On the Bayesian Estimation of Multivariate Regression

Author(s): George C. Tiao and Arnold Zellner

Source: *Journal of the Royal Statistical Society. Series B (Methodological)*, 1964, Vol. 26, No. 2 (1964), pp. 277-285

Published by: Wiley for the Royal Statistical Society

Stable URL: https://www.jstor.org/stable/2984424

On the Bayesian Estimation of Multivariate Regression

By George C. Tiao and Arnold Zellner

University of Wisconsin

[Received January 1964. Revised April 1964]

Summary

In this paper, we use a Bayesian approach to analyse sets of regression equations with correlated error terms. In the case that the matrix of "independent variables" is the same for all equations, our model reduces to the traditional multivariate regression model. For this case, the marginal posterior distribution of the regression coefficient vector for any equation is shown to be of the multivariate-t form. Further, the variances and covariances of the error terms have an "inverted" Wishart distribution *a posteriori*. Some properties of this distribution are given. Finally, the joint posterior distribution of the regression coefficients in the general model is derived and discussed.

1. Introduction and Specification of the Model

In this paper we discuss some Bayesian estimation procedures for the parameters in the following m-equation multivariate regression model:

$$\mathbf{y}_\alpha = \mathbf{X}_\alpha \boldsymbol{\beta}_\alpha + \mathbf{u}_\alpha \quad (\alpha = 1, 2, \ldots, m), \tag{1.1}$$

where $\mathbf{y}_\alpha$ is a $T \times 1$ vector of observations, $\mathbf{X}_\alpha$ a $T \times k_\alpha$ matrix of fixed elements with rank k_α, $\boldsymbol{\beta}_\alpha$ a $k_\alpha \times 1$ vector of regression coefficients with $\boldsymbol{\beta}_\alpha \neq \boldsymbol{\beta}_l$ for $\alpha \neq l$ and $\mathbf{u}_\alpha$ a $T \times 1$ vector of random disturbances. It is assumed that the $\mathbf{u}_\alpha$'s are jointly normally distributed with zero means and covariance matrix $\boldsymbol{\Sigma} \otimes \mathbf{I}_T$ where $\boldsymbol{\Sigma} = \{\sigma_{\alpha l}\}$ is a $m \times m$ positive definite matrix, $\mathbf{I}_T$ a $T \times T$ identity matrix and $\otimes$ denotes the Kronecker product. Write $\mathbf{y}' = (\mathbf{y}_1', \ldots, \mathbf{y}_m')$, $\mathbf{u}' = (\mathbf{u}_1', \ldots, \mathbf{u}_m')$, $\boldsymbol{\beta}' = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_m')$ and $\mathbf{Z}$ as a $Tm \times q$ block-diagonal matrix, where

$$q = \sum_{\alpha=1}^{m} k_\alpha,$$

and

$$\mathbf{Z} = \begin{bmatrix} \mathbf{X}_1 & & \\ & \mathbf{X}_2 & \\ & & \ddots \\ & & & \mathbf{X}_m \end{bmatrix}.$$

We have for the joint likelihood function of $\boldsymbol{\Sigma}$ and $\boldsymbol{\beta}$:

$$\begin{aligned} l(\boldsymbol{\beta}, \boldsymbol{\Sigma} \mid \mathbf{y}) &\propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}T} \exp\{-\tfrac{1}{2}\mathbf{u}'\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T \mathbf{u}\} \\ &\propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}T} \exp\{-\tfrac{1}{2}(\mathbf{y} - \mathbf{Z}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T (\mathbf{y} - \mathbf{Z}\boldsymbol{\beta})\}. \end{aligned} \tag{1.2}$$

The model in (1.1) can be utilized in analysing correlated observations on individual units through time or space with each unit having its own set of "determining" variables. For example, $\mathbf{y}_\alpha$ might represent observations on the αth firm's investment for years $t = 1, \ldots, T$, and $\mathbf{X}_\alpha$ the $T \times k_\alpha$ observations on the k_α variables specific to the αth firm which determine its investment. We should also expect that in a particular year investment outlays of different firms will be correlated since they might be subject to similar random shocks. Our specification of the covariance matrix above incorporates allowance for this possibility. Or it may be that concurrent observations are made on m characteristics of a physical system with each characteristic having its own determining variables. Again the physical set-up may be such that the concurrent observations are correlated. On the other hand, there are many situations in which $\mathbf{X}_1 = \mathbf{X}_2 = \ldots = \mathbf{X}_m$. For example, experiments may be performed to analyse the variation of m characteristics with the same experimental situation for every characteristic on each run. Also, in econometrics, this case is encountered in the analysis of certain "reduced form" equation systems.

We shall be concerned mainly with the situation where $\mathbf{X}_1 = \mathbf{X}_2 = \ldots = \mathbf{X}_m = \mathbf{X}$ (so that $k_1 = \ldots = k_m = k$). In this case our model is then the traditional multivariate regression model. Estimation and testing procedures in the Neyman–Pearson theory are fully discussed in, for example, Anderson (1958). For the special case that $\boldsymbol{\beta}_\alpha$ is a single parameter and $\mathbf{X}$ is a $T \times 1$ vector of ones, the problem has been considered by Savage in "The Subjective Basis of Statistical Practice" (unpublished manuscript to be denoted hereafter by SBSP), and Geisser and Cornfield (1963) from the Bayesian point of view. In the more general situation where the $\mathbf{X}$'s are not assumed to be identical, some recent work on the problem within the Neyman–Pearson framework has been done by Zellner (1962, 1963). The main difficulty seems to be that the minimum variance Aitken estimator for $\boldsymbol{\beta}$ involves the unknown $\boldsymbol{\Sigma}$; and the estimators proposed are "optimal" only in the asymptotic sense.

In Section 2, we discuss prior and posterior distributions of the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$. Properties of the posterior distributions of these parameters for the traditional model are derived in Sections 3–6. In Section 7, we give some finite Bayesian results for the general model.

2. Prior and Posterior Distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$

For the prior distribution of $\boldsymbol{\beta}$ and the $\frac{1}{2}m(m+1)$ distinct elements of $\boldsymbol{\Sigma}$, we assume that the experimental situation is such that little is known about these parameters. Adopting the invariance theory due to Jeffreys (1961, p. 179), we take

$$p(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = p(\boldsymbol{\beta})\, p(\boldsymbol{\Sigma}), \tag{2.1}$$

with

$$p(\boldsymbol{\beta}) = \text{constant}, \tag{2.2}$$

$$p(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}(m+1)}. \tag{2.3}$$

In the special case $m = 1$, (2.3) reduces to

$$p(\sigma_{11}) \propto 1/\sigma_{11}, \tag{2.4}$$

which coincides with the usual assumption about the prior distribution of a scale parameter—see e.g. Savage *et al.* (1962), Box and Tiao (1962, 1964). It is also interesting to notice that if we denote $\sigma^{\alpha l}$ as the (α, l)th element of the inverse of $\boldsymbol{\Sigma}$, then the Jacobian of the transformation of the $\frac{1}{2}m(m+1)$ variables

$$(\sigma_{11}, \ldots, \sigma_{mm}, \sigma_{12}, \ldots, \sigma_{m(m-1)}) \quad \text{to} \quad (\sigma^{11}, \ldots, \sigma^{mm}, \sigma^{12}, \ldots, \sigma^{m(m-1)})$$

is

$$J = \left| \frac{\partial(\sigma_{11}, \sigma_{12}, \ldots, \sigma_{mm})}{\partial(\sigma^{11}, \sigma^{12}, \ldots, \sigma^{mm})} \right| = |\boldsymbol{\Sigma}|^{m+1}. \tag{2.5}$$

This result can be established using the analysis in Anderson (1958, p. 162 and pp. 348–349). Consequently the prior distribution of the $\frac{1}{2}m(m+1)$ distinct elements of $\boldsymbol{\Sigma}^{-1}$ is

$$p(\boldsymbol{\Sigma}^{-1}) \propto |\boldsymbol{\Sigma}^{-1}|^{-\frac{1}{2}(m+1)}, \tag{2.6}$$

which is the prior distribution used by Savage (SBSP), arrived at through a slightly different argument.

Utilizing the prior distribution in (2.1), (2.2) and (2.3) in conjunction with the likelihood function in (1.2), the posterior distribution of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ is

$$p(\boldsymbol{\beta}, \boldsymbol{\Sigma} \mid \mathbf{y}) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}(T+m+1)} \exp\{-\tfrac{1}{2}(\mathbf{y}-\mathbf{Z}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T(\mathbf{y}-\mathbf{Z}\boldsymbol{\beta})\}. \tag{2.7}$$

In what follows we discuss the properties of this distribution.

3. Posterior Distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ when $\mathbf{X}_1 = \ldots = \mathbf{X}_m = \mathbf{X}$

In the situation where the $\mathbf{X}$'s are identical, it is shown in Anderson (1958) that the statistics

$$\hat{\boldsymbol{\beta}}_\alpha = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_\alpha \quad (\alpha = 1, \ldots, m), \tag{3.1a}$$

$$s_{\alpha l} = (\mathbf{y}_\alpha - \mathbf{X}\hat{\boldsymbol{\beta}}_\alpha)'(\mathbf{y}_l - \mathbf{X}\hat{\boldsymbol{\beta}}_l) \quad (\alpha, l = 1, \ldots, m) \tag{3.1b}$$

are jointly sufficient for $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, and the likelihood function in (1.2) can be written

$$l(\boldsymbol{\beta}, \boldsymbol{\Sigma} \mid \mathbf{y}) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}T} \exp\{-\tfrac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\mathbf{S} - \tfrac{1}{2}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\boldsymbol{\Sigma}^{-1} \otimes \mathbf{X}'\mathbf{X}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\} \tag{3.2}$$

where $\hat{\boldsymbol{\beta}}' = (\hat{\boldsymbol{\beta}}_1', \ldots, \hat{\boldsymbol{\beta}}_m')$ and $\mathbf{S} = \{s_{\alpha l}\}$ is proportional to the sample covariance matrix. Using (3.2), the posterior distribution of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ in (2.7) can be expressed as

$$p(\boldsymbol{\beta}, \boldsymbol{\Sigma} \mid \mathbf{y}) = p(\boldsymbol{\beta} \mid \boldsymbol{\Sigma}, \mathbf{y})\, p(\boldsymbol{\Sigma} \mid \mathbf{y}), \tag{3.3}$$

with

$$p(\boldsymbol{\beta} \mid \boldsymbol{\Sigma}, \mathbf{y}) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}k} \exp\{-\tfrac{1}{2}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\boldsymbol{\Sigma}^{-1} \otimes \mathbf{X}'\mathbf{X}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\}, \tag{3.4}$$

and

$$p(\boldsymbol{\Sigma} \mid y) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}\nu} \exp(-\tfrac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\mathbf{S}) \tag{3.5}$$

with $\nu = T - k + m + 1$. It is seen that the conditional posterior distribution of $\boldsymbol{\beta}$ for given value of $\boldsymbol{\Sigma}$ is multivariate normal with mean $\hat{\boldsymbol{\beta}}$ and covariance matrix $\boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1}$. In particular, if interest centres only on $\boldsymbol{\beta}_\alpha$, its conditional distribution is multivariate normal,

$$p(\boldsymbol{\beta}_\alpha \mid \boldsymbol{\Sigma}, \mathbf{y}) \propto \sigma_{\alpha\alpha}^{-\frac{1}{2}k} \exp\{-(\boldsymbol{\beta}_\alpha - \hat{\boldsymbol{\beta}}_\alpha)'\mathbf{X}'\mathbf{X}(\boldsymbol{\beta}_\alpha - \hat{\boldsymbol{\beta}}_\alpha)/2\sigma_{\alpha\alpha}\}, \tag{3.6}$$

which depends only on $\sigma_{\alpha\alpha}$. The posterior distribution of $\boldsymbol{\Sigma}$ in (3.5) may be called an "inverted" Wishart distribution. In the following Section, we discuss some properties of this distribution.

4. THE "INVERTED" WISHART DISTRIBUTION

We show in this Section that the marginal posterior distribution of the elements of any principal minor matrix of $\mathbf{\Sigma}$ is also in an "inverted" Wishart form. From this result we then deduce the distribution of $\sigma_{\alpha\alpha}$ and that of the correlation coefficient $\rho_{\alpha l}$.

Without loss of generality, we now derive the marginal distribution of the elements of $\mathbf{\Sigma}_{11}$ where $\mathbf{\Sigma}_{11}$ is the $p \times p$ upper left-hand principal minor matrix of $\mathbf{\Sigma}\,(p<m)$. Denoting

$$\mathbf{\Sigma} = \begin{array}{cc} \begin{array}{cc} p & m-p \end{array} & \\ \left[\begin{array}{c|c} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \hline \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{array}\right] & \begin{array}{l} p \\ m-p \end{array} \end{array}$$

and remembering that $\mathbf{\Sigma}$ is assumed positive definite, we can express the determinant and the inverse of $\mathbf{\Sigma}$ as

$$|\mathbf{\Sigma}| = |\mathbf{\Sigma}_{11}||\mathbf{\Sigma}_{22.1}| \quad \text{where} \quad \mathbf{\Sigma}_{22.1} = \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12} \tag{4.1}$$

and

$$\mathbf{\Sigma}^{-1} = \left[\begin{array}{c|c} \mathbf{\Sigma}_{11}^{-1} + \mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22.1}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1} & -\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22.1}^{-1} \\ \hline -\mathbf{\Sigma}_{22.1}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1} & \mathbf{\Sigma}_{22.1}^{-1} \end{array}\right] = \left[\begin{array}{c|c} \mathbf{\Sigma}_{11}^{-1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] + \mathbf{W} \tag{4.2}$$

say. (Equation (4.2) may be verified by showing $\mathbf{\Sigma}^{-1}\mathbf{\Sigma} = \mathbf{I}$.) Thus, the distribution in (3.5) can be written

$$p(\mathbf{\Sigma}\,|\,\mathbf{y}) \propto |\mathbf{\Sigma}_{11}|^{-\frac{1}{2}\nu}|\mathbf{\Sigma}_{22.1}|^{-\frac{1}{2}\nu} \exp(-\tfrac{1}{2}\operatorname{tr}\mathbf{\Sigma}_{11}^{-1}\mathbf{S}_{11} - \tfrac{1}{2}\operatorname{tr}\mathbf{WS}) \tag{4.3}$$

where $\mathbf{S}_{11}$ is the corresponding $p \times p$ upper left-hand principal minor of $\mathbf{S}$. For fixed $\mathbf{\Sigma}_{11}$, consider the transformation

$$\begin{cases} \mathbf{Y} = \mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}, \\ \mathbf{\Omega} = \mathbf{\Sigma}_{22.1} = \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}. \end{cases}$$

It is easy to verify that the Jacobian of the transformation is

$$J = \left|\frac{\partial(\mathbf{\Sigma}_{12}, \mathbf{\Sigma}_{22})}{\partial(\mathbf{Y}, \mathbf{\Omega})}\right| = |\mathbf{\Sigma}_{11}|^{m-p}.$$

Consequently, we have that

$$p(\mathbf{\Sigma}_{11}, \mathbf{Y}, \mathbf{\Omega}\,|\,\mathbf{y}) \propto |\mathbf{\Sigma}_{11}|^{-\frac{1}{2}\{\nu-2(m-p)\}}|\mathbf{\Omega}|^{-\frac{1}{2}\nu} \exp(-\tfrac{1}{2}\operatorname{tr}\mathbf{\Sigma}_{11}^{-1}\mathbf{S}_{11} - \tfrac{1}{2}\operatorname{tr}\mathbf{WS}). \tag{4.4}$$

This implies that the marginal distribution of the $\frac{1}{2}p(p+1)$ elements of $\mathbf{\Sigma}_{11}$ is

$$p(\mathbf{\Sigma}_{11}\,|\,\mathbf{y}) \propto |\mathbf{\Sigma}_{11}|^{-\frac{1}{2}\{\nu-2(m-p)\}} \exp(-\tfrac{1}{2}\operatorname{tr}\mathbf{\Sigma}_{11}^{-1}\mathbf{S}_{11}). \tag{4.5}$$

In particular, if $p = 1$, the distribution of σ_{11} is

$$p(\sigma_{11}\,|\,\mathbf{y}) \propto \sigma_{11}^{-\frac{1}{2}\{\nu-2(m-1)\}} \exp(-s_{11}/2\sigma_{11}) \tag{4.6}$$

which is in the form of an "inverted" χ^2 distribution. It is of interest to compare the result in (4.6) with the posterior distribution of σ_{11} in the single equation regression (i.e. $m = 1$). The latter is given in Savage (SBSP) and can be obtained by setting $m = 1$ in (3.6) to yield:

$$p(\sigma_{11}\,|\,\mathbf{y}) \propto \sigma_{11}^{-\frac{1}{2}(T-k+2)} \exp(-s_{11}/2\sigma_{11}). \tag{4.7}$$

We see that as the value of m increases, the distribution in (4.6) becomes less and less concentrated about s_{11}. This is an intuitively pleasing result because when m increases, a larger and larger part of the information from the sample is utilized to estimate $\sigma_{12}, \sigma_{13}, \ldots, \sigma_{1m}$. In fact the exponent of $\sigma_{11}^{-\frac{1}{2}}$ in (4.6) differs from that in (4.7) by $m-1$. We may say, as is usually done, that "one degree of freedom is lost for each of the $m-1$ elements $\sigma_{12}, \ldots, \sigma_{1m}$". By setting $p = 2$ in the distribution in (4.5), we can then follow the development in Jeffreys (1961, p. 174) to obtain the posterior distribution of the correlation coefficient ρ_{12} as:

$$p(\rho_{12} \mid \mathbf{y}) \propto \frac{(1-\rho_{12}^2)^{\frac{1}{2}(n-3)}}{(1-\rho_{12} r_{12})^{n-\frac{1}{2}}} S_n(\rho_{12} r_{12}) \tag{4.8}$$

with

$$n = T - k - (m-2),$$

$$r_{12} = s_{12}/(s_{11} s_{22})^{\frac{1}{2}},$$

and

$$S_n(\rho_{12} r_{12}) = 1 + \sum_{l=1}^{\infty} \frac{1}{l!} \frac{1^2 . 3^2 \ldots (2l-1)^2}{(n+\frac{1}{2}) \ldots (n+l-\frac{1}{2})} \left(\frac{1+\rho_{12} r_{12}}{8}\right).$$

Except for the changes in the "degrees of freedom", this distribution is in the same form as the one given in Jeffreys for the case of sampling from a bivariate normal population.

5. POSTERIOR DISTRIBUTION OF $\boldsymbol{\beta}_1$

In many practical applications of the multivariate regression model, an investigator's main interest may be centred on the regression coefficients of a particular equation. As we have seen in (3.6), the conditional posterior distribution of $\boldsymbol{\beta}_1$ given $\boldsymbol{\Sigma}$ depends upon only σ_{11}. Thus from (3.6) and (4.6), we have for the marginal posterior distribution of $\boldsymbol{\beta}_1$,

$$\begin{aligned} p(\boldsymbol{\beta}_1 \mid \mathbf{y}) &= \int p(\sigma_{11} \mid \mathbf{y}) p(\boldsymbol{\beta}_1 \mid \boldsymbol{\Sigma}, \mathbf{y}) \, d\sigma_{11} \\ &\propto \{s_{11} + (\boldsymbol{\beta}_1 - \hat{\boldsymbol{\beta}}_1)' \mathbf{X}'\mathbf{X} (\boldsymbol{\beta}_1 - \hat{\boldsymbol{\beta}}_1)\}^{-\frac{1}{2}\{T-(m-1)\}}. \end{aligned} \tag{5.1}$$

Expression (5.1) is then in the form of a multivariate t-distribution—see e.g. Dunnett and Sobel (1954). As in the case of the posterior distribution of σ_{11}, if we set $m = 1$ the distribution reduces to that for a single equation regression, cf. e.g. Savage (SBSP). The only difference is then the change in the "degrees of freedom" due to the inclusion of the $m-1$ parameters $\sigma_{12}, \ldots, \sigma_{1m}$ in the model.

We may mention that in some economic applications of the model, there may be reasons to restrict the values of a subset of the parameters $\boldsymbol{\beta}_1$, say

$$\boldsymbol{\beta}_1' = (\beta_{11}, \ldots, \beta_{1r}, \beta_{1(r+1)}, \ldots, \beta_{1k})$$

where $\beta_{1(r+1)} = \ldots = \beta_{1k} = 0$. This is sometimes called a "restricted" equation. From the properties of the multivariate t-distribution, the conditional posterior distribution of $\beta_{11}, \ldots, \beta_{1r}$ given that $\beta_{1(r+1)} = \ldots = \beta_{1k} = 0$ is again of the multivariate t-form—for details see e.g. Raiffa and Schlaifer (1961, p. 258).

6. THE JOINT POSTERIOR DISTRIBUTION OF $\boldsymbol{\beta}$

In this Section, we give an alternative derivation of the posterior distribution of $\boldsymbol{\beta}_1$ by first finding the joint posterior distribution of $\boldsymbol{\beta}' = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_m')$. From the joint posterior distribution of $\boldsymbol{\Sigma}$ and $\boldsymbol{\beta}$ in (3.3) and the Jacobian in (2.5), we immediately deduce the distribution of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}^{-1}$ as

$$p(\boldsymbol{\beta}, \boldsymbol{\Sigma}^{-1} \mid \mathbf{y}) \propto |\boldsymbol{\Sigma}^{-1}|^{\frac{1}{2}\{T-(m+1)\}} \exp\{-\tfrac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1}(\mathbf{S}+\mathbf{B})\} \tag{6.1}$$

where $\mathbf{B} = \{b_{\alpha l}\}$ is a $m \times m$ matrix with

$$b_{\alpha l} = (\boldsymbol{\beta}_\alpha - \hat{\boldsymbol{\beta}}_\alpha)' \mathbf{X}' \mathbf{X} (\boldsymbol{\beta}_l - \hat{\boldsymbol{\beta}}_l). \tag{6.2}$$

For fixed $\boldsymbol{\beta}$, the distribution in (6.1) is in the Wishart form. From the properties of the Wishart distribution, integrating over the elements of $\boldsymbol{\Sigma}^{-1}$ yields the marginal posterior distribution of $\boldsymbol{\beta}$ as

$$p(\boldsymbol{\beta} \mid \mathbf{y}) \propto |\mathbf{S}+\mathbf{B}|^{-\frac{1}{2}T}. \tag{6.3}$$

In the special case of sampling from a m-variate multinormal distribution, (6.3) is the joint posterior distribution of the m means, as first derived by Savage (SBSP). It was subsequently shown by Geisser and Cornfield (1963) to be in the form of a multivariate t-distribution with covariance matrix proportional to S. Unfortunately, in the multivariate regression case considered here, it is not possible to extend the result by putting (6.3) in the multivariate-t form, even though we have seen from (5.1) that the marginal distribution of $\boldsymbol{\beta}_1$ is of this form. We now show, however, that if we express the joint distribution of $\boldsymbol{\beta}$ as the product

$$p(\boldsymbol{\beta} \mid \mathbf{y}) = p(\boldsymbol{\beta}_1 \mid \mathbf{y}) p(\boldsymbol{\beta}_2 \mid \boldsymbol{\beta}_1, \mathbf{y}) \ldots p(\boldsymbol{\beta}_m \mid \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{m-1}, \mathbf{y}), \tag{6.4}$$

then each of the factors on the right of (6.4) can be expressed in terms of a multivariate t-distribution. We first derive an expression for the product

$$p(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{m-1} \mid \mathbf{y}) p(\boldsymbol{\beta}_m \mid \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{m-1}, \mathbf{y}).$$

Denote the determinant $|\mathbf{S}+\mathbf{B}|$ in (6.3) as:

$$|\mathbf{S}+\mathbf{B}| = \left|\begin{array}{c:c} \bar{\mathbf{S}}+\bar{\mathbf{B}} & (\mathbf{s}+\mathbf{b}) \\ \hdashline (\mathbf{s}+\mathbf{b})' & s_{mm}+b_{mm} \end{array}\right| \tag{6.5}$$

where $\bar{\mathbf{S}}+\bar{\mathbf{B}}$ is the $(m-1)\times(m-1)$ upper left-hand principal minor matrix of $(\mathbf{S}+\mathbf{B})$ and

$$\mathbf{s}' = (s_{m1}, \ldots, s_{m(m-1)}) \quad \text{and} \quad \mathbf{b}' = (b_{m1}, \ldots, b_{m(m-1)}). \tag{6.6}$$

Expanding the determinant in (6.5), we obtain

$$|\mathbf{S}+\mathbf{B}| = |\bar{\mathbf{S}}+\bar{\mathbf{B}}| \{s_{mm} + b_{mm} - (\mathbf{s}+\mathbf{b})'(\bar{\mathbf{S}}+\bar{\mathbf{B}})^{-1}(\mathbf{s}+\mathbf{b})\}. \tag{6.7}$$

In the second factor on the right of (6.7), let us write

$$b_{\alpha l} = \boldsymbol{\gamma}_\alpha' \boldsymbol{\gamma}_l \quad \text{with} \quad \boldsymbol{\gamma}_l = \mathbf{X}(\boldsymbol{\beta}_l - \hat{\boldsymbol{\beta}}_l) \quad (\alpha = 1, \ldots, m;\ l = 1, \ldots, m), \tag{6.8}$$
$$\mathbf{b}' = \boldsymbol{\gamma}_m' \boldsymbol{\gamma} \text{ where } \boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_{m-1}).$$

After a little algebraic rearrangement, we have that

$$|\mathbf{S}+\mathbf{B}| = |\bar{\mathbf{S}}+\bar{\mathbf{B}}| \{c_m + (\boldsymbol{\gamma}_m - \boldsymbol{\mu})' \mathbf{D}_m (\boldsymbol{\gamma}_m - \boldsymbol{\mu})\} \tag{6.9}$$

where

$$\mathbf{D}_m = \mathbf{I} - \boldsymbol{\gamma}(\bar{\mathbf{S}}+\bar{\mathbf{B}})^{-1}\boldsymbol{\gamma}',$$

$$\boldsymbol{\mu} = \mathbf{D}_m^{-1}\boldsymbol{\gamma}(\bar{\mathbf{S}}+\bar{\mathbf{B}})^{-1}\mathbf{s},$$

and

$$c_m = s_{mm} - \mathbf{s}'(\bar{\mathbf{S}}+\bar{\mathbf{B}})^{-1}\mathbf{s} - \boldsymbol{\mu}'\mathbf{D}_m\boldsymbol{\mu}.$$

We now make use of a theorem due to Tocher (1951) which says that if $\mathbf{A}$ is a $m \times n$ matrix and $\mathbf{B}$ is a $n \times m$ matrix, then

$$(\mathbf{I}_m - \mathbf{AB})^{-1} = \mathbf{I}_m + \mathbf{A}(\mathbf{I}_n - \mathbf{BA})^{-1}\mathbf{B}. \tag{6.10}$$

Applying (6.10) and noting that $\boldsymbol{\gamma}'\boldsymbol{\gamma} = \bar{\mathbf{B}}$, we obtain

$$\begin{aligned} \mathbf{D}_m^{-1} &= \mathbf{I} + \boldsymbol{\gamma}\bar{\mathbf{S}}^{-1}\boldsymbol{\gamma}', \\ \boldsymbol{\mu} &= \boldsymbol{\gamma}\bar{\mathbf{S}}^{-1}\mathbf{s}, \\ c_m &= s_{mm} - \mathbf{s}'\bar{\mathbf{S}}^{-1}\mathbf{s}, \end{aligned} \tag{6.11}$$

and c_m is in fact the reciprocal of the (m,m)th element of $\mathbf{S}^{-1}$. In terms of $\boldsymbol{\beta}_m$, the second factor on the right of (6.9) is

$$\{c_m + (\boldsymbol{\gamma}_m - \boldsymbol{\mu})'\mathbf{D}_m(\boldsymbol{\gamma}_m - \boldsymbol{\mu})\} = \{c_m + (\boldsymbol{\beta}_m - \boldsymbol{\eta}_m)'\mathbf{X}'\mathbf{D}_m\mathbf{X}(\boldsymbol{\beta}_m - \boldsymbol{\eta}_m)\} \tag{6.12}$$

with

$$\boldsymbol{\eta}_m = \hat{\boldsymbol{\beta}}_m + \mathbf{d}\bar{\mathbf{S}}^{-1}\mathbf{s} \quad \text{and} \quad \mathbf{d} = (\boldsymbol{\beta}_1 - \hat{\boldsymbol{\beta}}_1, \ldots, \boldsymbol{\beta}_{m-1} - \hat{\boldsymbol{\beta}}_{m-1}).$$

Using (6.9) and (6.12), we can therefore write the distribution in (6.3) as

$$p(\boldsymbol{\beta}\,|\,\mathbf{y}) = |\bar{\mathbf{S}}+\bar{\mathbf{B}}|^{-\frac{1}{2}T}\{c_m + (\boldsymbol{\beta}_m - \boldsymbol{\eta}_m)'\mathbf{X}'\mathbf{D}_m\mathbf{X}(\boldsymbol{\beta}_m - \boldsymbol{\eta}_m)\}^{-\frac{1}{2}T}. \tag{6.13}$$

Now the determinant of the matrix $\mathbf{X}'\mathbf{D}_m\mathbf{X}$ is

$$\begin{aligned} |\mathbf{X}'\mathbf{D}_m\mathbf{X}| &= |\mathbf{X}'\mathbf{X}| \;\; |\mathbf{I} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\gamma}(\bar{\mathbf{S}}+\bar{\mathbf{B}})^{-1}\boldsymbol{\gamma}'\mathbf{X}| \\ &= |\mathbf{X}'\mathbf{X}| \;\; |\mathbf{I} - \mathbf{d}(\bar{\mathbf{S}}+\bar{\mathbf{B}})^{-1}\boldsymbol{\gamma}'\mathbf{X}|. \end{aligned} \tag{6.14}$$

We now make use of a theorem given in Bellman (1960, p. 95) which says if $\mathbf{A}$ and $\mathbf{B}$ are two $n \times n$ matrices then $|\mathbf{I}-\mathbf{AB}| = |\mathbf{I}-\mathbf{BA}|$. As pointed out by a referee, this result can be generalized to the case where $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{B}$ is an $n \times m$ matrix. Suppose $m < n$, then

$$\begin{aligned} |\mathbf{I}_m - \mathbf{AB}| &= \left|\mathbf{I}_n - \begin{bmatrix} \mathbf{A} \\ \hdashline \mathbf{0} \end{bmatrix} (\mathbf{B} \,\vdots\, \mathbf{0})\right| \\ &= \left|\mathbf{I}_n - (\mathbf{B} \,\vdots\, \mathbf{0}) \begin{bmatrix} \mathbf{A} \\ \hdashline \mathbf{0} \end{bmatrix}\right| = |\mathbf{I}_n - \mathbf{BA}|. \end{aligned}$$

Using this result, we can write the second determinant on the right of (6.14) as

$$|\mathbf{I} - \mathbf{d}(\bar{\mathbf{S}}+\bar{\mathbf{B}})^{-1}\boldsymbol{\gamma}'\mathbf{X}| = |\mathbf{I} - (\bar{\mathbf{S}}+\bar{\mathbf{B}})^{-1}\boldsymbol{\gamma}'\mathbf{X}\mathbf{d}|.$$

Hence,

$$\begin{aligned} |\mathbf{X}'\mathbf{D}_m\mathbf{X}| &= |\mathbf{X}'\mathbf{X}| \;\; |\mathbf{I} - (\bar{\mathbf{S}}+\bar{\mathbf{B}})^{-1}\boldsymbol{\gamma}'\boldsymbol{\gamma}| \\ &= |\mathbf{X}'\mathbf{X}| \;\; |\bar{\mathbf{S}}| \;\; |\bar{\mathbf{S}}+\bar{\mathbf{B}}|^{-1}. \end{aligned} \tag{6.15}$$

Consequently, we can write the distribution in (6.13) as

$$p(\boldsymbol{\beta}\,|\,\mathbf{y}) = p(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{m-1}\,|\,\mathbf{y})\,p(\boldsymbol{\beta}_m\,|\,\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{m-1}, \mathbf{y}) \tag{6.16}$$

with

$$p(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{m-1}\,|\,\mathbf{y}) \propto |\bar{\mathbf{S}}+\bar{\mathbf{B}}|^{-\frac{1}{2}(T-1)} \tag{6.17}$$

and

$$p(\boldsymbol{\beta}_m\,|\,\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{m-1}, \mathbf{y}) \propto |\mathbf{X}'\mathbf{D}_m\mathbf{X}|^{\frac{1}{2}}\{c_m+(\boldsymbol{\beta}_m-\boldsymbol{\eta}_m)'\,\mathbf{X}'\mathbf{D}_m\mathbf{X}(\boldsymbol{\beta}_m-\boldsymbol{\eta}_m)\}^{-\frac{1}{2}T}. \tag{6.18}$$

The conditional distribution of $\boldsymbol{\beta}_m$ can therefore be expressed in terms of a multivariate t-distribution, while the marginal distribution of $(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{m-1})$, is of the same form as the original distribution of $\boldsymbol{\beta}$ except of course for the changes in the dimensions of the matrix $\bar{\mathbf{S}}+\bar{\mathbf{B}}$ and the value of the exponent of the determinant. Repeating the same process $m-1$ times, we can then express the joint distribution as the product

$$\begin{aligned} p(\boldsymbol{\beta}\,|\,\mathbf{y}) \propto {} & \{s_{11}+(\boldsymbol{\beta}_1-\hat{\boldsymbol{\beta}}_1)'\,\mathbf{X}'\mathbf{X}(\boldsymbol{\beta}_1-\hat{\boldsymbol{\beta}}_1)\}^{-\frac{1}{2}\{T-(m-1)\}} \\ & \times \prod_{\alpha=2}^{m} |\mathbf{X}'\mathbf{D}_\alpha\mathbf{X}|^{\frac{1}{2}}\{c_\alpha+(\boldsymbol{\beta}_\alpha-\boldsymbol{\eta}_\alpha)'\,\mathbf{X}'\mathbf{D}_\alpha\mathbf{X}(\boldsymbol{\beta}_\alpha-\boldsymbol{\eta}_\alpha)\}^{-\frac{1}{2}(T-m+\alpha)} \end{aligned} \tag{6.19}$$

where $\mathbf{D}_\alpha$, $\boldsymbol{\eta}_\alpha$ and c_α are defined in exactly the same way as in the case $\alpha = m$ given in (6.18). The factors in (6.19) correspond precisely to the distributions set out in (6.4), and clearly the first factor is the marginal distribution of $\boldsymbol{\beta}_1$ as obtained earlier in (5.1). This is an interesting example showing that even though the conditional distribution and the marginal distribution of certain subsets of variables are of the multivariate-t form the joint distribution fails to be of the same form.

At the end of Section 5 we have discussed restrictions on coefficients of a single equation. In general, there may be restrictions on the coefficients appearing in several or all equations. The joint posterior distribution of the remaining unrestricted coefficients can readily be obtained by inserting the values of the restricted coefficients in (6.3). This result may be regarded as the solution of a special case of our more general model set out in (1.1). That is, suppose for a two-equation model we have

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{X}_1\,\boldsymbol{\beta}_1+\mathbf{u}_1, \\ \mathbf{y}_2 &= \mathbf{X}_2\,\boldsymbol{\beta}_2+\mathbf{u}_2. \end{aligned} \tag{6.20}$$

If we form the augmented matrix $\mathbf{X}$ such that

$$\mathbf{X} = [\mathbf{X}_c \,\vdots\, \mathbf{X}_{1d} \,\vdots\, \mathbf{X}_{2d}]$$

where

$$\begin{aligned} \mathbf{X}_1 &= [\mathbf{X}_c \,\vdots\, \mathbf{X}_{1d}], \\ \mathbf{X}_2 &= [\mathbf{X}_c \,\vdots\, \mathbf{X}_{2d}], \end{aligned}$$

then (6.20) can be written as

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{X}\boldsymbol{\beta}_1^*+\mathbf{u}_1, \\ \mathbf{y}_2 &= \mathbf{X}\boldsymbol{\beta}_2^*+\mathbf{u}_2, \end{aligned}$$

provided that particular subsets of $\boldsymbol{\beta}_1^*$ and $\boldsymbol{\beta}_2^*$ compatible with the partitioning of $\mathbf{X}$ have zero values. Note, however, that this approach for the general model requires that the augmented matrix $\mathbf{X}$ must be of full rank.

7. POSTERIOR DISTRIBUTION OF $\boldsymbol{\beta}$ FOR THE GENERAL MODEL

We now return to the analysis of the general model considered in Sections 1 and 2. From the joint posterior distribution of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ in (2.7), it is clear that the conditional distribution of $\boldsymbol{\beta}$ given $\boldsymbol{\Sigma}$ is normal with mean

$$\tilde{\boldsymbol{\beta}} = (\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I}_T\,\mathbf{Z})^{-1}\,\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I}_T\,\mathbf{y}, \tag{7.1}$$

and covariance matrix

$$\operatorname{cov}(\boldsymbol{\beta}) = (\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I}_T\,\mathbf{Z})^{-1}. \tag{7.2}$$

It is seen that $\tilde{\boldsymbol{\beta}}$, the centre of the conditional distribution, depends upon $\boldsymbol{\Sigma}$ and only in the case in which the $\mathbf{X}_\alpha$'s are identical (or proportional) will (7.1) reduce to (3.1a). As regards the marginal distributions of $\boldsymbol{\Sigma}$ and of $\boldsymbol{\beta}$, unfortunately because of the dependence of $\boldsymbol{\beta}$ on $\boldsymbol{\Sigma}$, the analysis in Sections 3–6 cannot be extended here. However, by an argument similar to that given in Section 6, the posterior distribution of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}^{-1}$ can be written

$$p(\boldsymbol{\beta},\boldsymbol{\Sigma}^{-1}\,|\,\mathbf{y}) \propto |\boldsymbol{\Sigma}^{-1}|^{\frac{1}{2}\{T-(m-1)\}}\exp(-\tfrac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\mathbf{U}), \tag{7.3}$$

where

$$\mathbf{U} = \{\mathbf{u}'_\alpha\,\mathbf{u}_l\} \quad (\alpha = 1,\ldots,m;\ l = 1,\ldots,m),$$

and

$$\mathbf{u}_\alpha = \mathbf{y}_\alpha - \mathbf{X}_\alpha\,\boldsymbol{\beta}_\alpha \quad (\alpha = 1,\ldots,m).$$

From (7.3), we obtain the marginal posterior distribution of $\boldsymbol{\beta}$ as

$$p(\boldsymbol{\beta}\,|\,\mathbf{y}) \propto |\mathbf{U}|^{-\frac{1}{2}T}.$$

Properties of this distribution are currently being investigated.

ACKNOWLEDGEMENT

The authors wish to thank a referee for many helpful comments which have allowed them to improve the presentation of the paper.

REFERENCES

ANDERSON, T. W. (1958), *An Introduction to Multivariate Statistical Analysis.* New York: Wiley.

BELLMAN, R. (1960), *Introduction to Matrix Analysis.* New York: McGraw-Hill.

BOX, G. E. P. and TIAO, G. C. (1962), "A further look at robustness via Bayes' theorem", *Biometrika*, **49**, 419–432.

—— —— (1964), "A Bayesian approach to the importance of assumptions applied to the comparison of variances", *Biometrika*, **51** (to appear).

DUNNETT, C. W. and SOBEL, M. (1954), "A bivariate generalization of Student's *t*-distribution, with tables for certain special cases", *Biometrika*, **41**, 153–169.

GEISSER, S. and CORNFIELD, J. (1963), "Posterior distributions for multivariate normal parameters", *J. R. statist. Soc.* B, **25**, 368–376.

JEFFREYS, H. (1961), *Theory of Probability*, 3rd ed. Oxford: Clarendon Press.

RAIFFA, H. and SCHLAIFER, R. (1961), *Applied Statistical Decision Theory.* Boston: Harvard University Graduate School of Business Administration.

SAVAGE, L. J. *et al.* (1962), *The Foundations of Statistical Inference.* London: Methuen.

TOCHER, K. D. (1951), "Discussion on Mr Box and Dr Wilson's paper", *J. R. statist. Soc.* B, **13**, 39–42.

ZELLNER, A. (1962), "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias", *J. Amer. statist. Ass.*, **57**, 348–368.

—— (1963), "Estimators for seemingly unrelated regression equations: some exact finite sample results", *J. Amer. statist. Ass.*, **58**, 977–992.