The Annals of Statistics
1979, Vol. 7, No. 4, 823–837

ASYMPTOTIC NONNULL DISTRIBUTIONS FOR LIKELIHOOD RATIO STATISTICS IN THE MULTIVARIATE NORMAL PATTERNED MEAN AND COVARIANCE MATRIX TESTING PROBLEM[1]

BY TED H. SZATROWSKI

Rutgers University

The multivariate normal patterned mean and covariance matrix testing problem is studied for general one and k-population hypotheses. T. W. Anderson's iterative algorithm for finding the maximum likelihood estimates, the forms of the likelihood ratio tests, and asymptotic chi-square distributions of these tests under the null hypothesis are given. The nonnull asymptotic normal distribution is derived using the standard delta method. This derivation involves using several extensions of matrix identities given in Anderson, matrix derivatives and asymptotic likelihood equations. The forms of the variances are greatly simplified using a result of Szatrowski when the maximum likelihood estimates under the null hypothesis have explicit representations.

1. Introduction. Hypothesis tests involving linear patterned mean and covariance matrix assumptions for multivariate normal data have received much attention in the literature. Problems that motivate these studies arise in many applications including those in the areas of medicine, psychology, educational testing and reliability testing. In many of these studies, (e.g. Wilks (1946), Votaw (1948), Gleser and Olkin (1966, 1969), Olkin and Press (1969), Olkin (1972), Arnold (1973) and Szatrowski (1976)), the maximum likelihood estimates and likelihood ratio tests can be explicitly obtained. In some cases the exact null distribution of the likelihood ratio statistic is obtained (e.g. Consul (1968, 1969), Khatri and Srivastava (1971)). In other cases the asymptotic chi-square distribution is given (e.g. Mukherjee (1966, 1970), Anderson (1970)) or the moments of the likelihood ratio test are obtained and the null distribution approximated by linear combinations of chi-square variables (e.g. Wilks (1946), Votaw (1948), Gleser and Olkin (1966, 1969), Olkin and Press (1969), Olkin (1972), Szatrowski (1976)). Similarly, for the nonnull distributions, the exact nonnull distributions are sometimes obtained (e.g. Khatri and Srivastava (1971)); in other cases, the standard delta method is used to obtain asymptotic nonnull distributions (e.g. Olkin and Press (1969), Szatrowski (1976)) and, occasionally, additional terms of the expansion are also given (e.g. Sugiura (1969), Sugiura and Fujikoshi (1969)).

Received September 1977; revised May 1978.

[1]Research partially supported by the National Science Foundation under contract MPS-7509450-000, an N.S.F. Competitive Fellowship and by a Rutgers University Summer Fellowship.

AMS 1970 *subject classifications.* 62H10, 62H15.

Key words and phrases. Asymptotic nonnull distribution, delta method, hypothesis testing, matrix derivatives, patterned means, patterned covariance matrices.

In this study, the asymptotic nonnull distribution is derived for a general form of nested hypotheses used in one and k-population testing problems. These results can be simplified when maximum likelihood estimates used in the likelihood ratio statistics have explicit representations under the null hypothesis. These asymptotic nonnull results are useful for obtaining approximate powers and sample sizes for the tests under consideration.

In Section Two the problem and hypotheses under consideration are described. Section Three contains results on the derivation of maximum likelihood estimates and likelihood ratio tests. The usual asymptotic chi-square distribution for the null distributions of the likelihood ratio test is also given together with some results on matrix representations and derivatives. Sections Four and Five contain the asymptotic nonnull results for the one and k-population hypotheses, respectively.

2. Hypotheses, likelihood ratio tests, asymptotic null distributions. Let $\mathbf{X}$ be a p component column vector with multivariate normal distribution such that the mean vector $\boldsymbol{\mu} = \mathcal{E}\mathbf{X}$ and covariance matrix $\boldsymbol{\Sigma} = \text{Cov}(\mathbf{X}) = \mathcal{E}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'$ have the linear structure considered by Anderson (1969, 1970, 1973). Specifically, $\boldsymbol{\mu} = \Sigma_1^r \mathbf{z}_j \beta_j = \mathbf{Z}\boldsymbol{\beta}$, $\mathbf{Z} = [\mathbf{z}_1, \cdots, \mathbf{z}_r]$, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_r)'$, $\boldsymbol{\beta} \in \mathfrak{R}^r$, where the $\mathbf{z}$'s are known, linearly independent column vectors and the β's are unknown scalars. The covariance matrix is given by $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\sigma}) = \Sigma_0^m \sigma_g \mathbf{G}_g$, $\boldsymbol{\sigma} = (\sigma_0, \cdots, \sigma_m)'$, where the $\mathbf{G}$'s are known, linearly independent symmetric matrices and the σ's are unknown scalars, such that $\boldsymbol{\sigma} \in \boldsymbol{\Theta}$, $\boldsymbol{\Theta} = \{\boldsymbol{\theta} \in \mathfrak{R}^{m+1} | \boldsymbol{\Sigma}(\boldsymbol{\theta}) > 0\}$, where $\boldsymbol{\Sigma} > 0$ denotes $\boldsymbol{\Sigma}$ positive definite. We assume that $\boldsymbol{\Theta}$ is nonempty so that there exists at least one value of $\boldsymbol{\sigma}$ that results in $\boldsymbol{\Sigma}(\boldsymbol{\sigma})$ being positive definite.

DEFINITION 1. Let $\mathbf{A}$ be a $p \times p$ symmetric matrix. $\langle \mathbf{A} \rangle$ is defined to be a column vector consisting of the upper triangle of elements of $\mathbf{A}$,

$$\langle \mathbf{A} \rangle = (a_{11}, a_{22}, \cdots, a_{pp}, a_{12}, \cdots, a_{1p}, a_{23}, \cdots, a_{p-1,p})'.$$

Using Definition 1, and defining $\mathbf{W} = [\langle \mathbf{G}_0 \rangle, \langle \mathbf{G}_1 \rangle, \cdots, \langle \mathbf{G}_m \rangle]$, we observe that $\langle \boldsymbol{\Sigma} \rangle = \mathbf{W}\boldsymbol{\sigma}$.

2.1. *The one population problem.* Suppose we observe independent, identically distributed observations $\mathbf{x}_1, \cdots, \mathbf{x}_N$ from a multivariate normal distribution with patterned mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Identify substructures of $\mathbf{Z}$, $\boldsymbol{\beta}$, $\mathbf{W}$ and $\boldsymbol{\sigma}$ by

$$\mathbf{Z} = [\mathbf{Z}_0 : \mathbf{Z}_1], \quad \boldsymbol{\beta} = (\boldsymbol{\beta}_0' : \boldsymbol{\beta}_1')', \quad \mathbf{W} = [\mathbf{W}_0 : \mathbf{W}_1], \quad \boldsymbol{\sigma} = (\boldsymbol{\sigma}_0' : \boldsymbol{\sigma}_1')', \tag{2.1}$$

where $\mathbf{Z}_0$ and $\boldsymbol{\beta}_0$ are $p \times r_0$ and $r_0 \times 1$ respectively and $\mathbf{W}_0$ and $\boldsymbol{\sigma}_0$ are $\frac{1}{2}p(p+1) \times (m_0 + 1)$ and $(m_0 + 1) \times 1$ respectively. At least one of the inequalities $r_0 \leqslant r$ and $m_0 \leqslant m$ is assumed to be strict. The problem is to test the null hypothesis $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}, \boldsymbol{\sigma}_1 = \mathbf{0}$ versus the alternative hypothesis H_1 which does not so restrict $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$.

The likelihood ratio statistic λ for testing these hypotheses is of the form

$$\lambda^{(2/N)} = |\tilde{\Sigma}_1|/|\tilde{\Sigma}_0| \tag{2.2}$$

where $\tilde{\Sigma}_0$ and $\tilde{\Sigma}_1$ are the maximum likelihood estimates of Σ under H_0 and H_1 respectively. Methods used to find these estimates are discussed in Section 3.1.

In general, the exact distribution of the likelihood ratio statistic is difficult to derive. Often no explicit form of the likelihood ratio statistic exists. The usual asymptotic chi-square distribution applies under the null hypothesis assumption, yielding $\lim_{N\to\infty}\mathcal{L}(-2\log\lambda) = \chi_f^2$, where $f = m + r - (m_0 + r_0)$. We reject the null hypothesis when $-2\log\lambda$ is too large. The standard delta method is used to derive the asymptotic nonnull distribution in Section Four.

2.2. *The k-population problem.* Suppose we observe random samples from k multivariate normal populations. Specifically, we have observations $\mathbf{x}_{dj}, j = 1, \cdots, p_d, d = 1, \cdots, k$ where $\mathcal{L}(\mathbf{x}_{dj}) = \mathfrak{N}(\boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d)$ and $N = \Sigma_1^k p_d$. We assume all means and covariances have the same patterns, i.e., $\boldsymbol{\mu}_d = \mathbf{Z}\boldsymbol{\beta}_d$, $\langle\boldsymbol{\Sigma}_d\rangle = \mathbf{W}\boldsymbol{\sigma}_d$, $d = 1, \cdots, k$. Three sets of hypotheses motivated by Votaw (1948) are investigated. $H_k(\text{MVC}|\text{mvc})$ is the hypothesis that $\boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_k$. $H_k(\text{VC}|\text{mvc})$ is the hypothesis that $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_k$. Both of these hypotheses are tested against $H_{k,A}$, the alternative hypothesis without the equality constraint on the means and covariances. A third hypothesis, $H_k(M|\text{mVC})$ involves testing $H_k(\text{MVC}|\text{mvc})$ against the alternative hypothesis $H_k(\text{VC}|\text{mvc})$.

The likelihood ratio statistic is of the form

$$\lambda^{2/N} = \left(\Pi_1^k|\tilde{\Sigma}_d|^{p_d/N}\right)/|\tilde{\Sigma}| \tag{2.3}$$

for testing the null hypotheses $H_k(\text{MVC}|\text{mvc})$ or $H_k(\text{VC}|\text{mvc})$, where $\tilde{\Sigma}_d$, $d = 1, \cdots, k$, are the maximum likelihood estimates found under the alternative hypothesis $H_{k,A}$ and $\tilde{\Sigma}$ is the maximum likelihood estimate of the common covariance matrix found under the null hypothesis $H_k(\text{MVC}|\text{mvc})$ or $H_k(\text{VC}|\text{mvc})$, respectively. For testing the null hypothesis $H_k(M|\text{mVC})$, the likelihood ratio statistic is of the form $\lambda^{(2/N)} = |\tilde{\Sigma}_1|/|\tilde{\Sigma}_0|$ where $\tilde{\Sigma}_0$ and $\tilde{\Sigma}_1$ are the maximum likelihood estimates of the common covariance matrix found under the assumptions of $H_k(\text{MVC}|\text{mvc})$ and $H_k(\text{VC}|\text{mvc})$, respectively. Methods used to find these estimates are given in Section 3.1.

In general, as in the one population case, the exact distribution of these likelihood ratio statistics is difficult to derive. The usual asymptotic chi-square distribution applies under the null hypothesis assumption, yielding $\lim_{N\to\infty}\mathcal{L}(-2\log\lambda) = \chi_f^2$, where it is assumed that $\lim_{N\to\infty}(p_d/N) = f_d$, $0 < f_d < 1$, $d = 1, \cdots, k$. The degrees of freedom f are $(r + m + 1)(k - 1)$ for testing $H_k(\text{MVC}|\text{mvc})$, $(m + 1)(k - 1)$ for testing $H_k(\text{mVC}|\text{mvc})$ and $r(k - 1)$ for testing $H_k(M|\text{mVC})$. We reject the null hypothesis when $-2\log\lambda$ is too large. The standard delta method is used to derive the asymptotic nonnull distributions in Section Five.

3. Results used in nonnull distribution derivations.

3.1. *Maximum likelihood estimates.* Let $[d_{gh}]$ denote a matrix whose g, h element is d_{gh} and let (d_g) denote a column vector whose gth element is d_g. Anderson (1973) derives the likelihood equations for the one-population problem (Section 2.1),

$$\hat{\boldsymbol{\beta}} = (\mathbf{Z}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\hat{\boldsymbol{\Sigma}}^{-1}\bar{\mathbf{x}}, \tag{3.1}$$
$$\hat{\boldsymbol{\sigma}} = \left[\operatorname{tr} \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{G}_g\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{G}_h\right]^{-1}\left(\operatorname{tr} \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{G}_g\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{C}\right),$$
$$\bar{\mathbf{x}} = (1/N)\Sigma_{\alpha=1}^{N}\mathbf{x}_\alpha, \ \mathbf{C} = \mathbf{A} + (\bar{\mathbf{x}} - \hat{\boldsymbol{\mu}})(\bar{\mathbf{x}} - \hat{\boldsymbol{\mu}})',$$
$$\mathbf{A} = (1/N)\Sigma_{\alpha=1}^{N}(\mathbf{x}_\alpha - \bar{\mathbf{x}})(\mathbf{x}_\alpha - \bar{\mathbf{x}})',$$

where $[\operatorname{tr} \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{G}_g\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{G}_h]$ is an $(m + 1) \times (m + 1)$ matrix, $(\operatorname{tr} \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{G}_g\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{C})$ is an $(m + 1) \times 1$ column vector, and $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$ are $r \times 1$ and $(m + 1) \times 1$ column vectors of unknowns respectively. Anderson points out that the likelihood equations written in this form suggest an iterative scheme wherein from an initial estimate of $\boldsymbol{\Sigma}$, $\hat{\boldsymbol{\Sigma}}$, one can solve the linear equations in $\boldsymbol{\beta}$, compute $\mathbf{C}(\hat{\boldsymbol{\mu}} = \mathbf{Z}\hat{\boldsymbol{\beta}})$, and then solve the linear equations in $\boldsymbol{\sigma}$ to yield the next estimate of $\boldsymbol{\Sigma}$. Szatrowski (1980) shows that if the maximum likelihood estimates have explicit representations, then they are given by $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\sigma}}$ in (3.1) after one iteration from any allowable starting point $\hat{\boldsymbol{\Sigma}}$, i.e., any starting point $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\sigma}}) > 0$.

We say that the maximum likelihood estimate of the mean has an explicit representation if the maximum likelihood estimate for $\boldsymbol{\beta}$, $\tilde{\boldsymbol{\beta}}$, can be expressed by $\tilde{\boldsymbol{\beta}} = \mathbf{A}\bar{\mathbf{x}}$, where $\mathbf{A}$ is a function only of $\mathbf{Z}$ and $\mathbf{W}$. Similarly, the maximum likelihood estimate of $\boldsymbol{\sigma}$, $\tilde{\boldsymbol{\sigma}}$, has an explicit representation if and only if the mean does and $\tilde{\boldsymbol{\sigma}}$ can be expressed as $\tilde{\boldsymbol{\sigma}} = \mathbf{B}\langle\mathbf{C}\rangle$, where $\mathbf{B}$ is a function only of $\mathbf{W}$; $\mathbf{C}$ is given in (3.1).

For the one-population hypothesis, we use Anderson's algorithm for finding the maximum likelihood estimates under the alternative hypothesis. Under the null hypothesis, we repeat this procedure using the restricted forms of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. For the k-population hypotheses, one must pool some of the data before using equation (3.1). Under $H_k(\text{MVC}|\text{mvc})$, we use equation (3.1) with $\bar{\mathbf{x}} = \Sigma_1^k(p_d/N)\bar{\mathbf{x}}_d$, the grand mean, and $\mathbf{C}$ of the form

$$\mathbf{C} = \Sigma_1^k(p_d/N)\{\mathbf{A}_d + (\bar{\mathbf{x}}_d - \bar{\mathbf{x}})(\bar{\mathbf{x}}_d - \bar{\mathbf{x}})'\} + (\bar{\mathbf{x}} - \hat{\boldsymbol{\mu}})(\bar{\mathbf{x}} - \hat{\boldsymbol{\mu}})', \tag{3.2}$$
$$\mathbf{A}_d = (1/p_d)\Sigma_{j=1}^{p_d}(\mathbf{x}_{dj} - \bar{\mathbf{x}}_d)(\bar{\mathbf{x}}_{dj} - \bar{\mathbf{x}}_d)'.$$

Under $H_k(\text{mVC}|\text{mvc})$ we use $\mathbf{C}$ of the form

$$\mathbf{C} = \Sigma_1^k(p_d/N)\left[\mathbf{A}_d + (\bar{\mathbf{x}}_d - \hat{\boldsymbol{\mu}}_d)(\bar{\mathbf{x}}_d - \hat{\boldsymbol{\mu}}_d)'\right] = \Sigma_1^k(p_d/N)\mathbf{C}_d, \tag{3.3}$$

where the $\hat{\boldsymbol{\mu}}_d$'s are calculated using the $\bar{\mathbf{x}}_d$'s. Under $H_{k,A}$, we obtain maximum likelihood estimates for each of the k populations separately using equation (3.1).

3.2. *The standard delta method and derivative results.* To derive the asymptotic nonnull distribution, we use the standard delta method in the following form, given in Anderson (1958, page 76).

THEOREM 1. *Let* $\mathbf{Q}(n)$ *be an m-component random column vector and* $\mathbf{b}$ *a fixed vector. Assume* $n^{\frac{1}{2}}(\mathbf{Q}(n) - \mathbf{b})$ *has the asymptotic multivariate normal distribution* $\mathfrak{N}(\mathbf{0}, \mathbf{T})$. *Let* $w = f(\mathbf{q})$ *be a scalar function of the vector* $\mathbf{q}$ *with first and second derivatives existing in the neighborhood of* $\mathbf{q} = \mathbf{b}$. *Let* $(\partial f/\partial q_i)|_{\mathbf{q}=\mathbf{b}}$ *be the* ith *component of* $\boldsymbol{\phi}_b$. *Then the limiting distribution of* $n^{\frac{1}{2}}[f(\mathbf{Q}(n)) - f(\mathbf{b})]$ *is* $\mathfrak{N}(\mathbf{0}, \boldsymbol{\phi}_b'\mathbf{T}\boldsymbol{\phi}_b)$.

In our problems, $f(\mathbf{q}) = -(2/N)\log\lambda$, where $\mathbf{Q}(n) = (\langle\mathbf{A}\rangle', \bar{\mathbf{x}}')'$ in the one-population problems and $\mathbf{Q}(n) = (\langle\mathbf{A}_1\rangle', \cdots, \langle\mathbf{A}_k\rangle', \bar{\mathbf{x}}_1', \cdots, \bar{\mathbf{x}}_k')'$ in the k-population problems. The following is a well-known result on the asymptotic distribution of $\bar{\mathbf{x}}$ and $\mathbf{A}$ in the one-population problem.

THEOREM 2. $N^{\frac{1}{2}}\{(\bar{\mathbf{x}}', \langle\mathbf{A}\rangle') - (\boldsymbol{\mu}', \langle\boldsymbol{\Sigma}\rangle')\}'$ *is asymptotically normally distributed according to* $\mathfrak{N}(\mathbf{0}, \mathbf{T})$, *where* $\mathbf{T} = \text{diag}(\boldsymbol{\Sigma}, \boldsymbol{\Phi})$, *with* $\boldsymbol{\Phi}$ *given in Definition* 2 *below.*

DEFINITION 2. (Anderson, 1969). Let $\boldsymbol{\Phi}$ be a $\{p(p+1)/2\} \times \{p(p+1)/2\}$ symmetric matrix with elements $\boldsymbol{\Phi} \equiv \boldsymbol{\Phi}(\boldsymbol{\Sigma}) = (\phi_{ij,kl}) = (\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk})$, $i \leqslant j$, $k \leqslant l$, where σ_{ij} is the ij element of $\boldsymbol{\Sigma}$. The notation $\phi_{ij,kl}$ represents the element of $\boldsymbol{\Phi}$ with row in the same position as the element a_{ij} in $\langle\mathbf{A}\rangle$, where $\mathbf{A}$ is $p \times p$ symmetric, and column in the same position as a_{kl} in $\langle\mathbf{A}\rangle'$.

We observe that if the $p \times p$ matrix $\mathbf{W} > 0$ has a Wishart distribution with parameters $\boldsymbol{\Sigma} > 0$ and n ($\mathcal{L}(\mathbf{W}) = \mathfrak{W}(\boldsymbol{\Sigma}, n)$), then $n\boldsymbol{\Phi}(\boldsymbol{\Sigma}) = \text{Cov}(\langle\mathbf{W}\rangle)$. The following theorems are generalizations of the theorem in Anderson (1969, page 61).

THEOREM 3. *If* $\mathbf{E}$ *and* $\mathbf{F}$ *are* $p \times p$ *symmetric matrices, then*

$$\langle\mathbf{E}\rangle'\boldsymbol{\Phi}^{-1}(\boldsymbol{\Sigma})\langle\mathbf{F}\rangle = \tfrac{1}{2}\text{tr}\boldsymbol{\Sigma}^{-1}\mathbf{E}\boldsymbol{\Sigma}^{-1}\mathbf{F}. \tag{3.4}$$

PROOF. The proof is a straightforward extension of Anderson's (1969) proof. Techniques used in this proof appear also in the proof of Theorem 4 below. □

Examples of possible $\mathbf{E}$ and/or $\mathbf{F}$ matrices to be used in Theorem 3 are $\boldsymbol{\Sigma}$; $\mathbf{G}_f$, $f = 0, \cdots, m$; and $\mathbf{C}$, $\mathbf{A}$ and $\mathbf{C} - \mathbf{A}$ defined in equation (3.1).

Using Theorem 3, we obtain the following identities and an alternative form of the likelihood equation (3.1) ($\hat{\boldsymbol{\Phi}} \equiv \boldsymbol{\Phi}(\hat{\boldsymbol{\Sigma}})$),

$$\langle\mathbf{G}_h\rangle'\hat{\boldsymbol{\Phi}}^{-1}\langle\mathbf{C}\rangle = \tfrac{1}{2}\text{tr}\,\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{G}_h\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{C}, \qquad h = 0, 1, \cdots, m, \tag{3.5}$$

$$\langle\mathbf{G}_h\rangle'\hat{\boldsymbol{\Phi}}^{-1}\langle\mathbf{G}_f\rangle = \tfrac{1}{2}\text{tr}\,\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{G}_h\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{G}_f, \qquad f, h = 0, \cdots, m, \tag{3.6}$$

$$\hat{\boldsymbol{\sigma}} = (\mathbf{W}'\hat{\boldsymbol{\Phi}}^{-1}\mathbf{W})^{-1}\mathbf{W}'\hat{\boldsymbol{\Phi}}^{-1}\langle\mathbf{C}\rangle \equiv \hat{\mathbf{T}}\langle\mathbf{C}\rangle. \tag{3.7}$$

Before stating and proving Theorem 4, we state a matrix theory result used in the proofs of Theorems 3 and 4.

LEMMA 1. *If* $\mathbf{R}$ *is a nonsingular* $p \times p$ *matrix, then there exists a nonsingular matrix* $\mathbf{B}$ *such that* $\langle\mathbf{RSR}'\rangle = \mathbf{B}\langle\mathbf{S}\rangle$ *for any* $p \times p$ *symmetric matrix* $\mathbf{S}$. *If, in addition,* $\mathbf{S} > 0$, $\boldsymbol{\Phi}(\mathbf{RSR}') = \mathbf{B}\boldsymbol{\Phi}(\mathbf{S})\mathbf{B}'$.

PROOF. Let $\mathbf{H}_{ij}$ be a symmetric matrix of zeroes with a one in the ij and ji positions. Any $p \times p$ symmetric matrix can be expressed as a unique linear combination of the $p(p+1)/2$ symmetric matrices, $\{\mathbf{H}_{ij}, i \leqslant j = 1, \cdots, p\}$. Define $\mathbf{B}$ by the set of equations $\langle \mathbf{R}\mathbf{H}_{ij}\mathbf{R}' \rangle = \mathbf{B}\langle \mathbf{H}_{ij} \rangle$, $i \leqslant j = 1, \cdots, p$. By linearity, with $\mathbf{S} = \Sigma_{i \leqslant j} s_{ij}\mathbf{H}_{ij}$, we see $\langle \mathbf{RSR}' \rangle = \mathbf{B}\langle \mathbf{S} \rangle$ for all symmetric $\mathbf{S}$. If $\mathbf{B}$ were singular, there would exist a vector $\langle \mathbf{S} \rangle \neq 0$ such that $\mathbf{B}\langle \mathbf{S} \rangle = 0$. Since this contradicts the fact that $\langle \mathbf{RSR}' \rangle = 0$ for symmetric $\mathbf{S}$ if and only if $\mathbf{S} = 0$, we conclude $\mathbf{B}$ is nonsingular. The second part of the lemma follows by noting if $\mathcal{L}(\mathbf{W}) = \mathcal{W}(\mathbf{S}, n)$, then $\mathcal{L}(\mathbf{RWR}') = \mathcal{W}(\mathbf{RSR}', n)$ and $n\mathbf{\Phi}(\mathbf{RSR}') = \mathrm{Cov}(\langle \mathbf{RWR}' \rangle)$ $= \mathrm{Cov}(\mathbf{B}\langle \mathbf{W} \rangle) = \mathbf{B}\{\mathrm{Cov}(\langle \mathbf{W} \rangle)\}\mathbf{B}' = n\mathbf{B\Phi(S)B}'$. □

THEOREM 4. *If $\mathbf{\Sigma}_1$, $\mathbf{\Sigma}_2$, and $\mathbf{\Sigma}_3$ are $p \times p$ positive definite covariance matrices, then, with $\mathbf{\Phi}_i \equiv \mathbf{\Phi}(\mathbf{\Sigma}_i)$, $i = 1, 2, 3$,*

$$\langle \mathbf{\Sigma}_1 \rangle' \mathbf{\Phi}_1^{-1} \mathbf{\Phi}_3 \mathbf{\Phi}_2^{-1} \langle \mathbf{\Sigma}_2 \rangle = \tfrac{1}{2} \mathrm{tr}\, \mathbf{\Sigma}_1^{-1} \mathbf{\Sigma}_3 \mathbf{\Sigma}_2^{-1} \mathbf{\Sigma}_3. \tag{3.8}$$

PROOF. Using Lemma 1, we see that for $\mathbf{R}$ nonsingular, (3.8) is invariant under the transformation $(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \mathbf{\Sigma}_3) \rightarrow (\mathbf{R\Sigma}_1\mathbf{R}', \mathbf{R\Sigma}_2\mathbf{R}', \mathbf{R\Sigma}_3\mathbf{R}')$. Since the $\mathbf{\Sigma}$'s are positive definite, we can choose $\mathbf{R}$ so $\mathbf{\Sigma}_2 = \mathbf{I}$, and $\mathbf{\Sigma}_1 = \mathrm{diag}(a_1, \cdots, a_p)$. Decompose $\mathbf{\Phi}$ into

$$\mathbf{\Phi} = \begin{bmatrix} \mathbf{\Phi}_{11} & \mathbf{\Phi}_{12} \\ \mathbf{\Phi}_{21} & \mathbf{\Phi}_{22} \end{bmatrix},$$

where $\mathbf{\Phi}_{11}$ is $p \times p$ and observe that $\mathbf{\Phi}(\mathbf{I}) = \mathrm{diag}(2\mathbf{I}_p, \mathbf{I}_{p(p-1)/2})$. Also, for $\mathbf{\Sigma}_1$ in this diagonal form, $(\mathbf{\Phi}_1)_{12} = (\mathbf{\Phi}_1)'_{21} = 0$ and $(\mathbf{\Phi}_1)_{11} = \mathrm{diag}(2a_1^2, \cdots, 2a_p^2)$. Substitution into equation (3.8) yields

$$(a_1, \cdots, a_p)(\mathbf{\Phi}_1)_{11}^{-1}(\mathbf{\Phi}_3)_{11}(2\mathbf{I}_p)^{-1}(1, \cdots, 1)' = \tfrac{1}{2}\mathrm{tr}\left\{\left[\mathrm{diag}(a_1, \cdots a_p)\right]^{-1}\mathbf{\Sigma}_3\mathbf{\Sigma}_3\right\}. \tag{3.9}$$

Multiplication yields

$$\tfrac{1}{4}\left(a_1^{-1}, \cdots, a_p^{-1}\right)(\mathbf{\Phi}_3)_{11}(1, \cdots, 1)' = \tfrac{1}{2}\mathrm{tr}\left\{\left[\mathrm{diag}(a_1, \cdots, a_p)\right]^{-1}\mathbf{\Sigma}_3\mathbf{\Sigma}_3\right\}, \tag{3.10}$$

which, upon further simplification, yields the result

$$\tfrac{1}{2}\Sigma_{i,j=1}^{p}\left(\sigma_{ij}^2/a_i\right) = \tfrac{1}{2}\Sigma_{i,j=1}^{p}\left(\sigma_{ij}^2/a_i\right), \tag{3.11}$$

after inspecting the form of $(\mathbf{\Phi}_3)_{11}$. □

Finally, several well-known results on matrix derivatives are stated without proof.

LEMMA 2. *If $\mathbf{\Sigma}$ is a patterned covariance matrix, then*

$$\frac{\partial}{\partial \sigma_g} \log |\mathbf{\Sigma}| = \mathrm{tr}\, \mathbf{\Sigma}^{-1}\mathbf{G}_g = 2\langle \mathbf{\Sigma} \rangle' \mathbf{\Phi}^{-1}(\mathbf{\Sigma})\langle \mathbf{G}_g \rangle. \tag{3.12}$$

LEMMA 3. *If* $\mathbf{Y}$ *is a matrix function of a matrix* $\mathbf{X}$, *then*

$$\frac{\partial}{\partial x_{\alpha\beta}}\operatorname{tr}(\mathbf{AYB}) = \operatorname{tr}\mathbf{A}\frac{\partial \mathbf{Y}}{\partial x_{\alpha\beta}}\mathbf{B}, \tag{3.13}$$

$$\frac{\partial \mathbf{Y}^{-1}}{\partial x_{\alpha\beta}} = -\mathbf{Y}^{-1}\left(\frac{\partial \mathbf{Y}}{\partial x_{\alpha\beta}}\right)\mathbf{Y}^{-1}. \tag{3.14}$$

4. One-population asymptotic nonnull distribution. In this section we use the results from Section Three to prove the following theorem.

THEOREM 5. *For the one-population likelihood ratio statistic given by equation* (2.2), *the asymptotic nonnull distribution is given by*

$$\lim_{N\to\infty}\mathcal{L}\left[N^{\frac{1}{2}}\{-(2/N)\log\lambda - \log(|\boldsymbol{\Sigma}_0|/|\boldsymbol{\Sigma}^*|)\}\right] = \mathfrak{N}(0, v_\infty), \tag{4.1}$$

$$\begin{aligned} v_\infty = 2p &- 8\langle\boldsymbol{\Sigma}_0\rangle'\boldsymbol{\Phi}_0^{-1}\mathbf{W}_0(\mathbf{I}-\mathbf{M}_0)^{-1}\mathbf{T}_0\langle\boldsymbol{\Sigma}^*\rangle \\ &+4\langle\boldsymbol{\Sigma}_0\rangle'\boldsymbol{\Phi}_0^{-1}\mathbf{W}_0(\mathbf{I}-\mathbf{M}_0)^{-1}\mathbf{T}_0(\boldsymbol{\Phi}^*+\mathbf{F})\mathbf{T}_0'(\mathbf{I}-\mathbf{M}_0')^{-1}\mathbf{W}_0'\boldsymbol{\Phi}_0^{-1}\langle\boldsymbol{\Sigma}_0\rangle, \end{aligned} \tag{4.2}$$

with

$$\mathbf{M}_0 = \left[\operatorname{tr}\boldsymbol{\Sigma}_0^{-1}\mathbf{G}_g\boldsymbol{\Sigma}_0^{-1}\mathbf{G}_h\right]^{-1}\left[\operatorname{tr}\boldsymbol{\Sigma}_0^{-1}(\mathbf{F}_g^*+\mathbf{H}_g^*)\boldsymbol{\Sigma}_0^{-1}\mathbf{G}_h\right],$$

$$\mathbf{T}_0 = (\mathbf{W}_0'\boldsymbol{\Phi}_0^{-1}\mathbf{W}_0)^{-1}\mathbf{W}_0'\boldsymbol{\Phi}_0^{-1},\ \mathbf{R}_0 = \mathbf{Z}_0(\mathbf{Z}_0'\boldsymbol{\Sigma}_0^{-1}\mathbf{Z}_0)^{-1}\mathbf{Z}_0',\ \mathbf{B}_0 = (\boldsymbol{\mu}^*-\boldsymbol{\mu}_0)(\boldsymbol{\mu}^*-\boldsymbol{\mu}_0)',$$

$$\mathbf{F}_g^* = \mathbf{B}_0\boldsymbol{\Sigma}_0^{-1}\mathbf{G}_g\boldsymbol{\Sigma}_0^{-1}\mathbf{R}_0 + \mathbf{R}_0\boldsymbol{\Sigma}_0^{-1}\mathbf{G}_g\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_0,$$

$$\mathbf{H}_g^* = \mathbf{G}_g\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\Sigma}_0-\mathbf{C}^*) + (\boldsymbol{\Sigma}_0-\mathbf{C}^*)\boldsymbol{\Sigma}_0^{-1}\mathbf{G}_g,$$

$$\mathbf{C}^* = \boldsymbol{\Sigma}^*+\mathbf{B}_0,\ \mathbf{F} = \Sigma_{i,j=1}^{p}\boldsymbol{\Sigma}_{ij}^*\mathbf{b}_i\mathbf{b}_j',$$

$$\mathbf{b}_i = \langle(\mathbf{I}-\mathbf{R}_0\boldsymbol{\Sigma}_0^{-1})\mathbf{e}_i(\boldsymbol{\mu}^*-\boldsymbol{\mu}_0)' + (\boldsymbol{\mu}^*-\boldsymbol{\mu}_0)\mathbf{e}_i'(\mathbf{I}-\mathbf{R}_0\boldsymbol{\Sigma}_0^{-1})'\rangle,$$

where $\mathbf{e}_i$ *is a column vector of zeroes with one in the* ith *position,* $(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ *is the value assumed under the alternative hypothesis and not assumed under the null hypothesis,* $\boldsymbol{\Phi}_0 \equiv \boldsymbol{\Phi}(\boldsymbol{\Sigma}_0)$, $\boldsymbol{\Phi}^* \equiv \boldsymbol{\Phi}(\boldsymbol{\Sigma}^*)$, *and* $\boldsymbol{\mu}_0$ *and* $\boldsymbol{\Sigma}_0$ *are the "maximum likelihood estimates" under the null hypothesis derived from equations* (3.1) *with* $\bar{\mathbf{x}}$ *replaced by* $\boldsymbol{\mu}^*$, $\mathbf{A}$ *replaced by* $\boldsymbol{\Sigma}^*$ *and* $\mathbf{W}$ *and* $\mathbf{Z}$ *replaced by* $\mathbf{W}_0$ *and* $\mathbf{Z}_0$ *respectively. Note* $\boldsymbol{\mu}_0$ *and* $\boldsymbol{\Sigma}_0$ *are not statistics.*

Under the additional assumption that the maximum likelihood estimates have explicit representations under the null hypothesis, the form of the asymptotic variance simplifies, becoming

$$v_\infty = 2\left\{\operatorname{tr}(\mathbf{I}-\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}^*)^2 + 2(\boldsymbol{\mu}^*-\boldsymbol{\mu}_0)'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\mu}^*-\boldsymbol{\mu}_0)\right\}. \tag{4.3}$$

REMARK. The form of v_∞ in (4.3) is of the form obtained by Sugiura (1969) for testing the null hypothesis $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$ and $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus the alternative hypothesis $\boldsymbol{\Sigma} \neq \boldsymbol{\Sigma}_0$ or $\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$.

PROOF. Let $f = f(\langle \mathbf{A} \rangle, \bar{\mathbf{x}}) = -(2/N)\log \lambda = \log(|\tilde{\boldsymbol{\Sigma}}_0|/|\tilde{\boldsymbol{\Sigma}}_1|)$. We need to find the partial derivatives of $f(\langle \mathbf{A} \rangle, \bar{\mathbf{x}})$ with respect to $\langle \mathbf{A} \rangle$ and $\bar{\mathbf{x}}$ to use the standard delta method given in Theorem 1. Since $\langle \mathbf{A} \rangle$ and $\bar{\mathbf{x}}$ are independent by the normality assumption, we may express $v_\infty = v_{\langle \mathbf{A} \rangle} + v_{\bar{\mathbf{x}}}$. Here $v_{\langle \mathbf{A} \rangle} = \boldsymbol{\phi}'_{\langle \mathbf{A} \rangle}\boldsymbol{\Phi}^*\boldsymbol{\phi}_{\langle \mathbf{A} \rangle}$ and $v_{\bar{\mathbf{x}}} = \boldsymbol{\phi}'_{\bar{\mathbf{x}}}\boldsymbol{\Sigma}^*\boldsymbol{\phi}_{\bar{\mathbf{x}}}$. Let y be an element in $(\langle \mathbf{A} \rangle', \bar{\mathbf{x}}')$. Using Lemma 2 and the chain rule, we find

$$(4.4)\qquad \frac{\partial f}{\partial y} = \Sigma_{s=0}^{m_0} \frac{\partial \log |\tilde{\boldsymbol{\Sigma}}_0|}{\partial \tilde{\sigma}_s} \frac{\partial \tilde{\sigma}_s^{(0)}}{\partial y} - \Sigma_{s=0}^{m} \frac{\partial \log |\tilde{\boldsymbol{\Sigma}}_1|}{\partial \tilde{\sigma}_s} \frac{\partial \tilde{\sigma}_s^{(1)}}{\partial y}$$

$$= \Sigma_{s=0}^{m_0} \operatorname{tr}(\tilde{\boldsymbol{\Sigma}}_0^{-1}\mathbf{G}_s) \frac{\partial \tilde{\sigma}_s^{(0)}}{\partial y} - \Sigma_{s=0}^{m} \operatorname{tr}(\tilde{\boldsymbol{\Sigma}}_1^{-1}\mathbf{G}_s) \frac{\partial \tilde{\sigma}_s^{(1)}}{\partial y},$$

where $\tilde{\boldsymbol{\sigma}}^{(i)}$ are the maximum likelihood estimates of $\boldsymbol{\sigma}$ under H_i, $i = 0, 1$, derived from equation (3.1).

Note that the $\tilde{\boldsymbol{\sigma}}^{(i)}$ are functions of $\bar{\mathbf{x}}$ and $\mathbf{A}$. In general, it is not possible to obtain an explicit form for the $\tilde{\boldsymbol{\sigma}}^{(i)}$ or $\partial\tilde{\boldsymbol{\sigma}}^{(i)}/\partial y$. Instead we find $\partial\hat{\boldsymbol{\sigma}}^{(i)}/\partial y$ using implicit differentiation on the likelihood equations (3.1) (substituting $\mathbf{W}_0$, $\mathbf{Z}_0$, $\boldsymbol{\sigma}_0$ and $\boldsymbol{\beta}_0$ for $\mathbf{W}$, $\mathbf{Z}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\beta}$ respectively when $i = 0$.) Due to the linear structure of the patterned mean and covariance matrix, implicit differentiation yields a set of linear equations which can be explicitly solved for $\partial\hat{\boldsymbol{\sigma}}^{(i)}/\partial y$. We then can find $\partial\tilde{\boldsymbol{\sigma}}^{(i)}/\partial y$ evaluated at $(\bar{\mathbf{x}}, \mathbf{A}) = (\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$, $i = 0, 1$, to be $\partial\hat{\boldsymbol{\sigma}}^{(0)}/\partial y$ evaluated at $(\bar{\mathbf{x}}, \mathbf{A}) = (\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ and $\partial\hat{\boldsymbol{\sigma}}^{(1)}/\partial y$ evaluated at $(\bar{\mathbf{x}}, \mathbf{A}) = (\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$, respectively.

Here $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ are "maximum likelihood estimates" of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ obtained from the null hypothesis form of the likelihood equations (3.1) with $(\bar{\mathbf{x}}, \mathbf{A})$ replaced by $(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$. Note $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ are *not* statistics. If there are several roots of the likelihood equations, we use the root which maximizes the likelihood function. Since $(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ is a point in the alternative hypothesis region, it is the maximum likelihood estimate of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ under the alternative hypothesis.

To evaluate $\partial\hat{\sigma}_s^{(i)}/\partial y$, we rewrite equation (3.1) (dropping the superscript)

$$(4.5)\qquad \hat{\sigma}_s = \mathbf{e}_s'\hat{\boldsymbol{\sigma}} = \mathbf{e}_s'\mathbf{Y}_1^{-1}\mathbf{y}_2 = \mathbf{e}_s'\mathbf{T}\langle \mathbf{C} \rangle;\ \hat{\boldsymbol{\mu}} = \mathbf{Z}\hat{\boldsymbol{\beta}} = \mathbf{R}\hat{\boldsymbol{\Sigma}}^{-1}\bar{\mathbf{x}},$$

where we define $\mathbf{Y}_1$, $\mathbf{y}_2$, $\mathbf{T}$ and $\mathbf{R}$ by

$$(4.6)\qquad \mathbf{Y} = \left[\operatorname{tr} \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\sigma}})\mathbf{G}_g\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\sigma}})\mathbf{G}_h\right], \qquad \mathbf{y}_2 = \left(\operatorname{tr} \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\sigma}})\mathbf{G}_g\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\sigma}})\mathbf{C}(\hat{\boldsymbol{\sigma}})\right),$$

$$\mathbf{T} = (\mathbf{W}'\boldsymbol{\Phi}^{-1}(\hat{\boldsymbol{\Sigma}})\mathbf{W})^{-1}\mathbf{W}'\boldsymbol{\Phi}^{-1}(\hat{\boldsymbol{\Sigma}})', \qquad \mathbf{R} = \mathbf{Z}(\mathbf{Z}'\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\sigma}})\mathbf{Z})^{-1}\mathbf{Z}'.$$

The form of the derivative $\partial\hat{\sigma}_s/\partial y$ is given by

$$(4.7)\qquad \frac{\partial \hat{\sigma}_s}{\partial y} = b_{sy} + \Sigma_\pi a_{sy\pi}\frac{\partial \hat{\sigma}_\pi}{\partial y}.$$

These linear equations can be solved for the explicit form of $\partial\hat{\sigma}_s/\partial y$. After taking the appropriate derivatives, using Lemma 3 and the chain rule, the form of $\mathbf{C}$ given in (3.1), of $\hat{\boldsymbol{\mu}}$ in (4.5) and substitution of asymptotic forms of $(\bar{\mathbf{x}}, \mathbf{A})$, we are able to solve the linear equations (4.7) to yield

$$(4.8) \qquad \frac{\partial\hat{\sigma}}{\partial y} = (\mathbf{I} - \mathbf{M})^{-1}\mathbf{b}_y, \qquad \mathbf{M} = \mathbf{Y}_1^{-1}\left[\operatorname{tr} \hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{F}_g + \mathbf{H}_g)\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{G}_h\right],$$

$$\mathbf{F}_g = \mathbf{B}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{G}_g\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{R} + \mathbf{R}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{G}_g\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{B}, \qquad \mathbf{B} = (\bar{\mathbf{x}} - \hat{\boldsymbol{\mu}})(\bar{\mathbf{x}} - \hat{\boldsymbol{\mu}})',$$

$$\mathbf{H}_g = \mathbf{G}_g\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\Sigma}} - \mathbf{C}) + (\hat{\boldsymbol{\Sigma}} - \mathbf{C})\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{G}_g.$$

We now look at $\partial\sigma_s/\partial y = \mathbf{e}_s'(\partial\sigma/\partial y)$ given by (4.8) under the null and alternative hypotheses evaluated at $\mathbf{A} = \boldsymbol{\Sigma}^*$ and $\bar{\mathbf{x}} = \boldsymbol{\mu}^*$.

Under the alternative hypothesis, $\mathbf{M} = \mathbf{0}$ since $\tilde{\boldsymbol{\Sigma}} = \mathbf{C} = \boldsymbol{\Sigma}^*$ and $\tilde{\boldsymbol{\mu}} = \bar{\mathbf{x}} = \boldsymbol{\mu}^*$ yielding

$$(4.9) \quad \left.\frac{\partial\sigma_s^{(1)}}{\partial\langle\mathbf{A}\rangle}\right|_{(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)} = \mathbf{T}^{*\prime}\mathbf{e}_s, \quad \left.\frac{\partial\sigma_s^{(1)}}{\partial\bar{x}_i}\right|_{(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)} = 0, \mathbf{T}^* = (\mathbf{W}'\boldsymbol{\Phi}^{*-1}\mathbf{W})^{-1}\mathbf{W}'\boldsymbol{\Phi}^{-1}.$$

Under the null hypothesis, we find $\partial\sigma_s^{(0)}/\partial\langle\mathbf{A}\rangle$ and $\partial\sigma_s^{(0)}/\partial\bar{x}_i$ to be

$$(4.10) \qquad \left.\frac{\partial\sigma_s^{(0)}}{\partial\langle\mathbf{A}\rangle}\right|_{(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)} = \mathbf{T}_0'(\mathbf{I} - \mathbf{M}_0')^{-1}\mathbf{e}_s,$$

$$(4.11) \qquad \left.\frac{\partial\sigma_s^{(0)}}{\partial\bar{x}_i}\right|_{(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)} = \mathbf{e}_s'(\mathbf{I} - \mathbf{M}_0)^{-1}\mathbf{T}_0\langle(\mathbf{I} - \mathbf{R}_0\boldsymbol{\Sigma}_0^{-1})\mathbf{e}_i(\boldsymbol{\mu}^* - \boldsymbol{\mu}_0)' + (\boldsymbol{\mu}^* - \boldsymbol{\mu}_0)\mathbf{e}_i'(\mathbf{I} - \mathbf{R}_0\boldsymbol{\Sigma}_0^{-1})'\rangle,$$

where $\mathbf{M}_0$, $\mathbf{T}_0$ and $\mathbf{R}_0$ are given in (4.2).

Let ϕ_{0y} and ϕ_{1y} be the first and second terms of $\partial f/\partial y$ in equation (4.4) evaluated at $\mathbf{A} = \boldsymbol{\Sigma}^*$ and $\bar{\mathbf{x}} = \boldsymbol{\mu}^*$. Then $\phi_{\langle\mathbf{A}\rangle} = \phi_{0\langle\mathbf{A}\rangle} - \phi_{1\langle\mathbf{A}\rangle}$ and $v_{\langle\mathbf{A}\rangle}$ can be expressed as

$$(4.12) \qquad v_{\langle\mathbf{A}\rangle} = \phi_{0\langle\mathbf{A}\rangle}'\boldsymbol{\Phi}^*\phi_{0\langle\mathbf{A}\rangle} + \phi_{1\langle\mathbf{A}\rangle}'\boldsymbol{\Phi}^*\phi_{1\langle\mathbf{A}\rangle} - 2\boldsymbol{\Phi}_{0\langle\mathbf{A}\rangle}'\boldsymbol{\Phi}^*\phi_{1\langle\mathbf{A}\rangle}.$$

Using Lemma 2 and Theorem 3, these terms can be evaluated yielding

$$(4.13) \quad \phi_{0\langle\mathbf{A}\rangle}'\boldsymbol{\Phi}^*\phi_{0\langle\mathbf{A}\rangle} = 4\langle\boldsymbol{\Sigma}_0\rangle'\boldsymbol{\Phi}_0^{-1}\mathbf{W}_0(\mathbf{I} - \mathbf{M}_0)^{-1}\mathbf{T}_0\boldsymbol{\Phi}^*\mathbf{T}_0'(\mathbf{I} - \mathbf{M}_0')^{-1}\mathbf{W}_0'\boldsymbol{\Phi}_0^{-1}\langle\boldsymbol{\Sigma}_0\rangle,$$

$$\phi_{1\langle\mathbf{A}\rangle}'\boldsymbol{\Phi}^*\phi_{1\langle\mathbf{A}\rangle} = 2p; \ \phi_{0\langle\mathbf{A}\rangle}'\boldsymbol{\Phi}^*\phi_{1\langle\mathbf{A}\rangle} = 4\langle\boldsymbol{\Sigma}_0\rangle'\boldsymbol{\Phi}_0^{-1}\mathbf{W}_0(\mathbf{I} - \mathbf{M})^{-1}\mathbf{T}_0\langle\boldsymbol{\Sigma}^*\rangle.$$

The variance associated with the mean is given by $v_{\bar{\mathbf{x}}} = \Sigma_{i,j=1}^p \Sigma_{ij}^*\phi_i\phi_j$ where ϕ_i is the value of $\partial f/\partial\bar{x}_i$ evaluated at $\mathbf{A} = \boldsymbol{\Sigma}^*$ and $\bar{\mathbf{x}} = \boldsymbol{\mu}^*$ and is obtained by combining equations (4.4), (4.9) and (4.11). Using the forms of $\mathbf{b}_i$ and $\mathbf{F}$ of (4.2), we find that

$$(4.14) \quad v_{\bar{x}} = 4\Sigma_{i,j=1}^p\Sigma_{ij}^*\mathbf{W}_0'\boldsymbol{\Phi}_0^{-1}\langle\boldsymbol{\Sigma}_0\rangle\langle\boldsymbol{\Sigma}_0\rangle'\boldsymbol{\Phi}_0^{-1}\mathbf{W}_0(\mathbf{I} - \mathbf{M}_0)^{-1}\mathbf{T}_0\mathbf{b}_i\mathbf{b}_j'\mathbf{T}_0'(\mathbf{I} - \mathbf{M}_0')^{-1}$$

$$= 4\langle\boldsymbol{\Sigma}_0\rangle'\boldsymbol{\Phi}_0^{-1}\mathbf{W}_0(\mathbf{I} - \mathbf{M}_0)^{-1}\mathbf{T}_0\mathbf{F}\mathbf{T}_0'(\mathbf{I} - \mathbf{M}_0')^{-1}\mathbf{W}_0'\boldsymbol{\Phi}_0^{-1}\langle\boldsymbol{\Sigma}_0\rangle.$$

Combining (4.13) and (4.12), along with (4.14) yields the asymptotic variance v_∞ in (4.2).

The simplification of v_∞ (4.3) when there are explicit representations of the maximum likelihood estimates under the null hypothesis is now shown. From Szatrowski (1980), we find the maximum likelihood estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$ are given by (3.1), where $\hat{\boldsymbol{\Sigma}}$ is any allowable starting point. Thus $\hat{\boldsymbol{\Sigma}}$ is independent of $\mathbf{A}$ and $\bar{\mathbf{x}}$. This greatly simplifies taking the derivatives. In particular, equation (4.7) simplifies to $\partial\hat{\sigma}_s/\partial y = b_{sy}$. Thus we may set $\mathbf{M}_0 = \mathbf{0}$ in the earlier expressions. These facts along with applications of Theorems 3 and 4 lead to the simplifications of (4.13)

(4.15) $\quad \phi'_{0\langle\mathbf{A}\rangle}\boldsymbol{\Phi}^*\phi_{0\langle\mathbf{A}\rangle} = 2\,\mathrm{tr}\,\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}^*, \quad \phi'_{0\langle\mathbf{A}\rangle}\boldsymbol{\Phi}^*\phi_{1\langle\mathbf{A}\rangle} = 2\,\mathrm{tr}\,\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}^*.$

Similarly the expression for $v_{\bar{\mathbf{x}}}$ (4.14) simplifies to yield

(4.16) $\quad v_{\bar{\mathbf{x}}} = 4\langle\boldsymbol{\Sigma}_0\rangle'\boldsymbol{\Phi}_0^{-1}\mathbf{F}\boldsymbol{\Phi}_0^{-1}\langle\boldsymbol{\Sigma}_0\rangle = 4\Sigma_{i,j=1}^{p}\Sigma_{ij}^*\langle\boldsymbol{\Sigma}_0\rangle'\boldsymbol{\Phi}_0^{-1}\mathbf{b}_i\mathbf{b}_j'\boldsymbol{\Phi}_0^{-1}\langle\boldsymbol{\Sigma}_0\rangle.$

This expression can be further simplified using Theorem 3 by observing

(4.17)
$$\begin{aligned}\langle\boldsymbol{\Sigma}_0\rangle'\boldsymbol{\Phi}_0^{-1}\mathbf{b}_i &= \mathrm{tr}\{\boldsymbol{\Sigma}_0^{-1}(\mathbf{I} - \mathbf{R}_0\boldsymbol{\Sigma}_0^{-1})\mathbf{e}_i(\boldsymbol{\mu}^* - \boldsymbol{\mu}_0)'\}\\ &= (\boldsymbol{\mu}^* - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}_0^{-1}(\mathbf{I} - \mathbf{R}_0\boldsymbol{\Sigma}_0^{-1})\mathbf{e}_i = (\boldsymbol{\mu}^* - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}_0^{-1}\mathbf{e}_i.\end{aligned}$$

Substitution of (4.17) into (4.16) yields

(4.18)
$$v_{\bar{x}} = 4\,\mathrm{tr}\{\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\mu}^* - \boldsymbol{\mu}_0)(\boldsymbol{\mu}^* - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}_0^{-1}\} = 4(\boldsymbol{\mu}^* - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\mu}^* - \boldsymbol{\mu}_0).$$

Combining (4.18) and the simplified expressions (4.15) and (4.13) yields (4.3). □

5. *k*-population asymptotic nonnull distributions. In this section Theorems 6–8 giving the asymptotic nonnull distributions for the three k-population likelihood ratio statistics are presented along with their proofs. These proofs are short on detail as it is assumed that the reader is familiar with techniques used to prove Theorem 5. Recall that Section 2.2 contains a description of the k-population hypotheses and likelihood ratio tests.

5.1. H_k(MVC|mvc) *versus* $H_{k,A}$. In this case we test the null hypothesis H_k(MVC|mvc) that $\boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_k$ versus the alternative hypothesis $H_{k,A}$ without this restriction. Note that the structures of the mean and covariance matrix are the same under both null and alternative hypothesis and that the explicit maximum likelihood estimates involve different ways of pooling the data.

THEOREM 6. *The asymptotic nonnull distribution for the k-population likelihood ratio statistic given by* (2.3) *for testing* H_k (MVC|mvc) *versus* $H_{k,A}$ *is given by*

(5.1) $\quad \lim_{N\to\infty}\mathfrak{L}\left[N^{\frac{1}{2}}\{-(2/N)\log\lambda + \log[(\Pi_1^k|\boldsymbol{\Sigma}_d^*|^{f_d})/|\boldsymbol{\Sigma}_0|]\}\right] = \mathfrak{N}(0, v_\infty),$

(5.2)
$$\begin{aligned}v_\infty = 2\Sigma_{d=1}^k f_d^2\{p &- 4\langle\boldsymbol{\Sigma}_0\rangle'\boldsymbol{\Phi}_0^{-1}\mathbf{W}(\mathbf{I} - \mathbf{M}_0)^{-1}\mathbf{T}_0\langle\boldsymbol{\Sigma}_d^*\rangle\\ &+2\langle\boldsymbol{\Sigma}_0\rangle'\boldsymbol{\Phi}_0^{-1}\mathbf{W}(\mathbf{I} - \mathbf{M}_0)^{-1}\mathbf{T}_0(\boldsymbol{\Phi}_d^* + \mathbf{F}_d)\mathbf{T}_0'(\mathbf{I} - \mathbf{M}_0')^{-1}\mathbf{W}'\boldsymbol{\Phi}_0^{-1}\langle\boldsymbol{\Sigma}_0\rangle\},\end{aligned}$$

with

$$\mathbf{M}_0 = \left[\operatorname{tr} \Sigma_0^{-1}\mathbf{G}_g\Sigma_0^{-1}\mathbf{G}_h\right]^{-1}\left[\operatorname{tr} \Sigma_0^{-1}(\mathbf{F}_g^* + \mathbf{H}_g^*)\Sigma_0^{-1}\mathbf{G}_h\right],$$

$$\mathbf{T}_0 = (\mathbf{W}'\Phi_0^{-1}\mathbf{W})^{-1}\mathbf{W}'\Phi_0^{-1},\ \mathbf{R}_0 = \mathbf{Z}(\mathbf{Z}'\Sigma_0^{-1}\mathbf{Z})^{-1}\mathbf{Z}',\ \mathbf{B}_0 = (\bar{\mu}^* - \mu_0)(\bar{\mu}^* - \mu_0)',$$

$$\mathbf{F}_g^* = \mathbf{B}_0\Sigma_0^{-1}\mathbf{G}_g\Sigma_0^{-1}\mathbf{R}_0 + \mathbf{R}_0\Sigma_0^{-1}\mathbf{G}_g\mathbf{T}_0^{-1}\mathbf{B}_0,$$

$$\mathbf{H}_g^* = \mathbf{G}_g\Sigma_0^{-1}(\Sigma_0 - \mathbf{C}^*) + (\Sigma_0 - \mathbf{C}^*)\Sigma_0^{-1}\mathbf{G}_g,$$

$$\mathbf{F}_d = \Sigma_{i,j=1}^{p}(\Sigma_d^*)_{ij}\mathbf{b}_{id}\mathbf{b}_{jd}',$$

$$\mathbf{C}^* = \mathbf{B}_0 + \Sigma_{d=1}^{k} f_d\{\Sigma_d^* + (\mu_d^* - \bar{\mu}^*)(\mu_d^* - \bar{\mu}^*)'\},\quad f_d = p_d/N,$$

$$\mathbf{b}_{id} = \langle \mathbf{e}_i(\mu_d^* - \bar{\mu}^*)' + (\mu_d^* - \bar{\mu}^*)\mathbf{e}_i'$$
$$+ (\mathbf{I} - \mathbf{R}_0\Sigma_0^{-1})\mathbf{e}_i(\bar{\mu}^* - \mu_0)' + (\bar{\mu}^* - \mu_0)\mathbf{e}_i'(\mathbf{I} - \mathbf{R}_0\Sigma_0^{-1})'\rangle,\ \bar{\mu}^* = \Sigma_{d=1}^{k} f_d\mu_d^*,$$

where $((\mu_1^*, \Sigma_1^*), \cdots, (\mu_k^*, \Sigma_k^*))$ *is the value assumed under the alternative hypothesis and not under the null hypothesis,* $\Phi_d^* = \Phi(\Sigma_d^*)$, $\Phi_0 = \Phi(\Sigma_0)$, *and* μ_0 *and* Σ_0 *are the "maximum likelihood estimates" under the null hypothesis whose derivation is described in Section* 3.1 *with* $\bar{\mathbf{x}}$ *replaced by* $\bar{\mu}^*$, $\bar{\mathbf{x}}_d$ *replaced by* μ_d^* *and* $\mathbf{A}_d$ *replaced by* Σ_d^*.

Under the additional assumption that the maximum likelihood estimates have explicit representations under the null hypothesis, the form of the asymptotic variance simplifies to

$$(5.3)\quad v_\infty = 2\Sigma_{d=1}^{k} f_d^2\{\operatorname{tr}(\mathbf{I} - \Sigma_0^{-1}\Sigma^*)^2 + 2(\mu_d^* - \mu_0)'\Sigma_0^{-1}\Sigma_d^*\Sigma_0^{-1}(\mu_d^* - \mu_0)\}.$$

PROOF. Let $f = f(\langle \mathbf{A}_1\rangle, \cdots, \langle \mathbf{A}_k\rangle, \bar{\mathbf{x}}_1, \cdots, \bar{\mathbf{x}}_k) = -(2/N)\log\lambda = \log|\tilde{\Sigma}| - \Sigma_{d=1}^{k} f_d \log|\tilde{\Sigma}_d|$. We need to find the partial derivatives of f with respect to $\mathbf{A}_d$ and $\bar{\mathbf{x}}_d$, $d = 1, \cdots, k$. As in the proof of Theorem 5, noting that the observations in different populations are independent, we may express $v_\infty = \Sigma_{d=1}^{k}(v_{\langle \mathbf{A}_d\rangle} + v_{\bar{\mathbf{x}}_d})$ and calculate the variances $v_{\langle \mathbf{A}_d\rangle} = \phi'_{\langle \mathbf{A}_d\rangle}\Phi_d^*\phi_{\langle \mathbf{A}_d\rangle}$ and $v_{\bar{\mathbf{x}}_d} = \phi_{\bar{\mathbf{x}}_d}'\Sigma_d^*\phi_{\bar{\mathbf{x}}_d}$ separately. Let y be an element of $(\mathbf{A}_d', \bar{\mathbf{x}}_d')$.

$$(5.4)\qquad \frac{\partial f}{\partial y} = \Sigma_{s=0}^{m}\operatorname{tr}(\tilde{\Sigma}^{-1}\mathbf{G}_s)\frac{\partial\tilde{\sigma}_s}{\partial y} - f_d\Sigma_{s=0}^{m}\operatorname{tr}(\tilde{\Sigma}_d^{-1}\mathbf{G}_s)\frac{\partial\tilde{\sigma}_s^{(d)}}{\partial y}.$$

We use the form of $\hat{\sigma}_s^{(d)}$ given by

$$(5.5)\qquad \hat{\sigma}_s^{(d)} = \mathbf{e}_s'\hat{\sigma}^{(d)} = \mathbf{e}_s'\mathbf{Y}_1^{(d)-1}\mathbf{y}_2^{(d)} = \mathbf{e}_s'\mathbf{T}_d\langle \mathbf{C}_d\rangle,\ \hat{\mu}_d = \mathbf{Z}\hat{\beta}_d = \mathbf{R}_d\hat{\Sigma}_d^{-1}\bar{\mathbf{x}}_d,$$

where we define $\mathbf{Y}_1^{(d)}$, $\mathbf{y}_2^{(d)}$, $\mathbf{T}_d$, $\mathbf{R}_d$ and $\mathbf{C}_d$ by

$$(5.6)\quad \mathbf{Y}_1^{(d)} = \left[\operatorname{tr} \hat{\Sigma}_d^{-1}\mathbf{G}_g\hat{\Sigma}_d^{-1}\mathbf{G}_h\right],$$

$$\mathbf{y}_2^{(d)} = (\operatorname{tr} \hat{\Sigma}_d^{-1}\mathbf{G}_g\hat{\Sigma}_d^{-1}\mathbf{C}_d),\ \mathbf{T}_d = (\mathbf{W}'\hat{\Phi}_d^{-1}\mathbf{W})^{-1}\mathbf{W}'\hat{\Phi}_d^{-1},$$

$$\mathbf{R}_d = \mathbf{Z}(\mathbf{Z}'\hat{\Sigma}_d^{-1}\mathbf{Z})^{-1}\mathbf{Z}',\ \mathbf{C}_d = \mathbf{A}_d + (\bar{\mathbf{x}}_d - \hat{\mu}_d)(\bar{\mathbf{x}}_d - \hat{\mu}_d)',\ \hat{\Phi}_d \equiv \Phi(\hat{\Sigma}_d).$$

The form of $\hat{\sigma}_s$ is given by (4.5) and (4.6) with $\mathbf{C}$ defined by

$$\mathbf{C} = (\bar{\mathbf{x}} - \hat{\boldsymbol{\mu}})(\bar{\mathbf{x}} - \hat{\boldsymbol{\mu}})' + \mathbf{A},\ \mathbf{A} = \Sigma_{d=1}^{k} f_d(\mathbf{A}_d + (\bar{\mathbf{x}}_d - \bar{\mathbf{x}})(\bar{\mathbf{x}}_d - \bar{\mathbf{x}})'). \tag{5.7}$$

The derivation of $\partial\tilde{\sigma}_s^{(d)}/\partial y$ follows the derivation in the proof of Theorem 5. Thus, under the alternative hypothesis,

$$\left.\frac{\partial\tilde{\sigma}_s^{(d)}}{\partial\langle\mathbf{A}_d\rangle}\right| = \mathbf{T}_d^{*\prime}\mathbf{e}_s,\ \left.\frac{\partial\tilde{\sigma}_s^{(d)}}{\partial\bar{\mathbf{x}}_d}\right| = \mathbf{0},\ \mathbf{T}_d^* = (\mathbf{W}'\boldsymbol{\Phi}_d^*\mathbf{W})^{-1}\mathbf{W}'\boldsymbol{\Phi}_d^{*-1}. \tag{5.8}$$

The derivation of $\partial\hat{\sigma}_s/\partial y$ also parallels the derivation in the proof of Theorem 5 with only the slight modifications that

$$\frac{\partial\langle\mathbf{A}\rangle}{\partial y} = f_d\left\{\frac{\partial\langle\mathbf{A}_d\rangle}{\partial y} + \frac{\partial\bar{\mathbf{x}}_d}{\partial y}(\bar{\mathbf{x}}_d - \bar{\mathbf{x}})' + (\bar{\mathbf{x}}_d - \bar{\mathbf{x}})\left(\frac{\partial\bar{\mathbf{x}}_d}{\partial y}\right)'\right\};\ \frac{\partial\bar{\mathbf{x}}}{\partial y} = f_d\frac{\partial\bar{\mathbf{x}}_d}{\partial y}. \tag{5.9}$$

Thus, from (4.8) and (4.10), we have

$$\frac{\partial\tilde{\sigma}_s}{\partial\langle\mathbf{A}_d\rangle} = f_d\mathbf{T}_0'(\mathbf{I} - \mathbf{M}_0')^{-1}\mathbf{e}_s, \tag{5.10}$$

with $\mathbf{T}_0$ and $\mathbf{M}_0$ given in (5.2). Using (5.9), (4.8) and (4.11) we find that

$$\begin{aligned}\left.\frac{\partial\tilde{\sigma}_s}{\partial(\bar{x}_d)_i}\right| &= f_d\mathbf{e}_s'(\mathbf{I} - \mathbf{M})^{-1}\mathbf{T}_0\langle\mathbf{e}_i(\bar{\mathbf{x}}_d - \bar{\mathbf{x}})' + (\bar{\mathbf{x}}_d - \bar{\mathbf{x}})\mathbf{e}_i' \\ &\quad + (\mathbf{I} - \mathbf{R}_0\boldsymbol{\Sigma}_0^{-1})\mathbf{e}_i(\bar{\mathbf{x}} - \hat{\boldsymbol{\mu}})' + (\bar{\mathbf{x}} - \hat{\boldsymbol{\mu}})\mathbf{e}_i'(\mathbf{I} - \mathbf{R}_0\boldsymbol{\Sigma}_0^{-1})'\rangle.\end{aligned} \tag{5.11}$$

The calculation of $v_{\langle\mathbf{A}_d\rangle}$ yields the same result as derived in (4.12)–(4.14) with the addition of the factor f_d^2 on the right side. Inserting the factor f_d^2 on the right side of the expression for $v_{\bar{\mathbf{x}}}$ (4.15), replacing $\boldsymbol{\Sigma}^*$ by $\boldsymbol{\Sigma}_d^*$, dropping the zero subscripts on $\mathbf{W}$ and $\mathbf{Z}$ and replacing $\mathbf{b}_i$ with $\mathbf{b}_{id}$ (see 5.2) yields the expression for $v_{\bar{\mathbf{x}}d}$. The remainder of the derivation parallels that in the proof of Theorem 5 yielding (5.2).

The simplification of v_∞ (5.3) when there are explicit representations of the maximum likelihood estimates under the null hypothesis results in $\mathbf{M}_0 = \mathbf{0}$ as in the proof of Theorem 5. The simplifications parallel those in the earlier proof and lead to (5.3). □

5.2. $H_k(\text{VC}|\text{mvc})$ *versus* $H_{k,A}$. In this case we test the null hypothesis $H_k(\text{VC}|\text{mvc})$ that $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_k$ versus the alternative hypothesis $H_{k,A}$ without this restriction.

THEOREM 7. *The asymptotic nonnull distribution for the k-population likelihood ratio statistic given by* (2.3) *for testing* $H_k(\text{VC}|\text{mvc})$ *versus* $H_{k,A}$ *is of the form* (5.1) *where the variance* v_∞ *is given by* (5.2) *with* $\mathbf{F}_d = \mathbf{0}$, $\mathbf{F}_g^* = \mathbf{0}$, $\mathbf{B}_0 = \mathbf{0}$, $\mathbf{b}_i = \mathbf{0}$ *and* $\mathbf{C}^* = \Sigma_{d=1}^{k} f_d\boldsymbol{\Sigma}_d^*$, *where* $(\boldsymbol{\Sigma}_1^*, \cdots, \boldsymbol{\Sigma}_k^*)$ *is the value assumed under the alternative hypothesis and not under the null hypothesis, and* $\boldsymbol{\Sigma}_0$ *is the "maximum likelihood estimate" of the common covariance matrix under the null hypothesis whose derivation is described in Section* (3.1) *with* $(\bar{\mathbf{x}}_d, \mathbf{A}_d) = (\boldsymbol{\mu}_d^*, \boldsymbol{\Sigma}_d^*)$, $d = 1, \cdots, k$.

Under the additional assumption that the maximum likelihood estimates have explicit representations under the null hypothesis, the asymptotic variance simplifies, becoming (5.3) *after dropping the second term,* $2(\boldsymbol{\mu}_d^* - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_d^*\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\mu}_d^* - \boldsymbol{\mu}_0)$.

REMARK. The form of the asymptotic variance under this additional explicit representation assumption is similar to a form obtained by Sugiura (1969) for testing the null hypothesis $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_k$ in the nonpatterned case.

PROOF. The proof uses the same techniques used in the proofs of Theorems 5 and 6 with $\hat{\sigma}_s^{(d)}$ given by (5.5) and $\hat{\sigma}_s$ given by

$$\hat{\sigma}_s = \mathbf{e}_s'\left[\operatorname{tr} \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{G}_g\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{G}_h\right]^{-1} \left(\operatorname{tr} \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{G}_g\hat{\boldsymbol{\Sigma}}^{-1}\left\{\Sigma_{d=1}^k f_d\left\{\mathbf{A}_d + (\bar{\mathbf{x}}_d - \hat{\boldsymbol{\mu}}_d)(\bar{\mathbf{x}}_d - \hat{\boldsymbol{\mu}}_d)'\right\}\right\}\right). \tag{5.12}$$

In this case, $v_{\bar{\mathbf{x}}} = 0$ and the asymptotic nonnull distribution is independent of the values $\boldsymbol{\mu}_1^*, \cdots, \boldsymbol{\mu}_k^*$ as long as these values have the hypothesized structures. The details of the proof are omitted. □

5.3. $H_k(M|\mathrm{mVC})$ *versus* $H_k(\mathrm{VC}|\mathrm{mvc})$. In this case we test the null hypothesis that $\boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k$ versus the alternative hypothesis without this restriction. Under both null and alternative hypotheses we assume $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_k$.

THEOREM 8. *The asymptotic nonnull distribution for the k-population likelihood ratio statistic of the form* (2.2) *for testing* $H_k(M|\mathrm{mVC})$ *versus* $H_k(\mathrm{VC}|\mathrm{mvc})$ *is given by*

$$\lim_{N\to\infty}\mathcal{L}\left[N^{\frac{1}{2}}\{-(2/N)\log\lambda + \log(|\boldsymbol{\Sigma}^*|/|\boldsymbol{\Sigma}_0|)\}\right] = \mathfrak{N}(0, v_\infty), \tag{5.13}$$

where v_∞ *is given by* (5.2) *with* $\boldsymbol{\Sigma}_1^* = \cdots = \boldsymbol{\Sigma}_k^* \equiv \boldsymbol{\Sigma}^*$, $\boldsymbol{\Phi}_1^* = \cdots = \boldsymbol{\Phi}_k^* \equiv \boldsymbol{\Phi}^*$, $\boldsymbol{\Sigma}^*$ *is the common variance matrix of the k populations,* $(\boldsymbol{\mu}_1^*, \cdots, \boldsymbol{\mu}_k^*, \boldsymbol{\Sigma}^*)$ *is the value assumed under the alternative hypothesis and not under the null hypothesis, and* $\boldsymbol{\mu}_0$ *and* $\boldsymbol{\Sigma}_0$ *are the "maximum likelihood estimates" under the null hypothesis whose derivation is described in Section* (3.1) *with* $\bar{\mathbf{x}}$ *replaced by* $\bar{\boldsymbol{\mu}}^*$, $\bar{\mathbf{x}}_d$ *replaced by* $\boldsymbol{\mu}_d^*$ *and* $\mathbf{A}_d$ *replaced by* $\boldsymbol{\Sigma}^*$.

Under the additional assumption that the maximum likelihood estimates have explicit representations under the null hypothesis, the asymptotic variance simplifies becoming (5.3) *with* $\boldsymbol{\Sigma}_d^* \equiv \boldsymbol{\Sigma}^*$, $d = 1, \cdots, k$ *and* $\mathbf{B}_0$ *replaced by* $\mathbf{B}_d = (\boldsymbol{\mu}_d^* - \bar{\boldsymbol{\mu}}^*)(\boldsymbol{\mu}_d^* - \bar{\boldsymbol{\mu}}^*)'$.

PROOF. The forms of $\hat{\boldsymbol{\sigma}}$ under the null and alternative hypotheses are given by (4.5), (4.6), (5.7) and (5.12) respectively. The proof is straightforward using techniques from the proofs of Theorems 5 and 6. □

6. Discussion of results. Evaluating any of the asymptotic variances in Theorems 5–8 at a point in the null hypothesis region results in a zero variance. This is expected because the standard delta method yields first order results under the assumption that the variances are evaluated at a point belonging strictly to the alternative hypothesis under consideration and not to any alternative hypothesis that contains the null hypothesis.

In many studies with explicit representations of the maximum likelihood estimates the problems can be transformed to achieve a simple canonical form. The forms of the results in Theorems 5–8 for these special cases may be easily simplified for specific problems using the same transformations. While it is important that the mean and covariance have the linear structure described in Section Two, it is not necessary to explicitly represent them in this form in order to use the distributional results of this paper.

It is expected that a typical application of these nonnull results will be to an approximate power calculation or sample size calculation. Frequently in patterned hypothesis testing problems, the investigator hopes that the null hypothesis, the simpler pattern, is the true state of nature. This differs from the usual case in which the investigator hopes the alternative hypothesis is the true state of nature. Thus it becomes increasingly important that the test has high power against suitable points under the alternative hypothesis.

Acknowledgments. I wish to thank a referee for suggesting the second part of Lemma 1 and its proof as well as for simplifying the statement and proofs of Theorems 3 and 4. Also both the referee and Associate Editor made many helpful suggestions concerning notation used in this paper.

REFERENCES

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

[2] ANDERSON, T. W. (1969). Statistical inference for covariance matrices with linear structure. In *Proceedings of the Second International Symposium on Multivariate Analysis.* (P. R. Krishnaiah, ed.). 55–66. Academic Press, New York.

[3] ANDERSON, T. W. (1970). Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices. In *Essays in Probability and Statistics*, (R. C. Bose, et. al, eds.), 1–24. Univ. North Carolina Press, Chapel Hill.

[4] ANDERSON, T. W. (1973). Asymptotic efficient estimation of covariance matrices with linear structure. *Ann. Statist.* **1** 135–141.

[5] ARNOLD, S. F. (1970). Products of problems and patterned covariance matrices that arise from interchangeable random variables. Technical report No. 46, Statist. Dept., Stanford Univ.

[6] ARNOLD, S. F. (1973). Application of the theory of products of problems to certain patterned covariance matrices. *Ann. Statist.* **1** 682–699.

[7] ARNOLD, S. F. (1976). Applications of products to the generalized compound symmetry problem. *Ann. Statist.* **3** 227–233.

[8] CONSUL, P. C. (1968). On the distribution of Votaw's likelihood ratio criterion L for testing the bipolarity of a covariance matrix. *Math. Nachr.* **36** 1–13.

[9] CONSUL, P. C. (1969). On the exact distribution of Votaw's criteria for testing compound symmetry of a covariance matrix. *Ann. Math. Statist.* **40** 836–843.

[10] GLESER, L. J. and I. OLKIN (1966). A k-sample regression model with covariance. In *Proceedings of the First International Symposium on Multivariate Analysis.* (P. R. Krishnaiah, ed.). 59–72. Academic Press, New York.

[11] GLESER, L. J. and I. OLKIN (1969). Testing for equality of means, equality of variances, and equality of covariances under restrictions upon the parameter space. *Ann. Inst. Statist. Math.* **21** 33–48.

[12] JOHN, S. (1971). A test of equality of block-diagonal covariance matrices and its role of unification. *J. Roy. Statist. Soc. Ser. B* **33** 301–306.

[13] KHATRI, C. G. and SRIVASTAVA, M. S. (1971). On exact nonnull distribution of likelihood ratio criteria for sphericity test and equality of covariance matrices. *Sankhyā Ser. A.* **33** 201–206.

[14] MATHAI, A. M. and P. N. RATHIE (1970). The exact distribution of Votaw's criteria. *Ann. Inst. Statist. Math.* **22** 89–116.

[15] MAUCHLY, J. W. (1940). Significance test for sphericity of a normal n-variate distribution. *Ann. Math. Statist.* **11** 204–209.

[16] MUKHERJEE, B. N. (1966). Derivation of likelihood ratio tests for the Guttman quasi-simplex covariance structures. *Psychometrika* **31** 97–123.

[17] MUKHERJEE, B. N. (1970). Likelihood ratio tests of statistical hypotheses associated with patterned covariance matrices in psychology. *British J. Math. Statist. Psychology* **23** 87–108.

[18] OLKIN, I. (1972). Testing and estimation for structures which are circularly symmetric in blocks. In *Proceedings of the Symposium on Multivariate Analysis*. 183–195. Dalhousie, Nova Scotia.

[19] OLKIN, I. and PRESS, S. J. (1969). Testing and estimation for a circular stationary model. *Ann. Math. Statist.* **40** 1358–1373.

[20] SUGIURA, N. (1969). Asymptotic expansions of the distributions of the likelihood ratio criteria for covariance matrix. *Ann. Math. Statist.* **40** 2051–2063.

[21] SUGIURA, N. and FUJIKOSHI, Y. (1969). Asymptotic expansions of the nonnull distributions of the likelihood ratio criteria for multivariate linear hypothesis and independence. *Ann. Math. Statist.* **40** 942–952.

[22] SZATROWSKI, T. H. (1976). Estimation and testing for block compound symmetry and other patterned covariance matrices with linear and nonlinear structure. Technical report No. 107, Statist. Dept., Stanford Univ.

[23] SZATROWSKI, T. H. (1980). Necessary and sufficient conditions for explicit solutions in the multivariate normal estimation problem for patterned means and covariances. *Ann. Statist.* To appear.

[24] VOTAW, D. F. (1948). Testing compound symmetry in a normal multivariate distribution. *Ann. Math. Statist.* **19** 447–473.

[25] WILKS, S. S. (1946). Sample criteria for testing equality of means, equality of variances and equality of covariances in a normal multivariate distribution. *Ann. Math. Statist.* **17** 257–281.

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
HILL CENTER
NEW BRUNSWICK, NEW JERSEY 08903