The Annals of Mathematical Statistics
1971, Vol. 42, No. 4, 1412–1424

ON OBTAINING LARGE-SAMPLE TESTS FROM ASYMPTOTICALLY NORMAL ESTIMATORS[1]

by T. W. F. STROUD
Queen's University at Kingston

This is an extension of Wald's asymptotic test procedure based on unrestricted maximum-likelihood estimators. Wald showed that under certain regularity conditions the test statistic has a limiting central chi-square distribution under the hypothesis and a limiting noncentral chi-square distribution under a sequence of local alternatives. We extend this procedure, allowing it to be based on a broader class of estimators and to obey simpler and less restrictive conditions. Sufficient conditions for validity of the limiting distributions are local twice-differentiability of the left side of the hypothesis and, under a sequence of local alternatives, asymptotic normality of the estimator of the parameter defining the distribution and stochastic convergence (to the appropriate asymptotic value) of the estimator of the covariance matrix. The required asymptotic behavior is verified for the case of independent sampling from two normal distributions and formulas are presented which aid in computing the test statistic.

1. Introduction and notation. The topic of this paper is an extension of an asymptotic test procedure based on unrestricted maximum-likelihood estimators, which was formulated by Wald (1943) under some rather restrictive regularity conditions. The extension is to a wider class of estimators and to problems obeying weaker conditions.

The well-known properties of consistency and asymptotic normality of maximum-likelihood estimators have been proved, under certain regularity assumptions, by Doob (1934). These properties suggest that, for the problem of testing (under a parametric family of distributions) the hypothesis that a vector parameter is equal to zero, an asymptotic chi-square test may be based on the quadratic form obtained from the maximum-likelihood estimator of its asymptotic covariance matrix, suitably normalized by the sample size.

Wald (1943) showed that, under certain conditions, this quadratic form has in fact an asymptotic central chi-square distribution under the null hypothesis, and an asymptotic noncentral chi-square distribution under a sequence of local alternatives converging to the null hypothesis at the rate $n^{-\frac{1}{2}}$. Wald also claimed an extension of this to a certain global result involving uniform convergence, but this extension has been shown to be incorrect (Stroud, 1970).

Wald also showed under the same conditions that, for testing the hypothesis that the vector parameter is zero, the procedure that rejects when this quadratic

Received December 3, 1970.

[1] This work is a revision of part of the author's doctoral dissertation submitted to Stanford University, and was supported in part by the National Sciences Foundation grant GP-6681, and the National Research Council of Canada grant A7220.

form is greater than or equal to the constant necessary to produce a level-α test is asymptotically most stringent, and that this procedure has further asymptotically optimal properties concerning its power. In the same paper he showed that these asymptotically optimal properties are shared by the likelihood-ratio test.

One of Wald's conditions, which is particularly restrictive, is that the parameter involved be obtainable from the parameters defining the distributions through a transformation possessing uniformly continuous and bounded first and second order partial derivatives. Wald's results are based also on a number of assumptions which are difficult to verify in practice, such as the uniform consistency of the maximum-likelihood estimators.

In a paper by LeCam (1956), a number of theorems are proved which generalize the asymptotic properties of maximum-likelihood estimators to a wider class of estimators. It is stated there that, on the basis of these theorems, Wald's results can be generalized to a class of problems for which Wald's restrictions are not satisfied. However, the application of LeCam's results to the general problem of testing hypotheses has not been worked out in detail, so that these results have not been used here.

In Section 2 of the current work, conditions are given under which an asymptotically normal estimator yields an asymptotic χ^2-test, where the noncentral behavior is governed by a sequence of local alternatives of the order $n^{-\frac{1}{2}}$. The regularity conditions are simpler than Wald's and are local conditions; in particular, global uniform continuity and boundedness are not required. The use of maximum-likelihood estimators as estimators of the parameters involved is not essential. It is sufficient that the parameters defining the distributions be asymptotically normal under the sequence of local alternatives; the estimator of the covariance matrix need only converge stochastically under this sequence to its asymptotic value. (In applications, of course, one would want to use asymptotically efficient estimators whenever possible.) The transformation yielding the vector parameter is assumed to possess continuous and bounded second partial derivatives locally within a neighborhood of any parameter point.

The property of asymptotic stringency and the other asymptotically optimal properties treated by Wald are not investigated here.

In Sections 3 and 4 the results are applied to a class of hypothesis testing problems where the observations are assumed to be the mean vectors and covariance matrices of samples from two multivariate normal populations. Section 3 contains a verification that the conditions required in Section 2 are met, and in Section 4 a method of computing the test statistic is indicated. The class of problems covered is quite broad, encompassing any hypothesis which can be formulated by a twice-differentiable vector equation which is to be tested against general alternatives. The method is particularly valuable in cases where the likelihood-ratio test is difficult to compute, such as the Behrens–Fisher problem (Kendall and Stuart (1961)) (where the test generated by this method is equivalent to that of Welch (1938)), or the extension of this problem in which one compares two regressions under errors of measurement (Stroud (1968)).

Parameters are denoted by Greek letters, statistics by Latin letters or by Greek letters with the caret ($\wedge$) denoting an estimator. Bold-face capitals denote matrices, bold-face lower-case letters denote vectors and light-face letters denote scalars. The letter n is used to index a sequence of parameter values, whereas the letters M and N refer to sample sizes. Light-face lower-case letters with two extra subscripts are used to denote components of matrices labelled by the corresponding bold-face capitals.

The law of the random vector $\mathbf{x}$ is denoted throughout by $\mathscr{L}(\mathbf{x})$. In particular, $\mathscr{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ refers to a normal law with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. By $\mathscr{L}(\mathbf{x}_n) \to \mathscr{L}(\mathbf{y})$ or $\mathscr{L}(\mathbf{x}_n) \to \mathscr{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is meant, respectively, that the law of $\mathbf{x}_n$ converges to the law of $\mathbf{y}$ or to the stated normal law, as $n \to \infty$. The notation $\chi_r^{\,2}(\delta^2)$ refers to the noncentral chi-square distribution and is understood to mean the distribution of the squared norm of a normal random r-dimensional vector with covariance matrix equal to the identity and with mean vector having a norm of δ.

The definitions of the Mann–Wald symbols O_p and o_p may be found in Chernoff (1956, Section 2), as may the statements of some basic results of large-sample theory which are used freely in the theory of the following two sections.

2. The proposed test statistic and conditions under which it has a limiting chi-square distribution. Let $\mathbf{t}$ be a given estimator of a vector parameter $\boldsymbol{\theta}$ indexing a family of distributions. For the problem of testing that a certain continuously twice-differentiable vector-valued function γ, defined on the parameter space, is equal to zero when evaluated at $\boldsymbol{\theta}$, the following test procedure is proposed. The procedure is to reject for large values of the quadratic form

$$J = [\gamma(\mathbf{t})]'\mathbf{D}^{-1}\gamma(\mathbf{t}),$$

where $\mathbf{D}$ is an estimator of the covariance matrix $\boldsymbol{\Delta}$ of the column vector $\gamma(\mathbf{t}) - \gamma(\boldsymbol{\theta})$. We consider properties of $\mathbf{t}$ and $\mathbf{D}$, based on asymptotic theory, which yield an approximate central (noncentral) chi-square distribution for J when the hypothesis is true (false).

Consider a sequence $\boldsymbol{\theta}_n$ of parameter values such that $\boldsymbol{\delta}_n \equiv n^{\frac{1}{2}}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0)$ converges to some finite value $\boldsymbol{\delta}$, where $\gamma(\boldsymbol{\theta}_0) = \mathbf{0}$. For each n let $\mathbf{t}_n$ be some estimator of $\boldsymbol{\theta}_n$, and let $\mathbf{D}_n$ and J_n be defined relative to $\mathbf{t}_n$ and $\boldsymbol{\theta}_n$ in the manner of $\mathbf{D}$ and J above. It is shown in the following theorem that if for some positive definite $\boldsymbol{\Sigma}_0$ the limiting distribution of $n^{\frac{1}{2}}(\mathbf{t}_n - \boldsymbol{\theta}_n)$ is the normal law $\mathscr{N}(\mathbf{0}, \boldsymbol{\Sigma}_0)$, and if $\mathbf{D}_n = n^{-1} \times \mathbf{G}_n\mathbf{S}_n\mathbf{G}_n'$ where $\mathbf{S}_n$ converges in probability to $\boldsymbol{\Sigma}_0$ and $\mathbf{G}_n$ is the matrix of partial derivatives of γ evaluated at $\mathbf{t}_n$, then J_n has a limiting noncentral chi-square distribution (central if $\boldsymbol{\delta} = \mathbf{0}$). The form $\mathbf{D}_n = n^{-1}\mathbf{G}_n\mathbf{S}_n\mathbf{G}_n'$ is suggested by the fact that if, for fixed $\boldsymbol{\theta}$, the distribution of $n^{\frac{1}{2}}(\mathbf{t}_n - \boldsymbol{\theta})$ converges to $\mathscr{N}(\mathbf{0}, \boldsymbol{\Sigma})$, then the distribution of $n^{\frac{1}{2}}[\gamma(\mathbf{t}_n) - \gamma(\boldsymbol{\theta})]$ converges to $\mathscr{N}(\mathbf{0}, \boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}')$, where $\boldsymbol{\Gamma}$ is the matrix of partial derivatives of γ evaluated at $\boldsymbol{\theta}$.

If $\mathbf{t}_n$ is a maximum-likelihood estimator, and if $\gamma(\boldsymbol{\theta})$ can be thought of as a subvector of a vector $\tilde{\gamma}(\boldsymbol{\theta})$, where the function $\tilde{\gamma}$ is one-to-one, then, because maximum-likelihood estimators are invariant under one-to-one transformations,

$\gamma(\mathbf{t}_n)$ may be regarded as a maximum-likelihood estimator of $\gamma(\boldsymbol{\theta}_n)$. If, in addition, $\mathbf{D}_n$ is a maximum-likelihood estimator, then the testing procedure proposed here corresponds to the procedure described by Wald.

THEOREM. *Let* $\{\boldsymbol{\theta}_n\}$ *be a sequence of points in p-dimensional Euclidean space* $\mathscr{E}^p$ *of the form* $\boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + n^{-\frac{1}{2}}\boldsymbol{\delta}_n$, *where* $\lim \boldsymbol{\delta}_n = \boldsymbol{\delta}$ *and* $\boldsymbol{\theta}_0$, $\boldsymbol{\delta}$ *are fixed points. Let* $\{\mathbf{t}_n\}$ *be a sequence of p-dimensional random vectors such that* $\mathscr{L}[n^{\frac{1}{2}}(\mathbf{t}_n - \boldsymbol{\theta}_n)] \to \mathscr{N}(\mathbf{0}, \boldsymbol{\Sigma}_0)$, *where* $\boldsymbol{\Sigma}_0$ *is nonsingular; and let* $\{\mathbf{S}_n\}$ *be a sequence of* $p \times p$ *symmetric random matrices, nonsingular with probability one, such that* $\operatorname{plim} \mathbf{S}_n = \boldsymbol{\Sigma}_0$.

Suppose $\gamma : \mathscr{E}^p \to \mathscr{E}^r (r \leqq p)$ *is a function satisfying* $\gamma(\boldsymbol{\theta}_0) = \mathbf{0}$, *with bounded and continuous second partial derivatives in a sphere of radius* ρ *about* $\boldsymbol{\theta}_0$, *and such that the matrix*

$$\boldsymbol{\Gamma}_0 \equiv (\partial\gamma_i/\partial\theta_j)_{1 \leqq i \leqq r, 1 \leqq j \leqq p} \qquad \textit{(evaluated at } \boldsymbol{\theta}_0)$$

has rank r. Define

$$\mathbf{z}_n = \gamma(\mathbf{t}_n), \qquad J_n = n\mathbf{z}_n'(\mathbf{G}_n\mathbf{S}_n\mathbf{G}_n')^{-1}\mathbf{z}_n,$$

where

$$\mathbf{G}_n \equiv (\partial\gamma_i/\partial\theta_j)_{1 \leqq i \leqq r, 1 \leqq j \leqq p} \qquad \textit{(evaluated at } \mathbf{t}_n).$$

Then as $n \to \infty$ *the distribution of* J_n *converges to the noncentral chi-square distribution* $\chi_r^2(\boldsymbol{\delta}'\boldsymbol{\Gamma}_0'(\boldsymbol{\Gamma}_0\boldsymbol{\Sigma}_0\boldsymbol{\Gamma}_0')^{-1}\boldsymbol{\Gamma}_0\boldsymbol{\delta})$. *If* $\boldsymbol{\delta} = \mathbf{0}$, *the limit distribution is central chi-square.*

PROOF. The conclusion of the theorem will follow if the relation

$$\mathscr{L}[n^{\frac{1}{2}}(\mathbf{G}_n\mathbf{S}_n\mathbf{G}_n')^{-\frac{1}{2}}\mathbf{z}_n] \to \mathscr{N}((\boldsymbol{\Gamma}_0\boldsymbol{\Sigma}_0\boldsymbol{\Gamma}_0')^{-\frac{1}{2}}\boldsymbol{\Gamma}_0\boldsymbol{\delta}, \mathbf{I}) \tag{2.1}$$

can be established.

As a preliminary, the limit distribution of $n^{\frac{1}{2}}\mathbf{z}_n$ is obtained. By Taylor's theorem for p variables [e.g. Apostol (1957) page 124], the following representation holds for each $i = 1, \cdots, r$ and any $\boldsymbol{\theta}$ satisfying $|\boldsymbol{\theta} - \boldsymbol{\theta}_0| < \rho$:

$$\gamma_i(\boldsymbol{\theta}) - \gamma_i(\boldsymbol{\theta}_0) = \sum_{\alpha=1}^{p} \frac{\partial\gamma_i}{\partial\theta_{0\alpha}}(\theta_\alpha - \theta_{0\alpha}) + \frac{1}{2!}\sum_{\alpha=1}^{p}\sum_{\beta=1}^{p} \frac{\partial^2\gamma_i}{\partial\theta_\alpha{}^*\partial\theta_\beta{}^*}(\theta_\alpha - \theta_{0\alpha})(\theta_\beta - \theta_{0\beta}) \tag{2.2}$$

where $\boldsymbol{\theta}^*$ is some point on the open line segment joining $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$. Consider

$$n^{\frac{1}{2}}(\mathbf{t}_n - \boldsymbol{\theta}_0) = n^{\frac{1}{2}}(\mathbf{t}_n - \boldsymbol{\theta}_n) + n^{\frac{1}{2}}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0).$$

The first term on the right side converges in distribution to a $\mathscr{N}(\mathbf{0}, \boldsymbol{\Sigma}_0)$ random vector and the second term converges to $\boldsymbol{\delta}$. Hence

$$\mathscr{L}[n^{\frac{1}{2}}(\mathbf{t}_n - \boldsymbol{\theta}_0)] \to \mathscr{N}(\boldsymbol{\delta}, \boldsymbol{\Sigma}_0). \tag{2.3}$$

In (2.2), let $\boldsymbol{\theta}$ assume, as n increases, the values of the random vectors $\mathbf{t}_n$; then $\boldsymbol{\theta}^*$ assumes corresponding values, denoted by $\mathbf{t}_n{}^*$ on the open line segment joining $\boldsymbol{\theta}_0$ and $\mathbf{t}_n$. Consider terms of the form

$$n^{\frac{1}{2}} \frac{\partial^2\gamma_i}{\partial t_{n\alpha}^* \partial t_{n\beta}^*}(t_{n\alpha} - \theta_{0\alpha})(t_{n\beta} - \theta_{0\beta}).$$

From the continuity of the second partial derivative of γ_i and the fact that $\operatorname{plim} \mathbf{t}_n{}^* = \boldsymbol{\theta}_0$ (a consequence of (2.3)), it follows that $\partial^2\gamma_i/\partial t_{n\alpha}^* \partial t_{n\beta}^* = O_p(1)$. From (2.3), $n(t_{n\alpha}-\theta_{0\alpha})(t_{n\beta}-\theta_{0\beta}) = O_p(1)$, and hence $n^{\frac{1}{2}}(t_{n\alpha}-\theta_{0\alpha})(t_{n\beta}-\theta_{0\beta}) = o_p(1)$.

Hence

$$n^{\frac{1}{2}} \frac{\partial^2 \gamma_i}{\partial t_{n\alpha}^* \partial t_{n\beta}^*}(t_{n\alpha}-\theta_{0\alpha})(t_{n\beta}-\theta_{0\beta}) = o_p(1), \qquad i = 1,\cdots,r, \alpha = 1,\cdots,p, \beta = 1,\cdots,p.$$

If (2.2), for $\boldsymbol{\theta} = \mathbf{t}_n$, is multiplied by $n^{\frac{1}{2}}$, then in view of the above and of the fact that $\gamma_i(\boldsymbol{\theta}_0) = 0$, the result

$$n^{\frac{1}{2}}\mathbf{z}_n = n^{\frac{1}{2}}\boldsymbol{\Gamma}_0(\mathbf{t}_n-\boldsymbol{\theta}_0)+o_p(1)$$

is yielded. Hence, by (2.3),

$$\mathscr{L}(n^{\frac{1}{2}}\mathbf{z}_n) \to \mathscr{N}(\boldsymbol{\Gamma}_0\boldsymbol{\delta}, \boldsymbol{\Gamma}_0\boldsymbol{\Sigma}_0\boldsymbol{\Gamma}_0{}'). \tag{2.4}$$

Now to establish (2.1), write

$$n^{\frac{1}{2}}(\mathbf{G}_n\mathbf{S}_n\mathbf{G}_n{}')^{-\frac{1}{2}}\mathbf{z}_n = n^{\frac{1}{2}}[(\mathbf{G}_n\mathbf{S}_n\mathbf{G}_n{}')^{-\frac{1}{2}}-(\boldsymbol{\Gamma}_0\boldsymbol{\Sigma}_0\boldsymbol{\Gamma}_0{}')^{-\frac{1}{2}}]\mathbf{z}_n + n^{\frac{1}{2}}(\boldsymbol{\Gamma}_0\boldsymbol{\Sigma}_0\boldsymbol{\Gamma}_0{}')^{-\frac{1}{2}}\mathbf{z}_n. \tag{2.5}$$

It follows from (2.4) that the second term on the right side of (2.5) is convergent in distribution to $\mathscr{N}((\boldsymbol{\Gamma}_0\boldsymbol{\Sigma}_0\boldsymbol{\Gamma}_0{}')^{-\frac{1}{2}}\boldsymbol{\Gamma}_0\boldsymbol{\delta}, \mathbf{I})$. If it can be shown that the first term is stochastically convergent to zero, i.e., if

$$n^{\frac{1}{2}}[(\mathbf{G}_n\mathbf{S}_n\mathbf{G}_n{}')^{-\frac{1}{2}}-(\boldsymbol{\Gamma}_0\boldsymbol{\Sigma}_0\boldsymbol{\Gamma}_0{}')^{-\frac{1}{2}}]\mathbf{z}_n = o_p(1), \tag{2.6}$$

then (2.1) will be established.

From the fact that $\operatorname{plim} \mathbf{t}_n = \boldsymbol{\theta}_0$ (from (2.3)), and from the continuity of the partial derivatives $\{\partial\gamma_i/\partial\theta_j\}$, it follows that

$$\operatorname{plim} \mathbf{G}_n = \boldsymbol{\Gamma}_0. \tag{2.7}$$

By hypothesis, $\operatorname{plim} \mathbf{S}_n = \boldsymbol{\Sigma}_0$, which together with (2.7) yields

$$\operatorname{plim}(\mathbf{G}_n\mathbf{S}_n\mathbf{G}_n{}') = \boldsymbol{\Gamma}_0\boldsymbol{\Sigma}_0\boldsymbol{\Gamma}_0{}'. \tag{2.8}$$

Since $\boldsymbol{\Gamma}_0\boldsymbol{\Sigma}_0\boldsymbol{\Gamma}_0{}'$ is of full rank, the mapping f defined by $f(\mathbf{A}) = \mathbf{A}^{-\frac{1}{2}}$ is continuous at $\mathbf{A} = \boldsymbol{\Gamma}_0\boldsymbol{\Sigma}_0\boldsymbol{\Gamma}_0{}'$; hence, because of (2.8),

$$\operatorname{plim}(\mathbf{G}_n\mathbf{S}_n\mathbf{G}_n{}')^{-\frac{1}{2}} = (\boldsymbol{\Gamma}_0\boldsymbol{\Sigma}_0\boldsymbol{\Gamma}_0{}')^{-\frac{1}{2}}. \tag{2.9}$$

By (2.4), $n^{\frac{1}{2}}\mathbf{z}_n = O_p(1)$, which together with (2.9) implies that

$$n^{\frac{1}{2}}[(\mathbf{G}_n\mathbf{S}_n\mathbf{G}_n{}')^{-\frac{1}{2}}-(\boldsymbol{\Gamma}_0\boldsymbol{\Sigma}_0\boldsymbol{\Gamma}_0{}')^{-\frac{1}{2}}]\mathbf{z}_n = o_p(1)O_p(1) = o_p(1).$$

Thus (2.6), and hence (2.1), are established. □

The above theorem is designed for application in situations where the local differentiability conditions on the function $\boldsymbol{\gamma}$ and the full rank of its matrix of partial derivatives hold for every $\boldsymbol{\theta}_0$ in a given subset Θ of $\mathscr{E}^p$. Θ is to be regarded as the parameter space, and it may be smaller than the range of values of $\mathbf{t}_n$. If, for each $\boldsymbol{\theta}_0 \in \Theta$ and every sequence $\{\delta_n\}$ converging to a point $\delta \in \mathscr{E}^p$, the distributions of $\mathbf{t}_n$ and $\mathbf{S}_n$ are determined in such a way that the relevant conditions of the theorem are satisfied, then a test which rejects the hypothesis that $\gamma(\boldsymbol{\theta}) = \mathbf{0}$ for large values of J_n is an asymptotic χ^2-test of the hypothesis.

3. Application to normal testing problems: Verification of conditions. The theorem of the preceding section is applied to obtain asymptotic distributions of a proposed test statistic for the general hypothesis testing problem of the form $H: \gamma(\boldsymbol{\theta}) = 0$, where $\boldsymbol{\theta}$ represents the parameters of two multivariate normal populations, from which samples of unequal size are observed, and γ is any vector-valued function with continuous second partial derivatives and full linear rank. The parameter space must be an open subset of the full-dimensional space of conceivable values, such that the covariance matrices are always strictly positive definite. If γ involves the parameters of only one population the test statistic will be based on the corresponding sample, so that one-sample problems are included here as a special case. The method described below can clearly be extended to problems involving samples from more than two populations.

Let the q-dimensional mean vectors of the two populations be $\boldsymbol{\mu}$ and $\mathbf{v}$, and let the covariance matrices be $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, respectively, assumed positive definite. The parameter space Θ is a non-empty open subset of the $q(q+3)$-dimensional space of possible values of $\boldsymbol{\theta} = (\boldsymbol{\mu}, \mathbf{v}, \boldsymbol{\Phi}, \boldsymbol{\Psi})$. Samples of sizes M and N are observed, on the basis of which it is desired to test, against all alternatives, the hypothesis

$$H: \gamma(\boldsymbol{\theta}) = 0,$$

where for any $\boldsymbol{\theta} \in \Theta$ the function $\gamma: \Theta \to \mathscr{E}^r$ is assumed to have full rank $r \leqq q(q+3)$.

The proposed test statistic is based on the sufficient statistic composed of the sample means and sample covariance matrices. Denote this by $(\mathbf{x}^M, \mathbf{y}^N, \hat{\boldsymbol{\Phi}}^M, \hat{\boldsymbol{\Psi}}^N)$; then $\mathbf{x}^M$ and $\mathbf{y}^N$ have the normal distributions $\mathscr{N}(\boldsymbol{\mu}, M^{-1}\boldsymbol{\Phi})$ and $\mathscr{N}(\mathbf{v}, N^{-1}\boldsymbol{\Psi})$, respectively, $(M-1)\,\hat{\boldsymbol{\Phi}}^M$ has the Wishart distribution with $M-1$ degrees of freedom and expectation $(M-1)\boldsymbol{\Phi}$, denoted by $\mathscr{W}(\boldsymbol{\Phi}, M-1)$, and $(N-1)\hat{\boldsymbol{\Psi}}^N$ has the Wishart distribution $\mathscr{W}(\boldsymbol{\Psi}, N-1)$. The four quantities are stochastically independent.

The vector $(\mathbf{x}^M, \mathbf{y}^N, \hat{\boldsymbol{\Phi}}^M, \hat{\boldsymbol{\Psi}}^N)$ will be used in the theorem as the estimator of $\boldsymbol{\theta}$. (The results will be essentially unchanged if $\boldsymbol{\Phi}^M$ and $\hat{\boldsymbol{\Psi}}^N$ are rescaled in some asymptotically unimportant way, e.g., maximum-likelihood estimators.) Since for convenience it is desirable that all quantities depend on a single index N, let M depend on N in the manner $M = [N/\rho]$, where ρ is a constant and $[\cdot]$ is the greatest integer function. Denote

$$\mathbf{t}^N = (\mathbf{x}^M, \mathbf{y}^N, \hat{\boldsymbol{\Phi}}^M, \hat{\boldsymbol{\Psi}}^N).$$

Then [see Anderson (1958, page 75)] $\mathbf{t}^N$ has expectation $\boldsymbol{\theta}$ and the block-diagonal covariance matrix $\operatorname{diag}[M^{-1}\boldsymbol{\Phi}, N^{-1}\boldsymbol{\Psi}, (M-1)^{-1}\mathbf{H}(\boldsymbol{\Phi}), (N-1)^{-1}\mathbf{H}(\boldsymbol{\Psi})]$, where

$\mathbf{H}$ is the function which maps the $q \times q$ symmetric matrix $\mathbf{A} = (a_{ij})$ into the $q(q+1)/2 \times q(q+1)/2$ symmetric matrix whose components are

$$[\mathbf{H}(\mathbf{A})]_{ij,kl} = a_{ik}a_{jl} + a_{il}a_{jk}, \qquad 1 \leqq i \leqq j \leqq q,\ 1 \leqq k \leqq l \leqq g,$$

arranged in lexicographic order. Hence the covariance matrix of $N^{\frac{1}{2}}(\mathbf{t}^N - \boldsymbol{\theta})$ is

$$\boldsymbol{\Sigma}^N \equiv \operatorname{diag}[(N/M)\boldsymbol{\Phi}, \boldsymbol{\Psi}, (N/(M-1))\mathbf{H}(\boldsymbol{\Phi}), (N/(N-1))\mathbf{H}(\boldsymbol{\Psi})].$$

For $\boldsymbol{\Sigma}^N$ we have the obvious estimator

$$\mathbf{S}^N = \operatorname{diag}[(N/M)\hat{\boldsymbol{\Phi}}^M, \hat{\boldsymbol{\Psi}}^N, (N/(M-1))\mathbf{H}(\hat{\boldsymbol{\Phi}}^M), (N/(N-1))\mathbf{H}(\hat{\boldsymbol{\Psi}}^N)].$$

The proposed test statistic J^N is then constructed from $\mathbf{t}^N$ and $\mathbf{S}^N$ in the same manner that J_n was defined in the statement of the theorem.

The quantities $\mathbf{t}^N$ and $\mathbf{S}^N$ have been defined on the basis of observations distributed according to the parameter $\boldsymbol{\theta}$. The remainder of this section consists of a verification that the conditions of the theorem hold for the problem under consideration. We consider a sequence of parameter values $\boldsymbol{\theta}_n = (\boldsymbol{\mu}_n, \boldsymbol{\nu}_n, \boldsymbol{\Phi}_n, \boldsymbol{\Psi}_n)$ given by $\boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + n^{-\frac{1}{2}}\boldsymbol{\delta}_n$, where $\boldsymbol{\delta}_n$ converges to some fixed $\boldsymbol{\delta}$ and $\boldsymbol{\theta}_0 = (\boldsymbol{\mu}_0, \boldsymbol{\nu}_0, \boldsymbol{\Phi}_0, \boldsymbol{\Psi}_0)$ is such that $\gamma(\boldsymbol{\theta}_0) = 0$. For this purpose a subscript n will be appended to the observed quantities, which also have superscripts M or N. The superscript refers to the sample size, and the subscript n means that the distribution of the observations is determined by the parameter value $\boldsymbol{\theta}_n$. For example, the symbol $\mathbf{t}_n^N$ refers to the statistic $\mathbf{t}^N$ observed when the parameter value is $\boldsymbol{\theta}_n$.

Define

$$\boldsymbol{\Sigma}_0 = \operatorname{diag}[\rho\boldsymbol{\Phi}_0, \boldsymbol{\Psi}_0, \rho\mathbf{H}(\boldsymbol{\Phi}_0), \mathbf{H}(\boldsymbol{\Psi}_0)].$$

To apply the theorem we determine the sample size by setting $N = n$; the test statistic J_n^n is based on $\mathbf{t}_n^n$ and $\mathbf{S}_n^n$. The asymptotic distribution given by the theorem will be shown valid when the relations $\boldsymbol{\Sigma}_0 > 0$, $\operatorname{plim} \mathbf{S}_n^n = \boldsymbol{\Sigma}_0$ and

$$\mathscr{L}[n^{\frac{1}{2}}(\mathbf{t}_n^n - \boldsymbol{\theta}_n)] \to \mathscr{N}(\mathbf{0}, \boldsymbol{\Sigma}_0)$$

are established.

By the following lemma, $\boldsymbol{\Phi}_0 > 0$ and $\boldsymbol{\Psi}_0 > 0$ imply that $\mathbf{H}(\boldsymbol{\Phi}_0) > 0$ and $\mathbf{H}(\boldsymbol{\Psi}_0) > 0$, which together with the hypothesis $\rho > 0$ yields the result $\boldsymbol{\Sigma}_0 > 0$.

LEMMA 1. *Let* $\mathbf{A}$ *be a* $q \times q$ *symmetric matrix and let* $[\mathbf{H}(\mathbf{A})]_{ij,kl} = a_{ik}a_{jl} + a_{il}a_{jk} \equiv \tilde{a}_{ij,kl}$. *If* $\mathbf{A} > 0$, *then* $\mathbf{H}(\mathbf{A}) > 0$.

PROOF. It needs to be shown that

$$\sum_{1 \leqq i \leqq j \leqq q,\, 1 \leqq k \leqq l \leqq q} w_{ij}\tilde{a}_{ij,kl}w_{kl} > 0,$$

for all symmetric nonzero $\mathbf{W} = (w_{ij})$.

Define $\mathbf{X} = (x_{ij})$ by

$$x_{ii} = w_{ii}$$
$$x_{ij} = \tfrac{1}{2}w_{ij} \qquad (i \neq j)$$

Then

$$\textstyle\sum_{1 \leqq i \leqq j \leqq q, 1 \leqq k \leqq l \leqq q} w_{ij}\tilde{a}_{ij,kl}w_{kl} = 2 \operatorname{tr} \mathbf{XAXA}. \tag{3.1}$$

To see this, note that the operator identity

$$\textstyle\sum_{\alpha \leqq \beta} = \sum_{\alpha} + \frac{1}{2}\sum_{\alpha \neq \beta}$$

is valid when the operand is symmetric in α and β; (3.1) then follows by direct computation.

But, since $\mathbf{A} > 0$, $\operatorname{tr} \mathbf{XAXA} = \operatorname{tr} \mathbf{A}^{\frac{1}{2}}\mathbf{XAXA}^{\frac{1}{2}} > 0$. □

To show that plim $\mathbf{S}_n{}^n = \mathbf{\Sigma}_0$ it is sufficient to show plim $\hat{\mathbf{\Psi}}_n{}^n = \mathbf{\Psi}_0$, for then plim $\mathbf{H}(\hat{\mathbf{\Psi}}_n{}^n) = \mathbf{H}(\mathbf{\Psi}_0)$ follows from the continuity of $\mathbf{H}$, and similar reasoning can be applied to the $\mathbf{\Phi}$ quantities. Let $\varepsilon > 0$ be given; then

$$\begin{aligned} P\{\|\hat{\mathbf{\Psi}}_n{}^n - \mathbf{\Psi}_0\| > \varepsilon\} &\leqq P\{\|\hat{\mathbf{\Psi}}_n{}^n - \mathbf{\Psi}_n\| > \varepsilon/2\} + P\{\|\mathbf{\Psi}_n - \mathbf{\Psi}_0\| > \varepsilon/2\} \\ &\leqq \textstyle\sum_{i \leqq j} P\{|\hat{\psi}_{nij}^n - \psi_{nij}| > \varepsilon/q(q+1)\} + P\{\|\mathbf{\Psi}_n - \mathbf{\Psi}_0\| > \varepsilon/2\} \end{aligned}$$

where $\|\cdot\|$ represents the vector norm in $\mathscr{E}^{q(q+1)/2}$. The last term on the right is zero for sufficiently large n. Since $E(\hat{\psi}_{nij}^n) = \psi_{nij}$ and $\operatorname{Var}(\hat{\psi}_{nij}^n) = (\psi_{nii}\psi_{njj} + \psi_{nij}^2)/(n-1)$, it follows by Chebyshev's inequality that for each i, j

$$\begin{aligned} P\{|\hat{\psi}_{nij}^n - \psi_{nij}| > \varepsilon/q(q+1)\} &\leqq \frac{[q(q+1)]^2(\psi_{nii}\psi_{njj} + \psi_{nij}^2)}{(n-1)\varepsilon^2} \\ &\leqq \frac{4[q(q+1)]^2(\psi_{0ii}\psi_{0jj} + \psi_{0ij}^2)}{(n-1)\varepsilon^2} \end{aligned}$$

for n sufficiently large to ensure $\psi_{nij} \leqq 2\psi_{0ij}$ for all $i, j = 1, \cdots, q$. Thus $\sum_{i \leqq j} P\{|\psi_{nij}^n - \psi_{nij}| > \varepsilon/q(q+1)\} \leqq k/(n-1)\varepsilon^2$, where $k = 2[q(q+1)]^3 \max_{i,j} (\psi_{0ii}\psi_{0jj} + \psi_{0ij}^2)$, and hence this term can be made arbitrarily small by making n sufficiently large. Consequently plim $\hat{\mathbf{\Psi}}_n{}^n = \mathbf{\Psi}_0$.

The next step is to show that $\mathscr{L}[n^{\frac{1}{2}}(\mathbf{t}_n{}^n - \boldsymbol{\theta}_n)] \to \mathscr{N}(\mathbf{0}, \mathbf{\Sigma}_0)$, where $\mathbf{t}_n{}^n = (\mathbf{x}_n{}^{[n/\rho]}, \mathbf{y}_n{}^n, \hat{\mathbf{\Phi}}_n{}^{[n/\rho]}, \hat{\mathbf{\Psi}}_n{}^n)$. Since $\mathbf{x}_n{}^{[n/\rho]}$, $\mathbf{y}_n{}^n$, $\hat{\mathbf{\Phi}}_n{}^{[n/\rho]}$ and $\hat{\mathbf{\Psi}}_n{}^n$ are independent, it suffices to show that each of these quantities, centered at its expectation and normalized by $n^{\frac{1}{2}}$, is asymptotically normal as $n \to \infty$ with zero mean and covariance matrix, respectively, $\rho\mathbf{\Phi}_0$, $\mathbf{\Psi}_0$, $\rho\mathbf{H}(\mathbf{\Phi}_0)$ or $\mathbf{H}(\mathbf{\Psi}_0)$.

Since $n^{\frac{1}{2}}[\mathbf{\Psi}_n{}^{-\frac{1}{2}}(\mathbf{y}_n{}^n - \mathbf{v}_n)]$ has the distribution $\mathscr{N}(\mathbf{0}, \mathbf{I})$ and since $\mathbf{\Psi}_0{}^{-\frac{1}{2}}\mathbf{\Psi}_n{}^{\frac{1}{2}} \to \mathbf{I}$, it follows that $\mathscr{L}\{n^{\frac{1}{2}}[\mathbf{\Psi}_0{}^{-\frac{1}{2}}(\mathbf{y}_n{}^n - \mathbf{v}_n)]\} \to \mathscr{N}(\mathbf{0}, \mathbf{I})$, or equivalently,

$$\mathscr{L}[n^{\frac{1}{2}}(\mathbf{y}_n{}^n - \mathbf{v}_n)] \to \mathscr{N}(\mathbf{0}, \mathbf{\Psi}_0).$$

Similarly,

$$\mathscr{L}[n^{\frac{1}{2}}(\mathbf{x}_n{}^{[n/\rho]} - \boldsymbol{\mu}_n)] \to \mathscr{N}(\mathbf{0}, \rho\mathbf{\Phi}).$$

To obtain the limiting distribution of $n^{\frac{1}{2}}(\hat{\mathbf{\Psi}}_n{}^n - \mathbf{\Psi}_n)$, we adapt the method presented in Anderson ((1958) Theorem 4.2.4) to the needs of the present problem which involves parameter values changing with n. Where Anderson uses a multivariate adaptation of the classical central limit theorem, we similarly adapt a limit theorem for double sequences.

Consider the representation

$$(n-1)\hat{\mathbf{\Psi}}_n{}^n = \textstyle\sum_{\alpha=1}^{n-1} \mathbf{u}_\alpha \mathbf{u}_\alpha{}'$$

where $\mathbf{u}_1, \mathbf{u}_2, \cdots$ are independent q-dimensional random vectors each with the distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Psi}_n)$. Let $\mathbf{w}_\alpha$ be the column vector of the $q(q+1)/2$ products of components of $\mathbf{u}_\alpha$

$$\mathbf{w}_\alpha{}' = (u_{1\alpha}^2, u_{1\alpha}u_{2\alpha}, \cdots, u_{2\alpha}^2, \cdots, u_{q\alpha}^2), \qquad \alpha = 1, \cdots, n-1.$$

The components of the matrix $(n-1)\hat{\mathbf{\Psi}}_n{}^n$ are identical to the components of the vector $\sum_{\alpha=1}^{n-1} w_\alpha$. Note the dependence of $u_{i\alpha}$ and $\mathbf{w}_\alpha$ on n, which has been suppressed in the notation. The mean vector and covariance matrix of $\mathbf{w}_\alpha$ are given by

$$E(u_{i\alpha}u_{j\alpha}) = \psi_{nij}, \tag{3.2}$$

$$\operatorname{Cov}(u_{i\alpha}u_{j\alpha}, u_{k\alpha}u_{l\alpha}) = \psi_{nik}\psi_{njl} + \psi_{nil}\psi_{njk}.$$

Let $\mathbf{v}_n$ be the vector obtained from $\sum_{\alpha=1}^{n-1}\mathbf{w}_\alpha$ by subtracting away its expectation and dividing by $(n-1)^{\frac{1}{2}}$. Then the components of $\mathbf{v}_n$ are the components of $(n-1)^{\frac{1}{2}}(\hat{\mathbf{\Psi}}_n{}^n - \mathbf{\Psi}_n)$; i.e., these components are of the form $(n-1)^{-\frac{1}{2}}\sum_{\alpha=1}^{n-1}(u_{i\alpha}u_{j\alpha} - \psi_{nij})$. The limiting distribution of $\mathbf{v}_n$ is determined by the limit of its characteristic function $f_n(\boldsymbol{\tau}) = E[\exp(i\boldsymbol{\tau}'\mathbf{v}_n)]$, where $\boldsymbol{\tau}$ represents an arbitrary fixed vector. Using the Cramér–Wold technique (Cramér (1970) page 104), we consider the characteristic function of an arbitrary nontrivial linear functional of $\mathbf{v}_n$, say $\boldsymbol{\lambda}'\mathbf{v}_n$: $E[\exp(i\tau\boldsymbol{\lambda}'\mathbf{v}_n)] = f_n(\tau\boldsymbol{\lambda})$. Regard this as a function of an arbitrary scalar τ, with $\boldsymbol{\lambda}$ fixed. We proceed to establish that

$$f_n(\tau\boldsymbol{\lambda}) \to \exp((-\tau^2/2)\boldsymbol{\lambda}'[\mathbf{H}(\mathbf{\Psi}_0)]\boldsymbol{\lambda}), \tag{3.3}$$

which by setting $\boldsymbol{\tau} = \tau\boldsymbol{\lambda}$ implies $f_n(\boldsymbol{\tau}) \to \exp((-\frac{1}{2})\boldsymbol{\tau}'[\mathbf{H}(\mathbf{\Psi}_0)]\boldsymbol{\tau})$, proving that the asymptotic distribution of $\mathbf{v}_n$, or equivalently that of $n^{\frac{1}{2}}(\hat{\mathbf{\Psi}}_n{}^n - \mathbf{\Psi}_n)$, is $\mathcal{N}(\mathbf{0}, \mathbf{H}(\mathbf{\Psi}_0))$.

Relation (3.3) simply states that the limiting distribution of $\boldsymbol{\lambda}'\mathbf{v}_n$ is $\mathcal{N}(0, \boldsymbol{\lambda}'[\mathbf{H}(\mathbf{\Psi}_0)]\boldsymbol{\lambda})$. Since $\boldsymbol{\lambda}'[\mathbf{H}(\mathbf{\Psi}_n)]\boldsymbol{\lambda} \to \boldsymbol{\lambda}'[\mathbf{H}(\mathbf{\Psi}_0)]\boldsymbol{\lambda}$, it is sufficient to prove that

$$\mathscr{L}\{\boldsymbol{\lambda}'\mathbf{v}_n/[\boldsymbol{\lambda}'H(\mathbf{\Psi}_n)\boldsymbol{\lambda}]^{\frac{1}{2}}\} \to \mathcal{N}(0,1). \tag{3.4}$$

This is accomplished by the use of the Lévy–Feller normal convergence criterion (Loève (1963) page 295). In notation similar to Loève's, let

$$X_{n\alpha} = (n-1)^{-\frac{1}{2}}\textstyle\sum_{i \leq j} \lambda_{ij}(u_{i\alpha}u_{j\alpha} - \psi_{nij})/[\boldsymbol{\lambda}'\mathbf{H}(\mathbf{\Psi}_n)\boldsymbol{\lambda}]^{\frac{1}{2}}$$

for $\alpha = 1, \cdots, n-1$, where for $i \leq j$ the real numbers λ_{ij} are the components of the $q(q+1)/2$-dimensional vector $\boldsymbol{\lambda}$. Then (3.4) refers to convergence to $\mathcal{N}(0, 1)$ of

$\sum_{\alpha=1}^{n-1} X_{n\alpha}$, which has mean zero and variance one. By the cited criterion, this is assured if for every $\varepsilon > 0$

$$\sum_{\alpha=1}^{n-1} \int_{|x| \geqq \varepsilon} x^2 \, dF_{n\alpha} \to 0,$$

where $F_{n\alpha}$ is the cdf of $X_{n\alpha}$. Since for fixed n the $X_{n\alpha}$ are identically distributed (3.4) holds if we can show that

$$E[X_{n\alpha}^2 I_\varepsilon(X_{n\alpha})] = o(n^{-1}), \tag{3.5}$$

where I_ε is the indicator function of $\{x : |x| \geqq \varepsilon\}$. Relation (3.5) is established with the aid of an inequality given by the following lemma.

LEMMA 2. *If Y is any random variable such that $E(Y) = 0$ and* $\operatorname{Var}(Y^2) < \infty$, *and $I_\varepsilon(\cdot)$ is the indicator function of the set $\{x: |x| \geqq \varepsilon\}$, then*

$$E(Y^2 I_\varepsilon(Y)) \leqq [\operatorname{Var}(Y)]^2/\varepsilon^2 + [\operatorname{Var}(Y)\operatorname{Var}(Y^2)]^{\frac{1}{2}}/\varepsilon.$$

PROOF. Application of the inequality $E(UV) \leqq E(U)E(V) + [\operatorname{Var}(U)\operatorname{Var}(V)]^{\frac{1}{2}}$, i.e. that the correlation coefficient is bounded by unity, yields

$$E(Y^2 I_\varepsilon(Y)) \leqq E(Y^2)p_\varepsilon + [p_\varepsilon(1-p_\varepsilon)\operatorname{Var}(Y^2)]^{\frac{1}{2}},$$

where $p_\varepsilon = P\{|Y| \geqq \varepsilon\}$. The lemma follows by applying Chebyshev's inequality to p_ε and by using the fact that $p_\varepsilon(1-p_\varepsilon) \leqq p_\varepsilon$. □

We now show that

$$[\operatorname{Var}(X_{n\alpha})]^2 = o(n^{-1}), \tag{3.6}$$
$$\operatorname{Var}(X_{n\alpha})\operatorname{Var}(X_{n\alpha}^2) = o(n^{-2}),$$

which on the basis of Lemma 2, with $Y = X_{n\alpha}$, will establish (3.5). By (3.2), $\operatorname{Var}(u_{i\alpha}u_{j\alpha}) = \psi_{nii}\psi_{njj} + \psi_{nij}^2$, which remains bounded as $n \to \infty$ by virtue of $\mathbf{\Psi}_n \to \mathbf{\Psi}_0$. A similar identity may be derived expressing the variance of $u_{i\alpha}u_{j\alpha}u_{k\alpha}u_{l\alpha}$ as a fourth degree expression in the ψ_{nij}, also bounded as $n \to \infty$. It follows from $\mathbf{\Psi}_n \to \mathbf{\Psi}_0$ that $\boldsymbol{\lambda}'\mathbf{H}(\mathbf{\Psi}_n)\boldsymbol{\lambda} \to \boldsymbol{\lambda}'\mathbf{H}(\mathbf{\Psi}_0)\boldsymbol{\lambda} > 0$, so that the denominator of $X_{n\alpha}$ is bounded away from zero. Thus the orders of magnitude of the fourth and lower moments of $X_{n\alpha}$ are governed solely by the factor $(n-1)^{-\frac{1}{2}}$; specifically, $E(X_{n\alpha}^2) = O(n^{-1})$ and $E(X_{n\alpha}^4) = O(n^{-2})$. Hence $[\operatorname{Var}(X_{n\alpha})]^2 = O(n^{-2})$ and $\operatorname{Var}(X_{n\alpha})\operatorname{Var}(X_n{}^2) = O(n^{-3})$, thus proving (3.6), and hence (3.5) and (3.4). It has now been established that

$$[n^{\frac{1}{2}}(\hat{\mathbf{\Psi}}_n{}^n - \mathbf{\Psi}_n)] \to \mathcal{N}(\mathbf{0}, \mathbf{H}(\mathbf{\Psi})).$$

It is similarly shown that

$$\mathcal{L}[n^{\frac{1}{2}}(\mathbf{\Phi}^{[n/\rho]} - \mathbf{\Phi}_n)] \to \mathcal{N}(\mathbf{0}, \rho\mathbf{H}(\mathbf{\Phi})).$$

This completes the verification of the conditions of the theorem of Section 2.

4. Application to normal testing problems: Computation of the statistic J_N. Having verified that the asymptotic distribution, under $\{\boldsymbol{\theta}_n\}$, of the test statistic

$J_n{}^n$ as defined in Sections 2 and 3 is central (noncentral) chi-square when the hypothesis $\gamma(\boldsymbol{\theta}) = \mathbf{0}$ is true (false), we turn our attention to the computation of the test statistic for given sample sizes M and N. As in Section 3, we make the identification $N = n$, and for simplicity we write $\rho = M/N$. Instead of writing $\mathbf{t}_n{}^n$ and $\mathbf{S}_n{}^n$, we now simply write $\mathbf{t}_N = (\mathbf{x}_M, \mathbf{y}_N, \hat{\mathbf{\Phi}}_M, \hat{\mathbf{\Psi}}_N)$ and $\mathbf{S}_N = \text{diag}\,(\rho\hat{\mathbf{\Phi}}_M, \hat{\mathbf{\Psi}}_N, \rho\mathbf{H}(\hat{\mathbf{\Phi}}_M), \mathbf{H}(\hat{\mathbf{\Psi}}_N))$, where we use subscripts instead of superscripts for convenience. Similarly we write $\mathbf{z}_N = \gamma(\mathbf{t}_N)$ and we write $\mathbf{G}_N$ for the matrix of partial derivatives of γ evaluated at $\mathbf{t}_N$. The test statistic is

$$J_N = N\mathbf{z}_N(\mathbf{G}_N\mathbf{S}_N\mathbf{G}_N')^{-1}\mathbf{z}_N,$$

which can be readily calculated if the matrix $\hat{\mathbf{\Omega}}_N \equiv \mathbf{G}_N\mathbf{S}_N\mathbf{G}_N'$ is at hand. If the form of $\mathbf{\Omega} \equiv \mathbf{\Gamma\Sigma\Gamma}'$ is available, where $\mathbf{\Gamma}$ is the matrix of partial derivatives of γ evaluated at an arbitrary $\boldsymbol{\theta} = (\boldsymbol{\mu}, \mathbf{v}, \mathbf{\Phi}, \mathbf{\Psi})$, and

$$\mathbf{\Sigma} = \text{diag}(\rho\mathbf{\Phi}, \mathbf{\Psi}, \rho\mathbf{H}(\Phi), \mathbf{H}(\mathbf{\Psi})), \tag{4.1}$$

then $\hat{\mathbf{\Omega}}_N$ is obtained by simply evaluating $\mathbf{\Omega}$ at $\boldsymbol{\theta} = \mathbf{t}_N$. The remainder of this paper is devoted to obtaining a description of the matrix $\mathbf{\Omega}$ for the general normal testing problem of Section 3.

The (i,j)th component of $\mathbf{\Omega}$ is given by

$$\omega_{ij} = \mathbf{\Gamma}_{i\cdot}\mathbf{\Sigma\Gamma}_{j\cdot}' \qquad (i,j = 1, \cdots, r), \tag{4.2}$$

where $\mathbf{\Gamma}_{i\cdot}$ denotes the ith row of $\mathbf{\Gamma}$, and $\mathbf{\Gamma}_{i\cdot}'$ denotes its transpose. $\mathbf{\Gamma}_{i\cdot}$ is the row vector whose components are

$$(\mathbf{\Gamma}_{i\cdot})_u = \frac{\partial\gamma_i}{\partial\theta_u} \qquad (u = 1, \cdots, p),$$

where $p = q(q+3)$. Hence, by (4.2),

$$\omega_{ij} = \sum_{u=1}^{p}\sum_{v=1}^{p}\frac{\partial\gamma_i}{\partial\theta_u}\frac{\partial\gamma_j}{\partial\theta_v}\sigma_{uv} \qquad (i,j = 1, \cdots, r),$$

where $\{\sigma_{uv}\}$ are the components of the matrix $\mathbf{\Sigma}$. To streamline the notation, let $\alpha = \gamma_i$, $\beta = \gamma_j$, and $\omega(\alpha, \beta) = \omega_{ij}$. The problem of writing down the formula for $\mathbf{\Omega} = \mathbf{\Gamma\Sigma\Gamma}'$ is reduced to the problem of evaluating

$$\omega(\alpha, \beta) = \sum_{u=1}^{p}\sum_{v=1}^{p}\frac{\partial\alpha}{\partial\theta_u}\frac{\partial\beta}{\partial\theta_v}\sigma_{uv} \tag{4.3}$$

for arbitrary real variables α, β expressible as differentiable functions of the vector $\boldsymbol{\theta}$.

Referring to (4.1), we may rewrite (4.3) in terms of the components of $\boldsymbol{\mu}$, $\mathbf{v}$, $\mathbf{\Phi}$ and $\mathbf{\Psi}$ as follows:

$$\omega(\alpha, \beta) = \rho\sum_{a,b=1}^{s}\frac{\partial\alpha}{\partial\mu_a}\frac{\partial\beta}{\partial\mu_b}\phi_{ab} + \sum_{a,b=1}^{s}\frac{\partial\alpha}{\partial v_a}\frac{\partial\beta}{\partial v_b}\psi_{ab}$$

$$(4.4) \qquad +\rho \sum_{a\leq b,c\leq d} \frac{\partial\alpha}{\partial\phi_{ab}} \frac{\partial\beta}{\partial\phi_{cd}} (\phi_{ca}\phi_{bd}+\phi_{ad}\phi_{bc}) + \sum_{a\leq b,c\leq d} \frac{\partial\alpha}{\partial\psi_{ab}} \frac{\partial\beta}{\partial\psi_{cd}} (\psi_{ac}\psi_{bd}+\psi_{ad}\psi_{bc}).$$

Equation (4.4) may be simplified by the use of a convenient notation for the derivative of a real variable with respect to a matrix. In general, let f be a real-valued function of a matrix **A**, and let $y = f(\mathbf{A})$. By the matrix derivative $dy/d\mathbf{A}$ we shall mean one of the matrices described below, according to whether **A** is (i) $m \times n$, or (ii) $m \times m$ symmetric.

(i) If **A** is $m \times n$, then $dy/d\mathbf{A}$ is $m \times n$, and its (i, j)th component is

$$(4.5) \qquad (dy/d\mathbf{A})_{ij} = \frac{\partial y}{\partial a_{ij}} \qquad (i = 1, \cdots, m; j = 1, \cdots, n).$$

(ii) If **A** is $m \times m$ symmetric, then $dy/d\mathbf{A}$ is $m \times m$ symmetric, and its components are given by

$$(4.6) \qquad (dy/d\mathbf{A})_{ii} = \partial y/\partial a_{ii} \qquad (i = 1, \cdots, m);$$
$$(dy/d\mathbf{A})_{ij} = \tfrac{1}{2}\partial y/\partial a_{ij} = \tfrac{1}{2}\partial y/\partial a_{ji} \qquad (i \neq j; i, j = 1, \cdots, m).$$

The application of the factor $\frac{1}{2}$ in differentiation with respect to off-diagonal elements of symmetric matrices was used by Aitken (1953).

If f is a real-valued function of two matrices **A** and **B** with $y = f(\mathbf{A}, \mathbf{B})$, then the matrix partial derivative $y_A = \partial y/\partial\mathbf{A}$ is defined as above holding **B** constant; $y_B = \partial y/\partial\mathbf{B}$ is similarly defined.

It may be noted that the matrix derivative $dy/d\mathbf{A}$ defined in (4.6) can be thought of as corresponding to the matrix **X** used in the proof of Lemma 1, where the matrix **W** used in this proof corresponds to the matrix obtained simply by writing down all the scalar partial derivatives $\partial y/\partial a_{ij}$. Applying (3.1) to the third and fourth expressions on the right-hand side of (4.4), we may write (4.4), using partial matrix derivatives as defined by (4.5) and (4.6), in the form

$$(4.7) \qquad \omega(\alpha, \beta) = \rho\alpha_\mu'\mathbf{\Phi}\beta_\mu + \alpha_\nu'\mathbf{\Psi}\beta_\nu + 2\rho \operatorname{tr} \alpha_\Phi\mathbf{\Phi}\beta_\Phi\mathbf{\Phi} + 2 \operatorname{tr} \alpha_\Psi\mathbf{\Psi}\beta_\Psi\mathbf{\Psi}.$$

Since α and β represent components γ_i and γ_j, respectively, of the vector $\gamma(\boldsymbol{\theta})$, with $\omega(\alpha, \beta) = \omega_{ij}$, (4.7) exhibits the components ω_{ij} of the matrix $\mathbf{\Omega} = \mathbf{\Gamma\Sigma\Gamma}'$.

Acknowledgment. I am grateful to Professor Ingram Olkin for his encouragement and guidance during the writing of the dissertation. I wish to thank R. P. Singh for a piece of bibliographical assistance.

REFERENCES

[1] AITKEN, A. C. (1953). A note on trace-differentiation and the Ω-operator. *Proc. Edinburgh Math. Soc.* (2) **10** 1–4.

[2] ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis.* Wiley, New York.

[3] Apostol, T. M. (1957). *Mathematical Analysis*. Addison-Wesley, Reading.
[4] Chernoff, H. (1956). Large sample theory: parametric case. *Ann. Math. Statist.* **27** 1–22.
[5] Cramér, H. (1970). *Random Variables and Probability Distributions*, 3rd ed. Cambridge Tracts in Mathematics and Mathematical Physics, No. 36, Cambridge Univ. Press.
[6] Doob, J. L. (1934). Probability and Statistics. *Trans. Amer. Math. Soc.* **36** 759–775.
[7] Kendall, M. G. and Stuart, A. (1961). *The Advanced Theory of Statistics* **2**. Hafner, New York.
[8] LeCam, L. (1956). On the asymptotic theory of estimation and testing hypotheses. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 129–156.
[9] Loève, M. (1963). *Probability Theory*, 3rd ed. Van Nostrand, Princeton.
[10] Stroud, T. W. F. (1968). Comparing conditional distributions under measurement errors of known variances. Technical Report No. 28, Department of Statistics, Stanford Univ.
[11] Stroud, T. W. F. (1970). Fixed alternatives and Wald's formulation of the noncentral asymptotic behavior of the likelihood ratio statistic. Queen's Mathematical Preprint No. 1970–33, Queen's University at Kingston.
[12] Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* **54** 426–482.
[13] Welch, B. L. (1938). The significance of the differences between two means when the population variances are unequal. *Biometrika* **29** 360.