

ESTIMATING LINEAR RESTRICTIONS ON REGRESSION COEFFICIENTS FOR MULTIVARIATE NORMAL DISTRIBUTIONS¹

BY T. W. ANDERSON

Columbia University

Summary. In this paper linear restrictions on regression coefficients are studied. Let the $p \times q_2$ matrix of coefficients of regression of the p dependent variates on q_2 of the independent variates be² $\bar{\mathbf{B}}_2$. Maximum likelihood estimates of an $m \times p$ matrix $\mathbf{\Gamma}$ satisfying $\mathbf{\Gamma}'\bar{\mathbf{B}}_2 = \mathbf{0}$ and certain other conditions are found under the assumption that the rank of $\bar{\mathbf{B}}_2$ is $p - m$ and the dependent variates are normally distributed (Section 2). Confidence regions for $\mathbf{\Gamma}$ under various conditions are obtained (Section 5). The likelihood ratio test of the hypothesis that the rank of $\bar{\mathbf{B}}_2$ is a given number is obtained (Section 3). A test of the hypothesis that $\mathbf{\Gamma}$ is a certain matrix is given (Section 4). These results are applied to the “ q -sample problem” (Section 7) and are extended for certain econometric models (Section 6).

1. Introduction.

1.1. *Univariate analysis of variance.* A large number of problems of univariate statistics can be put into the form of analysis of variance or regression analysis. We assume that

$$(1.1) \quad \varepsilon x_\alpha = \boldsymbol{\beta}' \mathbf{z}_\alpha,$$

where $\boldsymbol{\beta}$ and \mathbf{z}_α are column vectors³ of q components, that $\varepsilon(x_\alpha - \boldsymbol{\beta}' \mathbf{z}_\alpha)^2 = \sigma^2$, and that x_α is uncorrelated with x_γ ($\alpha \neq \gamma$). On the basis of a sample, that is, a set⁴ $x_1, \dots, x_N; z_1, \dots, z_N$ (where there are q linearly independent \mathbf{z}_α), we may test hypotheses about $\boldsymbol{\beta}$, we may obtain point estimates of $\boldsymbol{\beta}$, or we may find a confidence region for the vector $\boldsymbol{\beta}$. It is well known that this model is sufficiently general to include the analysis of variance model for fixed effects. The usual point estimate \mathbf{b} of $\boldsymbol{\beta}$ is defined by

¹ By invitation parts of this paper were presented to the Cleveland Meeting of the Institute of Mathematical Statistics, December 30, 1948. Some of these results were included in the thesis [1] submitted to the Mathematics Department of Princeton University in partial fulfillment of the requirements for the degree of Doctor of Philosophy, June, 1945; some other results were included in dittoed papers at the Cowles Commission for Research in Economics (of which organization the author is a research consultant). This paper will be reprinted as Cowles Commission Paper, New Series, No. 50.

² We use $\bar{\mathbf{B}}$ to distinguish the capital “beta” (a matrix of parameters) from \mathbf{B} (a matrix of estimates). Matrices are indicated by boldface capital letters, and vectors are indicated by boldface lower case letters.

³ $\boldsymbol{\beta}'$ is the transpose of $\boldsymbol{\beta}$.

⁴ We use the same notation for observations as for random variables with the hope that the reader can easily distinguish between the uses on the basis of the context.

$$(1.2) \quad \mathbf{b}' \sum_{\alpha=1}^N \mathbf{z}_\alpha \mathbf{z}'_\alpha = \sum_{\alpha=1}^N x_\alpha \mathbf{z}'_\alpha.$$

The estimate s^2 of σ^2 is given by

$$(1.3) \quad (N - q)s^2 = \sum_{\alpha=1}^N (x_\alpha - \mathbf{b}'\mathbf{z}_\alpha)^2 = \sum_{\alpha=1}^N x_\alpha^2 - \mathbf{b}' \left(\sum_{\alpha=1}^N \mathbf{z}_\alpha \mathbf{z}'_\alpha \right) \mathbf{b}.$$

Consider testing the hypothesis $\boldsymbol{\beta}_2 = \mathbf{0}$, where $\boldsymbol{\beta}_2$ is a subvector of $\boldsymbol{\beta}$ with q_2 components; that is, $\boldsymbol{\beta}' = (\boldsymbol{\beta}'_1 \boldsymbol{\beta}'_2)$. We partition \mathbf{b} and \mathbf{z}_α similarly as $\mathbf{b}' = (\mathbf{b}'_1 \mathbf{b}'_2)$ and $\mathbf{z}'_\alpha = (\mathbf{z}'_{1\alpha} \mathbf{z}'_{2\alpha})$. We use the statistic

$$(1.4) \quad F = \frac{\mathbf{b}'_2 \mathbf{Q} \mathbf{b}_2}{q_2 s^2},$$

where

$$(1.5) \quad \mathbf{Q} = \sum_{\alpha=1}^N \mathbf{z}_{2\alpha} \mathbf{z}'_{2\alpha} - \sum_{\alpha=1}^N \mathbf{z}_{2\alpha} \mathbf{z}'_{1\alpha} \left(\sum_{\alpha=1}^N \mathbf{z}_{1\alpha} \mathbf{z}'_{1\alpha} \right)^{-1} \sum_{\alpha=1}^N \mathbf{z}_{1\alpha} \mathbf{z}'_{2\alpha}.$$

This statistic has the F -distribution with q_2 and $N - q$ degrees of freedom if the null hypothesis is true and if $\{x_\alpha\}$ are normally distributed. The above statistic is equivalent to

$$(1.6) \quad \frac{1}{\frac{q_2}{N - q} F + 1} = \frac{\sum_{\alpha=1}^N (x_\alpha - \mathbf{b}'\mathbf{z}_\alpha)^2}{\sum_{\alpha=1}^N (x_\alpha - \mathbf{b}'_1 \mathbf{z}_{1\alpha})^2},$$

where \mathbf{b}'_1 is defined by

$$(1.7) \quad \mathbf{b}'_1 \sum_{\alpha=1}^N \mathbf{z}_{1\alpha} \mathbf{z}'_{1\alpha} = \sum_{\alpha=1}^N x_\alpha \mathbf{z}'_{1\alpha}.$$

The numerator is proportional to the estimate of σ^2 when we do not believe the hypothesis to be true, and the denominator is proportional to the estimate of σ^2 when we do believe the hypothesis to be true. We reject the null hypothesis when the observed F is greater than $F_{q_2, N-q}(\epsilon)$, the significance point of the F -distribution with q_2 and $N - q$ degrees of freedom corresponding to a significance level ϵ .

We can also test the hypothesis that $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_2^0$, any arbitrary given vector. A confidence region for $\boldsymbol{\beta}_2$ consists of all those $\boldsymbol{\beta}_2^0$ such that the corresponding hypothesis is not rejected on the basis of our sample.

1.2. *Multivariate analysis of variance.* Now let us turn to the generalization of the analysis of variance to be used in the treatment of a vector variate. The expected value of a vector variate \mathbf{x}_α with p components is assumed to be

$$(1.8) \quad \mathbf{E} \mathbf{x}_\alpha = \bar{\mathbf{B}} \mathbf{z}_\alpha,$$

where $\bar{\mathbf{B}}$ is a $p \times q$ matrix of regression coefficients. We further assume that

$$(1.9) \quad \varepsilon(\mathbf{x}_\alpha - \bar{\mathbf{B}}\mathbf{z}_\alpha)(\mathbf{x}_\alpha - \bar{\mathbf{B}}\mathbf{z}_\alpha)' = \boldsymbol{\Sigma},$$

a positive definite matrix, and that \mathbf{x}_α is uncorrelated with $\mathbf{x}_\gamma (\alpha \neq \gamma)$. As in the univariate case we may wish to estimate the regression coefficients. The estimate of each row of $\bar{\mathbf{B}}$ is of the form (1.2). Thus \mathbf{B} , the estimate of $\bar{\mathbf{B}}$, is defined by

$$(1.10) \quad \mathbf{B} \sum_{\alpha=1}^N \mathbf{z}_\alpha \mathbf{z}'_\alpha = \sum_{\alpha=1}^N \mathbf{x}_\alpha \mathbf{z}'_\alpha.$$

The estimate of \mathbf{S} of $\boldsymbol{\Sigma}$ is given by

$$(1.11) \quad (N - q)\mathbf{S} = \sum_{\alpha=1}^N (\mathbf{x}_\alpha - \mathbf{B}\mathbf{z}_\alpha)(\mathbf{x}_\alpha - \mathbf{B}\mathbf{z}_\alpha)'$$

If $\{\mathbf{x}_\alpha\}$ are normally distributed, \mathbf{B} has a normal distribution with mean $\bar{\mathbf{B}}$ and $(N - q)\mathbf{S} = \mathbf{A}$ has a Wishart distribution with $\boldsymbol{\Sigma}$ covariance matrix and $N - q$ degrees of freedom.

To test the hypothesis $\bar{\mathbf{B}}_2 = \mathbf{0}$, where $\bar{\mathbf{B}}_2$ is the second submatrix in $\bar{\mathbf{B}} = (\bar{\mathbf{B}}_1 \bar{\mathbf{B}}_2)$, we use a generalization of the statistic $\left(\frac{q_2}{N - q} F + 1\right)^{-1}$, which was used in the univariate case, namely,

$$(1.12) \quad U = \frac{\left| \sum_{\alpha=1}^N (\mathbf{x}_\alpha - \mathbf{B}\mathbf{z}_\alpha)(\mathbf{x}_\alpha - \mathbf{B}\mathbf{z}_\alpha)' \right|}{\left| \sum_{\alpha=1}^N (\mathbf{x}_\alpha - \mathbf{B}_1^* \mathbf{z}_\alpha)(\mathbf{x}_\alpha - \mathbf{B}_1^* \mathbf{z}_\alpha)' \right|},$$

where \mathbf{B}_1^* is defined by

$$(1.13) \quad \mathbf{B}_1^* \sum_{\alpha=1}^N \mathbf{z}_{1\alpha} \mathbf{z}'_{1\alpha} = \sum_{\alpha=1}^N \mathbf{x}_\alpha \mathbf{z}'_{1\alpha}.$$

The matrix in the numerator of U is proportional to the estimate of $\boldsymbol{\Sigma}$ when the null hypothesis is not believed true, and the matrix in the denominator of U is proportional to the estimate of $\boldsymbol{\Sigma}$ when the null hypothesis is believed true. We reject the hypothesis $\bar{\mathbf{B}}_2 = \mathbf{0}$ when the observed U is less than $U_{p, q_2, N - q}(\epsilon)$, the significance point at significance level ϵ . The distribution of U has been given by Wilks in many special cases [16]; Rao has given an approximation to $U_{p, q_2, N - q}(\epsilon)$ based on an asymptotic expansion of the distribution of $\log U$ [14]. The likelihood ratio criterion is $U^{1/N}$.

We can also test the hypothesis that $\bar{\mathbf{B}}_2 = \bar{\mathbf{B}}_2^0$, an arbitrary given matrix. A confidence region for $\bar{\mathbf{B}}_2$ consists of all those $\bar{\mathbf{B}}_2^0$ such that the corresponding hypothesis is not rejected on the basis of our sample.

1.3. *Rank of the regression matrix; linear restrictions.* If $\bar{\mathbf{B}}_2 \neq \mathbf{0}$, there enters into the multivariate case a new feature which does not appear in the univariate case. There may be some of the rows of $\bar{\mathbf{B}}_2$ that are zero; that is, there may be some components of \mathbf{x}_α that have expected values independent of $\mathbf{z}_{2\alpha}$. More

generally, all of the elements of $\bar{\mathbf{B}}_2$ may be different from zero, but the rank of $\bar{\mathbf{B}}_2$ may be less than the maximum possible. That implies that it is possible to take a linear combination of components of \mathbf{x}_α such that the expected value of this linear combination is independent of $\mathbf{z}_{2\alpha}$. If the rank of $\bar{\mathbf{B}}_2$ is less than p , there is a vector $\boldsymbol{\gamma}$ (of p components) such that

$$(1.14) \quad \boldsymbol{\gamma}' \bar{\mathbf{B}}_2 = 0.$$

From (1.8) we obtain

$$(1.15) \quad \boldsymbol{\varepsilon} \boldsymbol{\gamma}' \mathbf{x}_\alpha = \boldsymbol{\gamma}' \bar{\mathbf{B}} \mathbf{z}_\alpha = \boldsymbol{\gamma}' \bar{\mathbf{B}}_1 \mathbf{z}_{1\alpha}.$$

In general, if r is the rank of $\bar{\mathbf{B}}_2$ there are $p - r$ linearly independent vectors $\boldsymbol{\gamma}$ satisfying (1.14).

In this paper we are primarily interested in estimating $p - r$ linearly independent vectors satisfying (1.14). In Section 2 we find that the maximum likelihood estimates of these vectors (under certain normalization conditions) are the characteristic vectors of $\mathbf{B}_2 \mathbf{Q} \mathbf{B}_2'$ in the metric of \mathbf{S} . In Section 5 we find confidence regions for these vectors using the theory summarized in Section 1.2.

1.4. *Examples.* A number of multivariate problems can be thrown into the above form, some naturally, some a little unnaturally. As a simple example, consider a sample $\mathbf{U}_1, \dots, \mathbf{U}_n$ from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and a sample of the same size $\mathbf{V}_1, \dots, \mathbf{V}_n$ from $N(\boldsymbol{\nu}, \boldsymbol{\Sigma})$, where $N(\boldsymbol{\lambda}, \boldsymbol{\Sigma})$ denotes the normal distribution with mean $\boldsymbol{\lambda}$ and covariance matrix $\boldsymbol{\Sigma}$. We can describe this by the above model. Let $\mathbf{x}_\alpha = \mathbf{U}_\alpha$, $\mathbf{x}_{n+\alpha} = \mathbf{V}_\alpha$ ($\alpha = 1, \dots, n$). Let $z_{1\alpha} \equiv 1$, $\alpha = 1, \dots, 2n = N$; and let $z_{2\alpha} = 1$ for $\alpha = 1, \dots, n$, and $z_{2\alpha} = -1$ for $\alpha = n + 1, \dots, 2n$. Then the regression model holds with the first column of $\bar{\mathbf{B}}$ being $\frac{1}{2}(\boldsymbol{\mu} + \boldsymbol{\nu})$ and the second being $\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\nu})$. The hypothesis $\boldsymbol{\mu} = \boldsymbol{\nu}$ is transformed into the hypothesis of the regression of \mathbf{x}_α on $\mathbf{z}_{2\alpha}$ being zero. In this case $\bar{\mathbf{B}}_2$ consists of one column and is either of rank 0 or 1. The T^2 test that it is of rank 0 is a special case of the tests given in Section 3. Estimation of linear restrictions on $\bar{\mathbf{B}}_2$ is trivial and is a special case of the treatment in Section 2.

In a similar fashion we may treat certain three-sample problems. Suppose we have samples of size n from $N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$, $N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$, and $N(\boldsymbol{\mu}_3, \boldsymbol{\Sigma})$. Then $z_{1\alpha}$, $z_{2\alpha}$, and $z_{3\alpha}$ can be chosen so that the three columns of $\bar{\mathbf{B}}$ are $\boldsymbol{\beta}_1 = \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2 + \boldsymbol{\mu}_3$, $\boldsymbol{\beta}_2 = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$, $\boldsymbol{\beta}_3 = \boldsymbol{\mu}_2 - \boldsymbol{\mu}_3$. If the means lie on a line, that is, if $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = k(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_3)$ for some constant k , the rank of $(\boldsymbol{\beta}_2 \boldsymbol{\beta}_3)$ is one instead of two. In such a case there are $p - 1$ vectors that are annihilated by this $p \times 2$ matrix. The intersection of the planes perpendicular to these $p - 1$ vectors is, of course, the line joining the means. This vector is simply proportional to one of the last two columns of $\bar{\mathbf{B}}$. In Section 7 we give a more general consideration of "q-sample problems."

Now we shall consider a more elaborate problem that is naturally put into this pattern. Suppose that the workings of the economic system are such that the vector of economic variates \mathbf{x}_α has a mean $\bar{\mathbf{B}} \mathbf{z}_\alpha$, where \mathbf{z}_α consists of non-economic variates. $\bar{\mathbf{B}} \mathbf{z}_\alpha$ might be called the vector of "systematic parts" (or

it is called the “reduced form”). Suppose there is a vector γ such that $\gamma' \bar{\mathbf{B}} = \mathbf{0}$. Then the variate $\gamma' \mathbf{x}_\alpha$ has mean zero. Under certain conditions the equation $\gamma' \mathbf{x}_\alpha = v_\alpha$ is called a “structural equation.” In many such economic models one would like to include in the \mathbf{z}_α “lagged” values of \mathbf{x}_α . Although many of the results in this paper apply to such a case, we shall exclude such considerations in order to simplify the discussion.

2. Maximum likelihood estimates of the coefficients of the restrictions. We assume that \mathbf{x}_α is normally distributed with mean (1.8), that is, $\varepsilon \mathbf{x}_\alpha = \bar{\mathbf{B}}_1 \mathbf{z}_{1\alpha} + \bar{\mathbf{B}}_2 \mathbf{z}_{2\alpha} (\alpha = 1, \dots, N)$. Since the $\{\mathbf{z}_\alpha\}$ are fixed, we can make the transformation

$$(2.1) \quad \mathbf{v}_\alpha = \mathbf{z}_{2\alpha} - \sum_{\beta=1}^N \mathbf{z}_{2\beta} \mathbf{z}'_{1\beta} \left(\sum_{\beta=1}^N \mathbf{z}_{1\beta} \mathbf{z}'_{1\beta} \right)^{-1} \mathbf{z}_{1\alpha},$$

$$(2.2) \quad \bar{\mathbf{B}}_1^* = \bar{\mathbf{B}}_1 + \bar{\mathbf{B}}_2 \left(\sum_{\beta=1}^N \mathbf{z}_{2\beta} \mathbf{z}'_{1\beta} \right) \left(\sum_{\beta=1}^N \mathbf{z}_{1\beta} \mathbf{z}'_{1\beta} \right)^{-1}.$$

Then the expectation of \mathbf{x}_α is

$$(2.3) \quad \varepsilon \mathbf{x}_\alpha = \bar{\mathbf{B}}_1^* \mathbf{z}_{1\alpha} + \bar{\mathbf{B}}_2 \mathbf{v}_\alpha,$$

and $\mathbf{z}_{1\alpha}$ and \mathbf{v}_α are orthogonal (in our sample). We want to estimate the $p \times m$ matrix $\mathbf{\Gamma}$ such that

$$(2.4) \quad \mathbf{\Gamma}' \bar{\mathbf{B}}_2 = \mathbf{0}.$$

In order to avoid trivial estimates we require

$$(2.5) \quad \mathbf{\Gamma}' \mathbf{\Sigma} \mathbf{\Gamma} = \mathbf{I}.$$

There is still an indeterminacy because (2.4) and (2.5) are satisfied if $\mathbf{\Gamma}$ is replaced by the product of $\mathbf{\Gamma}$ on the right by an orthogonal matrix.

Before proceeding formally with the method of maximum likelihood, let us consider a more intuitive approach. We can make an analysis of variance for a linear combination $\mathbf{c}' \mathbf{x}_\alpha$. Then

$$(2.6) \quad \begin{aligned} \sum_{\alpha=1}^N (\mathbf{c}' \mathbf{x}_\alpha)^2 &= \sum_{\alpha=1}^N \mathbf{c}' \mathbf{x}_\alpha \mathbf{x}'_\alpha \mathbf{c} \\ &= \mathbf{c}' \mathbf{B}_1^* \sum_{\alpha=1}^N \mathbf{z}_{1\alpha} \mathbf{z}'_{1\alpha} \mathbf{B}_1^{*'} \mathbf{c} + \mathbf{c}' \mathbf{B}_2 \sum_{\alpha=1}^N \mathbf{v}_\alpha \mathbf{v}'_\alpha \mathbf{B}_2' \mathbf{c} + (N - q) \mathbf{c}' \mathbf{S} \mathbf{c}, \end{aligned}$$

where \mathbf{B}_1^* is defined by (1.13) and \mathbf{B}_2 , the usual estimate of $\bar{\mathbf{B}}_2$, is given by (1.10) or by

$$(2.7) \quad \mathbf{B}_2 = \sum_{\alpha=1}^N \mathbf{x}_\alpha \mathbf{v}'_\alpha \mathbf{Q}^{-1},$$

where \mathbf{Q} , given by (1.5), is $\sum_{\alpha=1}^N \mathbf{v}_\alpha \mathbf{v}'_\alpha$. The second term on the right in (2.6) is the “sum of squares of the v effects” and the third term is the “error sum of squares.” Since a restriction γ is such that the expected value of a “ v effect”

is 0, a good estimate of a γ would seem to be the vector \mathbf{c} that minimizes the “ v effect sum of squares” relative to the “error sum of squares”, that is, the \mathbf{c} that minimizes

$$(2.8) \quad \lambda = \frac{\mathbf{c}'\mathbf{B}_2\mathbf{Q}\mathbf{B}'_2\mathbf{c}}{\mathbf{c}'(\mathbf{N} - \mathbf{q})\mathbf{S}\mathbf{c}}.$$

The minimum ratio is the smallest root of

$$(2.9) \quad |\mathbf{B}_2\mathbf{Q}\mathbf{B}'_2 - \phi\mathbf{A}| = 0,$$

where $\mathbf{A} = (\mathbf{N} - \mathbf{q})\mathbf{S}$, and the vector \mathbf{c} is the corresponding characteristic vector satisfying

$$(2.10) \quad (\mathbf{B}_2\mathbf{Q}\mathbf{B}'_2 - \phi\mathbf{A})\mathbf{c} = \mathbf{0}.$$

If there are m linearly independent restrictions, we use the characteristic vectors associated with the m smallest roots of (2.10) normalized by

$$(2.11) \quad \mathbf{c}'\mathbf{A}\mathbf{c} = N.$$

THEOREM 1. *Suppose \mathbf{x}_α (of p components) is distributed according to $N(\bar{\mathbf{B}}_1\mathbf{z}_{1\alpha} + \bar{\mathbf{B}}_2\mathbf{z}_{2\alpha}, \boldsymbol{\Sigma})$ ($\alpha = 1, \dots, N$). Suppose $\bar{\mathbf{B}}_2$ is of rank $p - m$. Then a set of maximum likelihood estimates of $\bar{\mathbf{B}}_1, \bar{\mathbf{B}}_2, \boldsymbol{\Sigma}$, and $\boldsymbol{\Gamma}$ satisfying (2.4) and (2.5) are*

$$(2.12) \quad \hat{\bar{\mathbf{B}}}_1 = \sum_{\alpha} \mathbf{x}_{\alpha} \mathbf{z}_{1\alpha} (\sum_{\alpha} \mathbf{z}_{1\alpha} \mathbf{z}'_{1\alpha})^{-1} - \hat{\bar{\mathbf{B}}}_2 \sum_{\beta} \mathbf{z}_{2\beta} \mathbf{z}'_{1\beta} (\sum_{\alpha} \mathbf{z}_{1\alpha} \mathbf{z}'_{1\alpha})^{-1},$$

$$(2.13) \quad \hat{\bar{\mathbf{B}}}_2 = (\mathbf{I} - \hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{\Gamma}}')\mathbf{B}_2,$$

$$(2.14) \quad \hat{\boldsymbol{\Sigma}} = \mathbf{H} + \mathbf{H}\hat{\boldsymbol{\Gamma}}(\mathbf{I} + \boldsymbol{\Phi}^*)\boldsymbol{\Phi}^*\hat{\boldsymbol{\Gamma}}'\mathbf{H},$$

$$(2.15) \quad \hat{\boldsymbol{\Gamma}} = (\hat{\gamma}_{p-m+1}, \dots, \hat{\gamma}_p),$$

where \mathbf{B}_2 is given by (2.7) or (1.10), $\mathbf{H} = [(\mathbf{N} - \mathbf{q})/\mathbf{N}]\mathbf{S}$ is given by (1.11), $\boldsymbol{\Phi}^*$ is the diagonal matrix whose nonzero elements $(\phi_{p-m+1}, \dots, \phi_p)$ are the m smallest roots of (2.9), and $\hat{\gamma}_i$ are the corresponding vectors defined by (2.10) and normalized by $\hat{\gamma}'_i\mathbf{H}\hat{\gamma}_i = 1/(1 + \phi_i)$. $\hat{\boldsymbol{\Gamma}}$ may be multiplied on the right by any orthogonal matrix to obtain another maximum likelihood estimate of $\boldsymbol{\Gamma}$.

PROOF. We shall maximize the logarithm of the likelihood function⁵

$$(2.16) \quad \log L = -\frac{1}{2}Np \log 2\pi + \frac{1}{2}N \log |\boldsymbol{\Sigma}^{-1}| - \frac{1}{2} \sum_{\alpha=1}^N (\mathbf{x}_{\alpha} - \bar{\mathbf{B}}_1^*\mathbf{z}_{1\alpha} - \bar{\mathbf{B}}_2\mathbf{v}_{\alpha})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_{\alpha} - \bar{\mathbf{B}}_1^*\mathbf{z}_{1\alpha} - \bar{\mathbf{B}}_2\mathbf{v}_{\alpha})$$

with respect to $\boldsymbol{\Sigma}, \bar{\mathbf{B}}_1^*, \bar{\mathbf{B}}_2$, and $\boldsymbol{\Gamma}$, subject to restrictions (2.4) and (2.5). Let $\boldsymbol{\Phi}$ be an $m \times q_2$ matrix, and $\boldsymbol{\Psi}$ an $m \times m$ symmetric matrix of Lagrange multipliers. We shall maximize

$$(2.17) \quad f = \log L + \text{tr}(\boldsymbol{\Phi}\bar{\mathbf{B}}_2'\boldsymbol{\Gamma}) + \frac{1}{2}\text{tr}[\boldsymbol{\Psi}(\boldsymbol{\Gamma}'\boldsymbol{\Sigma}\boldsymbol{\Gamma} - \mathbf{I})]$$

⁵ The method is similar to that used in [3].

by taking partial derivatives and setting these equal to zero ("tr" denoting trace). The partial derivatives with respect to the elements of Γ set equal to zero are

$$(2.18) \quad \hat{\Phi} \hat{B}'_2 + \hat{\Psi} \hat{\Gamma}' \hat{\Sigma} = 0.$$

Multiplication on the right by $\hat{\Gamma}$ gives

$$(2.19) \quad \hat{\Phi} \hat{B}'_2 \hat{\Gamma} + \hat{\Psi} \hat{\Gamma}' \hat{\Sigma} \hat{\Gamma} = 0.$$

In view of (2.4) and (2.5), which are to be satisfied by \hat{B}_2 , $\hat{\Gamma}$, and $\hat{\Sigma}$, this shows that $\hat{\Psi} = 0$.

The partial derivatives of f with respect to the elements of Σ , \bar{B}_1^* , \bar{B}_2 , and Γ lead to

$$(2.20) \quad N \hat{\Sigma} = \sum_{\alpha} (x_{\alpha} - \hat{B}_1^* z_{1\alpha} - \hat{B}_2 v_{\alpha})(x_{\alpha} - \hat{B}_1^* z_{1\alpha} - \hat{B}_2 v_{\alpha})',$$

$$(2.21) \quad \hat{\Sigma}^{-1} \sum_{\alpha} x_{\alpha} z'_{1\alpha} = \hat{\Sigma}^{-1} \hat{B}_1^* \sum_{\alpha} z_{1\alpha} z'_{1\alpha},$$

$$(2.22) \quad \hat{\Sigma}^{-1} \sum_{\alpha} x_{\alpha} v'_{\alpha} - \hat{\Sigma}^{-1} \hat{B}_2 \sum_{\alpha} v_{\alpha} v'_{\alpha} + \hat{\Gamma} \hat{\Phi} = 0,$$

$$(2.23) \quad \hat{\Phi} \hat{B}'_2 = 0.$$

The solution of (2.21) for \hat{B}_1^* gives B_1^* . From (2.22) we obtain

$$(2.24) \quad \hat{B}_2 = \sum_{\alpha} x_{\alpha} v'_{\alpha} Q^{-1} + \hat{\Sigma} \hat{\Gamma} \hat{\Phi} Q^{-1}.$$

Multiplication on the left by $\hat{\Gamma}'$ and on the right by Q gives

$$(2.25) \quad 0 = \hat{\Gamma}' \sum_{\alpha} x_{\alpha} v'_{\alpha} + \hat{\Gamma}' \hat{\Sigma} \hat{\Gamma} \hat{\Phi}.$$

Thus

$$(2.26) \quad \hat{\Phi} = -\hat{\Gamma}' \sum_{\alpha} x_{\alpha} v'_{\alpha}.$$

Substitution into (2.24) gives

$$(2.27) \quad \hat{B}_2 = (I - \hat{\Sigma} \hat{\Gamma} \hat{\Gamma}') B_2.$$

In view of (2.26), (2.27), and (2.7), we derive from (2.23)

$$(2.28) \quad (I - \hat{\Sigma} \hat{\Gamma} \hat{\Gamma}') B_2 Q B'_2 \hat{\Gamma} = 0.$$

We now consider (2.20), which can be written

$$(2.29) \quad N \hat{\Sigma} = A + \hat{\Sigma} \hat{\Gamma}' \hat{\Gamma} B_2 Q B'_2 \hat{\Gamma} \hat{\Gamma}' \hat{\Sigma}.$$

It is clear that if $\hat{\Gamma}_1$, $\hat{\Sigma}$, \hat{B}_2 is one solution of (2.4), (2.5), (2.27), and (2.29), another solution is $\hat{\Gamma}_2$, $\hat{\Sigma}$, \hat{B}_2 , where $\hat{\Gamma}_2 = \hat{\Gamma}_1 O$ and O is any orthogonal matrix. In this class of solutions let us consider those which make

$$(2.30) \quad \frac{1}{N} \hat{\Gamma}' B_2 Q B_2' \hat{\Gamma} = D,$$

say, diagonal. Then (2.29) can be written

$$(2.31) \quad N \hat{\Sigma} = A + N \hat{\Sigma} \hat{\Gamma} D \hat{\Gamma}' \hat{\Sigma}.$$

Multiplication on the right by $\hat{\Gamma}$ gives

$$(2.32) \quad N \hat{\Sigma} \hat{\Gamma} = A \hat{\Gamma} + N \hat{\Sigma} \hat{\Gamma} D,$$

or

$$(2.33) \quad N \hat{\Sigma} \hat{\Gamma} (I - D) = A \hat{\Gamma}.$$

We can write (2.28) as

$$(2.34) \quad B_2 Q B_2' \hat{\Gamma} - N \hat{\Sigma} \hat{\Gamma} D = 0.$$

Since D is diagonal, multiplication of (2.34) on the right by $(I - D)$ gives

$$(2.35) \quad B_2 Q B_2' \hat{\Gamma} (I - D) - N \hat{\Sigma} \hat{\Gamma} (I - D) D = 0.$$

Substitution from (2.33) gives

$$(2.36) \quad B_2 Q B_2' \hat{\Gamma} (I - D) - A \hat{\Gamma} D = 0,$$

or

$$(2.37) \quad B_2 Q B_2' \hat{\Gamma} = (B_2 Q B_2' + A) \hat{\Gamma} D.$$

Thus, the columns of $\hat{\Gamma}$ satisfy

$$(2.38) \quad [B_2 Q B_2' - d(B_2 Q B_2' + A)] \hat{\gamma} = 0,$$

where d is a root of

$$(2.39) \quad | B_2 Q B_2' - d(B_2 Q B_2' + A) | = 0.$$

Let the roots of (2.39) be $d_1 \geq d_2 \geq \dots \geq d_p \geq 0$. Each column of $\hat{\Gamma}$ satisfies (2.10), where ϕ is one of

$$(2.40) \quad \phi_i = \frac{d_i}{1 - d_i}.$$

Let the solutions of (2.10), normalized by

$$(2.41) \quad \mathbf{c}'_i A \mathbf{c}_j = N \delta_{ij}$$

(where δ_{ij} is the Kronecker delta), be $\mathbf{c}_1, \dots, \mathbf{c}_p$. If $\phi_i \neq \phi_j$, then (2.41) is satisfied automatically. Then $\hat{\gamma}'_i = k_i \mathbf{c}_i$. To determine k_i , multiply (2.33) on the left by $\hat{\gamma}'_i$. Then

$$(2.42) \quad N \hat{\gamma}'_i \hat{\Sigma} \hat{\Gamma} (I - D) = k_i \mathbf{c}_i A \hat{\Gamma}.$$

Thus

$$(2.43) \quad N\hat{\gamma}'_i \hat{\Sigma} \hat{\gamma}_j (1 - d_j) = k_i \mathbf{c}_i \mathbf{A} \mathbf{c}_j k_j = N \delta_{ij} k_i^2$$

for i and j being indices of $\hat{\gamma}_j$ in $\hat{\Gamma}$. Since $\hat{\gamma}_i$ is to be normalized by $\hat{\Sigma}$, we have $k_i^2 = 1 - d_i$. Thus

$$(2.44) \quad \hat{\gamma}_i = \sqrt{1 - d_i} \mathbf{c}_i = \frac{1}{\sqrt{1 + \phi_i}} \mathbf{c}_i.$$

Now we wish to show that we should take the vectors corresponding to the m smallest roots. Let $\mathbf{C} = (\mathbf{c}_1, \dots, \mathbf{c}_p)$, $\mathbf{C}^* = (\mathbf{c}_{p-m+1}, \dots, \mathbf{c}_p)$. From (2.29) we obtain

$$(2.45) \quad \begin{aligned} N\hat{\Sigma} &= \mathbf{A} + \frac{1}{N} \mathbf{A} \hat{\Gamma} (\mathbf{I} - \mathbf{D})^{-1} \mathbf{D} (\mathbf{I} - \mathbf{D})^{-1} \hat{\Gamma}' \mathbf{A} \\ &= \mathbf{A} + \frac{1}{N} \mathbf{A} \mathbf{C}^* (\mathbf{I} - \mathbf{D})^{-1} \mathbf{D} \mathbf{C}^{*'} \mathbf{A}, \end{aligned}$$

if $\hat{\Gamma} = \mathbf{C}^* (\sqrt{1 - d_i} \delta_{ij})$ for $i, j = p - m + 1, \dots, p$. Then

$$(2.46) \quad \mathbf{C}' \hat{\Sigma} \mathbf{C} = \mathbf{I} + \begin{pmatrix} 0 & 0 \dots 0 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 0 & 0 \dots 0 \\ 1 & 0 \dots 0 \\ 0 & 1 \dots 0 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 0 & 0 \dots 1 \end{pmatrix} \Phi^* \begin{pmatrix} 0 \dots 0 & 1 & 0 \dots 0 \\ 0 \dots 0 & 0 & 1 \dots 0 \\ \cdot & \cdot \\ \cdot & \cdot \\ 0 \dots 0 & 0 & 0 \dots 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 \dots 0 & 0 & \dots & 0 \\ 0 & 1 \dots 0 & 0 & \dots & 0 \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ 0 & 0 \dots 1 & 0 & \dots & 0 \\ 0 & 0 \dots 0 & 1 + \phi_{p-m+1} & \dots & 0 \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ 0 & 0 \dots 0 & 0 & \dots & 1 + \phi_p \end{pmatrix} = \mathbf{F},$$

say. Thus

$$(2.47) \quad |\hat{\Sigma}| = |\mathbf{C}|^{-2} |\mathbf{F}| = \left| \frac{1}{N} \mathbf{A} \right| \prod_{i=p-m+1}^p (1 + \phi_i).$$

The logarithm of the maximized likelihood function is

$$(2.48) \quad \log \hat{L} = -\frac{1}{2}pN \log(2\pi) - \frac{1}{2}N \log \left| \frac{1}{N} \mathbf{A} \right| - \frac{1}{2}N \log \prod_{i=p-m+1}^p (1 + \phi_i) - \frac{1}{2}pN.$$

Thus $\log \hat{L}$ is maximized by choosing the smallest ϕ_i . This completes the proof of the theorem.

It should be pointed out that $\mathbf{\Gamma}$ can be normalized in other ways. Since Ψ was shown to be zero, it is clear that the maximum likelihood estimate of $\mathbf{\Gamma}$ under other such linear conditions can be obtained from (2.15).

In the process of estimating $\mathbf{\Gamma}$ we obtained an estimate of $\bar{\mathbf{B}}_2$ of rank $p - m$. This can be written as

$$(2.49) \quad \hat{\bar{\mathbf{B}}}_2 = \hat{\Sigma} \hat{\mathbf{\Gamma}}^* \hat{\mathbf{\Gamma}}^{*'} \mathbf{B}_2,$$

where the columns of $\hat{\mathbf{\Gamma}}^*$ are the vectors satisfying (2.10) for the $p - m$ largest roots of (2.9). Fisher [7] obtained the same result for a special problem (see Section 7). The author has given a different proof of this in [1]. Tintner [15] and Geary [8] have considered the problem for Σ known.

The joint asymptotic distribution of the characteristic roots and vectors defined by (2.9), (2.10), and (2.11) has been given [2] under the general conditions that $q_2 \geq p$ and $\frac{1}{N} \sum_{\alpha=1}^N z_\alpha z'_\alpha$ approaches a limit in such a way that, roughly speaking, the multiplicities of the roots of $\left| \frac{1}{N} \bar{\mathbf{B}}_2 \mathbf{Q} \bar{\mathbf{B}}_2' - \lambda \Sigma \right| = 0$ are the same for all N (see [2] for the exact conditions). In particular, if the nonzero roots of $\left| \bar{\mathbf{B}}_2 \lim_{N \rightarrow \infty} \frac{1}{N} \mathbf{Q} \bar{\mathbf{B}}_2' - \lambda \Sigma \right| = 0$ are all different, the limiting distribution of \mathbf{C}^* is given. Since $p \lim_{N \rightarrow \infty} d_i = 0 (i = p - m + 1, \dots, p)$, it follows from the theorems of Rubin quoted in [2] that the limiting distribution of $\hat{\mathbf{\Gamma}}$ is the same as that of \mathbf{C}^* .

3. The likelihood ratio criterion for testing the number of linear restrictions.

If the number of independent restrictions on $\bar{\mathbf{B}}_2$ is m , the rank of $\bar{\mathbf{B}}_2$ is $r = p - m$. Testing the hypothesis that the rank of $\bar{\mathbf{B}}_2$ is r_1 against the alternative that it is not greater than $r_2 (> r_1)$ is equivalent to testing the hypothesis that the number of restrictions on $\bar{\mathbf{B}}_2$ is $m_1 = p - r_1$ against the alternative that it is $m_2 = p - r_2 (< m_1)$. The likelihood ratio criterion is the ratio of the likelihood maximized under the hypothesis of m_1 to that maximized under the hypothesis of m_2 . From (2.48) we see that the criterion is

$$(3.1) \quad \lambda = \prod_{i=r_1+1}^p (1 + \phi_i)^{-\frac{1}{2}N} / \prod_{i=r_2+1}^p (1 + \phi_i)^{-\frac{1}{2}N} = \prod_{i=r_1+1}^{r_2} (1 + \phi_i)^{-\frac{1}{2}N}.$$

THEOREM 2. *Suppose \mathbf{x}_α (of p components) is distributed according to $N(\bar{\mathbf{B}}_1 \mathbf{z}_{1\alpha} + \bar{\mathbf{B}}_2 \mathbf{z}_{2\alpha}, \Sigma) (\alpha = 1, \dots, N)$. The likelihood ratio criterion for testing the hypothesis that the rank of $\bar{\mathbf{B}}_2$ is r_1 against the alternatives that it is $r_2 (> r_1)$ is (3.1), where $\{\phi_i\}$ are the ordered roots (in descending order) of (2.10).*

In particular, the criterion for $r = r_1(m = m_1)$ against all other possible alternatives requires the product over all the $p - r_1(=m_1)$ smallest roots. In this case

$$(3.2) \quad -2 \log \lambda = N \sum_{i=r_1+1}^p \log (1 + \phi_i)$$

is for large samples approximately $N \sum_{i=r_1+1}^p \phi_i$, which has been suggested by Fisher [7] and Hsu [9], [10] for testing this hypothesis.

THEOREM 3. *Let $\frac{1}{N}Q$ approach a nonsingular limit as $N \rightarrow \infty$, and suppose \bar{B}_2 is of rank r_1 . Then $-2 \log \lambda$, where λ is the likelihood ratio criterion defined in Theorem 2 for testing the hypothesis that the rank of \bar{B}_2 is r_1 against the alternatives that it is not greater than p , is asymptotically distributed like χ^2 with $(p - r_1)(q_2 - r_1)$ degrees of freedom.*

PROOF. Let $\theta_i = N\phi_i$. It has been shown by Hsu [10] that $\sum_{i=r_1+1}^p N\phi_i = \sum_{i=r_1+1}^p \theta_i$ has an asymptotic χ^2 -distribution with $(p - r_1)(q_2 - r_1)$ degrees of freedom⁶. Let $\{\theta_{iN}\}$ be a sequence of real numbers such that $\theta_{iN} \rightarrow \theta_{i\infty}$ for each i . Then $N \sum_{i=r_1+1}^p \log (1 + \theta_{iN}/N) \rightarrow \sum_{i=r_1+1}^p \theta_{i\infty}$. The proof is concluded by applying the theorem of Rubin (see [2], Section 2).

It might be observed that Theorem 3 does not follow from the usual theorems concerning the asymptotic distribution of -2 times the logarithm of a likelihood ratio criterion. In fact, if $r_2 < \min(p, q_2)$, then $-2 \log \lambda$ does not have a limiting χ^2 -distribution. However, as $N \rightarrow \infty$, its distribution approaches the limiting distribution of $\sum_{i=r_1+1}^{r_2} \theta_i$. This distribution can be obtained from the limiting distribution of $\{\theta_i\}$ [9], [2].

4. Testing hypotheses about the linear restrictions.

4.1. *Case of one restriction.* Suppose we wish to test the hypothesis that

$$(4.1) \quad g'\bar{B}_2 = 0,$$

where g is a specified (p -dimensional) vector. The (q_2 -dimensional) vector B'_2g is distributed according to $N(\bar{B}'_2g, g'\Sigma g Q^{-1})$, where Q is given by (1.5). When the null hypothesis is true, $g'B_2$ has mean 0 and, therefore, $g'B_2QB'_2g/g'\Sigma g$ has a χ^2 -distribution with q_2 degrees of freedom. From the analysis of variance of $g'x_\alpha$ we see that $g'Ag/g'\Sigma g$ is distributed independently and according to a χ^2 -distribution with $N - q$ degrees of freedom. When the null hypothesis is true,

$$(4.2) \quad \frac{g'B_2QB'_2g}{g'Ag} \cdot \frac{N - q}{q_2} = \frac{1}{q_2} \frac{g'B_2QB'_2g}{g'Sg}$$

⁶ This also follows from [2] and an application of the theorem of Rubin mentioned before.

has the F -distribution with q_2 and $N - q$ degrees of freedom. Therefore, we have the following theorem.

THEOREM 4. *Let \mathbf{x}_α (with p components) be distributed according to $N(\bar{\mathbf{B}}_1 \mathbf{z}_{1\alpha} + \bar{\mathbf{B}}_2 \mathbf{z}_{2\alpha}, \boldsymbol{\Sigma})$, $\alpha = 1, \dots, N$. Define the $(p \times q_2)$ matrix \mathbf{B}_2 by (1.10), \mathbf{Q} by (1.5), and \mathbf{S} by (1.11). Then the critical region of a test of the hypothesis (4.1) at significance level ϵ is*

$$(4.3) \quad \frac{\mathbf{g}' \mathbf{B}_2 \mathbf{Q} \mathbf{B}_2' \mathbf{g}}{q_2 \mathbf{g}' \mathbf{S} \mathbf{g}} \geq F_{q_2, N-q}(\epsilon).$$

It may be noticed that we do not need to put any condition on the rank of $\bar{\mathbf{B}}_2$.

4.2. *Case of several restrictions.* Now consider testing the hypothesis that

$$(4.4) \quad \mathbf{g}'_i \bar{\mathbf{B}}_2 = 0, \quad i = 1, \dots, m,$$

where $\mathbf{g}_1, \dots, \mathbf{g}_m$ are given (p -dimensional) vectors. We shall assume that $(\mathbf{g}_1, \dots, \mathbf{g}_m) = \mathbf{G}$ is of rank m (otherwise some of (4.4) are redundant). The (q_2 -dimensional) vectors $\bar{\mathbf{B}}_2' \mathbf{g}_i$ are normally distributed with means $\bar{\mathbf{B}}_2' \mathbf{g}_i$ and covariances

$$(4.5) \quad \varepsilon(\mathbf{B}'_2 \mathbf{g}_i - \bar{\mathbf{B}}_2' \mathbf{g}_i)(\mathbf{B}'_2 \mathbf{g}_j - \bar{\mathbf{B}}_2' \mathbf{g}_j)' = \mathbf{g}'_i \boldsymbol{\Sigma} \mathbf{g}_j \mathbf{Q}^{-1}.$$

When the null hypothesis is true, the expected value of $\mathbf{B}'_2 \mathbf{g}_i$ is $\mathbf{0}$ and $\mathbf{G}' \mathbf{B}_2 \mathbf{Q} \mathbf{B}'_2 \mathbf{G}$ is distributed according to $W(\mathbf{G}' \boldsymbol{\Sigma} \mathbf{G}, q_2)$, that is, the Wishart distribution with covariance matrix $\mathbf{G}' \boldsymbol{\Sigma} \mathbf{G}$ and q_2 degrees of freedom.⁷ Also $\mathbf{G}' \mathbf{A} \mathbf{G}$ is independently distributed according to $W(\mathbf{G}' \boldsymbol{\Sigma} \mathbf{G}, N - q)$. When the null hypothesis is true,

$$(4.6) \quad \frac{|\mathbf{G}' \mathbf{B}_2 \mathbf{Q} \mathbf{B}'_2 \mathbf{G} + \mathbf{G}' \mathbf{A} \mathbf{G}|}{|\mathbf{G}' \mathbf{A} \mathbf{G}|}$$

has the $U_{m, q_2, N-q}$ distribution. The following theorem results:

THEOREM 5. *Let \mathbf{x}_α (with p components) be distributed according to $N(\bar{\mathbf{B}}_1 \mathbf{z}_{1\alpha} + \bar{\mathbf{B}}_2 \mathbf{z}_{2\alpha}, \boldsymbol{\Sigma})$, $\alpha = 1, \dots, N$. Define the $(p \times q_2)$ matrix \mathbf{B}_2 by (1.10), \mathbf{Q} by (1.5), and \mathbf{S} by (1.11). Then the critical region of a test of the hypothesis $\mathbf{G}' \bar{\mathbf{B}}_2 = \mathbf{0}$, where \mathbf{G} is $m \times p$, at significance level ϵ is*

$$(4.7) \quad \frac{|\mathbf{G}' \mathbf{A} \mathbf{G}|}{|\mathbf{G}' \mathbf{B}_2 \mathbf{Q} \mathbf{B}'_2 \mathbf{G} + \mathbf{G}' \mathbf{A} \mathbf{G}|} \leq U_{m, q_2, N-q}(\epsilon).$$

The above is the likelihood ratio test of hypothesis (4.4); this is based on Wilks' test for the general linear hypothesis. Another test based on the test suggested by Lawley (and later by Hotelling) for the general linear hypothesis is based on the statistic $\text{tr}[\mathbf{G}' \mathbf{B}_2 \mathbf{Q} \mathbf{B}'_2 \mathbf{G} (\mathbf{G}' \mathbf{A} \mathbf{G})^{-1}]$. The test (4.7) has the usual properties of the likelihood ratio test; it is consistent; $-2N$ times the logarithm of (4.6) has approximately the χ^2 -distribution with $m q_2$ degrees of freedom when N is large (under the assumption that $\frac{1}{N} \mathbf{Q}$ tends towards a nonsingular limit).

⁷ $W(\boldsymbol{\Psi}, r)$ is defined as the distribution of $\boldsymbol{\Sigma}'_1 \mathbf{Y}_o \mathbf{Y}'_o$, where \mathbf{Y}_o are independently distributed according to $N(\mathbf{0}, \boldsymbol{\Psi})$.

4.3. *Approximate test of rank.* If the rank of $\bar{\mathbf{B}}_2$ is p , then (4.1) cannot be satisfied for any vector \mathbf{g} ; that is, $\mathbf{B}'_2\mathbf{g}$ must have a mean different from $\mathbf{0}$ for every \mathbf{g} . This suggests that if we are interested in testing the hypothesis that the rank of $\bar{\mathbf{B}}_2$ is $p - 1$ against the hypothesis that it is p , a possible procedure is to reject the hypothesis if (4.3) holds for every \mathbf{g} . This will be true if the minimum of the left hand side of (4.3) with respect to \mathbf{g} is greater than $F_{q_2, N-q}(\epsilon)$. The minimum is the smallest root of

$$(4.8) \quad \left| \frac{1}{q_2} \mathbf{B}_2 \mathbf{Q} \mathbf{B}'_2 - f \mathbf{S} \right| = 0.$$

The smallest root is $f_p = [(N - q)/q_2]\phi_p$. Thus a critical region for this test is

$$(4.9) \quad f_p \geq F_{q_2, N-q}(\epsilon).$$

This test is "conservative"; that is, the probability is less than ϵ of rejecting the null hypothesis when it is true.

We can use the results of Section 4.2 to generalize this technique. Only if $\bar{\mathbf{B}}_2$ is of rank $p - m$ can (4.4) be true for m linearly independent \mathbf{g}_i . A possible test of the hypothesis that $\bar{\mathbf{B}}_2$ is of rank $p - m$ against alternatives that the rank is greater consists of rejecting the hypothesis when (4.7) holds for all \mathbf{G} of rank m , that is, if (4.7) holds when \mathbf{G} is chosen to maximize the left hand side. The maximum is obtained by taking as columns of \mathbf{G} the vectors satisfying

$$(4.10) \quad [\mathbf{A} - \psi(\mathbf{B}_2 \mathbf{Q} \mathbf{B}'_2 + \mathbf{A})]\mathbf{x} = \mathbf{0}$$

corresponding to the m largest roots of

$$(4.11) \quad |\mathbf{A} - \psi(\mathbf{B}_2 \mathbf{Q} \mathbf{B}'_2 + \mathbf{A})| = 0.$$

Let these roots be $\psi_1 \geq \psi_2 \geq \dots \geq \psi_p$. Then

$$(4.12) \quad \frac{|G'AG|}{|G'B_2QB'_2G + G'AG|} = \prod_{i=1}^m \psi_i.$$

It is easily seen that $\psi_i = 1/(1 + \phi_{p-i})$. Therefore, a critical region for testing the hypothesis that $\bar{\mathbf{B}}_2$ has rank $p - m$ is

$$(4.13) \quad \prod_{i=p-m+1}^p (1 + \phi_i)^{-1} \leq U_{m, q_2, N-q}(\epsilon).$$

In other words, $U_{m, q_2, N-q}(\epsilon)$ is an approximation to the significance point for the criterion at significance level ϵ .

5. Confidence regions for the coefficients of the restrictions.

5.1. *Case of one restriction.* If $\bar{\mathbf{B}}_2$ is of rank $p - 1$, there is one vector $\boldsymbol{\gamma}$ satisfying

$$(5.1) \quad \boldsymbol{\gamma}'\bar{\mathbf{B}}_2 = \mathbf{0}$$

and a normalization, say

$$(5.2) \quad \boldsymbol{\gamma}'\Phi\boldsymbol{\gamma} = 1$$

(with the rank of $(\bar{\mathbf{B}}_2 \Phi)$ being greater than that of $\bar{\mathbf{B}}_2$). If Φ is a given (known) matrix, then a confidence region for γ , given the statistics \mathbf{B}_2 and \mathbf{A} (and \mathbf{Q}), consists of the vectors \mathbf{g} (satisfying $\mathbf{g}'\Phi\mathbf{g} = 1$) for which the test in Section 4.1 does not lead to rejection. If $\Phi = \Sigma$ then we make use of the fact that in this case $\gamma' \mathbf{B}_2 \mathbf{Q}' \mathbf{B}_2 \gamma$ and $\gamma' \mathbf{A} \gamma$ have independent χ^2 -distributions.

THEOREM 6. *Let x_α (with p components) be distributed according to $N(\bar{\mathbf{B}}_1 \mathbf{z}_{1\alpha} + \bar{\mathbf{B}}_2 \mathbf{z}_{2\alpha}, \Sigma)$, $\alpha = 1, \dots, N$. Define the $p \times q_2$ matrix \mathbf{B}_2 by (1.10), the non-singular \mathbf{Q} by (1.5), and \mathbf{S} by (1.11). If the normalization of γ is (5.2) for Φ known, then a confidence region for γ defined by (5.1) with confidence coefficient $1 - \epsilon$ consists of the vectors \mathbf{g} satisfying*

$$(5.3) \quad \frac{\mathbf{g}' \mathbf{B}_2 \mathbf{Q} \mathbf{B}_2' \mathbf{g}}{q_2 \mathbf{g}' \mathbf{S} \mathbf{g}} \leq F_{q_2, N-q}(\epsilon)$$

and

$$(5.4) \quad \mathbf{g}' \Phi \mathbf{g} = 1.$$

If the normalization is $\gamma' \Sigma \gamma = 1$, then a confidence region of confidence $(1 - \epsilon_1)(1 - \epsilon_2)$ is the intersection of

$$(5.5) \quad \mathbf{g}' \mathbf{B}_2 \mathbf{Q} \mathbf{B}_2' \mathbf{g} \leq \chi_{q_2}^2(\epsilon_1)$$

and

$$(5.6) \quad \chi_{N-q}^2(\epsilon_2) \leq \mathbf{g}' \mathbf{A} \mathbf{g} \leq \bar{\chi}_{N-q}^2(\epsilon_2),$$

where $\chi_{q_2}^2(\epsilon_1)$ is chosen so that the probability of (5.5) is $1 - \epsilon_1$ when $\mathbf{g} = \gamma$ and $\chi_{N-q}^2(\epsilon_2)$ and $\bar{\chi}_{N-q}^2(\epsilon_2)$ are chosen so that the probability of (5.6) is $1 - \epsilon_2$ when $\mathbf{g} = \gamma$.

These kinds of confidence regions were developed by Rubin and the author in [3] following a suggestion by Girshick. Bartlett [5] has used this method in treating an econometric problem.

5.2. Case of several restrictions. When $\bar{\mathbf{B}}_2$ is of rank $p - m$, there are sets of m linearly independent vectors $\gamma_{p-m+1}, \dots, \gamma_p$ satisfying

$$(5.7) \quad \gamma_i' \bar{\mathbf{B}}_2 = 0.$$

Of course, if we take a set of m linearly independent linear combinations of $\gamma_{p-m+1}, \dots, \gamma_p$ we obtain another set of vectors satisfying (5.7). We can take out some of the indeterminacy in the definition of $\gamma_{p-m+1}, \dots, \gamma_p$ by requiring

$$(5.8) \quad \gamma_i' \Sigma \gamma_j = \delta_{ij}.$$

However, $\Gamma = (\gamma_{p-m+1}, \dots, \gamma_p)$ can still be multiplied on the right by an arbitrary orthogonal matrix. Let us suppose that there are $m(m - 1)/2$ more independent restrictions on Γ ,

$$(5.9) \quad f_v(\Gamma) = 0, \quad v = 1, \dots, m(m - 1)/2,$$

with $f_v(\Gamma)$ being a completely specified function. We assume that (5.7), (5.8), and (5.9) determine Γ uniquely.⁸

If $\mathbf{g}_{p-i} = \boldsymbol{\gamma}_i$, then $\mathbf{g}'_i \mathbf{B}_2 \mathbf{Q} \mathbf{B}'_2 \mathbf{g}_i$ is distributed as χ^2 with q_2 degrees of freedom and independently of $\mathbf{g}'_j \mathbf{B}_2 \mathbf{Q} \mathbf{B}'_2 \mathbf{g}_j$ ($i \neq j$). The matrix $(\mathbf{g}'_i \mathbf{A} \mathbf{g}_j)$ is independently distributed according to $W(I, N - q)$. Then a confidence region may consist of the sets of vectors $\mathbf{g}_1, \dots, \mathbf{g}_m$ satisfying (5.9) with \mathbf{G} for Γ ,

$$(5.10) \quad \mathbf{g}'_i \mathbf{B}_2 \mathbf{Q} \mathbf{B}'_2 \mathbf{g}_i \leq \chi^2_{q_2}(\epsilon_i), \quad i = 1, \dots, m,$$

$$(5.11) \quad \underline{d}_{ij}(\epsilon) \leq \mathbf{g}'_i \mathbf{A} \mathbf{g}_j \leq \bar{d}_{ij}(\epsilon),$$

where $\underline{d}_{ij}(\epsilon) \leq \delta_{ij} \leq \bar{d}_{ij}(\epsilon)$ are chosen so that the probability of (5.11) for all i and j is $1 - \epsilon$ when $\mathbf{G}' \boldsymbol{\Sigma} \mathbf{G} = \mathbf{I}$. The confidence coefficient is $(1 - \epsilon_1)(1 - \epsilon_2) \dots (1 - \epsilon_m)(1 - \epsilon)$. Unfortunately, since the Wishart distribution has not been tabulated, the intervals (5.11) could be obtained from present tables only for $m = 2$, in which case one could use the distribution of the variances and the correlation coefficient.

The confidence region defined by (5.10) and (5.11) has the same confidence coefficient as the region which is the intersection of this and (5.9). If we do not impose the conditions (5.9), there is the indeterminacy of orthogonal transformations in the regions; that is, if \mathbf{G} is in the region, $\mathbf{G}\mathbf{O}$ is in the region if \mathbf{O} is orthogonal (for most \mathbf{G} 's). If one is interested simply in estimating the linear subspace spanned by $\boldsymbol{\gamma}_{p-m+1}, \dots, \boldsymbol{\gamma}_p$, then this region (not imposing (5.9)) is adequate.

Under the restrictions imposed here we could replace (5.10) by

$$(5.12) \quad \sum_{i=1}^m \mathbf{g}'_i \mathbf{B}_2 \mathbf{Q} \mathbf{B}'_2 \mathbf{g}_i \leq \chi_{mq_2}(\epsilon^*),$$

and obtain a region with confidence coefficient $(1 - \epsilon^*)(1 - \epsilon)$. Other regions could also be constructed by replacing (5.12) by other inequalities which take into account that $\mathbf{g}'_i \mathbf{B}_2$ are normally distributed with mean $\mathbf{0}$ when $\mathbf{g}_i = \boldsymbol{\gamma}_i$. For example, (5.12) could be replaced by

$$(5.13) \quad | \mathbf{G}' \mathbf{B}_2 \mathbf{Q} \mathbf{B}'_2 \mathbf{G} | \leq V_{m, N-q}(\epsilon^*),$$

where $[1/(N - q)^m] V_{m, N-q}(\epsilon^*)$ is the ϵ^* significance point of the distribution of the generalized variance of m dimensions and $N - q$ degrees of freedom (for covariance matrix \mathbf{I}).

Another kind of sets of restrictions is

$$(5.14) \quad \boldsymbol{\gamma}'_i \boldsymbol{\Sigma} \boldsymbol{\gamma}_j = 0, \quad i \neq j,$$

and (5.9) for $v = 1, \dots, m(m + 1)/2$. When $\mathbf{G} = \Gamma$, $\mathbf{g}'_i \mathbf{B}_2 \mathbf{Q} \mathbf{B}'_2 \mathbf{g}_i / \mathbf{g}'_i \boldsymbol{\Sigma} \mathbf{g}_i$ has a χ^2 -distribution with q_2 degrees of freedom and independent of $\mathbf{g}'_j \mathbf{B}_2 \mathbf{Q} \mathbf{B}'_2 \mathbf{g}_j$ ($i \neq j$) and $\mathbf{g}'_j \mathbf{A} \mathbf{g}_k$. Thus $[\mathbf{g}'_i \mathbf{B}_2 \mathbf{Q} \mathbf{B}'_2 \mathbf{g}_i / \mathbf{g}'_i \mathbf{A} \mathbf{g}_i] \cdot [(N - q)/q_2]$ has the F -distribution with

⁸ This problem of "identification" has been studied by Koopmans, Rubin and Leipnik in [12].

q_2 and $N - q$ degrees of freedom. Moreover, the set of $m(m - 1)/2$ random variables $\mathbf{g}'_i \mathbf{A} \mathbf{g}_j / \sqrt{\mathbf{g}'_i \mathbf{A} \mathbf{g}_i \mathbf{g}'_j \mathbf{A} \mathbf{g}_j}$ ($i \neq j$) is independently distributed like the set of correlation coefficients r_{ij} ($i \neq j; i, j = 1, \dots, m$) based on $N - q + 1$ observations from $N(\mathbf{0}, \mathbf{I})$ using deviations from the sample mean. Define $r_{ij}(\epsilon) < 0 < \bar{r}_{ij}(\epsilon)$ for $i \neq j$ such that the probability $r_{ij}(\epsilon) < r_{ij} < \bar{r}_{ij}(\epsilon)$ ($i \neq j; i, j = 1, \dots, m$) is $1 - \epsilon$. Then a confidence region with coefficient $(1 - \epsilon_1) \cdots (1 - \epsilon_m)(1 - \epsilon)$ is the intersection of

$$(5.15) \quad \frac{\mathbf{g}'_i \mathbf{B}_2 \mathbf{Q} \mathbf{B}'_2 \mathbf{g}_i}{\mathbf{g}'_i \mathbf{S} \mathbf{g}_i} \leq q_2 F_{q_2, N-q}(\epsilon_i),$$

$$(5.16) \quad r_{ij}(\epsilon) \leq \frac{\mathbf{g}'_i \mathbf{A} \mathbf{g}_j}{\sqrt{\mathbf{g}'_i \mathbf{A} \mathbf{g}_i \mathbf{g}'_j \mathbf{A} \mathbf{g}_j}} \leq \bar{r}_{ij}(\epsilon), \quad i \neq j;$$

and $f_v(\mathbf{G}) = 0$ for $v = 1, \dots, m(m + 1)/2$. If one imposes $f_v(\mathbf{\Gamma}) = 0$ ($v = 1, \dots, m$) only for normalization of the vector $\boldsymbol{\gamma}_i$, there is the indeterminacy of orthogonal transformations. In this case a confidence region may be the intersection of (5.15), (5.16), and $f_v(\mathbf{G}) = 0, v = 1, \dots, m$. If there are no restrictions $f_v(\mathbf{\Gamma}) = 0$, the vectors $\boldsymbol{\gamma}_i$ are not normalized. A confidence region then may consist of the intersection of (5.15) and (5.16).

Now let us suppose that equation (5.14) does not hold. Instead, suppose the matrix $\mathbf{\Gamma}$ is determined uniquely by restrictions (5.9) for $v = 1, \dots, m^2$. Then, in general, $\mathbf{g}'_i \mathbf{B}_2 \mathbf{Q} \mathbf{B}'_2 \mathbf{g}_i$ is not distributed independently of $\mathbf{g}'_i \mathbf{B}_2 \mathbf{Q} \mathbf{B}'_2 \mathbf{g}_j$ ($i \neq j$). We now make use of the theory given in Section 4.2. A confidence region with confidence coefficient $1 - \epsilon$ is given by the intersection of

$$(5.17) \quad \frac{|\mathbf{G}' \mathbf{A} \mathbf{G}|}{|\mathbf{G}' \mathbf{B}_2 \mathbf{Q} \mathbf{B}'_2 \mathbf{G} + \mathbf{G}' \mathbf{A} \mathbf{G}|} \geq U_{m, q_2, N-q}(\epsilon)$$

and $f_v(\mathbf{G}) = 0$ for $v = 1, \dots, m^2$. If the restrictions $f_v(\mathbf{\Gamma}) = 0$ are less than enough for unique identification, (5.17) together with the restrictions imposed on \mathbf{G} constitute a confidence region with coefficient $1 - \epsilon$.

Let us summarize the above results for the cases of unique identification:

THEOREM 7. *Let \mathbf{x}_α (with p components) be distributed according to $N(\bar{\mathbf{B}}_1 \mathbf{z}_{1\alpha} + \bar{\mathbf{B}}_2 \mathbf{z}_{2\alpha}, \boldsymbol{\Sigma})$, $\alpha = 1, \dots, N$. Define the $p \times q_2$ matrix $\bar{\mathbf{B}}_2$ by (1.10), the nonsingular \mathbf{Q} by (1.5), and $\mathbf{S} = [1/(N - q)]\mathbf{A}$ by (1.11). (a) A confidence region for the $m \times p$ matrix $\mathbf{\Gamma}$, the unique solution of $\mathbf{\Gamma}' \bar{\mathbf{B}}_2 = \mathbf{0}$, $\mathbf{\Gamma}' \boldsymbol{\Sigma} \mathbf{\Gamma} = \mathbf{I}$, and $f_v(\mathbf{\Gamma}) = 0$ ($v = 1, \dots, m(m - 1)/2$) with confidence coefficient $(1 - \epsilon_1) \cdots (1 - \epsilon_m)(1 - \epsilon)$ consists of matrices \mathbf{G} satisfying (5.11), (5.10), and $f_v(\mathbf{G}) = 0$; a region of confidence $(1 - \epsilon^*)(1 - \epsilon)$ consists of the set of matrices \mathbf{G} satisfying (5.11), (5.12), and $f_v(\mathbf{G}) = 0$, or the set of matrices \mathbf{G} satisfying (5.11), (5.13), and $f_v(\mathbf{G}) = 0$. (b) A region for $\mathbf{\Gamma}$, the unique solution of $\mathbf{\Gamma}' \bar{\mathbf{B}}_2 = \mathbf{0}$, $\mathbf{\Gamma}' \boldsymbol{\Sigma} \mathbf{\Gamma}$ being diagonal, and $f_v(\mathbf{\Gamma}) = 0$ ($v = 1, \dots, m(m + 1)/2$) of confidence $(1 - \epsilon_1) \cdots (1 - \epsilon_m)(1 - \epsilon)$ consists of the matrices \mathbf{G} satisfying (5.15), (5.16), and $f_v(\mathbf{G}) = 0$. (c) A region of confidence $1 - \epsilon$ for $\mathbf{\Gamma}$, the unique solution of $\mathbf{\Gamma}' \bar{\mathbf{B}}_2 = \mathbf{0}$ and $f_v(\mathbf{\Gamma}) = 0$ ($v = 1, \dots, m^2$) consists of the matrices \mathbf{G} satisfying (5.17) and $f_v(\mathbf{G}) = 0$.*

In this section we have assumed that the restrictions on Γ were just sufficient to take out indeterminacy in the definition. If there are more restrictions (i.e., some restrictions are redundant), they may all be applied to the matrices \mathbf{G} which define the confidence regions.

It might be mentioned in passing that a confidence region for $\bar{\mathbf{B}}_2$ under the restriction that the rank of $\bar{\mathbf{B}}_2$ is r is given by the set of all matrices $(p \times q_2)\Phi^*$ of rank r satisfying

$$(5.18) \quad \frac{|A|}{|(\mathbf{B}_2 - \Phi^*)\mathbf{Q}(\mathbf{B}_2 - \Phi^*)' + A|} \geq U_{p, q_2, N-q}(\epsilon).$$

5.3. *Consistency of confidence regions.* It is clear from the preceding discussion that there are many ways of constructing confidence regions for Γ . One desirable property of a confidence region is that it is consistent. By consistency of a confidence region we mean that if $\frac{1}{N}\mathbf{Q}$ approaches a nonsingular limit (as $N \rightarrow \infty$) the confidence region for Γ is arbitrarily small with arbitrarily high probability for N large enough.

It is easy to verify that if there are restrictions on Γ sufficient for identification the regions given in Sections 5.1 and 5.2 are consistent. Consider, for example, the first region given in Section 5.2. The inequalities (5.10) can be written

$$(5.19) \quad \mathbf{g}'_i \mathbf{B}_2 \frac{1}{N} \mathbf{Q} \mathbf{B}'_2 \mathbf{g}_i \leq \frac{1}{N} \chi^2_{q_2}(\epsilon_i).$$

For N sufficiently large the right hand side of (5.19) is arbitrarily small, $\frac{1}{N}\mathbf{Q}$ is arbitrarily near $\lim_{N \rightarrow \infty} \frac{1}{N} \mathbf{Q}$, and \mathbf{B}_2 is arbitrarily near $\bar{\mathbf{B}}_2$ with probability arbitrarily near one. If $\mathbf{G}\bar{\mathbf{B}}_2 \neq 0$, then N can be chosen large enough so that $\mathbf{G}\mathbf{B}_2$ will have an arbitrarily small probability of satisfying (5.19).

As a matter of fact, consistency of the regions holds even if the assumptions, such as normality of \mathbf{x}_α , are relaxed. Moreover, the confidence coefficient has approximately the value given here if N is sufficiently large although some of the conditions are not fulfilled.⁹

6. Econometric models.

6.1. *Point estimates for certain "shock" models.* In many econometric models the relations between variables may be expressed in terms of a system of stochastic linear equations

$$(6.1) \quad \Phi \mathbf{x}_\alpha^* + \Psi \mathbf{z}_\alpha = \boldsymbol{\varepsilon}_\alpha^*,$$

where \mathbf{x}_α^* is a vector of p^* endogenous (economic) variables and \mathbf{z}_α is a vector of q exogenous (noneconomic) variables, and $\boldsymbol{\varepsilon}_\alpha^*$ is a vector of p^* disturbances. This model has been called a "shock model" [13]. For Φ square and nonsingular we

⁹ See [4] for the treatment of the special case $m = 1$.

can solve (6.1) for¹⁰ \mathbf{x}_α^* ,

$$(6.2) \quad \mathbf{x}_\alpha^* = -\Phi^{-1}\Psi\mathbf{z}_\alpha + \Phi^{-1}\boldsymbol{\varepsilon}_\alpha^*.$$

The distribution of $\boldsymbol{\varepsilon}_\alpha^*$ for fixed \mathbf{z}_α induces the distribution of \mathbf{x}_α^* . Let

$$(6.3) \quad -\Phi^{-1}\Psi = \bar{\mathbf{B}}^*,$$

$$(6.4) \quad \Phi^{-1}\boldsymbol{\varepsilon}_\alpha^* = \mathbf{n}_\alpha^*.$$

Equation (6.2) is the so-called “reduced form.”

Suppose there are m rows of Φ that are known to contain only p components of \mathbf{x}_α^* ; let the vector of these p components be \mathbf{x}_α , and let these m rows and p columns of Φ constitute a submatrix Γ' . Suppose that of these m rows of Ψ there are only q_1 columns different from zero; let the subvector of \mathbf{z}_α with nonzero coefficients be $\mathbf{z}_{1\alpha}$, and let the m rows and q_1 columns of Ψ constitute the matrix Θ : Let the remaining components of \mathbf{z}_α constitute $\mathbf{z}_{2\alpha}$. Thus we have partitioned Φ and Ψ as

$$(6.5) \quad \Phi = \begin{pmatrix} \Gamma' & 0 \\ \Phi_{21} & \Phi_{22} \end{pmatrix}, \quad \Psi = \begin{pmatrix} \Theta & 0 \\ \Psi_{21} & \Psi_{22} \end{pmatrix}.$$

The m equations we are primarily interested in are

$$(6.6) \quad \Gamma'\mathbf{x}_\alpha + \Theta\mathbf{z}_{1\alpha} = \boldsymbol{\varepsilon}_\alpha.$$

The part of (6.2) involving \mathbf{x}_α is

$$(6.7) \quad \mathbf{x}_\alpha = \bar{\mathbf{B}}_1\mathbf{z}_{1\alpha} + \bar{\mathbf{B}}_2\mathbf{z}_{2\alpha} + \mathbf{n}_\alpha.$$

We shall assume that \mathbf{n}_α is distributed according to $N(\mathbf{0}, \Sigma)$. Since the coefficients of $\mathbf{z}_{2\alpha}$ in (6.6) are zero,

$$(6.8) \quad \Gamma'\bar{\mathbf{B}}_2 = \mathbf{0},$$

$$(6.9) \quad \Gamma'\bar{\mathbf{B}}_1 = -\Theta.$$

In order that (6.8) have a unique solution for Γ' except for premultiplication by an arbitrary nonsingular $m \times m$ matrix, we shall assume that $q_2 \geq p - m$ and that the rank of $\bar{\mathbf{B}}_2$ is $p - m$. Then the block of m equations is *identified*. To completely determine Γ we may require that the columns $\boldsymbol{\gamma}_g$ satisfy some normalization conditions, for example,

$$(6.10) \quad \boldsymbol{\gamma}'_g \Sigma \boldsymbol{\gamma}_g = 1,$$

and that there are $m - 1$ coefficients in each row of $(\Gamma' \Theta)$ that are specified to be zero. It has been shown that a given equation is then identified if the rank of the matrix formed from $(\Gamma' \Theta)$ by taking the columns containing the zero coefficients is $m - 1$ and if the rank of the $(p^* - m) \times (p^* - p + q_2)$ matrix $(\Phi_{22} \Psi_{22})$ is $p^* - m$.

¹⁰ This could also be considered as an “error model” with $\Phi^{-1}\boldsymbol{\varepsilon}_\alpha^*$ the error part of \mathbf{x}_α^* and \mathbf{z}_α not subject to error.

Suppose we have observations $(x_\alpha, z_{1\alpha}, z_{2\alpha}), \alpha = 1, \dots, N$. Then we can obtain an estimate $\hat{\Gamma}^*$ of Γ^* satisfying the restrictions of Section 2 by the method described in Section 2. We similarly have

$$(6.11) \quad \hat{\Theta}^* = -\hat{\Gamma}^{*'} \hat{B}_1 = -\hat{\Gamma}^{*'} B_1^*.$$

To obtain $\hat{\Gamma}$ satisfying the restrictions above we must find a matrix D such that

$$(6.12) \quad (\hat{\Gamma}' \hat{\Theta}) = D(\hat{\Gamma}^{*'} \hat{\Theta}^*)$$

satisfies these restrictions. The identifying restrictions make D unique. Suppose a given row of $(\Gamma' \Theta)$ is $(\gamma', \theta, \phi', \theta)$ where there are $m - 1$ coefficients 0. Let $(\hat{\Gamma}^{*'} \hat{\Theta}^*)$ be partitioned similarly into

$$(6.13) \quad (\hat{\Gamma}_1^{*'} \hat{\Gamma}_2^{*'} \hat{\Theta}_1^* \hat{\Theta}_2^*).$$

Then the corresponding row d of D must satisfy

$$(6.14) \quad (\theta \theta) = d(\hat{\Gamma}_2^{*'} \hat{\Theta}_2^*).$$

The matrix on the right is of rank $m - 1$, and the solution of d is unique except for a proportionality constant. That is determined by the sample equivalent of (6.10).

The type of shock model considered in this section seems special. However, the method of estimation may be useful if a block of m equations is identified even though the restrictions on $(\Gamma' \Theta)$ are more than enough to identify each equation within the set of m . One could ignore the surplus restrictions.

In time series analysis the index α denotes the time. In many models components of z_α may be components of $x_{\alpha-1}, x_{\alpha-2}, \dots$ (i.e., lagged values). The estimates given above are nevertheless maximum likelihood estimates.

6.2. *Confidence regions for coefficients in "shock models."* The shock models treated in Section 6.1 are of a special sort in that a block of 0 coefficients is required to be given by a priori conditions. The idea of confidence regions considered in Section 5 can be used, however, in more general circumstances. We shall now discuss this subject in greater generality. For convenience we shall modify the notation of Section 6.1. Let x_α be the (p -component) vector of all the endogenous variables and z_α the (q -component) vector of all exogenous variables, and let us write the set of "structural" equations as

$$(6.15) \quad \Phi x_\alpha + \Psi z_\alpha = \epsilon_\alpha.$$

Let $\Phi^{-1}\Psi = -\bar{B}$ and $\Phi^{-1}\epsilon_\alpha = n_\alpha$. Then the "reduced form" of (6.15) is

$$(6.16) \quad x_\alpha = \bar{B} z_\alpha + n_\alpha.$$

We assume n_α to be distributed according to $N(0, \Sigma)$. We are interested in the first m rows of Φ , which we shall call Γ' , and the first m rows of Ψ which we shall call Θ .

We shall suppose that the restrictions for effecting identification are of one of

the three alternative kinds: (1) $\Gamma'\Sigma\Gamma = I$ and $\gamma_{ij} = 0, \theta_{kl} = 0$ for certain pairs (i, j) and (k, l) ; (2) $\Gamma'\Sigma\Gamma$ diagonal, $\gamma_{ij} = 0$ and $\theta_{kl} = 0$ for certain pairs (i, j) and (k, l) , and $f_v(\Gamma) = 0$ for $v = 1, \dots, m$ (for normalization); and (3) $\gamma_{ij} = 0, \theta_{kl} = 0$ for certain pairs (i, j) and (k, l) , and $f_v(\Gamma) = 0$ for $v = 1, \dots, m$ (for normalization). These kinds of restrictions have been studied extensively [12].

In any case the $(m \times q)$ matrix $\Gamma'B$ (where B is defined by (1.10)) has a normal distribution with expected value $\Gamma'\bar{B} = -\Theta$; the covariance between the i, j th element of $\Gamma'B$ and the k, l th element is $\gamma'_i \Sigma \gamma_k m^{jl}$, where $(m^{jl}) = M^{-1}$ and

$$(6.17) \quad M = \sum_{\alpha=1}^N z_{\alpha} z'_{\alpha}.$$

Thus $(\Gamma'B - \Theta)M(B'\Gamma - \Theta')$ is distributed according to $W(\Gamma'\Sigma\Gamma, q)$. Furthermore $\Gamma'A\Gamma$ is independently distributed according to $W(\Gamma'\Sigma\Gamma, N - q)$.

Let θ_i be the i th row of Θ , and let θ_i^* consist of the q_i^* components of θ_i which are not specified to be 0. Let \bar{B}_i^* be composed of the columns of \bar{B} corresponding to the components of θ_i^* so that $\gamma'_i \bar{B}_i^* = -\theta_i^*$. Let \bar{B}_i^{**} be composed of the other q_i^{**} columns of \bar{B} ; then $\gamma'_i \bar{B}_i^{**} = 0$. If B_i^* and B_i^{**} are formed similarly from B , then $\gamma'_1 B_1^{**}, \dots, \gamma'_m B_m^{**}$ are jointly normally distributed with means 0 , and the covariances involve only Γ, Σ , and M .

CASE 1. Since $\Gamma'\Sigma\Gamma = I$, the distribution $W(I, q)$ of $(\Gamma'B - \Theta)M(B'\Gamma - \Theta')$ and the distribution $W(I, N - q)$ of $\Gamma'A\Gamma$ have all parameters known. A confidence region for Γ and Θ of confidence $(1 - \epsilon)(1 - \epsilon_1) \dots (1 - \epsilon_m)$ consists of all G and T satisfying

$$(6.18) \quad (g'_i B - t_i)M(B'g_i - t'_i) \leq \chi^2_{q_i}(\epsilon_i)$$

and (5.11) and the identification conditions. It is understood that in g_i and t_i above we set those coefficients equal to 0 that are so specified in γ_i and θ_i , respectively, by the a priori identification restrictions. A confidence region of confidence $(1 - \epsilon)(1 - \epsilon^*)$ is the intersection of

$$(6.19) \quad \text{tr}(G'B - T)M(B'G - T') \leq \chi^2_{mq}(\epsilon^*)$$

and (5.11) with the identification conditions imposed on G and T . Thirdly, a confidence region of confidence $(1 - \epsilon)(1 - \epsilon^*)$ is the intersection of

$$(6.20) \quad |(G'B - T)M(B'G - T')| \leq V_{m, N-q}(\epsilon^*)$$

and (5.11) with the identification conditions imposed on G and T .

We can also construct confidence regions for Γ alone. A region of confidence $(1 - \epsilon)(1 - \epsilon_1) \dots (1 - \epsilon_m)$ consists of g_i satisfying (5.11), the identification conditions on γ_i , and

$$(6.21) \quad g'_i B_i^{**} Q_i B_i^{**'} g_i \leq \chi^2_{q_i^{**}}(\epsilon_i),$$

where Q_i^{-1} is composed of the rows and columns of M^{-1} according to the way

B_i^{**} is composed of the columns of B . A region of confidence $(1 - \epsilon)(1 - \epsilon^*)$ consists of g_i satisfying (5.11), the identification conditions, and

$$(6.22) \quad \sum_{i=1}^m g_i' B_i^{**} Q_i B_i^{**'} g_i \leq \chi_{\Sigma q_i}^2(\epsilon^*).$$

CASE 2. Since $\Gamma' \Sigma \Gamma$ is diagonal, $(\gamma_i' B - \theta_i) M(B' \gamma_i - \theta_i) / \gamma_i' \Sigma \gamma_i$ is distributed according to the χ^2 -distribution with q degrees of freedom independently of $(\gamma_j' B - \theta_j) M(B' \gamma_j - \theta_j)$ and of $\gamma_i' A \gamma_i / \gamma_i' \Sigma \gamma_i$, the latter being distributed according to the χ^2 -distribution with $N - q$ degrees of freedom. A confidence region of confidence coefficient $(1 - \epsilon)(1 - \epsilon_1) \cdots (1 - \epsilon_m)$ consists of G and T satisfying the identification conditions (including $f_v(G) = 0$), (5.16), and

$$(6.23) \quad \frac{(g_i' B - t_i) M(B' g_i - t_i)}{g_i' A g_i} \cdot \frac{N - q}{q} \leq F_{q, N-q}(\epsilon_i).$$

If identification is effected by the restrictions $\gamma_{ij} = 0$, $\theta_{ij} = 0$, and $f_v(\Gamma) = 0$, then (5.16) is unnecessary.

A confidence region for Γ alone of confidence $(1 - \epsilon)(1 - \epsilon_1) \cdots (1 - \epsilon_m)$ consists of G satisfying the identification conditions, (5.16), and

$$(6.24) \quad \frac{g_i' B_i^{**} Q_i B_i^{**'} g_i}{g_i' A g_i} \cdot \frac{N - q}{q_i} \leq F_{q_i, N-q}(\epsilon_i).$$

CASE 3. In this case a region of confidence $1 - \epsilon$ consists of G and T satisfying the identification conditions and

$$(6.25) \quad \frac{|G' A G|}{|(G' B - T) M(B' G - T') + G' A G|} \geq U_{m, q, N-q}(\epsilon).$$

A region for Γ alone could be given, but since it is more complicated than the above we shall not write it here.

It is clear that there are many ways of obtaining confidence regions. For other combinations of identification conditions we could give similar kinds of confidence regions. A property that all the regions given in this section have is consistency (except the region involving (6.20)). In fact, if some of the assumptions, such as that of normality of n_α , are relaxed, the regions are nevertheless consistent (see [4]). Furthermore, as shown in [4], the confidence coefficients of the regions given above when $m = 1$ approach $1 - \epsilon$ as $N \rightarrow \infty$ under certain conditions even though the variables are not normally distributed and even though some of the components of z_α are "lagged" values of components of x_α . Similarly, it can be shown for m greater than 1 that if the regions are used when the assumptions are relaxed in certain ways one can have confidence about $1 - \epsilon$ if N is large enough. Sufficient conditions are given in Theorem 6 of [4].

It might be remarked in passing that if the number of 0 coefficients in $(\gamma_i' \theta_i)$ are more than enough for identification some columns can be dropped from B_i^{**} ; if $p - 1$ columns remain one can determine confidence regions for γ_i' or

($\gamma'_i \theta_i$). It will also be noted that regions can easily be constructed when the identification equations are of other kinds. In any case they are simply imposed directly on \mathbf{G} and \mathbf{T} (as long as the restrictions do not involve Σ).

6.3. An "error" model. In an error model we consider each observed variable as composed of a "systematic part" and an "error." If x_α is the vector of observed values, ξ_α the vector of systematic parts, and v_α the vector of errors, then $x_\alpha = \xi_\alpha + v_\alpha$. The m linear relations are taken to hold on the systematic parts, that is,

$$(6.26) \quad \Gamma' \xi_\alpha = \mathbf{0}.$$

We shall assume that v_α is distributed according to $N(\mathbf{0}, \Sigma)$, and that ξ_α is "fixed."

If Σ is known, Tintner [15] has suggested estimating the columns of Γ as the vectors (or linear combinations of the vectors) satisfying

$$(6.27) \quad \left(\sum_{\alpha=1}^N x_\alpha x'_\alpha - \lambda \Sigma \right) c = \mathbf{0}$$

corresponding to the m smallest roots of

$$(6.28) \quad \left| \sum_{\alpha=1}^N x_\alpha x'_\alpha - \lambda \Sigma \right| = 0.$$

The obvious shortcoming of this procedure is that usually Σ is unknown.

As a modification of this procedure Bartlett [5] in a special case and Geary [8] when Σ is known have suggested that ξ_α be represented as $\bar{\mathbf{B}}z_\alpha$, where the components of z_α are given functions of time (preferably orthogonal functions). It is clear that the methods proposed here can be used in these circumstances.

7. Another example; a q -sample problem. Consider q multivariate normal distributions $N(\mathbf{y}_k, \Sigma)$ ($k = 1, \dots, q$) with common covariance matrix. The means may be represented as q points in a p -dimensional space. We may ask whether these points lie in an r -dimensional linear subspace, or we may ask what is this r -dimensional subspace. Fisher [7] considered a related problem; a theory about gene structure of three varieties of iris led to a hypothesis that the means of three populations were on a line. (Since the relative distances on the line were also specified, Fisher's hypothesis could be reduced to a hypothesis specifying rank zero.)

Suppose we have a sample $\{x_\alpha^{(k)}\}$ ($\alpha = 1, \dots, N_k$) from each population $N(\mathbf{y}_k, \Sigma)$ ($k = 1, \dots, q$). Let

$$(7.1) \quad \mathbf{u} = \frac{1}{N} (N_1 \mathbf{y}_1 + \dots + N_q \mathbf{y}_q),$$

where $N = N_1 + \dots + N_q$. The hypothesis that the \mathbf{y}_k lie in an r -dimensional space is equivalent to testing the hypothesis that

$$(7.2) \quad (\mathbf{y}_1 - \mathbf{u}, \mathbf{y}_2 - \mathbf{u}, \dots, \mathbf{y}_q - \mathbf{u})$$

is of rank r . It is well known (Hsu [11], for example) that this model can be put into the form of (1.8). Thus we can apply Theorem 2 for deriving a test function.

Let

$$(7.3) \quad \bar{x}_k = \sum_{\alpha=1}^{N_k} x_{\alpha}^{(k)} / N_k,$$

$$(7.4) \quad \bar{x} = \sum_{k=1}^q \sum_{\alpha=1}^{N_k} x_{\alpha}^{(k)} / N,$$

$$(7.5) \quad A = \sum_{k=1}^q \sum_{\alpha=1}^{N_k} (x_{\alpha}^{(k)} - \bar{x}_k)(x_{\alpha}^{(k)} - \bar{x}_k)'$$

Then $Z_k = \sqrt{N_k} \bar{x}_k$ is distributed according to $N(\sqrt{N_k} \boldsymbol{\mu}_k, \boldsymbol{\Sigma})$. Let F be an orthogonal matrix with first row $(\sqrt{N_1}/N, \dots, \sqrt{N_q}/N)$. Let

$$(7.6) \quad Y_l = \sum_{k=1}^q f_{lk} Z_k.$$

Then

$$(7.7) \quad \boldsymbol{\Sigma} Y_l = \sum f_{lk} \sqrt{N_k} \boldsymbol{\mu}_k = \mathbf{v}_l,$$

say. The Y_l are independently distributed according to $N(\mathbf{v}_l, \boldsymbol{\Sigma})$. We have $\mathbf{v}_1 = \sqrt{N} \boldsymbol{\mu}$ and

$$(7.8) \quad \mathbf{v}_l = \sum f_{lk} \sqrt{N_k} (\boldsymbol{\mu}_k - \boldsymbol{\mu}), \quad l = 2, \dots, q,$$

and the rank of $(\mathbf{v}_2, \dots, \mathbf{v}_q)$ is that of (7.2).

For the purposes of testing rank, a model equivalent to the one above is a model with N random variables, Y_1, \dots, Y_q , and $N - q$ others independently distributed according to $N(\mathbf{0}, \boldsymbol{\Sigma})$. Let the l th coordinate of \mathbf{z}_k be δ_{kl} , and let $\bar{\mathbf{B}} = (\mathbf{v}_1, \dots, \mathbf{v}_q)$. Then this model is that of Sections 2 and 3. $\mathbf{B} = (Y_1, \dots, Y_q)$ and $\mathbf{B}_2 \mathbf{Q} \mathbf{B}'_2$ is

$$(7.9) \quad \begin{aligned} \sum_{k=2}^{q-1} Y_k Y'_k &= \sum_{k=1}^q Y_k Y'_k - Y_1 Y'_1 = \sum_{k=1}^q N_k \bar{x}_k \bar{x}'_k - N \bar{x} \bar{x}' \\ &= \sum_{k=1}^q N_k (\bar{x}_k - \bar{x})(\bar{x}_k - \bar{x})'. \end{aligned}$$

Then \mathbf{A} is as defined in Section 1. The \mathbf{c}_i ($i = 1, \dots, p$) are the characteristic vectors of

$$(7.10) \quad \left(\sum_{k=1}^q N_k (\bar{x}_k - \bar{x})(\bar{x}_k - \bar{x})' - \phi_i \mathbf{A} \right) \mathbf{c} = \mathbf{0}$$

satisfying

$$(7.11) \quad \mathbf{c}'_i \mathbf{A} \mathbf{c}_j = N \delta_{ij},$$

where ϕ_i is the i th ordered root of

$$(7.12) \quad \left| \sum_{k=1}^q N_k (\bar{x}_k - \bar{x})(\bar{x}_k - \bar{x})' - \phi \mathbf{A} \right| = 0.$$

The estimate of $\mathbf{y} = \frac{1}{\sqrt{N}} \mathbf{v}_1$ is \bar{x} . Let

$$(7.13) \quad R = \hat{\Sigma} \left(\sum_{i=1}^r c_i c'_i \frac{1}{1 + \phi_i} \right),$$

where $\hat{\Sigma}$ is given by (2.14). Then

$$(7.14) \quad \hat{\mathbf{v}}_l = R Y_l, \quad l = 2, \dots, q.$$

Since F is orthogonal, we have

$$(7.15) \quad \begin{aligned} \sqrt{N_k} \hat{\mathbf{u}}_k &= \sum_{l=1}^q \hat{\mathbf{v}}_l f_{lk} = R \sum_{l=2}^q Y_l f_{lk} + Y_1 f_{1k} \\ &= R \sum_{l=1}^q Y_l f_{lk} + (I - R) Y_l f_{lk} = R \sqrt{N_k} \bar{x}_k + (I - R) \sqrt{N_k} \bar{x}. \end{aligned}$$

Thus

$$(7.16) \quad \hat{\mathbf{u}}_k = R(\bar{x}_k - \bar{x}) + \bar{x}.$$

The likelihood ratio criterion for testing the hypothesis that the rank of $(\mathbf{v}_2, \dots, \mathbf{v}_q)$, that is, the rank of (7.2), is r is given by

$$(7.17) \quad \prod_{i=r+1}^p (1 + \phi_i)^{-\frac{1}{2}N},$$

where $\phi_{r+1}, \dots, \phi_p$ are the $p - r$ smallest roots of (7.10).

An interesting example of the q -sample problem has been discussed by Cochran [6]. Let $x_{i\alpha}^{(k)}$ be a measurement on the α th replicate under the k th treatment, measured on scale i . Among other questions concerning comparisons of the scales, there is the problem of whether the scales are linearly related. This is exactly the problem considered above.

REFERENCES

- [1] T. W. ANDERSON, "The non-central Wishart distribution and its application to problems in multivariate statistics," unpublished thesis, Princeton University Library, Princeton, N. J., 1945.
- [2] T. W. ANDERSON, "The asymptotic distribution of certain characteristic vectors," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951.
- [3] T. W. ANDERSON AND HERMAN RUBIN, "Estimation of the parameters of a single equation in a complete system of stochastic equations," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 46-63.
- [4] T. W. ANDERSON AND HERMAN RUBIN, "The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equations," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 570-582.
- [5] M. S. BARTLETT, "A note on the statistical estimation of demand and supply relations from time series," *Econometrica*, Vol. 16 (1948), pp. 323-329.
- [6] W. G. COCHRAN, "The comparison of different scales of measurement for experimental results," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 205-216.

- [7] R. A. FISHER, "The statistical utilization of multiple measurements," *Annals of Eugenics*, Vol. 8 (1938), pp. 376-386.
- [8] R. C. GEARY, "Studies in relations between economic time series," *Jour. Roy. Stat. Soc., Ser. B*, Vol. 10 (1948), pp. 140-158.
- [9] P. L. HSU, "On the limiting distribution of roots of a determinantal equation," *Jour. London Math. Soc.*, Vol. 16 (1941), pp. 183-194.
- [10] P. L. HSU, "On the problem of rank and the limiting distribution of Fisher's test function," *Annals of Eugenics*, Vol. 11 (1941), pp. 39-41.
- [11] P. L. HSU, "Canonical reduction of the general regression problem," *Annals of Eugenics*, Vol. 11 (1941), pp. 42-46.
- [12] T. C. KOOPMANS, HERMAN RUBIN, AND R. B. LEIPNIK, "Measuring the equation systems of dynamic economics," *Statistical Inference in Dynamic Economic Models, Cowles Commission Monograph No. 10*, John Wiley & Sons, 1950, pp. 53-237.
- [13] J. MARSCHAK, "Statistical inference in economics: an introduction," *Statistical Inference in Dynamic Economic Models, Cowles Commission Monograph No. 10*, John Wiley & Sons, 1950, pp. 1-50.
- [14] C. R. RAO, "Tests of significance in multivariate analysis," *Biometrika*, Vol. 35 (1948), pp. 58-79.
- [15] GERHARD TINTNER, "A note on rank, multicollinearity and multiple regression," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 304-308.
- [16] S. S. WILKS, "On the independence of k sets of normally distributed statistical variables," *Econometrica*, Vol. 3 (1935), pp. 309-326.