Discussion Paper 7922 December 1979

On The Maximum Likelihood Method of Factor Analysis

by

Susan S. Navarro

ABSTRACT

Rao's solution of the estimation equations in the maximum likelihood method of factor analysis is derived in this paper in a model wherein Morrison's specific-factor variate e_i is replaced by $\delta_i U_i$ and the covariance structure, by the correlation pattern. The correlation pattern is used, at times, in classifying variables according to the criteria which are specified in section 1 of this paper.

The following innovations are recommended in this paper:

1. The use of δ_i^2 as an indicator of dependence or independence of the ith variable and the other variables in the given set.
2. The application of simultaneous tests of independence among variables having a multivariate normal distribution (see page 3) as part of the factor analysis technique (maximum likelihood method) to determine the validity of the classification of the variables and thereby solve the following problems:
   a. indeterminacy due to the non-uniqueness of solutions of the estimation equations
   b. subjectivity of analysis done with or without the common practice of rotating the factor loading matrix, as observed by Scott

   These tests may be used independently of factor analysis in classifying variables into independent groups. This implies the exclusion of variables which are correlated with independent variables.

ON THE MAXIMUM LIKELIHOOD METHOD OF FACTOR ANALYSIS*

by

Susan S. Navarro

Introduction

Factor analysis is a study of interdependence among variables. Referring to its origin, development and application, Harman says:[1]

> "The birth of factor analysis is generally ascribed to Charles Spearman. His monumental work in developing a psychological theory involving a single general factor and a number of specific factors goes back to 1904... Of course, his 1904 investigation was only the beginning of his work in developing the Two Factor Theory, and his work is not explicitly in terms of 'factors.' Perhaps a more crucial article, certainly insofar as the statistical aspects are concerned, is the 1901 paper by Karl Pearson [386] in which he sets forth 'the method of principal axes'...
>
> Factor analysis is a branch of statistical science, but because of its development and extensive use in psychology the technique itself is often mistakenly considered as psychological theory. The method came into being specifically to provide mathematical models for the explanation of psychological theories of human ability and behavior...
>
> The application of factor analysis techniques has been chiefly in the field psychology. This limitation has no foundation other than the fact that it had its origin in psychology and that accounts of the subject have tended to be '...so bound up with the psychological conception of mental factors that an ordinary statistician has difficulty in seeing it in a proper setting in relation to the general body of statistical method.'"

---

*Comments of Dean José Encarnación, Jr. and Dr. Marcelo Orense on an earlier draft of this paper, as well as suggestions of Dr. Ernesto Pernia and Dr. Roberto Mariano, are gratefully acknowledged.

[1] Harry Harman, Modern Factor Analysis, 2nd ed. (Chicago: The University of Chicago Press, 1967), pp. 3 and 6.

There are numerous methods of factor analysis. Maximum likelihood is the selected method for discussion in this paper for the following reasons:

1. The number of significant common factors may be determined rigorously in the maximum likelihood method.

   Referring to the different methods of factor analysis, Morrison says:[2]

   > "...The various approaches are discussed by Harman [16] in his scholarly and comprehensive text and in summary form by Solomon [38]. While many of the models included 'error' terms reflecting the sampling variation of the observed correlations, none actually used the results of the new discipline of statistical inference. It was not until 1940 that D.N. Lawley reduced the extraction of factor parameters to a problem in maximum likelihood estimation and by so doing eliminated the indeterminacies of the centroid method. Furthermore, the goodness of fit of a solution with just m factors could now be tested rigorously by the generalized likelihood - ratio principle."

   The above mentioned test for goodness of fit is for determining the number of significant common factors.

2. The validity of the classification of variables according to the criteria specified in section 1 of this paper may be verified in the maximum likelihood method, which is

---

[2] Donald Morrison, Multivariate Statistical Methods (New York: McGraw-Hill Book Co., 1967), p. 260.

applicable to a multinormal population, by applying simultaneous tests for independence among variables having a multivariate normal distribution.[3] These tests determine whether or not each variable is independent of each of the other variables in the set.

In practice, conclusions about the classification of the variables are drawn from the values of factor loadings as discussed in section 2. The validity of conclusions is not tested statistically. Scott says:[4]

> "Those familiar with factor analysis can observe the factor loading matrix A and make some subjective analysis of the data based on the factor loadings themselves. When there are more than three factors extracted, however, it becomes difficult even for the experienced factor analyst to draw many conclusions from the original factor loading matrix. Many factor analysts therefore make a rotation on the matrix A...the main advantage of rotation of factor loadings with an orthogonal matrix is in subjective analysis of the factor loadings themselves."

Oster says:[5]

> "To date, there exist no precise sampling error formulas for factor loadings. Approximate procedures, however, were developed by Holzinger and Harman (1941) under certain simplifying assumptions."

---

[3] Donald Morrison, Multivariate Statistical Methods, Chapter 3.

[4] John Scott, Jr. "Factor Analysis and Regression," Econometrica, 34 (1966), 557 and 558.

[5] Gerry Oster, "A Factor Analytic Test of the Theory of the Dual Economy," The Review of Economics and Statistics, 61 (1979), 35.

The exclusion of a variable from the classification into m (representing the number of common factors) groups implies that the variable is considered independent of the other variables in the set. The above mentioned simultaneous tests are useful in identifying these variables and consequently, determine the insignificance of the loadings concerned.

The discussion in this paper is divided into 5 sections. The criteria for classifying variables through factor analysis are specified in section 1. The properties of the factor analysis model (maximum likelihood method), which may help in understanding its uses, are discussed in section 2. The following are the differences between the model presented in this paper and the maximum likelihood model which Morrison presents:[6]

1. Morrison's specific-factor variate e_i is replaced by $\delta_i U_i$ in this paper.

   The important role of δ_i as an indicator of independence or non-independence of the ith variable and the p-1 other variables is discussed in section 2. The variance of e_i in Morrison's model is simply the difference between the variance of the ith variable and the sum of the squares of the loadings of the same variable with each of the m common factors.

---

[6] Donald Morrison, Multivariate Statistical Methods, Chapter 8.

2. Morrison's model explains the covariance structure; the model in this paper shows the correlation pattern among the p variables. The difference is caused by using the standardized value of Y_j, i.e.

$$X_j = \frac{Y_j - \mu_{y_j}}{\sigma_{y_j}}$$

in (2.1) instead of the deviation of Y_j from its mean,

$$Q_j = Y_j - \mu_{y_j}$$

as Morrison does. There is then a difference in the scales of the response variates. Morrison proved the following, called the invariance property of the maximum likelihood loading estimates:

> "Changes in the scales of the response variates only appear as scale changes of the loadings. In particular, the loadings extracted from the correlation matrix differ from those of the covariance matrix only by the factors
>
> $$\frac{1}{s_i}\ ."[7]$$

---

[7]This statement is valid if s_i, instead of σ_{y_i}, is used in determining the standardized variate X_i. Otherwise, the loadings differ by $\frac{1}{\sigma_{y_i}}$, instead of $\frac{1}{s_i}$, as shown in footnote number 14. In practice, σ_{y_i} is usually unknown.

In section 3, the invariance property is used to show that Morrison's estimation equations may be used to determine the solutions of the model in this paper.

The estimation equations are presented in section 4. The results indicate that the solution is not uniquely determined. Harman[8] and Morrison[9] did not present the proofs of the derivation process due to extensive algebraic manipulation and relatively higher mathematical level of discussion, respectively. Harman's estimation equations, which are different from those obtained in this paper, are for Lawley's iterative method for determining the factor loadings. The proofs of the derivation process are presented in Appendix A.

Morrison discussed - without proof - the mathematical foundation of a different method, RaO's[10] iterative solution. Morrison's formulas, transformed into the system presented in this paper through the invariance property of the loadings, are discussed in section 5. The derivation of these formulas is shown in Appendix B.

---

[8] Harry Harman, Modern Factor Analysis, pp. 214-217.

[9] Donald Morrison, Multivariate Statistical Methods, pp. 264-267.

[10] C. Radhakrishna RaO, "Estimation and Tests of Significance in Factor Analysis," Psychometrika 20 (1955), 105-106.

1. The Problem

A set of variables may be classified such that

1. those variables that are highly intercorrelated are in the same group and

2. variables that are in a group are independent of the variables that are not in the same group. Variables which belong to more than one group are not independent of the variables - which are in different groups - with which they are highly intercorrelated.

Factor analysis may be used to determine the classification in this case.

2. Factor Analysis Model (Maximum Likelihood Method)

Given the following variables

$$Y_1, Y_2 \dots Y_p$$

which are measured from the n elements of a sample from a given population. Let

$$X_j = \frac{Y_j - \mu_{y_j}}{\sigma_{y_j}}$$

where $\mu_{y_j} = E(Y_j)$ and

$$\sigma^2_{y_j} = E[Y_j - \mu_{y_j}]^2$$

The basic factor analysis model is

$$(2.1) \qquad X_j = \sum_{k=1}^{m} \alpha_{jk} F_k + \delta_j U_j \qquad (j = 1, 2, \ldots, p)$$

where each of the standardized variables X_j is expressed linearly in terms of m (usually less than p) common factors F_k, $k = 1, 2, \ldots, m$, and a unique factor U_j.[11] α_{jk}, which is called a "loading," and δ_j are population values and are estimated from a given sample. δ_j^2 is referred to as "uniqueness."

The main problems in factor analysis are:

1. to estimate α_{jk} and δ_j
2. to determine the number of significant common factors[12]
3. to test the independence of those variables which are classified under a group from those which are not in the same group.[13]

---

[11]In some models, X_j is replaced by $Q_j = Y_j - \mu_{y_j}$ in equation (2.1). Note that $E(Q_j) = 0$. For example, see Donald Morrison, Multivariate Statistical Methods, p. 261.

[12]Harry Harman, Modern Factor Analysis, pp. 219-221.

[13]This is a recommendation in this paper. The suggested tests are specified in footnote number 3.

A variable may be classified, together with other variables, under a common factor or as a single element under a unique factor. The manner of classification is discussed in pages 13 and 14.

Assumptions

1. Y_j is normally distributed with mean μ_{y_j} and variance $\sigma^2_{y_j}$, $j = 1, 2, \ldots, p$. Consequently, $X_1, X_2, \ldots, X_p$ are normally distributed with zero means and unit variances.

2. $F_1, F_2, \ldots, F_m, U_1, U_2, \ldots, U_p$ are mutually stochastically independent, normally distributed random variables with zero means and unit variances.

Notation for Matrices

The following symbols will be used in this paper:

Matrix		Order	Definition
Population	Sample		
$\Sigma = (\sigma_{jq})$	$S = (s_{jq})$	$p \times p$	covariance matrix
$\rho = (\rho_{jq})$	$R = (r_{jq})$	$p \times p$	correlation matrix
$\Lambda = (\alpha_{jk})$	$A = (a_{jk})$	$p \times m$	matrix of common factor coefficients
$\delta = (\delta^2_j)$	$D = (d^2_j)$	$p \times p$	diagonal matrix of uniquenesses
$\Omega = (\delta_j)$		$p \times p$	diagonal matrix of square roots of uniquenesses

Properties of the model[14]

1. α_{jq} is the correlation coefficient of X_j and F_q.

Proof:

$$\rho_{X_jF_q} = \frac{E(X_jF_q) - E(X_j)E(F_q)}{\sigma_{X_j}\ \sigma_{F_q}}$$

$$= E\ [\sum_{k=1}^{m} \alpha_{jk} F_k + \delta_j U_j]\ F_q$$

$$= E[\alpha_{j1}F_1F_q + \alpha_{j2}F_2F_q + \ldots + \alpha_{jq}F_qF_q + \ldots + \alpha_{jm}F_mF_q]$$

$$+ E\ [\delta_j U_j F_q]$$

$$= \alpha_{jq}$$

---

[14]If Q_j, instead of X_j, is used in equation 2.1 then

$$\rho_{Q_jF_q} = \frac{\alpha'_{jq}}{\sigma_{Q_j}}, \quad \rho_{Q_jU_j} = \frac{\delta'_j}{\sigma_{Q_j}}, \quad \rho_{Q_jQ_q} = \frac{\sum_{k=1}^{m} \alpha'_{jk}\ \alpha'_{qk}}{\sigma_{Q_j}\ \sigma_{Q_q}}$$

and $$\sigma^2_{Q_j} = \sum_{k=1}^{m} \alpha'^{\,2}_{jk} + \delta'^{\,2}_j$$

note that $\sigma_{Q_j} = \sigma_{y_j}$

2. δ_j is the correlation coefficient of X_j and U_j.

Proof:

$$\rho_{X_j U_j} = \frac{E(X_j U_j) - E(X_j)E(U_j)}{\sigma_{X_j}\sigma_{U_j}}$$

$$= E\left[\sum_{k=1}^{m} \alpha_{jk} F_k + \delta_j U_j\right] U_j$$

$$= E\left[\sum_{k=1}^{m} \alpha_{jk} F_k U_j\right] + E\left[\delta_j U_j^2\right]$$

$$= \delta_j$$

3. If $j \neq t$ then X_j is uncorrelated with U_t.

Proof:

$$\rho_{X_j U_t} = \frac{E(X_j U_t) - E(X_j)E(U_t)}{\sigma_{X_j}\sigma_{U_t}}$$

$$= E\left[\sum_{k=1}^{m} \alpha_{jk} F_k + \delta_j U_j\right] U_t$$

$$= E\left[\sum_{k=1}^{m} \alpha_{jk} F_k U_t\right] + E\left[\delta_j U_j U_t\right]$$

$$= 0$$

4. The correlation coefficient of X_j and X_q is

$$\sum_{k=1}^{m} \alpha_{jk}\, \alpha_{qk}$$

Proof:

$$\rho_{X_j X_q} = \frac{E(X_j X_q) - E(X_j)\, E(X_q)}{\sigma_{X_j}\, \sigma_{X_q}}$$

$$= E[\sum_{k=1}^{m} \alpha_{jk}\, F_k + \delta_j U_j]\, [\sum_{k=1}^{m} \alpha_{qk}\, F_k + \delta_q U_q]$$

$$= E[\sum_{k=1}^{m} \alpha_{jk}\, \alpha_{qk}\, F_k^2] + E[\delta_j \delta_q\, U_j\, U_q]$$

$$= \sum_{k=1}^{m} \alpha_{jk}\, \alpha_{qk}.$$

The model consists of the following set of p equations, called a factor pattern:

$$X_1 = \alpha_{11}\, F_1 + \alpha_{12}\, F_2 + \ldots + \alpha_{1m}\, F_m + \delta_1 U_1$$

$$X_2 = \alpha_{21}\, F_1 + \alpha_{22}\, F_2 + \ldots + \alpha_{2m}\, F_m + \delta_2 U_2$$

$$\vdots$$

$$X_p = \alpha_{p1}\, F_1 + \alpha_{p2}\, F_2 + \ldots + \alpha_{pm}\, F_m + \delta_p U_p$$

with the following properties:

1. In each of the equations in the pattern, the coefficient of a factor is its correlation coefficient with the variable in the given equation.

2. A unique factor U_j is uncorrelated with all the variables X_k where $k \neq j$.

3. $\sum \alpha_{jk}^2 + \delta_j^2 = \sigma_{X_j}^2 = 1.$

4. The correlation coefficient of X_j and X_q is

$$\sum_{k=1}^{m} \alpha_{jk} \alpha_{qk}$$

5. The common and the unique factors are uncorrelated among themselves.

In practice, researchers classify X_j under F_k if

$$\max \left(|\hat{\alpha}_{j1}|, |\hat{\alpha}_{j2}|, \ldots, |\hat{\alpha}_{jm}| \right) = |\hat{\alpha}_{jk}|$$

and $|\hat{\alpha}_{jk}|$ is not too small. All those variables which are classified under the same common factor F_k are considered as belonging to the same group.[15] If X_q is also classified under F_k then

---

[15] For example, see Emmanuel Velasco, "Span of Control: A Comparative Analytic Approach," The Philippine Review of Business and Economics, 10 (1973). The principal component method is used in this paper. The resulting factors are then rotated by the use of the varimax method to arrive at an orthogonal multiple factor solution.

$$\max\left(|\hat{\alpha}_{j1}\hat{\alpha}_{q1}|,\ |\hat{\alpha}_{j2}\hat{\alpha}_{q2}|,\ \ldots,\ |\hat{\alpha}_{jm}\hat{\alpha}_{qm}|\right) = |\hat{\alpha}_{jk}\hat{\alpha}_{qk}|.$$

For $|\hat{\alpha}_{jk}\hat{\alpha}_{qk}|$ to be at least 0.49, which is less than $\frac{1}{2}$ of the maximum absolute value of a correlation coefficient, each of $|\hat{\alpha}_{jk}|$ and $|\hat{\alpha}_{qk}|$ should be at least 0.70. If $|\hat{\alpha}_{jk}|$ is 0.9 then $|\hat{\alpha}_{qk}|$ should be at least 0.6 for $|\hat{\alpha}_{jk}\hat{\alpha}_{qk}|$ to be at least 0.54. The terms in $\sum_{k=1}^{m} \hat{\alpha}_{jk}\hat{\alpha}_{qk}$, which may differ in signs, determine the estimate of the population correlation coefficient of X_j and X_q. We have shown that the loadings may be used to formulate hypotheses regarding the classification of variables into groups, which implies non-zero population correlation coefficients of those belonging to the same group.

If $\hat{\delta}_j^2$ is very close to 1 then $\sum_{k=1}^{m} \hat{\alpha}_{jk}^2$ is very close to 0. Consequently,

$$\hat{\rho}_{X_jX_q} = \sum_{k=1}^{m} \hat{\alpha}_{jk}\hat{\alpha}_{qk}, \quad q = 1, 2, \ldots, j-1, \ j+1, \ldots, p$$

is expected to be small. X_j is then classified as a single element under the factor U_j. The independence of X_j from the $p-1$ other variables in the set may be hypothesized.

The tests of independence[16] among variables in a multivariate normal population should be used to determine the validity of the

---

[16] See Donald Morrison, Multivariate Statistical Methods, Chapter 3.

classification of the variables according to the criteria specified in section 1. Variables that should belong to more than one group, because they are highly correlated with independent groups of variables, may be identified through these tests.

3. The Invariance Property of the Maximum Likelihood Loading Estimates

A proof of the invariance property of the maximum likelihood loading estimates may be deduced from footnote number 14 and the corresponding derivations in section 2.

In Morrison's model,

$$V(Q_j) = \sum_{k=1}^{m} \alpha'^{2}_{jk} + V(e_j). \tag{3.1}$$

Dividing both sides of (3.1) by $\sigma^2_{y_j}$, we get

$$1 = \sum_{k=1}^{m} \alpha^2_{jk} + \frac{V(e_j)}{\sigma^2_{y_j}} \tag{3.2}$$

by the invariance property of the loadings. We note that $V(Q_j) = \sigma^2_{y_j}$.

In our model,

$$(3.3)\qquad V(X_j) = 1 = \sum_{k=1}^{m} \alpha_{jk}^2 + \delta_j^2$$

(3.2) and (3.3) imply that

$$(3.4)\qquad \delta_j^2 = \frac{V(e_j)}{\sigma_{y_j}^2} = \frac{\Psi_j}{\sigma_{y_j}^2}\,.$$ [17]

We have shown that Morrison's estimation equations may be used to determine the solution of the model in this paper.

4. Estimates of Factor Loadings

The likelihood function of the sample covariance matrix S is defined by Wishart's distribution function as

$$f(S) = K|\Sigma|^{-1/2\,(n-1)}\; S^{1/2\,(n-p-2)}\; \exp\left(-\frac{n-1}{2}\,\mathrm{tr}\Sigma^{-1}\,S\right).$$

In terms of logarithms,

$$\ln f(S) = \ln k - \frac{1}{2}(n-1)\ln|\Sigma| + \frac{1}{2}(n-p-2)\ln S - \frac{n-1}{2}\,\mathrm{tr}\Sigma^{-1}\,S$$

which may be simplified as follows:

---

[17] Morrison defines $V(e_j)$ as Ψ_j.

$$L = -\frac{2}{n-1} \ln f(S) = \ln |\Sigma| + tr\Sigma^{-1} S + Q$$

$$= \ln |\Lambda\Lambda' + \delta| + tr(\Lambda\Lambda' + \delta)^{-1} S + Q$$

where Q is a function which is independent of Σ.[18]

The maximum likelihood estimates of the loadings are obtained by imposing the conditions that

$$\frac{\partial L}{\partial \delta_j} = \frac{\partial L}{\partial \alpha_{jk}} = 0 \qquad j = 1, 2, \ldots, p \text{ and } k = 1, 2, \ldots, m.$$

The resulting estimation equations are as follows:[19]

If $\frac{\partial L}{\partial \delta_j} = 0$, $\qquad j = 1, 2, \ldots, p$, then

(4.1) $\qquad \text{diag}(\hat{\Sigma}^{-1}) = \text{diag}(\hat{\Sigma}^{-1} S \hat{\Sigma}^{-1})$.

If $\frac{\partial L}{\partial \alpha_{jk}}$, $\qquad j = 1, 2, \ldots, p$ and $k = 1, 2, \ldots, m$ then

(4.2) $\qquad S\hat{\Sigma}^{-1} \hat{\Lambda} = \hat{\Lambda}$.

Since $\hat{\Sigma}^{-1} \hat{\Lambda} = \hat{\delta}^{-1} \hat{\Lambda} (I + \hat{\Lambda}' \hat{\delta}^{-1} \hat{\Lambda})^{-1}$ then the estimation equation (4.2) may be written as

---

[18] $\Sigma = \Lambda \Lambda' + \delta$.

[19] (4.1) and (4.2) imply that diag $(\hat{\Sigma})$ = diag (S). Consequently, the solution is not unique.

$$(4.3) \qquad S\ \hat{\delta}^{-1}\ \hat{\Lambda}\ (I + \hat{\Lambda}'\ \hat{\delta}^{-1}\ \hat{\Lambda})^{-1} = \hat{\Lambda} \text{ or}$$

$$(4.4) \qquad (S-\hat{\delta})\ \hat{\delta}^{-1}\ \hat{\Lambda} = \hat{\Lambda}\ (\hat{\Lambda}'\ \hat{\delta}^{-1}\ \hat{\Lambda}).$$

Premultiplication of both sides of (4.4) by $\hat{\Omega}^{-1}$ yields

$$(4.5) \qquad [\hat{\Omega}^{-1}\ (S-\hat{\delta})\ \hat{\Omega}^{-1}]\ \hat{\Omega}^{-1}\ \hat{\Lambda} = \hat{\Omega}^{-1}\ \hat{\Lambda}\ (\hat{\Lambda}'\ \hat{\delta}^{-1}\ \hat{\Lambda}).$$

5. RaO's Iterative Solution of the Resulting Maximum Likelihood Estimation Equations

In section 4, we showed that

$$(4.5) \qquad [\hat{\Omega}^{-1}\ (S-\hat{\delta})\ \hat{\Omega}^{-1}]\ \hat{\Omega}^{-1}\ \hat{\Lambda} = \hat{\Omega}^{-1}\ \hat{\Lambda}\ (\hat{\Lambda}'\ \hat{\delta}^{-1}\ \hat{\Lambda}).$$

If $\hat{\Lambda}'\ \hat{\delta}^{-1}\ \hat{\Lambda}$ is a diagonal matrix[20] then the characteristic roots of $\hat{\Omega}^{-1}\ (S-\hat{\delta})\ \hat{\Omega}^{-1}$ are equal to the successive elements of $\hat{\Lambda}'\ \hat{\delta}^{-1}\ \hat{\Lambda}$ and hence the ith column of $\hat{\Omega}^{-1}\ \hat{\Lambda}$ is merely the characteristic vector corresponding to the ith largest root of $\hat{\Omega}^{-1}\ (S-\hat{\delta})\ \hat{\Omega}^{-1}$. The elements of $\hat{\delta}$ are also unknown and may be estimated from the equation $\hat{\delta} = \text{diag}\ (S-\hat{\Lambda}\hat{\Lambda}')$.

---

[20] If $\hat{\Lambda}'\ \hat{\delta}^{-1}\ \hat{\Lambda}$ is a diagonal matrix than

$$\sum_{q=1}^{p} \frac{\hat{\alpha}_{qs}\ \hat{\alpha}_{qr}}{\hat{\delta}_q^2} = 0$$

if $s \neq r$.

Numerical Solution of the Estimation Equations

The iterative process follows this plan:[21]

(a) Compute the greatest characteristic root ℓ_{10} and its vector a_{10} of S, where the elements of the vector have been scaled such that $a'_{10} a_{10} = \ell_{10}$.

(b) Approximate the specific variances from

$$\hat{\delta}_{10} = \text{diag}\,(S - a_{10} a'_{10})$$

where in the sequel the subscripts i, j of δ and a will denote the jth iterates of the i-factor solution.

(c) Form the matrix

$$\hat{\Omega}_{10}^{-1} (S - \hat{\delta}_{10}) \hat{\Omega}_{10}^{-1}$$

and extract the vector a_{11} associated with its greatest root ℓ_{11}. Scale the elements so that $a'_{11} a_{11} = \ell_{11}$ and premultiply the vector by $\hat{\Omega}_{10}$ to obtain the first approximation $\hat{\lambda}_{11}$ to the single column of $\hat{\Lambda}_1$

(d) Compute

$$\hat{\delta}_{11} = \text{diag}\,(S - \hat{\lambda}_{11} \hat{\lambda}'_{11})$$

and repeat the process for the second approximation to $\hat{\Lambda}_1$. Continue in this fashion until the corresponding elements of the successive iterates

---

[21]This is patterned after the presentation of the iterative process in Donald Morrison, Multivariate Statistical Methods, pp. 271-272.

$\hat{\lambda}_{1i}$ and $\hat{\lambda}_{1,\ i+1}$ do not differ by more than some predetermined amount. The resulting column vector $\hat{\Lambda}_1$ will contain the maximum likelihood estimates of the loadings for the one-factor model.

To obtain the estimated loadings of the second, third, ..., mth factors:

(e) Compute the residual matrix $S_1 = S - \hat{\Lambda}_1 \hat{\Lambda}_1'$ of the single factor solution.

(f) Compute the greatest characteristic root ℓ_{20} and its vector a_{20} of S_1, where the elements of the vector have been scaled such that $a_{20}' a_{20} = \ell_{20}$. a_{20} is taken as the initial approximation to the loadings of the second factor.

(g) From the single-factor solution and the new initial vector form the px2 matrix

$$\hat{\Lambda}_{20} = [\hat{\Lambda}_1 \quad a_{20}]$$

for the zero order approximation to the estimated loadings of the two factor model.

(h) Approximate δ_{20} from

$$\hat{\delta}_{20} = \text{diag}\ (S - \hat{\Lambda}_{20} \hat{\Lambda}_{20}')$$

(i) Form the matrix

$$\hat{\Omega}_{20}^{-1} \ (S - \hat{\delta}_{20})\ \hat{\Omega}_{20}^{-1}$$

and extract the first two of its largest characteristic roots, ℓ_{11} and ℓ_{21}. Compute the characteristic vectors, a_{11} and a_{21}, corresponding to ℓ_{11} and ℓ_{21} respectively. The elements of these vectors have been scaled to the usual loading form. It is essential to note that a_{11} is not equal to the first **iterate** a_{11} of the single factor solution.

(j) Premultiply $[a_{11} \;\; a_{21}]$ by $\hat{\Omega}_{20}$ to obtain the first approximation $\hat{\Lambda}_{21}$ to the loading estimates of the two factor model. Repeat the process until all the elements of the iterates $\hat{\Lambda}_{2i}$ have converged with specified accuracy to the two factor solution $\hat{\Lambda}_2$.

The solution of the m-factor model begins in like manner from the (m-1)-factor solution: those latter estimates provide the starting values for the m-1 factors of the new model, while the m^{th} trial vector is found from the characteristic vector of the greatest root of

$$S_{m-1} = S - \hat{\Lambda}_{m-1} \hat{\Lambda}_{m-1}'.$$

The iterative process is repeated until the elements of $\hat{\Lambda} \equiv \hat{\Lambda}$ have converged with appropriate accuracy.

Appendix A

The likelihood function of S is defined by Wishart's distribution function as

$$f(S) = K\ |\Sigma|^{-1/2\ (n-1)}\ S^{\ 1/2\ (n-p-2)}\ \exp\left(-\frac{n-1}{2}\ \mathrm{tr}\ \Sigma^{-1}\ S\right).$$

In terms of logarithms,

$$\ln f(S) = \ln k - \frac{1}{2}(n-1)\ln|\Sigma| + \frac{1}{2}(n-p-2)\ln S$$
$$- \frac{n-1}{2}\ \mathrm{tr}\ \Sigma^{-1}\ S$$

which may be simplified as follows:

$$L = -\frac{2}{n-1}\ln f(S) = \ln|\Sigma| + \mathrm{tr}\ \Sigma^{-1}\ S + Q$$
$$= \ln|\Lambda\Lambda' + \delta| + \mathrm{tr}\ (\Lambda\Lambda' + \delta)^{-1}\ S + Q$$

where Q is a function which is independent of Σ.

The maximum likelihood estimates of the loadings are obtained by imposing the conditions that

$$\frac{\partial L}{\partial \delta_j} = \frac{\delta L}{\partial \alpha_{jk}} = 0 \quad , \quad j = 1, 2, \ldots\ldots, p \text{ and}$$
$$k = 1, 2, \ldots\ldots, m.$$

The resulting equations are given in theorems A.1 and A.5.

Theorem A.1

If $\frac{\partial L}{\partial \delta_j} = 0$, $(j = 1, 2, \ldots., p)$ then

$$\operatorname{diag}(\hat{\Sigma}^{-1}) = \operatorname{diag}(\hat{\Sigma}^{-1} S \hat{\Sigma}^{-1}).$$

<u>Proof</u>:

$$\frac{\partial L}{\partial \delta_j} = \frac{1}{|\Sigma|} \frac{\partial |\Sigma|}{\partial \delta_j} - \operatorname{tr} \Sigma^{-1} \frac{\partial \Sigma}{\partial \delta_j} \Sigma^{-1} S$$

$$\text{(A.1)} \quad \frac{\partial L}{\partial \delta_j} = \frac{2\delta_j \sigma^{jj}}{|\Sigma|} - \operatorname{tr} \Sigma^{-1} (2\delta_j T_{jj}) \Sigma^{-1} S = 0$$

where σ^{jj} is the cofactor of the j^{th} diagonal element of Σ and T_{jj} is the pxp matrix with unity in its j^{th} diagonal position and zeros elsewhere. We note the following:

(a) $\frac{\sigma^{jj}}{|\Sigma|}$ is the j^{th} diagonal element of Σ^{-1}

(b) $\Sigma^{-1} T_{jj}$ = the pxp matrix all of whose elements are zeros except the j^{th} diagonal element which is equal to $\frac{\sigma^{jj}}{|\Sigma|}$

(c) Since S and Σ^{-1} are symmetric then $\Sigma^{-1} S = S \Sigma^{-1}$

(d) $\operatorname{tr} \Sigma^{-1} T_{jj} S\Sigma^{-1}$ = the j^{th} diagonal element of $\Sigma^{-1} S \Sigma^{-1}$.

Therefore, (A.1) implies that

$$\text{(A.2)} \quad \operatorname{diag}(\hat{\Sigma}^{-1}) = \operatorname{diag}(\hat{\Sigma}^{-1} S \hat{\Sigma}^{-1}).$$

The derivative of the likelihood function with respect to α_{jk} follows:

$$\text{(A.3)} \quad \frac{\partial L}{\partial \alpha_{jk}} = \frac{1}{|\Sigma|} \frac{\partial |\Sigma|}{\partial \alpha_{jk}} - \operatorname{tr} \Sigma^{-1} \frac{\partial \Sigma}{\partial \alpha_{jk}} \Sigma^{-1} S.$$

The second term in the differentiation of L is obtained by differentiating tr Σ^{-1}.

In theorem A.2, we shall show that the pm expressions in $\frac{1}{|\Sigma|}\frac{\partial|\Sigma|}{\partial\alpha_{jk}}$ can be expressed as $2\,\Sigma^{-1}\Lambda$. The pm expressions in tr $\Sigma^{-1}\frac{\partial\Sigma}{\partial\alpha_{jk}}\Sigma^{-1}S$ can be expressed as $2\,\Sigma^{-1}S\,\Sigma^{-1}\Lambda$. This statement is proved in theorem A.4.

By applying theorems A.2 and A.4, we shall prove in theorem A.5 that the pm equations $\frac{\partial L}{\partial\alpha_{jk}} = 0$ imply that $S\,\hat{\Sigma}^{-1}\hat{\Lambda}$.

Theorem A.2

The pm expressions $\frac{1}{|\Sigma|}\frac{\partial|\Sigma|}{\partial\alpha_{jk}}$, $j = 1, 2, \ldots, p$ and $k = 1, 2, \ldots, m$, can be expressed as $2\Sigma^{-1}\Lambda$.

Proof:

$$2\,\Sigma^{-1}\Lambda = \frac{2}{|\Sigma|}\begin{bmatrix} \sigma^{11} & \sigma^{12} & \cdots & \sigma^{1p} \\ \sigma^{21} & \sigma^{22} & \cdots & \sigma^{2p} \\ \cdot & \cdot & \cdots & \cdot \\ \sigma^{p1} & \sigma^{p2} & \cdots & \sigma^{pp} \end{bmatrix}\begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1m} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2m} \\ \cdot & \cdot & \cdots & \cdot \\ \alpha_{p1} & \alpha_{p2} & \cdots & \alpha_{pm} \end{bmatrix}$$

where $\sigma^{jk} = \sigma^{kj}$ = the cofactor of σ_{jk} = the cofactor of σ_{kj}. Therefore, the element in the j^{th} row, k^{th} column of $2\Sigma^{-1}\Lambda$ is $\frac{2}{|\Sigma|} \sum_{q=1}^{p} \alpha_{qk} \sigma^{qj}$. On the other hand,

$$|\Sigma| = \sum_{r_1 r_2 \cdots r_p} (-1)^{[r_1 r_2 \cdots r_p]} \sigma_{1r_1} \sigma_{2r_2} \cdots \sigma_{pr_p}$$

where $r_1 \, r_2 \ldots r_p$ is a permutation of the numbers 1, 2, ..., p. $[r_1 \, r_2 \ldots r_p]$ is the number of inversions in the permutation of the numbers 1, 2, ..., p. An inversion occurs when a larger number precedes a smaller number. Therefore,

$$\frac{\partial|\Sigma|}{\partial\alpha_{jk}} = \alpha_{1k}\sigma^{1j} + \alpha_{1k}\sigma^{j1} + \cdots + \alpha_{pk}\sigma^{pj} + \alpha_{pk}\sigma^{jp}$$

$$= 2\alpha_{1k}\sigma^{1j} + 2\alpha_{2k}\sigma^{2j} + \cdots + 2\alpha_{pk}\sigma^{pj}$$

or

$$\frac{1}{|\Sigma|}\frac{\partial|\Sigma|}{\partial\alpha_{jk}} = \frac{2}{|\Sigma|}\sum_{q=1}^{p} \alpha_{qk}\sigma^{qj}$$ = the element in the j^{th} row, k^{th} column of $2\,\Sigma^{-1}\Lambda$.

The following theorem will be used in proving theorem A.4 which simplifies the second expression of equation (A.3).

Theorem A.3

$\Sigma^{-1} S \Sigma^{-1}$ is symmetric.

Econ. 027 f

Proof:

$$\text{(A.4)} \quad \frac{1}{|\Sigma|^2} \sum_{r=1}^{p} \sum_{q=1}^{p} \sigma^{jq} \; S_{qr} \; \sigma^{rk}$$

is the element in the j^{th} row, k^{th} column of $\Sigma^{-1}\ S\Sigma^{-1}$. Since S is symmetric then $S_{qr} = S_{rq}$. Also $\sigma^{jp} = \sigma^{qj}$ and $\sigma^{rk} = \sigma^{kr}$. Therefore, (A.4) may be expressed as

$$\text{(A.5)} \quad \frac{1}{|\Sigma|^2} \sum_{q=1}^{p} \sum_{r=1}^{p} \sigma^{kr} \; S_{rq} \; \sigma^{qj}.$$

(A.5) is the element in the k^{th} row, j^{th} column of $\Sigma^{-1}\ S\Sigma^{-1}$. Therefore, $\Sigma^{-1}\ S\Sigma^{-1}$ is symmetric.

Theorem A.4

The pm expressions, tr $\Sigma^{-1} \dfrac{\partial \Sigma}{\partial \alpha_{jk}} \Sigma^{-1}$ S, can be expressed as elements of the matrix $2\Sigma^{-1}\ S\Sigma^{-1}\ \Lambda$.

Proof:

$$\frac{\partial \Sigma}{\partial \alpha_{jk}} = \begin{bmatrix} 0 & 0 & \cdots & \alpha_{1k} & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \alpha_{1k} & \cdot & \cdots & 2\alpha_{jk} & \cdots & \alpha_{pk} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \alpha_{pk} & \cdots & 0 \end{bmatrix}$$

is a symmetric matrix with zeros in all positions except those of the j^{th} row and column.

Since Σ^{-1}, S and $\frac{\partial \Sigma}{\partial \alpha_{jk}}$ are symmetric matrices then

$$\operatorname{tr} \Sigma^{-1} \frac{\partial \Sigma}{\partial \alpha_{jk}} \Sigma^{-1} S = \operatorname{tr} \Sigma^{-1} S \Sigma^{-1} \frac{\partial \Sigma}{\partial \alpha_{jk}} =$$

$$(A.6)\ \operatorname{tr} \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ c_{p1} & c_{p2} & \cdots & c_{pp} \end{bmatrix} \begin{bmatrix} 0 & \cdots & \alpha_{1k} & \cdots & 0 \\ \cdot & \cdots & \cdot & \cdots & \cdot \\ \alpha_{1k} & \cdots & 2\alpha_{jk} & \cdots & \alpha_{pk} \\ \cdot & \cdots & \cdot & \cdots & \cdot \\ 0 & \cdots & \alpha_{pk} & \cdots & 0 \end{bmatrix}$$

where c_{jq} is the element in the j^{th} row, q^{th} column of $\Sigma^{-1} S \Sigma^{-1}$.

(A.6) may be expressed as

$$\text{(A.7)}\ \operatorname{tr}\begin{bmatrix} c_{1j}\,\alpha_{1k}\ \cdots\ \sum_{q=1}^{p} c_{1q}\,\alpha_{qk} + c_{1j}\,\alpha_{jk}\ \cdots\ c_{1j}\,\alpha_{pk} \\ c_{2j}\,\alpha_{1k}\ \cdots\ \sum_{q=1}^{p} c_{2q}\,\alpha_{qk} + c_{2j}\,\alpha_{jk}\ \cdots\ c_{2j}\,\alpha_{pk} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ c_{pj}\,\alpha_{1k}\ \cdots\ \sum_{q=1}^{p} c_{pq}\,\alpha_{qk} + c_{qj}\,\alpha_{jk}\ \cdots\ c_{pj}\,\alpha_{pk} \end{bmatrix}$$

$$= c_{1j}\,\alpha_{1k} + c_{2j}\,\alpha_{2k} + \cdots + \sum_{q=1}^{p} c_{jq}\,\alpha_{qk} + c_{jj}\,\alpha_{jk} + \cdots + c_{pj}\,\alpha_{pk}.$$

Since $\Sigma^{-1}\ S\ \Sigma^{-1}$ is symmetric, then $c_{jq} = c_{qj}$ Therefore, (A.7) is equal to

$$2\ c_{j1}\,\alpha_{1k} + 2\ c_{j2}\,\alpha_{2k} + \cdots + 2\ c_{jp}\,\alpha_{pk},$$

the element in the j^{th} row, k^{th} column of $2\ c\ \Lambda = 2\ \Sigma^{-1}\ S\ \Sigma^{-1}\ \Lambda$.

<u>Theorem A.5</u>

If $\dfrac{\partial L}{\partial \alpha_{jk}} = 0$, $(j = 1, 2, \ldots, p$ and $k = 1, 2, \ldots, m)$, then

$$S\ \hat{\Sigma}^{-1}\ \hat{\Lambda} = \hat{\Lambda}.$$

<u>Proof</u>:

$$\text{(A.8)}\quad \frac{\partial L}{\partial \alpha_{jk}} = \frac{1}{|\Sigma|}\ \frac{\partial |\Sigma|}{\partial \alpha_{jk}} - \operatorname{tr}\ \Sigma^{-1}\ \frac{\partial \Sigma}{\partial \alpha_{jk}}\ \Sigma^{-1}\ S$$

$$= 0.$$

Applying theorems A.2 and A.4, we get

$$\frac{\partial L}{\partial \alpha_{jk}} = 2\,\Sigma^{-1}\,\Lambda \;-\; 2\,\Sigma^{-1}\,S\,\Sigma^{-1}\,\Lambda = 0$$

or

$$S\,\hat{\Sigma}^{-1}\,\hat{\Lambda} = \hat{\Lambda}. \tag{A.9}$$

(A.9) will be used in combination with (A.2) to get the estimates.

The maximum likelihood estimation equations have been derived in theorems A.1 and A.5. In the following theorem, we shall show that these equations imply that

$$\mathrm{diag}\,(\,\hat{\Sigma}\,) = \mathrm{diag}\,(S).$$

Theorem A.6

$$\mathrm{diag}\,(\,\hat{\Sigma}\,) = \mathrm{diag}\,(S)$$

Proof:

In theorems A.1 and A.5, we showed that the maximum likelihood estimates of the loadings should satisfy the following conditions:

$$\mathrm{diag}\,(\,\hat{\Sigma}^{-1}\,) = \mathrm{diag}\,(\,\hat{\Sigma}^{-1}\,S\,\Sigma^{-1}\,) \text{ and} \tag{A.2}$$

$$S\,\hat{\Sigma}^{-1}\,\hat{\Lambda} = \hat{\Lambda}. \tag{A.9}$$

Pre- and post-multiply (A.2) by $\hat{\delta} = \hat{\Sigma} - \hat{\Lambda}\hat{\Lambda}'$ to get

$$\text{diag}\ (\hat{\Sigma} - \hat{\Lambda}\hat{\Lambda}')\ \hat{\Sigma}^{-1}\ (\hat{\Sigma} - \hat{\Lambda}\hat{\Lambda}') =$$

$$\text{diag}\ (\hat{\Sigma} - \hat{\Lambda}\hat{\Lambda}')\ \hat{\Sigma}^{-1}\ S\ \hat{\Sigma}^{-1}\ (\hat{\Sigma} - \hat{\Lambda}\hat{\Lambda}')$$

$$\text{diag}\ (\hat{\Sigma} - 2\hat{\Lambda}\hat{\Lambda}' + \hat{\Lambda}\hat{\Lambda}'\hat{\Sigma}^{-1}\hat{\Lambda}\hat{\Lambda}') =$$

$$\text{diag}\ (S - S\hat{\Sigma}^{-1}\hat{\Lambda}\hat{\Lambda}' - \hat{\Lambda}\hat{\Lambda}'\hat{\Sigma}^{-1}S + \hat{\Lambda}\hat{\Lambda}'\hat{\Sigma}^{-1}S\hat{\Sigma}^{-1}\hat{\Lambda}\hat{\Lambda}')$$

Simplifying the terms on the right side of the equation by applying (A.8), we get

$$\text{diag}\ (\hat{\Sigma} - 2\hat{\Lambda}\hat{\Lambda}' + \hat{\Lambda}\hat{\Lambda}'\hat{\Sigma}^{-1}\hat{\Lambda}\hat{\Lambda}') =$$

$$\text{diag}\ (S - 2\hat{\Lambda}\hat{\Lambda}' + \hat{\Lambda}\hat{\Lambda}'\hat{\Sigma}^{-1}\hat{\Lambda}\hat{\Lambda}')$$

or

(A.10) $\text{diag}\ (\hat{\Sigma}) = \text{diag}\ (S)$.

Theorem A.7 may be used to obtain (A.11) and (A.12).

<u>Theorem A.7</u>

$$\hat{\Sigma}^{-1}\hat{\Lambda} = \hat{\delta}^{-1}\hat{\Lambda}\ (I + \hat{\Lambda}'\hat{\delta}^{-1}\hat{\Lambda})^{-1}$$

<u>Proof</u>:

Since $\hat{\Sigma} = \hat{\Lambda}\hat{\Lambda}' + \hat{\delta}$ then

$$\hat{\Sigma}^{-1}\hat{\Lambda} = (\hat{\Lambda}\hat{\Lambda}' + \hat{\delta})^{-1}\hat{\Lambda} = \hat{\Lambda}^{-1}(\hat{\delta} + \hat{\Lambda}\hat{\Lambda}')$$

$$= \hat{\Lambda}^{-1}\hat{\delta} + \hat{\Lambda}'$$

$$= \hat{\Lambda}^{-1}\hat{\delta} + \hat{\Lambda}'\hat{\delta}^{-1}\hat{\Lambda}\hat{\Lambda}^{-1}\hat{\delta}$$

$$= (I + \hat{\Lambda}' \hat{\delta}^{-1} \hat{\Lambda}) \ \hat{\Lambda}^{-1} \hat{\delta}$$

$$= (I + \hat{\Lambda}' \hat{\delta}^{-1} \hat{\Lambda}) (\hat{\delta}^{-1} \hat{\Lambda})^{-1}$$

$$= \hat{\delta}^{-1} \ \hat{\Lambda} (I + \hat{\Lambda}' \hat{\delta}^{-1} \hat{\Lambda})^{-1}$$

Using Theorem A.7, we can transform A.9 into

$$\text{(A.11)} \quad S \hat{\delta}^{-1} \hat{\Lambda} (I + \hat{\Lambda}' \hat{\delta}^{-1} \hat{\Lambda})^{-1} = \hat{\Lambda} \quad \text{or}$$

$$\text{(A.12)} \quad (S - \hat{\delta}) \delta^{-1} \hat{\Lambda} = \hat{\Lambda} (\hat{\Lambda}' \delta^{-1} \hat{\Lambda}).$$

Theorems B.1, B.2, and B.3 are used to prove theorem B.4, which is stated in section 5.

Theorem B.1

$$\delta^{-1} \Lambda (I_m + \Lambda' \delta^{-1} \Lambda)^{-1} = (\delta + \Lambda \Lambda')^{-1} \Lambda.$$

Proof:

$$\begin{aligned} \Lambda = I_p \Lambda &= (I_p + \Lambda \Lambda' \delta^{-1})^{-1} (I_p + \Lambda \Lambda' \delta^{-1}) \Lambda \\ &= (I_p + \Lambda \Lambda' \delta^{-1})^{-1} (\Lambda + \Lambda \Lambda' \delta^{-1} \Lambda) \\ &= (I_p + \Lambda \Lambda' \delta^{-1})^{-1} \Lambda (I_m + \Lambda' \delta^{-1} \Lambda) \\ &= (\delta \delta^{-1} + \Lambda \Lambda' \delta^{-1})^{-1} \Lambda (I_m + \Lambda' \delta^{-1} \Lambda) \\ &= [(\delta + \Lambda \Lambda')\delta^{-1}]^{-1} \Lambda (I_m + \Lambda' \delta^{-1} \Lambda) \\ &= \delta (\delta + \Lambda \Lambda')^{-1} \Lambda (I_m + \Lambda' \delta^{-1} \Lambda) \end{aligned}$$

Therefore,

$$\delta^{-1} \Lambda (I_m + \Lambda' \delta^{-1} \Lambda)^{-1} = (\delta + \Lambda \Lambda')^{-1} \Lambda.$$

Theorem B.2

$$(S - \hat{\delta}) \hat{\delta}^{-1} \hat{\Lambda} = \hat{\Lambda} (\hat{\Lambda}' \hat{\delta}^{-1} \hat{\Lambda}).$$

Proof:

Premultiplying both sides of the equation in theorem B.1 by S, we get

$$\text{(B.1)} \quad S \delta^{-1} \Lambda (I_m + \Lambda' \delta^{-1} \Lambda)^{-1} = S (\delta + \Lambda \Lambda')^{-1} \Lambda .$$

Since $\Lambda \Lambda' + \delta = \Sigma$, then, (B.1) may be expressed as

(B.2) $\quad S\, \delta^{-1} \Lambda (I_m + \Lambda' \delta^{-1} \Lambda)^{-1} = S\, \Sigma^{-1} \Lambda$

By theorem A.5, $S\, \hat{\Sigma}^{-1} \hat{\Lambda} = \hat{\Lambda}$. So,

$$S\, \hat{\delta}^{-1} \hat{\Lambda} = \hat{\Lambda} (I_m + \hat{\Lambda}' \hat{\delta}^{-1} \hat{\Lambda}) = \hat{\Lambda} + \hat{\Lambda} \hat{\Lambda}' \hat{\delta}^{-1} \hat{\Lambda} \quad \text{or}$$

$$S\, \hat{\delta}^{-1} \hat{\Lambda} - I \hat{\Lambda} = \hat{\Lambda} \hat{\Lambda}' \hat{\delta}^{-1} \hat{\Lambda} .$$

Therefore, $(S - \hat{\delta}) \hat{\delta}^{-1} \hat{\Lambda} = \hat{\Lambda} (\hat{\Lambda}' \hat{\delta}^{-1} \hat{\Lambda})$.

<u>Theorem B.3</u>

The characteristics roots of $(S - \delta)\, \delta^{-1}$ are equal to the characteristic roots of $\Omega^{-1} (S - \delta)\, \Omega^{-1}$ where

$$\Omega^{-1} = \begin{bmatrix} \frac{1}{\delta_1} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{\delta_2} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & \cdots & \frac{1}{\delta_p} \end{bmatrix}$$

<u>Proof</u>:

The characteristic equation of $(S - \delta)\, \delta^{-1}$ is

$$| (S - \delta)\, \delta^{-1} - \lambda I | = | (S - \delta) - \lambda\delta | \, | \delta^{-1} | = 0$$

The characteristic equation of $\Omega^{-1}(S - \delta)\Omega^{-1}$ is

$$|\Omega^{-1}(S - \delta)\Omega^{-1} - \lambda I| = |\Omega^{-1}(S - \delta) - \lambda\Omega|\,|\Omega^{-1}|$$

$$= |\Omega^{-1}|\,|S - \delta - \lambda\Omega\,\Omega|\,|\Omega^{-1}| = |S - \delta - \lambda\delta|\,|\delta^{-1}| = 0$$

$=$ the characteristic equation of $(S - \delta)\delta^{-1}$.

Theorem B.4

If $\hat{\Lambda}'\hat{\delta}^{-1}\hat{\Lambda}$ is a diagonal matrix, then the characteristic roots of $\hat{\Omega}^{-1}(S - \hat{\delta})\Omega^{-1}$ are equal to the successive elements of $\hat{\Lambda}'\hat{\delta}^{-1}\hat{\Lambda}$, and hence the i^{th} column of $\hat{\Omega}^{-1}\hat{\Lambda}$ is merely the characteristic vector corresponding to the i^{th} largest root of $\hat{\Omega}^{-1}(S - \hat{\delta})\hat{\Omega}^{-1}$.

Proof:

Premultiplying both sides of the equation in theorem B.2 by $\hat{\Omega}^{-1}$ we get

$$\hat{\Omega}^{-1}\hat{\Lambda}(\hat{\Lambda}'\hat{\delta}^{-1}\hat{\Lambda}) = \hat{\Omega}^{-1}(S - \hat{\delta})\hat{\delta}^{-1}\hat{\Lambda} \quad \text{or}$$

$$\hat{\Omega}^{-1}\hat{\Lambda}J = [\hat{\Omega}^{-1}(S - \hat{\delta})\hat{\Omega}^{-1}]\,\hat{\Omega}^{-1}\hat{\Lambda} \quad \text{where}$$

$J = \hat{\Lambda}'\hat{\delta}^{-1}\hat{\Lambda}$. The element in the s^{th} row, r^{th} column of J is equal to $\displaystyle\sum_{q=1}^{p} \frac{\hat{\alpha}_{qs}\hat{\alpha}_{qr}}{\hat{\delta}_q^2}$

Let

$$\hat{\Omega}^{-1}\,\Lambda = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \cdot & \cdot & \cdot & \cdot \\ b_{p1} & b_{p2} & \cdots & b_{pm} \end{bmatrix} \qquad J = \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & J_m \end{bmatrix}$$

$\hat{\Omega}^{-1}\,(S - \hat{\delta})\,\hat{\Omega}^{-1} = C$. Therefore,

$$\hat{\Omega}^{-1}\,\hat{\Lambda}\,J = \left[J_1 \begin{bmatrix} b_{11} \\ b_{21} \\ \cdot \\ \cdot \\ \cdot \\ b_{p1} \end{bmatrix} \cdots J_i \begin{bmatrix} b_{1i} \\ b_{2i} \\ \cdot \\ \cdot \\ \cdot \\ b_{pi} \end{bmatrix} \cdots J_m \begin{bmatrix} b_{1m} \\ b_{2m} \\ \cdot \\ \cdot \\ \cdot \\ b_{pm} \end{bmatrix} \right]$$

$$C\,\hat{\Omega}^{-1}\,\hat{\Lambda} = \begin{bmatrix} C \begin{bmatrix} b_{11} \\ b_{21} \\ \cdot \\ \cdot \\ \cdot \\ b_{p1} \end{bmatrix} & \cdots & C \begin{bmatrix} b_{1i} \\ b_{2i} \\ \cdot \\ \cdot \\ \cdot \\ b_{pi} \end{bmatrix} & \cdots & C \begin{bmatrix} b_{1m} \\ b_{2m} \\ \cdot \\ \cdot \\ \cdot \\ b_{pm} \end{bmatrix} \end{bmatrix}$$

Therefore,

$$J_i \begin{bmatrix} b_{1i} \\ b_{2i} \\ \cdot \\ \cdot \\ \cdot \\ b_{pi} \end{bmatrix} = C \begin{bmatrix} b_{1i} \\ b_{2i} \\ \cdot \\ \cdot \\ \cdot \\ b_{pi} \end{bmatrix}$$

or

$$\left| C - J_i I \right| \begin{bmatrix} b_{1i} \\ b_{2i} \\ \cdot \\ \cdot \\ \cdot \\ b_{pi} \end{bmatrix} = 0 = \left| \hat{\Omega}^{-1} (S - \hat{\delta}) \hat{\Omega}^{-1} - J_i I \right| \begin{bmatrix} b_{1i} \\ b_{2i} \\ \cdot \\ \cdot \\ \cdot \\ b_{pi} \end{bmatrix}$$