

MEDEDELINGEN LANDBOUWHOGESCHOOL
WAGENINGEN • NEDERLAND • 70-13 (1970)

ON A TEST FOR THE DIFFERENCE BETWEEN TWO CORRELATION COEFFICIENTS

L. C. A. CORSTEN

*Department of Mathematics, Agricultural University,
Wageningen, The Netherlands*

(Received 4-I-1970)

SUMMARY

Let $x' = (x_1, x_2, x_3)$ be a 3-dimensional normal vector with unknown means and unknown covariance matrix C . A likelihood ratio test for the null hypothesis: $\rho_{13} = \rho_{23}$ on the basis of n independent observations of x is developed. The calculation of the test statistic which asymptotically has a chi-square distribution with one degree of freedom requires the maximum likelihood estimation of C under H_0 . This estimate is found by using a representation of C derived from factor analysis i.e. $S(bb' + I)S$ where b' is a 3-tuple $(\beta_1, \beta_2, \beta_3)$ and S is a diagonal matrix with elements $\sigma_1, \sigma_2, \sigma_3$. The main difference from factor analytical methods is that in this case 5 instead of 6 parameters have to be estimated since $\beta_1 = \beta_2$, and that the parameters can be as well real as purely imaginary. As three mutually exclusive possibilities have to be distinguished, three different numerical procedures, all having some resemblance with the estimation procedure in factor analysis, have to be applied separately each time a test is required.

1. THE PROBLEM

Let $x' = (x_1, x_2, x_3)$ be a 3-dimensional random vector with unknown expectation vector μ and unknown covariance matrix C . We are concerned with the test on the basis of n independent observations of x of the null hypothesis that two of the three correlation coefficients, ρ_{13} and ρ_{23} , say, are equal against the alternative that they are not.

This problem arises in the situation where x_1 and x_2 are both possible linear

predictors for x_3 but, for example, for reasons of economy, one wishes to measure only one of these in order to make future predictions. It may happen that x_1 can be measured quite simply while x_2 is very difficult to determine. In choosing one of these two predicting variables it may be helpful to test whether one variable is more effective than the other in predicting x_3 or not. Since

$$\text{var}(x_3 | x_i = x_i) = (1 - \rho_{i3}^2) \text{var } x_3 \quad i = (1,2)$$

this difference in prediction efficiency may equivalently be tested by investigating

$$H_0 : \rho_{13}^2 = \rho_{23}^2 \text{ against } H_1 : \rho_{13}^2 \neq \rho_{23}^2 \quad (1.1)$$

by means of n triplets of joint observations of the variables concerned.

In particular when there is not a significant difference in efficiency there are no statistical reasons for a preference (at least on the basis of the sample considered) and one will choose the most easily measurable variable as predictor. When, on the other hand, there is a significant difference one should be guided, not only by considerations of easiness of measurement, but also by statistical arguments.

This problem (and some generalizations) had the attention of several authors such as Hotelling [4], Williams [6, 7, 8], Healy [2, 3], while Kendall and Stuart [5] vol II, p 374, refer, without criticism, to previous treatments. In the first place Hotelling and subsequent writers implicitly change the problem into that of testing

$$H_0 : \rho_{13} = \rho_{23} \text{ against } H_1 : \rho_{13} \neq \rho_{23} \quad (1.2)$$

as has been observed by William Kruskal and Ingram Olkin (personal communication). This different problem (1.2) is of course still an interesting one and it will be considered in this paper as a preliminary but most important step.

Next the hypothesis $\rho_{13} = \rho_{23}$ equivalent to $\alpha_1 \text{var}^{\dagger}x_1 = \alpha_2 \text{var}^{\dagger}x_2$, where α_1 and α_2 are the linear regression coefficients of x_1 and x_2 , respectively, in the joint (population) regression of x_3 on x_1 and x_2 is replaced by these authors by

$$\alpha_1 s_1 = \alpha_2 s_2 \quad (1.3)$$

where s_1^2 and s_2^2 are the sample mean squares of x_1 and x_2 , which opens the way to what is called a treatment of the problem conditional on the values of x_1 and x_2 actually observed. The linear relation (1.3) among linear regression coefficients, and with the normal distribution also among the coefficients of the regression function expressing the expectations of x_3 in the corresponding values of x_1 and x_2 , is subsequently tested in the traditional manner, leading to a test statistic with Student's t -distribution under this null hypothesis.

This remarkable 'elimination' of 'nuisance parameters' made Kruskal and

Olkin suspicious, while Williams [6, 7] also reveals certain doubts. The following objections to the procedure can be raised. (a) The null hypothesis (1.3) is simply different from (1.2) with probability one, and is dependent on the 'condition'. (b) One might hope that the 'conditional' test statistic proposed which, given the values of x_1 and x_2 , obviously has a non-central t-distribution under H_0 of (1.2), would have an unconditional central distribution, at least asymptotically. However, one can prove that the non-centrality does not converge to zero in probability, its variance being independent of n . Hence the 'unconditional' chance of an error of the first kind can be considerably larger than the significance level used in the 'conditional' test. (c) If one restricts the problem to the conditional situation only, i.e. testing the relation (1.3) with respect to the n values of the regression function at the n values of x_1 and x_2 actually observed, one is not allowed to see this as an investigation of the general prediction problem introduced in the beginning. Under very specific conditions this restricted problem may have some practical meaning, that is only with respect to the approximations of the n values of the expectation of x_3 by the corresponding values either of x_1 or of x_2 , and not elsewhere.

Another proposal advocated by Olkin (to be published) consists of using the asymptotic normal distribution and the asymptotic variance of the usual estimate of $\rho_{13}-\rho_{23}$ by K. Pearson in terms of the three correlation coefficients involved and equating population and sample correlation coefficients in that variance. Since the expression does not take in account the restrictions imposed by the null hypothesis it is expected to be too large, so that the resulting test will be too conservative. Some numerical checks were in agreement with this conjecture.

As the description above indicates, a basic difficulty of the testing problem was the occurrence of nuisance parameters in the 6 parameter covariance matrix C , more particular in the case of the null hypothesis (1.2) when there are still 4 independent parameters in addition to $\rho_{13} = \rho_{23}$.

We wish to consider a test of (1.2) based on the likelihood ratio principle which requires the maximization of the logarithm of the likelihood function

$$f = -\frac{1}{2}n(3 \ln 2\pi - \ln |C^{-1}| + \text{tr } A C^{-1}) \quad (1.4)$$

where A is the sample covariance matrix. The maximum likelihood estimation of C under H_0 by C_m , say, is the basic problem considered in this paper. The maximum likelihood estimate of C in the general framework is of course equal to A and the corresponding value of f equals.

$$f_1 = -\frac{1}{2}n(3 \ln 2\pi + 3 + \ln |A|), \quad (1.5)$$

while the value of f under H_0 equals

$$f_0 = -\frac{1}{2}n(3 \ln 2\pi - \ln |C_m^{-1}| + \text{tr } A C_m^{-1}) \quad (1.6)$$

The resulting test statistic

$$z = -2f_0 + 2f_1 = -n(\ln|A| - \ln|C_m| + 3 - \text{tr } A C_m^{-1}) \quad (1.7)$$

will have asymptotically a one-dimensional chi-square distribution as n goes to infinity.

In passing it may be noted that an alternative heuristic procedure has been proposed by Aitkin, Nelson and Reinfurt [1] in the form of their test statistic T_7 or rather its modification T_8 . They investigated its empirical sampling distribution, disposing of the likelihood ratio test as not feasible. Further, their preference of T_9 , that is Hotelling's statistic, to likelihood ratio statistics because of a higher power is to be judged carefully against the objections given above. So we continue the study of the likelihood ratio method which reveals some interesting aspects.

Once $C = C_m$ has been found and subsequently (1.7), we need only a few words for the (asymptotically valid) test of (1.1) i.e. $\rho_{13}^2 = \rho_{23}^2$. The other possible restriction of (1.1) is then in addition to that of (1.2): $\rho_{13} = -\rho_{23}$. But this can immediately be reduced into a case similar to (1.2) by changing the sign of the values of either x_1 or x_2 , that is of the nondiagonal elements of the first or second row and column of A , respectively. Next, the restriction which yields the smaller value of (1.7) will correspond to the maximum likelihood estimate under (1.1), and the test statistic, again a one-dimensional chi-square variate under (1.1), equals this smaller value.

2. PRELIMINARIES

Before going into the maximum likelihood estimation of C i.e. maximization of (1.4) with respect to the elements of C under the restriction $\rho_{13} = \rho_{23}$, we will introduce new parameters instead of the correlation coefficients ρ_{12} , ρ_{13} , ρ_{23} and the variances $\text{var } x_1$, $\text{var } x_2$ and $\text{var } x_3$ occurring in C or even worse in C^{-1} , since differentiation with respect to these old parameters and subsequent introduction of the restriction is impracticable and leads to intractable equations.

We derive this representation from factor analysis in which a canonical representation of a pxp covariance matrix C is given by

$$S(UU' + I)S \quad (2.1)$$

where S is a diagonal matrix of order p with positive diagonal elements, I the pxp identity matrix and U a pxk matrix of standardized factor loadings ($k < p$) satisfying

$$U'U = \Lambda \quad (2.2)$$

where Λ is a diagonal $k \times k$ matrix with positive diagonal elements. The ortho-

gonal columns of U span the standardized factor space. Each diagonal element of C is the sum of the corresponding element of S^2 , the specific variance, and that of $SU(SU)'$, the communality. S may be considered as a scale transformation while the addition of the term I makes UU' non-singular.

In order that (2.1) does not contain more than 6 parameters in our case with $p = 3$, k must not be greater than 1. This, however, does not mean that any positive definite 3×3 matrix can be written in the form (2.1) as defined above with $k = 1$. This is only the case for positive definite matrices with this special structure, where only non-negative specific variances and communalities are allowed.

To remove this restriction we must first allow S^2 to have negative elements; therefore the elements of S can be real or purely imaginary. Next, diagonal elements of $U'U + I$ can be negative; hence elements of the only column of U can be real or purely imaginary, and the only element of (2.2) can be negative.

So C will be represented by

$$S(bb' + I)S \quad (2.3)$$

where S is 3×3 diagonal matrix with real or purely imaginary non-zero diagonal elements σ_i ($i = 1, 2, 3$), b a column vector of three real or purely imaginary elements β_i ($i = 1, 2, 3$) and I the 3×3 identity matrix. Further let

$$b'b = \lambda \quad (2.4)$$

where λ is a real number.

Now the restriction of the null hypothesis $\rho_{13} = \rho_{23}$ is equivalent to

$$\frac{\sigma_1 \sigma_3 \beta_1 \beta_3}{\sqrt{\{\sigma_1^2 \sigma_3^2 (\beta_1^2 + 1) (\beta_3^2 + 1)\}}} = \frac{\sigma_2 \sigma_3 \beta_2 \beta_3}{\sqrt{\{\sigma_2^2 \sigma_3^2 (\beta_2^2 + 1) (\beta_3^2 + 1)\}}}$$

i.e. $\beta_3 = 0$ or, on squaring both sides but noting that β_1 and β_2 must have the same sign, $\beta_1 = \beta_2$. The first case leads to the null hypothesis $\rho_{13} = \rho_{23} = 0$ for which the well-known conditional test for joint linear regression of x_3 on x_1 and x_2 is available and in this context is not of interest. Hence we add to (2.3) and (2.4)

$$\beta_1 = \beta_2 \quad (2.5)$$

as the formulation of H_0 (1.2).

Now we consider the conditions under which (2.3), (2.4) and (2.5) define a positive definite matrix C . It is necessary and sufficient that the non-diagonal elements are real (the diagonal elements already are) and that the determinants of three principal matrices of increasing order, each of which contains the previous ones are positive. This leads to:

$$\sigma_1^2 \sigma_2^2 \beta_1^4 > 0 \quad (2.6a)$$

$$\sigma_1^2 \sigma_3^2 \beta_1^2 \beta_3^2 > 0 \quad (2.6b)$$

$$\sigma_2^2 \sigma_3^2 \beta_1^2 \beta_3^2 > 0 \quad (2.6c)$$

$$\sigma_1^2 (\beta_1^2 + 1) > 0 \quad (2.6d)$$

$$\sigma_1^2 \sigma_2^2 (2\beta_1^2 + 1) > 0 \quad (2.6e)$$

$$\sigma_1^2 \sigma_2^2 \sigma_3^2 (2\beta_1^2 + \beta_3^2 + 1) > 0 \quad (2.6f)$$

Suppose $\sigma_1^2 < 0$. Then (2.6d) $\Rightarrow \beta_1^2 < -1$ and (2.6a) $\Rightarrow \sigma_2^2 < 0$. These three inequalities contradict (2.6e). Hence, necessarily $\sigma_1^2 > 0$ and (2.6a) $\Rightarrow \sigma_2^2 > 0$. On defining $\sigma_1^2 > 0$ and $\sigma_2^2 > 0$ we may replace the set of inequalities (2.6) equivalently by:

$$\sigma_3^2 \beta_1^2 \beta_3^2 > 0 \quad (2.7a)$$

$$\beta_1^2 + 1 > 0 \quad (2.7b)$$

$$2\beta_1^2 + 1 > 0 \quad (2.7c)$$

$$\sigma_3^2 (2\beta_1^2 + \beta_3^2 + 1) > 0 \quad (2.7d)$$

Suppose $\sigma_3^2 > 0$. Then from (2.7a) two cases can arise.

(i) $\sigma_3^2 > 0$; β_1 and β_3 are real without limitation by the remaining equations.

Obviously from (2.4):

$$\lambda = 2\beta_1^2 + \beta_3^2 \geq 0 \quad (2.8)$$

where the equality sign corresponds to the case where C is a diagonal matrix, which is of no great interest. Note that $\rho_{12} \geq 0$.

(ii) $\sigma_3^2 > 0$; β_1 and β_3 are both imaginary and $\rho_{12} < 0$. Then the only restriction imposed by the other equations turns out to be $-1 < 2\beta_1^2 + \beta_3^2 < 0$, i.e.

$$-1 < \lambda < 0 \quad (2.9)$$

If $\sigma_3^2 < 0$ then, from (2.7a), $\beta_1^2 < 0 \Rightarrow \beta_3^2 \geq 0$. But this is impossible since $\beta_3^2 \geq 0$ makes (2.7c) and (2.7d) contradictory. So there remains only one case.

(iii) $\sigma_3^2 < 0$; then $\beta_1^2 \geq 0$, $\beta_3^2 < 0$, hence $\rho_{12} \geq 0$. These inequalities together with (2.7b, c, d) are equivalent to:

$$\beta_1^2 \geq 0$$

and

$$2\beta_1^2 + \beta_3^2 < -1 \text{ i.e. } \lambda < -1 \quad (2.10)$$

Note that we have obtained three mutually exclusive possibilities for λ , all values for λ being possible except $\lambda = -1$, which would make C singular, and a partition of the parameter space of C into three regions, to be called region (i); (ii) and (iii), respectively, in the sequel.

In the following we shall use a generalization of the well-known theorem that any real symmetric $p \times p$ matrix C can be represented by $\sum_{i=1}^p \lambda_i y_i y_i'$ where the y_i ($i = 1, \dots, p$) are real orthonormal eigenvectors of C and λ_i the associated real eigenvalues, or in the case that C is positive definite as $\sum_{i=1}^p x_i x_i'$ where the x_i are orthogonal eigenvectors of C satisfying $x_i' x_i = \lambda_i$. This is seen from the fact that the y_i form a basis of R^p and that each y_i is mapped by $\sum_{i=1}^p \lambda_i y_i y_i'$ as $\lambda_i y_i$. Using this representation, C^{-1} is given immediately by $\sum_{i=1}^p \lambda_i^{-1} y_i y_i'$.

If C is Hermitian and non-singular, a similar representation is possible, since the eigenvalues are also real, while there is an orthogonal set of p eigenvectors (real or imaginary), the inner products for complex-valued vectors being defined similar to that for real vectors. In particular if the real symmetric matrix has negative eigenvalues the corresponding terms $x_i x_i'$ will contain purely imaginary vectors x_i , with real but negative squared length. Now explicit inversion of C given by (2.3), (2.4) and (2.5) will be possible, for each of the three regions separately.

For region (i) $S^{-1}CS^{-1} = bb' + I$ can have the same set of eigenvectors as bb' while the associated eigenvalues are each increased by one. The eigenvectors of bb' are firstly column b , with eigenvalue λ greater than zero, in agreement with (2.8), and then an arbitrary pair of orthogonal vectors, orthogonal to b , and each with eigenvalue zero. Hence $(S^{-1}CS^{-1})^{-1} = SC^{-1}S$ has eigenvalues equal to $(\lambda + 1)^{-1}$ and to unity respectively. Since I can have the same set of eigenvectors as $SC^{-1}S$ and $1 - (\lambda + 1)^{-1} = \lambda(\lambda + 1)^{-1}$ it follows that

$$SC^{-1}S = I - (\lambda + 1)^{-1}bb'$$

or

$$C^{-1} = S^{-1}\{I - (\lambda + 1)^{-1}bb'\}S^{-1} \quad (2.11)$$

all elements being real.

For region (ii) the same reasoning can be applied as b is a purely imaginary vector and $-1 < b'b = \lambda < 0$. We have here

$$C^{-1} = S^{-1}\{I - (\lambda + 1)^{-1}bb'\}S^{-1} \quad (2.12)$$

where all elements are real except those of b , while $0 < \lambda + 1 < 1$. An alternative form of (2.12) is

$$S^{-1}\{I + (\lambda + 1)^{-1} \begin{pmatrix} b_1 \\ b_1 \\ b_3 \end{pmatrix} (b_1, b_1, b_3)\}S^{-1} \quad (2.13)$$

where all elements are real, the only difference from (2.11) being the positive sign before $(\lambda + 1)^{-1}$, while $2b_1^2 + b_3^2$ must be smaller than one.

For region (iii) we define C, with s_3 and b_3 real, as (2.3) where S has diagonal elements $\sigma_1, \sigma_2, -is_3$, and $b' = (\beta_1, \beta_1, ib_3)$ while from (2.10) $b'b < -1$. In the same way as above we obtain

$$C^{-1} = S^{-1}\{I - (\lambda + 1)^{-1}bb'\}S^{-1} \quad (2.14)$$

where S^{-1} has diagonal elements $\sigma_1^{-1}, \sigma_2^{-1}, is_3^{-1}$, and $\lambda + 1 < 0$.

In expanded form (2.14) equals

$$\begin{pmatrix} \sigma_1^{-1} & 0 & 0 \\ 0 & \sigma_2^{-1} & 0 \\ 0 & 0 & s_3^{-1} \end{pmatrix} \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} - (\lambda + 1)^{-1} \begin{pmatrix} \beta_1 \\ \beta_1 \\ -b_3 \end{pmatrix} (\beta_1, \beta_1, -b_3) \right] \begin{pmatrix} \sigma_1^{-1} & 0 & 0 \\ 0 & \sigma_2^{-1} & 0 \\ 0 & 0 & s_3^{-1} \end{pmatrix} \quad (2.15)$$

with real elements only.

3. THE MAXIMUM LIKELIHOOD EQUATIONS

In order that the expression (1.4) will be maximized with respect to C, the stationary points will be determined by setting the partial derivatives with respect to the five real parameters equal to zero, in each of the three regions for C separately. The maximum of the three regional maxima corresponding to the smallest value of z in (1.7) will be the required maximum.

We first consider *region (i)* with (2.11) for C^{-1} ; the other regions will be treated in a similar manner.

With $C = (c_{ij})$, $C^{-1} = (c^{ij})$ and $A = (a_{ij})$ we observe

$$\frac{\delta \ln |C|^{-1}}{\delta \beta_i} = \sum_u \sum_v \frac{\delta \ln |C|^{-1}}{\delta c^{uv}} \frac{\delta c^{uv}}{\delta \beta_i} \quad (i=1,3),$$

hence

$$\frac{\delta \ln |C|^{-1}}{\delta \beta_1} = -2(\lambda + 1)^{-1} \sum_{i=1}^2 \sum_{u=1}^3 c_{iu} \sigma_i^{-1} \beta_u \sigma_u^{-1}$$

$$\frac{\delta \ln |C|^{-1}}{\delta \beta_3} = -2(\lambda + 1)^{-1} \sum_{u=1}^3 c_{3u} \sigma_3^{-1} \beta_u \sigma_u^{-1} \text{ both with } \beta_2 = \beta_1.$$

Similarly

$$\frac{\delta \operatorname{tr} \mathbf{AC}^{-1}}{\delta \beta_1} = -2(\lambda + 1)^{-1} \sum_{i=1}^2 \sum_{u=1}^3 a_{iu} \sigma_i^{-1} \beta_u \sigma_u^{-1}$$

$$\frac{\delta \operatorname{tr} \mathbf{AC}^{-1}}{\delta \beta_3} = -2(\lambda + 1)^{-1} \sum_{u=1}^3 a_{3u} \sigma_3^{-1} \beta_u \sigma_u^{-1}$$

and

$$\frac{\delta \ln |\mathbf{C}|^{-1}}{\delta \sigma_i^{-1}} = 2 \sum_{u=1}^3 c_{iu} \sigma_i c^{iu}$$

$$\frac{\delta \operatorname{tr} \mathbf{AC}^{-1}}{\delta \sigma_i^{-1}} = 2 \sum_{u=1}^3 a_{iu} \sigma_i c^{iu}$$

It follows that

$$\frac{\delta f}{\delta \mathbf{b}} = n(\lambda + 1)^{-1} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} (\mathbf{S}^{-1} \mathbf{A} \mathbf{S}^{-1} \mathbf{b} - \mathbf{S}^{-1} \mathbf{C} \mathbf{S}^{-1} \mathbf{b}) \quad (3.1)$$

$$\frac{\delta f}{\delta \operatorname{diag} \mathbf{S}^{-1}} = n(\operatorname{diag} \mathbf{C} \mathbf{C}^{-1} \mathbf{S} - \operatorname{diag} \mathbf{A} \mathbf{C}^{-1} \mathbf{S}) \quad (3.2)$$

where $\operatorname{diag} \mathbf{A}$ is defined as the vector $(a_1, a_2, a_3)'$ with $a_i = a_{ii}$, the i -th diagonal element of \mathbf{A} .

It may be noted that (3.1) and (3.2) set equal to zero are almost the same equations as those occurring in factor analysis, but instead of each element of the column $(\mathbf{S}^{-1} \mathbf{A} \mathbf{S}^{-1} - \mathbf{S}^{-1} \mathbf{C} \mathbf{S}^{-1}) \mathbf{b}$ being zero, only the third element and the sum of the first and the second element need to be so, and also $\beta_1 = \beta_2$. The cofactor $(\lambda + 1)^{-1}$ occurring in (3.1) is not of interest and could as well have been absorbed all the time into \mathbf{b} by using $(\lambda + 1)^{-1} \mathbf{b}$ instead.

On substitution of (2.3) into (3.1) and use of (2.4), (3.1) set equal to zero reduces to

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \{ \mathbf{S}^{-1} \mathbf{A} \mathbf{S}^{-1} \mathbf{b} - (\lambda + 1) \mathbf{b} \} = 0 \quad (3.3)$$

Defining

$$\mathbf{R} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3.4)$$

so that $\mathbf{R} \mathbf{b} = \mathbf{b}$ and $\mathbf{R}^2 = \mathbf{R}$, we replace (3.3), in repeating the first equation, by

$$R\{S^{-1}AS^{-1} + (\lambda + 1)I\}Rb = 0$$

or

$$RS^{-1}AS^{-1}Rb = (\lambda + 1)b \quad (3.5)$$

Hence b must be an eigenvector of the symmetric matrix

$$B = RS^{-1}AS^{-1}R \quad (3.6)$$

with squared length equal to the corresponding eigenvalue minus one, although S is as yet unknown. We observe that B is obtained by a scale transformation S^{-1} of the sample covariance matrix, followed by the replacement of the four North-West elements by their arithmetic mean and similarly of the first two elements of the third row and column, while the third diagonal element remains unaltered.

As yet it is not clear which eigenvector of B (with rank ≤ 2 and positive trace, since the contribution of the first two diagonal elements is the sample variance of $(x_1 + x_2)/2$ after a scale transformation) has to be chosen. We shall show that the *largest* eigenvalue and the associated eigenvector has to be taken in order that f , given S , will be maximized.

Observe from the considerations about $S^{-1}CS^{-1}$ above,

$$\ln|C| = \ln|S^{-1}CS^{-1}| + \ln|S^2| = \ln(\lambda + 1) + \ln|S^2|$$

and from (2.11), (3.4), (3.6), (3.5) and (2.4)

$$\begin{aligned} \text{tr}(AC^{-1}) &= \text{tr}(S^{-1}AS^{-1})(SC^{-1}S) = \\ &\text{tr}\{S^{-1}AS^{-1}(I - (\lambda + 1)^{-1}bb')\} = \\ &\text{tr}(S^{-1}AS^{-1}) - (\lambda + 1)^{-1} \text{tr}\{S^{-1}AS^{-1}Rb(Rb)'\} = \\ &\text{tr}(S^{-1}AS^{-1}) - (\lambda + 1)^{-1} \text{tr}(b'Bb) = \\ &\text{tr}(S^{-1}AS^{-1}) - \lambda \end{aligned}$$

Hence, the stationary value of (1.4) with respect to b , given S , is

$$-\frac{1}{2}n\{3 \ln 2\pi + \ln(\lambda + 1) + \ln|S^2| + \text{tr}(S^{-1}AS^{-1}) - \lambda\} \quad (3.7)$$

Since $\ln(\lambda + 1) - \lambda$ is decreasing in $\lambda > 0$ the maximum for f , given S , is reached by choosing the maximum value of λ in (3.5), as was to be proved.

By now we see that C will be estimated by (2.3) where b is the eigenvector of $B = RS^{-1}AS^{-1}R$ with largest eigenvalue $\lambda + 1$, while $b'b = \lambda$.

We are left with the estimation of the unknown S to be found in principle by setting (3.2) equal to zero, which gives:

$$\text{diag}(AC^{-1}) = \text{diag} I \quad (3.8)$$

By a reduction similar to that to (3.7) we note that

$$\text{diag}(AC^{-1}) = \text{diag}(S^{-1}AS^{-1}) - (\lambda + 1)^{-1} \text{diag}(S^{-1}AS^{-1}bb') \quad (3.9)$$

On substitution into (3.2) of b_0 , the optimal solution for b of (3.5), given S , we obtain

$$n\{\text{diag} S - \text{diag}(S^{-1}A) + (\lambda + 1)^{-1} \text{diag}(S^{-1}AS^{-1}b_0b_0'S)\}$$

as the partial derivative with respect to $\text{diag} S^{-1}$ of the maximum of (1.4) given S . We prefer, however, the partial derivatives of the maximum of f , given S ,

with respect to $\ln \sigma_i^{-1}$, while $\frac{df}{dz} = \frac{1}{z} \cdot \frac{df}{d \ln z}$,

$$n \text{diag}(I - S^{-1}AS^{-1} + (\lambda + 1)^{-1}S^{-1}AS^{-1}b_0b_0') \quad (3.10)$$

a function not of b but of S only.

The solution of the equations, obtained by setting (3.10) equal to zero, as well as the actual calculation of b_0 , given S , will be considered in the following section.

Since by (3.8) $\text{tr}(AC^{-1}) = 3$ we find that the required test statistic (1.7) will be equal to

$$\begin{aligned} z &= n(\ln|C_m| - \ln|A|) = n(\ln|S^{-1}C_mS^{-1}| - \ln|S^{-1}AS^{-1}|) \\ &= n(\ln(\lambda + 1) - \sum_{i=1}^3 \ln \lambda_i) \end{aligned} \quad (3.11)$$

where $(\lambda + 1)$ is the required largest eigenvalue of $B = RS^{-1}AS^{-1}R$ and λ_i are the eigenvalues of $S^{-1}AS^{-1}$.

Further observe that (3.10) = 0 implies

$$\text{tr}(I - S^{-1}AS^{-1} + (\lambda + 1)^{-1}S^{-1}AS^{-1}b_0b_0') = 0$$

or

$$3 - \sum_{i=1}^3 \lambda_i + \lambda = 0$$

Hence $\text{tr}(S^{-1}AS^{-1})$ will be at least 3 in order that $\lambda \geq 0$. This implies the σ_i^2 to be restricted by $\sum_{i=1}^3 (a_{ii}/\sigma_i^2) \geq 3$. Since the third element of $\text{diag}(S^{-1}AS^{-1}b_0b_0')$ equals that of $\text{diag}(RS^{-1}AS^{-1}b_0b_0') = (\lambda + 1) \text{diag}(b_0b_0')$, the third element of (3.9) is non-negative, from which an additional restriction $\sigma_3^{-2} a_{33} - 1 \geq 0$ i.e. $\sigma_3^2 \leq a_{33}$ follows.

It should be noted that the estimation procedure is scale invariant in the following sense. When the elements of x are subject to a change of scale i.e. are replaced by Kx where K is an arbitrary non-singular real diagonal matrix, and so C and A are replaced by KCK and KAK , then on replacing S by KS the equations (3.5) and (2.10) do not change; hence b_0 , λ , and z in (3.11) are invariant. The method introduces a fundamental scale for x such that the eigenvector b spans the one-dimensional standardized factor space. As an important conclusion we observe that the method may be applied to the sample correlation matrix with the same result, since this is an arbitrarily scaled sample covariance matrix. Then, the maximum likelihood estimates of the variances will, however, not be equal to unity, in general.

For region (ii) we obtain in taking derivatives of (1.4) with respect to the real parameters in (2.13), equations of the form (3.1) = 0 and (3.2) = 0. Although in the counterpart of (3.1) b_1 , b_3 and $-(\lambda+1)^{-1}$ occur instead of β_1 , β_3 and $(\lambda+1)^{-1}$, multiplication by the imaginary factor $-i$, gives (3.1) = 0 with b purely imaginary.

The analogue of (3.5) will be again

$$RS^{-1}AS^{-1}Rb = (\lambda + 1)b \quad (3.12)$$

with $-1 < b'b = \lambda < 0$ and b imaginary. But also the corresponding real vector (b_1, b_1, b_3) satisfying $2b_1^2 + b_3^2 = -\lambda$ is an eigenvector of $B = RS^{-1}AS^{-1}R$.

The stationary value of (1.4) with respect to b , given S , is again equal to (3.7). Since $\ln(\lambda+1)-\lambda$ is increasing in λ between -1 and 0 the *smaller* non-zero eigenvalue $\lambda+1$ of B (with rank ≤ 2 and positive trace) will be chosen in (3.12).

When b_0 is defined as the imaginary optimal solution, given S , of (3.12), with $b_0'b_0 = \lambda$, the analogue of (3.10) will be

$$n \text{diag} (I - S^{-1}AS^{-1} + (\lambda + 1)^{-1}S^{-1}AS^{-1}b_0b_0') = 0 \quad (3.13)$$

which in principle will provide the solution of S . Since also here (3.8) is valid the test statistic is again (3.11) where $(\lambda+1)$ is the required smallest eigenvalue of $B = RS^{-1}AS^{-1}R$ and λ_i are the eigenvalues of $S^{-1}AS^{-1}$.

From (3.13) follows $3 - \sum_{i=1}^3 \lambda_i + \lambda = 0$ which, by $-1 < \lambda < 0$, leads to the restriction $2 < \sum_{i=1}^3 (a_{ii}/\sigma_i^2) < 3$ on the σ_i^2 , while in addition we have $\sigma_3^2 \geq a_{33}$.

For region (iii) we take derivatives of (1.4) by (2.15) with respect to σ_1^{-1} , σ_2^{-1} , σ_3^{-1} , β_1 and $-b_3$. Since (2.15) has the same structure as (2.11), the counterpart of (3.1) is

$$n(\lambda + 1)^{-1} R(T^{-1}AT^{-1} - T^{-1}CT^{-1}) \begin{pmatrix} \beta_1 \\ \beta_1 \\ -b_3 \end{pmatrix} \quad (3.14)$$

where

$$T = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & s_3 \end{pmatrix} \quad (3.15)$$

while

$$\frac{\delta f}{\delta \text{diag } T^{-1}} = n(\text{diag } T - \text{diag } AC^{-1}T) \quad (3.16)$$

The matrix C as defined in (2.3) can be written:

$$\begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & s_3 \end{pmatrix} \left[\begin{pmatrix} \beta_1 \\ \beta_1 \\ b_3 \end{pmatrix} (\beta_1, \beta_1, b_3) + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & s_3 \end{pmatrix} \right] \quad (3.17)$$

On substituting (3.17) into (3.14) set equal to zero, and using (2.4), i.e. $2\beta_1^2 - b_3^2 = \lambda$, we get

$$R(T^{-1}AT^{-1} \begin{pmatrix} \beta_1 \\ \beta_1 \\ -b_3 \end{pmatrix} - (\lambda + 1) \begin{pmatrix} \beta_1 \\ \beta_1 \\ b_3 \end{pmatrix}) = 0 \quad (3.18)$$

Now multiply the last equation of (3.18) by the imaginary i . This implies multiplication by i in the third row of $T^{-1}AT^{-1}$ or in the third row of the left T^{-1} . A similar multiplication in the third column of the right T^{-1} will be compensated by division of i into the third element $-b_3$ of the adjoining column. So we obtain

$$RS^{-1}AS^{-1}Rb = (\lambda + 1)b \quad (3.19)$$

analogously to (3.5) and (3.12) where b and S are partially real and partially imaginary, satisfying $\lambda = b'b < -1$.

The complex matrix $B = RS^{-1}AS^{-1}R$ (of rank ≤ 2) of which b is an eigenvector, is Hermitian and has real eigenvalues. The stationary value of (1.4) with respect to b , given S , is equal to

$$\begin{aligned} & -\frac{1}{2}n\{3\ln 2\pi + \ln| -S^2| + \ln| -S^{-1}CS^{-1}| + \text{tr}(S^{-1}AS^{-1})(SC^{-1}S)\} \\ & = -\frac{1}{2}n\{3\ln 2\pi + \ln \sigma_1^2 + \ln \sigma_2^2 + \ln s_3^2 + \ln(-\lambda - 1) + \text{tr}(S^{-1}AS^{-1}) - \lambda\} \end{aligned}$$

analogously to (3.7). The function to be maximized with respect to λ is $-\ln(-\lambda-1) + \lambda$ with derivative $\lambda(\lambda+1)^{-1}$ and is increasing in $\lambda < -1$. Hence in (3.19) we choose the larger negative eigenvalue $(\lambda+1)$, which in practice is the *smallest* eigenvalue, since the other non-zero eigenvalue will turn out to be positive. On substitution into (3.16) of this optimal solution b_0 , given S , we obtain, in analogy with (3.10) and (3.13), as the derivative of the maximum of f , given S , with respect to $\ln \sigma_1^{-1}$, $\ln \sigma_2^{-1}$, $\ln s_3^{-1}$

$$n \operatorname{diag} (\mathbf{I} - \mathbf{S}^{-1} \mathbf{A} \mathbf{S}^{-1} + (\lambda + 1)^{-1} \mathbf{S}^{-1} \mathbf{A} \mathbf{S}^{-1} \mathbf{b}_0 \mathbf{b}'_0) \quad (3.20)$$

which, set equal to zero, will provide the solution of S . Since (3.8) is valid the test statistic will be

$$\begin{aligned} z &= n(\ln|\mathbf{C}_m| - \ln|\mathbf{A}|) = n(\ln|-\mathbf{S}^{-1} \mathbf{C}_m \mathbf{S}^{-1}| - \ln|-\mathbf{S}^{-1} \mathbf{A} \mathbf{S}^{-1}|) \\ &= n(\ln(-\lambda-1) - \sum_{i=1}^3 \ln|\lambda_i|) \end{aligned} \quad (3.21)$$

where λ_i are the eigenvalues of $\mathbf{S}^{-1} \mathbf{A} \mathbf{S}^{-1}$.

From (3.20) follows $3 - \operatorname{tr}(\mathbf{S}^{-1} \mathbf{A} \mathbf{S}^{-1}) + \lambda = 0$, which by $\lambda < 1$, leads to the restriction $a_{11}/\sigma_1^2 + a_{22}/\sigma_2^2 < 2 + a_{33}/s_3^2$. Consideration of the third diagonal element separately gives the uninteresting inequality $1 + s_3^{-2} a_{33} = b_3^2 > 0$.

As to the equations (3.10), (3.13) and (3.20) it may be noted that in each the third term of the form $(\lambda+1)^{-1} \mathbf{S}^{-1} \mathbf{A} \mathbf{S}^{-1} \mathbf{b}_0 \mathbf{b}'_0$ can be considered as a modified approximation of the matrix $\mathbf{S}^{-1} \mathbf{A} \mathbf{S}^{-1} - \mathbf{I}$. If we would approximate the matrix $\mathbf{R}(\mathbf{S}^{-1} \mathbf{A} \mathbf{S}^{-1} - \mathbf{I})\mathbf{R}$, with eigenvalue λ and eigenvector \mathbf{b} , $\lambda+1$ being the corresponding non-zero eigenvalue of $\mathbf{B} = \mathbf{R} \mathbf{S}^{-1} \mathbf{A} \mathbf{S}^{-1} \mathbf{R}$, by one principal component, we would use $\lambda(\mathbf{b}' \mathbf{b})^{-1} \mathbf{b} \mathbf{b}'$. This, according to (2.4), equals $(\lambda+1)^{-1} (\lambda+1) \mathbf{b} \mathbf{b}' = (\lambda+1)^{-1} (\mathbf{B} \mathbf{b}) \mathbf{b}'$. Hence the modification consists of using an eigenvector from \mathbf{B} instead of $\mathbf{S}^{-1} \mathbf{A} \mathbf{S}^{-1}$ and next $\mathbf{S}^{-1} \mathbf{A} \mathbf{S}^{-1}$ instead of \mathbf{B} in the approximation.

The remaining equations (3.10), (3.13) and (3.20) express the requirement that this modified approximation be exact for the diagonal elements, by a suitable choice of S .

4. NUMERICAL SOLUTION OF THE EQUATIONS

The solution of the *equation* (3.5) and its counterparts which must provide repeatedly the optimal b_0 and λ , given S , and are all of type $\mathbf{B} \mathbf{b} = (\lambda+1) \mathbf{b}$ is fairly simple.

Let for regions (i) and (ii):

$$\mathbf{B} = \begin{pmatrix} p & p & q \\ p & p & q \\ q & q & r \end{pmatrix} \text{ and } \mathbf{b} = \begin{pmatrix} \beta_1 \\ \beta_1 \\ \beta_3 \end{pmatrix} \quad (4.1)$$

Then $Bb = \mu b$ is equivalent to

$$\begin{cases} 2p\beta_1 + b\beta_3 = \mu\beta_1 \\ 2q\beta_1 + r\beta_3 = \mu\beta_3 \end{cases}$$

i.e. $(\beta_1, \beta_3)'$ should be eigenvector with eigenvalue μ of the matrix

$$\begin{pmatrix} 2p & q \\ 2q & r \end{pmatrix}$$

Hence

$$\mu = \frac{1}{2} [2p + r \pm \{(2p-r)^2 + 8q^2\}^{\frac{1}{2}}] \quad (4.2)$$

and (β_1, β_3) is proportional to

$$(2p - r + \{(2p - r)^2 + 8q^2\}^{\frac{1}{2}}, 4q) \quad (4.3)$$

and

$$(2p - r - \{(2p - r)^2 + 8q^2\}^{\frac{1}{2}}, 4q) \quad (4.4)$$

corresponding to the larger and the smaller eigenvalue μ_1 and μ_2 respectively.

In region (i) the larger value, in (ii) the smaller will be taken, while (4.3) will be normalized such that $2\beta_1^2 + \beta_3^2 = \lambda = \mu_1 - 1$ and (4.4) such that $2\beta_1^2 + \beta_3^2 = \lambda = (\mu_2 - 1)$, or since we prefer to work with real numbers, $2b_1^2 + b_3^2 = -\lambda = -(\mu_2 - 1)$.

For region (iii) we have to solve

$$R \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & i \end{pmatrix} T^{-1} A T^{-1} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & i \end{pmatrix} R b = \mu b$$

equivalent to

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & i \end{pmatrix} R T^{-1} A T^{-1} R \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & i \end{pmatrix} b = \mu b \quad (4.5)$$

Let

$$R T^{-1} A T^{-1} R = \begin{pmatrix} p & p & q \\ p & p & q \\ q & q & r \end{pmatrix} \text{ and } b = \begin{pmatrix} \beta_1 \\ \beta_1 \\ i\beta_3 \end{pmatrix} \quad (4.6)$$

Then $Bb = \mu b$ is equivalent to

$$\begin{cases} 2p\beta_1 - qb_3 = \mu\beta_1 \\ 2iq\beta_1 - irb_3 = i\mu b_3 \end{cases}$$

Hence $(\beta_1, b_3)'$ must be eigenvector with eigenvalue μ of the matrix

$$\begin{pmatrix} 2p & -q \\ 2q & -r \end{pmatrix}$$

So

$$\mu = \frac{1}{2}[2p - r \pm \{(2p + r)^2 - 8q^2\}^{\frac{1}{2}}] \quad (4.7)$$

and (β_1, b_3) is accordingly proportional to

$$(2p + r \pm \{(2p + r)^2 - 8q^2\}^{\frac{1}{2}}, 4q) = (k_1, k_3) \quad (4.8)$$

As to the sign to be chosen, we note that the requirement $2k_1^2 - k_3^2 < 0$ on (4.8) can be reduced to $\{(2p + r)^2 - 8q^2\}^{\frac{1}{2}} \pm (2p + r) < 0$. Since $(2p + r)$ is positive ($T^{-1}AT^{-1}$ is positive definite) the inequality holds only with the minus sign. The other non-zero eigenvalue will be positive, since if it were not, it would provide a second and, according to the previous section, even better solution of the problem at hand. This proves our previous statement. So we must take the minus sign in (4.7) and (4.8), the last being normalized such that $2\beta_1^2 - b_3^2 = \lambda = \mu - 1$, i.e. multiplication of (4.8) by $\{(\mu - 1)/(2k_1^2 - k_3^2)\}^{\frac{1}{2}}$.

In conclusion it may be remembered that p, q and r as required in (4.2), (4.3), (4.4) or (4.7), (4.8) are immediately obtained by averaging elements of the real $S^{-1}AS^{-1}$ or $T^{-1}AT^{-1}$.

Coming to the solution of the equations (3.10), (3.13) and (3.20) for S , we first report that a method suggested by factor analysis i.e. iterative solution by

$$\text{diag} (I - S_{j+1}^{-2}A + (\lambda_j + 1)^{-1}S_j^{-1}AS_j^{-1}b_jb_j') = 0 \quad (4.9)$$

with $j = 1, 2, \dots$, S_1 being a suitably chosen initial value of S , and b_j the optimal solution for b (real or complex) of (3.5) given S_j , or modifications of this functional iteration fail to converge in many cases or are downright divergent. Therefore we have recourse to another method also found very useful in factor analysis where similar difficulties arose, namely the method of steepest ascent.

Let

$$u_j = \text{diag} \{I - (S_j^{-1}AS_j^{-1} - (\lambda_j + 1)^{-1}S_j^{-1}AS_j^{-1}b_jb_j')\} \quad (4.10)$$

with elements u_{ij} ($i = 1, 2, 3$), which has the direction of the real partial derivatives of the maximum of (1.4), given S_j , with respect to the elements of

$$y = \ln \text{diag } S^{-1} = (\ln \sigma_1^{-1}, \ln \sigma_2^{-1}, \ln \sigma_3^{-1})' \quad (4.11a)$$

for regions (i) and (ii), or

$$y = \ln \text{diag } T^{-1} = (\ln \sigma_1^{-1}, \ln \sigma_2^{-1}, \ln \sigma_3^{-1})' \quad (4.11b)$$

for region (iii). Then corresponding to each point y , equal to $\ln \text{diag } S_j^{-1}$ or $\ln \text{diag } T_j^{-1}$, and an initial value S_1 being given, one has to find a real number μ_j such that the function in the point given by

$$y_{j+1} = y_j + \mu_j u_j \quad (j = 1, 2, \dots) \quad (4.12)$$

which also defines S_{j+1}^{-1} by $\sigma_{i,j+1}^{-1} = \sigma_{i,j}^{-1} \exp(\mu_j u_{ij})$ and similarly for T_{j+1}^{-1} , has a maximum. This, in turn, requires the derivative with respect to μ_j

$$\varphi(\mu_j) = u_j' u_{j+1} \quad (4.13)$$

to vanish. Thus S_j and T_j will converge to a stationary point S or T , a solution of (3.10) or its counterparts. From the theory about factor analysis it follows readily that in each region there is only one maximum of f . Hence by the method of steepest ascent the only solution for each region will be obtained.

The solution of the non-linear equation $\varphi(\mu_j) = 0$ requires repeated computation, for each value of μ_j to be used, of the corresponding u_{j+1} and thus of λ_{j+1} and b_{j+1} , as has been considered in principle beforehand, although the actual computation of (4.10) needs some comment.

Before doing so, we report that the equation $\varphi(\mu_j) = 0$ could satisfactorily be solved by successively fitting a parabola through three points $\{\mu_j, \varphi(\mu_j)\}$ with $\mu_j = 0, 1, 2$ as a start, and replacing the point with the largest $|\varphi(\mu_j)|$ by the point with an abscissa equal to that of the intersection between parabola and μ_j -axis. Here we need to add that the intersection is defined as the intersection with the smallest positive abscissa, and that in case all three values of $\varphi(\mu_j)$ are positive the part of the parabola is taken on by the three straight lines through the three pairs of points.

In (4.10) the difference between the second and third term has to be unity elementwise. Consequently a reasonable stopping criterion is given by $u_j' u_j$ which has to be sufficiently small. The choice of (3.10) and its counterparts instead of (3.2) is justified by the resulting symmetry but still more by the balance in the requirements on its individual elements and thus on those of S .

In the applications the method of steepest ascent, described above, turned out to be very efficient in obtaining a local improvement of S . In many cases, however, the maximization of the function is similar to a trip along a surface with one very slowly rising and winding ridge, but with very steep slopes on both sides of that ridge. The method of steepest ascent corrects fast for slipping off from these slopes, which is the highly necessary local improvement referred

to above, but the progress along the edge of the ridge can be rather slow. In order to overcome this difficulty we mixed the iterative process above with application of Aitken's acceleration procedure on three steps of that process of which two consecutive steps are two cycles apart, that is on S_j^{-1} , S_{j+2}^{-1} and S_{j+4}^{-1} , giving a new start S_1^{-1} of the process. Each element a of such a S_1^{-1} is calculated from the corresponding elements a_1 , a_2 and a_3 of S_j^{-1} , S_{j+2}^{-1} , S_{j+4}^{-1} as $(a_1 a_3 - a_2^2) (a_1 + a_3 - 2a_2)^{-1}$ from the guess that $(a - a_1)/(a - a_2) = (a - a_2)/(a - a_3)$. We took the steps two cycles apart in order to decrease considerably the frequency of the event $|a_2 - a_3| \geq |a_1 - a_2|$. In that case we took $a = (a_1 + a_2 + a_3)/3$. This mixture turned out to be quite successful, as will be illustrated by the numerical results. When no other information is available, the identity matrix I might be taken for S_1 or T_1 .

As to the actual calculation of (4.10), this is direct from (4.2), (4.3) and (4.4) for the regions (i) and (ii) when keeping in mind that in (ii) b_j is understood to be imaginary. Hence when using a real vector (b_1, b_1, b_3) instead of b in that case, as will be done in practice in using (4.2) (with minus sign) and (4.4), the minus sign in front of the third term of (4.10) has to be changed into a plus sign. In region (iii) the expression (4.10) as well as (3.20) contains S and b , both partially real and partially imaginary. On substituting

$$S = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & i \end{pmatrix} T \text{ and } b = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & i \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_1 \\ b_3 \end{pmatrix}$$

into (3.20) and using the fact that, if A and C are diagonal matrices, then $\text{diag}(ABC) = \text{diag}(CAB)$ we obtain

$$\text{diag } I - \text{diag} \left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} T^{-1} A T^{-1} \right) +$$

$$(\lambda + 1)^{-1} \text{diag} \left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}^{-1} A T^{-1} \begin{pmatrix} \beta_1 \\ \beta_1 \\ -b_3 \end{pmatrix} (\beta_1, \beta_2, b_3) \right).$$

With

$$T^{-1} A T^{-1} = (\alpha_{ij}) \tag{4.14}$$

which is a real matrix, we get the following real expansion for the elements of (3.20) for region (iii) and similarly for (4.10):

$$\begin{aligned}
& 1 - \alpha_{11} + (\lambda + 1)^{-1} \beta_1 (\alpha_{11} \beta_1 + \alpha_{12} \beta_1 - \alpha_{13} b_3) \\
& 1 - \alpha_{22} + (\lambda + 1)^{-1} \beta_1 (\alpha_{21} \beta_1 + \alpha_{22} \beta_1 - \alpha_{23} b_3) \\
& 1 + \alpha_{33} - (\lambda + 1)^{-1} b_3 (\alpha_{31} \beta_1 + \alpha_{32} \beta_2 - \alpha_{33} b_3)
\end{aligned} \tag{4.15}$$

β_1, b_3 and λ being given by (4.7) and (4.8).

In obtaining the maximum likelihood estimate C_m for C by the parameters of (2.3) we have also estimates r_{12} for the nuisance parameter ρ_{12} , and r_{13} for $\rho_{13} = \rho_{23}$. They are for the regions (i), (ii) and (iii) explicitly:

	r_{12}	$r_{13} = r_{23}$
(i)	$\beta_1^2 (\beta_1^2 + 1)^{-1}$	$\beta_1 \beta_3 \{(\beta_1^2 + 1) (\beta_3^2 + 1)\}^{-\frac{1}{2}}$
(ii)	$b_1^2 (b_1^2 - 1)^{-1}$	$-b_1 b_3 \{(b_1^2 - 1) (b_3^2 - 1)\}^{-\frac{1}{2}}$
(iii)	$\beta_1^2 (\beta_1^2 + 1)^{-1}$	$\beta_1 \beta_3 \{(\beta_1^2 + 1) (b_3^2 - 1)\}^{-\frac{1}{2}}$

as to be deduced from the corresponding representations of C .

5. NUMERICAL ILLUSTRATION

In Table 1 the results of the computation are given for (sample) correlation matrices A all with $|a_{12}| = .15$, $|a_{13}| = .55$ and $|a_{23}| = .33$. This example is due to Dr. G. Hamming who presented it to the author as one that (although positive definite) will not fit the usual model of factor analysis with $k = 1$, if all elements are positive.

Four such matrices A (1, 2, 3, 4) have been considered with varying signs of a_{ij} , denoted $sg(i, j)$.

The four numbers in each cell corresponding to one A and one region (i, ii and iii) of C are successively from top to bottom:

- a. the value of z/n where z is the test statistic (1.7)
- b. the maximum likelihood estimate r_{12} of ρ_{12}
- c. the maximum likelihood estimate r_{13} for $\rho_{13} = \rho_{23}$
- d. the number of times Aitken's acceleration was necessary in order to make $u'_j u_j$, defined by (4.10) as a stopping criterion, smaller than 10^{-6} , with initial value for S_1 or T_1 equal to I .

The starred cells correspond to the maximum likelihood estimates, given A , under the null hypothesis (1.2): $\rho_{13} = \rho_{23}$. The best fitting region seems to take the least amount of computing time although the reverse is not true.

The cells with, in particular, a pair of stars correspond to the maximum likelihood estimates under the wider null hypothesis (1.1): $\rho_{13}^2 = \rho_{23}^2$. For this purpose the columns 1 and 4 belong together, as well as 2 and 3, since the corresponding matrices are reducible into each other by changing the sign of x_2 in both cases.

TABLE 1.

Values of the test statistic and corresponding maximum likelihood estimates of correlation coefficients for the three regions of the parameter space and at four combinations of signs of the elements of a sample correlation matrix with fixed absolute values.

		1	2	3	4
region \ A	sg(1,2)	+	+	-	-
	sg(1,3)	+	+	+	+
	sg(2,3)	+	-	+	-
(i)	a	.0889	1.2488*	.4765	.9046
	b	.1986	.1496	.1962	.0162
	c	.4424	.1202	.4421	.1184
	d	9	0	4	3
(ii)	a	.1247	1.2932	.0617*	.8438*
	b	-.0027	-.0017	-.1508*	-.1493
	c	.3929	.1023	.4435	.1243
	d	8	5	2	0
(iii)	a	.0788*	1.2897	.1070	.8899
	b	.1585*	.0094	.0021	.0003
	c	.4438	.1164	.5028	.1454
	d	1	1	10	2

Other combinations of signs are not given since the corresponding results follow trivially from those given here by changing the signs of both x_1 and x_2 , and thus of a_{13} and a_{23} ; consequently only the third element of b changes sign in C .

Although r_{13} turns out to be not too far removed from the average of a_{13} and a_{23} , this at itself would not be sufficient for the computation of the test statistic, since the estimation of more parameters is necessary.

Finally we present the estimates of $\text{diag } S^{-1}$ corresponding to the two maximum likelihood estimates under the hypothesis (1.1) i.e. in columns 1 and 3, respectively, viz.

$$(1.121; 1.059; 2.030 \text{ i}) \text{ and } (.969; .893; .659)$$

ACKNOWLEDGEMENTS

The author would like to thank Dr. F. Gebhardt and Dr. W. Böge for some stimulating remarks as to previous versions of the treatment of this subject. He is most grateful to Mr. L. P. Kamil and Mr. T. A. Reesink for their assistance in solving the numerical problems and the actual computations carried out on the IBM 1620 computer of the Department of Mathematics of the Agricultural University. Some discussion with Dr. Ingram Olkin, oral as well as in writing, is acknowledged in gratitude.

REFERENCES

1. AITKIN, M. A., NELSON, W. C. and REINFURT, K. H. (1968). Tests for correlation matrices *Biometrika* **55**: 327-334.
2. HEALY, M. J. R. (1955). A significance test for the difference in efficiency between two predictors. *J. Roy. Statist. Soc. Ser. B* **17**: 266-268.
3. HEALY, M. J. R. (1959). Comment on Dr. Williams paper. *J. Roy. Statist. Soc. Ser. B* **21**: 399.
4. HOTELLING, H. (1940). The selection of variates for use in prediction with some comments on the general problem of nuisance parameters. *Ann. Math. Statist.* **11**: 271-283.
5. KENDALL, M. G. and STUART, A. (1961). *The advanced Theory of Statistics*. **2** (2nd ed.). Griffin, London.
6. WILLIAMS, E. J. (1959). *Regression Analysis*. Wiley. New York.
7. WILLIAMS, E. J. (1959). The comparison of regression variables. *J. Roy. Statist. Soc. Ser. B* **21**: 396-399.
8. WILLIAMS, E. J. (1959). Significance of difference between two non-independent correlation coefficients (Query). *Biometrics* **15**: 135-136.