A UNIVARIATE FORMULATION OF THE MULTIVARIATE LINEAR MODEL

BU-554-M

by

S. R. Searle

April, 1975

Abstract

Both Goldberger [1964, p. 246] and Zellner [1962] present a univariate formulation of the multivariate linear model, but they deal only with estimation and do not consider hypothesis testing.

The Model

We consider a matrix $\underline{Y}_{N\times p}$ of N observations on each of p random variables, for which the expected value is

$$E(\underline{Y}_{N\times p}) = \underline{X}_{N\times q}\underline{B}_{q\times p} \ . \tag{1}$$

Define

$$\underline{y}_{Np\times 1} = \begin{bmatrix} \underline{y}_1 \\ \vdots \\ \underline{y}_p \end{bmatrix} \quad \text{and} \quad \underline{b}_{qp\times 1} = \begin{bmatrix} \underline{b}_1 \\ \vdots \\ \underline{b}_p \end{bmatrix} \tag{2}$$

where $\underline{y}$ is the $Np\times 1$ vector of the p columns $\underline{y}_j$ of $\underline{Y}$ written one under the other, and $\underline{b}$ is the $qp\times 1$ vector of the p columns of $\underline{B}$ written in the same manner. Then the familiar multivariate linear model (1) is equivalent to

$$E(\underline{y}) = \begin{bmatrix} \underline{X} & & \underline{0} \\ & \underline{X} & \\ & & \ddots & \\ \underline{0} & & & \underline{X} \end{bmatrix} \underline{b} \ . \tag{3}$$

Paper No. BU-554-M in the Biometrics Unit Mimeo Series, Department of Plant Breeding and Biometry, Cornell University, Ithaca, New York 14853.

Using Kronecker multiplication of matrices to write

$$\begin{bmatrix} \underline{X} & & & \\ & \underline{X} & & \\ & & \ddots & \\ & & & \underline{X} \end{bmatrix} = \underline{I} \otimes \underline{X}$$

gives (3) as

$$E(\underline{y}) = (\underline{I} \otimes \underline{X})\underline{\beta} \ . \tag{4}$$

The model (1) customarily includes the property that for $\underline{Y} = \{\underline{y}_j\}$ $j = 1, \cdots, p$ the covariance structure is

$$\mathrm{cov}(\underline{y}_i \underline{y}_j') = \sigma_{ij}\underline{I}_N, \quad \text{for } i,j = 1, \cdots, p. \tag{5}$$

Hence for $\underline{y}$ of (3) and (4)

$$\mathrm{var}(\underline{y}) = \begin{bmatrix} \sigma_{11}\underline{I}_N & \cdots & \sigma_{1p}\underline{I}_N \\ \vdots & & \vdots \\ \sigma_{p1}\underline{I}_N & \cdots & \sigma_{pp}\underline{I}_N \end{bmatrix} = \underline{V} \otimes \underline{I}_N \tag{6}$$

where $\underline{V} = \{\sigma_{ij}\}$ for $i,j = 1, \cdots, p$. Through (4) and (6) we now have a univariate formulation of the multivariate linear model.

Estimation

Generalized least squares estimation of $\underline{\beta}$, based on (4) and (6) gives

$$(\underline{I}_p \otimes \underline{X})'(\underline{V} \otimes \underline{I}_N)^{-1}(\underline{I}_p \otimes \underline{X})\hat{\underline{\beta}} = (\underline{I} \otimes \underline{X})'(\underline{V} \otimes \underline{I}_N)^{-1}\underline{y} \ . \tag{7}$$

Recalling the following properties of Kronecker products,

$$(\underline{A} \otimes \underline{B})' = \underline{A}' \otimes \underline{B}'; \quad (\underline{A} \otimes \underline{B})(\underline{P} \otimes \underline{Q}) = \underline{AP} \otimes \underline{BQ}; \quad \text{and } (\underline{A} \otimes \underline{B})^{-1} = \underline{A}^{-1} \otimes \underline{B}^{-1} \ ,$$

this last only when all the inverses exist, we have from (7)

$$(\underset{\sim}{I}_p \otimes \underset{\sim}{X}')(\underset{\sim}{V}^{-1} \otimes \underset{\sim}{I}_N)(\underset{\sim}{I}_p \otimes \underset{\sim}{X})\hat{\underset{\sim}{\ell}} = (\underset{\sim}{I} \otimes \underset{\sim}{X}')(\underset{\sim}{V}^{-1} \otimes \underset{\sim}{I}_N)\underset{\sim}{y}$$

$$\therefore \quad (\underset{\sim}{V}^{-1} \otimes \underset{\sim}{X}'\underset{\sim}{X})\hat{\underset{\sim}{\ell}} = (\underset{\sim}{V}^{-1} \otimes \underset{\sim}{X}')\underset{\sim}{y}$$

$$\hat{\underset{\sim}{\ell}} = [\underset{\sim}{V} \otimes (\underset{\sim}{X}'\underset{\sim}{X})^{-1}](\underset{\sim}{V}^{-1} \otimes \underset{\sim}{X}')\underset{\sim}{y} \; ,$$

and thus

$$\hat{\underset{\sim}{\ell}} = [\underset{\sim}{I} \otimes (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}']\underset{\sim}{y} \; . \qquad (8)$$

Hence from (2)

$$\hat{\underset{\sim}{b}}_i = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{y}_i \quad \text{for } i = 1, \cdots, p, \qquad (9)$$

and so

$$\underset{\sim}{B} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Y} \; , \qquad (10)$$

as expected.

The sampling variance of $\hat{\underset{\sim}{\ell}}$ from (8) is

$$\begin{aligned} \text{var}(\hat{\underset{\sim}{\ell}}) &= \text{var}\{[\underset{\sim}{I} \otimes (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}']\underset{\sim}{y}\} \\ &= [\underset{\sim}{I} \otimes (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'](\underset{\sim}{V} \otimes \underset{\sim}{I}_N)[I \otimes (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}']' \\ &= \underset{\sim}{V} \otimes (\underset{\sim}{X}'\underset{\sim}{X})^{-1} \end{aligned} \qquad (11)$$

so that

$$\text{cov}(\hat{\underset{\sim}{b}}_i, \hat{\underset{\sim}{b}}_j') = \sigma_{ij}(\underset{\sim}{X}'\underset{\sim}{X})^{-1} \; . \qquad (12)$$

From (9)

$$\underset{\sim}{y}_i - \underset{\sim}{X}\hat{\underset{\sim}{b}}_i = [\underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}']\underset{\sim}{y}_i \qquad (13)$$

with expected value zero. Hence

$$E[(y_i - X\hat{b}_i)'(y_j - X\hat{b}_j)] = E\{y_i'[I - X(X'X)^{-1}X']^2 y_j\}$$

$$= tr\{[I - X(X'X)^{-1}X']E(y_j y_i')\}$$

$$= tr\{[I - X(X'X)^{-1}X'][Xb_i b_j'X' + \sigma_{ij} I_N]\}$$

$$= \sigma_{ij}(N - q).$$

Hence an unbiased estimator of σ_{ij} is

$$\hat{\sigma}_{ij} = y_i'[I - X(X'X)^{-1}X']y_j/(N - q) \tag{14}$$

and correspondingly that for V is

$$\hat{V} = \frac{1}{N - q} Y'[I - X(X'X)^{-1}X']Y \; . \tag{15}$$

Independence under normality

Let e_i' denote the i^{th} row of I_N, and similarly write

$$E_i = [0 \cdots 0 \;\; I_N \;\; 0 \cdots 0]_{N \times Np}$$

a partitioned matrix which is null except for the i^{th} $N \times N$ submatrix being I_N. Then (14) is

$$(N - q)\hat{\sigma}_{ij} = y'E_i[I - X(X'X)^{-1}X']E_j'y \; . \tag{16}$$

Now assume normality,

$$y \sim N[(I \otimes X)\ell, \;\; V \otimes I_N], \tag{17}$$

and use the theorem (e.g., Searle [1971], p. 59) that when $x \sim N(\mu, W)$, then Kx and $x'Ax$ are independent if and only if $KWA = 0$. Applying this to (8) and (16), using (17), we find that

$$[\underline{I} \otimes (\underline{X}'\underline{X})^{-1}\underline{X}'](\underline{V} \otimes \underline{I})\underline{E}_i[\underline{I} - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}']\underline{E}_j$$

$$= [\underline{V} \otimes (\underline{X}'\underline{X})^{-1}\underline{X}'] \left[\begin{array}{l} \text{A matrix that is null except} \\ \text{for } \underline{I} - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}' \text{ as its } ij^{th} \text{ submatrix} \end{array}\right]$$

$$= \underline{0}.$$

Hence $\hat{\underline{\ell}}$ and $\hat{\sigma}_{ij}$ are independent; and so therefore are $\hat{\underline{\ell}}$ and $\hat{\underline{V}}$.

Hypothesis testing

When, in univariate analysis $\underline{y} \sim N(\underline{X}\underline{b}, \sigma^2\underline{I})$, the F-statistic for testing the hypothesis $H: \underline{K}'\underline{b} = \underline{m}$ is, for $\underline{K}'$ being of full row rank q,

$$F = Q/q\hat{\sigma}^2$$

with

$$Q = (\underline{K}'\hat{\underline{b}} - \underline{m})'[\underline{K}'(\underline{X}'\underline{X})^{-1}\underline{K}]^{-1}(\underline{K}'\hat{\underline{b}} - \underline{m})$$

and

$$\hat{\sigma}^2 = \underline{y}'[\underline{I} - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}']\underline{y}/[N - r(\underline{X})],$$

$\underline{X}$ being of full column rank. Adapted to our situation here Q is

$$Q = (\underline{K}'\hat{\underline{\ell}} - \underline{m})'\{\underline{K}'[\underline{V} \otimes (\underline{X}'\underline{X})^{-1}]\underline{K}\}^{-1}(\underline{K}'\hat{\underline{\ell}} - \underline{m}). \tag{18}$$

Further development of Q in (18) depends on the form of $\underline{K}'$. We consider the hypothesis discussed in Anderson [1956] where, on partitioning $\underline{B}$ as

$$\underline{B} = \begin{bmatrix} \underline{B}_1 \\ \underline{B}_2 \end{bmatrix} \quad \begin{array}{l} \underline{B}_1 \text{ of order } q_1 \times p \\ \underline{B}_2 \text{ of order } q_2 \times p \end{array} \tag{19}$$

with $q_1 + q_2 = q$, we test

$$H: \underline{B}_1 = \underline{B}_{10} . \tag{20}$$

Writing the matrices of (20) as vectors in the manner of (2) the hypothesis is restated as

$$H: \underline{\beta}_1 = \underline{\beta}_{10} \tag{21}$$

where $\underline{\beta}_1$ of order $q_1 p \times 1$ has the p columns of $\underline{B}_1$ written one under the other; and similarly for the columns of $\underline{B}_{10}$ in $\underline{\beta}_{10}$. It is then not difficult to see that (21) can be expressed as

$$H: \underline{K}'\underline{\beta} = \underline{\beta}_{10} \tag{22}$$

for

$$(\underline{K}')_{q_1 p \times qp} = \underline{I}_p \otimes [\underline{I}_{q_1} \quad \underline{0}_{q_1 \times q_2}]. \tag{23}$$

Example: For

$$\underline{B}_1 = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \qquad \underline{B}_{10} = \begin{bmatrix} b_{110} & b_{120} \\ b_{210} & b_{220} \end{bmatrix} \qquad \underline{B}_2 = \begin{bmatrix} b_{31} & b_{32} \\ b_{41} & b_{42} \\ b_{51} & b_{52} \end{bmatrix}$$

the hypothesis $H: \underline{B}_1 = \underline{B}_{10}$ is,

$$\left[\begin{array}{ccccc|ccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{array}\right] \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \\ b_{41} \\ b_{51} \\ b_{12} \\ b_{22} \\ b_{32} \\ b_{42} \\ b_{52} \end{bmatrix} = \begin{bmatrix} b_{110} \\ b_{210} \\ b_{120} \\ b_{220} \end{bmatrix}.$$

To substitute (23) into (18) define

$$\underline{L} = [\underline{I}_{q_1} \quad \underline{0}_{q_1 \times q}] \tag{24}$$

so that

$$\underline{K}' = \underline{I} \otimes \underline{L} .$$

Then in (18)

$$\underline{K}'[\underline{V} \otimes (\underline{X}'\underline{X})^{-1}]\underline{K} = \underline{V} \otimes \underline{L}(\underline{X}'\underline{X})^{-1}\underline{L}' . \tag{25}$$

Now partition $\underline{X} = [\underline{X}_1 \quad \underline{X}_2]$ conformable with the partitioning in (19) so that on also using (24) in (25) we get

$$\underline{K}'[\underline{V} \otimes (\underline{X}'\underline{X})^{-1}]\underline{K} = \underline{V} \otimes [\underline{I} \quad \underline{0}] \begin{bmatrix} \underline{X}_1'\underline{X}_1 & \underline{X}_1'\underline{X}_2 \\ \underline{X}_2'\underline{X}_1 & \underline{X}_2'\underline{X}_2 \end{bmatrix}^{-1} \begin{bmatrix} \underline{I} \\ \underline{0} \end{bmatrix}$$

$$= \underline{V} \otimes \underline{P}_{11\cdot 2}^{-1} \tag{26}$$

where $\underline{P}_{11\cdot 2}$ is, from the inverse of a partitioned matrix, defined as

$$\underline{P}_{11\cdot 2} = \underline{X}_1'\underline{X}_1 - \underline{X}_1'\underline{X}_2(\underline{X}_2'\underline{X}_2)^{-1}\underline{X}_2'\underline{X}_1 . \tag{27}$$

Hence (26) and (22) used in (18) gives

$$Q = (\underline{K}'\hat{\underline{\ell}} - \underline{\ell}_{10})'(\underline{V} \otimes \underline{P}_{11\cdot 2}^{-1})^{-1}(\underline{K}'\hat{\underline{\ell}} - \underline{\ell}_{10})$$

$$= (\hat{\underline{\ell}}_1 - \underline{\ell}_{10})'(\underline{V}^{-1} \otimes \underline{P}_{11\cdot 2})(\hat{\underline{\ell}}_1 - \underline{\ell}_{10}). \tag{28}$$

By general linear model theory, for $\underline{V}$ known, this has a distribution proportional to a χ^2. But it contains $\underline{V}^{-1}$ and cannot be so used when $\underline{V}$ is unknown, as is usually the case. In the univariate case $\underline{V}^{-1} = 1/\sigma^2$ and we estimate σ^2 by $\hat{\sigma}^2$, for which $\hat{\sigma}^2/\sigma^2$ has a χ^2-distribution independent of that of Q; then the σ^2's cancel in $Q/(q\hat{\sigma}^2/\sigma^2)$ and we are left with $Q/q\hat{\sigma}^2$ as an F-statistic. But in (28) no such cancelling of $\underset{\sim}{V}^{-1}$ can occur. However, suppose we replace $\underset{\sim}{V}^{-1}$ by $\hat{\underset{\sim}{V}}^{-1}$ and consider

$$\hat{Q} = (\hat{\underset{\sim}{\ell}}_1 - \underset{\sim}{\ell}_{10})'(\hat{\underset{\sim}{V}}^{-1} \otimes \underset{\sim}{P}_{11\cdot 2})(\hat{\underset{\sim}{\ell}}_1 - \underset{\sim}{\ell}_{10}). \tag{29}$$

Consider the distributional properties of terms in (29). First, it is not difficult to show that under $H: \underset{\sim}{\ell}_1 = \underset{\sim}{\ell}_{10}$

$$\hat{\underset{\sim}{\ell}}_1 - \underset{\sim}{\ell}_{10} \sim N(\underset{\sim}{0}, \underset{\sim}{V} \otimes \underset{\sim}{P}^{-1}_{11\cdot 2}). \tag{30}$$

Second, $\hat{\underset{\sim}{\ell}}_1$ and $\hat{\underset{\sim}{V}}$ are independent as has already been proved; and third, from the form in (15) we know that $(N - q)\hat{\underset{\sim}{V}} = \underset{\sim}{S}$, say, where $\underset{\sim}{S} \sim W(\underset{\sim}{V}, p, N - q)$ is a Wishart distribution with parameter $\underset{\sim}{V}$, p variables and $N - q$ degrees of freedom. Furthermore, in (29), because $\underset{\sim}{X}'\underset{\sim}{X}$ has full rank, $\underset{\sim}{P}_{11\cdot 2}$ of (27) is positive definite; let $\underset{\sim}{P}_{11\cdot 2} = \underset{\sim}{M}\underset{\sim}{M}'$ say. Then from (29)

$$(N - q)\hat{Q} = (\hat{\underset{\sim}{\ell}}_1 - \underset{\sim}{\ell}_{10})'(\underset{\sim}{S}^{-1} \otimes \underset{\sim}{M}\underset{\sim}{M}')(\hat{\underset{\sim}{\ell}}_1 - \underset{\sim}{\ell}_{10}). \tag{31}$$

Now define

$$(\hat{\underset{\sim}{\ell}}_1 - \underset{\sim}{\ell}_{10})' = [\underset{\sim}{w}_1' \cdots \underset{\sim}{w}_p'] \tag{32}$$

for $\underset{\sim}{w}_i'$ of order $1 \times q_1$, for $i = 1, \cdots, p$. Then with $\underset{\sim}{S}^{-1} = \{s^{ij}\}$ for $i,j = 1,\cdots,p$, (31) is

$$(N - q)\hat{Q} = [\underset{\sim}{w}_1' \; \underset{\sim}{w}_2' \cdots \underset{\sim}{w}_p'] \begin{bmatrix} s^{11}\underset{\sim}{M}\underset{\sim}{M}' & \cdots & s^{1p}\underset{\sim}{M}\underset{\sim}{M}' \\ \vdots & & \\ s^{p1}\underset{\sim}{M}\underset{\sim}{M}' & \cdots & s^{pp}\underset{\sim}{M}\underset{\sim}{M}' \end{bmatrix} \begin{bmatrix} \underset{\sim}{w}_1 \\ \underset{\sim}{w}_2 \\ \vdots \\ \underset{\sim}{w}_p \end{bmatrix}$$

$$= \sum_{i=1}^{p} \sum_{j=1}^{p} \underline{w}_i' s^{ij} \underline{M}\underline{M}' \underline{w}_j$$

$$= \sum_{i=1}^{p} \sum_{j=1}^{p} s^{ij} \underline{u}_i' \underline{u}_j \quad \text{for } \underline{u}_i = \underline{M}' \underline{w}_i \tag{33}$$

$$= \operatorname{tr}[S^{-1}\{\underline{u}_i' \underline{u}_j\}]$$

$$= \operatorname{tr}(\underline{S}^{-1} \underline{U}' \underline{U}) \quad \text{for } \underline{U} = \{\underline{u}_j\} \quad j = 1, \cdots, p \, . \tag{34}$$

Now from (32) and (30), we have in (33) $\underline{u}_i \sim N(\underline{0}, \sigma_{ii}\underline{M}'\underline{P}_{11\cdot 2}^{-1}\underline{M}) \sim N(\underline{0}, \sigma_{ii}\underline{I}_{q_1})$ because $\underline{P}_{11\cdot 2} = \underline{M}\underline{M}'$. Hence in (34), $\underline{U}'\underline{U}$ is a Wishart matrix; and with $\underline{u}_i$ of (33) being, through $\underline{w}_i$ of (32), a linear combination of elements of $\hat{\underline{b}}_1 - \underline{b}_{10}$, we have $\underline{U}'\underline{U}$ being independent of $\underline{S}$. Hence

$$\Lambda = \frac{|\underline{S}|}{|\underline{S} + \underline{U}'\underline{U}|} = \prod_i \frac{1}{1 + \lambda_i}$$

is a Wilks' Λ-statistic, with the λ_i being latent roots of $\underline{S}^{-1}\underline{U}'\underline{U}$. Thus (34) is

$$(N - q)\hat{Q} = \operatorname{tr}(\underline{S}^{-1}\underline{U}'\underline{U}) = \sum_i \lambda_i \, . \tag{35}$$

Suppose r_i^2 represents the square of a canonical correlation between the variables represented by the independent Wishart matrices $\underline{S}$ and $\underline{U}'\underline{U}$. Then

$$|-r_i^2(\underline{S} + \underline{U}'\underline{U}) + \underline{U}'\underline{U}| = 0 \, . \tag{36}$$

But by the definition of λ_i

$$|\underline{S}^{-1}\underline{U}'\underline{U} - \lambda_i \underline{I}| = 0 \, .$$

This is equivalent to

$$|\underline{U}'\underline{U} - \lambda_i \underline{S}| = 0,$$

i.e.,

$$|-\lambda_i(\underline{S} + \underline{U}'\underline{U}) + (1 + \lambda_i)\underline{U}'\underline{U}| = 0.$$

Comparison with (36) gives

$$r_i^2 = \lambda_i/(1 + \lambda_i)$$

and hence from (35)

$$(N - q)\hat{Q} = \sum_i \lambda_i = \sum_i r_i^2/(1 - r_i^2).$$

Thus $(N - q)\hat{Q}$ is distributed as Hotelling's generalized T_0^2 (see, for example, Kshirsagar [1972, p. 331]).

Acknowledgments

Thanks go to David M. Allen and Timothy D. Mount for assistance.

References

Anderson, T. W. [1956]. An Introduction to Multivariate Statistical Analysis. John Wiley and Sons, New York.

Goldberger, Arthur S. [1964]. Econometric Theory. John Wiley and Sons, New York.

Kshirsagar, A. M. [1972]. Multivariate Analysis. Marcell Dekker, New York.

Searle, S. R. [1971]. Linear Models. John Wiley and Sons, New York.

Zellner, Arnold [1962]. An efficient method of estimating unrelated regressions and tests for aggregation bias. J. of Am. Stat. Assoc. 57:338-342.