Ann. Inst. Statist. Math.
38 (1986), Part A, 285-296

SOME TEST STATISTICS FOR THE STRUCTURAL COEFFICIENTS OF THE MULTIVARIATE LINEAR FUNCTIONAL RELATIONSHIP MODEL

SERGE B. PROVOST

(Received June 11, 1984; revised Oct. 5, 1985)

Summary

For the testing problem concerning the coefficients of the multivariate linear functional relationship model, the distribution of a statistic previously proposed by A. P. Basu depends on the unknown covariance matrix V of errors, so limiting its applicability. This article proposes new test statistics with sampling distributions which are independent of the unknown parameters for the cases where V is either unknown or known only up to a proportionality factor. The exact distributions of the test statistics are also discussed.

1. Introduction

The problem of inference concerning the coefficients of a single linear relation among several unobserved "true" variables, when the observed vectors are contaminated with errors or fluctuations has a long history. The early writers on this functional model, notably Adcock [2], Kummel [10], Pearson [13] and van Uven [19] were mainly concerned with the derivation of least squares estimators. Modern statistical methods were used for the first time by Wald [22]. Particular aspect of this problem has been studied by Creasy [5] assuming the ratio of error variances to be known a priori; by Geary [6] using product-cumulants and by Theil [17] resorting to nonparametric methods.

In experimental work, it is usually possible to replicate the observations. Data coming from replicated experiments can be analyzed without much difficulty, because we can easily obtain from them an estimator of the covariance matrix of the experimental errors which can be assumed to have a known distribution. The case in which replicated observations are available was considered by Tukey [18], who showed how estimators of the linear relation could be easily derived

Key words and phrases: Linear functional relationships, tests of hypotheses, exact distributions.

from a variance component analysis.

Villegas showed in [20] that an estimator previously proposed by Acton [1] was the maximum likelihood estimator and derived in [21] a test based on the F-distribution for testing the hypothesis that an unknown linear relation between the expected values of the components of the observation vectors is a given linear relation.

Following Villegas' approach, Basu [4] proposed a test statistic for the parameters of k linear relations. This model is now referred to as the multivariate linear functional relationship (MLFR) model.

Unfortunately the applicability of that statistic is limited since, as is pointed out in the next section, its distribution depends on $\boldsymbol{V}$, the covariance matrix of errors associated with the vectors of observations.

New test statistics for the MLFR model are proposed in Sections 3 and 4 for the cases where $\boldsymbol{V}$ is either unknown or known only up to a proportionality factor. The exact distributions of these test statistics are also discussed.

Recently, the MLFR model and other errors-in-variables models have been analyzed in Keller and Wansbeck [9] where their connections to several multivariate statistical techniques were pointed out.

The MLFR model can be defined as follows. Let

$$M=\{\underset{p\times 1}{\boldsymbol{g}} \mid \underset{k\times 1}{\boldsymbol{a}}+\underset{k\times p}{\boldsymbol{B}'}\boldsymbol{g}=\boldsymbol{0}\}\ . \tag{1.1}$$

This set contains all the vectors $\boldsymbol{g}$ which are solutions of the non-homogeneous system of k linear equations in p variables specified by $\boldsymbol{a}$ and $\boldsymbol{B}$. We assume that $k<p$ and that $\boldsymbol{B}$ is of full rank k.

Let $\boldsymbol{g}_1,\cdots,\boldsymbol{g}_r$ be r points belonging to the $(p-k)$-dimensional set M and let

$$\boldsymbol{x}_{ij}=\boldsymbol{g}_i+\boldsymbol{e}_{ij}\ ,\qquad j=1,\cdots,n_i\ , \tag{1.2}$$

be n_i replicated measurements that are available for each of the $\boldsymbol{g}_i$'s $i=1,\cdots,r$, where

$$\boldsymbol{e}_{ij}\overset{\text{ind}}{\sim} N_p(\boldsymbol{0},\boldsymbol{V})\ ,\qquad \boldsymbol{V}>0\ , \tag{1.3}$$

that is, the $\boldsymbol{e}_{ij}$'s are independently and identically distributed random vectors, each having a p-variate normal distribution with mean vector zero and positive definite covariance matrix $\boldsymbol{V}$. The MLFR model is specified by (1.1), (1.2) and (1.3) and is a special case of a more general model for whose parameters the maximum likelihood estimators were derived in Healy [7].

2. An account of Basu's main result

In this section, we explain why a test statistic proposed in Basu

[4] for the structural coefficients of the MLFR model cannot be used for testing the null hypothesis:

$$H_0: \boldsymbol{a}=\boldsymbol{a}_0,\ \boldsymbol{B}=\boldsymbol{B}_0 \quad \text{where } \boldsymbol{V} \text{ is unknown}. \tag{2.1}$$

Let

$$\boldsymbol{W}=\boldsymbol{B}_0'\boldsymbol{S}\boldsymbol{B}_0 \tag{2.2}$$

where

$$\boldsymbol{S}=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_i)'/(n-r)\ ,$$

$$n=n_1+\cdots+n_r$$

and

$$\bar{\boldsymbol{x}}_i=\sum_{j=1}^{n_i}\boldsymbol{x}_{ij}/n_i\ , \tag{2.3}$$

and let

$$d_i^2=(\boldsymbol{a}_0+\boldsymbol{B}_0'\bar{\boldsymbol{x}}_i)'(\boldsymbol{B}_0'\boldsymbol{B}_0)^{-1}(\boldsymbol{a}_0+\boldsymbol{B}_0'\bar{\boldsymbol{x}}_i)\ , \qquad i=1,\cdots,r\ , \tag{2.4}$$

where d_i represents the distance between $\bar{\boldsymbol{x}}_i$ and the set

$$M_0=\{\boldsymbol{g}:\ \boldsymbol{a}_0+\boldsymbol{B}_0'\boldsymbol{g}=\boldsymbol{0}\}$$

with respect to the Euclidean metric

$$\{(\bar{\boldsymbol{x}}_i-\boldsymbol{g})'(\bar{\boldsymbol{x}}_i-\boldsymbol{g})\}^{1/2}\ .$$

This last result is proved in Basu [4] where the following statistic is proposed to test H_0 defined in (2.1):

$$R=(n-r)d^2/|\boldsymbol{W}|^{1/k} \tag{2.5}$$

where $|\boldsymbol{W}|$ denotes the determinant of $\boldsymbol{W}$ defined in (2.2) and

$$d^2=\sum_{i=1}^{r}n_i d_i^2$$

where d_i^2 is defined in (2.4).

The distribution of $|\boldsymbol{W}|$ is that of the product of k independent chi-square variables times a constant (see Anderson [3], p. 171):

$$|\boldsymbol{W}|\sim|\boldsymbol{B}_0'\boldsymbol{V}\boldsymbol{B}_0|\prod_{i=1}^{k}\chi^2_{n-r-i+1}$$

where $\chi^2_{n-r-i+1}$ denotes a chi-square variable with $n-r-i+1$ degrees of freedom.

At this point, we will give a systematic derivation of the distribution of d since some of the results are needed in Section 4.

It is seen from (1.2) and (2.3) that

$$\bar{\boldsymbol{x}}_i \overset{\text{ind}}{\sim} N_k(\boldsymbol{g}_i, \boldsymbol{V}/n_i)\ , \qquad i=1,\cdots,r\ .$$

Hence, under H_0,

$$\boldsymbol{a}_0+\boldsymbol{B}_0'\bar{\boldsymbol{x}}_i \overset{\text{ind}}{\sim} N_k(\boldsymbol{0}, \boldsymbol{B}_0'\boldsymbol{V}\boldsymbol{B}_0/n_i)\ , \qquad i=1,\cdots,r\ , \tag{2.6}$$

and since $\boldsymbol{B}_0'\boldsymbol{V}\boldsymbol{B}_0$ is symmetric and positive definite, there exists a non-singular matrix $\boldsymbol{K}=(\boldsymbol{B}_0'\boldsymbol{V}\boldsymbol{B}_0)^{-1/2}$ such that

$$\boldsymbol{K}'\boldsymbol{B}_0'\boldsymbol{V}\boldsymbol{B}_0\boldsymbol{K}=\boldsymbol{I}\ . \tag{2.7}$$

Now, letting

$$\boldsymbol{Z}_i=n_i\boldsymbol{K}'(\boldsymbol{a}_0+\boldsymbol{B}'\bar{\boldsymbol{x}}_i)\ , \qquad i=1,\cdots,r\ ,$$

it is seen from (2.6) that

$$\boldsymbol{Z}_i \overset{\text{ind}}{\sim} N_k(\boldsymbol{0}, \boldsymbol{I})\ , \qquad i=1,\cdots,r\ ,$$

and from (2.4) that

$$n_id_i^2=\boldsymbol{Z}_i'\boldsymbol{K}^{-1}(\boldsymbol{B}_0'\boldsymbol{B}_0)^{-1}\boldsymbol{K}'^{-1}\boldsymbol{Z}_i\ , \qquad i=1,\cdots,r\ .$$

Since $(\boldsymbol{B}_0'\boldsymbol{B}_0)^{-1}$ is symmetric and positive definite, so is $\boldsymbol{K}^{-1}(\boldsymbol{B}_0'\boldsymbol{B}_0)^{-1}\boldsymbol{K}'^{-1}$ and hence there exists an orthogonal matrix $\boldsymbol{U}$ such that

$$\boldsymbol{U}'\boldsymbol{K}^{-1}(\boldsymbol{B}_0'\boldsymbol{B}_0)^{-1}\boldsymbol{K}'^{-1}\boldsymbol{U}=\boldsymbol{M}=\operatorname{diag}(m_1,\cdots,m_k)\ , \tag{2.8}$$

where $m_i>0$, $i=1,\cdots,k$. The m_i's are therefore the eigenvalues of

$$(\boldsymbol{B}_0'\boldsymbol{V}\boldsymbol{B}_0)^{1/2}(\boldsymbol{B}_0'\boldsymbol{B}_0)^{-1}\{(\boldsymbol{B}_0'\boldsymbol{V}\boldsymbol{B}_0)^{1/2}\}'\ . \tag{2.9}$$

Now let

$$(f_{i1},\cdots,f_{ik})'=\boldsymbol{f}_i=\boldsymbol{U}'\boldsymbol{z}_i\ , \qquad i=1,\cdots,r\ ,$$

then

$$\boldsymbol{f}_i \overset{\text{ind}}{\sim} N_k(\boldsymbol{0}, \boldsymbol{I})$$

and

$$n_id_i^2=\boldsymbol{f}_i'\boldsymbol{M}\boldsymbol{f}_i=\sum_{j=1}^{k} m_jf_{ij}$$

where

$$f_{ij} \overset{\text{ind}}{\sim} N(0, 1)\ , \qquad i=1,\cdots,r;\ j=1,\cdots,k\ .$$

Hence,

(2.10) $$n_i d_i^2 \overset{\text{ind}}{\sim} \sum_{j=1}^{k} m_j \chi_j^2(1) , \qquad i=1,\cdots,r ,$$

and

(2.11) $$d^2=\sum_{i=1}^{r} n_i d_i^2 \sim \sum_{j=1}^{k} m_j \chi_j^2(r) ,$$

where $\chi_j^2(r)$ denotes a chi-square variate with r degrees of freedom, the m_j's are defined in (2.9) and all the chi-squares in (2.10) and (2.11) are independently distributed.

So clearly, when $\boldsymbol{V}$ is unknown, we are unable to test H_0 with R defined in (2.5) since its distribution is not free of $\boldsymbol{V}$ and therefore the tail probabilities cannot be computed. In the next section, we propose an alternate statistic to test H_0.

3. Alternate test statistics for the case where $\boldsymbol{V}$ is unknown

In this section, we derive two new statistics to test the hypothesis H_0 given in (2.1).

Let

$$\boldsymbol{t}_i=n_i^{1/2}\boldsymbol{B}_1(\boldsymbol{a}_0+\boldsymbol{B}_0'\bar{\boldsymbol{x}}_i) ,$$

where

$$\boldsymbol{B}_1=(\boldsymbol{B}_0'\boldsymbol{B}_0)^{-1/2} \qquad \text{and} \qquad \boldsymbol{B}_1'=\boldsymbol{B}_1 .$$

Then, in view of (2.6),

$$\boldsymbol{t}_i \overset{\text{ind}}{\sim} N_k(\boldsymbol{0}, \boldsymbol{V}^*) , \qquad i=1,\cdots,r ,$$

where under the null hypothesis

$$\boldsymbol{V}^*=\boldsymbol{B}_1\boldsymbol{B}_0'\boldsymbol{V}\boldsymbol{B}_0\boldsymbol{B}_1 .$$

Let

$$\bar{\boldsymbol{t}}=\sum_{i=1}^{r} \boldsymbol{t}_i/r \qquad \text{and} \qquad \boldsymbol{S}^*=\sum_{i=1}^{r} (\boldsymbol{t}_i-\bar{\boldsymbol{t}})(\boldsymbol{t}_i-\bar{\boldsymbol{t}})' ,$$

then for $r>k$, we propose the test statistic

$$Q=r(\bar{\boldsymbol{t}}'(\boldsymbol{S}^*)^{-1}\bar{\boldsymbol{t}}) ,$$

where

$$Q \sim \{k/(r-k)\}F_{k,r-k} ,$$

that is, the distribution of Q under H_0 is that of a constant times an

F-variable with k and $(r-k)$ degrees of freedom, (see Anderson [3], p. 107).

Now, since

$$\bar{\boldsymbol{t}}=\boldsymbol{B}_1\sum_{i=1}^{r}\{n_i^{1/2}(\boldsymbol{a}_0+\boldsymbol{B}_0'\bar{\boldsymbol{x}}_i)\}/r=\boldsymbol{B}_1\boldsymbol{u}^*\left\{\sum_{i=1}^{r}n_i^{1/2}/r\right\}$$

where

$$\boldsymbol{u}^*=\boldsymbol{a}_0+\left\{\boldsymbol{B}_0'\sum_{i=1}^{r}(n_i^{1/2}\bar{\boldsymbol{x}}_i)\Big/\sum_{i=1}^{r}n_i^{1/2}\right\},$$

it is seen from (2.4) that

$$\bar{\boldsymbol{t}}'\bar{\boldsymbol{t}}\left(\sum_{i=1}^{r}n_i^{1/2}/r\right)^{-2}=\boldsymbol{u}^{*\prime}(\boldsymbol{B}_0'\boldsymbol{B}_0)^{-1}\boldsymbol{u}^*$$

represents the square of the Euclidean distance between M_0 and a weighted average of the $\bar{\boldsymbol{x}}_i$'s.

We may also consider the following statistic to test H_0:

$$Q_c=n(\bar{\boldsymbol{c}}'\boldsymbol{S}_c^{*\,-1}\bar{\boldsymbol{c}})\sim(k/(n-k))F_{k,n-k},$$

where

$$n=\sum_{i=1}^{r}n_i;\quad \bar{\boldsymbol{c}}=\sum_{i=1}^{r}\sum_{j=1}^{n_i}\boldsymbol{c}_{ij}/n;\quad \boldsymbol{S}_c^*=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(\boldsymbol{c}_{ij}-\bar{\boldsymbol{c}})(\boldsymbol{c}_{ij}-\bar{\boldsymbol{c}})',$$

and

$$\boldsymbol{c}_{ij}=\boldsymbol{B}_1(\boldsymbol{a}_0+\boldsymbol{B}_0'\boldsymbol{x}_{ij})\sim N_k(0,\boldsymbol{V}^*).$$

Moreover, since

$$\bar{\boldsymbol{c}}=\sum_{i=1}^{r}\sum_{j=1}^{n_i}\boldsymbol{B}_1(\boldsymbol{a}_0+\boldsymbol{B}_0'\boldsymbol{x}_{ij})/n=\boldsymbol{B}_1(\boldsymbol{a}_0+\boldsymbol{B}_0'\bar{\boldsymbol{x}})$$

where

$$\bar{\boldsymbol{x}}=\sum_{i=1}^{r}\sum_{j=1}^{n_i}\boldsymbol{x}_{ij}/n,\qquad \bar{\boldsymbol{c}}'\bar{\boldsymbol{c}}=(\boldsymbol{a}_0+\boldsymbol{B}_0'\bar{\boldsymbol{x}})'(\boldsymbol{B}_0'\boldsymbol{B}_0)^{-1}(\boldsymbol{a}_0+\boldsymbol{B}_0'\bar{\boldsymbol{x}})$$

represents the square of the Euclidean distance between M_0 and the arithmetic mean of the $\boldsymbol{x}_{ij}$'s according to (2.4).

4. A test statistic for the case where V is known only up to a proportionality factor

In this section, we consider the null hypothesis:

$$H_0^p:\ \boldsymbol{a}=\boldsymbol{a}_0,\qquad \boldsymbol{B}=\boldsymbol{B}_0 \text{ and } \boldsymbol{V} \text{ is proportional to } \boldsymbol{V}_0,$$

that is, $\boldsymbol{V}=v^2\boldsymbol{V}_0$, where $\boldsymbol{a}_0$, $\boldsymbol{B}_0$ and $\boldsymbol{V}_0$ are known.

Letting

$$v\boldsymbol{K}=\boldsymbol{K}_1=(\boldsymbol{B}_0'\boldsymbol{V}_0\boldsymbol{B}_0)^{-1/2},$$

we can rewrite (2.7) as

$$\boldsymbol{K}_1'\boldsymbol{B}_0'\boldsymbol{V}_0\boldsymbol{B}_0\boldsymbol{K}_1=\boldsymbol{I}$$

and (2.8) as

$$v\boldsymbol{U}'\boldsymbol{K}_1^{-1}(\boldsymbol{B}_0'\boldsymbol{B}_0)^{-1}\boldsymbol{K}_1'^{-1}\boldsymbol{U}v=\boldsymbol{M}=v^2\boldsymbol{M}^*=v^2\,\mathrm{diag}\,(m_1^*,\cdots,m_k^*)\,,$$

where

$$v^2m_i^*=m_i\,,\qquad i=1,\cdots,k\,.$$

The m_i^*'s are therefore the eigenvalues of

$$(\boldsymbol{B}_0'\boldsymbol{V}_0\boldsymbol{B}_0)^{1/2}(\boldsymbol{B}_0'\boldsymbol{B}_0)^{-1}\{(\boldsymbol{B}_0'\boldsymbol{V}_0\boldsymbol{B}_0)^{1/2}\}'\,,$$

and (2.11) becomes

$$d^2=\sum_{i=1}^{r}n_id_i^2\sim v^2\sum_{j=1}^{k}m_j^*\chi_j^2(r)\,.$$

We propose the following statistic to test H_0^p:

$$R^p=d^2/\hat{v}^2\,,$$

where

$$\hat{v}^2=\sum_{i=1}^{r}\sum_{j=1}^{n_i}\{(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}})'\boldsymbol{V}_0^{-1}(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}})\}/\{p(n-r)\}\,. \tag{4.1}$$

Let us prove that $\hat{v}^2$ is an unbiased estimator of v^2. Let

$$\boldsymbol{P}=\boldsymbol{V}^{-1/2}=\boldsymbol{V}_0^{-1/2}/v\,;\qquad \boldsymbol{y}_{ij}=\boldsymbol{P}\boldsymbol{x}_{ij}\,;$$

$$\bar{\boldsymbol{y}}_i=\sum_{j=1}^{n_i}\boldsymbol{y}_{ij}/n_i\,;\qquad \boldsymbol{y}_{ij}=(y_{ij1},\cdots,y_{ijp})'\,;$$

and

$$\bar{\boldsymbol{y}}_i=(y_{i\cdot1},\cdots,y_{i\cdot p})'\qquad \text{where } y_{i\cdot l}=\sum_{j=1}^{n_i}y_{ijl}/n_i\,,$$

then $\mathrm{cov}\,(\boldsymbol{y}_{ij})=\boldsymbol{I}$ and therefore the y_{ijl}'s, $l=1,\cdots,p$, are all independent with variance equal to unity. It follows that

$$\sum_{j=1}^{n_i}(y_{ijl}-y_{i\cdot l})^2\sim\chi^2_{n_i-1}\,.$$

Thus (4.1) becomes

$$\hat{v}^2 = v^2 \sum_{i=1}^{r} \sum_{j=1}^{n_i} \{(\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}}_i)'(\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}}_i)\} / \{p(n-r)\}$$
$$= v^2 \sum_{i=1}^{r} \sum_{l=1}^{p} \sum_{j=1}^{n_i} (y_{ijl} - y_{i \cdot l})^2 / \{p(n-r)\} ,$$

and

$$\hat{v}^2 \sim v^2 \sum_{l=1}^{p} \sum_{i=1}^{r} \chi^2_{n_i - 1} / \{p(n-r)\} ,$$

that is,

$$\hat{v}^2 \sim v^2 \chi^2_{p(n-r)} / \{p(n-r)\}$$

since all the chi-squares are independent. Hence

$$E(\hat{v}^2) = v^2 .$$

We will now show that d^2 and $\hat{v}^2$ are independently distributed. Let

$$\boldsymbol{w}_{ij} = \boldsymbol{V}_0^{-1/2} \boldsymbol{x}_{ij} , \quad i=1, \cdots, r ; \; j=1, \cdots, n_i ; \qquad \boldsymbol{w}_{ij} = (w_{ij1}, \cdots, w_{ijp})' ;$$
$$\bar{\boldsymbol{w}}_i = (w_{i \cdot 1}, \cdots, w_{i \cdot p})' \qquad \text{where } w_{i \cdot l} = (w_{i1l} + \cdots + w_{in_i l}) / n_i , \; l=1, \cdots, p ,$$

and

$$s_{il}^2 = \sum_{j=1}^{n_i} (w_{ijl} - w_{i \cdot l})^2 .$$

Then

$$d^2 = \sum_{i=1}^{r} n_i (\boldsymbol{a}_0 + \boldsymbol{B}_0' \bar{\boldsymbol{x}}_i)' (\boldsymbol{B}_0' \boldsymbol{B}_0)^{-1} (\boldsymbol{a}_0 + \boldsymbol{B}_0' \bar{\boldsymbol{x}}_i)$$
$$= \sum_{i=1}^{r} n_i (\boldsymbol{a}_0 + \boldsymbol{B}_0' \boldsymbol{V}_0^{1/2} \bar{\boldsymbol{w}}_i)' (\boldsymbol{B}_0' \boldsymbol{B}_0)^{-1} (\boldsymbol{a}_0 + \boldsymbol{B}_0' \boldsymbol{V}_0^{1/2} \bar{\boldsymbol{w}}_i)$$

is a function of the $w_{i \cdot l}$'s which are all independent, and

$$\hat{v}^2 = \sum_{i=1}^{r} \sum_{j=1}^{n_i} \{(\boldsymbol{w}_{ij} - \bar{\boldsymbol{w}}_i)'(\boldsymbol{w}_{ij} - \bar{\boldsymbol{w}}_i)\} / \{p(n-r)\} = \sum_{i=1}^{r} \sum_{l=1}^{p} s_{il}^2 / \{p(n-r)\}$$

is a function of the s_{il}^2's which are all independent. Since each s_{il}^2 is independent of $w_{i \cdot l}$, we can conclude that $\hat{v}^2$ is distributed independently of d^2. So

$$R^p = d^2 / \hat{v}^2 \sim d^* / z$$

where

$$d^* \sim p(n-r) \sum_{j=1}^{k} \{m_j^* \chi_j^2(r)\} , \qquad z \sim \chi^2_{p(n-r)}$$

and all the chi-squares are independent.

The exact distribution of R^p can be obtained by considering the respective distributions of d^* and z and then by making the appropriate transformation.

Three representations of the density of a linear combination of chi-squares can be found in Mathai and Pillai [11]. One is given in terms of finite sums, another in terms of a confluent hypergeometric function of several variables and a third one in terms of zonal polynomials. The convolution formula yields another representation which is discussed in Robbins [15]. The density of d^* may also be derived from its moment generating function:

$$M_{d^*}(t)=\prod_{j=1}^{k}(1-\mu_j t)^{-r/2}\ , \qquad \text{where } \mu_j=2p(n-r)m_j^*$$

on expanding the latter as an infinite power series in $(1-t)^{-1}$. Writing

$$(1-\mu_j t)=\mu_j(1-t)(1-c_j(1-t)^{-1})\ , \qquad \text{where } c_j=(\mu_j-1)/\mu_j\ ,$$

we have, if $|c_j|<1$, that is, if $p(n-r)m_j^*>1/4$, $j=1,\cdots,k$,

$$(1-\mu_j t)^{-r/2}=\mu_j^{-r/2}\sum_{\nu_j=0}^{\infty}\{(r/2)_{\nu_j}/\nu_j!\}c_j^{\nu_j}(1-t)^{-(\nu_j+r/2)}\ ,$$

where $(\alpha)_{\nu_j}=(\alpha+\nu_j-1)(\alpha+\nu_j-2)\cdots(\alpha)$ and $(\alpha)_0=1$, so

$$M_{d^*}(t)=\left\{\prod_{j=1}^{k}\mu_j^{-r/2}\right\}\sum_{\nu_1=0}^{\infty}\cdots\sum_{\nu_k=0}^{\infty}(r/2)_{\nu_1}\cdots(r/2)_{\nu_k}c_1^{\nu_1}\cdots c_k^{\nu_k}(1-t)^{-\nu-kr/2}/(\nu_1!\cdots\nu_k!)$$

where

$$\nu=\sum_{j=1}^{k}\nu_j$$

which on inversion gives, for $d^*>0$,

$$f(d^*)=\sum_{\nu=0}^{\infty}k_\nu e^{-d^*}(d^*)^{\nu+(rk/2)-1}/\Gamma(\nu+rk/2) \tag{4.2}$$

where

$$k_\nu=\sum_{\nu_1+\cdots+\nu_k=\nu}\left(\prod_{j=1}^{k}\mu_j^{-r/2}\right)(r/2)_{\nu_1}\cdots(r/2)_{\nu_k}(c_1^{\nu_1}\cdots c_k^{\nu_k})/(\nu_1!\cdots\nu_k!)\ .$$

This representation has also been derived by Provost [14] using the technique of the inverse Mellin transform. If the condition $p(n-r)m_j^*>1/4$ is not satisfied for $j=1,\cdots,k$, we use the following technique. We multiply both d^* and z by B where B is a positive number such that $Bp(n-r)m_j^*>1/4$ for all j. This allows us to use (4.2) to express the density of $B\ d^*$ in a computable form. Then Bz would be distributed as a gamma variate with parameters $p(n-r)/2$ and $2\cdot B$, and the exact density of R^p would be obtained with a transformation of variables.

The representation of the density of d^* derived in this section can be written in the following form:

$$h(d^*)=\sum_i \theta_{3i} d^{*\theta_{1i}} e^{-d^*\theta_{2i}}$$

where $\theta_{ji}>0,\ j=1, 2, 3.$

Let

$$q=d^*/z\ , \tag{4.3}$$

where q represents the test statistic R^p. The density of d^* is given in (4.2) and the density of z is

$$u(z)=\{z^{(p(n-r)/2)-1}e^{-z/2}\}/\{2^{p(n-r)/2}\Gamma(p(n-r)/2\}\ .$$

Since d^* and z are independently distributed the joint density of d^* and z is $t(d^*, z)=h(d^*)u(z)$.

Now let

$$y=d^*\ . \tag{4.4}$$

The equations (4.3) and (4.4) define a one-to-one transformation which maps the set $\{(d^*, z);\ 0\leqq d^*<\infty,\ 0\leqq z<\infty\}$ onto the set $\{(q, y);\ 0\leqq q<\infty,\ 0\leqq y<\infty\}$. Clearly $d^*=y$ and $z=y/q$, and the Jacobian of the transformation is y/q^2. Thus the joint density of y and q is

$$\nu(y, q)=(y/q^2)t(y, y/q)=(y/q^2)h(y)u(y/q)$$

and the probability density function of q is

$$w(q)=\int_0^\infty \nu(y, q)dy\ .$$

The parts containing y and q coming from $\nu(y, q)$ are of the following form:

$$(y/q^2)y^{\theta_{1i}}e^{-y\theta_{2i}}(y/q)^{p(n-r)/2-1}e^{-y(2q)^{-1}}$$

that is

$$y^{\theta_{1i}+\varepsilon}e^{-y(\theta_{2i}+(2q)^{-1})}q^{-(\varepsilon+1)}$$

where $\varepsilon=p(n-r)/2,\ \theta_{1i}>0$ and $\theta_{2i}>0$.

Hence the integration over y yields

$$(\theta_{2i}+(2q)^{-1})^{-(\theta_{1i}+\varepsilon+1)}\Gamma(\theta_{1i}+\varepsilon+1)q^{-\varepsilon-1}$$

$$=\Gamma(\theta_{1i}+\varepsilon+1)2^{\theta_{1i}+\varepsilon+1}q^{\theta_{1i}}(2q\theta_{2i}+1)^{-(\theta_{1i}+\varepsilon+1)}$$
$$=K_{1i}(2q\theta_{2i})^{(\theta_{1i}+1)-1}(2q\theta_{2i}+1)^{-(\theta_{1i}+1+\varepsilon)},$$

where K_{1i} is a constant.

Therefore the distribution of R^p is expressible as a linear combination of the distributions of beta type-2 random variables. Moreover, in computational work, we can use the tables of the beta distribution:

$$I_{r,s}(x)=\{\Gamma(r+s)/(\Gamma(r)\Gamma(s))\}\int_0^x u^{r-1}(1-u)^{s-1}du$$

for $0<x<1$, together with the identity

$$B_{r,s}(x)=I_{r,s}(x/(1+x))$$

where $B_{r,s}(x)$ denotes the distribution function of a beta type-2 variable with parameters r and s whose probability density function is

$$\beta(x)=\{\Gamma(r+s)/(\Gamma(r)\Gamma(s))\}x^{r-1}(1+x)^{-(r+s)}.$$

An appoximation of the distribution of R^p can be obtained by approximating the distribution of d^*. Various approximations are available for the distribution of a linear combination of chi-square random variables, see for instance Oman and Zacks [12], Jensen and Solomon [8] for a Gaussian approximation, or Solomon and Stephens [16] where the distribution is to be fitted by $A\omega^D$, where ω has the χ_s^2 distribution and the constants A, s and D are found by matching moments.

Acknowledgements

The author would like to express sincere thanks to his supervisor Professor A. M. Mathai and to the referee whose valuable suggestions improved the presentation of the material in this article. This research was supported by the Natural Sciences and Engineering Research Council of Canada.

THE UNIVERSITY OF WESTERN ONTARIO, LONDON, CANADA

REFERENCES

[1] Acton, Forman S. (1959). *Analysis of Straight-Line Data*, Wiley, New York.
[2] Adcock, R. J. (1978). A problem in least squares, *Analyst.*, **5**, 53–54.
[3] Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*, Wiley, New York.
[4] Basu, A. P. (1969). On some tests for several linear relations, *J. Roy. Statist. Soc.*, B, **31**, 65–71,
[5] Creasy, M. A. (1958). Confidence limits for the gradient in the linear functional relationship, *J. Roy. Statist. Soc.*, B, **18**, 65–69.

[6] Geary, R. C. (1942). Inherent relations between random variables, *Proc. R. Irish Acad.*, Ser. A, **47**, 63-76.

[7] Healy, John D. (1980). Maximum likelihood estimation of a multivariate linear functional relationship, *J. Multivar. Anal.*, **10**, 243-251.

[8] Jensen, D. R. and Solomon, H. (1972). A Gaussian approximation to the distribution of a definite quadratic form, *J. Amer. Statist. Ass.*, **67**, 898-902.

[9] Keller, W. J. and Wansbeck, Tom (1983). Multivariate methods for quantitative and qualitative data, *J. Econometrics*, **22**, 91-111.

[10] Kummel, Chas H. (1879). Reduction of observed equations which contain more than one observed quantity, *Analyst.*, **6**, 97-105.

[11] Mathai, A. M. and Pillai, K. C. S. (1982). Further results on the trace of a noncentral Wishart matrix, *Commun. Statist. Theor. Meth.*, **11**(10), 1077-1086.

[12] Oman, D. S. and Zacks, S. (1981). A mixture approximation to the distribution of a weighted sum of chi-squared variables, *J. Statist. Comput. Simul.*, **13**, 215-224.

[13] Pearson, Karl (1901). On lines and planes of closest fit to systems of points in space, *Philos. Mag.*, **2**, 559-572.

[14] Provost, S. B. (1984). *Distribution Problems Connected with the Multivariate Linear Functional Relationship Model*, Ph.D. dissertation, McGill University, Montreal.

[15] Robbins, Herbert (1948). The distribution of a definite quadratic form, *Ann. Math. Statist.*, **19**, 266-270.

[16] Solomon, H. and Stephens, M. (1977). Distribution of a sum of weighted chi-squared variables, *J. Amer. Statist. Ass.*, **72**, Theor. Meth., 881-885.

[17] Theil, H. (1950). A rank invariant method of linear and polynomial regression analysis, *Nederl. Akad. Wetensch. Proc.*, Ser. A, **53**, 386-392, 521-525 and 1397-1412.

[18] Tukey, John W. (1951). Components in regression, *Biometrics*, **7**, 33-69.

[19] van Uven, M. J. (1930). Adjustment of N points (in n-dimensional space) to the best linear $(n-1)$-dimensional space, *K. Akad. Wetens. Amsterdam Proc. Sec. Sciences*, **33**, 143-157; 307-326.

[20] Villegas, C. (1961). Maximum likelihood estimation of a linear functional relationship, *Ann. Math. Statist.*, **32**, 1048-1062.

[21] Villegas, C. (1964). Confidence region for a linear relation, *Ann. Math. Statist.*, **35**, 780-788.

[22] Wald, Abraham (1940). The fitting of straight lines if both variables are subject to error, *Ann. Math. Statist.*, **11**, 284-300.