DISTRIBUTION OF SUM OF SQUARES AND PRODUCTS MATRICES FOR THE GENERALIZED MULTILINEAR MATRIX-T MODEL

Shahjahan Khan

Department of Mathematics & Computing Sciences
The University of Southern Queensland

RUNNING HEAD :
DISTRIBUTIONS IN MATRIX-T MODEL

Shahjahan Khan

Department of Mathematics & Computing Sciences
The University of Southern Queensland
Toowoomba, Qld. 4350, Australia
Email: khans@usq.edu.au

ABSTRACT

The generalized multilinear model with the matrix-T error distribution is introduced in this paper. The sum of squares and products (SSP) matrix, as a counterpart of the Wishart matrix for the multinormal model, and the regression matrix for the errors and the observed as well as future responses are defined. The distributions of the regression matrix as well as the SSP matrix, and the prediction distribution of the future regression matrix and the future SSP matrix are derived.

Key Words: Generalized multilinear model, matrix-T distribution, sum of squares and product matrix, regression matrix, prediction distribution, Wishart distribution, invariant differentials, generalized beta and gamma distributions.

AMS 1991 Subject Classification : Primary 62H10, secondary 62J12.

1. INTRODUCTION

The Wishart distribution plays a key role in multivariate statistical analysis. This important distribution was first derived by Wishart (1928) as the generalized product moment distribution in sampling from a multivariate normal model. Further derivations include Hsu (1939), Sverdrup (1947), Janbunathan (1965) and Fraser (1968a). In some sense the Wishart distribution is a generalization of the univariate gamma (or χ^2) distribution derived as a sum of squares of samples from a normal population. It has a wide range of applications in statistical inference problems, e.g. testing hypotheses in multivariate analysis, factor analysis, as a natural conjugate prior in Bayesian analysis for the precision matrix, to mention a few.

Traditionally most of the work on statistical theory is based on the normal or multinormal models. The Wishart distribution is also obtained for the multinormal model. But, the normality assumption is under increasing criticism for being non-robust. It fails to allow sufficient probability in the tail areas to make allowance for the outliers or extreme values. Furthermore, it can not handle the dependent but uncorrelated responses which are often common in time series and econometric studies. On the other hand, the matrix-T distribution as a generalization of the multivariate Student-t distribution can overcome both the problems of outliers as well as dependent but uncorrelated data. More importantly, the multivariate normal distribution is a special case of the multivariate Student-t distribution when the degrees of freedom parameter approaches infinity. It also covers the multivariate Cauchy distribution as a special case when there is only one degree of freedom available. For further justification of preference for the multivariate-t distribution over the multivariate normal distribution, readers may refer to Prucha and Kelejian (1984).

Dickey (1967) derived the Matrix-T distribution as a logical generalization of the multivariate Student-t distribution to deal with matrix variate problems. It has wide applications in multivariate inference, especially in Bayesian analysis, as has been appreciated by many authors including Box and Tiao (1992, sec. 8.4) and Press (1986).

The prediction distribution is of pivotal importance for predictive inference. It has many applications in real life inferential problems. Aitchison and Dunsmore (1975) emphasized the suitability of predictive inference, as opposed to the parametric inference in the form of estimation and tests of parameters. Recently, Geisser (1993) used the prediction distribution in many predictive inference applications. Some of the most common and popular usages of the prediction distribution is in the construction of tolerance regions, calibration, classification, test of goodness of fit, selection of best population, perturbation analysis, process control and optimization. Unlike the above normal based studies, Khan and Haq (1994) investigated the predictive inference for the future responses from a multilinear model with matrix-T errors.

In this paper, we introduce the generalized multilinear model with matrix-T error distribution. We define the sum of squares and products (SSP) matrix for the errors as well as the responses for the matrix-T model as a counter part of the Wishart matrix for multinormal models. The SSP matrix will have a Wishart distribution if the errors are normally distributed. Obviously, for the matrix-T model the SSP matrix does not follow a Wishart distribution. Since the matrix-T distribution approaches to matrix variate normal distribution as the degrees of freedom parameter tends to infinity, the matrix-T model under study in this paper encompases the matrix normal distribution as a special case as the limit. Haq and Rinco (1976) considered a similar model with independent normal errors to construct a β-expectation tolerance region for the future responses of the model using the structural distribution method. Here, we are interested in the distributions of the regression as well as the SSP matrices for the matrix-T model. The prediction distributions of the future regression matrix and the future SSP matrix are also of interest. In particular, the distributions of the regression and the SSP matrix are derived for the generalised matrix-T multilinear model using the invariant differentials as well as orthogonal and triangular factorisation. The prediction distributions of the future regression and the SSP matrix are also obtained for the model.

In section 2, the matrix-T distribution and the generalized multilinear model are introduced. Some preliminaries are given in section 3. Distributions of the SSP matrix and the regression matrix are obtianed in section 4. Section 5 derives the prediction distributions of the SSP matrix as well as the regression matrix of the future errors and responses.

2. THE MATRIX-T DISTRIBUTION AND GENERALIZED MULTILINEAR MODEL

Let $\boldsymbol{U}$ be an $m \times n$ matrix of random variables. Then it is said to have a matrix-T distribution if the joint density of the mn random elements of $\boldsymbol{U}$ is given by

$$
\begin{aligned}
p\Big(\boldsymbol{U};\, \boldsymbol{\mu}, A, \Omega, \nu\Big) &= \frac{\Gamma_n\Big(\frac{\nu+m+n-1}{2}\Big)}{(\pi)^{\frac{mn}{2}}\Gamma_n\Big(\frac{\nu+n-1}{2}\Big)} \frac{|\Omega|^{\frac{n}{2}}}{|A|^{-\frac{\nu+n-1}{2}}} \\
&\times \Big|A + \Big(\boldsymbol{U}-\boldsymbol{\mu}\Big)'\Omega^{-1}\Big(\boldsymbol{U}-\boldsymbol{\mu}\Big)\Big|^{-\frac{\nu+m+n-1}{2}}
\end{aligned}
\tag{2.1}
$$

where $\mathrm{E}(\boldsymbol{U}) = \boldsymbol{\mu}$, an $m \times n$ matrix of location parameters; Ω is an $m \times m$ scaled covariance matrix of each column of $\boldsymbol{U}$; A is a positive definite matrix of order $n \times n$; $\nu > 0$ is the shape parameter; and $\Gamma_b(\frac{c}{2}) = (\pi)^{\frac{b(b-1)}{4}} \prod_{i=1}^{b} \Gamma(\frac{c-i+1}{2})$ is the generalized gamma function. The matrix-T density was first obtained by Dickey (1967) and it can be equivalently written as the density of the transpose of $\boldsymbol{U}$. From the above density it is clear that the matrix-T distribution is a member of the elliptically symmetric family of distributions. In notation, we write $\boldsymbol{U} \sim T_{m\times n}\Big(\boldsymbol{\mu}, A, \Omega, \nu\Big)$. Note that the covariance of $\boldsymbol{U}$ is $\frac{1}{\nu-2}A\otimes\Omega$, an $mn \times mn$ matrix, where $\otimes$ is the Kronecker product between two matrices. Thus, for the covariance matrix to be finite, we need a restriction on ν, namely, $\nu > 2$. Since ν is a positive real number, for different value of ν we get a different distribution, and hence the matrix-T model under study indeed represents a class of elliptically symmetric distributions with varying shape. When the shape parameter of the matrix-T distribution tends to infinity, the distribution of $\boldsymbol{U}$ approaches matric-variate normal. Thus, $\lim_{\nu\to\infty} T_{m\times n}(\boldsymbol{\mu}, A, \Omega, \nu) \to N_{m\times n}(\boldsymbol{\mu}, A^* \otimes \Omega)$ where $A^* = \frac{A}{\nu}$. Moreover, for $\nu = 1$, the matrix-T distribution becomes matrix Cauchy distribution. It may also include a range of sub-Cauchy distributions when $0 < \nu < 1$.

The marginal and conditional distributions of any row (or column, if interested) and one row, given another follow matrix-T distribution with appropriate parameters. As a special case, if $n = 1$, the matrix $\boldsymbol{U}$ reduces to $\boldsymbol{u}$, just a column vector of m components, and hence $\boldsymbol{u} \sim t_m(\mu_{(m)}, a, \Omega, \nu)$ where $\mathrm{E}(\boldsymbol{u}) = \boldsymbol{\mu}_{(m)}$ and $\mathrm{Cov}(\boldsymbol{u}) = \frac{a}{\nu-2}\Omega$, in which a is a scalar quantity. Let $\boldsymbol{U}$, $\boldsymbol{\mu}$ and A be partitioned as follows: $\boldsymbol{U}_{m\times n} = [\boldsymbol{U}^1_{m\times n_1} \;\vdots\; \boldsymbol{U}^2_{m\times n_2}]$, $\boldsymbol{\mu}_{m\times n} = [\boldsymbol{\mu}^1_{m\times n_1} \;\vdots\; \boldsymbol{\mu}^2_{m\times n_2}]$ and $A_{n\times n} = \begin{pmatrix} A^{11}_{n_1\times n_1} & A^{12}_{n_1\times n_2} \\ A^{21}_{n_2\times n_1} & A^{22}_{n_2\times n_2} \end{pmatrix}$ such that $n = n_1 + n_2$ and $m = m_1 + m_2$.

Then the marginal distribution of $\boldsymbol{U}^1$ is matrix-T with appropriate parameters, that is, $\boldsymbol{U}^1 \sim T_{m\times n_1}\Big(\boldsymbol{\mu}^1, A^{11}, \Omega, \nu\Big)$. Also, the conditional distribution of $\boldsymbol{U}^2$, given $\boldsymbol{U}^1$ is matrix-T, that is, $(\boldsymbol{U}^2|\boldsymbol{U}^1) \sim T_{m\times n_2}\Big(\boldsymbol{\mu}_2, A_{22.1}, \Omega^*, \nu + n_1\Big)$ where $\boldsymbol{\mu}_2 = \boldsymbol{\mu}^2 + (\boldsymbol{\mu}^1 - \boldsymbol{U}^1){A^{11}}^{-1}A^{12}$, $\Omega^* = \Omega + (\boldsymbol{U}^1 - \boldsymbol{\mu}^1){A^{11}}^{-1}(\boldsymbol{U}^1 - \boldsymbol{\mu}^1)'$ and $A_{22.1} = A^{22} - A^{21}{A^{11}}^{-1}A^{12}$. The distribution of sub-blocks of $\boldsymbol{U}$ can also be obtained in a similar fashion.

Now consider the following generalized multilinear model

$$\boldsymbol{Y} = \boldsymbol{\beta}\boldsymbol{X} + \boldsymbol{\Gamma}\boldsymbol{E} \tag{2.2}$$

where $\boldsymbol{Y}$ is a matrix of order $m \times n$, each of its n columns may be viewed as the response on m characteristics of an experiment; $\boldsymbol{\beta}$ is an $m \times p$ matrix of regression parameters; $\boldsymbol{X}$ is the $p \times n$ matrix of regressors, usually known as design matrix; $\boldsymbol{\Gamma} > 0$ is the scale parameter matrix of order $m \times m$; and $\boldsymbol{E}$ is the $m \times n$ matrix of errors associated with the response matrix $\boldsymbol{Y}$.

Assume the errors in the model are dependent, but uncorrelated, and jointly follow a matrix-T probability distribution. Also, assume that the expectation of $\boldsymbol{E}$ is $\mathbf{0}$ and the covariance matrix of $\boldsymbol{E}$ is $\frac{1}{\nu-2}I_m \otimes I_n$, where $\mathbf{0}$ is an $m \times n$ matrix of 0's and $\otimes$ denotes the Kronecker product. Thus the covariance matrix of each column of $\boldsymbol{Y}$ is $\boldsymbol{\Sigma} = \frac{1}{\nu-2}\boldsymbol{\Gamma}\boldsymbol{\Gamma}'$, and that of the $\boldsymbol{Y}$ matrix is $\frac{1}{\nu-2}\boldsymbol{\Sigma} \otimes I_n$.

The joint density function of the mn random elements of $\boldsymbol{E}$ can be written as

$$p(\boldsymbol{E}) = \frac{\Gamma_m\left(\frac{r+m+n-1}{2}\right)}{\pi^{\frac{mn}{2}}\Gamma_m\left(\frac{r+m-1}{2}\right)}\left|I_m + \boldsymbol{E}\boldsymbol{E}'\right|^{-\frac{r+m+n-1}{2}} \tag{2.3}$$

where r is the number of degrees of freedom of the matrix-T distribution for the errors. The above matrix-T density appears in many textbooks including Johnson and Kotz (1972), and Press (1986). Fraser and Ng (1980) used such a density to analyse multilinear model under a structural distribution set-up.

In this paper, we derive the distribution of the SSP matrix, and the prediction distribution of the future regression as well as future SSP matrix for the generalized multilinear model as specified in (2.2) and (2.3). To guarantee the positive definiteness of the SSP matrix as well as the integrability on higher dimension we require that $n > m + p$.

3. SOME PRELIMINARIES

Let us denote the regression matrix of $\boldsymbol{E}$ on $\boldsymbol{X}$ by $\boldsymbol{B}(\boldsymbol{E})$ and the error SSP matrix by $\boldsymbol{S}(\boldsymbol{E})$. Then we have

$$\begin{aligned} \boldsymbol{B}(\boldsymbol{E}) &= \boldsymbol{E}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{X}')^{-1} \text{ and} \\ \boldsymbol{S}(\boldsymbol{E}) &= [\boldsymbol{E} - \boldsymbol{B}(\boldsymbol{E})\boldsymbol{X}][\boldsymbol{E} - \boldsymbol{B}(\boldsymbol{E})\boldsymbol{X}]'. \end{aligned} \tag{3.1}$$

Let $\boldsymbol{C}(\boldsymbol{E})$ be a nonsingular matrix such that the error SSP matrix $\boldsymbol{S}(\boldsymbol{E})$ can be written as $\boldsymbol{C}(\boldsymbol{E})\boldsymbol{C}'(\boldsymbol{E}) = \boldsymbol{S}(\boldsymbol{E})$, and $\boldsymbol{D}(\boldsymbol{E}) = \boldsymbol{C}^{-1}(\boldsymbol{E})[\boldsymbol{E} - \boldsymbol{B}(\boldsymbol{E})\boldsymbol{X}]$ is the 'standardized' residual matrix.

Now we can write the error matrix, $\boldsymbol{E}$, in the following way:

$$\boldsymbol{E} = \boldsymbol{B}(\boldsymbol{E})\boldsymbol{X} + \boldsymbol{C}(\boldsymbol{E})\boldsymbol{D}(\boldsymbol{E}) \tag{3.2}$$

and hence we get

$$\boldsymbol{E}\boldsymbol{E}' = \boldsymbol{B}(\boldsymbol{E})\boldsymbol{X}\boldsymbol{X}'\boldsymbol{B}'(\boldsymbol{E}) + \boldsymbol{C}(\boldsymbol{E})\boldsymbol{C}'(\boldsymbol{E}) \; , \tag{3.3}$$

since $\boldsymbol{D}(\boldsymbol{E})\boldsymbol{D}'(\boldsymbol{E}) = I_m$ and $\boldsymbol{X}\boldsymbol{D}'(\boldsymbol{E}) = \boldsymbol{O}$.

From (3.2) and (2.2), the following relations can easily be established:

$$\begin{aligned} \boldsymbol{B}(\boldsymbol{E}) &= \boldsymbol{\Gamma}^{-1}\{\boldsymbol{B}(\boldsymbol{Y}) - \boldsymbol{\beta}\} \\ \boldsymbol{C}(\boldsymbol{E}) &= \boldsymbol{\Gamma}^{-1}\boldsymbol{C}(\boldsymbol{Y}), \end{aligned} \tag{3.4}$$

where

$$\begin{aligned} \boldsymbol{B}(\boldsymbol{Y}) &= \boldsymbol{Y}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{X}')^{-1} \quad \text{and} \\ \boldsymbol{S}(\boldsymbol{Y}) &= \boldsymbol{C}(\boldsymbol{Y})\boldsymbol{C}'(\boldsymbol{Y}) \end{aligned}$$

are the regression matrix of $\boldsymbol{Y}$ on $\boldsymbol{X}$, and the SSP matrix for the observed responses respectively.

It may be mentioned here that both $\boldsymbol{C}(\boldsymbol{E})$ and $\boldsymbol{C}(\boldsymbol{Y})$ have the same structure since the definitions of $\boldsymbol{S}(\boldsymbol{E})$ in (3.1) and that of $\boldsymbol{S}(\boldsymbol{Y})$ in (3.4) ensure the same format of the two SSP matrices of the error and response respectively. For the derivation of some of the forthcoming results, it is required that the determinant of $\boldsymbol{C}(\boldsymbol{E})$ is positive in the sense that $\boldsymbol{S}(\boldsymbol{E}) = \boldsymbol{C}(\boldsymbol{E})\boldsymbol{C}'(\boldsymbol{E})$ is positive definite (cf. Fraser and Ng, 1980). It can be easily shown that $\boldsymbol{D}(\boldsymbol{E}) = \boldsymbol{D}(\boldsymbol{Y})$. In the next section, the distributions of $\boldsymbol{S}(\boldsymbol{E})$, $\boldsymbol{S}(\boldsymbol{Y})$, $\boldsymbol{B}(\boldsymbol{E})$ and $\boldsymbol{B}(\boldsymbol{Y})$ are obtained.

4. THE DISTRIBUTION OF THE SSP MATRIX

From the probability density of $\boldsymbol{E}$ in (2.3) and the relation (3.3) the joint probability density of $\boldsymbol{B}(\boldsymbol{E})$ and $\boldsymbol{C}(\boldsymbol{E})$, conditional on the $\boldsymbol{D}(\boldsymbol{E})$, is obtained by using the invariant differentials (see Eaton, 1983, p.194-206) as follows

$$p\Big(\boldsymbol{B}(\boldsymbol{E}),\ \boldsymbol{C}(\boldsymbol{E})\Big|\boldsymbol{D}(\boldsymbol{E})\Big) = K_1(\boldsymbol{D})\ \Big|\boldsymbol{C}(\boldsymbol{E})\Big|^{n-p-m} \Big|I_m + \boldsymbol{B}(\boldsymbol{E})\boldsymbol{X}\boldsymbol{X}'\boldsymbol{B}'(\boldsymbol{E}) + \boldsymbol{C}(\boldsymbol{E})\boldsymbol{C}'(\boldsymbol{E})\Big|^{-\frac{r+m+n-1}{2}} \tag{4.1}$$

where $K_1(\boldsymbol{D})$ is the normalizing constant.

It is convenient to factorise $\boldsymbol{C}(\boldsymbol{E})$ into its *orthogonal component* $\boldsymbol{O}(\boldsymbol{E})$ and the *positive lower triangular component* $\boldsymbol{L}(\boldsymbol{E})$ as follows:

$$\boldsymbol{C}(\boldsymbol{E}) = \boldsymbol{L}(\boldsymbol{E})\boldsymbol{O}(\boldsymbol{E}). \tag{4.2}$$

For detail on such factorisation see Fraser (1968b, Ch.3 Sec. 6). This kind of factorisation is essential to facilitate the multiple integrations. Now it can be shown that

$$\frac{d\boldsymbol{C}(\boldsymbol{E})}{|\boldsymbol{C}(\boldsymbol{E})|^m} = \frac{d\boldsymbol{L}(\boldsymbol{E})d\boldsymbol{O}(\boldsymbol{E})}{|\boldsymbol{L}(\boldsymbol{E})|_\Delta} \tag{4.3}$$

where $\mathrm{d}\,\boldsymbol{O}(\boldsymbol{E})$ is interpreted as the volume orthogonal to the orbits of the positive lower triangular scale group in $\Re^{m^2}$, $\boldsymbol{L}(\boldsymbol{E})$ is a lower triangular matrix, $|\boldsymbol{L}(\boldsymbol{E})|_\Delta$ is the increasing determinant of $\boldsymbol{L}(\boldsymbol{E})$ and is equal to the product of the diagonal elements of $\boldsymbol{L}(\boldsymbol{E})$ each being raised to the power of its position. Now the relation (3.3) can be written as

$$\begin{aligned}\boldsymbol{E}\boldsymbol{E}' &= \boldsymbol{B}(\boldsymbol{E})\boldsymbol{X}\boldsymbol{X}'\boldsymbol{B}'(\boldsymbol{E}) + \boldsymbol{L}(\boldsymbol{E})\boldsymbol{O}(\boldsymbol{E})\boldsymbol{O}'(\boldsymbol{E})\boldsymbol{L}'(\boldsymbol{E}) \\ &= \boldsymbol{B}(\boldsymbol{E})\boldsymbol{X}\boldsymbol{X}'\boldsymbol{B}'(\boldsymbol{E}) + \boldsymbol{L}(\boldsymbol{E})\boldsymbol{L}'(\boldsymbol{E})\ .\end{aligned} \tag{4.4}$$

Therefore, from (4.1), (4.3) and (4.4), the joint probability element of $\boldsymbol{B}(\boldsymbol{E})$, $\boldsymbol{L}(\boldsymbol{E})$ and $\boldsymbol{O}(\boldsymbol{E})$ becomes

$$\begin{aligned} p\Big(\boldsymbol{B}(\boldsymbol{E}), \boldsymbol{L}(\boldsymbol{E}), \boldsymbol{O}(\boldsymbol{E}) \Big| \boldsymbol{D}(\boldsymbol{E})\Big) d\boldsymbol{B}(\boldsymbol{E}) d\boldsymbol{L}(\boldsymbol{E}) d\boldsymbol{O}(\boldsymbol{E}) &= K_2(\boldsymbol{D}) \frac{|\boldsymbol{L}(\boldsymbol{E})|^{n-p}}{|\boldsymbol{L}(\boldsymbol{E})|_\triangle} \\ \times \Big| I_m + \boldsymbol{B}(\boldsymbol{E})\boldsymbol{X}\boldsymbol{X}'\boldsymbol{B}'(\boldsymbol{E}) + \boldsymbol{L}(\boldsymbol{E})\boldsymbol{L}'(\boldsymbol{E}) \Big|^{-\frac{r+m+n-1}{2}} & d\boldsymbol{B}(\boldsymbol{E}) d\boldsymbol{L}(\boldsymbol{E}) d\boldsymbol{O}(\boldsymbol{E}) \end{aligned} \tag{4.5}$$

where $K_2(\boldsymbol{D})$ is the appropriate normalizing constant. Then the marginal density of $\boldsymbol{B}(\boldsymbol{E})$ and $\boldsymbol{L}(\boldsymbol{E})$ is obtained from (4.5) by integrating out $\boldsymbol{O}(\boldsymbol{E})$:

$$\begin{aligned} p\Big(\boldsymbol{B}(\boldsymbol{E}), \boldsymbol{L}(\boldsymbol{E}) \Big| \boldsymbol{D}(\boldsymbol{E})\Big) = K_3(\boldsymbol{D}) \ & \frac{|\boldsymbol{L}(\boldsymbol{E})|^{n-p}}{|\boldsymbol{L}(\boldsymbol{E})|_\triangle} \\ & \times \Big| I_m + \boldsymbol{B}(\boldsymbol{E})\boldsymbol{X}\boldsymbol{X}'\boldsymbol{B}'(\boldsymbol{E}) + \boldsymbol{L}(\boldsymbol{E})\boldsymbol{L}'(\boldsymbol{E}) \Big|^{-\frac{r+m+n-1}{2}}. \end{aligned} \tag{4.6}$$

The SSP matrix of the error matrix $\boldsymbol{E}$ for the generalized multilinear model is obtained from the positive lower triangular matrix $\boldsymbol{L}(\boldsymbol{E})$ as follows:

$$\begin{aligned} \boldsymbol{L}(\boldsymbol{E})\boldsymbol{L}'(\boldsymbol{E}) &= \boldsymbol{L}(\boldsymbol{E})\boldsymbol{O}(\boldsymbol{E})\boldsymbol{O}'(\boldsymbol{E})\boldsymbol{L}'(\boldsymbol{E}) = \boldsymbol{C}(\boldsymbol{E})\boldsymbol{C}'(\boldsymbol{E}) \\ &= \boldsymbol{C}(\boldsymbol{E})\boldsymbol{D}(\boldsymbol{E})\boldsymbol{D}'(\boldsymbol{E})\boldsymbol{C}'(\boldsymbol{E}) \\ &= \{\boldsymbol{E} - \boldsymbol{B}(\boldsymbol{E})\boldsymbol{X}\}\{\boldsymbol{E} - \boldsymbol{B}(\boldsymbol{E})\boldsymbol{X}\}' = \boldsymbol{S}(\boldsymbol{E}). \end{aligned} \tag{4.7}$$

Substituting the relation (4.7) and utilizing the inverse Jacobian factor, $J\{\boldsymbol{S}(\boldsymbol{E}) \rightarrow \boldsymbol{L}(\boldsymbol{E})\} = |\boldsymbol{L}(\boldsymbol{E})|_\triangledown$, where $|\boldsymbol{L}(\boldsymbol{E})|_\triangledown$ is the decreasing determinant of $\boldsymbol{L}(\boldsymbol{E})$, and the relation $|\boldsymbol{S}(\boldsymbol{E})|^{\frac{m+1}{2}} = |\boldsymbol{L}(\boldsymbol{E})|_\triangle \times |\boldsymbol{L}(\boldsymbol{E})|_\triangledown$ in (4.6), the joint density of $\boldsymbol{B}(\boldsymbol{E})$ and $\boldsymbol{S}(\boldsymbol{E})$ is obtained as

$$\begin{aligned} p\Big(\boldsymbol{B}(\boldsymbol{E}),\, \boldsymbol{S}(\boldsymbol{E}) \Big| \boldsymbol{D}(\cdot)\Big) = K_4(\boldsymbol{D}) \ & \Big|\boldsymbol{S}(\boldsymbol{E})\Big|^{\frac{n-p-m-1}{2}} \\ \Big| I_m + \boldsymbol{B}(\boldsymbol{E})\boldsymbol{X}\boldsymbol{X}'\boldsymbol{B}'(\boldsymbol{E}) + & \boldsymbol{S}(\boldsymbol{E}) \Big|^{-\frac{r+m+n-1}{2}} \end{aligned} \tag{4.8}$$

where $K_4(\boldsymbol{D})$ is the appropriate normalizing constant.

Now the marginal density of the SSP matrix, $\boldsymbol{S}(\boldsymbol{E})$ is obtained from (4.8) by integrating out $\boldsymbol{B}(\boldsymbol{E})$ using the matrix-T integral. Thus we obtain the probability density function of $\boldsymbol{S}(\boldsymbol{E})$ as follows:

$$p\Big(\boldsymbol{S}(\boldsymbol{E}) \Big| \boldsymbol{D}(\boldsymbol{E})\Big) = K_5(\boldsymbol{D}) \ \Big| I_m + \boldsymbol{S}(\boldsymbol{E}) \Big|^{-\frac{r+m+n-1}{2}} \Big|\boldsymbol{S}(\boldsymbol{E})\Big|^{\frac{n-p-m-1}{2}} \tag{4.9}$$

where

$$K_5(\boldsymbol{D}) = B_m^{-1}\left(\frac{n-p}{2},\ \frac{r+m-1}{2}\right) = \left[\frac{\Gamma_m(\frac{n-p}{2})\Gamma_m(\frac{r+m-1}{2})}{\Gamma_m(\frac{r+m+n-p-1}{2})}\right]^{-1}$$

is the normalizing constant and is obtained by using the generalized beta integral of the second kind. The notation ${B_m}^{-1}(\cdot)$ stands for the inverse of the generalized beta function. The density of $\boldsymbol{S}(\boldsymbol{E})$ in (4.9) does not depend on $\boldsymbol{D}(\boldsymbol{E})$ and hence we can re-write it as follows:

$$p\Big(\boldsymbol{S}(\boldsymbol{E})\Big) = B_m^{-1}\left(\frac{n-p}{2},\ \frac{r+m-1}{2}\right)\Big|\boldsymbol{S}(\boldsymbol{E})\Big|^{\frac{n-p-m-1}{2}}\Big|I_m+\boldsymbol{S}(\boldsymbol{E})\Big|^{-\frac{r+m+n-1}{2}}. \quad (4.10)$$

The density in (4.10) gives the distribution of the SSP matrix for the realized but unobserved error matrix $\boldsymbol{E}$ for the generalized multilinear model with matrix-T error distribution. The density obtained in (4.10) is known as the generalized beta density (cf. Olkin, 1959). The degrees of freedom for the generalized beta density are $(n-p)$ and $(r+m-1)$. Khan (2000) provides an extension of the generalized beta distribution with a matrix argument.

To derive the distribution of the SSP matrix of the responses, $\boldsymbol{S}(\boldsymbol{Y})$, consider the transformation

$$\boldsymbol{C}(\boldsymbol{E}) = \boldsymbol{\Gamma}^{-1}\boldsymbol{C}(\boldsymbol{Y}),$$
$$\text{or equivalently } \boldsymbol{S}(\boldsymbol{E}) = \boldsymbol{\Gamma}^{-1}\boldsymbol{S}(\boldsymbol{Y})\boldsymbol{\Gamma}'^{-1}$$

in (4.10), the Jacobian of the transformation being $J\{\boldsymbol{S}(\boldsymbol{E}) \to \boldsymbol{S}(\boldsymbol{Y})\} = |\boldsymbol{\Gamma}^{-1}|^{m+1}$. (See, for instance, Deemer and Olkin (1951)). Further detail on the matrix calculus can be found in Magnus and Neudecker (1988).

The density function of $\boldsymbol{S}(\boldsymbol{Y})$ is then obtained as

$$p\Big(\boldsymbol{S}(\boldsymbol{Y})\Big|\boldsymbol{Y}\Big) = K_6\,\Big|\boldsymbol{S}(\boldsymbol{Y})\Big|^{\frac{n-p-m-1}{2}}\Big|I_m+\boldsymbol{\Sigma}^{-1}\boldsymbol{S}(\boldsymbol{Y})\Big|^{-\frac{r+m+n-p-1}{2}} \quad (4.11)$$

where $K_6 = |\boldsymbol{\Sigma}|^{\frac{n-p}{2}}B_m\left(\frac{n-p}{2},\ \frac{r+m-1}{2}\right)$ in which $B_m(\cdot)$ is the generalized beta function. The density in (4.11) can also be obtained directly from (4.8) by using the following transformations

$$\begin{aligned}\boldsymbol{B}(\boldsymbol{E}) &= \boldsymbol{\Gamma}^{-1}\{\boldsymbol{B}(\boldsymbol{Y})-\boldsymbol{\beta}\}\\ \boldsymbol{S}(\boldsymbol{E}) &= \boldsymbol{\Gamma}^{-1}\boldsymbol{S}(\boldsymbol{Y})\boldsymbol{\Gamma}'^{-1}\end{aligned} \quad (4.12)$$

with the Jacobian $J\{(\boldsymbol{B}(\boldsymbol{E}),\ \boldsymbol{S}(\boldsymbol{E})) \to (\boldsymbol{B}(\boldsymbol{Y}),\ \boldsymbol{S}(\boldsymbol{Y}))\} = |\boldsymbol{\Gamma}^{-1}|^{m+p+1}$, and then integrating out $\boldsymbol{B}(\boldsymbol{Y})$ from the joint p.d.f. of $\boldsymbol{B}(\boldsymbol{Y})$ and $\boldsymbol{S}(\boldsymbol{Y})$. For detail about the Jacobian of symmetric matrices see Henderson and Searle (1979). Fraser (1979, p.290) used a similar transformation to analyse a multilinear model with normal errors by using the structural method. The density function of $\boldsymbol{S}(\boldsymbol{Y})$, as given in (4.11), is the p.d.f. of the SSP matrix, $\boldsymbol{S}(\boldsymbol{Y})$ for the responses from a generalized multilinear model with matrix-T error distribution.

4.1 Distribution of the Regression Matrix

The distribution of the error regression matrix, $\boldsymbol{B}(\boldsymbol{E})$, can be derived from (4.8) by integrating out $\boldsymbol{S}(\boldsymbol{E})$ by using the generalised beta integral of the second kind as follows:

$$p\Big(\boldsymbol{B}(\boldsymbol{E})\Big|\boldsymbol{D}\Big) = K_7(\boldsymbol{D}) \left|I_m + \boldsymbol{B}(\boldsymbol{E})\boldsymbol{X}\boldsymbol{X}'\boldsymbol{B}'(\boldsymbol{E})\right|^{-\frac{r+m+p-1}{2}} \tag{4.13}$$

where the normalizing constant

$$K_7(\boldsymbol{D}) = \frac{|XX'|^{\frac{m}{2}}\Gamma_m(\frac{r+m+p-1}{2})}{(\pi)^{\frac{mp}{2}}\Gamma_m(\frac{r+m-1}{2})}$$

is obtained by the matrix-T integration. Note that the normalizing constant does not depend on $\boldsymbol{D}$, and hence the conditional distribution is the same as the unconditional distribution.

Now the distribution of $\boldsymbol{B}(\boldsymbol{Y})$ is found by applying the transformation

$$\boldsymbol{B}(\boldsymbol{E}) = \boldsymbol{\Gamma}^{-1}\{\boldsymbol{B}(\boldsymbol{Y}) - \boldsymbol{\beta}\}$$

as follows:

$$p\Big(\boldsymbol{B}(\boldsymbol{Y})|\boldsymbol{D}\Big) = K_7(\boldsymbol{D})\ |\boldsymbol{\Sigma}|^{-\frac{p}{2}}\left|I_m + \{\boldsymbol{B}(\boldsymbol{Y}) - \boldsymbol{\beta}\}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{X}'\{\boldsymbol{B}(\boldsymbol{Y}) - \boldsymbol{\beta}\}'\right|^{-\frac{r+m+p-1}{2}}. \tag{4.14}$$

Both the distributions of $\boldsymbol{S}(\boldsymbol{Y})$ and $\boldsymbol{B}(\boldsymbol{Y})$ depend on the original degrees of freedom parameter, r of the matrix-T distribution.

5. THE PREDICTION DISTRIBUTIONS

Consider $n' \geq 1$ future responses from the generalized multilinear model as defined in (2.2) and (2.3):

$$\boldsymbol{Y}_f = \boldsymbol{\beta}\boldsymbol{X}_f + \boldsymbol{\Gamma}\boldsymbol{E}_f \tag{5.1}$$

where $\boldsymbol{X}_f$ is a $p \times n'$ dimensional design matrix of the future values of the p regressors; $\boldsymbol{E}_f$ is an $m \times n'$ dimensional matrix of future errors associated with the future response matrix $\boldsymbol{Y}_f$ of the same order; and $\boldsymbol{\beta}$ and $\boldsymbol{\Gamma}$ are the regression and scale parameter matrices as defined in (2.2).

Assuming that $\boldsymbol{E}_f$ has the same distribution as $\boldsymbol{E}$, the joint density function of the realized and the future errors can be written as

$$p\Big(\boldsymbol{E}, \boldsymbol{E}_f\Big) = \frac{\Gamma_m(\frac{r+m+n+n'-1}{2})}{\pi^{\frac{m(n+n')}{2}}\Gamma_m(\frac{r+m-1}{2})}\left|I_m + \boldsymbol{E}\boldsymbol{E}' + \boldsymbol{E}_f\boldsymbol{E}_f'\right|^{-\frac{r+m+n+n'-1}{2}} \tag{5.2}$$

where r is the number of degrees of freedom.

Following the arguments used in the previous section, we define the following statistics in terms of the future error and design matrices

$$\boldsymbol{B}_f(\boldsymbol{E}_f) = \boldsymbol{E}_f\boldsymbol{X}_f'\left(\boldsymbol{X}_f\boldsymbol{X}_f'\right)^{-1} \text{ and } \boldsymbol{C}_f(\boldsymbol{E}_f) = [\boldsymbol{E}_f - \{\boldsymbol{B}_f(\boldsymbol{E}_f)\boldsymbol{X}_f][\boldsymbol{D}(\boldsymbol{E}_f)]^{-1},$$

which gives,

$$\boldsymbol{E}_f = \boldsymbol{B}_f(\boldsymbol{E}_f)\boldsymbol{X}_f + \boldsymbol{C}_f(\boldsymbol{E}_f)\boldsymbol{D}(\boldsymbol{E}_f). \tag{5.3}$$

Therefore, we can write,

$$\boldsymbol{E}_f\boldsymbol{E}_f' = \boldsymbol{B}_f(\boldsymbol{E}_f)\boldsymbol{X}_f\boldsymbol{X}_f'\boldsymbol{B}_f'(\boldsymbol{E}_f) + \boldsymbol{C}_f(\boldsymbol{E}_f)\boldsymbol{C}_f'(\boldsymbol{E}_f) \tag{5.4}$$

and

$$\boldsymbol{S}_f(\boldsymbol{E}_f) = \boldsymbol{C}_f(\boldsymbol{E}_f)\boldsymbol{C}_f'(\boldsymbol{E}_f),$$

as the SSP matrix for the future error variables associated with the unobserved future response matrix $\boldsymbol{Y}_f$.

5.1 Distribution of the Future Regression Matrix

In this sub-section we derive the prediction distribution of the future regression matrix, conditional on the observed responses. The joint density function of the error statistics $\boldsymbol{B}(\boldsymbol{E})$, $\boldsymbol{S}(\boldsymbol{E})$, $\boldsymbol{B}_f(\boldsymbol{E}_f)$ and $\boldsymbol{S}_f(\boldsymbol{E}_f)$, for given $\boldsymbol{D}$, is derived from (5.2) by applying the properties of invariant differentials, as follows:

$$\begin{aligned} p\left(\boldsymbol{B}(\boldsymbol{E}),\boldsymbol{S}(\boldsymbol{E}),\boldsymbol{B}_f(\boldsymbol{E}_f),\boldsymbol{S}_f(\boldsymbol{E}_f)\middle|\boldsymbol{D}\right) &= \Psi_1 \times \left|\boldsymbol{S}(\boldsymbol{E})\right|^{\frac{n-m-p-1}{2}}\left|\boldsymbol{S}_f(\boldsymbol{E}_f)\right|^{\frac{n'-m-p-1}{2}} \\ \times \left|I_m + g_1\left(\boldsymbol{B},\boldsymbol{X}\right) + \boldsymbol{S}(\boldsymbol{E}) + g_2\left(\boldsymbol{B}_f,\boldsymbol{X}_f\right) + \boldsymbol{S}_f(\boldsymbol{E}_f)\right|^{-\frac{r+m+n+n'-1}{2}} \end{aligned} \tag{5.5}$$

where $g_1\left(\boldsymbol{B},\boldsymbol{X}\right) = \boldsymbol{B}(\boldsymbol{E})\boldsymbol{X}\boldsymbol{X}'\boldsymbol{B}'(\boldsymbol{E})$ and $g_2\left(\boldsymbol{B}_f,\boldsymbol{X}_f\right) = \boldsymbol{B}_f(\boldsymbol{E}_f)\boldsymbol{X}_f\boldsymbol{X}_f'\boldsymbol{B}_f'(\boldsymbol{E}_f)$ and Ψ_1 is the normalizing constant.

The structural relation of the model yields

$$\boldsymbol{B}(\boldsymbol{E}) = \boldsymbol{\Sigma}^{-\frac{1}{2}}[\boldsymbol{B}(\boldsymbol{Y}) - \boldsymbol{\beta}] \text{ and } \boldsymbol{S}(\boldsymbol{E}) = \boldsymbol{\Sigma}^{-1}\boldsymbol{S}(\boldsymbol{Y})$$

where

$$\begin{aligned} &\boldsymbol{B}(\boldsymbol{Y}) = \boldsymbol{Y}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{X}')^{-1}, \\ &\boldsymbol{S}(\boldsymbol{Y}) = [\boldsymbol{Y} - \boldsymbol{B}(\boldsymbol{Y})\boldsymbol{X}][\boldsymbol{Y} - \boldsymbol{B}(\boldsymbol{Y})\boldsymbol{X}]' \text{ and} \\ &\boldsymbol{\Gamma}\boldsymbol{\Gamma}' = \boldsymbol{\Sigma}. \end{aligned}$$

The joint distribution of $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$, $\boldsymbol{B}_f(\boldsymbol{E}_f)$, and $\boldsymbol{S}_f(\boldsymbol{E}_f)$ is then obtained by using the Jacobian of the transformation,

$$J\left\{[\boldsymbol{B}(\boldsymbol{E}),\boldsymbol{S}(\boldsymbol{E})] \to [\boldsymbol{\beta},\boldsymbol{\Sigma}]\right\} = \left|\boldsymbol{S}(\boldsymbol{Y})\right|^{\frac{m+1}{2}}\left|\boldsymbol{\Sigma}^{-1}\right|^{\frac{p+2m+2}{2}},$$

as follows:

$$
\begin{aligned}
p\Big(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{B}_f(\boldsymbol{E}_f), \boldsymbol{S}_f(\boldsymbol{E}_f)\Big|\boldsymbol{D}\Big) &= \Psi_2 \times \Big|\boldsymbol{S}\Big|^{\frac{n-p}{2}} \\
&\Big|\boldsymbol{S}_f(\boldsymbol{Y}_f)\Big|^{\frac{n'-m-p-1}{2}} \Big|\boldsymbol{\Sigma}\Big|^{-\frac{n+m+1}{2}} \\
&\times \Big|I_m + \boldsymbol{\Sigma}^{-1}\Big\{(\boldsymbol{B}-\boldsymbol{\beta})\boldsymbol{X}\boldsymbol{X}'(\boldsymbol{B}-\boldsymbol{\beta})' + \boldsymbol{S}\Big\} \\
&+ \boldsymbol{B}_f(\boldsymbol{E}_f)\boldsymbol{X}_f\boldsymbol{X}_f'\boldsymbol{B}_f'(\boldsymbol{E}_f) + \boldsymbol{S}_f(\boldsymbol{E}_f)\Big|^{-\frac{r+m+n+n'-1}{2}}
\end{aligned} \tag{5.6}
$$

where $\boldsymbol{B} = \boldsymbol{B}(\boldsymbol{Y})$ and $\boldsymbol{S} = \boldsymbol{S}(\boldsymbol{Y})$.

Now, since

$$
\boldsymbol{B}_f(\boldsymbol{E}_f) = \boldsymbol{\Sigma}^{-\frac{1}{2}}[\boldsymbol{B}_f(\boldsymbol{Y}_f) - \boldsymbol{\beta}] \text{ and } \boldsymbol{S}_f(\boldsymbol{E}_f) = \boldsymbol{\Sigma}^{-1}\boldsymbol{S}_f(\boldsymbol{Y}_f)
$$

where

$$
\begin{aligned}
&\boldsymbol{B}_f(\boldsymbol{Y}_f) = \boldsymbol{Y}_f\boldsymbol{X}_f'(\boldsymbol{X}_f\boldsymbol{X}_f')^{-1}, \text{ and} \\
&\boldsymbol{S}_f(\boldsymbol{Y}_f) = [\boldsymbol{Y}_f - \boldsymbol{B}_f(\boldsymbol{Y}_f)\boldsymbol{X}_f][\boldsymbol{Y}_f - \boldsymbol{B}_f(\boldsymbol{Y}_f)\boldsymbol{X}_f]',
\end{aligned}
$$

the joint density of $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$, $\boldsymbol{B}_f(\boldsymbol{Y}_f)$ and $\boldsymbol{S}_f(\boldsymbol{Y}_f)$ is obtained as

$$
\begin{aligned}
p\Big(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{B}_f, \boldsymbol{S}_f\Big|\boldsymbol{D}\Big) &= \Psi_3(\cdot) \times \Big|\boldsymbol{S}\Big|^{\frac{n-p}{2}} \\
&\times \Big|\boldsymbol{S}_f(\boldsymbol{Y}_f)\Big|^{\frac{n'-m-p-1}{2}} \Big|\boldsymbol{\Sigma}\Big|^{-\frac{n+n'+m+1}{2}} \\
&\times \Big|I_m + \boldsymbol{\Sigma}^{-1}\Big\{\Big(\boldsymbol{B}-\boldsymbol{\beta}\Big)\boldsymbol{X}\boldsymbol{X}'\Big(\boldsymbol{B}-\boldsymbol{\beta}\Big)' \\
&\quad + \boldsymbol{S} + \Big(\boldsymbol{B}_f - \boldsymbol{\beta}\Big)\boldsymbol{X}_f\boldsymbol{X}_f'\Big(\boldsymbol{B}_f - \boldsymbol{\beta}\Big)' + \boldsymbol{S}_f\Big\}\Big|^{-\frac{r+m+n+n'-1}{2}}
\end{aligned} \tag{5.7}
$$

where $\boldsymbol{B}_f = \boldsymbol{B}_f(\boldsymbol{Y}_f)$ and $\boldsymbol{S}_f = \boldsymbol{S}_f(\boldsymbol{Y}_f)$.

Note that the Jacobian of the transformation is

$$
J\Big\{[\boldsymbol{B}_f(\boldsymbol{E}_f), \boldsymbol{S}_f(\boldsymbol{E}_f)] \to [\boldsymbol{B}_f(\boldsymbol{Y}_f), \boldsymbol{S}_f(\boldsymbol{Y}_f)]\Big\} = \Big|\boldsymbol{\Sigma}^{-1}\Big|^{\frac{p+m+1}{2}}.
$$

To evaluate the normalizing constant $\Psi_3(\cdot)$, note the following.

Let

$$
\begin{aligned}
I_{\boldsymbol{\Sigma}} &= \int_{\boldsymbol{\Sigma}} p\Big(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{B}_f,\ bS_f|\boldsymbol{D}\Big)d\boldsymbol{\Sigma} \\
&= \Big|\boldsymbol{S}_f\Big|^{\frac{n'-m-p-1}{2}} \int_{\boldsymbol{\Sigma}} \Big|\boldsymbol{\Sigma}^{-1}\Big|^{-\frac{n+n'+m+1}{2}} \\
&\qquad \Big|I_m + \boldsymbol{\Sigma}^{-1}\boldsymbol{Q}\Big|^{-\frac{r+m+n+n'-1}{2}} d\boldsymbol{\Sigma}
\end{aligned} \tag{5.8}
$$

where

$$\boldsymbol{Q} = \Big(\boldsymbol{B} - \boldsymbol{\beta}\Big)\boldsymbol{X}\boldsymbol{X}'\Big(\boldsymbol{B} - \boldsymbol{\beta}\Big)' + \boldsymbol{S} + \Big(\boldsymbol{B}_f - \boldsymbol{\beta}\Big)\boldsymbol{X}_f\boldsymbol{X}_f'\Big(\boldsymbol{B}_f - \boldsymbol{\beta}\Big)' + \boldsymbol{S}_f.$$

Putting $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Lambda}$, we have

$$d\boldsymbol{\Sigma} = \Big|\boldsymbol{\Lambda}^{-1}\Big|^{m+1} d\boldsymbol{\Lambda}.$$

Therefore,

$$\begin{aligned} I_{\boldsymbol{\Sigma}} &= \Big|\boldsymbol{S}_f\Big|^{\frac{n'-m-p-1}{2}} \int_{\boldsymbol{\Lambda}} \Big|\boldsymbol{\Lambda}^{-1}\Big|^{-\frac{n+n'-m-1}{2}} \\ &\quad \times \Big|I_m + \boldsymbol{\Lambda}\boldsymbol{Q}\Big|^{-\frac{r+m+n+n'-1}{2}} d\boldsymbol{\Lambda} \\ &= \boldsymbol{B}_m\Big(\frac{n+n'}{2}, \frac{r+m-1}{2}\Big)\Big|\boldsymbol{S}_f\Big|^{\frac{n'-m-p-1}{2}} \Big|\boldsymbol{Q}\Big|^{\frac{n'-m-p-1}{2}}. \end{aligned} \tag{5.9}$$

Now, the terms involving $\boldsymbol{\beta}$ in $\boldsymbol{Q}$ can be expressed as follows:

$$\begin{aligned} &\Big(\boldsymbol{B} - \boldsymbol{\beta}\Big)\boldsymbol{X}\boldsymbol{X}'\Big(\boldsymbol{B} - \boldsymbol{\beta}\Big)' + \Big(\boldsymbol{B}_f - \boldsymbol{\beta}\Big)\boldsymbol{X}_f\boldsymbol{X}_f'\Big(\boldsymbol{B}_f - \boldsymbol{\beta}\Big)' = \\ &\qquad \Big(\boldsymbol{\beta} - \boldsymbol{F}\boldsymbol{A}^{-1}\Big)\boldsymbol{A}\Big(\boldsymbol{\beta} - \boldsymbol{F}\boldsymbol{A}^{-1}\Big)' + \Big(\boldsymbol{B}_f - \boldsymbol{B}\Big)\boldsymbol{F}^{-1}\Big(\boldsymbol{B}_f - \boldsymbol{B}\Big)' \end{aligned}$$

where

$$\boldsymbol{F} = \boldsymbol{B}\boldsymbol{X}\boldsymbol{X}' + \boldsymbol{B}_f\boldsymbol{X}_f\boldsymbol{X}_f', \quad \boldsymbol{A} = \boldsymbol{X}\boldsymbol{X}' + \boldsymbol{X}_f\boldsymbol{X}_f', \quad \text{and } \boldsymbol{H} = [\boldsymbol{X}\boldsymbol{X}']^{-1} + [\boldsymbol{X}_f\boldsymbol{X}_f']^{-1}.$$

Then, let

$$\begin{aligned} I_{\boldsymbol{\Sigma}\boldsymbol{\beta}} &= \int_{\boldsymbol{\beta}} I_{\boldsymbol{\Sigma}}\ d\boldsymbol{\beta} \\ &= \boldsymbol{B}_m\Big(\frac{n+n'}{2}, \frac{r+m-1}{2}\Big)\Big|\boldsymbol{S}_f\Big|^{\frac{n'-m-p-1}{2}} \\ &\quad \times \int_{\boldsymbol{\beta}} \Big|\Big(\boldsymbol{B}_f - \boldsymbol{B}\Big)H^{-1}\Big(\boldsymbol{B}_f' - \boldsymbol{B}\Big) + \boldsymbol{S} + \boldsymbol{S}_f + g\Big(\boldsymbol{\beta}, \boldsymbol{A}\Big)\Big|^{-\frac{n+n'}{2}} d\boldsymbol{\beta} \\ &= \frac{(\pi)^{\frac{mp}{2}}\Gamma_m(\frac{r+m-1}{2})\Gamma_m(\frac{n+n'-p}{2})}{\Big|\boldsymbol{A}\Big|^{\frac{m}{2}}\Gamma_m(\frac{r+m_n+n'-1}{2})}\Big|\boldsymbol{S}_f\Big|^{\frac{n'-m-p-1}{2}} \\ &\quad \times \Big|\Big(\boldsymbol{B}_f - \boldsymbol{B}\Big)H^{-1}\Big(\boldsymbol{B}_f' - \boldsymbol{B}\Big) + \boldsymbol{S} + \boldsymbol{S}_f + g\Big(\boldsymbol{\beta}, \boldsymbol{A}\Big)\Big|^{-\frac{n+n'-p}{2}} d\boldsymbol{\beta} \end{aligned} \tag{5.10}$$

where

$$g\Big(\boldsymbol{\beta}, \boldsymbol{A}\Big) = \Big(\boldsymbol{\beta} - \boldsymbol{F}\boldsymbol{A}^{-1}\Big)\boldsymbol{A}\Big(\boldsymbol{\beta} - \boldsymbol{F}\boldsymbol{A}^{-1}\Big)'.$$

In the same way, let

$$
\begin{aligned}
I_{\boldsymbol{\Sigma}\boldsymbol{\beta}\boldsymbol{B}_f} &= \int_{\boldsymbol{B}_f} I_{\boldsymbol{\Sigma}\beta}\ d\boldsymbol{B}_f \\
&= \frac{(\pi)^{\frac{mp}{2}}\Gamma_m(\frac{r+m-1}{2})\Gamma_m(\frac{n+n'-p}{2})}{\left|\boldsymbol{A}\right|^{\frac{m}{2}}\Gamma_m(\frac{r+m_n+n'-1}{2})}\left|\boldsymbol{S}_f\right|^{\frac{n'-m-p-1}{2}} \\
&\quad\times\int_{\boldsymbol{B}_f}\left|\Big(\boldsymbol{B}_f-\boldsymbol{B}\Big)H^{-1}\Big(\boldsymbol{B}_f'-\boldsymbol{B}\Big)+\boldsymbol{S}+\boldsymbol{S}_f+g\Big(\boldsymbol{\beta},\boldsymbol{A}\Big)\right|^{-\frac{n+n'-p}{2}}d\boldsymbol{B}_f \\
&= \frac{(\pi)^{mp}\Gamma_m(\frac{r+m-1}{2})\Gamma_m(\frac{n+n'-2p}{2})}{\left|\boldsymbol{A}\right|^{\frac{m}{2}}\left|\boldsymbol{H}\right|^{-\frac{m}{2}}\Gamma_m(\frac{r+m_n+n'-1}{2})}\left|\boldsymbol{S}_f\right|^{\frac{n'-m-p-1}{2}}\left|\boldsymbol{S}+\boldsymbol{S}_f\right|^{-\frac{n+n'-2p}{2}}.
\end{aligned}
\tag{5.11}
$$

Finally, let

$$
\begin{aligned}
I_{\boldsymbol{\Sigma}\boldsymbol{\beta}\boldsymbol{B}_f\boldsymbol{S}_f} &= \int_{\boldsymbol{S}_f} I_{\boldsymbol{\Sigma}\boldsymbol{\beta}\boldsymbol{B}_f}\ d\boldsymbol{S}_f \\
&= \frac{(\pi)^{mp}\Gamma_m(\frac{r+m-1}{2})\Gamma_m(\frac{n+n'-2p}{2})}{\left|\boldsymbol{A}\right|^{\frac{m}{2}}\left|\boldsymbol{H}\right|^{-\frac{m}{2}}\Gamma_m(\frac{r+m_n+n'-1}{2})} \\
&\quad\times\int_{\boldsymbol{S}_f}\left|\boldsymbol{S}_f\right|^{\frac{n'-m-p-1}{2}}\left|\boldsymbol{S}+\boldsymbol{S}_f\right|^{-\frac{n+n'-2p}{2}}d\boldsymbol{S}_f \\
&= \frac{(\pi)^{mp}\Gamma_m(\frac{r+m-1}{2})\Gamma_m(\frac{n-p}{2})\Gamma_m(\frac{n'-p}{2})\left|\boldsymbol{S}\right|^{-\frac{n-p}{2}}}{\left|\boldsymbol{A}\right|^{\frac{m}{2}}\left|\boldsymbol{H}\right|^{-\frac{m}{2}}\Gamma_m(\frac{r+m_n+n'-1}{2})}.
\end{aligned}
\tag{5.12}
$$

Thus, the normalizing constant becomes,

$$
\Psi_3(\cdot) = \frac{\left|\boldsymbol{A}\right|^{\frac{m}{2}}\left|\boldsymbol{H}\right|^{-\frac{m}{2}}\Gamma_m(\frac{r+m_n+n'-1}{2})}{(\pi)^{mp}\Gamma_m(\frac{r+m-1}{2})\Gamma_m(\frac{n-p}{2})\Gamma_m(\frac{n'-p}{2})}\left|\boldsymbol{S}\right|^{-\frac{n-p}{2}}. \tag{5.13}
$$

The marginal density of $\boldsymbol{\beta}$, $\boldsymbol{B}_f$ and $\boldsymbol{S}_f$ is derived by integrating out $\boldsymbol{\Sigma}$ from (5.7). Thus, we have,

$$
\begin{aligned}
p\Big(\boldsymbol{\beta},\boldsymbol{B}_f,\boldsymbol{S}_f\Big) &= \Psi_4\times\left|\boldsymbol{S}\right|^{\frac{n-p}{2}}\left|\boldsymbol{S}_f\right|^{\frac{n'-m-p-1}{2}}\Big|\boldsymbol{S}+\Big(\boldsymbol{\beta}-\boldsymbol{F}\boldsymbol{A}^{-1}\Big)\boldsymbol{A} \\
&\quad\times\Big(\boldsymbol{\beta}-\boldsymbol{F}\boldsymbol{A}^{-1}\Big)'+\Big(\boldsymbol{B}_f-\boldsymbol{B}\Big)\boldsymbol{H}^{-1}\Big(\boldsymbol{B}_f-\boldsymbol{B}\Big)+\boldsymbol{S}_f\Big|^{-\frac{n+n'}{2}}
\end{aligned}
\tag{5.14}
$$

where Ψ_4 is the normalizing constant.

Similarly, the marginal density of $\boldsymbol{B}_f$ and $\boldsymbol{S}_f$ is obtained by integrating out $\boldsymbol{\beta}$ over $\Re^{mp}$ from (5.14). This gives

$$
\begin{aligned}
p\Big(\boldsymbol{B}_f,\boldsymbol{S}_f\Big) &= \Psi_5\times\left|\boldsymbol{S}\right|^{\frac{n-p}{2}}\left|\boldsymbol{S}_f\right|^{\frac{n'-m-p-1}{2}} \\
&\quad\times\left|\boldsymbol{S}+\boldsymbol{S}_f+\Big(\boldsymbol{B}_f-\boldsymbol{B}\Big)\boldsymbol{H}^{-1}\Big(\boldsymbol{B}_f-\boldsymbol{B}\Big)\right|^{-\frac{n+n'-p}{2}}
\end{aligned}
\tag{5.15}
$$

where

$$\Psi_5 = \frac{\Gamma_m(\frac{n+n'-p}{2})}{(\pi)^{\frac{mp}{2}}\Gamma_m(\frac{n-p}{2})\Gamma_m(\frac{n'-p}{2})}\Big|\boldsymbol{H}\Big|^{-\frac{m}{2}}$$

is the normalizing constant.

The prediction distribution of the future regression matrix, $\boldsymbol{B}_f = \boldsymbol{B}_f(\boldsymbol{Y}_f)$, can now be obtained by integrating out $\boldsymbol{S}_f$ from (5.15). The integration yields

$$p\Big(\boldsymbol{B}_f\Big|\boldsymbol{Y}\Big) = \Psi_6 \times \Big|\boldsymbol{S} + \Big(\boldsymbol{B}_f - \boldsymbol{B}\Big)\boldsymbol{H}^{-1}\Big(\boldsymbol{B}_f - \boldsymbol{B}\Big)'\Big|^{-\frac{n}{2}} \tag{5.16}$$

where $\Psi_6 = \Psi_4 \times \boldsymbol{B}_m\Big(\frac{n'-p}{2}, \frac{n}{2}\Big)\Big|\boldsymbol{S}\Big|^{\frac{n-p}{2}}$. On simplification we get

$$\Psi_6 = \frac{\Gamma_m(\frac{n}{2})}{(\pi)^{\frac{mp}{2}}\Gamma_m(\frac{n-p}{2})\Big|\boldsymbol{H}\Big|^{\frac{m}{2}}}.$$

The prediction distribution of $\boldsymbol{B}_f$ can be written in the usual matrix-T form as follows:

$$\begin{aligned} p\Big(\boldsymbol{B}_f\Big|\boldsymbol{Y}\Big) &= \Psi_6 \times \Big|\boldsymbol{S}\Big|^{-\frac{n}{2}} \\ &\quad \times \Big|\boldsymbol{I}_m + \Big(\boldsymbol{B}_f - \boldsymbol{B}\Big)\Big[\boldsymbol{S}\boldsymbol{H}\Big]^{-1}\Big(\boldsymbol{B}_f - \boldsymbol{B}\Big)'\Big|^{-\frac{n}{2}} \end{aligned} \tag{5.17}$$

in which $n > p + m - 1$. The density in (5.17) is a matrix-T density. Therefore, the prediction distribution of the future regression matrix, $\boldsymbol{B}_f$, conditional on the observed responses, is a matrix-T distribution of dimension $m \times p$, and $(n-p-m+1)$ degrees of freedom. It is observed that unlike the distribution of $\boldsymbol{B}(\boldsymbol{Y})$ the prediction distribution of $\boldsymbol{B}_f$ does not depend on the number of degrees of freedom, r of the model.

5.2 Distribution of the Future SSP Matrix

The prediction distribution of the future SSP matrix, $\boldsymbol{S}_f(\boldsymbol{Y}_f)$, based on the future responses, $\boldsymbol{Y}_f$, conditional on the observed responses, $\boldsymbol{Y}$, is obtained by integrating out $\boldsymbol{B}_f$ from (5.15) as follows:

$$p\Big(\boldsymbol{S}_f(\boldsymbol{Y}_f)\Big|\boldsymbol{Y}\Big) = \Psi_7 \times \Big|\boldsymbol{S}_f(\boldsymbol{Y}_f)\Big|^{\frac{n'-p-m-1}{2}}\Big|\boldsymbol{S} + \boldsymbol{S}_f(\boldsymbol{Y}_f)\Big|^{-\frac{n+n'-2p}{2}}. \tag{5.18}$$

The density in (5.18) can be written in the usual matrix-T form as follows:

$$p\Big(\boldsymbol{S}_f\Big|\boldsymbol{Y}\Big) = \Psi_7 \times \Big|\boldsymbol{S}\Big|^{\frac{n'-p}{2}}\Big|\boldsymbol{S}_f\Big|^{\frac{n'-p-m-1}{2}}\Big|\boldsymbol{I}_m + \boldsymbol{S}^{-1}\boldsymbol{S}_f\Big|^{-\frac{n+n'-2p}{2}} \tag{5.19}$$

where $\Psi_7 = \frac{\Gamma_m(\frac{n}{2})|\boldsymbol{S}|^{-\frac{n}{2}}}{(\pi)^{\frac{mp}{2}}\Gamma_m(\frac{n-p}{2})|\boldsymbol{H}|^{\frac{m}{2}}}$.

This is the prediction distribution of the SSP matrix based on the future response $\boldsymbol{Y}_f$, conditional on the observed responses, from a generalized multilinear model with matrix-T error variable. The density in (5.19) is a modified form of generalized beta density with $(n' - p)$ and $(n - p)$ degrees of freedom. Once again note that unlike the distribution of $\boldsymbol{S}(\boldsymbol{Y})$ the prediction distribution of the future SSP matrix does not depend on the degrees of freedom, r of the model.

ACKNOWLEDGEMENTS

The author thankfully acknowledges a valuable suggestion of a reviewer that has contributed to the improvement of the quality of the results of the paper.

REFERENCES

Aitchison, J. and Dunsmore, I.R. (1975). Statistical Prediction Analysis. Cambridge University Press, Cambridge.

Box, J.E.P. and Tiao, G.C. (1992). Bayesian Inference in Statistical Analysis. Addison Wesley, London.

Deemer, W.L. and Olkin, I. (1951). The Jacobian of certain matrix transformations useful in multivariate analysis. *Biometrika*, **38** 345-367.

Dickey, J.M. (1967). Matrix variate generalization of the multivariate t-distribution. *Ann. Math. Statist.,* **38** 511-518.

Eaton, M.L. (1983). Multivariate Statistics - A Vector Space Approach. Wiley, New York.

Fraser, D.A.S. (1968a). The conditional Wishart: normal and non-normal. *Ann. Math. Statist.*, **39,** 593–605.

_______________ (1968b). The Structure of Inference. Wiley, New York.

_______________ (1979). Inference and Linear Models. McGraw-Hill, New York.

Fraser, D.A.S. and Ng, K.W. (1980). Multivariate regression analysis with spherical error. Multivariate Analysis-V. P.R. Krishnaiah, ed., North-Holland Publishing Co., New York, 369-386.

Geisser, S. (1993). Predictive Inference : An Introduction, Chapman and Hall, London

Haq, M.S. and Rinco, S.(1976). β-expectation tolerance regions for a generalised multivariate model with normal error variables. *Jou. Multiv. Analysis*, **6**, 414-432.

Henderson, H.V. and Searle, S.R.(1979). Vec and Vech operators for matrices, with some uses in Jacobians and multivariate statistics. *The Canadian Journal of Statistics*, **7**, 65-81.

Hsu, P.L. (1939). A new proof of the joint product moment distribution. *Proc. of Cambridge Phil. Soc.*, **35**, 336-338.

Jambunathan, M.V. (1965). A quick method of deriving Wishart's distribution. *Current Series*, **34**, 78.

Johnson, N.L. and Kotz, S. (1972). Distributions in Statistics : Continuous Multivariate Distributions. Wiley, New York.

Khan, S. (2000). An extension of the generalized beta integral with matrix argument, *Pak. Jou. Statist.*, Saleh-Aly Special Edition, 2000, **16(2)**, 163-167.

Khan, S. and Haq, M.S. (1994). Prediction distribution for the multilinear model with matrix-T error: The case of spherical error distribution. *Journal of Applied Statistical Sciences*, **1**, 239-250.

Magnus, J.R. and Neudecker, H. (1988). Matric Differential Calculus with Application in Statistics and Econometrics. Wiley, New York.

Olkin, I. (1959). A class of integral identities with matrix argument. *Duke Mathematical Journal*, **26,** 207–214.

Press, S.J. (1986). Applied Multivariate Analysis. Holt, Rinehart and Winston, Inc., New York.

Prucha, I.R. and Kelejian, H.H. (1984). The structure of simultaneous equation estimators : A generalization towards non-normal disturbances. *C*, **52**, 721-736.

Svendrup, E. (1947). Derivation of the Wishart distribution of the second order sample moments by straight forward integration of a multiple integral. *Skandinavisk Aktuarietidskrift*, **30**, 151-166.

Wishart, J. (1928). The generalized product moment distribution in sample from a normal multivariate population. *Biometrika*, **20**, 32-52.

Proof of Equation (4.1)

The error matrix $\boldsymbol{E}$ in $\mathcal{R}^{mn}$ can be viewed as a sequence of m vectors, $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_m$ in $\mathcal{R}^n$. Let $\mathcal{L}^+(\boldsymbol{X}, \boldsymbol{E})$ be a $(p+m)$-dimensional subspace spanned by the row vectors of $\boldsymbol{X}$ and $\boldsymbol{E}$. Project an arbitrary but fixed sequence of m linearly independent vectors onto $\mathcal{L}^+(\boldsymbol{X}, \boldsymbol{E})$ together with the same order as the positive orientation. Then orthogonalise the m vectors in that sequence to obtain orthogonalised vectors, $\boldsymbol{d}_1, \boldsymbol{d}_2, \cdots, \boldsymbol{d}_m$ that are also orthogonal to $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_p$. These provide a basis $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_p, \boldsymbol{d}_1, \boldsymbol{d}_2, \cdots, \boldsymbol{d}_m$ for $\mathcal{L}^+(\boldsymbol{X}, \boldsymbol{E})$ except for a set of measure zero for which the projections are linearly dependent. Note that $\boldsymbol{D}(\boldsymbol{E})$ can be written as the collection of the sequence of orthonormal vectors, $\boldsymbol{d}_1, \boldsymbol{d}_2, \cdots, \boldsymbol{d}_m$.

Using the same notations as in the paper, we can write the error matrix, $\boldsymbol{E}$, in the following way:

$$\boldsymbol{E} = \boldsymbol{B}(\boldsymbol{E})\boldsymbol{X} + \boldsymbol{C}(\boldsymbol{E})\boldsymbol{D}(\boldsymbol{E}) \tag{1}$$

and hence

$$\boldsymbol{E}\boldsymbol{E}' = \boldsymbol{B}(\boldsymbol{E})\boldsymbol{X}\boldsymbol{X}'\boldsymbol{B}'(\boldsymbol{E}) + \boldsymbol{C}(\boldsymbol{E})\boldsymbol{C}'(\boldsymbol{E}). \tag{2}$$

For the matrix-T model, the joint density function of the mn random elements of $\boldsymbol{E}$ is given by

$$p(\boldsymbol{E}) = \frac{\Gamma_m\left(\frac{r+m+n-1}{2}\right)}{\pi^{\frac{mn}{2}}\Gamma_m\left(\frac{r+m-1}{2}\right)}\left|I_m + \boldsymbol{E}\boldsymbol{E}'\right|^{-\frac{r+m+n-1}{2}}. \tag{3}$$

To derive the joint distribution of $\boldsymbol{B}(\boldsymbol{E})$ and $\boldsymbol{C}(\boldsymbol{E})$, conditional on $\boldsymbol{D}(\boldsymbol{E})$, we note the following argument regarding invariant differentials (cf. Fraser and Ng, 1980). As $\boldsymbol{E}$ varies in $\mathcal{R}^{mn}$, $\boldsymbol{D}(\boldsymbol{E})$ traces out smoothly the set of all m dimensional subspaces of $\mathcal{R}^{n-p}$, which is a copy of Grassman manifold $\mathcal{G}_{m,n-p}$. Let $d\boldsymbol{D}(\boldsymbol{E})$ denote the volume element of $\mathcal{G}_{m,n-p}$ orthogonal to the subspace $\mathcal{L}^+(\boldsymbol{X}, \boldsymbol{E})$. Hence the relationship between the volume elements of the error matrix, $\boldsymbol{E}$ and that of the $\boldsymbol{B}(\boldsymbol{E})$ and $\boldsymbol{C}(\boldsymbol{E})$, conditional on $\boldsymbol{D}(\boldsymbol{E})$, is

$$d\boldsymbol{E} = |\boldsymbol{X}\boldsymbol{X}'|^{\frac{m}{2}}|\boldsymbol{C}(\boldsymbol{E})|^{n-m-p}d\boldsymbol{B}(\boldsymbol{E})d\boldsymbol{C}(\boldsymbol{E})d\boldsymbol{D}(\boldsymbol{E}). \tag{4}$$

For further details, please refer to Fraser (1978, p.283-285). Now using the results in (2) and (4), the density in (3) becomes

$$\begin{aligned} p\Big(\boldsymbol{B}(\boldsymbol{E}),\ \boldsymbol{C}(\boldsymbol{E})\Big|\boldsymbol{D}(\boldsymbol{E})\Big) = K_1(\boldsymbol{D})\ &\Big|\boldsymbol{C}(\boldsymbol{E})\Big|^{n-p-m} \\ &\Big|I_m + \boldsymbol{B}(\boldsymbol{E})\boldsymbol{X}\boldsymbol{X}'\boldsymbol{B}'(\boldsymbol{E}) + \boldsymbol{C}(\boldsymbol{E})\boldsymbol{C}'(\boldsymbol{E})\Big|^{-\frac{r+m+n-1}{2}}. \end{aligned} \tag{5}$$

Hence the proof.