Estimation of a Linear Transformation: Large Sample Results*

by

Leon Jay Gleser
Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #453

May, 1976

*Research partially supported by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant AFOSR-72-2350B at Purdue University. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright hereon.

SUMMARY

ESTIMATION OF A LINEAR TRANSFORMATION: LARGE SAMPLE RESULTS

The present paper provides large sample strong consistency and distributional results for the maximum likelihood estimators $\hat{B}$ and $\hat{\sigma}^2$ of the regression slope matrix B and error variance in the multivariate "errors in variables" regression model introduced by Gleser and Watson (1973), and generalized by A. K. Bhargava (1975). In Bhargava's model, n independent observations $x_i' = (x_{1i}', x_{2i}')$ are taken on pairs of random vectors x_{1i}: px1 and x_{2i}: rx1, $r \leq p$. It is assumed that for each $i = 1,2,\ldots,n$, $\mathscr{E}(x_{2i}) = B\mathscr{E}(x_{1i})$ and that x_i has a (p+r)-variate normal distribution with covariance matrix $\sigma^2 I_{p+r}$. We wish to estimate B, σ^2, and $\mathscr{E}(x_{1i})$, $i = 1,2,\ldots,n$. Under a reasonable assumption concerning the sequence $\{\mathscr{E}(x_{1i})\}$, we show that $\hat{B}$ and $r^{-1}(p+r)\hat{\sigma}^2$ are strongly consistent estimators of B and σ^2, respectively, as $n \to \infty$. We also obtain the limiting distributions of $n^{\frac{1}{2}}(\hat{B}-B)$ and $n^{\frac{1}{2}}(r^{-1}(p+r)\hat{\sigma}^2-\sigma^2)$. Using these asymptotic distributions, approximate confidence region procedures for estimating B and σ^2 are suggested. In the course of our derivations, we establish large sample strong convergence and distributional results for the noncentral Wishart distribution.

AMS 1970 subject classifications. Primary 62H10, 62E20; Secondary 62F10, 62F25, 62H25, 62P15, 62P20.

Key words and phrases. Linear functional relationship, regression with errors in variables, asymptotic distributions, strong consistency, approximate confidence regions, principal components, noncentral Wishart distribution.

ESTIMATION OF A LINEAR TRANSFORMATION: LARGE SAMPLE RESULTS[1]

by

Leon Jay Gleser
Purdue University

1. Introduction. It is well known that the presence of errors of measurement in the independent variables in univariate linear regression (i.e., one dependent variable) makes the ordinary least squares estimators inconsistent and biased. Models of regression which incorporate "errors in variables" have been studied, and an extensive literature exists which deals with maximum likelihood and generalized least squares estimators of the parameters of univariate "errors in variables" regression models [Madansky (1959), Moran (1971), Sprent (1966), Williams (1955)]. Less is known concerning the estimation of the parameters in multivariate "errors in variables" regression models, although Gleser and Watson (1973) have considered maximum likelihood estimators (MLE) of the parameters in a multivariate "errors in variables" regression model in which the number of dependent variables equals the number of independent variables. Recently, A. K. Bhargava (1975) has found the MLE of the parameters in a multivariate "errors in variables" regression model in which the number r of dependent variables is no greater than the number p of independent variables ($r \leq p$).

It should be noted that many of the papers dealing with "errors in variables" regression models speak instead of "estimating linear functional relationships" or, in the case of Gleser and Watson (1973) and Bhargava (1975), of "estimating linear transformations". Because the present paper is concerned with the model discussed by Bhargava, we have adopted his terminology for the sake of continuity. The references at the end of this paper, particularly Moran (1971), should be sufficient for the reader to track down related papers.

The model which we adopt in the present paper is the following. We observe n independent pairs of random vectors $\underset{\sim}{x}_i' = (\underset{\sim}{x}_{1i}', \underset{\sim}{x}_{2i}')$, where $\underset{\sim}{x}_{1i}$ is px1 and $\underset{\sim}{x}_{i2}$ is rx1, $r \leq p$, $i = 1,2,\ldots,n$. We assume that

$$(1.1) \qquad \underset{\sim}{x}_i = \begin{pmatrix} \underset{\sim}{x}_{1i} \\ \underset{\sim}{x}_{2i} \end{pmatrix} = \begin{pmatrix} \underset{\sim}{\xi}_{1i} \\ \underset{\sim}{\xi}_{2i} \end{pmatrix} + \begin{pmatrix} \underset{\sim}{e}_{1i} \\ \underset{\sim}{e}_{2i} \end{pmatrix} = \underset{\sim}{\xi}_i + \underset{\sim}{e}_i,$$

where

$$(1.2) \qquad \underset{\sim}{\xi}_{2i} = \underset{\sim}{B}\, \underset{\sim}{\xi}_{1i},$$

$i = 1,2,\ldots,n$. We also assume that the vectors $\underset{\sim}{e}_i$, $i = 1,2,\ldots,n$, are i.i.d. with

$$(1.3) \qquad \mathscr{E}(\underset{\sim}{e}_i) = \underset{\sim}{0}, \qquad \mathscr{E}(\underset{\sim}{e}_i \underset{\sim}{e}_i') = \sigma^2 \underset{\sim}{I}_{p+r},$$

$i = 1,2,\ldots,n$. For the purpose of inference, the common distribution of the $\underset{\sim}{e}_i$'s is assumed to be multivariate normal. The parameters $\underset{\sim}{B}$: rxp, $\sigma^2 > 0$, and $\underset{\sim}{\xi}_{1i}$: px1, $i = 1,2,\ldots,n$, are assumed to be unknown, and are to be estimated.

Now let us adopt a more compact notation. Let

$$\underset{\sim}{X} = \begin{pmatrix} \underset{\sim}{X}_1 \\ \underset{\sim}{X}_2 \end{pmatrix} = \begin{pmatrix} \underset{\sim}{x}_{11} & \underset{\sim}{x}_{12} & \cdots & \underset{\sim}{x}_{1n} \\ \underset{\sim}{x}_{21} & \underset{\sim}{x}_{22} & \cdots & \underset{\sim}{x}_{2n} \end{pmatrix},$$

$$\underset{\sim}{\Xi} = \begin{pmatrix} \underset{\sim}{\Xi}_1 \\ \underset{\sim}{\Xi}_2 \end{pmatrix} = \begin{pmatrix} \underset{\sim}{\xi}_{11} & \underset{\sim}{\xi}_{12} & \cdots & \underset{\sim}{\xi}_{1n} \\ \underset{\sim}{\xi}_{21} & \underset{\sim}{\xi}_{22} & \cdots & \underset{\sim}{\xi}_{2n} \end{pmatrix},$$

and

$$\underset{\sim}{E} = \begin{pmatrix} \underset{\sim}{E}_1 \\ \underset{\sim}{E}_2 \end{pmatrix} = \begin{pmatrix} \underset{\sim}{e}_{11} & \underset{\sim}{e}_{12} & \cdots & \underset{\sim}{e}_{1n} \\ \underset{\sim}{e}_{21} & \underset{\sim}{e}_{22} & \cdots & \underset{\sim}{e}_{2n} \end{pmatrix},$$

where $\underset{\sim}{X}_1$, $\underset{\sim}{\Xi}_1$, $\underset{\sim}{E}_1$ are pxn; $\underset{\sim}{X}_2$, $\underset{\sim}{\Xi}_2$, $\underset{\sim}{E}_2$ are rxn; and $r \leq p$. In terms of these matrices, our model becomes

(1.4) $$\underset{\sim}{X} = \underset{\sim}{\Xi} + \underset{\sim}{E},$$

(1.5) $$\underset{\sim}{\Xi}_2 = B\, \underset{\sim}{\Xi}_1,$$

where the columns of $\underset{\sim}{E}$ are i.i.d. with mean vector $\underset{\sim}{0}$ and covariance matrix $\sigma^2 \underset{\sim}{I}_{p+r}$.

Let

(1.6) $$\underset{\sim}{W} = \underset{\sim}{X}\underset{\sim}{X}'.$$

Let $d_1 \geq d_2 \geq \ldots \geq d_p \geq d_{p+1} \geq \ldots \geq d_{p+r} \geq 0$ be the (ordered) eigenvalues of W, and let

(1.7) $$\underset{\sim}{D} = \begin{pmatrix} \underset{\sim}{D}_{max} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{D}_{min} \end{pmatrix} = \mathrm{diag}(d_1, d_2, \ldots, d_{p+r}),$$

where $\underset{\sim}{D}_{max} = \mathrm{diag}(d_1, d_2, \ldots, d_p)$, $\underset{\sim}{D}_{min} = \mathrm{diag}(d_{p+1}, \ldots, d_{p+r})$.

Finally, let

$$\underset{\sim}{G} = \begin{pmatrix} \underset{\sim}{G}_{11} & \underset{\sim}{G}_{12} \\ \underset{\sim}{G}_{21} & \underset{\sim}{G}_{22} \end{pmatrix}: \quad (p+r)x(p+r), \qquad G_{11}: \quad pxp,$$

satisfy

(1.8) $$\underset{\sim}{G}'\underset{\sim}{G} = \underset{\sim}{G}\underset{\sim}{G}' = I_{p+r},$$

(1.9) $$\underset{\sim}{W} = \underset{\sim}{G}\underset{\sim}{D}\underset{\sim}{G}'.$$

That is, $\underset{\sim}{G}$ is an orthogonal matrix whose ith column is the eigenvector corresponding to d_i, $i = 1,2,\ldots,p+r$.

<u>Theorem 1.1</u> [Bhargava (1975)]. Under the assumption that $n \geq p+r$ and that the common distribution of the columns of $\underset{\sim}{E}$ is multivariate normal, the MLE of $\underset{\sim}{B}$, $\underset{\sim}{\Xi}_1$, and σ^2 are respectively:

(1.10) $$\hat{\underset{\sim}{B}} = \underset{\sim}{G}_{21}\underset{\sim}{G}_{11}^{-1} = -(\underset{\sim}{G}_{22}')^{-1}\underset{\sim}{G}_{12}',$$

(1.11) $$\begin{aligned}\hat{\underset{\sim}{\Xi}}_1 &= \underset{\sim}{G}_{11}\underset{\sim}{G}_{11}'\underset{\sim}{X}_1 + \underset{\sim}{G}_{11}\underset{\sim}{G}_{21}'\underset{\sim}{X}_2 \\ &= (\underset{\sim}{I}_p - \underset{\sim}{G}_{12}\underset{\sim}{G}_{12}')\underset{\sim}{X}_1 - \underset{\sim}{G}_{12}\underset{\sim}{G}_{22}'\underset{\sim}{X}_2,\end{aligned}$$

and

(1.12) $$\hat{\sigma}^2 = n^{-1}(p+r)^{-1}\mathrm{tr}\ \underset{\sim}{D}_{min}.$$

In Section 2, we show that $\hat{\underset{\sim}{B}}$ and $r^{-1}(p+r)\hat{\sigma}^2$ are sequences of strongly consistent estimators of $\underset{\sim}{B}$ and σ^2, respectively. Our results are obtained without assuming that the common distribution of the columns of $\underset{\sim}{E}$ is the multivariate normal distribution. All that is needed is that

(1.13) $$\underset{\sim}{\Delta} = \lim_{n\to\infty} n^{-1}\ \underset{\sim}{\Xi}_1\underset{\sim}{\Xi}_1'$$

exists and, in the case of the strong convergence of $\hat{\underset{\sim}{B}}$, is positive definite. Interestingly, $n^{-1}\ \hat{\underset{\sim}{\Xi}}_1\hat{\underset{\sim}{\Xi}}_1'$ is not a sequence of consistent estimators of $\underset{\sim}{\Delta}$. However, at least one sequence of strongly consistent estimators of $\underset{\sim}{\Delta}$ does exist, as we show in Section 2.

Section 3 considers large sample distributional results for $n^{\frac{1}{2}}(\hat{\underset{\sim}{B}}-\underset{\sim}{B})$ and $n^{\frac{1}{2}}(r^{-1}(p+r)\hat{\sigma}^2 - \sigma^2)$. If the elements of any column of $\underset{\sim}{E}$ have finite fourth moments, both $n^{\frac{1}{2}}(\underset{\sim}{B}-\underset{\sim}{B})$ and $n^{\frac{1}{2}}(r^{-1}(p+r)\hat{\sigma}^2 - \sigma^2)$ are asymptotically normal. The covariance matrix of the asymptotic multivariate normal distribution of $n^{\frac{1}{2}}(\underset{\sim}{B}-\underset{\sim}{B})$ is determined in the special case when the columns of $\underset{\sim}{E}$ have a common multivariate normal distribution; and a strongly consistent sequence of estimators of this covariance matrix is established. These results lead to an approximate large sample 100(1-α)% elliptical confidence region for $\underset{\sim}{B}$. In Section 3, we also obtain an approximate large sample 100(1-α)% confidence interval for σ^2.

The method of proof that we use in Sections 2 and 3 (particularly Section 3) makes use of explicit representations of our estimators in terms of weighted matrix sums of the elements of $\underset{\sim}{W}$, and thus requires us to establish strong consistency and large sample distributional results for the elements of this matrix. We do this both under general assumptions about the common distribution of the columns $\underset{\sim}{e}_1$, $\underset{\sim}{e}_2$,... of $\underset{\sim}{E}$, and under the particular assumption that the columns of $\underset{\sim}{E}$ have a common multivariate normal distribution. In the latter case, $\underset{\sim}{W}$ has a noncentral Wishart distribution, and our results in Sections 2 and 3 provide large sample strong consistency and distributional results for the noncentral Wishart matrix.

2. Strong Consistency. We begin by investigating the strong convergence of $\underset{\sim}{W}$.

Lemma 2.1. Assume that $\underset{\sim}{e}_1$, $\underset{\sim}{e}_2$,... are i.i.d. with common mean vector $\underset{\sim}{0}$ and common covariance matrix $\sigma^2 \underset{\sim}{I}_{p+r}$. Let

(2.1)
$$\underset{\sim}{\Theta} = \sigma^2 \underset{\sim}{I}_{p+r} + \begin{pmatrix} \underset{\sim}{I}_p \\ \underset{\sim}{B} \end{pmatrix} \underset{\sim}{\Delta} \begin{pmatrix} \underset{\sim}{I}_p \\ \underset{\sim}{B} \end{pmatrix}',$$

where $\underset{\sim}{\Delta}$ is defined by (1.13), and is assumed to exist. Then

(2.2)
$$\lim_{n\to\infty} n^{-1}\underset{\sim}{W} = \underset{\sim}{\Theta}, \quad \text{a.s.}$$

Proof. From (1.4) and (1.6),

(2.3)
$$n^{-1}\underset{\sim}{W} = n^{-1}\underset{\sim}{E}\underset{\sim}{E}' + n^{-1}\underset{\sim}{\Xi}\underset{\sim}{E}' + n^{-1}\underset{\sim}{E}\underset{\sim}{\Xi}' + n^{-1}\underset{\sim}{\Xi}\underset{\sim}{\Xi}'.$$

Since the n columns of $\underset{\sim}{E}$ are i.i.d. with common mean vector $\underset{\sim}{0}$ and common covariance matrix $\sigma^2 \underset{\sim}{I}_{p+r}$, we have from the SLLN that

$$\lim_{n\to\infty} n^{-1}\underset{\sim}{E}\underset{\sim}{E}' = \sigma^2 \underset{\sim}{I}_{p+r}, \quad \text{a.s.} \tag{2.4}$$

From (1.5) and (1.13),

$$\lim_{n\to\infty} n^{-1}\underset{\sim}{\Xi}\underset{\sim}{\Xi}' = \begin{pmatrix} \underset{\sim}{I}_p \\ \underset{\sim}{B} \end{pmatrix} \underset{\sim}{\Delta} \begin{pmatrix} \underset{\sim}{I}_p \\ \underset{\sim}{B} \end{pmatrix}'. \tag{2.5}$$

Thus, (2.2) holds if

$$\lim_{n\to\infty} n^{-1}\underset{\sim}{\Xi}'\underset{\sim}{E}' = \lim_{n\to\infty} (n^{-1}\underset{\sim}{E}\underset{\sim}{\Xi}')' = \underset{\sim}{0}, \quad \text{a.s.} \tag{2.6}$$

Let $\underset{\sim}{\Xi} = ((\xi_{ij}))$, $\underset{\sim}{E} = ((e_{ij}))$, $\underset{\sim}{A}(n) = n^{-1}\underset{\sim}{\Xi}'\underset{\sim}{E} = ((a_{ij}^{(n)}))$. Finally, let

$h_{ij}^{(n)} = [\sum_{k=1}^{n} (\xi_{ik})^2]^{-\frac{1}{2}}\xi_{ij}$. Then for all (i,j), i, j = 1,2,...,p+r,

$$a_{ij}^{(n)} = n^{-1} \sum_{k=1}^{n} \xi_{ik}e_{kj} = [n^{-1} \sum_{k=1}^{n} \xi_{ik}^2]^{\frac{1}{2}}\, n^{-\frac{1}{2}} \sum_{k=1}^{n} h_{ik}^{(n)} e_{kj}.$$

By (2.5), $n^{-1} \sum_{k=1}^{n} \xi_{ik}^2$ converges to a finite nonnegative number. But by Lemma 2 of Gleser (1966), noting that $\sum_{k=1}^{n} (h_{ik}^{(n)})^2 = 1$, we have

$$\lim_{n\to\infty} n^{-\frac{1}{2}} \sum_{k=1}^{n} h_{ik}^{(n)} e_{kj} = 0, \text{ a.s.}$$

Thus for all (i,j), i,j = 1,2,...,p+r, $\lim_{n\to\infty} a_{ij}^{(n)} = 0$, a.s., proving (2.6), and thus (2.2). □

<u>Remark</u>. If $\underset{\sim}{e}_1, \underset{\sim}{e}_2, \ldots$ have common covariance matrix $\underset{\sim}{\Sigma}$: $(p+r)\times(p+r)$ and if $\lim_{n\to\infty} n^{-1}\underset{\sim}{\Xi}\underset{\sim}{\Xi}' = \underset{\sim}{T}$ exists, then a proof identical to that of Theorem 2.1 can be used to demonstrate that $\lim_{n\to\infty} n^{-1}\, \underset{\sim}{W} = \underset{\sim}{\Sigma} + \underset{\sim}{T}$, a.s.

Let $\gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_p \geq 0$ be the eigenvalues of $(\underset{\sim}{I}_p+\underset{\sim}{B}'\underset{\sim}{B})^{\frac{1}{2}}\underset{\sim}{\Delta}(\underset{\sim}{I}_p+\underset{\sim}{B}'\underset{\sim}{B})^{\frac{1}{2}}$, where $(\underset{\sim}{I}_p+\underset{\sim}{B}'\underset{\sim}{B})^{\frac{1}{2}}$ is the symmetric square root of $(\underset{\sim}{I}_p+\underset{\sim}{B}'\underset{\sim}{B})$. Let $\underset{\sim}{D}_\gamma = \text{diag}(\gamma_1, \gamma_2, \ldots, \gamma_p)$ and let $\underset{\sim}{\psi}$ be a pxp orthogonal matrix satisfying

$$(2.7) \qquad (\underset{\sim}{I}_p+\underset{\sim}{B}'\underset{\sim}{B})^{\frac{1}{2}}\underset{\sim}{\Delta}(\underset{\sim}{I}_p+\underset{\sim}{B}'\underset{\sim}{B})^{\frac{1}{2}} = \underset{\sim}{\psi}\underset{\sim}{D}_\gamma\underset{\sim}{\psi}'.$$

Note that if

$$(2.8) \qquad \underset{\sim}{\Gamma} = \begin{pmatrix} (\underset{\sim}{I}_p+\underset{\sim}{B}'\underset{\sim}{B})^{-\frac{1}{2}}\underset{\sim}{\psi} & \underset{\sim}{B}'(\underset{\sim}{I}_r+\underset{\sim}{B}\underset{\sim}{B}')^{-\frac{1}{2}} \\ \underset{\sim}{B}(\underset{\sim}{I}_p+\underset{\sim}{B}'\underset{\sim}{B})^{-\frac{1}{2}}\underset{\sim}{\psi} & -(\underset{\sim}{I}_r+\underset{\sim}{B}\underset{\sim}{B}')^{-\frac{1}{2}} \end{pmatrix},$$

where $(\underset{\sim}{I}_r+\underset{\sim}{B}\underset{\sim}{B}')^{\frac{1}{2}}$ is the symmetric square root of $(\underset{\sim}{I}_r+\underset{\sim}{B}\underset{\sim}{B}')$, then $\underset{\sim}{\Gamma}$: (p+r)x(p+r) is orthogonal, and

$$(2.9) \qquad \underset{\sim}{\Theta}\underset{\sim}{\Gamma} = \underset{\sim}{\Gamma}\begin{pmatrix} \sigma^2\underset{\sim}{I}_p+\underset{\sim}{D}_\gamma & \underset{\sim}{0} \\ \underset{\sim}{0} & \sigma^2\underset{\sim}{I}_r \end{pmatrix}.$$

We conclude that the columns of $\underset{\sim}{\Gamma}$ are eigenvectors of $\underset{\sim}{\Theta}$, and that the eigenvalues $\theta_1 \geq \theta_2 \geq \ldots \geq \theta_{p+r} \geq 0$ of Θ are:

$$(2.10) \qquad \begin{aligned} \theta_i &= \sigma^2 + \gamma_i, & i &= 1,2,\ldots,p, \\ \theta_{p+j} &= \sigma^2 \quad , & j &= 1,2,\ldots,r. \end{aligned}$$

<u>Lemma 2.2</u>. Under the conditions of Lemma 2.1,

$$(2.11) \qquad \lim_{n\to\infty} n^{-1}\underset{\sim}{D} = \underset{\sim}{D}_\theta = \text{diag}(\theta_1,\theta_2,\ldots,\theta_{p+r}), \quad \text{a.s.}$$

<u>Proof</u>. Under our assumptions about the vectors $\underset{\sim}{e}_1, \underset{\sim}{e}_2, \ldots,$ we know that $n^{-1}\underset{\sim}{W}$ is positive definite for all $n \geq p+r$ [Perlman and Eaton (1973)]. The ith eigenvalue of a positive definite matrix is a continuous function of the elements of that matrix. Since $n^{-1}\underset{\sim}{W}$ a.s. converges to a positive definite matrix $\underset{\sim}{\Theta}$ by Lemma 2.1, the result (2.11) immediately follows. □

In the following argument, we will need to notationally indicate the dependence of our sample quantities on the sample size n. Thus, for sample size n, let $\underset{\sim}{W}^{(n)}$, $\underset{\sim}{D}^{(n)}$, $\underset{\sim}{D}_{max}^{(n)}$, $\underset{\sim}{D}_{min}^{(n)}$, and

$$\underset{\sim}{G}^{(n)} = \begin{pmatrix} \underset{\sim}{G}_{11}^{(n)} & \underset{\sim}{G}_{12}^{(n)} \\ \underset{\sim}{G}_{21}^{(n)} & \underset{\sim}{G}_{22}^{(n)} \end{pmatrix}$$

be the quantities defined by (1.6), (1.7), and (1.9) respectively. Further, let $\hat{\underset{\sim}{B}}^{(n)}$ be the estimator of $\underset{\sim}{B}$ for sample size n given by (1.10).

Lemma 2.3. Under the assumptions of Lemma 2.1, plus the additional assumption that $\underset{\sim}{\Delta}$ is positive definite, we have

$$\lim_{n\to\infty} \hat{\underset{\sim}{B}}^{(n)} = \underset{\sim}{B}, \quad \text{a.s.,} \tag{2.12}$$

$$\lim_{n\to\infty} n^{-1}(\underset{\sim}{G}_{22}^{(n)})(\underset{\sim}{D}_{min}^{(n)})(\underset{\sim}{G}_{22}^{(n)})' = \sigma^2(\underset{\sim}{I}_r+\underset{\sim}{B}\underset{\sim}{B}')^{-1}, \quad \text{a.s.,} \tag{2.13}$$

$$\lim_{n\to\infty} n^{-1}(\underset{\sim}{G}_{11}^{(n)})(\underset{\sim}{D}_{max}^{(n)})(\underset{\sim}{G}_{11}^{(n)})' = \underset{\sim}{\Delta} + \sigma^2(\underset{\sim}{I}_p+\underset{\sim}{B}'\underset{\sim}{B})^{-1}, \quad \text{a.s.} \tag{2.14}$$

Proof. Note that the columns of $\underset{\sim}{G}^{(n)}$ are orthogonal and of length 1 for all $n \geq p+r$. Let

$$\omega = (\underset{\sim}{e}_1, \underset{\sim}{e}_2, \ldots)$$

be a fixed point in the underlying probability space. For fixed ω such that (2.2) and (2.11) hold, the sequence $\{\underset{\sim}{G}^{(n)}\}$ lies in a compact subspace of $(p+r)^2$-dimensional Euclidean space. Thus, each subsequence of $\{\underset{\sim}{G}^{(n)}\}$ has a convergent sub-subsequence. Suppose that the limit of this sub-subsequence is

$$\underset{\sim}{Q} = \begin{pmatrix} \underset{\sim}{Q}_{11} & \underset{\sim}{Q}_{12} \\ \underset{\sim}{Q}_{21} & \underset{\sim}{Q}_{22} \end{pmatrix}.$$

Then since for all n,

$$(n^{-1}\underset{\sim}{W}^{(n)}) \begin{pmatrix} \underset{\sim}{G}_{11}^{(n)} \\ \underset{\sim}{G}_{21}^{(n)} \end{pmatrix} = \begin{pmatrix} \underset{\sim}{G}_{11}^{(n)} \\ \underset{\sim}{G}_{21}^{(n)} \end{pmatrix} (n^{-1}\underset{\sim}{D}_{max}^{(n)}),$$

we can take limits over the indices of the sub-subsequence on both sides of this equality and obtain [see (2.2), (2.10) and (2.11)]

$$\underset{\sim}{\Theta}\begin{pmatrix} \underset{\sim}{Q}_{11} \\ \underset{\sim}{Q}_{21} \end{pmatrix} = \begin{pmatrix} \underset{\sim}{Q}_{11} \\ \underset{\sim}{Q}_{21} \end{pmatrix} (\sigma^2 \underset{\sim}{I}_p + \underset{\sim}{D}_\gamma).$$

Thus, $(\underset{\sim}{Q}_{11}', \underset{\sim}{Q}_{21}')'$ is in the eigensubspace corresponding to the largest p roots of $\underset{\sim}{\Theta}$. Since our additional assumption (that $\underset{\sim}{\Delta}$ is positive definite) implies that $\theta_p = \sigma^2+\gamma_p > \sigma^2 = \theta_{p+1}$, this eigensubspace is unique. Hence, from (2.8) and (2.9) there exists a nonsingular matrix $\underset{\sim}{T}$ such that

$$(2.15) \qquad \begin{pmatrix} \underset{\sim}{Q}_{11} \\ \underset{\sim}{Q}_{21} \end{pmatrix} = \begin{pmatrix} (\underset{\sim}{I}_p+\underset{\sim}{B}'\underset{\sim}{B})^{-\frac{1}{2}}\underset{\sim}{\psi} \\ \underset{\sim}{B}(\underset{\sim}{I}_p+\underset{\sim}{B}'\underset{\sim}{B})^{-\frac{1}{2}}\underset{\sim}{\psi} \end{pmatrix}\underset{\sim}{T}.$$

Again, since

$$\hat{\underset{\sim}{B}}^{(n)} = \underset{\sim}{G}_{21}^{(n)}(\underset{\sim}{G}_{11}^{(n)})^{-1},$$

taking limits on both sides of this equality over the indices of the sub-subsequence results, by (2.15), in the limiting value $\underset{\sim}{B}$. Thus, we have shown that for every value ω such that (2.2) and (2.11) holds, every subsequence of $\{\hat{\underset{\sim}{B}}^{(n)}\}$ has a subsubsequence converging to $\underset{\sim}{B}$. It then follows from facts about limits of sequences in Euclidean space that $\lim_{n\to\infty} \hat{\underset{\sim}{B}}^{(n)} = \underset{\sim}{B}$ for all ω such that (2.2) and (2.11) hold, and thus that (2.12) holds. The results (2.13) and (2.14) follow by similar arguments using the identities [see (1.9) and (1.10)]

$$\begin{aligned}(2.16) \qquad & (\underset{\sim}{B},-\underset{\sim}{I}_r)(n^{-1}\underset{\sim}{W}^{(n)})(\underset{\sim}{B},-\underset{\sim}{I}_r)' \\ &= (\hat{\underset{\sim}{B}}^{(n)}-\underset{\sim}{B})(\underset{\sim}{G}_{11}^{(n)})(n^{-1}\underset{\sim}{D}_{max}^{(n)})(\underset{\sim}{G}_{11}^{(n)})'(\hat{\underset{\sim}{B}}^{(n)}-\underset{\sim}{B})' \\ &\quad + (\underset{\sim}{I}_r+\underset{\sim}{B}\hat{\underset{\sim}{B}}^{(n)\prime})(\underset{\sim}{G}_{22}^{(n)})(n^{-1}\underset{\sim}{D}_{min}^{(n)})(\underset{\sim}{G}_{22}^{(n)})'(\underset{\sim}{I}_r+\hat{\underset{\sim}{B}}^{(n)}\underset{\sim}{B}')\end{aligned}$$

and

$$(2.17) \qquad (\underset{\sim}{I}_p, \underset{\sim}{B}')(n^{-1}\underset{\sim}{W}^{(n)})(\underset{\sim}{I}_p, \underset{\sim}{B}')'$$
$$= (\underset{\sim}{I}_p + \hat{\underset{\sim}{B}}^{(n)}{}'\underset{\sim}{B})(\underset{\sim}{G}_{11}^{(n)})(n^{-1}\underset{\sim}{D}_{max})(\underset{\sim}{G}_{11}^{(n)})'(\underset{\sim}{I}_p + \underset{\sim}{B}'\hat{\underset{\sim}{B}}^{(n)})$$
$$+ (\hat{\underset{\sim}{B}}^{(n)} - \underset{\sim}{B})'(\underset{\sim}{G}_{22}^{(n)})(n^{-1}\underset{\sim}{D}_{min})(\underset{\sim}{G}_{22}^{(n)})'(\hat{\underset{\sim}{B}}^{(n)} - \underset{\sim}{B}),$$

respectively. □

From (1.11), (1.6), and (1.9), we see that

$$(2.18) \qquad \hat{\underset{\sim}{\Xi}}_1 \hat{\underset{\sim}{\Xi}}_1' = \underset{\sim}{G}_{11}\underset{\sim}{D}_{max}\underset{\sim}{G}_{11}'.$$

It thus follows from (2.14) that $\{n^{-1}\hat{\Xi}_1\hat{\Xi}_1'\}$ is not a consistent sequence of estimators for $\underset{\sim}{\Delta}$. Since $\underset{\sim}{\Delta}$ helps to determine the covariance matrix of the asymptotic distribution of $n^{\frac{1}{2}}(\hat{\underset{\sim}{B}} - \underset{\sim}{B})$, we will need a consistent sequence of estimators for $\underset{\sim}{\Delta}$ in order to construct an approximate large-sample confidence region for $\underset{\sim}{B}$. The following theorem, which follows directly from Lemmas 2.2 and 2.3, both summarizes our strong consistency results for $\hat{\underset{\sim}{B}}$ and $r^{-1}(p+r)\hat{\sigma}^2$, and provides us with a strongly consistent sequence of estimators for $\underset{\sim}{\Delta}$.

<u>Theorem 2.1</u>. Under the conditions of Lemma 2.1,

$$(2.19) \qquad \lim_{n\to\infty} r^{-1}(p+r)\hat{\sigma}^2 = \sigma^2, \qquad \text{a.s.},$$

so that $r^{-1}(p+r)\hat{\sigma}^2$ is a strongly consistent (sequence of) estimator(s) for σ^2. Under the conditions of Lemma 2.3,

$$(2.20) \qquad \lim_{n\to\infty} \hat{\underset{\sim}{B}} = \underset{\sim}{B}, \qquad \text{a.s.},$$

and

$$(2.21) \qquad \lim_{n\to\infty} n^{-1}(\underset{\sim}{G}_{11}\underset{\sim}{D}_{max}\underset{\sim}{G}_{11}' - r^{-1}(p+r)\hat{\sigma}^2(\underset{\sim}{I}_p + \hat{\underset{\sim}{B}}'\hat{\underset{\sim}{B}})^{-1}) = \underset{\sim}{\Delta}, \qquad \text{a.s.},$$

so that $\hat{\underset{\sim}{B}}$ is a strongly consistent (sequence of) estimator(s) for $\underset{\sim}{B}$, and

$$(2.22) \qquad \hat{\underset{\sim}{\Delta}} = n^{-1}(\underset{\sim}{G}_{11}\underset{\sim}{D}_{max}\underset{\sim}{G}_{11}' - r^{-1}(p+r)n\hat{\sigma}^2(\underset{\sim}{I}_p + \hat{\underset{\sim}{B}}'\hat{\underset{\sim}{B}})^{-1})$$

is a strongly consistent (sequence of) estimator(s) for $\underset{\sim}{\Delta}$.

Remark I. Weak consistency results (i.e., convergence in probability) for $\hat{\underset{\sim}{B}}$ and $r^{-1}(p+r)\hat{\sigma}^2$ have been obtained previously by Gleser and Watson (1973) when $r = p$, and by Bhargava (1975) in the general case $r \leq p$. Their proof of consistency for $r^{-1}(p+r)\hat{\sigma}^2$ is given under slightly weaker conditions $[n^{-2}\underset{\sim}{\Xi}_1\underset{\sim}{\Xi}_1' = o(1)]$ than the conditions of Lemma 2.1, but their proof of the consistency of $\hat{B}$ requires the condition (1.13), and also has a theoretical gap [noted in Gleser and Watson (1973)]. The full strength of the almost sure convergence results given in this section are not really needed for deriving the large-sample distributional results of the next section. However, the methods and conclusions in this section are of interest in their own right (particularly Lemma 2.1 and the proof of Lemma 2.3), and Theorem 2.1 may be of use in future work concerning the construction of asymptotically consistent and efficient fixed-diameter sequential confidence regions [see Gleser (1965)] and asymptotically optimal Bayesian sequential regional estimators [see Gleser and Kunte (1976)] for $\underset{\sim}{B}$.

Remark II. We once again call attention to the fact that no argument in the present section requires us to assume that the common distribution of $\underset{\sim}{e}_1, \underset{\sim}{e}_2, \ldots,$ is multivariate normal.

3. Asymptotic distributions. We begin by finding the large sample distribution of $n^{-\frac{1}{2}}(\underset{\sim}{W} - \mathscr{E}(\underset{\sim}{W}))$. Let $\underset{\sim}{e}' = (e_1, e_2, \ldots, e_{p+r})$ be a random vector having the same distribution as $\underset{\sim}{e}_1, \underset{\sim}{e}_2, \ldots, \underset{\sim}{e}_n$ (the columns of $\underset{\sim}{E}$). We assume that $\mathscr{E}(e_i^4) < \infty$, $i = 1,2,\ldots,p+r$. Let

$$\phi_{ijk\ell} = \mathscr{E}(e_i e_j e_k e_\ell), \qquad i,j,k,\ell = 0,1,2,\ldots,p+r, \tag{3.1}$$

with the understanding that $e_0 \equiv 1$. Thus, $\phi_{0iii} = \mathscr{E}(e_i^3)$ and so forth. Now, let

$$e_k = \begin{pmatrix} e_{k1} \\ e_{k2} \\ \vdots \\ e_{k(p+r)} \end{pmatrix}, \qquad \xi_k = \begin{pmatrix} \xi_{k1} \\ \xi_{k2} \\ \vdots \\ \xi_{k(p+r)} \end{pmatrix}, \qquad k = 1,2,\ldots,n.$$

Note from (2.3) that

$$\mathscr{E}(W) = n(\sigma^2 I_{p+r} + n^{-1}\Xi\Xi'),$$

and thus that

$$n^{-\frac{1}{2}}(W - \mathscr{E}(W)) = n^{-\frac{1}{2}}(EE' - n\sigma^2 I_{p+r} + \Xi E' + E\Xi')$$

(3.2) $$= n^{-\frac{1}{2}} \sum_{k=1}^{n} Z_k,$$

where $Z_k = ((z_{kij}))$,

(3.3) $$z_{kij} = e_{ki}e_{kj} - \sigma^2\delta_{ij} + \xi_{ki}e_{kj} + e_{hi}\xi_{kj},$$

and δ_{ij} is the Kronecker delta. The matrices Z_1, $Z_2,\ldots,$ are mutually statistically independent (but not identically distributed) with $\mathscr{E}(Z_k) = 0$, $k = 1,2,\ldots,n$, and

$$\begin{aligned} \operatorname{cov}(z_{kij}, z_{ki'j'}) = {} & \phi_{iji'j'} - \sigma^4\delta_{ij}\delta_{i'j'} + \xi_{ki'}\phi_{0ijj'} \\ & + \xi_{kj'}\phi_{0iji'} + \xi_{ki}\phi_{0i'j'j} + \xi_{kj}\phi_{0i'j'i} \\ & + \sigma^2(\xi_{kj}\xi_{kj'}\delta_{ii'} + \xi_{ki}\xi_{ki'}\delta_{jj'} + \xi_{kj}\xi_{ki'}\delta_{ij'} + \xi_{ki}\xi_{kj'}\delta_{i'j}). \end{aligned}$$

Let

(3.5) $$\kappa_{(i,j),(i',j')} = \lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} \operatorname{cov}(z_{kij}, z_{ki'j'}).$$

Then for all (i,j), (i',j'); i,j,i',j' = 1,2,...,p+r; we have

$$\kappa_{(i,j),(i',j')} = \phi_{iji'j'} - \sigma^4\delta_{ij}\delta_{i'j'} + \bar{\xi}_{i'}\phi_{0ijj'} + \bar{\xi}_{j'}\phi_{0iji'}$$
$$(3.6) \qquad + \bar{\xi}_i\phi_{0i'j'j} + \bar{\xi}_j\phi_{0i'j'i} + \sigma^2(\tau_{jj'}\delta_{ii'}$$
$$+ \tau_{ii'}\delta_{jj'} + \tau_{ji'}\delta_{ij'} + \tau_{ij'}\delta_{i'j}),$$

where the existence of

$$(3.7) \qquad \bar{\xi}_i = \lim_{n\to\infty} n^{-1}\sum_{k=1}^{n} \xi_{ki}, \qquad \tau_{ij} = \lim_{n\to\infty} n^{-1}\sum_{k=1}^{n} \xi_{ki}\xi_{kj}$$

is guaranteed by (1.13).

<u>Theorem 3.1</u>. Under the assumptions that (1.13) exists (and is finite) and that $\mathscr{E}(e_i^4) < \infty$, i = 1,2,...,p+r, the elements on and below the diagonal (the subdiagonal elements) of $n^{-\frac{1}{2}}(\underset{\sim}{W} - \mathscr{E}(\underset{\sim}{W}))$ have a limiting joint (p+r)(p+r+1)/2-dimensional normal distribution with mean vector $\underset{\sim}{0}$ and covariance matrix $\underset{\sim}{\mathscr{K}} = ((\kappa_{(i,j),(i',j')}))$.

<u>Proof</u>. Let $\underset{\sim}{W} = ((w_{ij}))$. Consider any linear combination

$$(3.8) \qquad n^{-\frac{1}{2}}\sum_{i\le j} c_{ij}(w_{ij} - \mathscr{E}(w_{ij})) = n^{-\frac{1}{2}}\sum_{i\le j}\sum_{k=1}^{n} c_{ij}z_{kij}$$
$$= n^{-\frac{1}{2}}\sum_{k=1}^{n}\left(\sum_{i\le j} c_{ij}z_{kij}\right)$$

of the subdiagonal elements of $n^{-\frac{1}{2}}(\underset{\sim}{W} - \mathscr{E}(\underset{\sim}{W}))$. We recognize this as a normalized sum of independent random variables. Using (3.3), (3.4), (3.5) and the assumption that the fourth moments of $\underset{\sim}{e}$ exist, it is straightforward to prove that

$$\operatorname*{plim}_{n\to\infty}\left[\sum_{k=1}^{n} \operatorname{var}\left(\sum_{i\le j} c_{ij}z_{kij}\right)\right]^{-1}\sum_{k=1}^{n}\left(\sum_{i\le j} c_{ij}z_{kij}\right)^2 = 1.$$

It then follows from Raikov's Theorem [Gnedenko and Kolmogorov (1954; p. 143)] that

$$(3.9) \qquad n^{-\frac{1}{2}} \sum_{k=1}^{n} \left(\sum_{i \le j} c_{ij} z_{kij} \right) \xrightarrow{L} N\left(0, \sum_{i \le j} \sum_{i' \le j'} c_{ij} c_{i'j'} \kappa_{(i,j),(i',j')}\right).$$

Since (3.9) holds for all linear combinations (3.8), the conclusion of the theorem follows. □

Remark. Our implicit assumption that the covariance matrix of $\underset{\sim}{e}$ is $\sigma^2 \underset{\sim}{I}_{p+r}$ is unnecessary for the proof of asymptotic normality. If the covariance matrix of $\underset{\sim}{e}$ is $\underset{\sim}{\Sigma} = ((\sigma_{ij}))$, then the same conclusion holds, except that $\underset{\sim}{\Sigma}$ replaces $\sigma^2 \underset{\sim}{I}_{p+r}$ in the formula for $\mathscr{L}(\underset{\sim}{W})$, and in the formula for $\kappa_{(i,j),(i',j')}$ in (3.6) we have

$$\kappa_{(i,j),(i',j')} = \phi_{iji'j'} - \sigma_{ij}\sigma_{i'j'} + \bar{\xi}_{i'}\phi_{0ijj'} + \bar{\xi}_{j'}\phi_{0iji'} + \bar{\xi}_{i}\phi_{0i'j'j} + \bar{\xi}_{j}\phi_{0i'j'i} + (\tau_{jj'}\sigma_{ii'} + \tau_{ii'}\sigma_{jj'} + \tau_{ji'}\sigma_{ij'} + \tau_{ij'}\sigma_{i'j}).$$

Corollary 3.1. If $\underset{\sim}{e}_1, \underset{\sim}{e}_2, \ldots,$ are i.i.d. multivariate normal with mean vector $\underset{\sim}{0}$ and covariance matrix $\underset{\sim}{\Sigma}$, and if (1.13) exists (and is finite), then the subdiagonal elements of $n^{-\frac{1}{2}}(\underset{\sim}{W} - n\underset{\sim}{\Sigma} - \underset{\sim}{\Xi}\underset{\sim}{\Xi}')$ have a limiting joint $(p+r)(p+r+1)/2$-variate normal distribution with mean vector $\underset{\sim}{0}$ and covariance matrix $\underset{\sim}{\mathscr{K}} = ((\kappa_{(i,j),(i',j')}))$ given by

$$\kappa_{(i,j),(i',j')} = \sigma_{ii'}\sigma_{jj'} + \sigma_{ij'}\sigma_{i'j} + \tau_{jj'}\sigma_{ii'} + \tau_{ii'}\sigma_{jj'} + \tau_{ji'}\sigma_{ij'} + \tau_{ij'}\sigma_{i'j}.$$

When $\underset{\sim}{\Sigma} = \sigma^2 \underset{\sim}{I}_{p+r}$,

$$(3.10)\qquad \kappa_{(i,j),(i',j')} = \sigma^4(\delta_{ii'}\delta_{jj'} + \delta_{ij'}\delta_{i'j}) + \sigma^2(\tau_{jj'}\delta_{ii'} + \tau_{ii'}\delta_{jj'} + \tau_{ji'}\delta_{ij'} + \tau_{ij'}\delta_{i'j}).$$

Proof. Because $\underset{\sim}{e}_1, \underset{\sim}{e}_2, \ldots$ are i.i.d. $N(\underset{\sim}{0},\underset{\sim}{\Sigma})$, we have

$$\phi_{iji'j'} = \begin{cases} 0, & \text{if } i = 0,\ i' = 0,\ j = 0, \text{ or } j' = 0, \\ \sigma_{ii'}\sigma_{jj'} + \sigma_{ij'}\sigma_{i'j} + \sigma_{ij}\sigma_{i'j'}, & \text{otherwise.} \end{cases}$$

The result of the Corollary now is a direct consequence of Theorem 3.1. □

We note that Corollary 3.1 gives the asymptotic distribution of the noncentral Wishart matrix in cases where the noncentrality parameter is $O(n)$.

To find the asymptotic distribution of $n^{\frac{1}{2}}(\hat{\underset{\sim}{B}}-\underset{\sim}{B})$, it is sufficient to note that (1.9) and (1.10) yield the representation:

$$(3.11)\qquad \begin{aligned} &(\underset{\sim}{I}_p,\underset{\sim}{B}')[n^{-\frac{1}{2}}(\underset{\sim}{W}-\mathscr{E}(\underset{\sim}{W}))](\underset{\sim}{B},-\underset{\sim}{I}_r)' \\ &= (\underset{\sim}{I}_p+\underset{\sim}{B}'\hat{\underset{\sim}{B}})(n^{-1}\underset{\sim}{G}_{11}\underset{\sim}{D}_{max}\underset{\sim}{G}_{11}')n^{\frac{1}{2}}(\underset{\sim}{B}-\hat{\underset{\sim}{B}})' \\ &\quad - n^{\frac{1}{2}}(\underset{\sim}{B}-\hat{\underset{\sim}{B}})'(n^{-1}\underset{\sim}{G}_{22}\underset{\sim}{D}_{min}\underset{\sim}{G}_{22}')(\underset{\sim}{I}_r+\hat{\underset{\sim}{B}}\underset{\sim}{B}'). \end{aligned}$$

Assuming that $\underset{\sim}{\Delta}$ is positive definite, and using (2.13), (2.14) and (3.11), we conclude that $n^{\frac{1}{2}}(\hat{\underset{\sim}{B}}-\underset{\sim}{B})'$ and

$$(3.12)\qquad \underset{\sim}{F} = -\underset{\sim}{\Delta}^{-1}(\underset{\sim}{I}_p+\underset{\sim}{B}'\underset{\sim}{B})^{-1}(\underset{\sim}{I}_p,\underset{\sim}{B}')[n^{-\frac{1}{2}}(\underset{\sim}{W}-\mathscr{E}(\underset{\sim}{W}))](\underset{\sim}{B},-\underset{\sim}{I}_r)'$$

have the same asymptotic distribution. Since the elements of $\underset{\sim}{F}$ are linear combinations of the subdiagonal elements of $n^{-\frac{1}{2}}(\underset{\sim}{W}-\mathscr{E}(\underset{\sim}{W}))$, we conclude that <u>when the assumptions of Theorem 3.1 hold and $\underset{\sim}{\Delta}$ is positive definite, the elements of $n^{\frac{1}{2}}(\hat{\underset{\sim}{B}}-\underset{\sim}{B})'$ have a limiting rp-variate normal distribution with $\underset{\sim}{0}$ mean vector and a covariance matrix that can be calculated using (3.6) and (3.12).</u> Since the covariance matrix of the limiting distribution of

$n^{1/2}(\hat{\underset{\sim}{B}}-\underset{\sim}{B})$ under the general conditions of Theorem 3.1 involves fourth-order cross moments of $\underset{\sim}{e}$, and thus is both complicated and hard to estimate, and since we are primarily interested in the case where $\underset{\sim}{e}_1, \underset{\sim}{e}_2, \ldots$ are i.i.d. $N(\underset{\sim}{0}, \sigma^2 \underset{\sim}{I}_{p+r})$, we content ourselves with the following.

Theorem 3.2. If $\underset{\sim}{e}_1, \underset{\sim}{e}_2, \ldots$ are i.i.d. $N(\underset{\sim}{0}, \sigma^2 \underset{\sim}{I}_{p+r})$, and if (1.13) exists and is positive definite, then the elements of $n^{1/2}(\hat{\underset{\sim}{B}}-\underset{\sim}{B})'$ have a limiting joint rp-variate normal distribution with zero means and covariance between the (i,j)th and (i',j')th elements given by:

$$\sigma^2[\sigma^2(\underset{\sim}{\Delta}^{-1}(\underset{\sim}{I}_p+\underset{\sim}{B}'\underset{\sim}{B})^{-1}\underset{\sim}{\Delta}^{-1}) + \underset{\sim}{\Delta}^{-1}]_{ii'}[\underset{\sim}{I}_r+\underset{\sim}{B}\underset{\sim}{B}']_{jj'} . \tag{3.13}$$

Proof. The asymptotic normality follows from the preceeding arguments. The formula (3.13) may be obtained from (3.10), (3.12), and straightforward calculation. In the computation, it is helpful to note that if $\underset{\sim}{T} = ((\tau_{ij}))$ is defined by (3.7), then

$$\underset{\sim}{T} = \begin{pmatrix} \underset{\sim}{I}_p \\ \underset{\sim}{B} \end{pmatrix} \underset{\sim}{\Delta} \begin{pmatrix} \underset{\sim}{I}_p \\ \underset{\sim}{B} \end{pmatrix}' . \tag{3.14}$$ □

We note that from (2.14) and (2.22),

$$\lim_{n\to\infty} \hat{\underset{\sim}{\Delta}}^{-1}(n^{-1}\underset{\sim}{G}_{11}\underset{\sim}{D}_{max}\underset{\sim}{G}'_{11})\hat{\underset{\sim}{\Delta}}^{-1} = \underset{\sim}{\Delta}^{-1}(\underset{\sim}{\Delta} + \sigma^2(\underset{\sim}{I}_p+\underset{\sim}{B}'\underset{\sim}{B})^{-1})\underset{\sim}{\Delta}^{-1}, \quad \text{a.s.}$$

and from (2.12),

$$\lim_{n\to\infty} (\underset{\sim}{I}_r+\hat{\underset{\sim}{B}}\hat{\underset{\sim}{B}}') = (\underset{\sim}{I}_r+\underset{\sim}{B}\underset{\sim}{B}').$$

It then follows from Theorem 3.2 that an asymptotic 100(1-α)% elliptical confidence region for $\underset{\sim}{B}$ is:

$$\{\underset{\sim}{B}:\ \mathrm{tr}[n(\underset{\sim}{I}_r+\hat{\underset{\sim}{B}}\hat{\underset{\sim}{B}}')^{-1}(\hat{\underset{\sim}{B}}-\underset{\sim}{B})\hat{\underset{\sim}{\Delta}}(n^{-1}\underset{\sim}{G}_{11}\underset{\sim}{D}_{max}\underset{\sim}{G}'_{11})^{-1}\hat{\underset{\sim}{\Delta}}(\hat{\underset{\sim}{B}}-\underset{\sim}{B})'] \tag{3.15}$$
$$\leq r^{-1}(p+r)\hat{\sigma}^2\chi^2_{rp}[1-\alpha]\},$$

where $\chi^2_{rp}[1-\alpha]$ is the 100(1-α)th percentile of the χ^2_{rp} distribution.

Turning next to the question of the asymptotic distribution of $n^{\frac{1}{2}}(r^{-1}(p+r)\hat{\sigma}^2-\sigma^2)$, we note from (2.16), Lemma 2.3, and Theorem 3.1 that

$$n^{-1}\underset{\sim}{D}_{min} = \underset{\sim}{G}_{22}^{-1}(\underset{\sim}{I}_r+\hat{\underset{\sim}{B}}\hat{\underset{\sim}{B}}')^{-1}(\underset{\sim}{B},-\underset{\sim}{I}_r)(n^{-1}\underset{\sim}{W})(\underset{\sim}{B},-\underset{\sim}{I}_r)'(\underset{\sim}{I}_r+\hat{\underset{\sim}{B}}\underset{\sim}{B}')^{-1}(\underset{\sim}{G}'_{22})^{-1} + o_p(n^{-\frac{1}{2}}).$$

Since it also follows directly from Lemma 2.3 that

$$\lim_{n\to\infty} (\underset{\sim}{I}_r+\hat{\underset{\sim}{B}}\underset{\sim}{B}')^{-1}(\underset{\sim}{G}'_{22})^{-1}(\underset{\sim}{G}_{22})^{-1}(\underset{\sim}{I}_r+\underset{\sim}{B}\underset{\sim}{B}')^{-1}$$
$$= (I_r+BB')^{-1}, \qquad \text{a.s.},$$

we conclude that

$$n^{-1}\mathrm{tr}\underset{\sim}{D}_{min} = \mathrm{tr}[(\underset{\sim}{I}_r+\underset{\sim}{B}\underset{\sim}{B}')^{-\frac{1}{2}}(\underset{\sim}{B},-I_r)(n^{-1}\underset{\sim}{W})(\underset{\sim}{B},-\underset{\sim}{I}_r)'(\underset{\sim}{I}_r+\underset{\sim}{B}\underset{\sim}{B}')^{-\frac{1}{2}}] + o_p(n^{-\frac{1}{2}}),$$

or that

$$n^{\frac{1}{2}}(r^{-1}(p+r)\hat{\sigma}^2-\sigma^2)=r^{-1}\mathrm{tr}\{(I_r+\underset{\sim}{B}\underset{\sim}{B}')^{-\frac{1}{2}}(\underset{\sim}{B},-I_r)[n^{-\frac{1}{2}}(\underset{\sim}{W}-\mathscr{E}(\underset{\sim}{W}))](\underset{\sim}{B},-\underset{\sim}{I}_r)'(\underset{\sim}{I}_r+\underset{\sim}{B}\underset{\sim}{B}')^{-\frac{1}{2}}\} + o_p(1).$$

It now follows directly from Theorem 3.1 that the limiting distribution of $n^{\frac{1}{2}}(r^{-1}(p+r)\hat{\sigma}^2-\sigma^2)$ is univariate normal with zero mean, and a variance involving $\underset{\sim}{B}$ and the fourth-order moments of $\underset{\sim}{e}$. [<u>Note</u>. To obtain this result we need not only the assumptions of Theorem 3.1, but also the assumption that $\underset{\sim}{\Delta}$ is positive definite.] In the case when the $\underset{\sim}{e}_i$'s are i.i.d. $N(0,\sigma^2\underset{\sim}{I}_{p+r})$, the variance of the asymptotic distribution greatly simplifies, and we obtain the result:

<u>Theorem 3.3</u>. Under the assumptions of Theorem 3.2,

$$n^{\frac{1}{2}}(r^{-1}(p+r)\hat{\sigma}^2-\sigma^2) \overset{L}{\to} N(0,2\sigma^4r^{-1}). \tag{3.17}$$

Thus, an approximate 100(1-α)% confidence interval for σ^2 is

$$\{\sigma^2: \ |\sigma^2 - (nr)^{-1}\mathrm{tr}\underset{\sim}{D}_{min}| \leq (2\chi_1^2[1-\alpha]/rn)^{\frac{1}{2}}(nr)^{-1}\mathrm{tr}\underset{\sim}{D}_{min}\}. \tag{3.18}$$

Remark. The methods of proof used in this section differ from those usually used to prove asymptotic normality of principal components [see Anderson (1963)] or of factor loadings [see Anderson and Rubin (1956)]. There is, of course, considerable resemblance between the model (1.4) used in this paper, and the kinds of estimators derived, and the models and estimators of principal component analysis and of factor analysis. Indeed, a first step in computing $\hat{\underset{\sim}{B}}$ and $\hat{\sigma}^2$ is to obtain a principal components breakdown of the cross-product matrix $\underset{\sim}{W}$; but we must note that in our model, $\underset{\sim}{W}$ is noncentral Wishart with covariance matrix parameter $\sigma^2\underset{\sim}{I}_{p+r}$, while principal components analysis deals with a central Wishart matrix with a general covariance matrix $\underset{\sim}{\Sigma}$. The analogy of our model to factor analysis with fixed factor values [see Anderson and Rubin (1956) and Lawley (1953)] is much closer, although our model makes very restrictive assumptions about the form of the factor loadings and error covariance matrix. Even though it is probably possible to obtain our large sample results by specializing the more general results of Anderson and Rubin (1956), our approach in this section has the advantage of directness. Further, the representations which we have used may yield information about the accuracy of our large sample approximations in finite samples.

REFERENCES

[1] Anderson, T. W. (1963). Asymptotic theory for principal component analysis. Ann. Math. Statist. 34, 122-148.

[2] Anderson, T. W. and Rubin, H. (1956). Statistical inference in factor analysis. Proc. Third Berkeley Symp. on Math. Statist. and Prob. 5, 111-150. Univ. of California Press.

[3] Bhargava, A. K. (1975). Estimation of a linear transformation, and associated distributional problems. Ph.D. Thesis, Purdue University.

[4] Gleser, L. J. (1965). On the asymptotic theory of fixed-size sequential confidence bounds for linear regression parameters. Ann. Math. Statist. 36, 463-467.

[5] Gleser, L. J. (1966). Correction to "On the asymptotic theory of fixed-size sequential confidence bounds for linear regression parameters. Ann. Math. Statist. 37, 1053-1055.

[6] Gleser, L. J. and Kunte, S. (1976). On asymptotically optimal sequential Bayes interval estimation procedures. Ann. Statist. 4,

[7] Gleser, L. J. and Watson, G. S. (1973). Estimation of a linear transformation. Biometrika 60, 525-534.

[8] Gnedenko, V. and Kolmogorov, A. N. (1954). Limit Distributions for Sums of Independent Random Variables, (translated from the Russian by K. L. Chung). Addison-Wesley, Cambridge, Mass.

[9] Lawley, D. N. (1953). A modified method of estimation in factor analysis and some large sample results. Uppsala Symposium on Psychological Factor Analysis, 17-19 March 1953, 35-42. Almqvist and Wicksell, Uppsala.

[10] Madansky, A. (1959). The fitting of straight lines when both variables are subject to error. J. Amer. Statist. Assoc. 54, 173-205.

[11] Moran, P. A. P. (1971). Estimating structural and functional relationships. J. Multivariate Anal. 1, 232-255.

[12] Perlman, M. D. and Eaton, M. L. (1973). The non-singularity of generalized sample covariance matrices. Ann. Statist. 1, 710-717.

[13] Sprent, P. (1966). A generalized least squares approach to linear functional relationships. J. Roy. Statist. Soc. Ser. B 28, 278-297.

[14] Williams, E. J. (1955). Significance tests for discriminant functions and linear functional relationships. Biometrika 42, 308-381.

FOOTNOTES

[1]Research partially supported by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant AFOSR-72-2350B at Purdue University. The United States Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation hereon.