A UNIVARIATE FORMULATION OF THE MULTIVARIATE LINEAR MODEL

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 $Y_{N \times p}$ of $N$ observations on each of $p$ random variables, for which the expected value is

$$E(Y_{N \times p}) = X_{N \times q} B_{q \times p}.$$  

Define

$$
\begin{bmatrix}
Y_1 \\
\vdots \\
Y_p 
\end{bmatrix}
\quad \text{and} \quad
\begin{bmatrix}
b_1 \\
\vdots \\
b_p 
\end{bmatrix}
$$

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

$$E(y) = \begin{bmatrix} X & \mathbf{0} \\ \mathbf{0} & X \end{bmatrix}.$$  

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

\[
\begin{bmatrix}
X \\
X \\
\vdots \\
X
\end{bmatrix}
= I \otimes X
\]

gives (3) as

\[
E(y) = (I \otimes X) \hat{\beta} .
\] (4)

The model (1) customarily includes the property that for \( Y_j = [y_{ij}] \) \( j = 1, \ldots, p \)
the covariance structure is

\[
\text{cov}(y_iy_j') = \sigma_{ij}I_N, \quad \text{for } i, j = 1, \ldots, p.
\] (5)

Hence for \( y \) of (3) and (4)

\[
\text{var}(y) = 
\begin{bmatrix}
\sigma_{11}I_N & \cdots & \sigma_{1p}I_N \\
\vdots & \ddots & \vdots \\
\sigma_{p1}I_N & \cdots & \sigma_{pp}I_N
\end{bmatrix} = Y \otimes I_N
\] (6)

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

**Estimation**

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

\[
(1_p \otimes X)'(Y \otimes I_N)^{-1}(1_p \otimes X)\hat{\beta} = (I \otimes X)'(Y \otimes I_N)^{-1}y .
\] (7)

Recalling the following properties of Kronecker products,

\( (A \otimes B)' = A' \otimes B' \); \( (A \otimes B)(P \otimes Q) = AP \otimes BQ \); \quad \text{and} \quad (A \otimes B)^{-1} = A^{-1} \otimes B^{-1} \),
this last only when all the inverses exist, we have from (7)

\[(I_p \otimes X')(Y^{-1} \otimes I_N)(I_p \otimes X)\hat{\beta} = (I \otimes X')(Y^{-1} \otimes I_N)\beta\]

\[
\therefore (Y^{-1} \otimes X'X)\hat{\beta} = (Y^{-1} \otimes X')\beta
\]

and thus

\[
\hat{\beta} = [I \otimes (X'X)^{-1}](Y^{-1} \otimes X')\beta,
\]

Hence from (2)

\[
\hat{b}_1 = (X'X)^{-1}X'y_1 \quad \text{for } i = 1, \cdots, p, \quad (9)
\]

and so

\[
\beta = (X'X)^{-1}X'y,
\]

as expected.

The sampling variance of \(\hat{\beta}\) from (8) is

\[
\text{var}(\hat{\beta}) = \text{var}\left([I \otimes (X'X)^{-1}X']\beta\right)
\]

\[
= [I \otimes (X'X)^{-1}X'](Y \otimes I_N)[I \otimes (X'X)^{-1}X']'
\]

\[
= Y \otimes (X'X)^{-1}
\]

so that

\[
\text{cov}(\hat{b}_1, \hat{b}_1) = \sigma_{11}(X'X)^{-1}.
\]

From (9)

\[
y_i - \hat{X}\hat{b}_1 = [I - X(X'X)^{-1}X']y_i
\]

with expected value zero. Hence
\[
E[(\mathbf{y}_1 - \mathbf{\hat{x}}_1)^T(\mathbf{y}_3 - \mathbf{\hat{x}}_3)] = E[\mathbf{y}_1[I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']^T\mathbf{y}_3]
\]
\[
= \text{tr}[[I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']E(\mathbf{y}_3\mathbf{y}_1^T)]
\]
\[
= \text{tr}[[I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'][\mathbf{X}_1\mathbf{\hat{x}}_1\mathbf{X}' + \sigma_{ij}\mathbf{I}_N]]
\]
\[
= \sigma_{ij}(N - q).
\]

Hence an unbiased estimator of \(\sigma_{ij}\) is
\[
\hat{\sigma}_{ij} = \mathbf{y}_1[I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}_3/(N - q) \tag{14}
\]
and correspondingly that for \(\mathbf{y}\) is
\[
\hat{\mathbf{y}} = \frac{1}{N - q} \mathbf{y}'[I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y} \tag{15}
\]

**Independence under normality**

Let \(e_i\) denote the \(i\)th row of \(\frac{1}{\sqrt{N}}\), and similarly write
\[
\mathbf{E}_i = \begin{bmatrix} 0 & \cdots & 0 & \mathbf{I}_N & 0 & \cdots & 0 \end{bmatrix}_{N\times N_p}
\]
a partitioned matrix which is null except for the \(i\)th \(N\times N\) submatrix being \(\frac{1}{\sqrt{N}}\).

Then (14) is
\[
(N - q)\hat{\sigma}_{ij} = \mathbf{y}'\mathbf{E}_i[I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{E}_j\mathbf{y} \tag{16}
\]

Now assume normality,
\[
\mathbf{y} \sim \mathcal{N}(\mathbf{I} \otimes \mathbf{X})\mathbf{y}, \quad (\mathbf{V} \otimes \mathbf{I}_N) \tag{17}
\]
and use the theorem (e.g., Searle [1971], p. 59) that when \(\mathbf{x} \sim \mathcal{N}(\mu, \mathbf{V})\), then \(\mathbf{Vx}\) and \(\mathbf{x}'\mathbf{Ax}\) are independent if and only if \(\mathbf{Kx}\) contains \(\mathbf{x}'\mathbf{Ax}\) are independent if and only if \(\mathbf{Kx}\) is

Using (17), we find that
\[
[ I \otimes (X'X)^{-1}X'](V \otimes I)E_{j}[I - X(X'X)^{-1}X']E_{j}
\]

\[
= [V \otimes (X'X)^{-1}X'] \begin{bmatrix}
\text{A matrix that is null except for } I - X(X'X)^{1/2}X' \text{ as its } ij\text{th submatrix} \\
\end{bmatrix}
\]

\[
= 0.
\]

Hence \( \hat{\beta} \) and \( \hat{\sigma}_{ij} \) are independent; and so therefore are \( \hat{\beta} \) and \( \hat{\gamma} \).

**Hypothesis testing**

When, in univariate analysis \( Y \sim N(\mu, \sigma^2 I) \), the F-statistic for testing the hypothesis \( H: \beta = \mu \) is, for \( \beta \) being of full row rank \( q \),

\[
F = \frac{Q}{q\hat{\sigma}^2}
\]

with

\[
Q = (\beta - \mu)'[\beta'(X'X)^{-1}\beta]^{-1}(\beta - \mu)
\]

and

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

\( X \) being of full column rank. Adapted to our situation here \( Q \) is

\[
Q = (\beta - \mu)'[V \otimes (X'X)^{-1}X']^{-1}(\beta - \mu).
\]

Further development of \( Q \) in (18) depends on the form of \( \beta \). We consider the hypothesis discussed in Anderson [1956] where, on partitioning \( \beta \) as

\[
\beta = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}
\]

\[
= \begin{pmatrix} B_1 \text{ of order } q_1 \times p \\ B_2 \text{ of order } q_2 \times p \end{pmatrix}
\]

with \( q_1 + q_2 = q \), we test

\[
H: B_1 = B_{10}.
\]
Writing the matrices of (20) as vectors in the manner of (2) the hypothesis is restated as

\[ H : \mathbf{\lambda}_1 = \mathbf{\lambda}_{10} \]  

(21)

where \( \mathbf{\lambda}_1 \) of order \( q_1p \times 1 \) has the \( p \) columns of \( \mathbf{B}_1 \) written one under the other; and similarly for the columns of \( \mathbf{B}_{10} \) in \( \mathbf{\lambda}_{10} \). It is then not difficult to see that (21) can be expressed as

\[ H : \mathbf{K}' \mathbf{\lambda}_1 = \mathbf{\lambda}_{10} \]  

(22)

for

\[ (\mathbf{K}')_{q_1p \times q_0p} = \mathbf{I}_p \otimes \begin{bmatrix} \mathbf{I}_{q_1} & 0_{q_1 \times q_0} \end{bmatrix}. \]  

(23)

Example: For

\[
\mathbf{B}_1 = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}, \quad \mathbf{B}_{10} = \begin{bmatrix} b_{110} & b_{120} \\ b_{210} & b_{220} \end{bmatrix}, \quad \mathbf{B}_2 = \begin{bmatrix} b_{31} & b_{32} \\ b_{41} & b_{42} \\ b_{51} & b_{52} \end{bmatrix}
\]

the hypothesis \( H : \mathbf{B}_1 = \mathbf{B}_{10} \) is,

\[
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
\end{bmatrix}
\begin{bmatrix}
b_{11} \\
b_{21} \\
b_{31} \\
b_{41} \\
b_{51} \\
b_{12} \\
b_{22} \\
b_{32} \\
b_{42} \\
b_{52} \\
\end{bmatrix}
= \begin{bmatrix}
b_{110} \\
b_{210} \\
b_{120} \\
b_{220} \\
\end{bmatrix}.
\]
To substitute (23) into (18) define

$$\mathbf{L} = \begin{bmatrix} I_{q_1} & 0 \end{bmatrix}$$

so that

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

Then in (18)

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

(25)

Now partition \( \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{bmatrix} \) conformable with the partitioning in (19) so that on also using (24) in (25) we get

$$\mathbf{K}'[\mathbf{V} \otimes (\mathbf{X}'\mathbf{X})^{-1}]\mathbf{K} = \mathbf{V} \otimes \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{X}'\mathbf{X}_{11} & \mathbf{X}'\mathbf{X}_{12} \\ \mathbf{X}'\mathbf{X}_{21} & \mathbf{X}'\mathbf{X}_{22} \end{bmatrix}^{-1} \begin{bmatrix} I \\ 0 \end{bmatrix}$$

$$= \mathbf{V} \otimes \mathbf{P}_{11.2}^{-1}$$

(26)

where \( \mathbf{P}_{11.2} \) is, from the inverse of a partitioned matrix, defined as

$$\mathbf{P}_{11.2} = \mathbf{X}'\mathbf{X}_{11} - \mathbf{X}'\mathbf{X}_{12}(\mathbf{X}'\mathbf{X}_{22})^{-1}\mathbf{X}'\mathbf{X}_{21}.$$  

(27)

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

$$\mathbf{Q} = (\mathbf{K}'\hat{\beta} - \hat{\beta}_{10})'(\mathbf{V} \otimes \mathbf{P}_{11.2}^{-1})^{-1}(\mathbf{K}'\hat{\beta} - \hat{\beta}_{10})$$

$$= (\hat{\beta}_{1} - \hat{\beta}_{10})'(\mathbf{V}^{-1} \otimes \mathbf{P}_{11.2})(\hat{\beta}_{1} - \hat{\beta}_{10}).$$  

(28)

By general linear model theory, for \( \mathbf{V} \) known, this has a distribution proportional to a \( \chi^2 \). But it contains \( \mathbf{V}^{-1} \) and cannot be so used when \( \mathbf{V} \) is unknown, as is usually the case. In the univariate case \( \mathbf{V}^{-1} = 1/\sigma^2 \) and we estimate \( \sigma^2 \) by \( \hat{\sigma}^2 \), for which \( \hat{\sigma}^2/\sigma^2 \) has a \( \chi^2 \)-distribution independent of that of \( \mathbf{Q} \); then the \( \sigma^2 \)'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 \( \hat{\gamma}^{-1} \) can occur. However, suppose we replace \( \gamma^{-1} \) by \( \hat{\gamma}^{-1} \) and consider

\[
\hat{Q} = (\hat{x}_1 - \hat{x}_{10})'(\hat{\gamma}^{-1} \otimes P_{11.2})(\hat{x}_1 - \hat{x}_{10}).
\]  

(29)

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

\[
\hat{x}_1 - \hat{x}_{10} \sim N(\theta, \gamma \otimes \gamma_{11.2}).
\]  

(30)

Second, \( \hat{x}_1 \) and \( \hat{\gamma} \) are independent as has already been proved; and third, from the form in (15) we know that \( (N - q)\hat{\gamma} = S \), say, where \( S \sim W(\gamma, p, N - q) \) is a Wishart distribution with parameter \( \gamma, p \) variables and \( N - q \) degrees of freedom. Furthermore, in (29), because \( \chi'\chi \) has full rank, \( P_{11.2} \) of (27) is positive definite; let \( P_{11.2} = \gamma_{MM}' \) say. Then from (29)

\[
(N - q)\hat{Q} = (\hat{x}_1 - \hat{x}_{10})'(S^{-1} \otimes \gamma_{MM}')(\hat{x}_1 - \hat{x}_{10}).
\]  

(31)

Now define

\[
(\hat{x}_1 - \hat{x}_{10})' = [\gamma_1' \cdots \gamma_p']
\]  

(32)

for \( \gamma_i \) of order \( 1 \times q_i \), for \( i = 1, \cdots, p \). Then with \( S^{-1} = \{s_{ij}\} \) for \( i, j = 1, \cdots, p \), (31) is

\[
(N - q)\hat{Q} = [\gamma_1' \gamma_2' \cdots \gamma_p']\begin{bmatrix}
  s_{11}MM' & \cdots & s_{1p}MM' \\
  \vdots & \ddots & \vdots \\
  s_{p1}MM' & \cdots & s_{pp}MM'
\end{bmatrix}\begin{bmatrix}
  \gamma_1 \\
  \gamma_2 \\
  \vdots \\
  \gamma_p
\end{bmatrix}
\]
\[ p = \sum_{i=1}^{p} \sum_{j=1}^{p} s_{i,j} \]  
\[ \sum_{i=1}^{p} \sum_{j=1}^{p} s_{i,j} = \text{for } u_1 = M'w_1 \]  
\[ = \text{tr}(S^{-1}(u_1'u_1')) \]  
\[ = \text{tr}(S^{-1}YY) \text{ for } Y = \{u_j\} \text{ } j = 1, \ldots, p. \]  

Now from (32) and (30), we have in (33) \( u_1 \sim N(Q, \sigma_{ii}M'^{-1}P^{-1}M) \sim N(Q, \sigma_{ii}I) \) because \( P_{11}P_{22} = MM'. \) Hence in (34), \( U'U \) is a Wishart matrix, and with \( u_1 \) of (33) being, through \( w_1 \) of (32), a linear combination of elements of \( \hat{s}_1 - \hat{s}_{10}, \) we have \( U'U \) being independent of \( S. \) Hence

\[ A = \frac{|S|}{|S + U'U|} = \prod_{i=1}^{1} \frac{1}{1 + \lambda_i} \]

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

\[ (N - q)Q = \text{tr}(S^{-1}YY) = \sum_{i} \lambda_i. \]  

Suppose \( r_{1}^{2} \) represents the square of a canonical correlation between the variables represented by the independent Wishart matrices \( S \) and \( U'U. \) Then

\[ |r_{1}^{2}(S + U'U) + U'U| = 0. \]  

But by the definition of \( \lambda_i \)

\[ |S^{-1}U'U - \lambda_i I| = 0. \]
This is equivalent to

\[ |y'y - \lambda_1 s| = 0, \]

i.e.,

\[ |-\lambda_1 (s + y'y) + (1 + \lambda_1) y'y| = 0. \]

Comparison with (36) gives

\[ r^2_1 = \lambda_1/(1 + \lambda_1) \]

and hence from (35)

\[ (N - q) \hat{Q} = \sum_1 \lambda_1 = \sum_1 r^2_1/(1 - r^2_1). \]

Thus \((N - q) \hat{Q}\) is distributed as Hotelling's generalized \(T^2\) (see, for example, Kshirsagar [1972, p. 331]).

Acknowledgments

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

References


