21 : CHAPTER

The Analysis of Covariance

The Models

In the previous chapter, we have considered a model in the form of

$$y_{tj} = \mu + \gamma_t + \delta_j + \varepsilon_{tj} \tag{1}$$

wherein $t = 1, \ldots, T$ and $j = 1, \ldots, M$ are respectively indices of temporal and spatial location. In this chapter, we shall relate the index j to individual persons or to individual units of production such as farms or factories.

We shall also elaborate the model by introducing a function $x_{t.}\beta_{.tj} = \sum_k x_{tk}\beta_{ktj}$ comprising K explanatory variables or regressors. Of course, if the elements β_{ktj} were to vary across all of the indices, then we should have no chance of making any reasonable inference about their values unless some further assumptions were made as to the nature of this variation.

Without further ado, we shall make the very restrictive assumption that $\beta_{ktj} = \beta_{kj}$ for all t, which is to say that there is no temporal variation in these coefficients. If, in addition, we assume that $\gamma_t = 0$ for all t, then our model can be written as

$$\begin{aligned}(y_{tj}e_{jt}) &= \mu(e_{jt}) + (\delta_j e_{jt}) + (\{x_{tk}\beta_{kj}\}e_{jt}) + (\varepsilon_{tj}e_{jt}) \\ &= \mu(e_{jt}) + (\delta_j e_{jt}) + (x_{tk}e_{jt}^{jk})(\beta_{kj}e_{jk}) + (\varepsilon_{tj}e_{jt}).\end{aligned} \tag{2}$$

Here the braces which surround the expression $\{x_{tk}\beta_{kj}\}$ are to indicate that a sum has been taken over the repeated index k.

The set of T realisations of the jth equation can be written as

$$\begin{aligned}y_{.j} &= \mu\iota_T + \delta_j\iota_T + X\beta_{.j} + \varepsilon_{.j} \\ &= \alpha_j\iota_T + X\beta_{.j} + \varepsilon_{.j},\end{aligned} \tag{3}$$

where $\alpha_j = \mu + \delta_j$; and we can compile the full set of M such equations to give the following system:

(4)

$$\begin{bmatrix} y_{.1} \\ y_{.2} \\ \vdots \\ y_{.M} \end{bmatrix} = \begin{bmatrix} \iota_T & 0 & \ldots & 0 \\ 0 & \iota_T & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \iota_T \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_M \end{bmatrix} + \begin{bmatrix} X & 0 & \ldots & 0 \\ 0 & X & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & X \end{bmatrix} \begin{bmatrix} \beta_{.1} \\ \beta_{.2} \\ \vdots \\ \beta_{.M} \end{bmatrix} + \begin{bmatrix} \varepsilon_{.1} \\ \varepsilon_{.2} \\ \vdots \\ \varepsilon_{.M} \end{bmatrix}.$$

A useful elaboration of this model, which costs little in terms of added difficulty, is to allow the matrix X to vary between the M equations. Then in place of the variables x_{tk} we have elements x_{tkj} bearing the spatial subscript j. In that case, equation (4) is replaced by

$$\begin{aligned}\begin{bmatrix} y_{.1} \\ y_{.2} \\ \vdots \\ y_{.M} \end{bmatrix} &= \begin{bmatrix} \iota_T & 0 & \ldots & 0 \\ 0 & \iota_T & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \iota_T \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_M \end{bmatrix} \\ &+ \begin{bmatrix} X_1 & 0 & \ldots & 0 \\ 0 & X_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & X_M \end{bmatrix} \begin{bmatrix} \beta_{.1} \\ \beta_{.2} \\ \vdots \\ \beta_{.M} \end{bmatrix} + \begin{bmatrix} \varepsilon_{.1} \\ \varepsilon_{.2} \\ \vdots \\ \varepsilon_{.M} \end{bmatrix}. \end{aligned} \tag{5}$$

It may be, for example, that our equations, which explain farm production in M regions, comprise explanatory variables whose measured values vary from region to region.

Within the context of this model, we shall consider some more restrictive hypotheses. The first of these, which is denoted by

$$H_\beta : \beta_1 = \beta_2 = \cdots = \beta_M, \tag{6}$$

asserts that the slope parameters of all M of the regression equations are equal. This condition gives rise to a model in the form of

$$\begin{bmatrix} y_{.1} \\ y_{.2} \\ \vdots \\ y_{.M} \end{bmatrix} = \begin{bmatrix} \iota_T & 0 & \ldots & 0 \\ 0 & \iota_T & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \iota_T \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_M \end{bmatrix} + \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_M \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_{.1} \\ \varepsilon_{.2} \\ \vdots \\ \varepsilon_{.M} \end{bmatrix} \tag{7}$$

wherein each of the M equations is distinguished by having a particular value for the intercept.

The second hypothesis, which is denoted by

$$H_\alpha : \alpha_1 = \alpha_2 = \cdots = \alpha_M, \tag{8}$$

asserts that all of the intercepts have the same value. It is unlikely that we should ever wish to maintain this hypothesis without asserting H_β at the same time. The combined hypothesis $H_\gamma = H_\alpha \cap H_\beta$ gives rise to a model in the form of

$$\begin{bmatrix} y_{.1} \\ y_{.2} \\ \vdots \\ y_{.M} \end{bmatrix} = \begin{bmatrix} \iota_T \\ \iota_T \\ \vdots \\ \iota_T \end{bmatrix} \alpha + \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_M \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_{.1} \\ \varepsilon_{.2} \\ \vdots \\ \varepsilon_{.M} \end{bmatrix}. \tag{9}$$

The Least-Squares Estimates of the Models

We shall assume that the disturbances ε_{tj} are distributed independently and identically with $E(\varepsilon_{tj}) = 0$ and $V(\varepsilon_{tj}) = \sigma^2$ for all t, j. Under these assumptions, the equations of the model represented by (5) are wholly separable, and we can easily see that the parameters of the jth equation may be estimated efficiently by the formulae

$$(10)\qquad \begin{aligned} &\hat{\beta}_{.j} = \left\{X_j'(I - P_T)X_j\right\}^{-1} X_j'(I - P_T)y_{.j} \quad \text{and} \\ &\hat{\alpha}_j = \bar{y}_j - \bar{x}_{j.}\hat{\beta}_{.j}, \end{aligned}$$

where

$$(11)\qquad I - P_T = I - \iota_T(\iota_T'\iota_T)^{-1}\iota_T'$$

is the operator which transforms a vector of T observations into the vector of their deviations about the mean.

The residual sum of squares from the jth regression is given by

$$(12)\qquad S_j = y_{.j}'(I - P_T)y_{.j} - y_{.j}'(I - P_T)X_j\{X_j'(I - P_T)X_j\}^{-1}X_j'(I - P_T)y_{.j};$$

and therefore, from the separability of the M regressions, it follows that the residual sum of squares obtained from fitting the model of (5) to the data is just

$$(13)\qquad S = \sum_j S_j.$$

The formulae under (10) are familiar from our treatment of the linear regression model $(y; i\alpha + X\beta, \sigma^2 I)$, which was regarded as a particular instance of the partitioned model $(y; X_1\beta_1 + X_2\beta_2, \sigma^2 I)$.

We may recall that one way of developing the ordinary least-squares estimator of β_2 in the partitioned model depends on transforming the equation $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$ by the matrix $I - P_1$ where $P_1 = X_1(X_1'X_1)^{-1}X_1'$ is the orthogonal projector on the manifold of X_1. The effect of this transformation is to annihilate the term $X_1\beta_1$, which leads to the equation $(I - P_1)y = (I - P_1)X_2\beta_2 + (I - P_1)\varepsilon$. When ordinary least-squares regression is applied to the transformed equation, we get

$$(14)\qquad \hat{\beta}_2 = \left\{X_2'(I - P_1)X_2\right\}^{-1} X_2'(I - P_1)y.$$

The same estimator may be derived by applying ordinary least-squares regression to the equation $Q'y = Q'X_2\beta_2 + Q'\varepsilon$ where Q is a matrix of orthonormal vectors such that $QQ' = (I - P_T)$. Since $D(Q'\varepsilon) = \sigma^2 I_{T-k_1}$, it follows that the equation fulfils the assumptions of the classical linear model; and we can use a standard form of the Gauss–Markov theorem to demonstrate the efficiency of the estimator $\hat{\beta}_2$.

In the case of system under (5) the intercept term α_j is eliminated from each equation by premultiplying the equation by $(I - P_T)$. The intercept terms may be eliminated from the full system of of equations by premultiplying it by

$$W = I_M \otimes (I_T - P_T) = I_{MT} - (I_M \otimes P_T). \tag{15}$$

Now, let us consider fitting the model under (7) which may be regraded as a variant of the model under (5) which has been subjected to the restrictions of H_β of (6). It can be seen, in a variety of ways, that the efficient estimates of the parameters are given by the equations

$$\begin{aligned}\hat{\beta} &= \Big[\sum_j X_j'(I - P_T)X_j\Big]^{-1}\Big[\sum_j X_j'(I - P_T)y_j\Big]\\ &= \Big[X'\big\{I_M \otimes (I_T - P_T)\big\}X\Big]^{-1}X'\big\{I_M \otimes (I_T - P_T)\big\}y \quad \text{and}\\ \hat{\alpha}_j &= \bar{y}_{j.} - \bar{x}_j\hat{\beta},\end{aligned} \tag{16}$$

where $X' = [X_1', X_2', \ldots, X_M']$ and $y = [y_{.1}' y_{.2}', \ldots, y_{.M}']$. The estimator is clearly the result of applying ordinary least-squares regression to an equation derived by premultiplying (7) by the matrix W of (15) which serves to annihilate the intercept terms.

The residual sum of squares from fitting the model of (7) is given by

$$\begin{aligned}&S_\alpha = y'Wy - y'WX(X'WX)^{-1}X'Wy \quad \text{where}\\ &W = I_M \otimes (I_T - P_T)\end{aligned} \tag{17}$$

Finally, let us consider fitting the model of (9) which makes no distinction between the structures of the M equations. Let P_{MT} to denote the projector in the form of $P_{MT} = \iota_{MT}(\iota_{MT}'\iota_{MT})^{-1}\iota_{MT}'$ when ι_{MT} is the summation vector of order MT. Then we can write the estimators of the parameters of the model under (9) as

$$\begin{aligned}&\hat{\beta} = \{X'(I - P_{MT})X\}^{-1}X'(I - P_{MT})y\\ &\hat{\alpha} = \bar{y} - \bar{x}\hat{\beta}.\end{aligned} \tag{18}$$

The residual sum of squares from fitting the restricted model of (9) is given by

$$S_\gamma = (y - \alpha\iota_{MT} - X\hat{\beta})'(y - \alpha\iota_{MT} - X\hat{\beta}). \tag{19}$$

The Tests of the Restrictions

In order to test the various hypotheses, we need the following results concerning the distribution of the residual sum of squares from each of the regressions that we have considered:

$$(20) \qquad \begin{aligned} &1. \quad \frac{1}{\sigma^2}S \sim \chi^2\{MT - M(K+1)\},\\ &2. \quad \frac{1}{\sigma^2}S_\beta \sim \chi^2\{MT - (K+M)\},\\ &3. \quad \frac{1}{\sigma^2}S_\gamma \sim \chi^2\{MT - (K+1)\}. \end{aligned}$$

The number of degrees of freedom in each of these cases is easily explained. It is simply the number of observations available in the vector $y' = [y'_{.1}, y'_{.2}, \ldots, y'_{.M}]$ less the number of parameters that are estimated in the particular model.

We test the hypothesis H_β by assessing the loss of fit which results from imposing the restrictions $\beta_1 = \beta_2 = \cdots = \beta_M$. The loss is given by $S_\beta - S$. The residual sum of squares S from the unrestricted model is the standard against which we measure this loss. The appropriate test statistic is therefore

$$(21) \qquad F = \left\{ \frac{S_\beta - S}{(M-1)K} \bigg/ \frac{S}{MT - M(K+1)} \right\}$$

which has a F distribution of $(M-1)K$ and $MT - M(K+1)$ degrees of freedom.

If the hypothesis H_β is accepted, then we might proceed to test the more stringent hypothesis $H_\gamma = H_\beta \cap H_\alpha$ which entails the additional restrictions of $H_\alpha : \alpha_1 = \alpha_2 = \cdots = \alpha_M$. The relevant test statistic in this case is given by

$$(22) \qquad F = \left\{ \frac{S_\gamma - S_\beta}{M-1} \bigg/ \frac{S_\beta}{MT - (K+M)} \right\}$$

which has a F distribution of $M-1$ and $MT - (K+M)$ degrees of freedom. The numerator of this statistic embodies a measure of the loss of fit that comes from imposing the additional restrictions of H_α.

The statistic under (22) tests the hypothesis H_γ within the context of an assumption that H_β is true. We might decide to test additionally, or even alternatively, the joint hypothesis $H_\gamma = H_\alpha \cap H_\beta$ within the context of the unrestricted model. The relevant statistic in that case would be given by

$$(23) \qquad F = \left\{ \frac{S_\gamma - S}{(M-1)(K+1)} \bigg/ \frac{S}{MT - M(K+1)} \right\}$$

We have to consider the possibility that, having accepted the hypotheses H_β and H_α on the strength of the values the F statistics under (21) and (22), we shall then discover that value of the statistic of (23) casts doubt on the joint hypothesis $H_\gamma = H_\beta \cap H_\alpha$. The possibility arises from the fact that critical region of the test of H_γ can never coincide with the critical region of the joint test implicit in the sequential procedure. However, if the critical value of the test H_γ has been appropriately chosen, then such a conflict in the results of the tests is an unlikely eventuality.

Models with Two-way Fixed Effects

Let us now elaborate the model of H_β by including the parameters γ_t which represent the temporal variation which is experienced by all J individuals. Our equation now assumes the form of

$$(y_{tj}e_{jt}) = \mu(e_{jt}) + (x_{tjk}e_{jt}^{k})(\beta_k e_k) + (e_{jt}^{t})(e_t\gamma_t) + (e_{jt}^{j})(e_j\delta_j) + (\varepsilon_{tj}e_{jt}). \tag{24}$$

Comparison with equation (2) shows that we are assuming that $\beta_{kj} = \beta_k$ for all k which is to say that all j individuals share the same slope coefficients. Therefore, the system of equations as a whole can be represented by

$$Y^c = \mu\iota_{MT} + X\beta + (\iota_M \otimes I_T)\gamma + (I_M \otimes \iota_T)\delta + \mathcal{E}^c. \tag{25}$$

The matrix $[\iota_{MT}, X, \iota_M \otimes I_T, I_M \otimes \iota_T]$ which contains the regressors of the model is, in fact, singular by virtue of the linear dependence which exists between the columns of its submatrix $[\iota_{MT}, \iota_M \otimes I_T, I_M \otimes \iota_T]$. This dependence is made clear by writing the equation

$$(\iota_M \otimes I_T)(1 \otimes \iota_T) = (I_M \otimes \iota_T)(\iota_M \otimes 1) = \iota_{MT}. \tag{26}$$

Therefore, the parameters μ, β,γ, δ are not estimable as a whole unless we are prepared to introduce some restrictions. The obvious conditions are that $\iota_T'\gamma = \sum_t \gamma_t = 0$ and that $\iota_{MT}'\delta = \sum_j \delta_j = 0$.

To derive the ordinary least-squares estimate of β, we can begin by transforming our chosen equation in such a way as to eliminate the parameters γ, δ. The can be accomplished by premultiplying the equation by the matrix

$$\begin{aligned} W &= [I_{MT} - (I_M \otimes P_T)][I_{MT} - (P_M \otimes I_M)] \\ &= I_{MT} - (I_M \otimes P_T) - (P_M \otimes I_M) + (I_M \otimes P_T)(P_M \otimes I_M). \end{aligned} \tag{27}$$

The two factors commute. The first factor $I_{MT} - (I_M \otimes P_T)$ has the effect of annihilating the term $(I_M \otimes \iota_T)\delta$ whilst the second factor $I_{MT} - (P_M \otimes I_M)$ has the effect of annihilating $(\iota_M \otimes I_T)\gamma$. However, since W is a symmetric idempotent matrix, it can be written in the form of $W = QQ'$ where Q is a matrix of order $MT \times (MT - M - T)$ consisting or orthonormal vectors. Therefore, we may, with equal effect transform, our equation by premultiplying by Q' to obtain the system

$$Q'Y^c = Q'X\beta + \mathcal{E}^c. \tag{28}$$

The latter fulfils the assumptions of the classical linear model. It follows that the efficient estimator of β is given by

$$\begin{aligned}\beta &= (X'QQ'X)^{-1}X'QQ'y \\ &= (X'WX)^{-1}X'Wy.\end{aligned} \tag{29}$$

Models with Random Effects

An alternative way of accommodating temporal and individual effects is to regard them as random variables rather than as fixed constants. To signify the difference in approach that is entailed by adopting a random-effects model, we shall denote the model by

$$(y_{tj}e_{jt}) = \mu e_{jt} + (x_{jtk}e^k_{jt})(\beta_k e_k) + (e^t_{jt})(e_t\zeta_t) + (e^j_{jt})(e_j\eta_j) + (\varepsilon_{jt}e_{jt}) \tag{30}$$

A set of T realisations on all M equations is now written as

$$y = \mu\iota_{MT} + X\beta + (\iota_M \otimes I_T)\zeta + (I_M \otimes \iota_T)\eta + \varepsilon \tag{31}$$

We assume that the random variables ζ_t, η_j and ε_{tj} are independently distributed with $V(\zeta_t) = \sigma^2_\zeta, V(\eta_j) = \sigma^2_\eta$ and $V(\varepsilon_{tj}) = \sigma^2_\varepsilon$. It follows that the dispersion matrix of the vector of disturbances in this model is given by

$$\Omega = \sigma^2_\zeta(\iota_M\iota'_M \otimes I_T) + \sigma^2_\eta(I_M \otimes \iota_T\iota'_T) + \sigma^2_\varepsilon I_{MT}. \tag{32}$$

A special case that is often considered arises when $\sigma^2_\eta = 0$, which is to say that there is no intertemporal variation in the structure of the stochastic disturbances. In that case, the dispersion matrix is of the form

$$\begin{aligned}\Omega &= \sigma^2_\eta(I_M \otimes \iota_T\iota'_T) + \sigma^2_\varepsilon I_{MT} \\ &= I_M \otimes (\sigma^2_\varepsilon I_T + \sigma^2_\eta\iota_T\iota_T) = I_M \otimes V.\end{aligned} \tag{33}$$

It can be confirmed by direct multiplication that the inverse of the matrix $V = \sigma^2\varepsilon I_T + \sigma^2_\eta\iota_T\iota_T$ is

$$V^{-1} = \frac{1}{\sigma^2_\varepsilon}\left(I_T - \frac{\sigma^2_\eta}{\sigma^2_\varepsilon + T\sigma^2_\eta}\iota_T\iota'_T\right). \tag{34}$$

The inverse of the matix Ω of (32) has a somewhat complicated structure. It takes the form of the form of

(35)

$$\Omega^{-1} = \frac{1}{\sigma_\varepsilon^2}\{I_{MT} - \lambda_1(\iota_M\iota'_{MT} \otimes I_T) + \lambda_2(I_M \otimes \iota_T\iota'_T) + \lambda_3(\iota_M\iota'_{MT} \otimes \iota_T\iota_T)\}$$

$$\text{where} \quad \lambda_1 = \sigma_\zeta^2(\sigma_\varepsilon^2 - M\sigma_\zeta^2)^{-1},$$

$$\lambda_2 = \sigma_\eta^2(\sigma_\varepsilon^2 - T\sigma_\eta^2)^{-1},$$

$$\lambda_3 = \lambda_1\lambda_2(2\sigma_\varepsilon^2 + M\sigma_\zeta^2 + T\sigma_\eta^2)(\sigma_\varepsilon^2 + M\sigma_\zeta^2 + T\sigma_\eta^2)^{-1}.$$