde{X}_i\hat{\boldsymbol{\delta}}^*_{\text{GLS}} - \mathbf{e}\bar{\mathbf{x}}_i'\hat{\mathbf{a}}^*_{\text{GLS}})'V^{*-1}\cdot(\mathbf{y}_i - \tilde{X}_i\hat{\boldsymbol{\delta}}^*_{\text{GLS}} - \mathbf{e}\bar{\mathbf{x}}_i'\hat{\mathbf{a}}^*_{\text{GLS}})\right]/K}{\Sigma_{i=1}^{N}(\mathbf{y}_i - \tilde{X}_i\hat{\boldsymbol{\delta}}^*_{\text{GLS}} - \mathbf{e}\bar{\mathbf{x}}_i'\hat{\mathbf{a}}^*_{\text{GLS}})'V^{*-1}(\mathbf{y}_i - \tilde{X}_i\hat{\boldsymbol{\delta}}^*_{\text{GLS}} - \mathbf{e}\bar{\mathbf{x}}_i'\hat{\mathbf{a}}^*_{\text{GLS}})/[NT - (2K+1)]} \tag{3.5.1}$$

under H_0 has a central F distribution with K and $NT - (2K + 1)$ degrees of freedom, where $\hat{\boldsymbol{\delta}}^*_{\text{GLS}} = (\hat{\mu}_{\text{GLS}}, \hat{\boldsymbol{\beta}}'_{\text{GLS}})'$, and $\hat{\mathbf{a}}^*_{\text{GLS}}$ are given by (3.4.5)–(3.4.7), $V^{*-1} = (1/\sigma_u^2)[Q + \psi^*(1/T)\mathbf{e}\mathbf{e}']$, and $\psi^* = \sigma_u^2/(\sigma_u^2 + T\sigma_\omega^2)$. Hence, (3.5.1) can be used to test H_0 against H_1.[12]

An alternative testing procedure suggested by Hausman (1978) notes that under H_0 the GLS for (3.3.5) achieves the Cramer–Rao lower bounds, but under H_1, the GLS is a biased estimator. In contrast, the CV estimator of $\boldsymbol{\beta}$ is consistent under both H_0 and H_1. Hence, the Hausman test basically asks if the CV and GLS estimates of $\boldsymbol{\beta}$ are significantly different.

To derive the asymptotic distribution of the differences of the two estimates, Hausman makes use of the following lemma:[13]

Lemma 3.5.1: Based on a sample of N observations, consider two estimates $\hat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\beta}}_1$ that are both consistent and asymptotically normally distributed, with $\hat{\beta}_0$ attaining the asymptotic Cramer–Rao bound so that $\sqrt{N}(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta})$ is asymptotically normally distributed with variance–covariance matrix V_0. $\sqrt{N}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta})$ is asymptotically normally distributed, with mean 0 and variance–covariance matrix V_1. Let $\hat{\mathbf{q}} = \hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_0$. Then the limiting distribution of $\sqrt{N}(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta})$ and $\sqrt{N}\hat{\mathbf{q}}$ has 0 covariance: Cov $(\hat{\boldsymbol{\beta}}_0, \hat{\mathbf{q}}) = \mathbf{0}$, a zero matrix.

[12] When ψ^* is unknown, we substitute it by an estimated value and treat (3.5.1) as having an approximate F distribution.

[13] For proof, see Hausman (1978) or Rao (1973, p. 317).

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

From this lemma, it follows that $\text{Var}(\hat{\mathbf{q}}) = \text{Var}(\hat{\boldsymbol{\beta}}_1) - \text{Var}(\hat{\boldsymbol{\beta}}_0)$. Thus, Hausman suggests using the statistic[14]

$$m = \hat{\mathbf{q}}' \text{Var}(\hat{\mathbf{q}})^{-1} \hat{\mathbf{q}}, \tag{3.5.2}$$

where $\hat{\mathbf{q}} = \hat{\boldsymbol{\beta}}_{\text{CV}} - \hat{\boldsymbol{\beta}}_{\text{GLS}}$, $\text{Var}(\hat{\mathbf{q}}) = \text{Var}(\hat{\boldsymbol{\beta}}_{\text{CV}}) - \text{Var}(\hat{\boldsymbol{\beta}}_{\text{GLS}})$, to test the null hypothesis $E(\alpha_i \mid X_i) = 0$ against the alternative $E(\alpha_i \mid X_i) \neq 0$. Under the null hypothesis, this statistic is distributed asymptotically as central χ^2, with K degrees of freedom. Under the alternative, it has a noncentral χ^2 distribution with noncentrality parameter $\bar{\mathbf{q}}' \text{ Var}(\hat{\mathbf{q}})^{-1} \bar{\mathbf{q}}$, where $\bar{\mathbf{q}} = \text{plim}(\hat{\boldsymbol{\beta}}_{\text{CV}} - \hat{\boldsymbol{\beta}}_{\text{GLS}})$.

When N is fixed and T tends to infinity, $\hat{\boldsymbol{\beta}}_{\text{CV}}$ and $\hat{\boldsymbol{\beta}}_{\text{GLS}}$ become identical. However, it was shown by Ahn and Moon (2001) that the numerator and denominator of (3.5.2) approach 0 at the same speed. Therefore the ratio remains χ^2 distributed, although in this situation the fixed-effects and random-effects models become indistinguishable for all practical purposes. The more typical case in practice is that N is large relative to T, so that differences between the two estimators or two approaches are important problems.

We can use either (3.5.1) or (3.5.2) to test whether a fixed-effects or random-effects formulation is more appropriate for the wage equation cited at the beginning of Section 3.4 (Table 3.3). The advantage of the Hausman approach is that no $f(\alpha_i \mid \mathbf{x}_i)$ needs to be postulated. The χ^2 statistic for (3.5.2) computed by Hausman (1978) is 129.9. The critical value for the 1 percent significance level at 10 degrees of freedom is 23.2, a very strong indication of misspecification in the conventional random-effects model (3.3.3). Similar conclusions are also obtained by using (3.5.1). The F value computed by Hausman (1978) is 139.7, which well exceeds the 1 percent critical value. These tests imply that in the Michigan survey, important individual effects are present that are correlated with the right-hand variables. Because the random-effects estimates appear to be significantly biased with high probability, it may well be important to take account of permanent unobserved differences across individuals in estimating earnings equations using panel data.

3.6 MODELS WITH TIME- AND/OR INDIVIDUAL-INVARIANT EXPLANATORY VARIABLES AND BOTH INDIVIDUAL- AND TIME-SPECIFIC EFFECTS

3.6.1 Estimation of Models with Individual-Specific Variables

Model (3.2.12) can be generalized to a number of different directions with no fundamental change in the analysis. For instance, we can include a $1 \times p$ vector $\mathbf{z}_i'$ of individual-specific variables (such as sex, race, socioeconomic

[14] Strictly speaking, the Hausman test is more general than a test of $\Sigma_t \mathbf{x}_{it}' \mathbf{a}_t = 0$ versus $\Sigma_t \mathbf{x}_{it}' \mathbf{a}_t \neq 0$. The null of $f(\alpha_i \mid \mathbf{x}_i) = f(\alpha_i)$ implies that $\Sigma_t \mathbf{x}_{it}' \mathbf{a}_t = 0$, but not necessarily the converse. For a discussion of the general relationship between Hausman's specification testing and conventional testing procedures, see Holly (1982).

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

background variables, which vary across individual units but do not vary over time) in the specification of the equation for y_{it} and consider

$$\begin{aligned}\underset{T\times 1}{y_i} &= \underset{T\times 1}{\mathbf{e}}\ \underset{1\times 1}{\mu} + \underset{T\times p}{Z_i}\ \underset{p\times 1}{\boldsymbol{\gamma}} + \underset{T\times K}{X_i}\ \underset{K\times 1}{\boldsymbol{\beta}} \\ &\quad + \underset{T\times 1}{\mathbf{e}}\ \underset{1\times 1}{\alpha_i} + \underset{T\times 1}{\mathbf{u}_i}\end{aligned}\ , i = 1, \ldots, N, \tag{3.6.1}$$

where

$$Z_i = \underset{T\times 1}{\mathbf{e}} \quad \underset{1\times p.}{\mathbf{z}_i'}$$

If we assume that the α_i are fixed constants, model (3.6.1) is subject to perfect multicollinearity because $Z = (Z_1', \ldots, Z_N')'$ and $(I_N \otimes \mathbf{e})$ are perfectly correlated.[15] Hence, $\boldsymbol{\gamma}$, μ, and α_i are not separately estimable. However, $\boldsymbol{\beta}$ may still be estimated by the covariance method (provided $\Sigma_{i=1}^N X_i' Q X_i$ is of full rank). Premultiplying (3.6.1) by the (covariance) transformation matrix Q [(3.2.6)], we sweep out Z_i, $\mathbf{e}\mu$, and $\mathbf{e}\alpha_i$ from (3.6.1), so that

$$Q\mathbf{y}_i = QX_i\boldsymbol{\beta} + Q\mathbf{u}_i, \quad i = 1, \ldots, N. \tag{3.6.2}$$

Applying OLS to (3.6.2), we obtain the CV estimate of $\boldsymbol{\beta}$, (3.2.8).

There is no way one can separately identify $\boldsymbol{\gamma}$ and α_i^* under a fixed-effects formulation. However, if $\mathbf{z}_i$ and α_i^* are uncorrelated across i, one can treat $\alpha_i = \alpha_i^* - \mu$ as a random variable, where $\mu = \lim \frac{1}{N}\sum_{i=1}^N \alpha_i^*$. When the α_i are assumed random and uncorrelated with X_i and Z_i, CV uses the same method to estimate $\boldsymbol{\beta}$ (3.2.8). To estimate $\boldsymbol{\gamma}$, we note that the individual mean over time can be written in the form

$$\bar{y}_i - \bar{\mathbf{x}}_i'\boldsymbol{\beta} = \mu + \mathbf{z}_i'\boldsymbol{\gamma} + \alpha_i + \bar{u}_i, \quad i = 1, \ldots, N. \tag{3.6.3}$$

Treating $(\alpha_i + \bar{u}_i)$ as the error term and minimizing $\Sigma_{i=1}^N(\alpha_i + \bar{u}_i)^2$, we obtain

$$\hat{\boldsymbol{\gamma}} = \left[\sum_{i=1}^N (\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})'\right]^{-1} \left\{\sum_{i=1}^N (\mathbf{z}_i - \bar{\mathbf{z}})[(\bar{y}_i - \bar{y}) - (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})'\boldsymbol{\beta}]\right\}, \tag{3.6.4}$$

$$\hat{\mu} = \bar{y} - \bar{\mathbf{x}}'\boldsymbol{\beta} - \bar{\mathbf{z}}'\hat{\boldsymbol{\gamma}}, \tag{3.6.5}$$

where

$$\bar{\mathbf{z}} = \frac{1}{N}\sum_{i=1}^N \mathbf{z}_i, \quad \bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^N \bar{\mathbf{x}}_i, \quad \bar{y} = \frac{1}{N}\sum_{i=1}^N \bar{y}_i.$$

[15] We use $\otimes$ to denote the Kronecker product of two matrices (Theil 1971, their Chapter 7). Suppose that $A = (a_{ij})$ is an $m \times n$ matrix and B is a $p \times q$ matrix; $A \otimes B$ is defined as an $mp \times nq$ matrix of

$$\begin{bmatrix} a_{11}B & a_{12}B & \ldots & a_{1n}B \\ \vdots & & & \vdots \\ a_{m1}B & a_{m2}B & \ldots & a_{mn}B \end{bmatrix}.$$

Substituting the CV estimate of $\boldsymbol{\beta}$ into (3.6.4) and (3.6.5), we obtain estimators of $\boldsymbol{\gamma}$ and μ. When N tends to infinity, this two-step procedure is consistent. When N is fixed and T tends to infinity, $\boldsymbol{\beta}$ can still be consistently estimated by (3.2.8). But $\boldsymbol{\gamma}$ can no longer be consistently estimated, because when N is fixed, we have a limited amount of information on α_i and $\mathbf{z}_i$. To see this, note that the OLS estimate of (3.6.3) after substituting $\operatorname{plim}_{T\to\infty}\hat{\boldsymbol{\beta}}_{cv} = \boldsymbol{\beta}$ converges to

$$\begin{aligned}\hat{\boldsymbol{\gamma}}_{\text{OLS}} = \boldsymbol{\gamma} &+ \left[\sum_{i=1}^{N}(\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})'\right]^{-1}\left[\sum_{i=1}^{N}(\mathbf{z}_i - \bar{\mathbf{z}})(\alpha_i - \bar{\alpha})\right] \\ &+ \left[T\sum_{i=1}^{N}(\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})'\right]^{-1}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{z}_i - \bar{\mathbf{z}})(u_{it} - \bar{u})\right],\end{aligned} \tag{3.6.6}$$

where

$$\bar{u} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}, \qquad \bar{\alpha} = \frac{1}{N}\sum_{i=1}^{N}\alpha_i.$$

It is clear that

$$\underset{T\to\infty}{\text{plim}} \quad \tfrac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i - \bar{\mathbf{z}})\tfrac{1}{T}\sum_{t=1}^{T}(u_{it} - \bar{u}) = 0,$$

but $(1/N)\Sigma_{i=1}^{N}(\mathbf{z}_i - \bar{\mathbf{z}})(\alpha_i - \bar{\alpha})$ is a random variable, with mean 0 and covariance $\sigma_\alpha^2[\Sigma_{i=1}^{N}(z_i - \bar{z})(z_i - \bar{z})'/N^2] \neq 0$ for finite N, so that the second term in (3.6.6) does not have zero plim.

When α_i are random and uncorrelated with X_i and Z_i, the CV is not the BLUE. The BLUE of (3.6.1) is the GLS estimator

$$\begin{bmatrix}\hat{\mu}\\ \hat{\boldsymbol{\gamma}}\\ \hat{\boldsymbol{\beta}}\end{bmatrix} = \begin{bmatrix} NT\psi & NT\psi\bar{\mathbf{z}}' & NT\psi\bar{\mathbf{x}}' \\ NT\psi\bar{\mathbf{z}} & T\psi\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i' & T\psi\sum_{i=1}^{N}\mathbf{z}_i\bar{\mathbf{x}}_i' \\ NT\psi\bar{\mathbf{x}} & T\psi\sum_{i=1}^{N}\bar{\mathbf{x}}_i\mathbf{z}_i' & \sum_{i=1}^{N}X_i'QX_i + \psi T\sum_{i=1}^{N}\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i'\end{bmatrix}^{-1} \cdot \begin{bmatrix} NT\psi\bar{y} \\ \psi T\sum_{i=1}^{N}\mathbf{z}_i\bar{y}_i \\ \sum_{i=1}^{N}X_i'Q\mathbf{y}_i + \psi T\sum_{i=1}^{N}\bar{\mathbf{x}}_i\bar{\mathbf{y}}_i\end{bmatrix} \tag{3.6.7}$$

If ψ in (3.6.7) is unknown, we can substitute a consistent estimate for it. When T is fixed, the GLS is more efficient than the CV. When N is fixed and T tends to infinity, the GLS estimator of $\boldsymbol{\beta}$ converges to the CV estimator because V^{-1} (3.3.7) converges to $\frac{1}{\sigma_u^2}Q$; for details, see Lee (1978b).

One way to view (3.6.1) is that by explicitly incorporating time-invariant explanatory variables, $\mathbf{z}_i$, we can eliminate or reduce the correlation between α_i and $\mathbf{x}_{it}$. However, if α_i remains correlated with $\mathbf{x}_{it}$ or $\mathbf{z}_i$, the GLS will be a biased estimator. The CV will produce an unbiased estimate of $\boldsymbol{\beta}$, but the OLS estimates of $\boldsymbol{\gamma}$ and μ in (3.6.3) are inconsistent even when N tends to infinity if α_i is correlated with $\mathbf{z}_i$.[16] Thus, Hausman and Taylor (1981) suggested estimating $\boldsymbol{\gamma}$ in (3.6.3) by two-stage least squares, using those elements of $\bar{\mathbf{x}}_i$ that are uncorrelated with α_i as instruments for $\mathbf{z}_i$. A necessary condition to implement this method is that the number of elements of $\bar{\mathbf{x}}_i$ that are uncorrelated with α_i must be greater than the number of elements of $\mathbf{z}_i$ that are correlated with α_i.

3.6.2 Estimation of Models with Both Individual and Time Effects

We can further generalize model (3.6.1) to include time-specific variables and effects. Let

$$y_{it} = \mu + \underset{1\times p}{\mathbf{z}_i'}\ \underset{p\times 1}{\boldsymbol{\gamma}} + \underset{1\times l}{\mathbf{r}_t'}\ \underset{l\times 1}{\boldsymbol{\rho}} + \underset{1\times K}{\mathbf{x}_{it}'}\ \underset{K\times 1}{\boldsymbol{\beta}} + \alpha_i + \lambda_t + u_{it}, \qquad \begin{array}{l} i = 1, \ldots, N, \\ t = 1. \ldots, T, \end{array} \tag{3.6.8}$$

where $\mathbf{r}_t$ and λ_t denote $l \times 1$ and 1×1 time-specific variables and effects. Stacking (3.6.8) over i and t, we have

$$\underset{NT\times 1}{Y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{bmatrix} = \begin{bmatrix} \mathbf{e} & Z_1 & R & X_1 \\ \mathbf{e} & Z_2 & R & X_2 \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{e} & Z_N & R & X_N \end{bmatrix} \begin{bmatrix} \mu \\ \boldsymbol{\gamma} \\ \boldsymbol{\rho} \\ \boldsymbol{\beta} \end{bmatrix} + (I_N \otimes \mathbf{e})\boldsymbol{\alpha} + (\mathbf{e}_N \otimes I_T)\boldsymbol{\lambda} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_N \end{bmatrix}, \tag{3.6.9}$$

where $\boldsymbol{\alpha}' = (\alpha_1, \ldots, \alpha_N)$, $\boldsymbol{\lambda}' = (\lambda_1. \ldots, \lambda_T)$, $R' = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_T)$, $\mathbf{e}_N$ is an $N \times 1$ vector of ones, and $\otimes$ denotes the Kronecker product.

When both α_i and λ_t are present, estimators ignoring the presence of λ_t could be inconsistent no matter how large N is if T is finite. Take the simple case where $\mathbf{z}_i \equiv 0$ and $\mathbf{r}_t \equiv 0$, then the CV estimator of $\boldsymbol{\beta}$ ignoring the presence

[16] This is because Q sweeps out α_i from (3.6.1).

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

of λ_t (3.2.8) leads to

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{cv} &= \left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it}-\bar{\mathbf{x}}_i)(\mathbf{x}_{it}-\bar{\mathbf{x}}_i)'\right]^{-1}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it}-\bar{\mathbf{x}}_i)(y_{it}-\bar{y}_i)\right]\\ &= \boldsymbol{\beta}+\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it}-\bar{\mathbf{x}}_i)(\mathbf{x}_{it}-\bar{\mathbf{x}}_i)'\right]^{-1}\\ &\quad\cdot\left\{\frac{1}{NT}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it}-\bar{\mathbf{x}}_i)(\lambda_t-\bar{\lambda})+\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it}-\bar{\mathbf{x}}_i)(u_{it}-\bar{u}_i)\right]\right\},\end{aligned} \tag{3.6.10}$$

where $\bar{\lambda}=\frac{1}{T}\sum_{t=1}^{T}\lambda_t$. Under the assumption that x_{it} and u_{it} are uncorrelated, the last term after the second equality converges to 0. But

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it}-\bar{\mathbf{x}}_i)(\lambda_t-\bar{\lambda})=\frac{1}{T}\sum_{t=1}^{T}(\bar{\mathbf{x}}_t-\bar{\mathbf{x}})(\lambda_t-\bar{\lambda}), \tag{3.6.11}$$

where $\bar{\mathbf{x}}_t=\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_{it}$, $\bar{\mathbf{x}}=\frac{1}{N}\sum_{i=1}^{N}\bar{\mathbf{x}}_i=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}$, will converge to 0 only if λ_t are uncorrelated with $\bar{\mathbf{x}}_t$ and $T\longrightarrow\infty$. If λ_t is correlated with $\bar{\mathbf{x}}_t$ or even $E(\lambda_t\mathbf{x}_{it}')=0$, if T is finite, (3.6.11) will not converge to 0 no matter how large N is. To obtain a consistent estimator of $\boldsymbol{\beta}$, both α_i and λ_t need to be considered.

If $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$ are treated as fixed constants, there is a multi-collinearity problem, for the same reasons stated for model (3.6.1). The coefficients $\boldsymbol{\alpha}$, $\boldsymbol{\lambda}$, $\boldsymbol{\gamma}$, $\boldsymbol{\rho}$, and μ cannot be separately estimated. The coefficient $\boldsymbol{\beta}$ can still be estimated by the covariance method. Using the $NT\times NT$ (covariance) transformation matrix

$$\tilde{Q}=I_{NT}-I_N\otimes\frac{1}{T}\mathbf{e}\mathbf{e}'-\frac{1}{N}\mathbf{e}_N\mathbf{e}_N'\otimes I_T+\frac{1}{NT}J, \tag{3.6.12}$$

where J is an $NT\times NT$ matrix of ones, we can sweep out μ, $\mathbf{z}_i$, $\mathbf{r}_t$, α_i, and λ_t and estimate $\boldsymbol{\beta}$ by

$$\tilde{\boldsymbol{\beta}}_{cv}=[(X_1',\ldots,X_N')\tilde{Q}(X_1',\ldots,X_N')']^{-1}[(X_1',\ldots,X_N')\tilde{Q}Y]. \tag{3.6.13}$$

In other words, $\tilde{\boldsymbol{\beta}}$ is obtained by applying the least-squares regression to the model

$$\begin{aligned}(y_{it}-\bar{y}_i-\bar{y}_t+\bar{y}) &= (\mathbf{x}_{it}-\bar{\mathbf{x}}_i-\bar{\mathbf{x}}_t+\bar{\mathbf{x}})'\boldsymbol{\beta}\\ &\quad+(u_{it}-\bar{u}_i-\bar{u}_t+\bar{u}),\end{aligned} \tag{3.6.14}$$

where $\bar{y}_t=\frac{1}{N}\sum_{i=1}^{N}y_{it}$, $\bar{y}=\frac{1}{N}\sum_{i=1}^{N}\bar{y}_i=\frac{1}{T}\sum_{t=1}^{T}\bar{y}_t=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}y_{it}$, and $\bar{u}_i,\bar{u}_t,\bar{u}$ are similarly defined.

When u_{it} is independently, identically distributed with constant variance, the variance–covariance matrix of the CV estimator (3.6.13) is equal to

$$\text{Cov}\,(\hat{\boldsymbol{\beta}}_{cv})=\sigma_u^2\left[(X_1',\ldots,X_N')\tilde{Q}(X_1',\ldots,X_N')'\right]^{-1}. \tag{3.6.15}$$

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

To estimate μ, $\boldsymbol{\gamma}$, and $\boldsymbol{\rho}$, we note that the individual-mean (over time) and time-mean (over individuals) equations are of the form

$$\bar{y}_i - \bar{\mathbf{x}}_i'\boldsymbol{\beta} = \mu_c^* + \mathbf{z}_i'\boldsymbol{\gamma} + \alpha_i + \bar{u}_i, \quad i = 1, \ldots, N, \tag{3.6.16}$$

$$\bar{y}_t - \bar{\mathbf{x}}_t'\boldsymbol{\beta} = \mu_T^* + \mathbf{r}_t'\boldsymbol{\rho} + \lambda_t + \bar{u}_t, \quad t = 1.\ldots, T, \tag{3.6.17}$$

where

$$\mu_c^* = \mu + \bar{\mathbf{r}}'\boldsymbol{\rho} + \bar{\lambda}, \tag{3.6.18}$$

$$\mu_T^* = \mu + \bar{\mathbf{z}}'\boldsymbol{\gamma} + \bar{\alpha}, \tag{3.6.19}$$

and

$$\bar{\mathbf{r}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{r}_t, \quad \bar{\mathbf{z}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i, \quad \bar{\lambda} = \frac{1}{T}\sum_{t=1}^{T}\lambda_t, \quad \bar{\alpha} = \frac{1}{N}\sum_{i=1}^{N}\alpha_i,$$

$$\bar{y}_t = \frac{1}{N}\sum_{i=1}^{N}y_{it}, \quad \bar{\mathbf{x}}_t = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_{it}, \quad \bar{u}_t = \frac{1}{N}\sum_{i=1}^{N}u_{it}.$$

Replacing $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}_{cv}$, we can estimate $(\mu_c^*, \boldsymbol{\gamma}')$ and $(\mu_T^*, \boldsymbol{\rho}')$ by applying OLS to (3.6.16) and (3.6.17) over i and t, respectively, if α_i and λ_t are uncorrelated with $\mathbf{z}_i$, $\mathbf{r}_t$, and $\mathbf{x}_{it}$. To estimate μ, we can substitute estimated values of $\boldsymbol{\gamma}$, $\boldsymbol{\rho}$, and $\boldsymbol{\beta}$ into any of

$$\hat{\mu} = \hat{\mu}_c^* - \bar{\mathbf{r}}'\hat{\boldsymbol{\rho}}, \tag{3.6.20}$$

$$\hat{\mu} = \hat{\mu}_T^* - \bar{\mathbf{z}}'\hat{\boldsymbol{\gamma}}, \tag{3.6.21}$$

$$\hat{\mu} = \bar{y} - \bar{\mathbf{z}}'\hat{\boldsymbol{\gamma}} - \bar{\mathbf{r}}'\hat{\boldsymbol{\rho}} - \bar{\mathbf{x}}'\hat{\boldsymbol{\beta}}, \tag{3.6.22}$$

or apply the least-squares method to the combined equations (3.6.20)–(3.6.22). When both N and T go to infinity, $\hat{\mu}$ is consistent.

If α_i and λ_t are random, we can still estimate $\boldsymbol{\beta}$ by the CV estimator (3.6.13). However, if α_i and λ_t are uncorrelated with z_i, $\mathbf{r}_t$, and $\mathbf{x}_{it}$, the BLUE is the GLS estimator. Assuming α_i and λ_t satisfy (3.3.4), the $NT \times NT$ variance–covariance matrix of the error term, $\mathbf{u} + (I_N \otimes \mathbf{e})\boldsymbol{\alpha} + (\mathbf{e}_N \otimes I_T)\boldsymbol{\lambda}$, is

$$\tilde{V} = \sigma_u^2 I_{NT} + \sigma_\alpha^2 I_N \otimes \mathbf{e}\mathbf{e}' + \sigma_\lambda^2 \mathbf{e}_N\mathbf{e}_N' \otimes I_T. \tag{3.6.23}$$

Its inverse (Henderson 1971; Nerlove 1971b; Wallace and Hussain 1969) (see Appendix 3B) is

$$\tilde{V}^{-1} = \frac{1}{\sigma_u^2}[I_{NT} - \eta_1 I_N \otimes \mathbf{e}\mathbf{e}' - \eta_2 \mathbf{e}_N\mathbf{e}_N' \otimes I_T + \eta_3 J], \tag{3.6.24}$$

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

where

$$\eta_1 = \frac{\sigma_\alpha^2}{\sigma_u^2 + T\sigma_\alpha^2}, \quad \eta_2 = \frac{\sigma_\lambda^2}{\sigma_u^2 + N\sigma_\lambda^2},$$

$$\eta_3 = \frac{\sigma_\alpha^2 \sigma_\lambda^2}{(\sigma_u^2 + T\sigma_\alpha^2)(\sigma_u^2 + N\sigma_\lambda^2)} \left(\frac{2\sigma_u^2 + T\sigma_\alpha^2 + N\sigma_\lambda^2}{\sigma_u^2 + T\sigma_\alpha^2 + N\sigma_\lambda^2} \right).$$

When $N \to \infty$, $T \to \infty$, and the ratio N over T tends to a nonzero constant, Wallace and Hussain (1969) have shown that the GLS estimator converges to the CV estimator. It should also be noted that, contrary to the conventional linear regression model without specific effects, the speed of convergence of $\boldsymbol{\beta}_{\text{GLS}}$ to $\boldsymbol{\beta}$ is $(NT)^{1/2}$, whereas the speed of convergence for $\hat{\mu}$ is $N^{1/2}$. This is because the effect of a random component can be averaged out only in the direction of that random component. For details, see Kelejian and Stephan (1983).

For the discussion of the MLE of the two-way error components models, see Baltagi (1995) and Baltagi and Li (1992).

3.7 HETEROSCEDASTICITY AND AUTOCORRELATION

3.7.1 Heteroscedasticity

So far we have confined our discussion to the assumption that the variances of the errors across individuals are identical. However, many panel studies involve cross-sectional units of varying size. In an error-components setup, heteroscedasticity can arise because the variance of α_i, $\sigma_{\alpha i}^2$, varies with i (e.g., Baltagi and Griffin 1983; Mazodier and Trognon 1978) or the variance of u_{it}, σ_{ui}^2, varies with i, or both $\sigma_{\alpha i}^2$ and σ_{ui}^2 vary with i. Then

$$E\mathbf{v}_i\mathbf{v}_i' = \sigma_{ui}^2 I_T + \sigma_{\alpha i}^2 \mathbf{e}\mathbf{e}' = V_i. \tag{3.7.1}$$

The V_i^{-1} is of the same form as equation (3.3.5) with σ_{ui}^2 and $\sigma_{\alpha i}^2$ in place of σ_u^2 and σ_α^2. The GLS estimator of $\boldsymbol{\delta}$ is obtained by replacing V by V_i in (3.3.7).

When σ_{ui}^2 and $\sigma_{\alpha i}^2$ are unknown, substituting the unknown true values by their estimates, a feasible (or two-step) GLS estimator can be implemented. Unfortunately, with a single realization of α_i, there is no way one can get a consistent estimator for $\sigma_{\alpha i}^2$ even when $T \to \infty$. The conventional formula

$$\hat{\sigma}_{\alpha i}^2 = \hat{\bar{v}}_i^2 - \frac{1}{T}\hat{\sigma}_{ui}^2, \quad i = 1, \ldots, N, \tag{3.7.2}$$

where $\hat{v}_{it}$ is the initial estimate of v_{it}, say, the least-squares or CV estimated residual of (3.3.3), converges to α_i^2, not $\sigma_{\alpha i}^2$. However, σ_{ui}^2 can be consistently estimated by

$$\hat{\sigma}_{ui}^2 = \frac{1}{T-1}\sum_{t=1}^{T}(\hat{v}_{it} - \hat{\bar{v}}_i)^2, \tag{3.7.3}$$

as T tends to infinity. In the event that $\sigma_{\alpha i}^2 = \sigma_\alpha^2$ for all i, we can estimate σ_α^2 by taking the average of (3.7.2) across i as their estimates.

It should be noted that when T is finite, there is no way we can get consistent estimates of σ_{ui}^2 and $\sigma_{\alpha i}^2$ even when N tends to infinity. This is the classical incidental parameter problem of Neyman and Scott (1948). However, if $\sigma_{\alpha i}^2 = \sigma_\alpha^2$ for all i, then we can get consistent estimates of σ_{ui}^2 and σ_α^2 when both N and T tend to infinity. Substituting $\hat{\sigma}_{ui}^2$ and $\hat{\sigma}_\alpha^2$ for σ_{ui}^2 and σ_α^2 in V_i, we obtain its estimation $\hat{V}_i$. Alternatively, one may assume that the conditional variance of α_i conditional on $\mathbf{x}_i$ has the same functional form across individuals, var $(\alpha_i \mid \mathbf{x}_i) = \sigma^2(\mathbf{x}_i)$, to allow for the consistent estimation of heteroscadastic variance, $\sigma_{\alpha i}^2$. The feasible GLS estimator of $\boldsymbol{\delta}$,

$$\hat{\boldsymbol{\delta}}_{\text{FGLS}} = \left[\sum_{i=1}^{N} \tilde{X}_i' \hat{V}_i^{-1} \tilde{X}_i\right]^{-1} \left[\sum_{i=1}^{N} \tilde{X}_i' \hat{V}_i^{-1} \mathbf{y}_i\right] \tag{3.7.4}$$

is asymptotically equivalent to the GLS estimator when both N and T approach infinity. The asymptotic variance–covariance matrix of the $\hat{\boldsymbol{\delta}}_{\text{FGLS}}$ can be approximated by $(\sum_{i=1}^{N} \tilde{X}_i' \hat{V}_i^{-1} \tilde{X}_i)^{-1}$.

In the case that both $\sigma_{\alpha i}^2$ and σ_{ui}^2 vary across i, another way to estimate the model is to treat α_i as fixed by taking the covariance transformation to eliminate the effect of α_i, then apply the feasible weighted least-squares method. That is, we first weigh each individual observation by the inverse of σ_{ui}, $\mathbf{y}_i^* = \frac{1}{\sigma_{ui}}\mathbf{y}_i$, $X_i^* = \frac{1}{\sigma_{ui}}X_i$ and then apply the CV estimator to the transformed data

$$\hat{\boldsymbol{\beta}}_{cv} = \left[\sum_{i=1}^{N} X_i^{*'} Q X_i^*\right]^{-1} \left[\sum_{i=1}^{N} X_i^{*'} Q \mathbf{y}_i^*\right]. \tag{3.7.5}$$

3.7.2 Models with Serially Correlated Errors

The fundamental assumption we made with regard to the variable-intercept model was that the error term is serially uncorrelated conditional on the individual effects α_i. But there are cases in which the effects of unobserved variables vary systematically over time, such as the effect of serially correlated omitted variables or the effects of transitory variables whose effects last more than one period. The existence of these variables is not well described by an error term that is either constant or independently distributed over time periods. To provide for a more general autocorrelation scheme, one can relax the restriction that u_{it} are serially uncorrelated (e.g., Lillard and Weiss 1979; Lillard and Willis

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

1978).[17] Anderson and Hsiao (1982) have considered the MLE of the model (3.3.5) with u_{it} following a first-order autoregressive process,

$$u_{it} = \rho u_{i,t-1} + \epsilon_{it}, \tag{3.7.6}$$

where ϵ_{it} are independently, identically distributed, with 0 mean and variance σ_ϵ^2. However, computation of the MLE is complicated. But if we know ρ, we can transform the model into a standard variance–components model,

$$y_{it} - \rho y_{i,t-1} = \mu(1-\rho) + \boldsymbol{\beta}'(\mathbf{x}_{it} - \rho \mathbf{x}_{i,t-1}) + (1-\rho)\alpha_i + \epsilon_{it}. \tag{3.7.7}$$

Therefore, we can obtain an asymptotically efficient estimator of $\boldsymbol{\beta}$ by the following multistep procedure:

Step 1. Eliminate the individual effect α_i by subtracting the individual mean from (3.3.5). We have

$$y_{it} - \bar{y}_i = \boldsymbol{\beta}'(\mathbf{x}_{it} - \bar{\mathbf{x}}_i) + (u_{it} - \bar{u}_i). \tag{3.7.8}$$

Step 2. Use the least-squares residual of (3.7.8) to estimate the serial correlation coefficient ρ, or use the Durbin (1960) method by regressing $(y_{it} - \bar{y}_i)$ on $(y_{i,t-1} - \bar{y}_{i,-1})$, and $(\mathbf{x}_{i,t-1} - \bar{\mathbf{x}}_{i,-1})$, and treat the coefficient of $(y_{i,t-1} - \bar{y}_{i,-1})$ as the estimated value of ρ, where $\bar{y}_{i,-1} = (1/T)\Sigma_{t=1}^T y_{i,t-1}$ and $\bar{\mathbf{x}}_{i,-1} = (1/T)\Sigma_{t=1}^T \mathbf{x}_{i,t-1}$. (For simplicity, we assume that y_{i0} and x_{i0} are observable.)

Step 3. Estimate σ_ϵ^2 and σ_α^2 by

$$\begin{aligned}\hat{\sigma}_\epsilon^2 = & \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \{(y_{it} - \bar{y}_i) \\ & - \hat{\rho}(y_{i,t-1} - \bar{y}_{i,-1}) \\ & - \hat{\boldsymbol{\beta}}'[(\mathbf{x}_{it} - \bar{\mathbf{x}}_i) - (\mathbf{x}_{i,t-1} - \bar{\mathbf{x}}_{i,-1})\hat{\rho}]\}^2\end{aligned} \tag{3.7.9}$$

and

$$\begin{aligned}\hat{\sigma}_\alpha^2 = & \frac{1}{(1-\hat{\rho})^2}\left\{\frac{1}{N}\sum_{i=1}^N [\bar{y}_i - \hat{\mu}(1-\hat{\rho})\right. \\ & \left. - \hat{\rho}\bar{y}_{i,-1} - \hat{\boldsymbol{\beta}}'(\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_{i,-1}\hat{\rho})]^2 - \frac{1}{T}\hat{\sigma}_\epsilon^2\right\}.\end{aligned} \tag{3.7.10}$$

Step 4. Substituting $\hat{\rho}$, (3.7.9), and (3.7.10) for ρ, σ_ϵ^2, and σ_α^2 in the variance–covariance matrix of $\epsilon_{it} + (1-\rho)\alpha_i$, we estimate (3.7.7) by the feasible GLS method.

The above multistep or feasible generalized least-squares procedure treats the initial u_{i1} as fixed constants. A more efficient, but computationally more

[17] See Li and Hsiao (1998) for a test of whether the serial correlation in the error is caused by an individual-specific time invariant component or by the inertia in the shock and Hong and Kao (2004) for testing of serial correlation of unknown form.

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

burdensome feasible GLS is to treat initial u_{i1} as random variables with mean 0 and variance $\frac{\sigma_\epsilon^2}{1-\rho^2}$ (e.g., Baltagi and Li 1991). Premultiplying (3.3.5) by the $T \times T$ transformation matrix

$$R = \begin{pmatrix} (1-\rho^2)^{1/2} & 0 & 0 & \cdot & \cdot & 0 \\ -\rho & 1 & 0 & \cdot & \cdot & \cdot \\ 0 & -\rho & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & -\rho & 1 \end{pmatrix},$$

transforms $\mathbf{u}_i$ into serially uncorrelated homoscedastic error terms, but also transforms $\mathbf{e}_T\alpha_i$ into $(1-\rho)\boldsymbol{\ell}_T\alpha_i$, where $\boldsymbol{\ell}_T = [(\frac{1+\rho}{1-\rho})^{1/2}, 1, \ldots, 1]'$. Therefore, the transformed error terms will have covariance matrix

$$V^* = \sigma_\epsilon^2 I_T + (1-\rho)^2\sigma_\alpha^2 \boldsymbol{\ell}_T\boldsymbol{\ell}_T', \tag{3.7.11}$$

with inverse

$$V^{*-1} = \frac{1}{\sigma_\epsilon^2}[I_T - \frac{(1-\rho)^2\sigma_\alpha^2}{[T-(T-1)\rho-\rho^2]\sigma_\alpha^2 + \sigma_\epsilon^2}\boldsymbol{\ell}_T\boldsymbol{\ell}_T']. \tag{3.7.12}$$

Substituting initial estimates of ρ, σ_α^2, and σ_ϵ^2 into (3.7.12), one can apply the GLS procedure using (3.7.12) to estimate $\boldsymbol{\delta}$.

When T tends to infinity, the GLS estimator of $\boldsymbol{\beta}$ converges to the covariance estimator of the transformed model (3.7.7). In other words, an asymptotically efficient estimator of $\boldsymbol{\beta}$ is obtained by finding a consistent estimate of ρ, transforming the model to eliminate the serial correlation, and then applying the covariance method to the transformed model (3.7.7).

MaCurdy (1982) has considered a similar estimation procedure for (3.3.5) with a more general time series process of u_{it}. His procedure essentially involves eliminating α_i by first differencing and treating $y_{it} - y_{i,t-1}$ as the dependent variable. He then modeled the variance–covariance matrix of $\mathbf{u}_i$ by using a standard Box–Jenkins (1970) type of procedure to model the least-squares predictor of $u_{it} - u_{i,t-1}$, and estimated the parameters by an efficient algorithm.

Kiefer (1980) considered estimation of fixed-effects models of (3.2.1) with arbitrary intertemporal correlations for u_{it}. When T is fixed, the individual effects cannot be estimated consistently. He suggested that we first eliminate the individual effects by transforming the model to the form (3.7.8) using the transformation matrix $Q = I_T - (1/T)\mathbf{ee}'$. Then estimate the intertemporal variance–covariance matrix of $Q\mathbf{u}_i$ by

$$\hat{\Sigma}^* = \frac{1}{N}\sum_{i=1}^{N}[Q(\mathbf{y}_i - X_i\hat{\boldsymbol{\beta}})][Q(\mathbf{y}_i - X_i\hat{\boldsymbol{\beta}})]', \tag{3.7.13}$$

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

where $\hat{\boldsymbol{\beta}}$ is any arbitrary consistent estimator of $\boldsymbol{\beta}$ (e.g., CV of $\boldsymbol{\beta}$). Given an estimate of $\hat{\Sigma}^*$ one can estimate $\boldsymbol{\beta}$ by the GLS method,

$$\boldsymbol{\beta}^* = \left[\sum_{i=1}^{N} X_i' Q \hat{\Sigma}^{*-} Q X_i\right]^{-1} \left[\sum_{i=1}^{N} X_i' Q \hat{\Sigma}^{*-} Q \mathbf{y}_i\right], \qquad (3.7.14)$$

where $\hat{\Sigma}^{*-}$ is a generalized inverse of Σ^*, because Σ^* has only rank $T-1$. The asymptotic variance–covariance matrix of $\hat{\boldsymbol{\beta}}^*$ is

$$\text{Var}\,(\hat{\boldsymbol{\beta}}^*) = \left[\sum_{i=1}^{N} X_i' Q \hat{\Sigma}^{*-} Q X_i\right]^{-1}. \qquad (3.7.15)$$

Although any generalized inverse can be used for $\hat{\Sigma}^*$, a particularly attractive choice is

$$\hat{\Sigma}^{*-} = \begin{bmatrix} \hat{\Sigma}^{*-1}_{T-1} & \mathbf{0} \\ \mathbf{0}' & 0 \end{bmatrix}, \qquad (3.7.16)$$

where $\hat{\Sigma}^*_{T-1}$ is the $(T-1) \times (T-1)$ full-rank submatrix of $\hat{\Sigma}^*$ obtained by deleting the last row and column from $\hat{\Sigma}^*$. Using this generalized inverse simply amounts to deleting the Tth observation from the transformed observations $Q\mathbf{y}_i$ and QX_i, and then applying GLS to the remaining subsample. However, it should be noted that this is not the GLS estimator that would be used if the variance–covariance matrix of $\mathbf{u}_i$ were known.

3.7.3 Heteroscedasticity Autocorrelation Consistent Estimator for the Covariance Matrix of the CV Estimator

The previous two subsections discuss the estimation procedures when the patterns of heteroscedasticity or serial correlations are known. In the case that the errors u_{it} have unknown heteroscedasticity (across individuals and over time) and/or autocorrelation patterns, one may still use the covariance estimator (3.2.5) or (3.6.13) to obtain a consistent estimate of $\boldsymbol{\beta}$. However, the covariance matrix of the CV estimator of $\boldsymbol{\beta}$ no longer has the form (3.2.11) or $\sigma_u^2(X'\tilde{Q}X)^{-1}$, where $X' = (X_1', \ldots, X_N')$. For instance, when u_{it} has heteroscedasticity of unknown form, $\sqrt{NT}(\hat{\boldsymbol{\beta}}_{cv} - \boldsymbol{\beta})$ is asymptotically normally distributed with mean 0 and covariance matrix of the form (e.g., Arellano 2003)

$$\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}\tilde{\mathbf{x}}_{it}'\right)^{-1} \Omega \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}\tilde{\mathbf{x}}_{it}'\right)^{-1}, \qquad (3.7.17)$$

where

$$\tilde{\mathbf{x}}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_i, \qquad (3.7.18)$$

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

for model (3.2.1) and

$$\tilde{\mathbf{x}}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_i - \bar{\mathbf{x}}_t + \bar{\mathbf{x}} \tag{3.7.19}$$

for model (3.6.8), and

$$\Omega = \frac{1}{T}\sum_{t=1}^{T} E\left(\tilde{\mathbf{x}}_{it}\tilde{\mathbf{x}}_{it}' u_{it}^2\right). \tag{3.7.20}$$

It is shown by Stock and Watson (2008) that

$$\begin{aligned}\hat{\Omega} = \left(\frac{T-1}{T-2}\right)\Bigg\{ &\frac{1}{NT-N-K}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}\tilde{\mathbf{x}}_{it}'\hat{\hat{u}}_{it}^2 \\ &-\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}\tilde{\mathbf{x}}_{it}'\right)\left(\frac{1}{T-1}\sum_{s=1}^{T}\hat{\hat{u}}_{is}^2\right)\Bigg\},\end{aligned} \tag{3.7.21}$$

is a consistent estimator of Ω for any sequence of N or $T \longrightarrow \infty$. Where $\hat{\hat{u}}_{it} = \tilde{y}_{it} - \tilde{\mathbf{x}}_{it}'\hat{\boldsymbol{\beta}}_{cv}$, and $\tilde{y}_{it} - \bar{y}_i$ or $\tilde{y}_{it} = y_{it} - \bar{y}_i - \bar{y}_t + \bar{y}$.

When both N and T are large, Vogelsang (2012) suggests a robust estimator of the variance–covariance matrix of the CV estimator of $\boldsymbol{\beta}$ as

$$T\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}\tilde{\mathbf{x}}_{it}'\right)^{-1}\left(\sum_{i=1}^{N}\hat{\Omega}_i\right)\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}\tilde{\mathbf{x}}_{it}'\right)^{-1}, \tag{3.7.22}$$

where

$$\hat{\Omega}_i = \frac{1}{T}\left[\sum_{t=1}^{T}\hat{u}_{it}^2\tilde{\mathbf{x}}_{it}\tilde{\mathbf{x}}_{it}' + \sum_{t=2}^{T}\sum_{j=1}^{t-1}k\left(\frac{j}{m}\right)\hat{u}_{it}\hat{u}_{i,t-j}\left(\tilde{\mathbf{x}}_{it}\tilde{\mathbf{x}}_{i,t-j}' + \tilde{\mathbf{x}}_{i,t-j}\tilde{\mathbf{x}}_{it}'\right)\right], \tag{3.7.23}$$

where $k(\frac{j}{m})$ denotes the kernel such that $k(\frac{j}{m}) = 1 - \frac{j}{m}$ if $\mid \frac{j}{m} \mid \leq 1$ and $k(\frac{j}{m}) = 0$ if $\mid \frac{j}{m} \mid \geq 1$. If $M = T$, then all the sample autocorrelations are used for (3.7.23). If $M < T$, a truncated kernel is used.

3.8 MODELS WITH ARBITRARY ERROR STRUCTURE – CHAMBERLAIN π-APPROACH

The focus of this chapter is formulation and estimation of linear regression models when there exist time-invariant and/or individual-invariant omitted (latent) variables. In Sections 3.1–3.7 we have been assuming that the variance–covariance matrix of the error term possesses a known structure. In fact, when N tends to infinity, the characteristics of short panels allow us to exploit the unknown structure of the error process. Chamberlain (1982, 1984) has proposed treating each period as an equation in a multivariate setup to transform the problems of estimating a single-equation model involving two dimensions (cross sections and time series) into a one-dimensional problem of estimating a T-variate regression model with cross-sectional data. This formulation avoids imposing restrictions a priori on the variance–covariance matrix, so that serial correlation and certain forms of heteroscedasticity in the error process, which covers certain kinds of random-coefficient models (see Chapter 6), can be incorporated. The multivariate setup also provides a link between the single-equation and simultaneous-equations models (see Chapter 5). Moreover, the extended view of the Chamberlain method can also be reinterpreted in terms of the generalized method of moments (GMM) to be discussed in Chapter 4 (Crépon and Mairesse 1996).

For simplicity, consider the following model:

$$\begin{aligned} y_{it} = \alpha_i^* + \boldsymbol{\beta}'\mathbf{x}_{it} + u_{it}, \quad & i = 1, \ldots, N, \\ & t = 1, \ldots, T, \end{aligned} \tag{3.8.1}$$

and

$$E(u_{it} \mid \mathbf{x}_{i1}. \ldots, \mathbf{x}_{iT}, \alpha_i^*) = 0. \tag{3.8.2}$$

When T is fixed and N tends to infinity, we can stack the T time period observations of the ith individual's characteristics into a vector $(\mathbf{y}_i', \mathbf{x}_i')$, where $\mathbf{y}_i' = (y_{i1}, \ldots, y_{iT})$ and $\mathbf{x}_i' = (\mathbf{x}_{i1}'. \ldots, \mathbf{x}_{iT}')$ are $1 \times T$ and $1 \times KT$ vectors, respectively. We assume that $(\mathbf{y}_i', \mathbf{x}_i')$ is an independent draw from a common (unknown) multivariate distribution function with finite fourth-order moments and with $E\mathbf{x}_i\mathbf{x}_i' = \Sigma_{xx}$ positive definite. Then each individual observation vector corresponds to a T-variate regression

$$\underset{T \times 1}{\mathbf{y}_i} = \mathbf{e}\alpha_i^* + (I_T \otimes \boldsymbol{\beta}')\mathbf{x}_i + \mathbf{u}_i, \quad i = 1. \ldots, N. \tag{3.8.3}$$

To allow for the possible correlation between α_i^* and $\mathbf{x}_i$, Chamberlain, following the idea of Mundlak (1978), assumes that

$$E(\alpha_i^* \mid \mathbf{x}_i) = \mu + \sum_{t=1}^{T} \mathbf{a}_t'\mathbf{x}_{it} = \mu + \mathbf{a}'\mathbf{x}_i, \tag{3.8.4}$$

where $\mathbf{a}' = (\mathbf{a}_1'. \ldots, \mathbf{a}_T')$. While $E(\mathbf{y}_i \mid \mathbf{x}_i, \alpha_i^*)$ is assumed linear, it is possible to relax the assumption of $E(\alpha_i^* \mid \mathbf{x}_i)$ being linear for the linear model. In the case in which $E(\alpha_i^* \mid \mathbf{x}_i)$ is not linear, Chamberlain (1984) replaces (3.8.4) by

$$E^*(\alpha_i^* \mid \mathbf{x}_i) = \mu + \mathbf{a}'\mathbf{x}_i, \tag{3.8.5}$$

where $E^*(\alpha_i^* \mid \mathbf{x}_i)$ refers to the (minimum mean square error) linear predictor (or the projection) of α_i^* onto $\mathbf{x}_i$. Then,[18]

$$\begin{aligned} E^*(\mathbf{y}_i \mid \mathbf{x}_i) &= E^*\{E^*(\mathbf{y}_i \mid \mathbf{x}_i, \alpha_i^*) \mid \mathbf{x}_i\} \\ &= E^*\{\mathbf{e}\alpha_i^* + (I_T \otimes \boldsymbol{\beta}')\mathbf{x}_i \mid \mathbf{x}_i\} \\ &= \mathbf{e}\mu + \Pi\mathbf{x}_i, \end{aligned} \tag{3.8.6}$$

where

$$\underset{T\times KT}{\Pi} = I_T \otimes \boldsymbol{\beta}' + \mathbf{e}\mathbf{a}'. \tag{3.8.7}$$

Rewrite equations (3.8.3) and (3.8.6) as

$$\mathbf{y}_i = \mathbf{e}\mu + [I_T \otimes \mathbf{x}_i']\boldsymbol{\pi} + \boldsymbol{\nu}_i, \quad i = 1.\dots, N, \tag{3.8.8}$$

where $\boldsymbol{\nu}_i = \mathbf{y}_i - E^*(\mathbf{y}_i \mid \mathbf{x}_i)$ and $\boldsymbol{\pi}' = \text{vec}\,(\Pi)' = [\boldsymbol{\pi}_1', \dots, \boldsymbol{\pi}_T']$ is a $1 \times KT^2$ vector with $\boldsymbol{\pi}_t'$ denoting the tth row of Π'. Treating the coefficients of (3.8.8) as if they were unconstrained, we regress $(\mathbf{y}_i - \bar{\mathbf{y}}^*)$ on $[I_T \otimes (\mathbf{x}_i - \bar{\mathbf{x}}^*)']$ and obtain the least-squares estimate of $\boldsymbol{\pi}$ as[19]

$$\begin{aligned} \hat{\boldsymbol{\pi}} &= \left\{\sum_{i=1}^N [I_T \otimes (\mathbf{x}_i - \bar{\mathbf{x}}^*)][I_T \otimes (\mathbf{x}_i - \bar{\mathbf{x}}^*)']\right\}^{-1} \\ &\quad \cdot \left\{\sum_{i=1}^N [I_T \otimes (\mathbf{x}_i - \bar{\mathbf{x}}^*)](\mathbf{y}_i - \bar{\mathbf{y}}^*)\right\} \\ &= \boldsymbol{\pi} + \left\{\frac{1}{N}\sum_{i=1}^N [I_T \otimes (\mathbf{x}_i - \bar{\mathbf{x}}^*)][I_T \otimes (\mathbf{x}_i - \bar{\mathbf{x}}^*)']\right\}^{-1} \\ &\quad \cdot \left\{\frac{1}{N}\sum_{i=1}^N [I_T \otimes (\mathbf{x}_i - \bar{\mathbf{x}}^*)]\boldsymbol{\nu}_i\right\}, \end{aligned} \tag{3.8.9}$$

where $\bar{\mathbf{y}}^* = (1/N)\Sigma_{i=1}^N \mathbf{y}_i$ and $\bar{\mathbf{x}}^* = (1/N)\Sigma_{i=1}^N \mathbf{x}_i$.

By construction, $E(\boldsymbol{\nu}_i \mid \mathbf{x}_i) = 0$, and $E(\boldsymbol{\nu}_i \otimes \mathbf{x}_i) = 0$. The law of large numbers implies that $\hat{\boldsymbol{\pi}}$ is a consistent estimator of $\boldsymbol{\pi}$ when T is fixed and N tends to infinity (Rao 1973, Chapter 2). Moreover, because

$$\begin{aligned} \underset{N\to\infty}{\text{plim}} \ \frac{1}{N}\sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}}^*)(\mathbf{x}_i - \bar{\mathbf{x}}^*)' &= E[\mathbf{x}_i - E\mathbf{x}_i][\mathbf{x}_i - E\mathbf{x}_i]' \\ &= \Sigma_{xx} - (E\mathbf{x})(E\mathbf{x})' = \Phi_{xx}, \end{aligned}$$

[18] If $E(\alpha_i^* \mid \mathbf{x}_i)$ is linear, $E^*(\mathbf{y}_i \mid \mathbf{x}_i) = E(\mathbf{y}_i \mid \mathbf{x}_i)$.

[19] Of course, we can obtain the least-squares estimate of π by imposing the restriction that all T equations have identical intercepts μ. But this only complicates the algebraic equation of the least-squares estimate without a corresponding gain in insight.

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

we have $\sqrt{N}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi})$ converging in distribution to (Rao 1973, Chapter 2)

$$\begin{aligned}&\left[I_T \otimes \Phi_{xx}^{-1}\right]\left\{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[I_T \otimes (\mathbf{x}_i - \bar{\mathbf{x}}^*)]\boldsymbol{\nu}_i\right\}\\ &\quad = \left[I_T \otimes \Phi_{xx}^{-1}\right]\left\{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\boldsymbol{\nu}_i \otimes (\mathbf{x}_i - \bar{\mathbf{x}}^*)]\right\}.\end{aligned} \tag{3.8.10}$$

So the central-limit theorem implies that $\sqrt{N}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi})$ is asymptotically normally distributed, with mean 0 and variance–covariance matrix Ω, where[20]

$$\begin{aligned}\Omega =& E[(\mathbf{y}_i - \mathbf{e}\mu - \Pi\mathbf{x}_i)(\mathbf{y}_i - \mathbf{e}\mu - \Pi\mathbf{x}_i)'\\ &\otimes \Phi_{xx}^{-1}(\mathbf{x}_i - E\mathbf{x})(\mathbf{x}_i - E\mathbf{x})'\Phi_{xx}^{-1}].\end{aligned} \tag{3.8.11}$$

A consistent estimator of Ω is readily available from the corresponding sample moments,

$$\begin{aligned}\hat{\Omega} =& \frac{1}{N}\sum_{i=1}^{N}\Big\{\left[(\mathbf{y}_i - \bar{\mathbf{y}}^*) - \hat{\Pi}(\mathbf{x}_i - \bar{\mathbf{x}}^*)\right][(\mathbf{y}_i - \bar{\mathbf{y}}^*)\\ &- \hat{\Pi}(\mathbf{x}_i - \bar{\mathbf{x}}^*)]' \otimes S_{xx}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}^*)(\mathbf{x}_i - \bar{\mathbf{x}}^*)'S_{xx}^{-1}\Big\},\end{aligned} \tag{3.8.12}$$

where

$$S_{xx} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{x}}^*)(\mathbf{x}_i - \bar{\mathbf{x}}^*)'.$$

Equation (3.8.7) implies that Π is subject to restrictions. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \mathbf{a}')$. We specify the restrictions on Π [equation (3.8.7)] by the conditions that

$$\boldsymbol{\pi} = \mathbf{f}(\boldsymbol{\theta}). \tag{3.8.13}$$

We can impose these restrictions by using a minimum-distance estimator. Namely, choose $\boldsymbol{\theta}$ to minimize

$$[\hat{\boldsymbol{\pi}} - \mathbf{f}(\boldsymbol{\theta})]'\hat{\Omega}^{-1}[\hat{\boldsymbol{\pi}} - \mathbf{f}(\boldsymbol{\theta})]. \tag{3.8.14}$$

Under the assumptions that $\mathbf{f}$ possesses continuous second partial derivatives and the matrix of first partial derivatives

$$F = \frac{\partial\,\mathbf{f}}{\partial\boldsymbol{\theta}'} \tag{3.8.15}$$

has full column rank in an open neighborhood containing the true parameter $\boldsymbol{\theta}$, the minimum-distance estimator of (3.8.14), $\hat{\boldsymbol{\theta}}$, is consistent, and $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$,

[20] For details, see White (1980) or Chamberlain (1982).

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

is asymptotically normally distributed, with mean 0 and variance–covariance matrix

$$(F'\Omega^{-1}F)^{-1}. \tag{3.8.16}$$

The quadratic form

$$N[\hat{\boldsymbol{\pi}} - \mathbf{f}(\boldsymbol{\theta})]'\hat{\Omega}^{-1}[\hat{\boldsymbol{\pi}} - \mathbf{f}(\boldsymbol{\theta})] \tag{3.8.17}$$

converges to a χ^2 distribution, with $KT^2 - K(1+T)$ degrees of freedom.[21]

The advantage of the multivariate setup is that we need only to assume that the T period observations of the characteristics of the ith individual are independently distributed across cross-sectional units with finite fourth-order moments. We do not need to make specific assumptions about the error process. Nor do we need to assume that $E(\alpha_i^* \mid \mathbf{x}_i)$ is linear.[22] In the more restrictive case that $E(\alpha_i^* \mid \mathbf{x}_i)$ is indeed linear, [then the regression function is linear, that is, $E(\mathbf{y}_i \mid \mathbf{x}_i) = \mathbf{e}\mu + \Pi\mathbf{x}_i$], and $\text{Var}(\mathbf{y}_i \mid \mathbf{x}_i)$ is uncorrelated with $\mathbf{x}_i\mathbf{x}_i'$, (3.8.12) will converge to

$$E[\text{Var}(\mathbf{y}_i \mid \mathbf{x}_i)] \otimes \Phi_{xx}^{-1}. \tag{3.8.18}$$

If the conditional variance–covariance matrix is homoscedastic, so that $\text{Var}(\mathbf{y}_i \mid \mathbf{x}_i) = \Sigma$ does not depend on $\mathbf{x}_i$, then (3.8.12) will converge to

$$\Sigma \otimes \Phi_{xx}^{-1}. \tag{3.8.19}$$

The Chamberlain procedure of combining all T equations for a single individual into one system, obtaining the matrix of unconstrained linear-predictor coefficients and then imposing restrictions by using a minimum-distance estimator, also has a direct analog in the linear simultaneous-equations model, in which an efficient estimator is provided by applying a minimum-distance procedure to the reduce form (Malinvaud 1970, Chapter 19). We demonstrate this by considering the standard simultaneous-equations model for the time series data,[23]

$$\Gamma\mathbf{y}_t + B\mathbf{x}_t = \mathbf{u}_t, \quad t = 1.\dots, T, \tag{3.8.20}$$

and its reduced form

$$\mathbf{y}_t = \Pi\mathbf{x}_t + \mathbf{v}_t, \quad \Pi = -\Gamma^{-1}B, \quad \mathbf{v}_t = \Gamma^{-1}\mathbf{u}_t, \tag{3.8.21}$$

where Γ, B, and Π are $G \times G$, $G \times K$, and $G \times K$ matrices of coefficients, $\mathbf{y}_t$ and $\mathbf{u}_t$ are $G \times 1$ vectors of observed endogenous variables and unobserved disturbances, respectively, and $\mathbf{x}_t$ is a $K \times 1$ vector of observed exogenous variables. The $\mathbf{u}_t$ is assumed to be serially independent, with bounded variances and covariances.

[21] For proof, see Appendix 3A, Chamberlain (1982), Chiang (1956), or Malinvaud (1970).

[22] If $E(\alpha_i^* \mid \mathbf{x}_i) \neq E^*(\alpha_i^* \mid \mathbf{x}_i)$, then there will be heteroscedasticity, because the residual will contain $E(\alpha_i^* \mid \mathbf{x}_i) - E^*(\alpha_i^* \mid \mathbf{x}_i)$.

[23] For fitting model (3.8.20) to panel data, see Chapter 5.

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

In general, there are restrictions on Γ and B. We assume that the model (3.8.20) is identified by zero restrictions (e.g., Hsiao 1983) so that the gth structural equation is of the form

$$y_{gt} = \mathbf{w}'_{gt}\boldsymbol{\theta}_g + v_{gt}, \tag{3.8.22}$$

where the components of $\mathbf{w}_{gt}$ are the variables in $\mathbf{y}_t$ and $\mathbf{x}_t$ that appear in the gth equation with unknown coefficients. Let $\Gamma(\boldsymbol{\theta})$ and $B(\boldsymbol{\theta})$ be parametric representations of Γ and B that satisfy the zero restrictions and the normalization rule, where $\boldsymbol{\theta}' = (\boldsymbol{\theta}'_1, \ldots, \boldsymbol{\theta}'_G)$. Then $\boldsymbol{\pi} = \mathbf{f}(\boldsymbol{\theta}) = \text{vec}\{[-\Gamma^{-1}(\boldsymbol{\theta})B(\boldsymbol{\theta})]'\}$.

Let $\hat{\Pi}$ be the least-squares estimator of Π, and

$$\tilde{\Omega} = \frac{1}{T}\sum_{t=1}^{T}\left[(\mathbf{y}_t - \hat{\Pi}\mathbf{x}_t)(\mathbf{y}_t - \hat{\Pi}\mathbf{x}_t)' \otimes S_x^{*-1}(\mathbf{x}_t\mathbf{x}'_t)S_x^{*-1}\right], \tag{3.8.23}$$

where $S_x^* = (1/T)\Sigma_{t=1}^T \mathbf{x}_t\mathbf{x}'_t$. The generalization of the Malinvaud (1970) minimum-distance estimator is to choose $\hat{\boldsymbol{\theta}}$ to

$$\min\, [\hat{\boldsymbol{\pi}} - \mathbf{f}(\boldsymbol{\theta})]'\tilde{\Omega}^{-1}[\hat{\boldsymbol{\pi}} - \mathbf{f}(\boldsymbol{\theta})]. \tag{3.8.24}$$

Then we have $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ being asymptotically normally distributed, with mean 0 and variance–covariance matrix $(F'\tilde{\Omega}^{-1}F)^{-1}$, where $F = \partial\,\mathbf{f}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$.

The formula for $\partial\boldsymbol{\pi}/\partial\boldsymbol{\theta}'$ is given in Rothenberg (1973, p. 69):

$$F = \frac{\partial\boldsymbol{\pi}}{\partial\boldsymbol{\theta}'} = -(\Gamma^{-1} \otimes I_K)\left[\Sigma_{wx}\left(I_G \otimes \Sigma_{xx}^{-1}\right)\right]', \tag{3.8.25}$$

where Σ_{wx} is block-diagonal: $\Sigma_{wx} = \text{diag}\{E(\mathbf{w}_{1t}\mathbf{x}'_t), \ldots, E(\mathbf{w}_{Gt}\mathbf{x}'_t)\}$ and $\Sigma_{xx} = E(\mathbf{x}_t\mathbf{x}'_t)$. So we have

$$(F'\tilde{\Omega}^{-1}F)^{-1} = \{\Sigma_{wx}[E(\mathbf{u}_t\mathbf{u}'_t \otimes \mathbf{x}_t\mathbf{x}'_t)]^{-1}\Sigma'_{wx}\}^{-1}, \tag{3.8.26}$$

If $\mathbf{u}_t\mathbf{u}'_t$ is uncorrelated with $\mathbf{x}_t\mathbf{x}'_t$, then (3.8.26) reduces to

$$\{\Sigma_{wx}\Big[[E(\mathbf{u}_t\mathbf{u}'_t)]^{-1} \otimes \Sigma_{xx}^{-1}\Big]\Sigma'_{xw}\}^{-1}, \tag{3.8.27}$$

which is the conventional asymptotic covariance matrix for the three-stage least-squares (3SLS) estimator (Zellner and Theil 1962). If $\mathbf{u}_t\mathbf{u}'_t$ is correlated with $\mathbf{x}_t\mathbf{x}'_t$, then the minimum-distance estimator of $\hat{\boldsymbol{\theta}}$ is asymptotically equivalent to the Chamberlain (1982) generalized 3SLS estimator,

$$\hat{\boldsymbol{\theta}}_{G3SLS} = (S_{wx}\hat{\Psi}^{-1}S'_{wx})^{-1}(S_{wx}\hat{\Psi}^{-1}\mathbf{s}_{xy}), \tag{3.8.28}$$

where

$$S_{wx} = \text{diag}\left\{\frac{1}{T}\sum_{t=1}^{T}\mathbf{w}_{1t}\mathbf{x}'_t, \ldots, \frac{1}{T}\sum_{t=1}^{T}\mathbf{w}_{Gt}\mathbf{x}'_t\right\},$$

$$\hat{\Psi} = \frac{1}{T}\sum_{t=1}^{T}\{\hat{\mathbf{u}}_t\hat{\mathbf{u}}'_t \otimes \mathbf{x}_t\mathbf{x}'_t\}, \qquad \mathbf{s}_{xy} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{y}_t \otimes \mathbf{x}_t,$$

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

and

$$\hat{\mathbf{u}}_t = \hat{\Gamma}\mathbf{y}_t + \hat{B}\mathbf{x}_t,$$

where $\hat{\Gamma}$ and $\hat{B}$ are any consistent estimators for Γ and B. When certain equations are exactly identified, then just as in the conventional 3SLS case, applying the generalized 3SLS estimator to the system of equations, excluding the exactly identified equations, yields the same asymptotic covariance matrix as the estimator obtained by applying the generalized 3SLS estimator to the full set of G equations.[24]

However, as with any generalization, there is a cost associated with it. The minimum-distance estimator is efficient only relative to the class of estimators that do not impose a priori restrictions on the variance–covariance matrix of the error process. If the error process is known to have an error-component structure, as assumed in previous sections, the least-squares estimate of Π is not efficient (see Section 5.2), and hence the minimum-distance estimator, ignoring the specific structure of the error process, cannot be efficient, although it remains consistent.[25] The efficient estimator is the GLS estimator. Moreover, computation of the minimum-distance estimator can be quite tedious, whereas the two-step GLS estimation procedure is fairly easy to implement.

APPENDIX 3A: CONSISTENCY AND ASYMPTOTIC NORMALITY OF THE MINIMUM-DISTANCE ESTIMATOR

In this appendix we briefly sketch the proof of consistency and asymptotic normality of the minimum-distance estimator.[26] For completeness we shall state the set of conditions and properties that they imply in general forms.

Let

$$S_N = [\hat{\boldsymbol{\pi}}_N - \mathbf{f}(\boldsymbol{\theta})]' A_N [\hat{\boldsymbol{\pi}}_N - \mathbf{f}(\boldsymbol{\theta})]. \tag{3A.1}$$

Assumption 3A.1: The vector $\hat{\boldsymbol{\pi}}_N$ converges to $\boldsymbol{\pi} = f(\boldsymbol{\theta})$ in probability.[27] The matrix A_N converges to Ψ in probability, where Ψ is positive definite.

[24] This follows from examining the partitioned inverse of (3.8.26).

[25] If $\hat{\boldsymbol{\pi}}^*$ is another estimator of $\boldsymbol{\pi}$ with asymptotic variance–covariance matrix Ω^*, then the minimum-distance estimator of $\boldsymbol{\theta}$ by choosing $\hat{\boldsymbol{\theta}}^*$ to minimize $[\hat{\boldsymbol{\pi}}^* - \mathbf{f}(\boldsymbol{\theta})]'\Omega^{*-1}[\hat{\boldsymbol{\pi}}^* - \mathbf{f}(\boldsymbol{\theta})]$ has asymptotic variance–covariance matrix $(F'\Omega^{*-1}F)^{-1}$. Suppose $\Omega - \Omega^*$ is positive semidefinite; then $F'\Omega^{*-1}F - F'\Omega^{-1}F = F'(\Omega^{*-1} - \Omega^{-1})F$ is positive semidefinite. Thus, the efficiency of the minimum-distance estimator depends crucially on the efficiency of the (unconstrained) estimator of $\boldsymbol{\pi}$.

[26] For a comprehensive discussion of the Chamberlain π-approach and the GMM method, see Crépon and Mairesse (1996).

[27] In fact, a stronger result can be established for the proposition that $\hat{\boldsymbol{\pi}}$ converges to $\boldsymbol{\pi}$ almost surely. In this monograph we do not attempt to distinguish the concept of convergence in probability and convergence almost surely (Rao 1973, their Section 2.c), because the stronger result requires a lot more rigor in assumptions and derivations without much gain in intuition.

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

Assumption 3A.2: The vector $\boldsymbol{\theta}$ belongs to a compact subset of p-dimensional space. The functions $\mathbf{f}(\boldsymbol{\theta})$ possess continuous second partial derivatives, and the matrix of the first partial derivatives [equation (3.8.15)] has full column rank p in an open neighborhood containing the true parameter $\boldsymbol{\theta}$.

Assumption 3A.3: $\sqrt{N}[\hat{\boldsymbol{\pi}}_N - \mathbf{f}(\boldsymbol{\theta})]$ is asymptotically normally distributed with mean zero and variance–covariance matrix Δ.

The minimum-distance estimator chooses $\hat{\boldsymbol{\theta}}$ to minimize S_N.

Proposition 3A.1: *If assumptions 3A.1 and 3A.2 are satisfied, $\hat{\boldsymbol{\theta}}$ converges to $\boldsymbol{\theta}$ in probability.*

Proof: Assumption 3.A.1 implies that S_N converges to $S = [\mathbf{f}(\boldsymbol{\theta}) - \mathbf{f}(\hat{\boldsymbol{\theta}})]'\Psi[\mathbf{f}(\boldsymbol{\theta}) - \mathbf{f}(\hat{\boldsymbol{\theta}})] = h \geq 0$. Because min $S = 0$ and the rank condition [assumption 3A.2 or (3.8.15)] implies that in the neighborhood of the true $\boldsymbol{\theta}$, $\mathbf{f}(\boldsymbol{\theta}) = \mathbf{f}(\boldsymbol{\theta}^*)$ if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ (Hsiao 1983, p. 256), $\hat{\boldsymbol{\theta}}$ must converge to $\boldsymbol{\theta}$ in probability. Q.E.D.

Proposition 3A.2: *If assumptions 3A.1–3A.3 are satisfied, $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is asymptotically normally distributed, with mean 0 and variance–covariance matrix*

$$(F'\Psi F)^{-1}F'\Psi\Delta\Psi F(F'\Psi F)^{-1}. \tag{3A.2}$$

Proof: $\hat{\boldsymbol{\theta}}$ is the solution of

$$\mathbf{d}_N(\hat{\boldsymbol{\theta}}) = \frac{\partial S_N}{\partial \boldsymbol{\theta}}\,|_{\hat{\boldsymbol{\theta}}} = -2\left(\frac{\partial\,\mathbf{f}'}{\partial\hat{\boldsymbol{\theta}}}\right)A_N[\hat{\boldsymbol{\pi}}_N - \mathbf{f}(\hat{\boldsymbol{\theta}})] = \mathbf{0}. \tag{3A.3}$$

The mean-value theorem implies that

$$\mathbf{d}_N(\hat{\boldsymbol{\theta}}) = \mathbf{d}_N(\boldsymbol{\theta}) + \left(\frac{\partial d_N(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}'}\right)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}), \tag{3A.4}$$

where $\boldsymbol{\theta}^*$ is on the line segment connecting $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}$. Because $\hat{\boldsymbol{\theta}}$ converges to $\boldsymbol{\theta}$, direct evaluation shows that $\partial d_N(\boldsymbol{\theta}^*)/\partial\boldsymbol{\theta}'$ converges to

$$\frac{\partial d_N(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} = 2\left(\frac{\partial\,\mathbf{f}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right)'\Psi\left(\frac{\partial\,\mathbf{f}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right) = 2F'\Psi F.$$

Hence, $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ has the same limiting distribution as

$$-\left[\frac{\partial d_N(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right]^{-1}\cdot\sqrt{N}\mathbf{d}_N(\boldsymbol{\theta}) = (F'\Psi F)^{-1}F'\Psi\cdot\sqrt{N}[\hat{\boldsymbol{\pi}}_N - \mathbf{f}(\boldsymbol{\theta})]. \tag{3A.5}$$

Assumption 3A.3 says that $\sqrt{N}[\hat{\boldsymbol{\pi}}_N - \mathbf{f}(\boldsymbol{\theta})]$ is asymptotically normally distributed, with mean 0 and variance–covariance Δ. Therefore, $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is asymptotically normally distributed, with mean 0 and variance–covariance matrix given by (3A.2). Q.E.D.

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

Proposition 3A.3: *If Δ is positive definite, then*

$$(F'\Psi F)^{-1}F'\Psi\Delta\Psi F(F'\Psi F)^{-1} - (F'\Delta^{-1}F)^{-1} \tag{3A.6}$$

is positive semidefinite; hence, an optimal choice for Ψ is Δ^{-1}.

Proof: Because Δ is positive definite, there is a nonsingular matrix $\tilde{C}$ such that $\Delta = \tilde{C}\tilde{C}'$. Let $\tilde{F} = \tilde{C}^{-1}F$ and $\tilde{B} = (F'\Psi F)^{-1}F'\Psi\tilde{C}$. Then (3A.6) becomes $\tilde{B}[I - \tilde{F}(\tilde{F}'\tilde{F})^{-1}\tilde{F}']\tilde{B}'$, which is positive semidefinite. Q.E.D.

Proposition 3A.4: *Assumptions 3A.1–3A.3 are satisfied, if Δ is positive definite, and if A_N converges to Δ^{-1} in probability, then*

$$N[\hat{\boldsymbol{\pi}}_N - \mathbf{f}(\hat{\boldsymbol{\theta}})]'A_N[\hat{\boldsymbol{\pi}}_N - \mathbf{f}(\hat{\boldsymbol{\theta}})] \tag{3A.7}$$

converges to χ^2 distribution, with $KT^2 - p$ degrees of freedom.

Proof: Taking Taylor-series expansion of $\mathbf{f}(\hat{\boldsymbol{\theta}})$ around $\boldsymbol{\theta}$, we have

$$\mathbf{f}(\hat{\boldsymbol{\theta}}) \simeq \mathbf{f}(\boldsymbol{\theta}) + \frac{\partial\,\mathbf{f}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}). \tag{3A.8}$$

Therefore, for sufficiently large N, $\sqrt{N}[\mathbf{f}(\hat{\boldsymbol{\theta}}) - \mathbf{f}(\boldsymbol{\theta})]$ has the same limiting distribution as $F \cdot \sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$. Thus,

$$\sqrt{N}[\hat{\boldsymbol{\pi}}_N - \mathbf{f}(\hat{\boldsymbol{\theta}})] = \sqrt{N}[\hat{\boldsymbol{\pi}}_N - \mathbf{f}(\boldsymbol{\theta})] - \sqrt{N}[\mathbf{f}(\hat{\boldsymbol{\theta}}) - \mathbf{f}(\boldsymbol{\theta})] \tag{3A.9}$$

converges in distribution to $Q^*\tilde{C}u^*$, where $Q^* = I_{KT^2} - F(F'\Delta^{-1}F)^{-1}F'\Delta^{-1}$, $\tilde{C}$ is a nonsingular matrix such that $\tilde{C}\tilde{C}' = \Delta$, and $\mathbf{u}^*$ is normally distributed, with mean 0 and variance–covariance matrix I_{KT^2}. Then the quadratic form, (3A.7), converges in distribution of $\mathbf{u}^{*'}\tilde{C}'Q^{*'}\Delta^{-1}Q^*\tilde{C}\mathbf{u}^*$. Let $\tilde{F} = \tilde{C}^{-1}F$ and $M = I_{KT^2} - \tilde{F}(\tilde{F}'\tilde{F})^{-1}\tilde{F}'$; then M is a symmetric idempotent matrix with rank $KT^2 - p$, and $\tilde{C}'Q^{*'}\Delta^{-1}Q^*\tilde{C} = M^2 = M$; hence, (3A.7) converges in distribution to $\mathbf{u}^{*'}M\mathbf{u}^*$, which is χ^2, with $KT^2 - p$ degrees of freedom. Q.E.D.

APPENDIX 3B: CHARACTERISTIC VECTORS AND THE INVERSE OF THE VARIANCE–COVARIANCE MATRIX OF A THREE-COMPONENT MODEL

In this appendix we derive the inverse of the variance–covariance matrix (3.6.23) for a three-component model (3.6.8) by means of its characteristic roots and vectors. The material is drawn from the work of Nerlove (1971b).

The matrix $\tilde{V}$ (3.6.23) has three terms, one in I_{NT}, one in $I_N \otimes \mathbf{ee}'$, and one in $\mathbf{e}_N\mathbf{e}_N' \otimes I_T$. Thus, the vector $(\mathbf{e}_N/\sqrt{N}) \otimes (\mathbf{e}/\sqrt{T})$ is a characteristic vector, with the associated root $\sigma_u^2 + T\sigma_\alpha^2 + N\sigma_\lambda^2$. To find $NT - 1$ other characteristic vectors, we note that we can always find $N - 1$ vectors, $\boldsymbol{\psi}_j$, $j = 1, \ldots, N - 1$, each $N \times 1$ that are orthonormal and orthogonal to $\mathbf{e}_N$:

$$\begin{aligned} &\mathbf{e}'_N \boldsymbol{\psi}_j = 0, \\ &\boldsymbol{\psi}'_j \boldsymbol{\psi}_{j'} = \begin{cases} 1, & \text{if } j = j', \\ 0, & \text{if } j \neq j', \quad j = 1. \ldots, N-1, \end{cases} \end{aligned} \tag{3B.1}$$

and $T - 1$ vectors Φ_k, $k = 1, \ldots, T - 1$, each $T \times 1$, that are orthonormal and orthogonal to $\mathbf{e}$:

$$\begin{aligned} &\mathbf{e}' \Phi_k = 0 \\ &\Phi'_k \Phi_{k'} = \begin{cases} 1 & \text{if } k = k', \\ 0, & \text{if } k \neq k', \quad k = 1, \ldots, T-1, \end{cases} \end{aligned} \tag{3B.2}$$

Then the $(N - 1)(T - 1)$ vectors $\boldsymbol{\psi}_j \otimes \Phi_k$, $j = 1, \ldots, N - 1$, $k = 1, \ldots, T - 1$, the $N - 1$ vectors $\boldsymbol{\psi}_j \otimes (\mathbf{e}/\sqrt{T})$, $j = 1, \ldots, N - 1$, and the $T - 1$ vectors $\mathbf{e}_N/\sqrt{N} \otimes \Phi_k$, $k = 1, \ldots, T - 1$, are also characteristic vectors of $\tilde{V}$, with the associated roots σ_u^2, $\sigma_u^2 + T\sigma_\alpha^2$, and $\sigma_u^2 + N\sigma_\lambda^2$, which are of multiplicity $(N - 1)(T - 1)$, $(N - 1)$, and $(T - 1)$, respectively.

Let

$$\begin{aligned} C_1 &= \frac{1}{\sqrt{T}}[\boldsymbol{\psi}_1 \otimes \mathbf{e}, \ldots, \boldsymbol{\psi}_{N-1} \otimes \mathbf{e}], \\ C_2 &= \frac{1}{\sqrt{N}}[\mathbf{e}_N \otimes \Phi_1, \ldots, \mathbf{e}_N \otimes \Phi_{T-1}], \\ C_3 &= [\boldsymbol{\psi}_1 \otimes \Phi_1, \boldsymbol{\psi}_1 \otimes \Phi_2, \ldots, \boldsymbol{\psi}_{N-1} \otimes \Phi_{T-1}], \\ C_4 &= \left(\mathbf{e}_N/\sqrt{N}\right) \otimes \left(\mathbf{e}/\sqrt{T}\right) = \frac{1}{\sqrt{NT}}\mathbf{e}_{NT}, \end{aligned} \tag{3B.3}$$

and

$$C = [C_1 \quad C_2 \quad C_3 \quad C_4]. \tag{3B.4}$$

Then

$$CC' = C_1C'_1 + C_2C'_2 + C_3C'_3 + C_4C'_4 = I_{NT}, \tag{3B.5}$$

$$C\tilde{V}C' =$$

$$\begin{bmatrix} (\sigma_u^2 + T\sigma_\alpha^2)I_{N-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_u^2 + N\sigma_\lambda^2 I_{T-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_u^2 I_{(N-1)(T-1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_u^2 + T\sigma_\alpha^2 + N\sigma_\lambda^2 \end{bmatrix} = \Lambda, \tag{3B.6}$$

and

$$\tilde{V} = C\Lambda C'.$$

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press

Let $A = I_N \otimes \mathbf{e}\mathbf{e}'$, $D = \mathbf{e}_N \mathbf{e}_N' \otimes I_T$, and $J = \mathbf{e}_{NT}\mathbf{e}_{NT}'$. From

$$C_4 C_4' = \frac{1}{NT} J, \tag{3B.7}$$

Nerlove (1971b) showed that by premultiplying (3B.5) by A, we have

$$C_1 C_1' = \frac{1}{T} A - \frac{1}{NT} J, \tag{3B.8}$$

and premultiplying (3B.5) by D,

$$C_2 C_2' = \frac{1}{N} D - \frac{1}{NT} J. \tag{3B.9}$$

Premultiplying (3B.5) by A and D and using the relations (3B.5), (3B.7), (3B.8), and (3B.9), we have

$$C_3 C_3' = I_{NT} - \frac{1}{T} A - \frac{1}{N} D + \frac{1}{NT} J = \tilde{Q}. \tag{3B.10}$$

Because $\tilde{V}^{-1} = C \Lambda^{-1} C'$, it follows that

$$\begin{aligned} \tilde{V}^{-1} = {} & \frac{1}{\sigma_u^2 + T\sigma_\alpha^2}\left(\frac{1}{T} A - \frac{1}{NT} J\right) + \frac{1}{\sigma_u^2 + N\sigma_\lambda^2}\left(\frac{1}{N} D - \frac{1}{NT} J\right) \\ & + \frac{1}{\sigma_u^2}\tilde{Q} + \frac{1}{\sigma_u^2 + T\sigma_\alpha^2 + N\sigma_\lambda^2}\left(\frac{1}{NT} J\right). \end{aligned} \tag{3B.11}$$

https://doi.org/10.1017/CBO9781139839327.006 Published online by Cambridge University Press