MULTIVARIATE REGRESSION ANALYSIS
WITH SPHERICAL ERROR

D. A. S. FRASER and Kai W. NG
University of Toronto, Toronto, Canada

In a recent publication Arnold Zellner [8] has examined the regression model using the multivariate student distribution for the error: tests and confidence regions are derived for the regression coefficients and error variance and are found to correspond closely to those under normal theory; flat and conjugate priors are also used for a Bayesian analysis. This paper examines tests and confidence methods for the multivariate regression model with error distribution given by the multivariate student or more generally by spherical error. The model is based on recent results concerning the identification of distribution form and the analysis uses the parameter and distribution factorization methods developed in D. A. S. Fraser and Jock MacKay [2] and organized for multivariate regression in D. A. S. Fraser and Kai W. Ng [4].

0. Introduction

The formation and validity of statistical models has been discussed in Fraser [5]. As part of this, the distribution form can in certain contexts be identified objectively and correspondingly the context requires the statistical model to use explicitly the distribution describing that distribution form. A model with an explicit (and objective) distribution for error or variation is a structural-type model (Fraser, [5]). This paper examines multivariate regression with a multivariate student or more generally spherical distribution for the explicit error distribution.

The determination of tests and confidence regions for structural models has been developed in D. A. S. Fraser and Jock MacKay [2], and organized for the multivariate regression model in D. A. S. Fraser and Kai W. Ng [4]. The use of the more extensive structural model leads to essentially unique tests and confidence regions for the primary parameters of the model: the order in which parameters are examined is the one available option; and the uniqueness is present without the introduction of the usual reduction principle such as sufficiency and conditionality. The factorization methods developed in the cited papers produced the unique tests and confidence regions. These methods are used in this paper to produce in a direct, straightforward manner the appropriate tests and confidence regions for multivariate regression with student or spherical error.
1. Preliminaries

Arnold Zellner [8] has examined the common regression model using a multivariate student distribution for the error. The multivariate-\(t\) distribution in standardized form can be written as

\[
f_\lambda(z) = \frac{\Gamma((\lambda + n)/2)}{\Gamma(\lambda/2)\pi^{n/2}} (1 + \lambda^{-1}z'z)^{-(\lambda+n)/2}\lambda^{-n/2}
\]

\[
= \frac{A_\lambda}{A_{\lambda+n}} (1 + \lambda^{-1}z'z)^{-(\lambda+n)/2}\lambda^{-n/2}
\]

(1)

where the use of \(A_\lambda = 2\pi^{\lambda/2}/\Gamma(\lambda/2)\) for the surface volume of a unit sphere in \(\mathbb{R}^\lambda\) leads to simpler and tidier formulas. Some arguments for the use of students distribution in special applications are given in Zellner [8]. When we proceed to the matrix generalization for the multivariate models we will avoid the nuisance of the scaling factor \(\sqrt{\lambda}\) in the density expression and use an alternative standardization given by

\[
f_\lambda(z) = \frac{A_\lambda}{A_{\lambda+n}} (1 + z'z)^{-(\lambda+n)/2};
\]

(2)

this corresponds to a reparametrization of the scale parameters in the full statistical model.

Some of the results for the multivariate student distributions extend easily to a general spherical distribution for the error. For this we take

\[
f(z) = g(z'z),
\]

(3)

where \(g(\cdot)\) is a non-negative function over the non-negative real numbers. This is the general form of a distribution that is invariant under rotation in the sample space \(\mathbb{R}^n\).

Consider the regression model

\[
y = X\beta + \sigma z,
\]

(4)

where \(y\) is an \(n \times 1\) vector of observable responses, \(X\) is an \(n \times r\) design matrix with full rank \(r < n\), \(\beta\) is an \(r \times 1\) vector of regression parameters,
\( \sigma > 0 \) is a scale parameter, and \( z \) is an \( n \times 1 \) vector of unobservable random variables representing the (standardized) error with distribution (1). Note that the shape parameter \( \lambda \) in (1) is generally unknown. For properties of the multivariate-\( t \) distribution, see Johnson and Kotz [6], chapter 37 and the references cited within.

Let \( y^0 \) be the observed response. Then we have an inference base

\[
\mathcal{I} = (\mathcal{M}, y^0),
\]

where the model \( \mathcal{M} \) consists of (4) and (1), delineating the generation of the data \( y^0 \) from the random error \( z \). For detailed discussion of an inference base, see Fraser [5]. We start the analysis with the basic question: what do we know about the unobservable \( z \) through the observable \( y^0 \)? Or, in other words, what portion of the unobservable \( z \) can be observed and what is the appropriate probabilistic description for the unobservable portion? To answer this, we use some transparent coordinates system for the sample space \( \mathbb{R}^n \) such that a part of the coordinates represents the observable portion of \( z \) and the remaining part represents the unobservable portion of \( z \). It should be emphasized that there are many choices of the transparent coordinates and that they are all one-to-one equivalent, giving the same probabilistic description.

For each vector \( z \) in the sample space \( \mathbb{R}^n \), let \( b(z) \) be the \( r \times 1 \) vector of regression coefficients of \( z \) on the space spanned by the columns of \( X \), \( d(z) \) be the unit residual vector, and \( s(z) \) the residual length:

\[
b(z) = (X'X)^{-1}X'z
\]
\[
s^2(z) = (z - Xb(z))'(z - Xb(z))
\]
\[
d(z) = s^{-1}(z)(z - Xb(z)).
\]

This amounts to writing the \( \mathbb{R}^n \) as the algebraic direct sum of the subspace \( \mathcal{L}(X) \) and its orthogonal complement \( \mathcal{L}^\perp(X) \), where in \( \mathcal{L}(X) \) we use the coordinates relative to the basis \( (x_1, x_2, \ldots, x_r) \) while in \( \mathcal{L}^\perp(X) \) we use \( (n-r) \) dimensional spherical coordinates with \( s(\cdot) \) as radius and \( d(\cdot) \) as directional unit vector. Thus the volume element in terms of the new coordinates is given by

\[
dz = |X'X|^{1/2} db \cdot s^{(n-r)-1} ds da,
\]

where \( da \) is the area element on the unit sphere in \( \mathcal{L}^\perp(X) \).
In terms of these coordinates for $\mathbb{R}^n$, the model $\mathcal{M}$ consisting of (4) and (1) is given by

$$
d(y) = d(z)$$

$$
b(y) = \beta + \sigma b(z)$$

$$
s(y) = \sigma s(z)$$

$$
f_\lambda(z) dz = \frac{A_\lambda}{A_{\lambda+n}} \frac{|X'X|^{1/2} s^{-r-1}}{(1+\lambda^{-1}(s^2 + b'X'Xb))^{(\lambda+n)/2}} d\mathbf{b} ds da,
$$

where we have used (7) and the relation

$$
z = X \mathbf{b}(z) + s(z) d(z)
$$

It is clear in this presentation of the model that the observable portion of the random $z$ is given by $d(z) = d(y)$ and that the appropriate probabilistic description for the unobservable portion of $z$ is the conditional distribution of $b(z), s(z)$ given $d(z)$. This leads to the quantities

$$
\begin{align*}
\frac{b(y) - \beta}{\sigma} &= b \\
\frac{s(y)}{\sigma} &= s
\end{align*}
$$

or

$$
\begin{align*}
\frac{b(y) - \beta}{s(y)} &= t \\
\frac{s(y)}{\sigma} &= s
\end{align*}
$$

(10)

to be examined conditionally given the observed $d(z) = d(y^0)$. Thus we have answered the basic question posed earlier concerning the random error $z$.

The shape parameter $\lambda$ is the only parameter involved in the error distribution (1) and the observed portion of the error $z$ is $d(z)$, so the probability of the observed $d(z) = d(y^0)$, as a function of $\lambda$, is the likelihood function of $\lambda$. With a given value of $\lambda$, the inference of $\beta$ and $\sigma$ are based on the conditional distribution of the quantities (10).

2. Inference of component parameters

Consider first the inference of the parameters in the order $\lambda$, $\sigma$, $\beta$. The factorization procedure from Fraser and MacKay [2] separates the error variable $z$ in the corresponding order $b$, $s$, $d$ with the marginal distribution
for \( \mathbf{d} \), conditional distribution for \( s \) given \( \mathbf{d} \) and the conditional distribution for \( \mathbf{b} \) given \( \mathbf{d}, s \). In this case, the error distribution (1) becomes

\[
f_{\lambda}(z) \, dz = \frac{A_{\lambda}}{A_{\lambda+n}} \frac{|X'X|^{(\lambda+n)/2}}{(1 + \lambda^{-1}(s^2 + \mathbf{b}'X'X\mathbf{b}))^{(\lambda+n)/2}} \, db \, ds \, da
\]

\[
= \frac{A_{\lambda+n-r}}{A_{\lambda+n-r+2}} \frac{(\lambda+s^2)^{-r/2}}{(1 + (\lambda+s^2)^{-1}\mathbf{b}'X'X\mathbf{b})^{\lambda+(n-r)+2/2}} \, db
\]

\[
\times \frac{A_{\lambda}A_{n-r}}{A_{\lambda+n-r}} \frac{s^{n-r-1} \lambda^{-(n-r)/2}}{(1 + s^2/\lambda)^{\lambda+(n-r)/2}} \, ds
\]

\[
\times \frac{1}{A_{n-r}} \, da; \tag{13}
\]

and the spherical error distribution (3) becomes

\[
g(z'z) \, dz = h^{-1}(s^2)g(s^2 + \mathbf{b}'X'X\mathbf{b})|X'X|^{1/2} \, db
\]

\[
\times A_{n-r}h(s^2)s^{n-r-1} \, ds \tag{12a}
\]

\[
\times \frac{1}{A_{n-r}} \, da, \tag{13a}
\]

where the non-negative function \( h(\cdot) \) over the non-negative real values is obtained through the normalization of the conditional distribution (11a) given \( \mathbf{d}, s \).

In both cases the marginal probability distribution for the observable \( \mathbf{d} \) is a uniform distribution on the unit sphere in \( \mathbb{C}^{\perp}(X) \), not depending on \( \lambda \). Thus there is no discriminatory information available concerning the values of the shape parameter \( \lambda \) alone. With several runs of the regression model we can test the goodness of fit for the spherical error (against non-spherical error) using the tests of uniform distributions on the surface of unit hyperspheres. See Prentice [7] for such tests.

For the parameter \( \sigma \) given \( \lambda \) we use the relation

\[
\frac{s(y)}{\sigma} = s(z),
\]

where the random variable \( s = s(z) \) has the distribution (12) in the case of
Student’s error and (12a) in the case of spherical error. Note that if we write the above relation as

\[ \frac{s^2(y)}{(n-r)\sigma^2} = \frac{s^2}{n-r} \]

then \( s^2/(n-r) \) has an \( F(n-r, \lambda) \) distribution for the case of Student’s error, and it becomes a Chi-square with \( (n-r) \) degrees of freedom as \( \lambda \to \infty \), which corresponds to the case of normal error. So the tests or confidence intervals for \( \sigma^2 \) are based on \( F \)-distribution. It agrees with Zellner’s results for \( \sigma^2 \) in [8].

For the tests and confidence regions of \( \beta \) we use the relation

\[ \frac{b(y) - \beta}{\sigma} = b \]

and the distribution (13) in the case of Student’s error and (13a) in the case of spherical error. Note that in the case of Student’s error, the distribution (13) is a multivariate Student \( (\lambda + n - r; 0, (\lambda + s^2)^{-1} (X'X)^{-1}) \), which becomes a multivariate normal as \( \lambda \to \infty \) (the case of a normal error).

Now let us consider the inference of the parameters in a second order: \( \lambda, \beta, \sigma \). Here we use the second presentation of (10) and the variables of the error will be considered in the order \( d, t, s \). In the case of Student’s error (1), we have

\[ f_\lambda(z) \, dz = \frac{A_\lambda}{A_{\lambda+n}} \left[ \frac{|X'X|^{1/2} n^{-1} \lambda^{-n/2}}{1 + \lambda^{-1} (1 + t'X'Xt) s^2} \right]^{(\lambda+n)/2} \, ds \, dt \, da \]

\[ = \frac{A_\lambda A_n}{A_{\lambda+n}} \left[ \frac{(1 + t'X'Xt)^{n/2} n^{-1} \lambda^{-n/2}}{1 + \lambda^{-1} (1 + t'X'Xt) s^2} \right]^{(\lambda+n)/2} \, ds \]  \hspace{1cm} (14)

\[ \times \frac{A_{n-r}}{A_n} \left[ \frac{|X'X|^{1/2}}{1 + t'X'Xt} \right]^{(n-r+r)/2} \, dt \]  \hspace{1cm} (15)

\[ \times \frac{1}{A_{n-r}} \, da, \]  \hspace{1cm} (16)
and in the case of spherical error (3) we have

\[
g(z'z) \, dz = A_n \left(1 + t'X'Xt\right)^{n/2} g(z'z) s^{n-1} \lambda^{-n/2} \, ds
\]

\[
\times \frac{A_{n-r}}{A_n} \frac{|X'X|^{1/2}}{(1 + t'X'Xt)^{(n-r+r)/2}} \, dt
\]

\[
\times \frac{1}{A_{n-r}} \, da.
\]

The only observable of \( z \) is \( d(z) = d(y) \) which has the uniform distribution (16), (16a) on the unit sphere in \( \mathcal{L}^\perp(X) \). The analysis is the same as the first ordering.

For the tests and confidence regions of \( \beta \) given \( \lambda \) we use the relation

\[
\frac{b(y) - \beta}{s(y)} = t,
\]

where the variable \( t \) has distribution (12) which is exactly the same as (12a). Note that this distribution is a multivariate Student \((n-r; 0, (X'X)^{-1})\), not depending on the shape parameter \( \lambda \). It coincides with the result from normal theory and it agrees with Zellner's analysis for \( \beta \) in [8].

For the tests and confidence intervals of \( \sigma \) given \( \lambda, \beta \) we use the relation

\[
\frac{s(y)}{\sigma} = s(z),
\]

where \( s = s(z) \) has distribution (14) in the case of Student's error and (14a) in the case of spherical error. Note that the distribution (14) for \( s \) is equivalent to an \( F(n, \lambda) \) distribution for \( s^2(1 + t'X'Xt)/n \).

This second order of examining may be the more natural way. It corresponds in the ordinary analysis of variance to the sequential pooling of error variance. For any amenable ordering of the parameters, the shape \( \lambda \) is first considered, using the likelihood function; and in the error distribution here we have the peculiarity that there is no observable information concerning \( \lambda \) alone.

Other than the two orders of parameters, one may examine \( \sigma \) using the first and examine \( \beta \) using the second. This does not control the overall
confidence level as with the sequential procedures of either of the two orders (see [5] for general discussion), but it does allow the separate examination of the location and the scale parameters. Computer simulations to date give some preference for this alternative procedure.

3. Multivariate regression with linear scaling

The methods of presentation and analysis given for the univariate regression model can now be mechanized routinely to give the methods and distributions theory for multivariate regression. For this we distinguish two cases corresponding to different degree of identifiability of the error distribution involved. In this section we examine the case where the error distribution is identifiable only up to a positive linear transformation. Then in section 4 we consider the case in which the error distribution is further identified up to a positive triangular transformation; this corresponds to a special case in which there is a known ordering to the response variables and the distributional information is relative to this ordering.

Now consider the model

\[ Y = \Theta X + \Gamma Z \]

\[ f_\lambda(Z) \, dZ \]  

(17)

where \( Y \) is a \( p \times n \) matrix of observable responses, \( \Theta \) a \( p \times r \) matrix of regression parameters, \( X \) a \( r \times n \) design matrix of regressors with full rank \( r < n \), \( \Gamma \) a \( p \times p \) matrix of scaling parameters with \( |\Gamma| > 0 \), \( Z \) a \( p \times n \) matrix of unobservable error variables with the density \( f_\lambda(Z) \). We assume that the density is a canonical matrix student:

\[ f_\lambda(Z) = \frac{A^{(p)}_\lambda}{A^{(p)}_{\lambda+n}} |I_p + Z^T Z|^{-(\lambda+n)/2} \]  

(18)

where we denote \( A^{(p)}_k = A_k A_{k-1} \cdots A_{k-p+1} \) for the decreasing product of surface areas of the unit spheres in \( \mathbb{R}^k, \mathbb{R}^{k-1}, \ldots, \mathbb{R}^{k-p+1} \) (please see (1)). Or more generally, we assume that the density has the spherical form:

\[ f_\lambda(Z) = g(ZZ') \]  

(19)

where \( g(\cdot) \) is any non-negative function over the \( p \times p \) positive definite matrices such that \( f_\lambda(Z) \) is a density. This is the general form for a
distribution invariant under rotations through the sample: $Z$ and $ZO$ have identical distribution for any $n\times n$ orthogonal matrix $O$.

Following Section 1, we want to present the model (17) in a more transparent coordinates system in which the observable portion and the unobservable portion of the random error $Z$ are readily recognized. For this purpose, it is convenient to view a matrix $Z$ in $\mathbb{R}^{mn}$ as a sequence $z_1, z_2, \ldots, z_p$ of $p$ vectors in $\mathbb{R}^n$. Now for any $Z$, project an arbitrary but fixed sequence of $p$ linearly independent vectors (say the first $p$ of the $n$ unit vectors in $\mathbb{R}^n$) onto the $(r+p)$-dimensional subspace $\mathcal{L}^+(x_1, \ldots, x_r; z_1, \ldots, z_p)$, or simply $\mathcal{L}^+(X; Z)$, spanned by the row vectors of $X$ and $Z$ together with that order as the positive orientation. The $p$ projections are then orthonormalized in that sequence to obtain $p$ orthonormal vectors $d_1, d_2, \ldots, d_p$, which are orthogonal to $x_1, \ldots, x_r$. This procedure gives a basis $x_1, \ldots, x_r, d_1, \ldots, d_p$ for the subspace $\mathcal{L}^+(X; Z)$ except for a set of measure zero for which the projections are linearly dependent. Note that

$$D(Z) = \begin{bmatrix} d_1 \\ \vdots \\ d_p \end{bmatrix}$$

depends on $\mathcal{L}^+(X; Z)$ but not otherwise on $Z$. Thus $D(Z) = D(Y)$ if and only if $\mathcal{L}^+(X; Z) = \mathcal{L}^+(X; Y)$.

We denote the regression coefficients of $z_j$ on $x_1, \ldots, x_r$ and $d_1, \ldots, d_p$ by $b_j = (b_{j1}, \ldots, b_{jr})$ and $c_j = (c_{j1}, \ldots, c_{jp})$, for $j = 1, \ldots, p$. Writing

$$B(Z) = \begin{bmatrix} b_1 \\ \vdots \\ b_p \end{bmatrix}, \quad C(Z) = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix},$$

we have the relation between the old and new coordinates:

$$Z = B(Z)X + C(Z)D(Z) = BX + CD. \quad (20)$$

As $Z$ varies in $\mathbb{R}^{mn}$, $D(\cdot)$ traces out smoothly the set of all $p$-dimensional subspaces of $\mathbb{R}^{n-r}$, which is a copy of the Grassman manifold $\mathcal{G}_{p, n-r}$. Let $dD$ denote the volume element of $\mathcal{G}_{p, n-r}$ orthogonal to the subspaces
Let \( \mathbb{Q}^{+}(X; Z) \), the relation among the volume element is

\[
dZ = |XX'|^{p/2}|C|^{n-r-p}dBdCdD.
\]

(21)

In terms of the new coordinates, the model (17) becomes

\[
D(Y) = D(Z)
\]

\[
B(Y) = \mathbb{B} + \Gamma B(Z)
\]

\[
C(Y) = \Gamma C(Z)
\]

\[
f_{\lambda}(BX + CD)|XX'|^{p/2}|C|^{n-r-p}dBdCdD,
\]

(22)

where we have abbreviated \( D(Z), B(Z), C(z) \) by \( D, B, C \) in the probability element. The variable \( D(Z) \) is directly observable, and, given data \( \mathbb{Q}^0 \), the appropriate probabilistic description for the unobservable \( B(Z), C(Z) \) is the conditional distribution of \( B, C \) given \( D = D(Y^0) \). This leads to the quantities

\[
\begin{cases}
\Gamma^{-1}(\mathbb{B} - B) = B & \text{or} & C^{-1}(Y)(B(Y) - \mathbb{B}) = H \\
\Gamma^{-1}C(Y) = C
\end{cases}
\]

(23)

to be examined conditionally given \( D = D(Y^0) \). Note that in the second presentation of (23), we have made a further change of variable: \( H = C^{-1}B \).

We first examine the parameters in the order: \( \lambda, \Gamma, \mathbb{B} \). The corresponding order for the quantities is then: \( D, C, B \). According to this order, the error distribution (18) is factored

\[
f_{\lambda}(Z)dZ = \frac{A^{(p)}_{\lambda}}{A^{(p)}_{\lambda+n}} \frac{|XX'|^{p/2}|C|^{n-r-p}}{|I + CC' + BXX'B'|^{(\lambda+n)/2}} dBdCdD
\]

\[
= \frac{A^{(p)}_{\lambda+n-r}}{A^{(p)}_{\lambda+n}} \frac{|I + CC'|^{(\lambda+(n-r))/2}|XX'|^{p/2}}{|I + CC' + BXX'B'|^{(\lambda+(n-r)+r)/2}} dB
\]

(24)

\[
\times \frac{A^{(p)}_{\lambda,p} A^{(p)}_{n-r:p}}{A^{(p)}_{\lambda+n-r:p}} \frac{|C|^{n-r-p}}{|I + CC'|^{(\lambda+(n-r))/2}} dC
\]

(25)

\[
\times \frac{1}{A^{(p)}_{n-r:p}} dD,
\]

(26)
where we have written

\[ A_{m:k}^{(p)} = A_m^{(p)} / A_k^{(p)}. \]

For the spherical error (19), we have

\[ g(ZZ') \, dZ = h^{-1}(CC')g(CC' + BXX'B')|XX'|^{p/2} \, dB \tag{24a} \]
\[ \times A_{n-r:p}^{(p)}h(CC')|C|^{n-r-p} \, dC \tag{25a} \]
\[ \times \frac{1}{\lambda A_n^{(p)}} \, dD, \tag{26a} \]

where

\[ h(CC') = \int g(CC' + BXX'B')|X'X|^{p/2} \, dB \]

is the norming constant of (24a)

In both cases, the marginal distribution for \( D \) is the uniform distribution with respect to the volume measure of the Grassman manifold, not depending on the shape parameter \( \lambda \). The likelihood function of \( \lambda \), like the case in section 2, is not informative.

For the parameter \( \Gamma \) given \( \lambda \) we use the relation

\[ \Gamma^{-1}C(Y) = C(Z) \]

where the random variable \( C = C(Z) \) has the distribution (25) in the case of Student's error (18), and (25a) in the case of spherical error (19). The distribution (25) is a positive root \( F \) distribution \( F_p^{1/2}(n-r, \lambda) \) for the \( p \times p \) square matrices \( C \) with positive determinant, in the sense that the positive definite matrix \( F = CC' \) has a matrix-\( F \) distribution:

\[ \frac{A_{n-r:p}^{(p)}A_{\lambda}^{(p)}}{A_{n-r+\lambda}^{(p)}} \cdot \frac{|F|^{(n-r)/2}}{|1 + F|^{(n-r+\lambda)/2}} \cdot \frac{dF}{2^p |F|^{(p+1)/2}} \tag{27} \]

For the population covariance matrix \( \Sigma = \Gamma \Gamma' \), we note that \( C'(Y)\Sigma^{-1}C(Y) = C'C \) and that \( C'C \) has the same distribution as \( CC' \), which is just the matrix-\( F \) in (27).

For the parameter \( \beta \) given \( \lambda, \Gamma \), we use

\[ \Gamma^{-1}(B(Y) - \beta) = B(Z), \]
where the random variable \( B = B(Z) \) has the distribution (24) in the case of Student’s error, and (24a) in the case of spherical error. Note that the distribution of (24) is a matrix-\( t \) distribution, \( T_p \times (n-r+\lambda; O, I + CC', (XX')^{-1}) \).

Now we examine the parameters in a more natural order for the regression purpose: \( \Gamma, \mathcal{B}, \lambda \). The quantities in the second presentation of (23) are thus examined in the order: \( D, H, C \). For the Student’s error (18), we have

\[
f_\lambda(Z) dZ = \frac{A^{(p)}_\lambda}{A^{(p)}_{\lambda+n}} \frac{|XX'|^{p/2}|C|^{n-p}}{|I + C(I + HXX'H')C'|^{(\lambda+n)/2}} dC dH dD
\]  
(28)

\[
= \frac{A^{(p)}_\lambda A^{(p)}_{n:p}}{A^{(p)}_{\lambda+n:p}} \frac{|I + HXX'H'|^{n/2}|C|^{n-p}}{|I + C(I + HXX'H')C'|^{(\lambda+n)/2}} dC
\]  
(29)

\[
\times \frac{A^{(p)}_{n-r}}{A^{(p)}_n} \frac{|XX'|^{p/2}}{|I + HXX'H'|^{(n-r+r)/2}} dH
\]  
(30)

\[
\times \frac{1}{A^{(p)}_{n-r:p}} dD,
\]  
(31)

where (28) is obtained using a change of variable \( (B, C) \rightarrow (H, C) \) with \( B = CH \),

\[
Z = C(HX + D), dZ = |XX'|^{p/2}|C|^{n-p} dC dH dD.
\]  
(32)

And for the spherical error (19) we have

\[
g(ZZ') dZ
= A^{(p)}_n |I + HXX'H'|^{n/2} g_\lambda(C(I + HXX'H')C')|C|^{n-p} dC
\]  
(29a)

\[
\times \frac{A^{(p)}_{n-r}}{A^{(p)}_n} \frac{|XX'|^{p/2}}{|I + HXX'H'|^{(n-r+r)/2}} dH
\]  
(30a)

\[
\times \frac{1}{A^{(p)}_{n-r:p}} dD
\]  
(31a)

In both error cases, the marginal distribution for the observable \( D(Z) = \)}
$D(Y)$ is, as in the first order considered above, a uniform distribution on the Grassman manifold $\mathfrak{g}_{p,n-r}$, not depending on the shape parameter $\lambda$. Hence the likelihood function of $\lambda$ is a constant function, giving no discriminatory information about $\lambda$.

For the parameter $\mathfrak{g}$ given $\lambda$, we use

$$C^{-1}(Y)(B(Y) - \mathfrak{g}) = H(Z),$$

where $H = (H(Z)$ has a matrix-$t$ distribution, $T_{p \times n}(n-r; O, I, (XX')^{-1})$, given by (30) and (30a). Note that this distribution does not depend on $\lambda$ and that the inference about $\mathfrak{g}$ is identical with the case of normal error.

For the parameter of $\Gamma$ given $\lambda$, $\mathfrak{g}$, we use

$$\Gamma^{-1}C(Y) = C(Z),$$

where $C = C(Z)$ has the distribution (29) and (29a) respectively for the Student’s and spherical error. Note that (29) is a positive root $F$ distribution, $F_{p, \lambda}^{1/2}(n, \lambda)$, as defined earlier.

In the second order of examining the parameters, the conditional tests and confidence regions for $\Gamma$ correspond to the pooling of error variance. One may prefer the ‘mixed’ procedure of using the first order for the inference of $\Gamma$ while using the second for the inference of $\mathfrak{g}$; but then the overall confidence level cannot be controlled. With this approach we have a matrix-$t$ distribution for the inference of $\mathfrak{g}$ as in the normal case while we have a matrix-$F$ distribution for the inference of $\Sigma^{-1} = (\Gamma \Gamma')^{-1}$ instead of a Wishart distribution as in the normal case.

4. Multivariate regression with triangular scaling

In this section we consider the multivariate regression model for the rather special case in which the response variables have a known sequential pattern and the error term can be identified up to a positive triangular transformation. The structural model has the form

$$Y = \mathfrak{g}X + \mathfrak{f}Z,$$

$$f_\lambda(Z) \, dZ$$

(32)

where $Y$, $\mathfrak{g}$, $X$ and $Z$ are matrices as defined in section 3, and $\mathfrak{f}$ is a $p \times p$ positive lower triangular matrix of scaling parameters. The error distribution $f_\lambda(Z)$ is assumed to be (18) or (19) in section 3.
For the space \( \mathbb{R}^{pn} \) of all \( p \times n \) matrices \( Z \), we change coordinates in the following way. Let \( B(Z) \) be the \( p \times r \) matrix of regression coefficients of the \( p \) row vectors of \( Z \) on \( X \),

\[
B(Z) = ZX'(XX')^{-1}.
\]  

(33)

The matrix \( Z - B(Z)X \) then consists of the \( p \) residual vectors. We ortho-normalize these \( p \) residual vectors and keep that orientation, getting

\[
Z - B(Z)X = T(Z)D(Z),
\]  

(34)

where \( T(Z) \) is a \( p \times p \) positive lower triangular matrix and \( D(Z) \) is a \( p \times n \) semi-orthogonal matrix with row vectors orthogonal to the row vectors of \( X \)

\[
D(Z)D'(Z) = I_p, \quad XD'(Z) = 0
\]  

(35)

so that

\[
Z = B(Z)X + T(Z)D(Z).
\]  

(36)

As \( Z \) varies in \( \mathbb{R}^{pn} \), \( D(\cdot) \) traces out smoothly the set of all those \( p \)-frames (i.e. ordered sequences of orthonormal vectors) in \( \mathbb{R}^{n} \) that are orthogonal to \( \mathcal{E}(X) \). This set is a copy of the Stiefel submanifold in \( \mathbb{R}^{p(n-r)} \) with dimension \( p(n-r) - p(p+1)/2 \). Let \( dD \) denotes its volume measure. The volume elements are then related

\[
dZ = |XX'|^{p/2}|T|^{n-r}|T|^{-1}_\Delta dBdTdD,
\]  

(37)

where \( |T|_\Delta = t_{11}t_{22}^{2} \cdots t_{pp}^{p} \) denotes the product of the diagonal elements of \( T \) in increasing power. Note that we have abbreviated \( B(Z) \), \( T(Z) \) and \( D(Z) \) by \( B \), \( T \), \( D \) respectively. We do this whenever it is convenient.

The model (32) can now be presented as

\[
D(Y) = D(Z)
\]

\[
B(Y) = \mathcal{B} + \mathcal{T}B(Z)
\]

\[
T(Y) = \mathcal{T}T(Z)
\]  

(38)

\[
\int_{\mathcal{A}} (BX + TD)|XX'|^{p/2}|T|^{n-r}|T|^{-1}_\Delta dBdTdD
\]

It is clear in this presentation that \( D(Z) \) is the only observable portion and
that the quantities

\[
\begin{cases}
    B(Y) = \mathcal{B} + \mathcal{J}B \\
    T(Y) = \mathcal{J}T
\end{cases}
\quad \text{or} \quad
\begin{cases}
    T^{-1}(Y)(B(Y) - \mathcal{B})T^{-1}B = H \\
    \mathcal{J}^{-1}T(Y)
\end{cases}
\tag{39}
\]

would be examined together with the conditional distribution of \( B, T \) given \( D = D(Y) \).

We first examine the parameters in the order: \( \lambda, \mathcal{J}, \mathcal{B} \). This leads to the sequence \( D, T, B \) for the error variable. The error distribution (18) becomes

\[
f_\lambda(Z) \, dZ = \frac{A^{(p)}_{\lambda + n - r}}{A^{(p)}_{\lambda + n}} \frac{|I + TT'|^{(\lambda + n - r)/2} |XX'|^{p/2}}{|I + TT' + BXX'B'|^{(\lambda + n - r + \lambda)/2}} \, dB \tag{40}
\]

\[
\times \frac{A^{(p)}_{n - r}}{A^{(p)}_{n - r + \lambda}} \frac{|T|^{n-r}|T|^{-1}}{|I + TT'|(n-r+\lambda)/2} \, dT \tag{41}
\]

\[
\times \frac{1}{A^{(p)}_{n - r}} \, dD; \tag{42}
\]

and the distribution (19) becomes

\[
g(ZZ') \, dZ = h^{-1}(TT')g(TT' + BXX'B') |XX'|^{p/2} \, dB \tag{40a}
\]

\[
\times A^{(p)}_{n - r} h(TT') |T|^{n-r} |T|^{-1} \, dT \tag{41a}
\]

\[
\times \frac{1}{A^{(p)}_{n - r}} \, dD. \tag{42a}
\]

The marginal distribution of the observable \( D \) is uniform with respect to the volume measure, in a Stiefel manifold providing no discriminatory information of the shape parameter \( \lambda \).

For the parameter \( \mathcal{J} \) given \( \lambda \), we use

\[
\mathcal{J}^{-1}T(Y) = T,
\]

where \( T = T(Z) \) is subject to the distribution (41) or (41a). The distribution (41) is a triangular root-\( F \) distribution \( \Delta F_{p}^{1}(n-r, \lambda) \) in the sense that the positive definite matrix \( F = TT' \) has the matrix-\( F \) distribution (27). For the covariance matrix \( \Sigma = \mathcal{J} \mathcal{J}' \), we note that \( T'(Y)\Sigma^{-1}T(Y) = T'T = V \) and
that $V$ has what we call here ‘disguised matrix-$F$’ distribution:

$$
\frac{A^{(p)}_{n-r, \lambda} A^{(p)}_{\lambda}}{A^{(p)}_{n-r + \lambda}} \frac{|V|^{(n-r)/2}}{|I + V|^{(n-r + \lambda)/2}} \frac{|T|_\nu}{|T|_\Delta} \frac{dV}{2^p |V|^{(p+1)/2}},
$$

(43)

where $|T|_\nu = t_{11}^p t_{22}^{-1} \cdots t_{pp}$ is the product of the diagonal elements of $T$ in decreasing power. Note that the only difference between (27) and (43) is the ‘disguising factor’ $|T|_\nu / |T|_\Delta$. So the inference of $\Sigma^{-1}$ is based on a disguised matrix-$F$ distribution.

For the parameter $\mathfrak{B}$ given $\lambda$, $\mathfrak{T}$, we use

$$
\mathfrak{T}^{-1}(B(Y) - \mathfrak{B}) = B(Z),
$$

where $B = B(Z)$ is subject to the distribution (40) or (40a). The distribution (40) is a matrix-$t$ distribution, $T_{p \times p}(n-r+\lambda; O, I + TT', (XX')^{-1})$.

At this point we introduce some notation developed in [1], for uses in the sequel. For a $p \times p$ matrix $V$ let $V^{(k)}$ and $V_{(k)}$ be the upper-left and low-right $k \times k$ matrix and define respectively the decreasing determinant and increasing determinant of $V$ as

$$
|V|_\nu = \prod_{k=1}^{p} |V^{(k)}|, |V|_\Delta = \prod_{k=1}^{p} |V_{(k)}|.
$$

(44)

These definitions generalized the decreasing and increasing determinants of triangular matrices. If $V$ is a positive definite matrix such that $V = LL'$, $V = T'T$, where both $L$ and $T$ are positive lower triangular, then

$$
|V|_\nu = |L|_\nu^2, \quad |V|_\Delta = |T|_\Delta^2
$$

$$
|V|^{(p+1)/2} = |L|_\Delta |L|_\nu = |T|_\Delta |T|_\nu,
$$

(45)

so that

$$
|L|_\nu / |L|_\Delta = |V|_\nu / |V|^{- (p+1)/2}
$$

$$
|T|_\nu / |T|_\Delta = |V|_\nu / |V|^{(p+1)/2}.
$$

(46)

For instance, the disguised matrix-$F$ distribution (43) can be expressed completely in terms of $V$:

$$
\frac{A^{(p)}_{n-r, \lambda} A^{(p)}_{\lambda}}{A^{(p)}_{n-r + \lambda}} \frac{|V|^{(n-r)/2}}{|I + V|^{(n-r + \lambda)/2}} \frac{dV}{2^p |V|_\nu}
$$

(47)
Consider now the parameters in the order: \( \lambda, \beta, \gamma \). The sequence of error variables is then: \( D, H, T \), using the second presentation of (39). The error distribution (18) becomes

\[
\begin{align*}
  f_{\lambda}(Z) \, dZ &= \frac{A^{(p)}_{\lambda}}{A^{(p)}_{\lambda + n}} \cdot \frac{|XX'|^{p/2}|T|^n|T|_\Delta^{-(\lambda + n)/2}}{|I + T(I + HXX'H')T'|^{(\lambda + n)/2}} \, dT \, dH \, dD \\
  &= \frac{A^{(p)}_{\lambda}}{A^{(p)}_{\lambda + n}} \frac{A^{(p)} |I + HXX'H'|^{(n - p - 1)/2}|1 + HXX'H'|_\gamma |T|^n|T|_\Delta^{-(\lambda + n)/2}}{|I + T(I + HXX'H')T'|^{(\lambda + n)/2}} \, dT \\
  &\times \frac{A^{(p)}_{n - r}}{A^{(p)}_n} \frac{|XX'|^{p/2}}{|I + HXX'H'|^{n/2}} \frac{|I + HXX'H'|^{(p + 1)/2}}{|I + HXX'H'|_\gamma} \, dH \\
  &\times \frac{1}{A^{(p)}_{n - r}} \, dD,
\end{align*}
\]

(48)

(49)

(50)

where the conditional distribution (48) for \( T \) is obtained by transforming the standard triangular root-\( F \) distribution (41). The distribution (49) is a ‘disguised matrix-\( t \)’ with ‘disguising’ factor

\[
q = \frac{|I + HXX'H'|^{(p + 1)/2}}{|I + HXX'H'|_\gamma}.
\]

(51)

A disguised matrix-\( t \) distribution is related to a disguised Wishart distribution in the same way that a matrix-\( t \) distribution is related to a Wishart distribution; see [9], [4] and [1].

Similarly, the error distribution (19) becomes

\[
\begin{align*}
  g(ZZ') \, dZ &= A^{(p)}_n |I + HXX'H'|^{n/2} g(T(I + HXX'H')T') |T|^n |T|_\Delta^{-1} q^{-1} \, dT \\
  &\times \frac{A^{(p)}_{n - r}}{A^{(p)}_n} \frac{|XX'|^{p/2}}{|I + HXX'H'|^{n/2}} \cdot q \, dH \\
  &\times \frac{1}{A^{(p)}_{n - r}} \, dD
\end{align*}
\]

(48a)

(49a)

(50a)
The marginal distribution for the observable $D$ is uniform with respect to the volume measure in the Stiefel manifold as in the first order of parameters.

For the parameter $\mathbb{B}$ given $\lambda$, we have the disguised matrix-$t$ distribution (49), which is the same as (49a), together with the relation

$$T^{-1}(Y)(B(Y) - \mathbb{B}) = H.$$ 

This coincides with the corresponding result for a normal error.

For the parameter $\mathbb{T}$ given $\lambda$, $\mathbb{B}$, we have the distribution (48) or (48a), together with the relation

$$\mathbb{T}^{-1}T(Y) = T.$$ 

As for the model in section 3, one may prefer the mixed procedure of using the first order for inference of $\mathbb{T}$ and the second order for the inference of $\mathbb{B}$. With this approach we will have a disguised matrix-$t$ distribution for the inference of $\mathbb{B}$ as in the normal case, while we will have a disguised matrix-$F$ for $\Sigma^{-1} = (\mathbb{T}\mathbb{T}^\prime)^{-1}$ instead of a disguised Wishart distribution as in the normal case.

References