CLEARINGHOUSE FOR FEDERAL SCIENTIFIC AND TECHNICAL INFORMATION CFSTI
DOCUMENT MANAGEMENT BRANCH 410.11

LIMITATIONS IN REPRODUCTION QUALITY

ACCESSION # AD 604 260

- [x] 1. WE REGRET THAT LEGIBILITY OF THIS DOCUMENT IS IN PART UNSATISFACTORY. REPRODUCTION HAS BEEN MADE FROM BEST AVAILABLE COPY.
- [ ] 2. A PORTION OF THE ORIGINAL DOCUMENT CONTAINS FINE DETAIL WHICH MAY MAKE READING OF PHOTOCOPY DIFFICULT.
- [ ] 3. THE ORIGINAL DOCUMENT CONTAINS COLOR, BUT DISTRIBUTION COPIES ARE AVAILABLE IN BLACK-AND-WHITE REPRODUCTION ONLY.
- [ ] 4. THE INITIAL DISTRIBUTION COPIES CONTAIN COLOR WHICH WILL BE SHOWN IN BLACK-AND-WHITE WHEN IT IS NECESSARY TO REPRINT.
- [ ] 5. LIMITED SUPPLY ON HAND: WHEN EXHAUSTED, DOCUMENT WILL BE AVAILABLE IN MICROFICHE ONLY.
- [ ] 6. LIMITED SUPPLY ON HAND: WHEN EXHAUSTED DOCUMENT WILL NOT BE AVAILABLE.
- [ ] 7. DOCUMENT IS AVAILABLE IN MICROFICHE ONLY.
- [ ] 8. DOCUMENT AVAILABLE ON LOAN FROM CFSTI (TT DOCUMENTS ONLY).
- [ ] 9.

NBS 9/64

PROCESSOR: DL

(1)

DEPARTMENT OF STATISTICS

University of Wisconsin

Madison, Wisconsin

Technical Report No. 23

December, 1963

On the Bayesian Estimation of Multivariate Regression

by

George C. Tiao and Arnold Zellner

This research was supported in part by the National Science Foundation under Grant GS-151 and by the United States Navy through the Office of Naval Research, under Contract NONR-1202, Project NR 042 222. Reproduction in whole or in part is permitted for any purpose of the United States Government.

---

*Also issued as Systems Formulation and Methodology Workshop Paper No. 6315.

On the Bayesian Estimation of Multivariate Regression

by

George C. Tiao and Arnold Zellner

University of Wisconsin

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:

$$(1.1) \qquad y_\alpha = X_\alpha \beta_\alpha + u_\alpha$$

$$\alpha = 1, 2, \ldots, m.$$

where y_α is a $T \times 1$ vector of observations, X_α a $T \times k_\alpha$ matrix of fixed elements with rank k_α, β_α a $k_\alpha \times 1$ vector of regression coefficients and u_α a $T \times 1$ vector of random disturbances. It is assumed that the u_α's are jointly normally distributed with zero means and covariance matrix $\Sigma \times I_T$ where $\Sigma = \sigma_{\alpha l}$, is a $m \times m$ positive definite matrix and I_T a $T \times T$ identity matrix. By writing $y' = (y_1', \ldots, y_m')$, $u' = (u_1', \ldots, u_m')$, ($\beta = \beta_1', \ldots, \beta_m'$) and Z as the $Tm \times \underline{q}$ block-diagonal matrix, ($\underline{q} = \sum_{\alpha=1}^{m} k_\alpha$)

$$Z = \begin{bmatrix} X_1 & & & \\ & X_2 & & \\ & & \ddots & \\ & & & X_m \end{bmatrix}$$

We have for the jointly likelihood function of Σ and β :

$$(1.2) \qquad \ell(\beta, \Sigma \mid y) \propto |\Sigma|^{-\frac{T}{2}} \exp\left\{-\tfrac{1}{2} u' \Sigma^{-1} \times I_T u\right\}$$

$$\propto |\Sigma|^{-\frac{T}{2}} \exp\left\{-\tfrac{1}{2} (y - Z\beta)' \Sigma^{-1} \otimes I_T (y - Z\beta)\right\}$$

We shall be concerned mainly with the situation where $X_1 = X_2 = \ldots = X_m = X$ 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 $X\beta_\alpha$ is a vector of means, i.e. β_α is a single parameter and X is a $(T \times 1)$ vector of ones, the problem has been considered by Savage (1961) and Geisser and Cornfield (1963) from the Bayesian pcint of view. In the more general situation where the 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) and Telser (1963). The main difficulty seems to be that the minimum variance Aitken estimator for β involves the unknown Σ; and the estimators proposed by both these authors are "optimal" only in the asymptotic sense.

In Section 2, we discuss the prior and posterior distribution of the parameters β and Σ. 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 sample Bayesian results for the general model.

2. Prior and posterior distribution of β and Σ.

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

$$(2.1) \qquad p(\beta, \Sigma) = p(\beta)\, p(\Sigma)$$

with

$$(2.2) \qquad p(\beta) \propto k$$

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

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

$$(2.4) \qquad p(\sigma_{11}) \propto \frac{1}{\sigma_{11}}$$

which coincides with the usual assumption about the prior distribution of a scale parameter--see e.g. Savage (1962), Box and Tiao (1962, 1963). It is also interesting to notice that if we denote $\sigma^{\alpha\ell}$ as the (α, ℓ)th element of the inverse Σ^{-1}, then the Jacobian of the transformation of the $\frac{m(m+1)}{2}$ variables $(\sigma_{11}, \ldots, \sigma_{mm}, \sigma_{12}, \ldots, \sigma_{m(m-1)})$ to $(\sigma^{11}, \ldots, \sigma^{mm}, \sigma^{12}, \ldots, \sigma^{(m-1),m})$ is

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

Consequently the prior distribution of the $\frac{m(m+1)}{2}$ distinct elements of Σ^{-1} is

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

which is the prior distribution used by Savage (1961), arrived at through a slightly different argument.

Utilizing the prior distribution in (2.2) and (2.3) in conjunction with likelihood function in (1.2) the posterior distribution of β and Σ is

$$(2.7) \quad p(\beta, \Sigma \mid y) \propto |\Sigma|^{-\frac{T+(m+1)}{2}} \exp\left\{-\tfrac{1}{2}(y-Z\beta)' \Sigma^{-1} \otimes I_T (y-Z\beta)\right\} .$$

In what follows we discuss the properties of this distribution.

3. Posterior distribution of β and Σ when $X_1 = \ldots = X_m = X$.

In the situation where the X's are identical, it is well known that the statistics

$$(3.1a) \quad \hat{\beta}_\alpha = (X'X)^{-1} X'y_\alpha \qquad \alpha = 1, \ldots, m$$

$$(3.1b) \quad s_{\alpha\ell} = (y_\alpha - X\hat{\beta}_\alpha)'(y_\ell - X\hat{\beta}_\ell) \qquad \ell = 1, \ldots, m$$

are jointly sufficient for β and Σ, and the likelihood function in (1.2) can be written

$$(3.2) \quad \ell(\beta, \Sigma \mid y) \propto |\Sigma|^{-\frac{T}{2}} \exp\left\{-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1} S - \tfrac{1}{2}(\beta-\hat{\beta})' \Sigma^{-1} \otimes I_k (\beta-\hat{\beta})\right\}$$

where I_k is a $k \times k$ identity matrix and $S = \{s_{\alpha\ell}\}$ is proportional to the sample covariance matrix. Using (3.2), the posterior distribution of β and Σ in (2.7) can be expressed as

$$(3.3) \quad p(\beta, \Sigma \mid y) = p(\beta \mid \Sigma\ y)\, p(\Sigma \mid y)$$

with

$$(3.4) \quad p(\beta \mid \Sigma, y) \propto |\Sigma|^{-\frac{k}{2}} \exp\left\{-\tfrac{1}{2}(\beta-\hat{\beta})' \Sigma^{-1} \otimes I_k (\beta-\hat{\beta})\right\} .$$

and

$$(3.5) \qquad p(\Sigma \mid y) \propto |\Sigma|^{-\frac{\nu}{2}} \exp\left\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}S\right\} .$$

with $\nu = T-k+m+1$. It is seen that the conditional posterior distribution of β for given value of Σ is multivariate normal with mean $\hat{\beta}$ and covariance matrix $\Sigma \otimes I_k$ as noted by Dre'ze (1962). In particular, if interest centers only on β_α, its conditional distribution is multivariate normal,

$$(3.6) \qquad p(\beta_\alpha \mid \Sigma, y) \propto \sigma_{\alpha\alpha}^{-\frac{k}{2}} \exp\left\{-\frac{1}{2\sigma_{\alpha\alpha}} (\beta_\alpha - \hat{\beta}_\alpha)' X'X (\beta - \hat{\beta}_\alpha)\right\} .$$

which depends only on $\sigma_{\alpha\alpha}$. The posterior distribution of Σ 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 Σ 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 Σ_{11} where Σ_{11} is the $p \times p$ upper left-hand principal minor matrix of Σ $(p<m)$. Denoting

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

and remembering that Σ is assumed positive definite, we can express the determinant and the inverse of Σ as

(4.1) $$|\Sigma| = |\Sigma_{11}|\,|\Sigma_{22.1}| \qquad \text{where } \Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

and

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

say. Thus, the distribution in (3.5) can be written

(4.3) $$p(\Sigma|y) \propto |\Sigma_{11}|^{-\frac{\nu}{2}}\,|\Sigma_{22.1}|^{-\frac{\nu}{2}} \exp\left\{-\tfrac{1}{2}\mathrm{tr}\,\Sigma_{11}^{-1}S_{11} - \tfrac{1}{2}\mathrm{tr}\,W\,S\right\}$$

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

$$\begin{cases} Y = \Sigma_{11}^{-1}\Sigma_{12} \\ \Omega = \Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{cases}$$

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

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

Consequently, we have that

(4.4) $$p(\Sigma_{11}, Y, \Omega|y) \propto |\Sigma_{11}|^{-\frac{\nu-2(m-p)}{2}}\,|\Omega|^{-\frac{\nu}{2}} \exp\left\{-\tfrac{1}{2}\mathrm{tr}\,\Sigma_{11}^{-1}S_{11} - \tfrac{1}{2}\mathrm{tr}\,W\,S\right\}.$$

is implies that the marginal distribution of the $\frac{p(p+1)}{2}$ elements of Σ_{11} is,

(4.5) $$p(\Sigma_{11}|y) \propto |\Sigma_{11}|^{-\frac{\nu-2(m-p)}{2}} \exp\left\{-\tfrac{1}{2}\mathrm{tr}\,\Sigma_{11}^{-1}S_{11}\right\}$$

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

$$(4.6) \quad p(\sigma_{11}|y) \propto \sigma_{11}^{-\frac{\nu-2(m-1)}{2}} \exp\left\{-\frac{s_{11}}{2\sigma_{11}}\right\}$$

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 (1962) and can be obtained by setting $m=1$ in (3.6) to yield:

$$(4.7) \quad p(\sigma_{11}|y) \propto \sigma_{11}^{-\frac{T-k+2}{2}} \exp\left\{-\frac{s_{11}}{2\sigma_{11}}\right\} .$$

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 pay, 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 pp 174) to obtain the posterior distribution of the correlation coefficient ρ_{12} as:

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

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

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

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

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. <u>Posterior distribution of β_1</u>

In many practical applications of the multivariate regression model, an investigator's main interest may be centered on the regression coefficients of a particular equation. As we have seen in (3.6), the conditional posterior distribution of β_1 given Σ depends upon only σ_{11}. Thus from (3.6) and (4.6), we have for the marginal posterior distribution of β_1,

$$(5.1) \qquad p(\beta_1|y) = \int p(\sigma_{11}|y)\, p(\beta_1|\Sigma, y)\, d\sigma_{11}$$

$$\propto \left\{ s_{11} + (\beta_1 - \hat{\beta}_1)' X'X(\beta_1 - \hat{\beta}_1) \right\}^{-\frac{T-(m-1)}{2}} .$$

Expression (5.1) is then in the form of a multivariate -t distribution --- see e.g. Dunnet 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, e.g. Savage (1961), the only difference is then the change in the "degrees of freedom" due to the inclusion of the m-1 parameters $\sigma_{12}, \ldots,$ σ_{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 paramters β_1, say $\beta_1' = (\beta_{11}, \ldots, \beta_{1r}, \beta_{1(r+1)}, \ldots, \beta_{1k})$ 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 $\beta_{1(r+1)} = \ldots = \beta_{rk} = 0$ is again of the multivariate t form-- for details see e.g. Raiffa and Schlaifer (1961, pp. 258).

6. The Joint Posterior Distribution of β.

In this section, we give an alternative derivation of the posterior distribution of β_1 by first finding the joint posterior distribution of $\beta = (\beta_1',\ldots,\beta_m')$. From the joint posterior distribution of Σ and β in (3.3) and the Jacobian in (2.5), we immediately deduce the distribution of β and Σ^{-1} as

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

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

$$b_{\alpha\ell} = (\beta_\alpha - \hat{\beta}_\alpha)' X' X (\beta_\ell - \hat{\beta}_\ell).$$

For fixed B, the distribution in (5.1) is in the Wishart form. From the properties of the Wishart distribution, integrating over the elements of Σ^{-1} yields the marginal posterior distribution of β as

$$(6.2)\qquad p(\beta|y) \propto |S+B|^{-\frac{T}{2}} .$$

In the special case of sampling from a m-variate multinormal distribution, (6.2) is the joint posterior distribution of the m means, as first derived by Savage (1961). 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 (6.2) in the multivariate-t form, even though we have seen from (5.1) that the marginal distribution of β_1 is of this form. We now show, however, that if we express the joint distribution of β as the product

(6.3) $p(\beta|y) = p(\beta_1|y)p(\beta_2|\beta_1, y) \ldots\ldots p(\beta_m|\beta_1, \ldots, \beta_{m-1}, y)$,

then each of the factors on the right of (5.3) can be expressed in terms of a multivariate-t distribution. We first derive an expression for the product $p(\beta_1 \ldots \beta_{m-1}|y)\, p(\beta_m|\beta_1, \ldots, \beta_{m-1}, y)$.

Denote the determinant $|S+B|$ in (6.2) as:

(6.5) $$|S+B| = \left|\begin{array}{c|c} \overline{S}+\overline{B} & (s+b) \\ \hline (s+b)' & s_{mm}+b_{mm} \end{array}\right|$$

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

(6.6) $s' = (s_{m_1}, \ldots, s_{m(m-1)})$ and $b' = (b_{m1}, \ldots b_{m(m-1)})$.

Using Cauchey's expansion formula for determinants, we can write (6.5) as

(6.7) $$|S+B| = |\overline{S}+\overline{B}| \left\{ s_{mm} + b_{mm} - (s+b)'(\overline{S}+\overline{B})^{-1}(s+b) \right\} .$$

In the second factor on the right of (5.7), let us write

(6.8) $$\begin{array}{llll} b_{\alpha\ell} = \gamma'_\alpha \gamma_\ell & \text{with} & \gamma_\ell = X(\beta_\ell - \hat{\beta}_\ell) & \alpha=1,\ldots,m \\ & & & \ell=1,\ldots,m \\ b' = \gamma'_m \gamma & \text{where} & \gamma = (\gamma_1, \ldots, \gamma_{m-1}) . & \end{array}$$

After a little alegbraic rearrangement, we have that

$$(6.9)\qquad |S+B| = |\bar{S}+\bar{B}| \left\{ c_m + (\gamma_m - \mu)' D_m (\gamma_m - \mu) \right\}$$

where $\qquad D_m = I - \gamma(\bar{S}+\bar{B})^{-1} \gamma'$

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

and $\qquad c_m = s_{mm} - s'(\bar{S}+\bar{B})^{-1} s - \mu' D_m \mu .$

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

$$(6.10)\qquad (I_m - AB)^{-1} = I_m + A(I_n - BA)^{-1} B$$

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

$$D_m^{-1} = I + \gamma' \bar{S}^{-1} \gamma$$

$$(6.11)\qquad \mu = \gamma \bar{S}^{-1} s$$

$$c_m = s_{mm} - s' \bar{S}^{-1} s$$

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

$$(6.12)\qquad \left\{ c_m + (\gamma_m - \mu)' D_m (\gamma_m - \mu) \right\} = \left\{ c_m + (\beta_m - \eta_m)' X' D_m X (\beta_m - \eta_m) \right\}$$

with $\qquad \eta_m = \hat{\beta}_m + d \bar{S}^{-1} s \qquad$ and $\qquad d = (\beta_1 - \hat{\beta}_1, \ldots, \beta_{m-1} - \hat{\beta}_{m-1}) .$

Using (6.9) and (6.12), we can therefore write the distribution in (6.2) as

$$(6.13)\qquad p(\beta | y) = |\bar{S}+\bar{B}|^{-\frac{T}{2}} \left\{ c_m + (\beta_m - \eta_m)' X' D_m X (\beta_m - \eta_m) \right\}^{-\frac{T}{2}}$$

Now the determinant of the matrix $X'D_m X$

$$(6.14) \qquad |X'D_m X| = |X'X| \; |I - (X'X)^{-1} X'\gamma(\bar{S}+\bar{D})^{-1} \gamma' X|$$

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

Since $X'D_m X$ is positive definite (this is seen by noting D_m^{-1} in (6.11) is clearly positive definite), the second determinant on the right of (6.14) is positive. By taking logarithms, it is easily verified that

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

Hence,

$$(6.15) \qquad |X'DX| = |X'X| \; |I - (\bar{S}+\bar{B})^{-1} \gamma'\gamma|$$

$$= |X'X| \; |\bar{S}| \; |\bar{S}+\bar{B}|^{-1} .$$

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

$$(6.16) \qquad p(\beta|y) = p(\beta_1, \ldots, \beta_{m-1}|y) \, p(\beta_m|\beta_1, \ldots, \beta_{m-1}, y)$$

with

$$(6.17) \qquad p(\beta_1, \ldots, \beta_{m-1}|y) \; \alpha \; |\bar{S}+\bar{B}|^{-\frac{[illegible]}{2}}$$

and

$$(6.18) \qquad p(\beta_m|\beta_1, \ldots, \beta_{m-1}, y) \; \alpha \; |X'D_m X|^{\frac{1}{2}} \left\{ c_m + (\beta_m - \eta_m)' X'D_m X (\beta_m - \eta_m) \right\}^{-}$$

The conditional distribution of β_m can therefore be expressed in terms of a multivariate t distribution, while the marginal distribution of $\beta_1, \ldots, \beta_{m-1}$ is of the same form as the original distribution of β except of course for the changes in the dimensions of the matrix $\bar{S}+\bar{B}$ 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

$$(6.19)\qquad p(\beta|y) \propto \left\{s_{11} + (\beta_1-\hat{\beta}_1)'X'X(\beta_1-\hat{\beta}_1)\right\}^{-\frac{T-(m-1)}{2}} \prod_{\alpha=2}^{m} |X'D_\alpha X|^{\frac{1}{2}} \left\{c_\alpha + (\beta_\alpha-\eta_\alpha)'X'D_\alpha X(\beta_\alpha-\eta_\alpha)\right\}^{-\frac{T-m+\alpha}{2}}$$

where D_α, η_α 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 (6.3), and clearly the first factor is the marginal distribution of β_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, these 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 (6.2). 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

$$(6.20)\qquad \begin{aligned} y_1 &= X_1\beta_1 + u_1 \\ y_2 &= X_2\beta_2 + u_2 \end{aligned}$$

If we form the augumented matrix X such that

$$X = [X_c \,\vdots\, X_{1d} \,\vdots\, X_{2d}]$$

where $X_1 = [X_c \,\vdots\, X_{1d}]$

$$X_2 = [X_c \,\vdots\, X_{2d}] \quad .$$

Then (6.20) can be written as

$$\begin{cases} y_1 = X\beta_1^* + u_1 \\ y_2 = X\beta_2^* + u_2 \end{cases}$$

provided that particular subsets of β_1^* and β_2^* compatible with the partitioning of X have zero values. Note however that this approach for the general model requires that the augmented matrix X must be of full rank.

7. Posterior distribution of β 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 β and Σ in (2.7), it is clearly that the conditional distribution of β given Σ is normal with mean

(7.1) $$\tilde{\beta} = \left\{Z' \Sigma^{-1} \otimes I_T\, Z\right\}^{-1} Z'\Sigma^{-1} \otimes I_T\, y$$

and covariance matrix

(7.2) $$\mathrm{Cov}(\beta) = \left\{Z' \Sigma^{-1} \otimes I_T\, Z\right\}^{-1}$$

It is seen that $\tilde{\beta}$, the center of the conditional distribution, depends upon Σ and only in the case in which the X_α's are identical (or proportional) will (7.1) reduces to (3.1a). As regard the marginal distributions of Σ and of β, unfortunately because of the dependence of β on Σ the analysis in Sections 3-6 can not be extended here. However, by an argument similar to that given in Section 6, the posterior distribution of β and Σ^{-1} can be written

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

where $U = \left\{u'_\alpha u_\ell\right\}$

and $u_\alpha = y_\alpha - X_\alpha \beta_\alpha$ $\quad \alpha = 1, \ldots, m$, $\quad \ell = 1, \ldots, m$.

From (7.3), we obtain the marginal posterior distribution of β as

$$p(\beta|y) \propto |U|^{-\frac{T}{2}}$$

Properties of this distribution are currently being investigated.

REFERENCES

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

Box, G. E. P. and Tiao, G. C. (1962). A further look at robustness via Bayes' Theorem. Biometrika, 49, 419-432.

Box, G. E. P. and Tiao, G. C. (1963). A Bayesian Approach to the importance of assumptions applied to the comparison of variances. Tech. Report #16, Department of Statistics, University of Wisconsin.

Drēze, J. (1962). The Bayesian approach to simultaneous equations estimation. O. N. R. Research Memo #67, The Technological Institute, Northwestern University.

Geisser, S. and Cornfield, J. (1963). Posterior distribution for multivariate normal parameters (abstract). Biometrics, 19, 505.

Jeffreys, H. (1961). Theory of Probability (3rd edition). Oxford: Clarendon Press.

Savage, L. J. (1961). The Subjective Basis of Statistical Practice (manuscript). University of Michigan.

Telser, L. G. (1963). Iterative estimation of a set of linear regression equations (manuscript). University of Chicago.

Zellner, A. (1962). An efficient method of estimation seemingly unrelated regression and test for aggregation bias. J. A. S. A., 57, 348-68.

Zellner, A. (1963). Estimators for seemingly unrelated regression equations: some exact finite sample results. J. A. S. A., 58, 977-92.

Tocher, K. D. (1951). Discussion on Mr. Box and Dr. Wilson's paper, J. R. Stat. Soc. (B), 13, 39-42.