The Annals of Statistics
1983, Vol. 11, No. 3, 947–958

MISSING DATA IN THE ONE-POPULATION MULTIVARIATE NORMAL PATTERNED MEAN AND COVARIANCE MATRIX TESTING AND ESTIMATION PROBLEM

BY TED H. SZATROWSKI[1]

Rutgers University

In this paper the multivariate normal linear patterned mean and covariance matrix testing and estimation problem is studied in the presence of missing data for general one-population hypotheses. The Newton-Raphson, Method of Scoring and EM algorithms are given for finding the maximum likelihood estimates. The asymptotic joint distribution of the maximum likelihood estimates under null and alternative hypotheses are derived along with the form of the likelihood ratio statistic and its asymptotically Chi squared null and asymptotically normal nonnull distributions. The distributions of the maximum likelihood estimates and nonnull distributions of the likelihood ratio tests are derived using the standard multivariate and univariate delta method respectively, and may be evaluated at a parameter point under the alternative hypothesis parameter space or at a parameter point in a parameter space that contains the null and alternative hypothesis parameter spaces. New results for these problems in the presence of complete data as well as known results (Szatrowski, 1979) are special cases of the results of this paper.

1. Introduction. The problem of testing and estimation for the unknown mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ from a multivariate normal distribution, $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ given a random sample $\mathbf{x}_1, \cdots, \mathbf{x}_N$ from this distribution has been studied when $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ have linear patterns by many authors. (See Anderson (1973) and Szatrowski (1979) for a partial list of references.) Maximum likelihood estimates (MLE), likelihood ratio statistics (LRS) and various null and nonnull, exact and asymptotic distributions have been obtained for both this one-population case and for generalizations to k-populations. In this paper, we extend these one-population results to the case in which we do not observe every component of every observation. This "missing data" or "incomplete data" can arise in any problem in which one is collecting multivariate observations. Szatrowski (1981a) obtains similar results for testing and estimation problems involving linearly patterned means and linearly patterned correlations (versus linear patterned covariances considered in this paper).

As a motivating example, consider the following problem which arises in test equating (Holland and Wightman, 1982). Let exams A and B each consist of three separately timed parallel sections, A_1, A_2, A_3 and B_1, B_2, B_3 respectively. Students are randomly assigned to take one of three forms of a five section exam, three sections of which always consist of A_1, A_2, and A_3. The other two sections are either (B_1, B_2), (B_1, B_3), or (B_2, B_3). Students are scored based on their performance on the operational sections, A_1, A_2, and A_3. The sections of test B, the experimental sections, are included for equating purposes only. Thus, it is not necessary for each student to take all sections of test B. Students do not know which sections are operational or experimental and order of sections may be randomized. Thus, each observation has as missing data one of the sections of test B.

The analyst's task is to estimate the mean vector and covariance matrix for the scores on the six test sections. Standard methods of estimation (MLE) that ignore possible

---

Received July 1982; revised February 1983.

[1] Research partially supported by National Science Foundation Grant MCS77-28184 and the Program Statistics Research Project of the Educational Testing Service Research Statistics Group.

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

Key words and phrases. Asymptotic efficiency, asymptotic nonnull distribution, delta method, EM algorithm, hypothesis testing, iterative computations, matrix derivatives, maximum likelihood estimates, missing data, patterned means, patterned covariance matrices.

patterns in the population mean and covariance estimates lead to less efficient estimates. In the case where only one rather than two sections of test B are given to each student, standard methods that ignore patterns in the mean and covariance matrix do not lead to unique estimates of the population mean vector and covariance matrix.

Let $\mathbf{X}$ be a p-component column vector with multivariate normal distribution such that the mean vector $\boldsymbol{\mu} = \mathscr{E}(\mathbf{X})$ and covariance matrix $\boldsymbol{\Sigma} = \mathrm{Cov}(\mathbf{X}) = \mathscr{E}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'$ have the linear structure considered by Anderson (1969, 1970, 1973). Specifically, $\boldsymbol{\mu} = \sum_1^r \mathbf{z}_j\beta_j = \mathbf{Z}\boldsymbol{\beta}$, $\mathbf{Z} = [\mathbf{z}_1, \cdots, \mathbf{z}_r]$, $\boldsymbol{\beta} \in R^r$, where the $\mathbf{z}$'s are known, linearly independent column vectors and the β's are unknown scalars. The covariance matrix is given by $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\sigma}) = \sum_1^m \sigma_g \mathbf{G}_g$, $\boldsymbol{\sigma} = (\sigma_1, \cdots, \sigma_m)'$, where the $\mathbf{G}$'s are known, linearly independent symmetric matrices and the σ's are unknown scalars, such that $\boldsymbol{\sigma} \in \boldsymbol{\Theta}$, $\boldsymbol{\Theta} = \{\boldsymbol{\theta} \in R^m \mid \boldsymbol{\Sigma}(\boldsymbol{\theta}) > 0\}$, 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. In some cases, there exists an orthogonal matrix $\boldsymbol{\Gamma}$ such that $\boldsymbol{\Gamma}\mathbf{G}_g\boldsymbol{\Gamma}' = \boldsymbol{\Lambda}_g$, $g = 1, \cdots, m$, diagonal matrices. This greatly simplifies some of the results obtained in this paper.

Definition 1.1. 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 1.1, and defining $\mathbf{W} = [\langle \mathbf{G}_1 \rangle, \langle \mathbf{G}_2 \rangle, \cdots, \langle \mathbf{G}_m \rangle]$, we observe $\langle \boldsymbol{\Sigma} \rangle = \mathbf{W}\boldsymbol{\sigma}$.

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

$$\mathbf{Z} = [\mathbf{Z}_0 : \mathbf{Z}_1],\ \boldsymbol{\beta} = (\boldsymbol{\beta}_0', \boldsymbol{\beta}_1')',\ \mathbf{W} = [\mathbf{W}_0 : \mathbf{W}_1],\ \boldsymbol{\sigma} = (\boldsymbol{\sigma}_0', \boldsymbol{\sigma}_1')', \tag{1.1}$$

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 $\frac{1}{2}p(p+1) \times m_0$ and $m_0 \times 1$ respectively. At least one of the inequalities $r_0 \leq r$ and $m_0 \leq m$ 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 $\boldsymbol{\sigma}$. To do this we find the MLE under each hypothesis and form the LRS.

However, instead of observing $\mathbf{x}_i$, we observe $\mathbf{E}_{\alpha(i)}\mathbf{x}_i$, $i = 1, \cdots, N$ where $\mathbf{E}_\alpha$, $\alpha = 1, \cdots, q$ are known $u_\alpha \times p$ matrices of full rank with $u_\alpha \leq p$. The function $\alpha(i)$ is given by $\alpha(i) = j$ for $i = m_{j-1} + 1, \cdots, m_j$; $j = 1, \cdots, q$, $m_0 \equiv 0$. For example, if $q = 2$, the observed data would be of the form $\mathbf{E}_1\mathbf{x}_1, \cdots, \mathbf{E}_1\mathbf{x}_{m_1}, \mathbf{E}_2\mathbf{x}_{m_1+1}, \cdots, \mathbf{E}_2\mathbf{x}_{m_2}$. We let $n_\alpha = m_\alpha - m_{\alpha-1}$ be the number of observations of the form $\mathbf{E}_\alpha\mathbf{x}$ and $f_\alpha = n_\alpha/N$.

The following Condition 1.1 relating the structure of the missing data and the patterned mean and covariance structures is assumed to hold for the data collected in the one-population missing data problems under consideration. If this condition did not hold, data would not be available for estimating one or more of the unknown parameters.

Condition 1.1. For each j, there exists an α such that $\mathbf{E}_\alpha\mathbf{z}_j \neq \mathbf{0}$, $j = 1, \cdots, r$, and for each g, there exists an α such that $\mathbf{E}_\alpha\mathbf{G}_g\mathbf{E}_\alpha' \neq \mathbf{0}$, $g = 1, \cdots, m$.

In Section 2, iterative algorithms for finding the MLE are given. Asymptotic distributions of the MLE are derived in Section 3. These asymptotic distribution results are used in Section 4 in the derivation of the asymptotic nonnull distributions of the likelihood ratio statistic. Details of many of the proofs of Lemmas and Theorems in this paper may be found in Szatrowski (1981b).

2. Maximum likelihood estimates and iterative algorithms. In this section, the forms of the Newton-Raphson, Method of Scoring and EM algorithms (e.g. Dempster, Laird and Rubin, 1977) are given for finding MLE for the various hypotheses in the one-population problem. Several algorithms are given because none of them have been shown to have guaranteed convergence or to be always better than the others. Some brief comments are made concerning starting points and convergence of these algorithms at the end of this section.

The likelihood function for our observations $\mathbf{E}_1\mathbf{x}_1, \cdots, \mathbf{E}_q\mathbf{x}_N$ is of the form

$$L(\boldsymbol{\beta}, \boldsymbol{\sigma}) = K \prod_{\alpha=1}^{q} |\mathbf{E}_\alpha\boldsymbol{\Sigma}\mathbf{E}_\alpha'|^{-n_\alpha/2}\exp\{-\tfrac{1}{2}\, n_\alpha \mathrm{tr}(\mathbf{E}_\alpha\boldsymbol{\Sigma}\mathbf{E}_\alpha')^{-1}\mathbf{E}_\alpha\mathbf{C}_\alpha\mathbf{E}_\alpha'\}, \tag{2.1}$$

with $n_\alpha\mathbf{C}_\alpha$ the usual sample cross product matrix for $\mathbf{x}_{m_{\alpha-1}+1}, \cdots, \mathbf{x}_{m_\alpha}$ given by

$$\mathbf{A}_\alpha = \textstyle\sum_{j=m_{\alpha-1}+1}^{m_\alpha}(\mathbf{x}_j - \bar{\mathbf{x}}_\alpha)(\mathbf{x}_j - \bar{\mathbf{x}}_\alpha)', \quad \bar{\mathbf{x}}_\alpha = (1/n_\alpha)\textstyle\sum_{m_{\alpha-1}+1}^{m_\alpha}\mathbf{x}_j,$$

plus a term involving the patterned mean, $n_\alpha(\bar{\mathbf{x}}_\alpha - \boldsymbol{\mu})(\bar{\mathbf{x}}_\alpha - \boldsymbol{\mu})'$, i.e.

$$n_\alpha\mathbf{C}_\alpha = \mathbf{A}_\alpha + n_\alpha(\bar{\mathbf{x}}_\alpha - \boldsymbol{\mu})(\bar{\mathbf{x}}_\alpha - \boldsymbol{\mu})', \quad \alpha = 1, \cdots, q, \tag{2.2}$$

with K being a generic for constants independent of $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$. Letting $\boldsymbol{\Sigma}_\alpha = \mathbf{E}_\alpha\boldsymbol{\Sigma}\mathbf{E}_\alpha'$, and $\mathbf{C}_\alpha^* = \mathbf{E}_\alpha\mathbf{C}_\alpha\mathbf{E}_\alpha'$, we find the loglikelihood function to be

$$l(\boldsymbol{\beta}, \boldsymbol{\sigma}) = K - \Sigma_1^q(n_\alpha/2)\log|\boldsymbol{\Sigma}_\alpha| - \Sigma_1^q(n_\alpha/2)\,\mathrm{tr}\,\boldsymbol{\Sigma}_\alpha^{-1}\mathbf{C}_\alpha^*. \tag{2.3}$$

To derive the first and second partial derivatives of $l(\boldsymbol{\beta}, \boldsymbol{\sigma})$, we use the following well-known matrix derivative results, given in Lemma 2.1 and Lemma 2.2.

LEMMA 2.1. *If $\boldsymbol{\Sigma}$ is a patterned covariance matrix, then*

$$\partial \log|\boldsymbol{\Sigma}|/\partial\sigma_g = \mathrm{tr}\,\boldsymbol{\Sigma}^{-1}\partial(\boldsymbol{\Sigma})/\partial\sigma_g = \mathrm{tr}\,\boldsymbol{\Sigma}^{-1}\mathbf{G}_g.$$

LEMMA 2.2. *If $\mathbf{Y}$ is a matrix function of a matrix $\mathbf{X}$, then*

$$\partial\,\mathrm{tr}(\mathbf{AYB})/\partial x_{ij} = \mathrm{tr}\,\mathbf{A}(\partial\mathbf{Y}/\partial x_{ij})\mathbf{B}, \quad \partial\,\mathbf{Y}^{-1}/\partial x_{ij} = -\mathbf{Y}^{-1}(\partial\mathbf{Y}/\partial x_{ij})\mathbf{Y}^{-1}.$$

Using Lemmas 2.1 and 2.2, we find the first partial derivatives of $l(\boldsymbol{\beta}, \boldsymbol{\sigma})$ are given by

$$(\partial l/\partial\boldsymbol{\beta}) = \Sigma_1^q n_\alpha\mathbf{Z}_\alpha'\,\boldsymbol{\Sigma}_\alpha^{-1}\mathbf{E}_\alpha(\bar{\mathbf{x}}_\alpha - \boldsymbol{\mu}), \tag{2.4}$$

$$(\partial l/\partial\sigma_g) = \tfrac{1}{2}\,\Sigma_1^q n_\alpha \mathrm{tr}\,\boldsymbol{\Sigma}_\alpha^{-1}\mathbf{G}_{g\alpha}\boldsymbol{\Sigma}_\alpha^{-1}(\mathbf{C}_\alpha^* - \boldsymbol{\Sigma}_\alpha), \quad g = 1, \cdots, m \tag{2.5}$$

with $\mathbf{Z}_\alpha = \mathbf{E}_\alpha\mathbf{Z}$ and $\mathbf{G}_{g\alpha} = \mathbf{E}_\alpha\mathbf{G}_g\mathbf{E}_\alpha'$, $\alpha = 1, \cdots, q$. Continuing to take partial derivatives yields the second partial derivatives given by

$$-(\partial^2 l/\partial\boldsymbol{\beta}^2) = \Sigma_1^q n_\alpha\mathbf{Z}_\alpha'\boldsymbol{\Sigma}_\alpha^{-1}\mathbf{Z}_\alpha, \tag{2.6}$$

$$-(\partial^2 l/\partial\beta_j\partial\sigma_h) = \Sigma_1^q n_\alpha\mathbf{z}_j'\mathbf{E}_\alpha'\boldsymbol{\Sigma}_\alpha^{-1}\mathbf{G}_{h\alpha}\boldsymbol{\Sigma}_\alpha^{-1}\mathbf{E}_\alpha(\bar{\mathbf{x}}_\alpha - \boldsymbol{\mu}), \quad j = 1, \cdots, r;\ h = 1, \cdots, m; \tag{2.7}$$

$$-(\partial^2 l/\partial\sigma_g\partial\sigma_h) = \tfrac{1}{2}\,\Sigma_1^q n_\alpha \mathrm{tr}\,\boldsymbol{\Sigma}_\alpha^{-1}\mathbf{G}_{g\alpha}\boldsymbol{\Sigma}_\alpha^{-1}\mathbf{G}_{h\alpha}\boldsymbol{\Sigma}_\alpha^{-1}(2\mathbf{C}_\alpha^* - \boldsymbol{\Sigma}_\alpha), \quad g, h = 1, \cdots, m. \tag{2.8}$$

Taking the expected values of the second partial derivatives after observing that $\mathscr{E}\bar{\mathbf{x}}_\alpha = \boldsymbol{\mu}_\alpha$ and $\mathscr{E}\mathbf{C}_\alpha^* = \boldsymbol{\Sigma}_\alpha$, $\alpha = 1, \cdots, q$, yields

$$-\mathscr{E}(\partial^2 l/\partial\boldsymbol{\beta}^2) = \Sigma_1^q n_\alpha\mathbf{Z}_\alpha'\boldsymbol{\Sigma}_\alpha^{-1}\mathbf{Z}_\alpha, \tag{2.9}$$

$$-\mathscr{E}(\partial^2 l/\partial\beta_j\partial\sigma_h) = 0, \quad j = 1, \cdots, r,\ h = 1, \cdots, m, \tag{2.10}$$

$$-\mathscr{E}(\partial^2 l/\partial\sigma_g\partial\sigma_h) = \tfrac{1}{2}\,\Sigma_1^q n_\alpha \mathrm{tr}\,\boldsymbol{\Sigma}_\alpha^{-1}\mathbf{G}_{g\alpha}\boldsymbol{\Sigma}_\alpha^{-1}\mathbf{G}_{h\alpha}, \quad g, h = 1, \cdots, m. \tag{2.11}$$

The Newton-Raphson and Method of Scoring iterative procedures are of the form

$$\hat{\boldsymbol{\theta}}^{(1)} = \hat{\boldsymbol{\theta}}^{(0)} + a\mathbf{S}^{-1}(\hat{\boldsymbol{\theta}}^{(0)})\mathbf{s}(\hat{\boldsymbol{\theta}}^{(0)}) \tag{2.12}$$

with $a = 1$, (a different from one is used for monitoring step size to improve convergence), $\mathbf{s}(\boldsymbol{\theta}) = \partial l(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, and with $\hat{\boldsymbol{\theta}}^{(0)}$ and $\hat{\boldsymbol{\theta}}^{(1)}$ the old and new values respectively in the iteration scheme. Let $\boldsymbol{\theta}$ and $\mathbf{s}$ in (2.12) be given by $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\sigma}')'$, $\mathbf{s} = (\mathbf{s}_1', \mathbf{s}_2')'$, $\mathbf{s}_1 = (\partial l/\partial\boldsymbol{\beta})$ and $(s_2)_g = (\partial l/\partial\sigma_g)$, $g = 1, \cdots, m$, in (2.4) and (2.5) respectively. Partition $\mathbf{S}(r + m) \times (r + m)$ into $\mathbf{S}_{ij}$, $i, j = 1, 2$ with $\mathbf{S}_{11}$ $r \times r$.

LEMMA 2.3. *The Newton-Raphson algorithm for finding the* MLE *in the one-population problem is given by* (2.12) *with* $a = 1$, $\mathbf{S}_{11} = (-\partial^2 l/\partial\boldsymbol{\beta}^2)$ *in* (2.6), $(S_{12})_{jh} = (-\partial^2 l/\partial\beta_j\partial\sigma_h)$ *in* (2.7), $\mathbf{S}_{21} = \mathbf{S}'_{12}$ *and* $(S_{22})_{gh} = (-\partial^2 l/\partial\sigma_g\partial\sigma_h)$ *in* (2.8).

For the Method of Scoring algorithm, the components of $\mathbf{S}_{11}$, $(S_{12})_{jh}$ and $(S_{22})_{gh}$ are given in (2.9)–(2.11) respectively. Because $\mathbf{S}_{12} = \mathbf{0}$, the Method of Scoring using (2.12) with $a = 1$ simplifies.

LEMMA 2.4. *The Method of Scoring algorithm for finding* MLE *in the one-population problem is given by*

$$\hat{\boldsymbol{\beta}} = (\Sigma_1^q n_\alpha \mathbf{Z}'_\alpha \hat{\boldsymbol{\Sigma}}_\alpha^{-1} \mathbf{Z}_\alpha)^{-1} \Sigma_1^q n_\alpha \mathbf{Z}'_\alpha \hat{\boldsymbol{\Sigma}}_\alpha^{-1} \mathbf{E}_\alpha \bar{\mathbf{x}}_\alpha, \tag{2.13}$$

$$n_\alpha \hat{\mathbf{C}}_\alpha = \mathbf{A}_\alpha + n_\alpha(\bar{\mathbf{x}}_\alpha - \mathbf{Z}\hat{\boldsymbol{\beta}})(\bar{\mathbf{x}}_\alpha - \mathbf{Z}\hat{\boldsymbol{\beta}})', \quad \mathbf{C}^*_\alpha = \mathbf{E}_\alpha \mathbf{C}_\alpha \mathbf{E}'_\alpha, \tag{2.14}$$

$$\hat{\boldsymbol{\sigma}} = [\Sigma_1^q n_\alpha \operatorname{tr} \hat{\boldsymbol{\Sigma}}_\alpha^{-1} \mathbf{G}_{g\alpha} \hat{\boldsymbol{\Sigma}}_\alpha^{-1} \mathbf{G}_{h\alpha}]_{gh}^{-1} (\Sigma_1^q n_\alpha \operatorname{tr} \hat{\boldsymbol{\Sigma}}_\alpha^{-1} \mathbf{G}_{g\alpha} \hat{\boldsymbol{\Sigma}}_\alpha^{-1} \hat{\mathbf{C}}^*_\alpha)_g, \tag{2.15}$$

where the right sides of (2.13)–(2.15) *use the current value of* $\hat{\boldsymbol{\beta}}$ *and* $\hat{\boldsymbol{\sigma}}$ *to yield a new value given on the left-hand side. Since* $\mathbf{S}_{12} = \mathbf{0}$ allows us to get new values of $\hat{\boldsymbol{\beta}}$ *and* $\hat{\boldsymbol{\sigma}}$ *sequentially at each iteration, we may revise the value of* $\hat{\mathbf{C}}_\alpha$ *in* (2.14) *with the new value of* $\hat{\boldsymbol{\beta}}$ *before using* (2.15).

Note in (2.15) that the term $[\ ]_{gh}$ is an $m \times m$ matrix and the term $(\ )_g$ is an $m \times 1$ column vector. The gh and g terms are given respectively inside [ ] and ( ). Equations (2.13) and (2.15) follow directly from (2.4) and (2.5). Anderson (1973) used this derivation for the complete data case with $\mathbf{E}_\alpha = \mathbf{I}$, $\alpha = 1, \cdots, q$. The matrices $(\Sigma_1^q n_\alpha \mathbf{Z}'_\alpha \hat{\boldsymbol{\Sigma}}_\alpha^{-1} \mathbf{Z}_\alpha)$ and $\mathbf{Y} = [\Sigma_1^q \operatorname{tr} \hat{\boldsymbol{\Sigma}}_\alpha^{-1} \mathbf{G}_{g\alpha} \hat{\boldsymbol{\Sigma}}_\alpha^{-1} \mathbf{G}_{h\alpha}]_{gh}$ in (2.13) and (2.15) are easily shown to be positive definite for $\hat{\boldsymbol{\Sigma}}$ positive definite. For example, for $\mathbf{x} \neq \mathbf{0}$, $\mathbf{x}'\mathbf{Y}\,\mathbf{x}$ is given by

$$\mathbf{x}'\mathbf{Y}\,\mathbf{x} = \textstyle\sum_{g,h} x_g Y_{gh} x_h = \Sigma_1^q \operatorname{tr} \hat{\boldsymbol{\Sigma}}_\alpha^{-1} \boldsymbol{\Sigma}_\alpha(\mathbf{x}) \hat{\boldsymbol{\Sigma}}_\alpha^{-1} \boldsymbol{\Sigma}_\alpha(\mathbf{x}) = \sum_{\alpha=1}^q \sum_{i,j} (\hat{\boldsymbol{\Sigma}}_\alpha^{-1/2} \boldsymbol{\Sigma}_\alpha(\mathbf{x}) \hat{\boldsymbol{\Sigma}}_\alpha'^{-1/2})_{ij}^2 > 0.$$

The EM algorithm has an E step (conditional expectation) and an M step (maximization). Let $\bar{\mathbf{x}} = (1/N)\Sigma_1^N \mathbf{x}_i$, $\mathbf{A} = \Sigma_i^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$, $N\hat{\mathbf{C}} = \mathbf{A} + N(\bar{\mathbf{x}} - \hat{\boldsymbol{\mu}})(\bar{\mathbf{x}} - \hat{\boldsymbol{\mu}})'$, statistics which we cannot calculate directly from our observed data $\mathbf{E}_1\mathbf{x}_1, \cdots, \mathbf{E}_q\mathbf{x}_N$. For each $\mathbf{E}_\alpha$, $(u_\alpha \times p)$, form a matrix $\mathbf{F}_\alpha$, $((p - u_\alpha) \times p)$ so that $(\mathbf{E}'_\alpha, \mathbf{F}'_\alpha)'$ is of full rank. The E step involves evaluating the two conditional expectations given in equations (2.16) and (2.17) below. The derivation of these equalities is straightforward (e.g. Szatrowski, 1981b).

$$\begin{aligned}&\mathscr{E}(\bar{\mathbf{x}} \mid \mathbf{E}_1\mathbf{x}_1, \cdots, \mathbf{E}_q\mathbf{x}_N; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &\qquad = \frac{1}{N} \Sigma_1^q n_\alpha \begin{pmatrix} \mathbf{E}_\alpha \\ \mathbf{F}_\alpha \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{E}_\alpha \bar{\mathbf{x}}_\alpha \\ \mathbf{F}_\alpha \boldsymbol{\mu} + (\mathbf{F}_\alpha \boldsymbol{\Sigma} \mathbf{E}'_\alpha)(\mathbf{E}_\alpha \boldsymbol{\Sigma} \mathbf{E}'_\alpha)^{-1} \mathbf{E}_\alpha(\bar{\mathbf{x}}_\alpha - \boldsymbol{\mu}) \end{pmatrix}.\end{aligned} \tag{2.16}$$

$$\begin{aligned}&\mathscr{E}(\mathbf{A} \mid \mathbf{E}_1\mathbf{x}_1, \cdots, \mathbf{E}_q\mathbf{x}_N; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &\quad = \Sigma_1^q \begin{pmatrix} \mathbf{E}_\alpha \\ \mathbf{F}_\alpha \end{pmatrix}^{-1} \left\{ \begin{pmatrix} (\boldsymbol{\Xi}_{11})_\alpha (\boldsymbol{\Xi}_{12})_\alpha \\ (\boldsymbol{\Xi}_{21})_\alpha (\boldsymbol{\Xi}_{22})_\alpha \end{pmatrix} + n_\alpha(1 - (n_\alpha/N)) \boldsymbol{\eta}_\alpha \boldsymbol{\eta}'_\alpha \right\} \begin{pmatrix} \mathbf{E}_\alpha \\ \mathbf{F}_\alpha \end{pmatrix}'^{-1} \\ &\qquad - \mathop{\Sigma_1^q \Sigma_1^q}_{\alpha \neq \beta} \frac{n_\alpha n_\beta}{N} \begin{pmatrix} \mathbf{E}_\alpha \\ \mathbf{F}_\alpha \end{pmatrix}^{-1} (\boldsymbol{\eta}_\alpha \boldsymbol{\eta}'_\beta) \begin{pmatrix} \mathbf{E}_\beta \\ \mathbf{F}_\beta \end{pmatrix}'^{-1},\end{aligned} \tag{2.17}$$

$$(\boldsymbol{\Xi}_{11})_\alpha = \mathbf{E}_\alpha \mathbf{A}_\alpha \mathbf{E}'_\alpha, \; (\boldsymbol{\Xi}_{21})'_\alpha = (\boldsymbol{\Xi}_{12})_\alpha = (\mathbf{E}_\alpha \mathbf{A}_\alpha \mathbf{E}'_\alpha)(\mathbf{E}_\alpha \boldsymbol{\Sigma} \mathbf{E}'_\alpha)^{-1}(\mathbf{E}_\alpha \boldsymbol{\Sigma} \mathbf{F}'_\alpha),$$

$$\begin{aligned}(\boldsymbol{\Xi}_{22})_\alpha &= n_\alpha(1 - N^{-1})(\mathbf{F}_\alpha \boldsymbol{\Sigma} \mathbf{F}'_\alpha - (\mathbf{F}_\alpha \boldsymbol{\Sigma} \mathbf{E}'_\alpha)(\mathbf{E}_\alpha \boldsymbol{\Sigma}_\alpha \mathbf{E}'_\alpha)^{-1}(\mathbf{E}_\alpha \boldsymbol{\Sigma} \mathbf{F}'_\alpha)) \\ &\quad + (\mathbf{F}_\alpha \boldsymbol{\Sigma} \mathbf{E}'_\alpha)(\mathbf{E}_\alpha \boldsymbol{\Sigma} \mathbf{E}'_\alpha)^{-1}(\mathbf{E}_\alpha \mathbf{A}_\alpha \mathbf{E}'_\alpha)(\mathbf{E}_\alpha \boldsymbol{\Sigma} \mathbf{E}'_\alpha)^{-1}(\mathbf{E}_\alpha \boldsymbol{\Sigma} \mathbf{F}'_\alpha).\end{aligned}$$

$$\boldsymbol{\eta}_{1\alpha} = (\mathbf{E}_\alpha \bar{\mathbf{x}}_\alpha), \; \boldsymbol{\eta}_{2\alpha} = \mathbf{F}_\alpha \boldsymbol{\mu} + (\mathbf{F}_\alpha \boldsymbol{\Sigma} \mathbf{E}'_\alpha)(\mathbf{E}_\alpha \boldsymbol{\Sigma} \mathbf{E}'_\alpha)^{-1} \mathbf{E}_\alpha(\bar{\mathbf{x}}_\alpha - \boldsymbol{\mu}).$$

LEMMA 2.5. The EM algorithm for finding MLE in the one-population problem is given by (1) *E step*: from an initial estimate of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ finding the conditional expectations (2.16) and (2.17), (2) *M-step*: and using these conditional expectations as complete data in the complete data versions of Lemma 2.3 or Lemma 2.4 (dropping all sums on $\alpha = 1, \ldots, q$ and $\mathbf{E}_\alpha$ in these results.) Each iteration of the EM algorithm involves both an E and M step although more than one iteration may take place in an M step. The M step yields new estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ which are used in the next E step.

Rubin and Szatrowski (1982a, b) suggest that in those cases in which the M-step does not have an explicit noniterative solution, one may sometimes add additional fully missing variables which result in the M-step having an explicit solution.

We next discuss the effect of the starting point on the convergence of the Newton-Raphson, Method of Scoring and EM algorithms for several special cases. These cases indicate the importance of using good initial starting points for the Newton-Raphson and EM algorithms. The Method of Scoring appears to be less sensitive to the choice of starting point. It should be noted that none of these algorithms appear to have guaranteed convergence to a solution of the likelihood equations under all mean and covariance patterns and all missing data configurations. When they do converge to a root of the likelihood equation, this root is not always the MLE. The Method of Scoring has been shown (Szatrowski, 1980) to converge in one iteration to the MLE when an explicit MLE exists from any positive definite covariance starting point. Under certain conditions when the complete-data model is a regular exponential family, the EM algorithm using the Method of Scoring for the M step will have guaranteed convergence to a root of the likelihood equations. (Dempster, Laird and Rubin, 1977).

Consider the problem of estimating $\boldsymbol{\Sigma} = \sigma\, \mathbf{I}_p$ in the complete data one-population problem assuming that the mean vector is known. The Method of Scoring in this case involves the simplification of (2.15)

$$\hat{\boldsymbol{\sigma}} = [\operatorname{tr} \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{G}_g\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{G}_h]_{gh}^{-1}(\operatorname{tr} \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{G}_g\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{C})_g$$

which further simplifies under $\boldsymbol{\Sigma} = \sigma\, \mathbf{I}_p$, with $c = \operatorname{tr} \mathbf{C}/p$, to yield

$$\hat{\sigma}^{(1)} = (\hat{\sigma}^2/p)pc/\hat{\sigma}^2 = c.$$

Thus we see that from any positive definite value of $\hat{\sigma}\, \mathbf{I}_p$, convergence occurs in one iteration. The Newton-Raphson algorithm (Lemma 2.3) for $p = 1$ simplifies to

$$\hat{\sigma}^{(1)} = \hat{\sigma}(1 + (c - \hat{\sigma})/(2c - \hat{\sigma}))$$

with $\hat{\sigma}$ the old value, $\hat{\sigma}^{(1)}$ the new value. Suppose we choose $\hat{\boldsymbol{\Sigma}} = \mathbf{I}$, i.e. $\hat{\sigma} = 1$ as the starting point when c is very large. Then the iterations are $\sigma^{(0)} = 1$, $\sigma^{(1)} \cong 3/2$, $\sigma^{(2)} \cong 9/4$ until the estimates of σ get close to c. This poor performance occurs even when there is an explicit MLE, $\hat{\sigma} = c$. Finally, consider the EM algorithm using the Method of Scoring for the M-step when we view the random sample of scalars $x_1, \cdots, x_N$ as the first components in a bivariate normal distribution with mean $\mathbf{0}$ and covariance $\boldsymbol{\Sigma} = \sigma\, \mathbf{I}_2$. The sufficient statistic for the complete data problem is $\mathbf{C} = (\Sigma_1^N \mathbf{x}_\alpha \mathbf{x}'_\alpha)/N$ where the $\mathbf{x}$'s are 2×1. The E-step yields $\mathscr{E}(\mathbf{C} \mid x_1, \ldots, x_N, \boldsymbol{\Sigma} = \hat{\sigma}\, \mathbf{I}) = \operatorname{diag}(c_{11}, \hat{\sigma})$.

The M-step using the Method of Scoring gives $\sigma^{(1)} = (c_{11} + \hat{\sigma})/2$. The MLE is $\hat{\sigma} = c_{11}$. Thus if c_{11} is very large and we choose $\hat{\boldsymbol{\Sigma}} = \mathbf{I}$, i.e. $\hat{\sigma} = 1$ as a starting point, the iterations are $\sigma^{(0)} = 1$, $\sigma^{(1)} \cong c_{11}/2$, $\sigma^{(2)} \cong 3c_{11}/4$, etc.

These special cases suggest it is important to start the Newton-Raphson and EM algorithms at good starting points for $\boldsymbol{\Sigma}$. The Method of Scoring is less sensitive to the starting point for $\boldsymbol{\Sigma}$. One possible starting point would be to use $\hat{\boldsymbol{\beta}} = \mathbf{0}$ and $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\sigma}})$, i.e. any positive definite $\boldsymbol{\Sigma}$ with the given patterned structure in the Method of Scoring in equations (2.14) and (2.15) to yield a starting point for $\hat{\sigma}$ to be used in any of the algorithms. Since $\mathscr{E}\mathbf{C}^* = \boldsymbol{\Sigma}_\alpha$, $\hat{\boldsymbol{\sigma}}$ would be an unbiased and consistent estimate of $\boldsymbol{\sigma}$. However, $\boldsymbol{\Sigma}(\hat{\boldsymbol{\sigma}})$ is not guaranteed to be positive definite. One can also average elements in the mean and covariance patterns. However, this can lead to a covariance estimate which is not positive definite (e.g. Szatrowski, 1978).

3. Asymptotic distributions of the maximum likelihood estimates. In this section asymptotic distributions of the MLE are given under both the null hypothesis (for any true parameter that is included in the parameterization under which the MLE is derived) and any alternative hypothesis (for any true parameter that is not included in the parameterization under which the MLE is derived). These asymptotic distributions are derived using the standard delta method and facilitate the derivation of the asymptotic nonnull distributions of the LRS given in Section 4. The results under the alternative hypothesis are useful for forming confidence intervals. Anderson's (1973) result, that one iteration of the Method of Scoring algorithm starting from a consistent estimate of $\boldsymbol{\Sigma}$ yields an asymptotic efficient estimate of $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$ in the one-population complete data problem even when the sampled population is not normal as long as the estimates obtained as the solutions to the likelihood equations are asymptotically efficient, is extended to the missing data problem using the Method of Scoring. The asymptotic distributions are given as N, the total sample size, goes to infinity. This concept needs to be clarified in the one-population problem because of the missing data. Roughly stated, we would like the number of data points useful for the estimation of each parameter divided by the total sample size N to have as a limit $N \to \infty$, a number, possibly different for each parameter, that is greater than zero. This condition is stated in Condition 3.1 below after we clarify some notation. Condition 3.1 is assumed for all one-population asymptotic distributions when $\lim_{N\to\infty}$ is stated. Let $1(\cdot)$ be one if the condition in ( ) is true, zero otherwise. Define $N_1(j) = \Sigma_{\alpha=1}^{q} n_\alpha 1(\mathbf{E}_\alpha \mathbf{z}_j \neq \mathbf{0}), j = 1, \cdots, r$, and $N_2(g) = \Sigma_{\alpha=1}^{q} n_\alpha 1(\mathbf{E}_\alpha \mathbf{G}_g \mathbf{E}_\alpha' \neq \mathbf{0}), g = 1, \cdots, m$.

CONDITION 3.1. $\lim_{N\to\infty}(N_s(t)/N) = \eta_{rs} > 0$ for $s = 1, t = 1, \cdots, r$ and $s = 2, t = 1, \cdots, m$.

To facilitate the simplification of matrix derivative expressions, we use the notation of Definition 3.1 and Lemmas 3.1–3.4 given below.

DEFINITION 3.1. (Anderson, 1969). Let $\boldsymbol{\Phi}$ be a $\{p(p+1)/2\} \times \{p(p+1)/2\}$ symmetric matrix with elements $\boldsymbol{\Phi} \equiv \boldsymbol{\Phi}(\boldsymbol{\Sigma}) = (\phi_{ij,kl}) = (\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}), i \leq j, k \leq l$, where σ_{ij} is the ij element of $\boldsymbol{\Sigma}$. The notation $\sigma_{ij,kl}$ 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 $\langle A \rangle'$.

We observe that if the $p \times p$ matrix $\mathbf{W} > 0$ has a Wishart distribution with parameters $\boldsymbol{\Sigma} > \mathbf{0}$ and n ($\mathscr{L}(\mathbf{W}) = \mathscr{W}(\boldsymbol{\Sigma}, n)$), then $n\boldsymbol{\Phi}(\boldsymbol{\Sigma}) = \operatorname{Cov}(\langle \mathbf{W} \rangle)$.

LEMMA 3.1. (Szatrowski, 1979). *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,* $\mathbf{S} > 0$, $\boldsymbol{\Phi}(\mathbf{RSR}') = \mathbf{B}\boldsymbol{\Phi}(\mathbf{S})\mathbf{B}'$.

LEMMA 3.2. (Szatrowski, 1979). *If* $\mathbf{E}$ *and* $\mathbf{F}$ *are* $p \times p$ *symmetric matrices, then*

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

LEMMA 3.3. *If* $\boldsymbol{\Sigma}$ *and* $\boldsymbol{\Sigma}_0$ *are* $p \times p$ *positive definite covariance matrices, then*

$$\boldsymbol{\Phi}(\boldsymbol{\Sigma}_0)\boldsymbol{\Phi}^{-1}(\boldsymbol{\Sigma})\boldsymbol{\Phi}(\boldsymbol{\Sigma}_0) = \boldsymbol{\Phi}(\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0).$$

PROOF. Choose $\mathbf{R}$ so that it simultaneously diagonalizes $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_0$ with the property that $\mathbf{R\Sigma R}' = \mathbf{I}$ and $\mathbf{R\Sigma_0 R}' = \mathbf{D}$, a diagonal matrix with positive diagonal elements. From Lemma 3.1 there exists a matrix $\mathbf{B} = \mathbf{B}(\mathbf{R})$ with the property $\mathbf{B\Phi(S)B}' = \mathbf{\Phi(RSR')}$ for any positive definite $\mathbf{S}$. We then have

$$\begin{aligned}\mathbf{\Phi(\Sigma_0)\Phi^{-1}(\Sigma)\Phi(\Sigma_0)} &= \mathbf{B}^{-1}((\mathbf{B\Phi(\Sigma_0)B}')(\mathbf{B\Phi(\Sigma)B}')^{-1}(\mathbf{B\Phi(\Sigma_0)B}'))\mathbf{B}^{-1\prime}\\ &= \mathbf{B}^{-1}(\mathbf{\Phi(D)\Phi^{-1}(I)\Phi(D)})\mathbf{B}^{-1\prime} = \mathbf{B}^{-1}\mathbf{\Phi(D^2)}\mathbf{B}^{-1\prime} = \mathbf{\Phi}(\mathbf{R}^{-1}\mathbf{D}^2\mathbf{R}^{-1\prime})\\ &= \mathbf{\Phi}((\mathbf{R}^{-1}\mathbf{DR}^{-1\prime})(\mathbf{R'R})(\mathbf{R}^{-1}\mathbf{DR}^{-1\prime})) = \mathbf{\Phi}(\mathbf{\Sigma_0\Sigma^{-1}\Sigma_0}),\end{aligned}$$

with the third equality following by multiplication after noting

$$\mathbf{\Phi(I)} = \operatorname{diag}(2\mathbf{I}_p, \mathbf{I}_{p(p-1)/2}) \quad \text{and} \quad \mathbf{\Phi(D)} = \operatorname{diag}(2d_{11}^2, \cdots, 2d_{pp}^2; d_{12}^2, d_{13}^2, \cdots, d_{p-1,p}^2). \quad \square$$

LEMMA 3.4. *Let $\mathbf{\Sigma}$, $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ be positive definite matrices and let $\mathbf{E}$ and $\mathbf{F}$ be symmetric matrices, all of dimension $p \times p$. Then*

$$\langle\mathbf{E}\rangle' \, \mathbf{\Phi}^{-1}(\mathbf{\Sigma}_1)\mathbf{\Phi}(\mathbf{\Sigma})\mathbf{\Phi}^{-1}(\mathbf{\Sigma}_2)\langle\mathbf{F}\rangle = \tfrac{1}{2} \operatorname{tr} \mathbf{E\Sigma}_1^{-1}\mathbf{\Sigma\Sigma}_2^{-1}\mathbf{F\Sigma}_2^{-1}\mathbf{\Sigma\Sigma}_1^{-1}. \tag{3.1}$$

PROOF. Using Lemma 3.1, we see that for $\mathbf{R}$ nonsingular, (3.1) is invariant under the transformation $(\mathbf{E}, \mathbf{F}, \mathbf{\Sigma}_1, \mathbf{\Sigma}, \mathbf{\Sigma}_2) \to (\mathbf{RER}', \mathbf{RFR}', \mathbf{R\Sigma_1R}', \mathbf{R\Sigma R}', \mathbf{R\Sigma_2R}')$.

Since $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are positive definite, we can choose $\mathbf{R}$ so $\mathbf{R\Sigma_2R}' = \mathbf{I}$ and $\mathbf{R\Sigma_1R}' = \operatorname{diag}(d_1, \cdots, d_p) = \mathbf{D}$, and since $\mathbf{E}$ and $\mathbf{F}$ and thus $\mathbf{RER}'$ and $\mathbf{RFR}'$ are symmetric matrices, it is sufficient to show for $\mathbf{E}$ and $\mathbf{F}$ symmetric, $\mathbf{\Sigma}$ and $\mathbf{D}$ positive definite with $\mathbf{D}$ diagonal that

$$\langle\mathbf{E}\rangle'\mathbf{\Phi}^{-1}(\mathbf{D})\mathbf{\Phi}(\mathbf{\Sigma})\mathbf{\Phi}^{-1}(\mathbf{I})\langle\mathbf{F}\rangle = \tfrac{1}{2} \operatorname{tr} \mathbf{D}^{-1}\mathbf{ED}^{-1}\mathbf{\Sigma F\Sigma}. \tag{3.2}$$

Let $\mathbf{J}_{ij}$ be a symmetric matrix of zeroes with a one in the ij and ji positions. By linearity, it suffices to prove (3.2) for $\mathbf{E} = \mathbf{J}_{ij}$ and $\mathbf{F} = \mathbf{J}_{kl}$, $1 \le i \le j \le p$ and $1 \le k \le l \le p$. We consider four cases: (1) $i = j$, $k = l$, (2) $i = j$, $k \ne l$, (3) $i \ne j$, $k = l$, and (4) $i \ne j$, $k \ne l$.

The right hand side of (3.2) with $E = J_{ij}$ and $F = J_{kl}$ simplifies noting $\mathbf{D}$ is diagonal to:

$$\tfrac{1}{2} \operatorname{tr} \mathbf{D}^{-1}\mathbf{J}_{ij}\mathbf{D}^{-1}\mathbf{\Sigma J}_{kl}\mathbf{\Sigma} = \Sigma_{r,s,t,u}\, d_r^{-1}(J_{ij})_{rs}\, d_s^{-1}\Sigma_{st}(J_{kl})_{tu}\Sigma_{ur}. \tag{3.3}$$

This further simplifies in the four cases to (1) $\tfrac{1}{2}\, d_i^{-2}\sigma_{ik}^2$ for $i = j$, $k = l$; (2) $d_i^{-2}\sigma_{ik}\sigma_{il}$ for $i = j$, $k \ne l$; (3) $d_i^{-1}\, d_j^{-1}\sigma_{ik}\sigma_{jk}$ for $i \ne j$, $k = l$; and (4) $d_i^{-1}\, d_j^{-1}(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk})$ for $i \ne j$, $k \ne l$.

The left hand side simplifies, noting $\mathbf{\Phi(D)}$ and $\mathbf{\Phi(I)}$ are diagonal with $\mathbf{\Phi(I)} = \operatorname{diag}(2\mathbf{I}_p, \mathbf{I}_{p(p-1)/2})$. If we decompose $\mathbf{\Phi}$ into blocks $\mathbf{\Phi}_{ij}$, $i, j = 1, 2$, with $\mathbf{\Phi}_{11}$, $p \times p$ then $\mathbf{\Phi}^{-1}(\mathbf{D})\mathbf{\Phi}(\mathbf{\Sigma})\mathbf{\Phi}^{-1}(\mathbf{I})$ is given by

$$\mathbf{\Phi}^{-1}(\mathbf{D})\Phi(\mathbf{\Sigma})\mathbf{\Phi}^{-1}(\mathbf{I}) = \begin{bmatrix} \tfrac{1}{4}\, \mathbf{D}^{-2}\mathbf{\Phi}_{11}(\mathbf{\Sigma}) & \tfrac{1}{2}\, \mathbf{D}^{-2}\mathbf{\Phi}_{12}(\mathbf{\Sigma}) \\ \tfrac{1}{2}\, \mathbf{\Phi}_{22}^{-1}(\mathbf{D})\mathbf{\Phi}_{21}(\mathbf{\Sigma}) & \mathbf{\Phi}_{22}^{-1}(\mathbf{D})\mathbf{\Phi}_{22}(\mathbf{\Sigma}) \end{bmatrix}. \tag{3.4}$$

The verification that the right and left sides of (3.2) are equal follows directly for each of the four cases using the results in (3.4), the definition of $\mathbf{\Phi}$ and the right hand side results given after (3.3). $\square$

The asymptotic distribution of the MLE derived under the null hypothesis assumptions when the true value of the parameters, $(\boldsymbol{\mu}^*, \mathbf{\Sigma}^*)$, does not necessarily belong to the parameter space of the null hypothesis is given in Theorem 3.1. These results are used in this paper when the asymptotic nonnull distribution of the LRS is derived. Theorem 3.1 also allows us to study the asymptotic behavior of the MLE derived under a reduced model assumed under the null hypothesis when a model under the alternative hypothesis is the true model. Thus, one can study the robustness of estimates, confidence intervals, etc. when the null hypothesis modelling assumption is violated.

THEOREM 3.1. *The asymptotic joint distribution of the* MLE *derived under the null and alternative hypotheses in the one-population missing data problem evaluated at the* *true parameter value* (μ^*, Σ^*) *where* μ^* *and* Σ^* *need not be patterned is given by* $(r, s = 0, 1)$,

$$\lim_{N\to\infty}\mathscr{L}\{N^{1/2}((\hat{\beta}_0', \hat{\sigma}_0', \hat{\beta}_1', \hat{\sigma}_1')' - (\beta_0', \sigma_0', \beta_1', \sigma_1')')\} = \mathscr{N}(\mathbf{0}, \mathbf{\Xi}), \tag{3.5}$$

$$\mathbf{\Xi}(\hat{\sigma}_r, \hat{\sigma}_s) = 2(\mathbf{I} - \mathbf{F}_r)^{-1}\mathbf{Y}_r^{-1}(2\mathbf{H}_{rs} + \mathbf{J}_{rs})\mathbf{Y}_s^{-1}(\mathbf{I} - \mathbf{F}_s)'^{-1}, \tag{3.6}$$

$$\mathbf{\Xi}(\hat{\beta}_r, \hat{\sigma}_s) = \mathbf{Q}_r\mathbf{\Xi}(\hat{\sigma}_r, \hat{\sigma}_s) + 2\mathbf{T}_{rs}\mathbf{Y}_s^{-1}(\mathbf{I} - \mathbf{F}_s)'^{-1}, \tag{3.7}$$

$$\begin{aligned}\mathbf{\Xi}(\hat{\beta}_r, \hat{\beta}_s) = {} & \mathbf{Q}_r\mathbf{\Xi}(\hat{\sigma}_r, \hat{\sigma}_s)\mathbf{Q}_s' + \mathbf{S}_r^{-1}\mathbf{L}_{rs}\mathbf{S}_s^{-1} \\ & + 2(\mathbf{T}_{rs}\mathbf{Y}_s^{-1}(\mathbf{I} - \mathbf{F}_s)'^{-1}\mathbf{Q}_s' + \mathbf{Q}_r(\mathbf{I} - \mathbf{F}_r)^{-1}\mathbf{Y}_r^{-1}\mathbf{T}_{sr}'),\end{aligned} \tag{3.8}$$

with

$$\begin{aligned}\mathbf{F}_r = 2\mathbf{Y}_r^{-1} \textstyle\sum_{\alpha=1}^q f_\alpha[\operatorname{tr} & \mathbf{\Sigma}_{r\alpha}^{-1}\mathbf{G}_{g\alpha}\mathbf{\Sigma}_{r\alpha}^{-1}(\mathbf{G}_{h\alpha}\mathbf{\Sigma}_{r\alpha}^{-1}(\mathbf{\Sigma}_{r\alpha} - \mathbf{\Sigma}_\alpha^* - \mathbf{B}_{r\alpha\alpha}) \\ & + \mathbf{Z}_{r\alpha}\mathbf{S}_r^{-1}(\textstyle\sum_{\gamma=1}^q f_\gamma \mathbf{Z}_{r\gamma}'\mathbf{\Sigma}_{r\gamma}^{-1}\mathbf{G}_{h\gamma}\mathbf{\Sigma}_{r\gamma}^{-1}\mathbf{B}_{r\gamma\alpha}))]_{gh},\end{aligned} \tag{3.9}$$

$$\mathbf{Y}_r = [\textstyle\sum_{\alpha=1}^q f_\alpha \operatorname{tr} \mathbf{\Sigma}_{r\alpha}^{-1}\mathbf{G}_{g\alpha}\mathbf{\Sigma}_{r\alpha}^{-1}\mathbf{G}_{h\alpha}]_{gh}, \quad \mathbf{B}_{r\gamma\alpha} = (\mu_\gamma^* - \mu_{r\gamma})(\mu_\alpha^* - \mu_{r\alpha})', \tag{3.10}$$

$$\mathbf{S}_r = \textstyle\sum_{\alpha=1}^q f_\alpha \mathbf{Z}_{r\alpha}'\mathbf{\Sigma}_{r\alpha}^{-1}\mathbf{Z}_{r\alpha}, \quad \mathbf{H}_{rs} = \textstyle\sum_{\alpha=1}^q f_\alpha[\mathbf{u}_{rg\alpha}\mathbf{\Sigma}_{r\alpha}^{-1}\mathbf{\Sigma}_\alpha^*\mathbf{\Sigma}_{s\alpha}^{-1}\mathbf{u}_{sh\alpha}']_{gh}, \tag{3.11}$$

$$\mathbf{u}_{rg\alpha} = ((\mu_\alpha^* - \mu_{r\alpha})'\mathbf{\Sigma}_{r\alpha}^{-1}\mathbf{G}_{g\alpha} - \textstyle\sum_{\gamma=1}^q f_\gamma(\mu_\gamma^* - \mu_{r\gamma})'\mathbf{\Sigma}_{r\gamma}^{-1}\mathbf{G}_{g\gamma}\mathbf{\Sigma}_{r\gamma}^{-1}\mathbf{Z}_{r\gamma}\mathbf{S}_r^{-1}\mathbf{Z}_{r\alpha}')_g, \tag{3.12}$$

$$\mathbf{J}_{rs} = \textstyle\sum_{\alpha=1}^q f_\alpha[\operatorname{tr} \mathbf{G}_{g\alpha}\mathbf{\Sigma}_{r\alpha}^{-1}\mathbf{\Sigma}_\alpha^*\mathbf{\Sigma}_{s\alpha}^{-1}\mathbf{G}_{h\alpha}\mathbf{\Sigma}_{s\alpha}^{-1}\mathbf{\Sigma}_\alpha^*\mathbf{\Sigma}_{r\alpha}^{-1}]_{gh}, \tag{3.13}$$

$$\mathbf{Q}_r = \mathbf{S}_r^{-1}(\textstyle\sum_{\alpha=1}^q f_\alpha \mathbf{Z}_{r\alpha}'\mathbf{\Sigma}_{r\alpha}^{-1}\mathbf{G}_{g\alpha}\mathbf{\Sigma}_{r\alpha}^{-1}(\mu_{r\alpha} - \mu_\alpha^*))_g, \tag{3.14}$$

$$\mathbf{T}_{rs} = (\mathbf{S}_r^{-1}\textstyle\sum_{\alpha=1}^q f_\alpha \mathbf{Z}_{r\alpha}'\mathbf{\Sigma}_{r\alpha}^{-1}\mathbf{\Sigma}_\alpha^*\mathbf{\Sigma}_{s\alpha}^{-1}\mathbf{u}_{sh\alpha}')_h, \quad \mathbf{L}_{rs} = \textstyle\sum_{\alpha=1}^q f_\alpha \mathbf{Z}_{r\alpha}'\mathbf{\Sigma}_{r\alpha}^{-1}\mathbf{\Sigma}_\alpha^*\mathbf{\Sigma}_{s\alpha}^{-1}\mathbf{Z}_{s\alpha}, \tag{3.15}$$

$(g, h = 1, \cdots, m_r$ in $\mathbf{F}_r, \mathbf{Y}_r, \mathbf{Q}_r, \mathbf{u}_{rg\alpha}; g = 1, \cdots, m_r, h = 1, \cdots, m_s$ *in* $\mathbf{J}_{rs}; m_1 \equiv m)$, *where* (β_r, σ_r) *are the* "MLE" *derived under* H_r *with* $\bar{\mathbf{x}}_\alpha^*$ *replaced by* μ_α^* *and* $\mathbf{A}_\alpha^*$ *replaced by* $n_\alpha\mathbf{\Sigma}_\alpha^*$. *These* "MLE" *are not statistics. Also* $\mu_\alpha^* = \mathbf{E}_\alpha\mu^*$ *and* $\mathbf{\Sigma}_\alpha^* = \mathbf{E}_\alpha\mathbf{\Sigma}^*\mathbf{E}_\alpha'$ *and* $\mathbf{Z}_{r\alpha} = \mathbf{E}_\alpha\mathbf{Z}_r$ *where* $\mathbf{Z}_0$ *is defined in* (1.1) *and* $\mathbf{Z}_1$ *is* $\mathbf{Z}$ *in* (1.1).

PROOF. Complete details of this proof may be found in Szatrowski (1981b). We use the form of the standard multivariate delta method given in Bishop, Fienberg, and Holland, (1975), Theorem 14.6.2. To use the delta method theorem, we must find $\partial f/\partial\theta_i$. We use the equations in Lemma 2.4. Note that $\hat{\mathbf{\Sigma}} = \mathbf{\Sigma}(\hat{\sigma})$. Let $\theta_{i\eta}$ be one of the elements of $(\bar{\mathbf{x}}_\eta^{*\prime}, \langle\mathbf{A}_\eta^*\rangle') \equiv ((\mathbf{E}_\eta\bar{\mathbf{x}}_\eta)', \langle\mathbf{E}_\eta\mathbf{A}_\eta\mathbf{E}_\eta'\rangle')$. Taking derivatives with respect to $\theta_{i\eta}$, we get a set of linear equations of the form

$$(\partial\hat{\beta}/\partial\theta_{i\eta}) = \mathbf{p}_{i\eta} + \mathbf{Q}(\partial\hat{\sigma}/\partial\theta_{i\eta}), \quad (\partial\hat{\sigma}/\partial\theta_{i\eta}) = \mathbf{m}_{i\eta} + \mathbf{F}(\partial\hat{\sigma}/\partial\theta_{i\eta}). \tag{3.16}$$

The second set of these equations can be solved yielding

$$(\partial\hat{\sigma}/\partial\theta_{i\eta}) = (\mathbf{I} - \mathbf{F})^{-1}\mathbf{m}_{i\eta}, \quad (\partial\hat{\beta}/\partial\theta_{i\eta}) = \mathbf{p}_{i\eta} + \mathbf{Q}(\mathbf{I} - \mathbf{F})^{-1}\mathbf{m}_{i\eta}. \tag{3.17}$$

We then rewrite these equations in terms of matrices $\mathbf{M}$ and $\mathbf{P}$ with columns given by $\mathbf{m}_{i\eta}$ and $\mathbf{p}_{i\eta}$ respectively, evaluating separately the cases when $\theta_{i\eta}$ is an element of $\bar{\mathbf{x}}_\eta^*$ and $\langle\mathbf{A}_\eta^*\rangle$ and then evaluate these expressions at the value $\bar{\mathbf{x}}_\alpha^* = \mathbf{u}_\alpha^*$, $\mathbf{A}_\alpha^* = n_\alpha\mathbf{\Sigma}_\alpha^*$, after replacing μ, β, $\mathbf{\Sigma}$ and σ with μ_r, β_r, $\mathbf{\Sigma}_r$ and σ_r respectively, $r = 0, 1$. These results are then substituted into the following results on the form of the asymptotic covariance derived from the multivariate delta method and the asymptotic multivariate distribution of the sample mean and covariance (e.g., Theorem 2, Szatrowski, 1979), with the understanding that all derivatives are evaluated at the parameter value $(\mu^*, \mathbf{\Sigma}^*)$,

$$\begin{aligned}\operatorname{Cov}(\hat{\beta}_r, \hat{\beta}_s) = \Sigma_1^q\{ & (\partial\hat{\beta}_r/\partial\bar{\mathbf{x}}_\eta^*)(\mathbf{\Sigma}_\eta^*/n_\eta)(\partial\hat{\beta}_s/\partial\bar{\mathbf{x}}_\eta^*)' \\ & + (\partial\hat{\beta}_r/\partial\langle\mathbf{A}_\eta^*\rangle)(n_\eta\mathbf{\Phi}(\mathbf{\Sigma}_\eta^*))(\partial\hat{\beta}_s/\partial\langle\mathbf{A}_\eta^*\rangle)'\},\end{aligned} \tag{3.18}$$

$$\mathrm{Cov}(\hat{\sigma}_r, \hat{\sigma}_s) = \Sigma_1^q \{(\partial\hat{\boldsymbol{\sigma}}_r/\partial\bar{\mathbf{x}}_\eta^*)(\boldsymbol{\Sigma}_\eta^*/n_\eta)(\partial\hat{\boldsymbol{\sigma}}_s/\partial\bar{\mathbf{x}}_\eta^*)' + (\partial\hat{\boldsymbol{\sigma}}_r/\partial\langle\mathbf{A}_\eta^*\rangle)(n_\eta\boldsymbol{\Phi}(\boldsymbol{\Sigma}_\eta^*))(\partial\hat{\boldsymbol{\sigma}}_s/\partial\langle\mathbf{A}_\eta^*\rangle)'\}, \tag{3.19}$$

$$\mathrm{Cov}(\hat{\beta}_r, \hat{\sigma}_s) = \Sigma_1^q \{(\partial\hat{\boldsymbol{\beta}}_r/\partial\bar{\mathbf{x}}_\eta^*)(\boldsymbol{\Sigma}_\eta^*/n_\eta)(\partial\hat{\boldsymbol{\sigma}}_s/\partial\bar{\mathbf{x}}_\eta^*)' + (\partial\hat{\boldsymbol{\beta}}_r/\partial\langle\mathbf{A}_\eta^*\rangle)(n_\eta\boldsymbol{\Phi}(\boldsymbol{\Sigma}_\eta^*))(\partial\hat{\boldsymbol{\sigma}}_s/\partial\langle\mathbf{A}_\eta^*\rangle)'\}. \square \tag{3.20}$$

A simplified version of Theorem 3.1 for the complete data problem is given by Theorem 3.2.

THEOREM 3.2. *Under the conditions of Theorem* 3.1 *with complete data, the asymptotic joint distributions of* $(\hat{\boldsymbol{\beta}}_0', \hat{\boldsymbol{\sigma}}_0', \hat{\boldsymbol{\beta}}_1', \hat{\boldsymbol{\sigma}}_1')'$ *is given by* (3.5)–(3.8) *with* (3.9)–(3.15) *replaced by*

$$\mathbf{F}_r = 2\mathbf{Y}_r^{-1}[\mathrm{tr}\,\boldsymbol{\Sigma}_r^{-1}\mathbf{G}_g\boldsymbol{\Sigma}_r^{-1}(\mathbf{G}_h\boldsymbol{\Sigma}_r^{-1}(\boldsymbol{\Sigma}_r - \boldsymbol{\Sigma}^* - \mathbf{B}_r) + \mathbf{R}_r\boldsymbol{\Sigma}_r^{-1}\mathbf{G}_h\boldsymbol{\Sigma}_r^{-1}\mathbf{B}_r)]_{gh}, \tag{3.21}$$

$$\mathbf{Y}_r = [\mathrm{tr}\,\boldsymbol{\Sigma}_r^{-1}\mathbf{G}_g\boldsymbol{\Sigma}_r^{-1}\mathbf{G}_h]_{gh}, \quad \mathbf{B}_r = (\boldsymbol{\mu}^* - \boldsymbol{\mu}_r)(\boldsymbol{\mu}^* - \boldsymbol{\mu}_r)', \quad \mathbf{S}_r = \mathbf{Z}_r'\boldsymbol{\Sigma}_r^{-1}\mathbf{Z}_r, \tag{3.22}$$

$$\mathbf{H}_{rs} = [\mathbf{u}_{rg}\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}_s^{-1}\mathbf{u}_{sh}']_{gh}, \quad \mathbf{R}_r = \mathbf{Z}_r(\mathbf{Z}_r'\boldsymbol{\Sigma}_r^{-1}\mathbf{Z}_r)^{-1}\mathbf{Z}_r', \tag{3.23}$$

$$\mathbf{u}_{rg} = (\boldsymbol{\mu}^* - \boldsymbol{\mu}_r)'\boldsymbol{\Sigma}_r^{-1}\mathbf{G}_g(\mathbf{I} - (\mathbf{R}_r\boldsymbol{\Sigma}_r^{-1})'), \tag{3.24}$$

$$\mathbf{J}_{rs} = [\mathrm{tr}\,\mathbf{G}_g\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}_s^{-1}\mathbf{G}_h\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}_r^{-1}]_{gh}, \tag{3.25}$$

$$\mathbf{Q}_r = (\mathbf{S}_r^{-1}\mathbf{Z}_r'\boldsymbol{\Sigma}_r^{-1}\mathbf{G}_g\boldsymbol{\Sigma}_r^{-1}(\boldsymbol{\mu}_r - \boldsymbol{\mu}^*))_g, \tag{3.26}$$

$$\mathbf{T}_{rs} = (\mathbf{S}_r^{-1}\mathbf{Z}_r'\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}_s^{-1}\mathbf{u}_{rg}')_g, \quad \mathbf{L}_{rs} = \mathbf{Z}_r'\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}_s^{-1}\mathbf{Z}_s. \tag{3.27}$$

Under the additional assumption that the MLE *have explicit representations under the null hypothesis (e.g., Szatrowski,* 1980), *the form of the asymptotic covariance* $\boldsymbol{\Xi}$ *simplifies, becoming*

$$\boldsymbol{\Xi}(\hat{\boldsymbol{\sigma}}_r, \hat{\boldsymbol{\sigma}}_s) = 2\mathbf{Y}_r^{-1}(2\mathbf{H}_{rs} + \mathbf{J}_{rs})\mathbf{Y}_s^{-1},$$

$$\boldsymbol{\Xi}(\hat{\boldsymbol{\beta}}_r, \hat{\boldsymbol{\sigma}}_s) = 2\mathbf{T}_{rs}\mathbf{Y}_s^{-1}, \quad \boldsymbol{\Xi}(\hat{\boldsymbol{\beta}}_r, \hat{\boldsymbol{\beta}}_s) = \mathbf{S}_r^{-1}\mathbf{Z}_r'(\boldsymbol{\Sigma}_r\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_s)^{-1}\mathbf{Z}_s\mathbf{S}_s^{-1}.$$

PROOF. In the complete data case we can drop all $\sum_1^q$, all subscripts α, η, γ, and all f_α become ones. When there is an explicit solution, further simplification follows after noting $\mathbf{Q}_0$ and $\mathbf{F}_0$ are zero matrices. □

Finally, we extend Anderson's (1973) results on asymptotic efficient estimates obtained from one iteration of the Method of Scoring algorithm in the one-population complete data problem to the one-population missing data problem. We parallel Anderson's (1973) Section 3 presentation. Note that standard asymptotic distribution results for MLE yield the following Lemma.

LEMMA 3.5. *The asymptotic distributions for the* MLE *of* $\hat{\boldsymbol{\beta}}$ *and* $\hat{\boldsymbol{\sigma}}$ *in the one-population missing data problem evaluated at a value* $\boldsymbol{\beta}$ *and* $\boldsymbol{\sigma}$ *in the null hypothesis region are independent with marginal distributions given by*

$$\lim_{N\to\infty}\mathscr{L}(N^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})) = \mathscr{N}(\mathbf{0}, (\Sigma_1^q n_\alpha \mathbf{Z}_\alpha'\boldsymbol{\Sigma}_\alpha^{-1}\mathbf{Z}_\alpha)^{-1}), \tag{3.28}$$

$$\lim_{N\to\infty}\mathscr{L}(N^{1/2}(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})) = \mathscr{N}(\mathbf{0}, [\tfrac{1}{2}\,\Sigma_1^q n_\alpha \mathrm{tr}\,\boldsymbol{\Sigma}_\alpha^{-1}\mathbf{G}_{g\alpha}\boldsymbol{\Sigma}_\alpha^{-1}\mathbf{G}_{h\alpha}]^{-1}). \tag{3.29}$$

Let $\hat{\boldsymbol{\beta}}(N)$ be given by (2.13) with $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}$ on the right hand side. Then $N^{1/2}(\hat{\boldsymbol{\beta}}(N) - \boldsymbol{\beta})$ has a limiting normal distribution given in (3.28) even when the population distribution is not normal assuming the population mean and covariance exist.

THEOREM 3.3. *If the data in the one-population missing data problem come from a population whose distribution is not necessarily multivariate normal with mean* $\boldsymbol{\mu} = \mathbf{Z}\boldsymbol{\beta}$ *and covariance* $\boldsymbol{\Sigma}$, *and* $\hat{\boldsymbol{\Sigma}}(N)$ *is a consistent estimate of* $\boldsymbol{\Sigma}$, then if $\hat{\boldsymbol{\beta}}^*(N)$ *is given by* (2.13) *with* $\hat{\boldsymbol{\Sigma}}$ *a consistent estimate of* $\boldsymbol{\Sigma}$, $N^{1/2}(\hat{\boldsymbol{\beta}}^*(N) - \boldsymbol{\beta})$ *has the limiting normal distribution given in* (3.28). *If* $\hat{\boldsymbol{\beta}}(N)$ *is asymptotically efficient, so is* $\hat{\boldsymbol{\beta}}^*(N)$.

PROOF.

$$N^{1/2}(\hat{\boldsymbol{\beta}}^*(N) - \hat{\boldsymbol{\beta}}(N)) = N^{1/2}\{\hat{\boldsymbol{\beta}}^*(N) - \boldsymbol{\beta} - (\hat{\boldsymbol{\beta}}(N) - \boldsymbol{\beta})\}$$
$$= \textstyle\sum_{\alpha=1}^{q} \{(\sum_{\gamma=1}^{q} f_\gamma \mathbf{Z}_\gamma' \hat{\boldsymbol{\Sigma}}_\gamma^{-1} Z_\gamma)^{-1}\mathbf{Z}_\alpha'\hat{\boldsymbol{\Sigma}}_\alpha^{-1}(N) - (\sum_{\gamma=1}^{q} f_\gamma \mathbf{Z}_\gamma' \boldsymbol{\Sigma}_\gamma^{-1}\mathbf{Z}_\gamma)^{-1}\mathbf{Z}_\alpha'\boldsymbol{\Sigma}_\alpha^{-1}(N)\} f_\alpha N^{1/2}\{\bar{\mathbf{x}}_\alpha - \mathbf{Z}_\alpha\boldsymbol{\beta}\}$$

converges stochastically to $\mathbf{0}$, where $f_\alpha = n_\alpha/N$ because for $\alpha = 1, \cdots, q$,

$$\operatorname{plim}_{N\to\infty}(\textstyle\sum_{\gamma=1}^{q} f_\gamma \mathbf{Z}_\gamma'\hat{\boldsymbol{\Sigma}}_\gamma^{-1}(N)\mathbf{Z}_\gamma)^{-1}\mathbf{Z}_\alpha'\hat{\boldsymbol{\Sigma}}_\alpha^{-1}(N) = (\sum_{\gamma=1}^{q} f_\gamma \mathbf{Z}_\gamma'\boldsymbol{\Sigma}_\gamma^{-1}\mathbf{Z}_\gamma)^{-1}\mathbf{Z}_\alpha'\boldsymbol{\Sigma}_\alpha^{-1},$$

and $f_\alpha N^{1/2}\{\bar{\mathbf{x}}_\alpha(N) - \mathbf{Z}_\alpha\boldsymbol{\beta}\} = f_\alpha^{1/2} n_\alpha^{1/2}\{\bar{\mathbf{x}}_\alpha(N) - \mathbf{Z}_\alpha\boldsymbol{\beta}\}$ is either a random variable that converges stochastically to zero or to a limiting normal distribution when either $n_\alpha/N \to 0$ or $n_\alpha/N \to \eta_\alpha > 0$. □

Using the notation of Anderson (1969, 1973) or Szatrowski (1979, 1980), we can write the scoring equations for $\hat{\boldsymbol{\sigma}}$ in (2.15) in the same form as the equation for $\hat{\boldsymbol{\beta}}$ in (2.13), and thus generalize the results of Theorem 3.3 for covariances.

THEOREM 3.4. *If the data in the one-population missing data problem come from a population with a linearly patterned population mean and covariance, then one iteration of the Method of Scoring algorithms given in Section* 2 *from any consistent estimate of the population covariance matrix yields estimates with asymptotic normal distributions given in* (3.28) *and* (3.29). *If the estimates derived as solutions to the scoring equations are asymptotically efficient, then the one iteration solution from a consistent estimate of the covariance matrix are also asymptotically efficient.*

We note that when the estimates derived as solutions to the scoring equations are MLE, such as when the sampled population is normal, then these estimates are asymptotically efficient in the sense of attaining the Cramér-Rao lower bound for the covariance matrix of unbiased estimates.

4. Likelihood ratio statistic and its asymptotic distributions. In this section, the form of the LRS and its asymptotic distributions are given. The asymptotic null distributions given are the usual asymptotic Chi squared distributions for LRS. The likelihood ratio statistic, λ, for testing the null hypothesis $H_0\colon \boldsymbol{\beta}_1 = \mathbf{0}$, $\boldsymbol{\sigma}_1 = \mathbf{0}$ against the alternative hypothesis H_1 which does not so restrict $\boldsymbol{\beta}_1$ and $\boldsymbol{\sigma}_1$ is easily shown by substitution of the MLE into the likelihood function to be given by

$$\lambda^{(2/N)} = \textstyle\prod_{\alpha=1}^{q} \{|\mathbf{E}_\alpha\tilde{\boldsymbol{\Sigma}}_1\mathbf{E}_\alpha'|/|\mathbf{E}_\alpha\tilde{\boldsymbol{\Sigma}}_0\mathbf{E}_\alpha'|\}^{f_\alpha}, \tag{4.1}$$

where $\tilde{\boldsymbol{\Sigma}}_0$ and $\tilde{\boldsymbol{\Sigma}}_1$ are the MLE of $\boldsymbol{\Sigma}$ under H_0 and H_1 respectively. Methods used to find these MLE are discussed in Section 2.

In general, the exact distribution of the LRS is difficult to derive. Often no explicit form of the LRS exists. The usual asymptotic Chi squared distribution applies under the null hypothesis assumption, yielding $\lim_{N\to\infty}\mathscr{L}(-2 \log \lambda) = \chi_f^2$, where $f = m + r - (m_0 + r_0)$. We reject the null hypothesis when $-2 \log \lambda$ is too large.

Finally, asymptotic nonnull distributions of the LRS are given using the standard delta method for values of the true parameter which are not in the null hypothesis region. The asymptotic null distribution results cannot be obtained from the nonnull results by assuming the true parameter is in the null hypothesis region. Such parameter values yield zero variance values below since the standard delta method first order derivative is zero under the null hypothesis. The values that we consider for the true parameter are not in the null hypothesis region, but are either in the alternative hypothesis region or in a parameter region that may include the null and alternative hypothesis regions. These results are useful for power and sample size calculations and for studying the behavior of the LRS for true parameter values that are neither in the null or alternative hypothesis region. For example, one may wish to know the distribution of the LRS in the one-population problem at a true value $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ which may have no patterned structure when testing a null and alternative hypothesis which are both patterned.

The asymptotic nonnull results in this section follow immediately using the standard delta method and the Section 3 asymptotic distribution of the MLE. They simplify to the results for the complete data case given by Szatrowski (1979) with the advantage that the simplified results in the present paper are in terms of $\boldsymbol{\Sigma}$ rather then $\boldsymbol{\Phi}(\boldsymbol{\Sigma})$, thus allowing for a simplification of calculations.

THEOREM 4.1. *The asymptotic nonnull distribution for the* LRS (4.1) *in the one population missing data hypothesis testing problem is given by*

$$\lim_{N\to\infty}\mathscr{L}[N^{1/2}\{-(2/N)\log\lambda - \textstyle\sum_{\alpha=1}^{q} f_\alpha \log(|\boldsymbol{\Sigma}_{0\alpha}|/|\boldsymbol{\Sigma}_{1\alpha}|)\}] = \mathscr{N}(0, v_\infty), \tag{4.2}$$

$$v_\infty = \textstyle\sum_{r,s=0}^{1}(-1)^{r+s}\sum_{\alpha,\gamma=1}^{q} f_\alpha f_\gamma \sum_{g=1}^{m_r}\sum_{h=1}^{m_s}(\operatorname{tr}\boldsymbol{\Sigma}_{r\alpha}^{-1}\mathbf{G}_{g\alpha})(\boldsymbol{\Xi}(\hat{\boldsymbol{\sigma}}_r, \hat{\boldsymbol{\sigma}}_s))_{gh}(\operatorname{tr}\boldsymbol{\Sigma}_{s\gamma}^{-1}\mathbf{G}_{h\gamma}), \tag{4.3}$$

where $\boldsymbol{\Xi}(\hat{\boldsymbol{\sigma}}_r, \hat{\boldsymbol{\sigma}}_s)$ *is given in* (3.6) *in Theorem* 3.1, *under the assumptions of Theorem* 3.1 *with the additional assumption that the true parameter* $(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ *does not lie in the null hypothesis region.*

PROOF. We wish to find the asymptotic distribution of $f(\hat{\boldsymbol{\sigma}}_0, \hat{\boldsymbol{\sigma}}_1) = -(2/N)\log\lambda$ using the standard delta method. Taking derivatives (e.g. Szatrowski, 1979) yields

$$\partial f/\partial(\hat{\sigma}_r)_g = (-1)^r \textstyle\sum_{\alpha=1}^{q} f_\alpha \operatorname{tr}\boldsymbol{\Sigma}_{r\alpha}^{-1}\mathbf{G}_{g\alpha}.$$

The variance term is

$$v_\infty = \Sigma_{r,s=0}^{1}(\partial f/\partial\hat{\boldsymbol{\sigma}}_r)'\boldsymbol{\Xi}(\hat{\boldsymbol{\sigma}}_r, \hat{\boldsymbol{\sigma}}_s)(\partial f/\partial\hat{\boldsymbol{\sigma}}_s). \qquad \square$$

Theorem 4.1 can be further simplified by making the usual assumption that $(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ belongs in the alternative hypothesis region. This greatly simplifies the form of $\boldsymbol{\Xi}(\hat{\boldsymbol{\sigma}}_r, \hat{\boldsymbol{\sigma}}_s)$ when r or s is equal to one since $\boldsymbol{\mu}_{1\alpha} = \boldsymbol{\mu}_\alpha^*$ and $\boldsymbol{\Sigma}_{1\alpha} = \boldsymbol{\Sigma}_\alpha^*$. The complete data case follows from Theorem 4.1 by using $\boldsymbol{\Xi}(\hat{\boldsymbol{\sigma}}_r, \hat{\boldsymbol{\sigma}}_s)$ from Theorem 3.2, replacing f_α with one in (4.2) and omitting the subscripts α and γ and the summations on α and γ from (4.2) and (4.3). Further simplifications for the complete data problem occur when we assume that $(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ is in the alternative parameter region (see the above similar modification to the missing data case) and/or when we make the further assumption that the null hypothesis has an explicit MLE.

For example, consider the complete data case when the MLE under the null hypothesis has an explicit representation and $(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ is a value under the alternative hypothesis. This covariance matrix in Theorem 3.2 simplifies to

$$\boldsymbol{\Xi}(\hat{\boldsymbol{\sigma}}_r, \hat{\boldsymbol{\sigma}}_s) = 2\mathbf{Y}_r^{-1}(2\mathbf{H}_{rs} + \mathbf{J}_{rs})\mathbf{Y}_s^{-1}$$

with $\mathbf{H}_{rs} = \mathbf{0}$ unless $r = s = 0$ in which case we have

$$\mathbf{H}_{00} = (\boldsymbol{\mu}^* - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}_0^{-1}\mathbf{G}_g\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}_0^{-1}\mathbf{G}_h\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\mu}^* - \boldsymbol{\mu}_0),$$

and $\mathbf{J}_{rs}$ is given by $\mathbf{J}_{11} = \mathbf{Y}_1$, $\mathbf{J}_{01} = \mathbf{J}_{10}$,

$$\mathbf{J}_{10} = [\operatorname{tr}\mathbf{G}_g\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}_0^{-1}\mathbf{G}_h\boldsymbol{\Sigma}^{*-1}]_{gh}, \quad \mathbf{J}_{00} = [\operatorname{tr}\mathbf{G}_g\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}_0^{-1}\mathbf{G}_h\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}_0^{-1}].$$

The asymptotic variance simplifies to the known result (Szatrowski, 1979),

$$v_\infty = 2\{\mathrm{tr}(\mathbf{I} - \mathbf{\Sigma}_0^{-1}\mathbf{\Sigma}^*)^2 + 2(\boldsymbol{\mu}^* - \boldsymbol{\mu}_0)'\mathbf{\Sigma}_0^{-1}\mathbf{\Sigma}^*\mathbf{\Sigma}_0^{-1}(\boldsymbol{\mu}^* - \boldsymbol{\mu}_0)\}$$

by noting that

$$v_\infty = \Sigma_{r,s=0}^1 (-1)^{r+s} \Sigma_{g=1}^{m_r} \Sigma_{h=1}^{m_s} (\mathrm{tr}\, \mathbf{\Sigma}_r^{-1}\mathbf{G}_g)(\mathbf{\Xi}(\hat{\boldsymbol{\sigma}}_r, \hat{\boldsymbol{\sigma}}))_{gh}(\mathrm{tr}\, \mathbf{\Sigma}_s^{-1}\mathbf{G}_h)$$

$$= 2\, \Sigma_{r,s=0}^1 (-1)^{r+s}\boldsymbol{\sigma}_r'(2\mathbf{H}_{rs} + \mathbf{J}_{rs})\boldsymbol{\sigma}_s$$

since

$$\Sigma_{g=1}^{m_r}(\mathrm{tr}\, \mathbf{\Sigma}_r^{-1}\mathbf{G}_g)(\mathbf{Y}_r^{-1})_{gi} = \{\boldsymbol{\sigma}_r'\mathbf{Y}_r\mathbf{Y}_r^{-1}\}_i = \{\boldsymbol{\sigma}_r'\}_i.$$

REFERENCES

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

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

ANDERSON, T. W. (1970). Estimation of covariance matrices which are linear combinations or whose inverse are linear combinations of given matrices, in *Essays in Probability and Statistics*, ed. R. C. Bose et. al. University of North Carolina Press, 1–24.

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

BISHOP, Y. M. M., FIENBERG, S. E., and HOLLAND, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice.* MIT Press, Cambridge, Massachusetts.

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38.

HOLLAND, P. W. and WIGHTMAN, L. (1982). Section pre-equating: a preliminary investigation, in *Test Equating*, ed. P. W. Holland and D. B. Rubin. Academic, New York, 271–297.

RUBIN, D. B. and SZATROWSKI, T. H. (1982a). Discussion of "Section pre-equating: a preliminary investigation", in *Test Equating*, ed. P. W. Holland and D. B. Rubin. Academic, New York, 301–306.

Rubin, D. B. and SZATROWSKI, T. H. (1982b). Finding maximum likelihood estimates of patterned covariance matrices by the EM algorithm. *Biometrika* **69** 657–660.

SZATROWSKI, T. H. (1978). Explicit solutions, one iteration convergence and averaging in the multivariate normal estimation problem for patterned means and covariances. *Ann. Statist. Math.* **30** Part A, 81–88.

SZATROWSKI, T. H. (1979). Asymptotic nonnull distributions for likelihood ratio statistics in the multivariate patterned mean and covariance problem. *Ann. Statist.* **7** 823–837.

SZATROWSKI, T. H. (1980). Necessary and sufficient conditions for explicit solutions in the multivariate normal estimation problem for patterned means and covariances. *Ann. Statist.* **8** (4), 802–810.

SZATROWSKI, T. H. (1981a). Missing data in the multivariate normal patterned mean and correlation matrix testing and estimation problem. Educational Testing Service Technical Report PSR No. 81-16.

SZATROWSKI, T. H. (1981b). Missing data in the multivariate normal patterned mean and covariance matrix testing and estimation problem. Educational Testing Service Technical Report PSR No. 81-17.

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

GRADUATE SCHOOL OF MANAGEMENT
RUTGERS UNIVERSITY
NEWARK, NEW JERSEY 07102