MATRIX KRONECKER PRODUCTS

Consider the matrix equation $Y = AXB'$. When all of the factors are 2×2 matrices, this becomes

$$\begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} = \left\{ \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \right\} \begin{bmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{bmatrix}$$

$$= \left\{ A \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix}, A \begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix} \right\} \begin{bmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{bmatrix}$$

$$= \left\{ b_{11} A \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} + b_{12} A \begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix}, b_{21} A \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} + b_{22} A \begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix} \right\}.$$

Stacking the columns of Y vertically yields

$$\begin{bmatrix} y_{11} \\ y_{21} \\ y_{12} \\ y_{22} \end{bmatrix} = \begin{bmatrix} b_{11} A & b_{12} A \\ b_{21} A & b_{22} A \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{21} \\ x_{12} \\ x_{22} \end{bmatrix}.$$

The *vectorisation* of Y can be denoted by writing

$$Y^c = (AXB')^c = (B \otimes A) X^c.$$

The Kronecker product of the matrices A and B is defined by

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{11}B & \dots & a_{1n}B \\ a_{21}B & a_{21}B & \dots & a_{2m}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{bmatrix}.$$

The following rules govern Kronecker products:

$$\begin{aligned} &\text{(i)} && (A \otimes B)(C \otimes D) = AC \otimes BD, \\ &\text{(ii)} && (A \otimes B)' = A' \otimes B', \\ &\text{(iii)} && A \otimes (B + C) = (A \otimes B) + (A \otimes C), \\ &\text{(iv)} && \lambda(A \otimes B) = \lambda A \otimes B = A \otimes \lambda B, \\ &\text{(v)} && (A \otimes B)^{-1} = (A^{-1} \otimes B^{-1}). \end{aligned}$$

The product is non-commutative, which is to say that $A \otimes B \neq B \otimes A$. However, observe that

$$A \otimes B = (A \otimes I)(I \otimes B) = (I \otimes B)(A \otimes I).$$

ECONOMETRIC PANEL DATA

The Unrestricted Model with Common Regressors

Consider M firms or individuals indexed by $j = 1, \ldots, M$. These react differently to the same measured influences, as indicated by the parameters α_j and $\beta_{.j}$. The set of T realisations of the jth equation can be written as

$$y_{.j} = \alpha_j \iota_T + X\beta_{.j} + \varepsilon_{.j}. \tag{8}$$

This is a classical regression equation that can be estimated by OLS. The full set of M such equations can be compiled as the system:

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

Using the Kronecker product, this can be rendered as

$$Y^c = (I_M \otimes \iota_T)\alpha + (I_M \otimes X)B^c + \mathcal{E}^c. \tag{10}$$

The General Model

A useful elaboration is to allow the matrix X to vary between the M equations. Then, in place of the variables x_{tk}, there are elements x_{tkj} bearing the individual-specific subscript j. Then, equation (9) is replaced by

$$
\begin{aligned}
\begin{bmatrix} y_{.1} \\ y_{.2} \\ \vdots \\ y_{.M} \end{bmatrix} &= \begin{bmatrix} \iota_T & 0 & \ldots & 0 \\ 0 & \iota_T & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \iota_T \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_M \end{bmatrix} \\
&\quad + \begin{bmatrix} X_1 & 0 & \ldots & 0 \\ 0 & X_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & X_M \end{bmatrix} \begin{bmatrix} \beta_{.1} \\ \beta_{.2} \\ \vdots \\ \beta_{.M} \end{bmatrix} + \begin{bmatrix} \varepsilon_{.1} \\ \varepsilon_{.2} \\ \vdots \\ \varepsilon_{.M} \end{bmatrix}.
\end{aligned} \tag{11}
$$

For example, the equations, explaining farm production in M regions, may comprise explanatory variables whose measured values vary from region to region.

The jth equation of the general model can be written as

$$y_{.j} = \iota_T \alpha_j + X_j \beta.j + \varepsilon_{.j}. \tag{12}$$

The Model with Individual Fixed Effects

Within this model, some restrictions can be imposed. Thus

$$H_\beta : \beta_{.1} = \beta_{.2} = \cdots = \beta_{.M}, \tag{13}$$

asserts that the slope parameters of all M of the regression equations are equal. This condition gives rise to the following model:

$$\begin{bmatrix} y_{.1} \\ y_{.2} \\ \vdots \\ y_{.M} \end{bmatrix} = \begin{bmatrix} \iota_T & 0 & \ldots & 0 \\ 0 & \iota_T & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \iota_T \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_M \end{bmatrix} + \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_M \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_{.1} \\ \varepsilon_{.2} \\ \vdots \\ \varepsilon_{.M} \end{bmatrix}, \tag{14}$$

Here, each of the M equations has a particular value for the intercept. This system of equations can be rendered as

$$Y^c = (I_M \otimes \iota_T)\alpha + X\beta + \mathcal{E}^c, \tag{15}$$

where $X' = [X_1', X_2', \ldots, X_M']$. It is common, in many many textbooks, to write $(I_M \otimes \iota_T) = D$ for the so-called matrix of dummy variables associated with the intercept terms.

The Pooled Model

A further hypothesis is that all of the intercepts have the same value:

$$H_\alpha : \alpha_1 = \alpha_2 = \cdots = \alpha_M. \tag{16}$$

It is unlikely that one would maintain this hypothesis without asserting H_β at the same time. The combined hypothesis $H_\gamma = H_\alpha \cap H_\beta$ gives rise to

$$\begin{bmatrix} y_{.1} \\ y_{.2} \\ \vdots \\ y_{.M} \end{bmatrix} = \begin{bmatrix} \iota_T \\ \iota_T \\ \vdots \\ \iota_T \end{bmatrix} \alpha + \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_M \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_{.1} \\ \varepsilon_{.2} \\ \vdots \\ \varepsilon_{.M} \end{bmatrix}. \tag{17}$$

This system of equations can be rendered as

$$Y^c = \iota_{MT}\alpha + X\beta + \mathcal{E}^c, \tag{18}$$

where, as before, $X' = [X_1', X_2', \ldots, X_M']$ and where ι_{MT} is a long vector consisting of MT units. This has the structure of the equation of a classical regression model, for which OLS is efficient.

The Least-Squares Estimation: the General Unrestricted Model

We 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. Then, the equations of the general model of (11) are separable, and the jth equation may be estimated efficiently by OLS.

The regression can be applied directly to the individual equations of (11). Alternatively, the intercept terms can be eliminated by taking the deviations of the data about their respective sample means. Then, the estimates for the jth equation are

$$\begin{aligned} \hat{\beta}_{.j} &= \left\{X_j'(I - P_T)X_j\right\}^{-1} X_j'(I - P_T)y_{.j} \quad \text{and} \\ \hat{\alpha}_j &= \bar{y}_j - \bar{x}_{j.}\hat{\beta}_{.j}, \end{aligned} \tag{19}$$

where

$$I - P_T = I - \iota_T(\iota_T'\iota_T)^{-1}\iota_T' \tag{20}$$

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

The intercept terms would be eliminated by premultiplying the full system of equations by

$$\begin{aligned} W &= I_M \otimes (I_T - P_T) = I_{MT} - (I_M \otimes P_T) \\ &= I_{MT} - D(D'D)^{-1}D'. \end{aligned} \tag{24}$$

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

$$S_j = y'_{.j}(I - P_T)y_{.j} - y'_{.j}(I - P_T)X_j\{X'_j(I - P_T)X_j\}^{-1}X'_j(I - P_T)y_{.j}. \tag{21}$$

From the separability of the M regressions, it follows that the residual sum of squares, obtained from fitting the multi-equation model of (11) to the data, is

$$S = \sum_j S_j. \tag{22}$$

Estimation of the Model with Individual Fixed Effects

Now, consider fitting the model under (14), which arises from the general model (11) when the slope parameters of the regressions are the same in every region. Then, the efficient estimates are obtained by treating the system of equations as a whole.

To eliminate the intercept terms, the individual equations $y_{.j} = \alpha_j \iota_T + X_j\beta + \varepsilon_j$ are multiplied by the operator $I - P_T$, to create deviations of the data about sample means:

$$(I - P_T)y_{.j} = (I - P_T)X_j\beta + (I - P_T)\varepsilon_j, \tag{25}$$

The equation at time t is

$$y_{tj} - \bar{y}_j = (x_{.tj} - \bar{x}_{.j})\beta + (\varepsilon_{tj} - \bar{\varepsilon}_j). \tag{26}$$

To obtain an efficient estimate of β, the full set of TM mean-adjusted equations must be taken together. Once the estimate of β available, the individual intercept terms can be obtained.

The efficient system-wide estimates for the slope parameters and the intercepts are

$$
\begin{aligned}
\hat{\beta}_W &= \Big[\sum_j X_j'(I - P_T)X_j\Big]^{-1}\Big[\sum_j X_j'(I - P_T)y_j\Big] \\
&= \Big[X'\{I_M \otimes (I_T - P_T)\}X\Big]^{-1} X'\{I_M \otimes (I_T - P_T)\}y \quad \text{and}
\end{aligned} \tag{27}
$$

$$
\hat{\alpha}_j = \bar{y}_{j.} - \bar{x}_j\hat{\beta}_W, \quad j = 1, \ldots, M,
$$

where $X' = [X_1', X_2', \ldots, X_M']$ and $y' = [y_{.1}'y_{.2}', \ldots, y_{.M}']$. Here, $\hat{\beta}_W$ is the result of applying OLS to an equation derived by premultiplying (14) by the matrix W of (24), to annihilate the intercept terms.

The residual sum of squares from fitting the model of (14) is given by

$$
\begin{aligned}
S_\beta &= y'Wy - y'WX(X'WX)^{-1}X'Wy, \quad \text{where} \\
W &= I_M \otimes (I_T - P_T).
\end{aligned} \tag{28}
$$

The estimator $\hat{\beta}_W$ of (27) makes use only of the information conveyed by the deviations of the data points from their means within the M groups. For this reason, it is often called the *within-groups estimator.*

The operator that averages over the sample of T observations is $P_T = \iota_T(\iota_T'\iota_T)^{-1}\iota_T'$. Applying it to the generic equation of (14) gives

$$\begin{aligned} P_T y_{.j} &= P_T\iota_T\alpha_{.j} + P_T X_j\beta + P_T\varepsilon_j \quad \text{or} \\ \iota_T\bar{y} &= \iota_T\alpha_j + \iota_T\bar{x}_{.j}\beta + \bar{\iota}_T\varepsilon_j. \end{aligned} \tag{29}$$

This equation comprises T redundant repetitions of the equation

$$\bar{y}_j = \alpha + \bar{x}_{.j}\beta + (\alpha_j - \alpha + \bar{\varepsilon}_j), \tag{30}$$

where $\alpha = \sum_j \alpha_j/M$ is a global or averaged intercept term, and where the deviation of the jth intercept α_j from global value has been joined with the averaged disturbance term.

Having gathering together the M equations of this form, an ordinary least-squares estimate of β, denoted $\hat{\beta}_B$ can be obtained which is described as the *between-groups estimator.* This estimator uses only the information conveyed by the variation amongst the M group means.

The Pooled Estimator

Finally, consider fitting the model of (17) which, on the basis of the hypothesis H_γ, makes no distinction between the structures of the M equations. Let P_{MT} denote the projector $P_{MT} = \iota_{MT}(\iota'_{MT}\iota_{MT})^{-1}\iota'_{MT}$, where ι_{MT} is the summation vector of order MT. Then, the estimators of the parameters of the model can be written as

$$\begin{aligned} \hat{\beta}_G &= \{X'(I - P_{MT})X\}^{-1}X'(I - P_{MT})y \\ \hat{\alpha} &= \bar{y} - \bar{x}\hat{\beta}. \end{aligned} \tag{31}$$

The residual sum of squares from fitting the restricted model of (17) is given by

$$S_\gamma = (y - \alpha\iota_{MT} - X\hat{\beta})'(y - \alpha\iota_{MT} - X\hat{\beta}). \tag{32}$$

The Tests of the Restrictions

In order to test the various hypotheses, the following results are needed, which concern the distribution of the residual sum of squares from each of the regressions that have been considered:

$$(33)\qquad \begin{aligned} &1. \quad \frac{1}{\sigma^2}S \sim \chi^2\{MT - M(K+1)\},\\ &2. \quad \frac{1}{\sigma^2}S_\beta \sim \chi^2\{MT - (K+M)\},\\ &3. \quad \frac{1}{\sigma^2}S_\gamma \sim \chi^2\{MT - (K+1)\}. \end{aligned}$$

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.

The hypothesis H_β can be tested by assessing the loss of fit from imposing the restrictions $\beta_1 = \beta_2 = \cdots = \beta_M$. The loss is given by $S_\beta - S$. This is measured in comparison with residual sum of squares S from the unrestricted model. The appropriate test statistic is

$$F = \left\{ \frac{S_\beta - S}{(M-1)K} \Big/ \frac{S}{MT - M(K+1)} \right\}, \tag{34}$$

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 one 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 is

$$F = \left\{ \frac{S_\gamma - S_\beta}{M-1} \Big/ \frac{S_\beta}{MT - (K+M)} \right\}, \tag{35}$$

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 of (35) tests the hypothesis H_γ within the context of an assumption that H_β is true. One 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

$$F = \left\{ \frac{S_\gamma - S}{(M-1)(K+1)} \bigg/ \frac{S}{MT - M(K+1)} \right\}. \tag{36}$$

The possibility has to be considered that, having accepted the hypotheses H_β and H_α on the strength of the values the F statistics under (34) and (35), we shall then discover that value of the statistic of (36) 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

The model of H_β can be elaborated by including the parameters γ_t which represent the temporal variation that is experienced by all J individuals. Assuming a common slope parameter, the system of equations as a whole is

$$Y^c = \mu\iota_{MT} + X\beta + (\iota_M \otimes I_T)\gamma + (I_M \otimes \iota_T)\delta + \mathcal{E}^c. \tag{38}$$

The matrix $[\iota_{MT}, X, \iota_M \otimes I_T, I_M \otimes \iota_T]$, containing the regressors is singular, because of the linear dependence between the columns of $[\iota_{MT}, \iota_M \otimes I_T, I_M \otimes \iota_T]$. This dependence is evident from the equation

$$(\iota_M \otimes I_T)(1 \otimes \iota_T) = (I_M \otimes \iota_T)(\iota_M \otimes 1) = \iota_{MT}. \tag{39}$$

Therefore, the parameters μ, β, γ, δ will not be estimable as a whole unless some restrictions are introduced. It is natural to impose the conditions that $\iota_T'\gamma = \sum_t \gamma_t = 0$ and that $\iota_{MT}'\delta = \sum_j \delta_j = 0$.

The parameters γ, δ, μ can be eliminated by premultiplying the equation (39) by the matrix

$$\begin{aligned} W &= [I_{MT} - (I_M \otimes P_T)][I_{MT} - (P_M \otimes I_M)] \\ &= I_{MT} - (I_M \otimes P_T) - (P_M \otimes I_M) + (I_M \otimes P_T)(P_M \otimes I_M). \end{aligned} \tag{40}$$

The two factors of W commute. The first factor $I_{MT} - (I_M \otimes P_T)$ annihilates the term $(I_M \otimes \iota_T)\delta$. The second factor $I_{MT} - (P_M \otimes I_M)$ annihilates $(\iota_M \otimes I_T)\gamma$.

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, the equation may, be transformed, with equal effect, by premultiplying it by Q' to obtain the system

$$Q'Y^c = Q'X\beta + \mathcal{E}^c. \tag{41}$$

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

$$\begin{aligned}\hat{\beta} &= (X'QQ'X)^{-1}X'QQ'y \\ &= (X'WX)^{-1}X'Wy.\end{aligned} \tag{42}$$

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. These effects, which are now part of the disturbance structure of the model, must be uncorrelated with the systematic part of its structure.

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

$$y = \mu\iota_{MT} + X\beta + (\iota_M \otimes I_T)\zeta + (I_M \otimes \iota_T)\eta + \varepsilon. \tag{44}$$

It is assumed that the random variables ζ_t, η_j and ε_{tj} are independently distributed with expectations of zero and with $V(\zeta_t) = \sigma^2_\zeta, V(\eta_j) = \sigma^2_\eta$ and $V(\varepsilon_{tj}) = \sigma^2_\varepsilon$. Then, the dispersion matrix of the vector of disturbances is

$$\Omega = \sigma^2_\zeta(\iota_M\iota'_M \otimes I_T) + \sigma^2_\eta(I_M \otimes \iota_T\iota'_T) + \sigma^2_\varepsilon I_{MT}. \tag{45}$$

A special case is when $\sigma^2_\zeta = 0$, which is when there is no intertemporal variation in the disturbances. Then

$$\begin{aligned} \Omega &= \sigma^2_\eta(I_M \otimes \iota_T\iota'_T) + \sigma^2_\varepsilon I_{MT} \\ &= I_M \otimes (\sigma^2_\varepsilon I_T + \sigma^2_\eta\iota_T\iota_T) = I_M \otimes V. \end{aligned} \tag{46}$$

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

$$V^{-1} = \frac{1}{\sigma_\varepsilon^2}\left(I_T - \frac{\sigma_\eta^2}{\sigma_\varepsilon^2 + T\sigma_\eta^2}\iota_T\iota_T'\right). \tag{47}$$

The inverse of the matrix Ω of (35) has a somewhat complicated structure. It takes the form of the form of

(48)

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

$$\begin{aligned} \text{where}\quad \lambda_1 &= \sigma_\zeta^2(\sigma_\varepsilon^2 - M\sigma_\zeta^2)^{-1}, \\ \lambda_2 &= \sigma_\eta^2(\sigma_\varepsilon^2 - T\sigma_\eta^2)^{-1}, \\ \lambda_3 &= \lambda_1\lambda_2(2\sigma_\varepsilon^2 + M\sigma_\zeta^2 + T\sigma_\eta^2)(\sigma_\varepsilon^2 + M\sigma_\zeta^2 + T\sigma_\eta^2)^{-1}. \end{aligned}$$

Feasible Least-Squares Estimator of the Random Effects Model

In order to realise the generalised least squares estimators of the random effect models, it is necessary to derive estimators of the variances σ_ζ^2, σ_η^2 and σ_ε^2 of the error components. For simplicity, we shall continue to assume that $\sigma_\zeta^2 = 0$. Then, the appropriate estimator can be derived from the the *within-groups* and between groups estimators associated with the fixed effects model of equation (9). The estimators are

$$\hat{\sigma}_\varepsilon^2 = \frac{(y - X\hat{\beta}_W)'(y - X\hat{\beta}_W)}{MT - (K + M)},$$

(49)
$$\hat{\sigma}_B^2 = \frac{(y - X\hat{\beta}_B)'(y - X\hat{\beta}_B)}{M - K} \quad \text{and}$$

$$\hat{\sigma}_\eta^2 = \hat{\sigma}_B^2 - \frac{\hat{\sigma}_\varepsilon^2}{T}.$$