UNCLASSIFIED

CENTRAL RESEARCH LIBRARY
DOCUMENT COLLECTION

MARTIN MARIETTA ENERGY SYSTEMS LIBRARIES
3 4456 0360983 3

Cy 1

ORNL-1919
Physics

SOME CONTRIBUTIONS TO FACTOR ANALYSIS

W. G. Howe

CENTRAL RESEARCH LIBRARY
DOCUMENT COLLECTION
LIBRARY LOAN COPY
DO NOT TRANSFER TO ANOTHER PERSON
If you wish someone else to see this document, send in name with document and the library will arrange a loan.

OAK RIDGE NATIONAL LABORATORY
OPERATED BY
CARBIDE AND CARBON CHEMICALS COMPANY
A DIVISION OF UNION CARBIDE AND CARBON CORPORATION
UCC
POST OFFICE BOX P
OAK RIDGE, TENNESSEE

UNCLASSIFIED

ORNL 1919

Copy No. 4

Contract No. W-7405-eng-26

MATHEMATICS PANEL

SOME CONTRIBUTIONS TO FACTOR ANALYSIS

William Gerow Howe

DATE ISSUED JUL 12 1955

OAK RIDGE NATIONAL LABORATORY
Operated by
CARBIDE AND CARBON CHEMICALS COMPANY
A Division of Union Carbide and Carbon Corporation
Post Office Box P
Oak Ridge, Tennessee

MARTIN MARIETTA ENERGY SYSTEMS LIBRARIES

3 4456 0360983 3

ORNL 1919
Physics

INTERNAL DISTRIBUTION

1. C. E. Center
2. Biology Library
3. Health Physics Library
4-5. Central Research Library
6. Reactor Experimental Engineering Library
7-11. Laboratory Records Department
12. Laboratory Records, ORNL R.C.
13. C. E. Larson
14. L. B. Emlet (K-25)
15. J. P. Murray (Y-12)
16. A. M. Weinberg
17. E. H. Taylor
18. E. D. Shipley
19-20. A. S. Householder
21. F. C. VonderLage
22. C. P. Keim
23. W. H. Jordan
24. J. A. Swartout
25. F. L. Culler
26. R. S. Livingston
27. S. C. Lind
28. A. H. Snell
29. A. Hollaender
30. M. T. Kelley
31. G. H. Clewett
32. J. H. Frye, Jr.
33. K. Z. Morgan
34. T. A. Lincoln
35. C. S. Harrill
36. D. W. Cardwell
37. C. E. Winters
38. E. M. King
39. Lewis Nelson
40. D. D. Cowen
41. J. A. Lane
42. R. A. Charpie
43. S. J. Cromer
44. M. J. Skinner
45. R. W. Johnson
46. G. J. Atta
47. U. Bents
48. S. G. Campbell
49. R. Cornell
50. T. Hildebrandt
51. C. P. Hubbard
52. A. W. Kimball
53. G. Medlin
54. R. Plunkett
55. R. Sheffield
56. W. G. Howe
57. ORNL Document Reference Library, Y-12 Branch

EXTERNAL DISTRIBUTION

58. R. F. Bacher, California Institute of Technology
59. Division of Research and Medicine, AEC, ORO
60. T. W. Anderson, Columbia University
61. E. E. Cureton, University of Tennessee
62. R. J. Hader, North Carolina State College
63. H. Hotelling, University of North Carolina
64. M. G. Kendall, University of London
65. D. N. Lawley, University of Edinburgh
66. G. E. Nicholson, University of North Carolina
67. S. N. Roy, University of North Carolina
68. W. Smith, University of North Carolina
69. L. L. Thurstone, University of North Carolina

70-378. Given distribution as shown in TID-4500 under Physics category

379-478. Oak Ridge Institute of Nuclear Studies

DISTRIBUTION PAGE TO BE REMOVED IF REPORT IS GIVEN PUBLIC DISTRIBUTION

TABLE OF CONTENTS

CHAPTER		PAGE
I.	INTRODUCTION.	1
	1.1 Statement of the problem.	1
	1.2 Importance of the study	3
	1.3 Review of related literature.	4
	1.4 Organization of the study by chapters	6
II.	THE MODEL	8
	2.1 Derivation of the model	8
	2.2 Advantages of the model	11
	2.3 Comparison with previous models	13
	2.4 The maximum likelihood equations.	15
	2.5 An alternate approach to factor analysis. . .	19
	2.6 Alternate interpretations of the maximum likelihood equations	23
III.	COMPUTATIONAL METHODS IN THE GENERAL CASE.	27
	3.1 Lawley's methods of computation.	27
	3.2 Modification of Lawley's methods	31
	3.3 A Gauss-Seidel iterative method.	33
	3.4 Other methods of computation	40
IV.	ORTHOGONAL FACTORS AND ROTATION.	43
	4.1 General comments on rotation	43
	4.2 The maximum likelihood equations	47

CHAPTER PAGE

IV. (continued)

4.3 Special simple structure hypotheses and the resulting likelihood equations. 49

4.4 Indeterminancy in the model and in the likelihood equations. 54

4.5 Calculation of the estimates. 56

4.6 General remarks 76

V. OBLIQUE FACTORS AND ROTATION. 77

5.1 Oblique factors 77

5.2 The maximum likelihood equations. 82

5.3 Indeterminancy for oblique factors. 85

5.4 Solution of the maximum likelihood equations. 87

VI. PREDICTION. 97

6.1 Prediction of Y from X. 97

6.2 Non-linearity and monotonicity. 101

6.3 Prediction of the factor loadings 102

VII. TESTING 105

7.1 Test of the fit of the model. 105

7.2 Determination of m. 114

7.3 Asymptotic variances and covariances of the maximum likelihood estimates. 115

VIII. SUMMARY AND CONCLUSIONS 119

BIBLIOGRAPHY 121

CHAPTER I

INTRODUCTION

Factor analysis has been a recognized problem among psychologists for the past fifty years. Since the very early 1900's mathematical statisticians have generally avoided the field, only in recent times again becoming interested in the problem. Beginning with Hotelling [15] [1] and continuing through to Anderson and Rubin [1], factor analysis has been attacked in many ways usually yielding quite different results. It is not the purpose of this paper to give the history and details of the various approaches, since this has been very adequately covered in several papers [7, 8, 11]. Nor will relations between factor analysis and newly discovered techniques in multivariate analysis be discussed at any length, since they too have been covered in papers by Bartlett [3, 4], Burt [6], Kendall [19], and in the previously mentioned thesis of Danford [11]. They will be considered only in cases where there is a direct bearing on the model herein proposed.

1.1 Statement of the Problem

The usual method of stating the problem is: can a p-dimensional complex of random variables be represented adequately by $m < p$ variables? Obviously, "adequately" is the ambiguous word which causes the different interpretations. A linear relationship is then assumed

---

[1] Numbers in square brackets refer to bibliography.

and the model is expressed:

$$x_{ij} = \sum_{k=1}^{m} a_{ik} y_{kj} + e_{ij} \qquad \begin{array}{l} i = 1, \ldots, p \\ j = 1, \ldots, N \end{array}$$

where N is the sample size; x_{ij} is the value of the random variable x_i for the jth individual; y_{kj} is the value of the kth common factor for the jth individual; a_{ik} is the factor loading of the ith random variable for the kth common factor; and e_{ij} is a random error such that $E\left[e_{ij}\right] = 0 = E\left[e_{ij}\, e_{nj}\right] = E\left[e_{ij}\, y_{nj}\right]$, $i \neq n$.

From this point on, assumptions differ. For example, some consider the y_{kj} fixed parameters [35] , others consider the y_{kj} random normal variates with variance one and mean zero [20] . However, most authors agree that m , a_{ik} and y_{kj} are to be estimated.

Once the estimates are obtained, the procedures diverge even farther. American psychologists are inclined to "rotate" to simple structure, a concept developed by Thurstone and carefully discussed in his book Multiple Factor Analysis [33] . Simple structure is equivalent to a new model wherein each x_{ij} may be expressed as a linear combination of not m common factors plus an error, but $r < m$ common factors plus an error; however, m common factors still exist in the model as a whole. For example, suppose two common factors had been estimated. Then simple structure in this case would mean that the p variables could be divided into two sets such that:

$$x_{ij} = a_{i1} y_{1j} + e_{ij} \qquad i = 1, \ldots, q \quad ;$$

$$x_{ij} = a_{i2} y_{2j} + e_{ij} \qquad i = q+1, \ldots, p \; .$$

The rotation consists in applying a linear transformation $S_{m \times m}$, where $S S'$ has unities in the diagonal, to the matrix $(a_{ik})_{p \times m}$ to obtain new loadings $(a_{ik})S$. If simple structure does exist, some of these new loadings will be close to zero. Except under certain rather restrictive assumptions no statistical formulation or solution of this particular problem has yet been published even for the case where the y's are assumed orthogonal $\left(E\left[y_{ij}\, y_{nj}\right] = 0 \;,\; n \neq i\right)$. In Chapters IV and V, this problem will be considered more extensively.

A concise statement of the problem is as follows:

1. Derive a satisfactory model both for orthogonal and oblique (correlated) factors with as few assumptions as possible and yet still obtain a solution.

2. Estimate the parameters in the model.

3. Test hypotheses concerning these parameters, particularly m.

4. Estimate the parameters under the a priori assumption that some of the a_{ik} are zero.

5. Test the simple structure hypothesis.

1.2 Importance of the Study

It cannot be denied that rotation and to a lesser extent factor analysis as a whole have been long considered in disrepute by mathematicians

and mathematical statisticians [16] ; there is good cause for this attitude. To say that simple structure is evident because the rotated loadings are close to zero is not enough; sampling considerations must be taken into account. It is the endeavor of this paper to accomplish this and to derive a simple model on which most factor analysts can agree and one which will allow tests of various hypotheses. In the Uppsala Symposium on Psychological Factor Analysis, this is listed as one of the recommended directions for research.

1.3 Review of Related Literature

Lawley's papers [20, 21] have been closely followed throughout the thesis. The model to be proposed, although more general than his, reduces, in the case of orthogonal factors, to the same population covariance or correlation matrix. Estimation and test procedures are also similar, although a new method is proposed whose convergence is a good bit faster for all the examples attempted. In reality, this paper is an extension and generalization of his 1940 and 1941 work at least as regards his Method I. In Lawley's later papers [23, 24, 25] formulae for asymptotic variances and covariances are derived under the assumption that the residual variances $\left(\underset{j}{E}\left[e_{ij}^2\right] = \sigma_i^2 = \text{residual variance}\right)$ are known. The 1953 paper is to be regarded as superseding the 1949 and 1950 papers.

Whittle [35] , Young [37] , and Lawley [21] consider the y_{kj} as fixed parameters. If the residual variances or at least their ratios are not assumed known, no method of solution is available.

A priori knowledge of the residual variances seems a rather drastic restriction. In Chapter II, the advantages and disadvantages of this will be further discussed.

Rippe [29] is concerned only with explaining the covariance matrix, not with estimation of the parameters. Thus, given a solution arrived at by any method, one tests the generated covariance matrix against the sample covariance matrix to see if they are significantly different. For example, the characteristic vector associated with the largest root may explain the covariance matrix while perhaps three vectors obtained by Thurstone's centroid method may not be sufficient. There is also quite a bit of trouble involved in going from the covariance matrix to the correlation matrix since the sample values r_{ii} must be regarded as having a sampling distribution. The r_{ii} are, of course, equal to one, since they are the diagonal elements of the sample correlation matrix. Therefore, this method does not answer the questions involved at all.

In reading through the literature one is amazed at the many approximate methods that have been devised to escape the computation involved in the more rigorous methods. This is particularly applicable to Lawley's Method I and to the computation of principal components. With the advent of high speed electronic computers to which many psychologists have or shortly will have access, this difficulty has been largely overcome. Also, as Emmett [13] points out, if the large amount of labor and expense involved in devising, administering, and

scoring the tests together with that in the computation of the correlation matrix, is considered, the extra time involved in analysis by a much better method is certainly worth while.

In this vein, Rao [28], in a discussion of the problem, proposes a solution equivalent to Lawley's. This has been coded and run on the University of Illinois digital computer and the code is available.

Perhaps the most comprehensive paper on the subject is that of Anderson and Rubin [1]. Unfortunately, it did not come to the author's attention until this paper was close to completion and therefore, there is some overlap between the two. However, the results were obtained independently and each paper contains new material not discussed in the other. Moreover, this thesis is concerned with computational methods, while Anderson and Rubin do not consider this aspect to any extent.

1.4 Organization of the Study by Chapters

In Chapter II the model is proposed and compared with previous models. The maximum likelihood equations are derived and an alternate approach is considered which results in the same equations. Finally, two other conceptions of the likelihood equations are discussed.

Chapter III considers computational methods in the general case. By the general case shall be meant that in which no a priori zeros are assumed. The estimation equations and computational methods for orthogonal factors after rotation are considered in Chapter IV. Chapter V utilizes some of the results in Anderson and Rubin's paper [1];

in the original Chapter V the likelihood equations were obtained for two special cases only. In Chapter VI the prediction of the y's from the x's is considered. This chapter also presents a method for obtaining approximate estimates if one variable is added to the complex, without going through the whole factorization again.

Chapter VII is entitled Testing and contains a discussion of possible tests of the fit of the model and some remarks on asymptotic variances and covariances of the maximum likelihood estimates. Throughout Chapters III-VI numerical examples are given to illustrate the computational techniques.

Chapter VIII contains general conclusions and suggestions for further research.

CHAPTER II

THE MODEL

2.1 Derivation of the Model

The first question that arises in a factor analysis is, does correlation exist? Or, is the population correlation matrix different from the identity matrix? If not, there is no point in proceeding further. The next step, if it were proved there were correlation, would be to ask, is there a random variable y_1 such that the partial correlation coefficients between pairs of the original p variables after eliminating the effect of y_1 are zero? ($\rho_{ij \cdot y_1} = 0$; $i, j = 1, 2, \ldots, p$; $i \neq j$). If the population correlation matrix were still unexplained, then the question would be, are there two random variables, y_1 and y_2, correlated or uncorrelated, such that the partial correlation coefficients between pairs of the original p variables after eliminating y_1 and y_2, are zero? ($\rho_{ij \cdot y_1 y_2} = 0$; $i, j = 1, 2, \ldots, p$; $i \neq j$). One may proceed further, say up to y_m. The y_k $(k = 1, 2, \ldots, m)$ would be the common factors. Thus, the hypothesis of m random variables explaining the correlation matrix of the original p variables is equivalent to

$$\rho_{ij \cdot y_1 y_2 \cdots y_m} = 0 \qquad i, j = 1, 2, \ldots, p \; ; \; i \neq j \; .$$

It is convenient to express the hypothesis in terms of matrices, and this involves the following two vectors: X , the $p \times 1$ vector of the random variables x_i $(i = 1, 2, \ldots, p)$, and Y , the $m \times 1$ vector of random variables y_k . Now, without any loss of generality, it can be assumed that

$$\sigma_{y_k y_k} = 1 \qquad k = 1, 2, \ldots, m$$

$$E(X) = 0$$

$$E(Y) = 0 \quad ,$$

since the y's are unknown in practice. Then the population partial correlation matrix of the x_i after eliminating the y's is defined as

$$U^{-\frac{1}{2}} E \left\{ (X - \beta Y)(X - \beta Y)' \right\} U^{-\frac{1}{2}} \quad .$$

Here U is a $p \times p$ diagonal matrix whose typical element, u_{ii} , is the corresponding diagonal element of $E\left\{(X - \beta Y)(X - \beta Y)'\right\}$. $\beta_{p \times m}$ is defined by

$$\beta E(Y Y') = E(X Y') = \lambda' \quad .$$

The hypothesis may then be stated

$$E\left\{(X - \beta Y)(X - \beta Y)'\right\} = \psi \quad ,$$

where ψ is a diagonal matrix. Then,

$$E(X\,X') = C = \psi + \beta\, E(Y\,Y')\,\beta' \quad ,$$

and C is obviously the population covariance matrix. If $E(Y\,Y') = I$, the identity matrix, then $\beta = \lambda'$, and hence,

$$C = \psi + \lambda'\,\lambda \quad ,$$

which is equivalent to Lawley's model. However, if $E(Y\,Y') = F \neq I$, then $\beta = \lambda'\,F^{-1}$, and $C = \psi + \lambda'\,F^{-1}\,\lambda$.

Now, if $W_{m \times 1} = S_{m \times m} Y$, where S is any non-singular linear transformation,

$$E(X - \beta^{(1)}W)(X - \beta^{(1)}W)' = E(X\,X') - \beta^{(1)}\,E(W\,W')\,\beta^{(1)'} \quad ,$$

where $\beta^{(1)} = \beta\,S^{-1}$. Hence,

$$E(X\,X') - \beta^{(1)}\,E(W\,W')\,\beta^{(1)'} = E(X\,X') - \beta\,F\,\beta' = \psi \quad .$$

This was to be expected, since, if m common factors result in zero partial correlation coefficients, then m factors, formed by a non-

singular linear transformation of the original m factors, will also accomplish this. Hence, in the general case, there is no necessity of employing correlated factors, since orthogonal factors will do just as well, and involve fewer parameters. Yet the problem does exist when the simple structure hypothesis is specified. This problem and other forms of the population covariance matrix that arise under simple structure hypotheses, will be discussed in succeeding chapters.

2.2 Advantages of the Model

It will be advantageous to give a brief discussion of the relationship between the usual model and that proposed in this paper. For a more complete coverage of partial correlation and linear mean square regression, Cramér [10, pp. 302-307] is recommended.

If $\eta_{p \times 1} = X - \beta Y$, where β, X, and Y are defined as before, then

$$E(\eta \; Y') = E(X \; Y') - \beta \, E(Y \; Y') = 0 \; .$$

Now it has been assumed in the hypothesis that $E(\eta \; \eta') = \psi$, where ψ is a diagonal matrix. Therefore, the usual model in factor analysis is obtained,

$$X = \beta Y + \eta \quad ,$$

subject to the restrictions $E(\eta \; Y') = 0$, and $E(\eta \; \eta')$ is a diagonal

matrix. On the other hand, starting with $X = GY + \varphi$, $E(\varphi Y') = 0$, and $E(\varphi \varphi') = \zeta$, a diagonal matrix, one can easily show, in the following manner, that the population partial covariance matrix of X after eliminating Y , is a diagonal matrix:

$$E(X Y') = G E(Y Y') = \beta E(Y Y') \quad .$$

Thus, $\beta = G$, and hence,

$$E\left\{(X - \beta Y)(X - \beta Y)'\right\} = E(\varphi \varphi') = \zeta \quad .$$

The two models are therefore equivalent. However, throughout this derivation no assumption, except finite second order moments, has been made on the joint distribution of X and Y . Now, let it be assumed that X has a multivariate normal distribution. Then the model still holds and there is no need to assume that either Y or η is normally distributed. Herein lies the chief advantage of this formulation; Y and η are not necessarily normal, and $E(Y Y')$ is not necessarily the identity matrix. However, if it is assumed that the y's are independent among themselves and that Y and η are independently distributed, then it follows from a theorem of Cramér [10, pp. 213] , that Y and η are normally distributed. Under the assumption that X alone has a multivariate normal distribution, estimates and tests may be obtained for the various hypotheses, including those of simple

structure. In addition, for the maximum likelihood estimates, the problem of standardization does not arise in this model, since, as will be shown, results are independent of the scale of measurement.

From the model it is evident that either the matrix β or the matrix λ can be estimated, the two being identical if $E(Y\,Y') = I$. Simple structure hypotheses are usually made on β, but the equivalence permits the use of either.

It should be noted that the model still contains the restrictions of linearity and monotonicity; this point will be further considered in Chapter VI.

2.3 Comparison with Previous Models

Lawley, in Method I, assumes that Y and η are distributed normally with zero expectations. This assumption has been criticized quite strongly [35, 36, 37]. Wold states:

> This requirement is a drastic one, since there are cases in practice, the analysis of a truncated population is the most striking example, where the factor values are definitely far from normal.

The proposed model seems to avoid this objection. Moreover, as Lawley states in the same discussion, there is a great distinction between conceiving the factor values as statistical variates and as fixed parameters. Both Lawley (Method II) and Whittle have worked under the assumption of the y's as fixed parameters. As it was noted in Chapter I, a method of solving the resulting maximum likelihood equations has not been found. Kendall [18, 19] and others have pointed out that this

is probably because there are more parameters than observations. However, Whittle [35], by additional a priori assumptions on the residual variances has been able to reach a solution.

The question is similar to that of Model I or Model II in analysis of variance [12]. If, as Thurstone says [33, pp. xii], badly biased samples are used, there would seem to be no recourse but to Lawley's Method II (unacceptable solution) or to Whittle if enough were known about the residual variances. In this case, of course, nothing could be inferred beyond the actual group of individuals involved. But, if it can be assumed that within some particular group, the individuals are selected at random and the distribution of X within this group is multivariate normal, then the proposed model still applies. Also, conclusions may be extended to the entire group, not merely restricted to the observed group.

The factor analyst himself must decide on the method to be used which is determined by his selection of samples and the assumptions he is willing to make. It is not possible to state that one or the other of the two assumptions is the only correct interpretation.

Bartlett [4, 5] has minor objections to Lawley's Method I, in that for $p = 2$, $m = 1$, the method does not work; in addition, that for $p = 3$, no test is available. This is entirely owing to insufficient data; there is just not enough information to test or draw any conclusions. He gives an example [18] of a matrix,

$$\begin{bmatrix} 1 & .60 & -.28 \\ & 1 & .60 \\ & & 1 \end{bmatrix} ,$$

which is incompatible with one general factor and asks, "How do we decide whether, on the basis of one general factor, it arose by chance, when the large sample χ^2 is inoperative?". The question may be answered by another question. If there are two observations from a bivariate normal population, how is it decided if $\rho = 0$? Under the model, the Wishart distribution furnishes the information, not the multivariate normal distribution.

2.4 The Maximum Likelihood Equations

Lawley derived the maximum likelihood equations in his 1940 paper; however, for later use it is advantageous to arrive at them in a different way.

The logarithm of the likelihood function for Wishart's distribution may be written as follows:

$$\text{Log } L = -\frac{N-1}{2}(\log|C| + \text{trace } C^{-1}A) + \text{a function independent of the elements of } C ,$$

where C is the population covariance matrix, A is the sample covariance matrix, and N is the sample size. Then,

$$\frac{N-1}{2} \text{ Trace } C^{-1} A = \frac{N-1}{2} \sum_{i,j=1}^{p} c^{ij} a_{ij}$$

$$= \frac{1}{2} \sum_{i,j=1}^{p} c^{ij} \sum_{n=1}^{N} (x_{in} - \bar{x}_i)(x_{jn} - \bar{x}_j)$$

$$= \frac{1}{2} \sum_{n=1}^{N} x_n C^{-1} x_n' \quad .$$

c^{ij} and a_{ij} are the elements of C^{-1} and A respectively; x_n is the $1 \times p$ row vector with elements $x_{in} - \bar{x}_i$ $(i = 1, 2, \ldots, p)$.

The partial derivative of log L with respect to some element, z , of C is

$$- \frac{N-1}{2|C|} \frac{\partial |C|}{\partial z} + \frac{1}{2} \sum_{n=1}^{N} x_n C^{-1} \frac{\partial C}{\partial z} C^{-1} x_n' \quad .$$

Now

$$x_n C^{-1} \frac{\partial C}{\partial z} C^{-1} x_n' = \text{Trace } \frac{\partial C}{\partial z} C^{-1} x_n' x_n C^{-1} \quad .$$

Therefore,

$$\sum_{n=1}^{N} x_n C^{-1} \frac{\partial C}{\partial z} C^{-1} x_n' = (N-1) \text{ Trace } \frac{\partial C}{\partial z} C^{-1} A C^{-1} \quad .$$

Then, if $z = \psi_{ii}$, the likelihood equations are

$$\frac{\hat{\delta}_{ii}}{|\hat{C}|} - \hat{\xi}_{ii} = 0 \qquad i = 1, 2, \ldots, p \ ;$$

where $\hat{\delta}_{ij}$ is the cofactor of $\hat{c}_{ij}$ in $\hat{C}$, and $\hat{\xi}_{ij}$ is the element in the <u>ith</u> row and <u>jth</u> column of $\hat{C}^{-1} A \hat{C}^{-1}$. Similarly, if $z = \lambda_{ik}$, then

$$\sum_{j=1}^{p} \hat{\lambda}_{jk} \frac{\hat{\delta}_{ij}}{|\hat{C}|} - \sum_{j=1}^{p} \hat{\lambda}_{jk} \hat{\xi}_{ij} = 0 \qquad \begin{array}{l} i = 1, 2, \ldots, p \\ k = 1, 2, \ldots, m \end{array} .$$

If $B = \hat{C}^{-1} A \hat{C}^{-1} - \hat{C}^{-1}$, with typical element b_{ij} , the equations may be written

$$\hat{\lambda} B = 0 \qquad \text{or} \qquad \hat{\lambda} \hat{C}^{-1} A = \hat{\lambda}$$

and

$$b_{ii} = 0 \qquad i = 1, 2, \ldots, p \ .$$

This implies that $\hat{\psi} B$ has zeros in the diagonal and that $\hat{\lambda}' \hat{\lambda} B = 0$. $\hat{C} = \hat{\psi} + \hat{\lambda}' \hat{\lambda}$, and hence $\left[\hat{\lambda}' \hat{\lambda} + \hat{\psi}\right] B = \hat{C} B = A \hat{C}^{-1} - I$ has zeros in the diagonal. Therefore, $A \hat{C}^{-1}$ has ones in the diagonal.

To show that $a_{ii} = \hat{c}_{ii}$ $(i = 1, 2, \ldots, p)$, one need only consider the diagonal elements of

$$\hat{\lambda}' \hat{\lambda} \hat{C}^{-1} A = A - \hat{\psi} \hat{C}^{-1} A = \hat{C} - \hat{\psi} \quad ,$$

since $\hat{C}^{-1} A$ has unities in the diagonal.

Moreover, by defining transformations

$$\hat{P} = V^{-\frac{1}{2}} \hat{C} V^{-\frac{1}{2}} \quad , \qquad R = V^{-\frac{1}{2}} A V^{-\frac{1}{2}} \quad ,$$

and

$$\hat{\ell} = \hat{\lambda} V^{-\frac{1}{2}} \quad ,$$

where V is the diagonal matrix

$$\begin{bmatrix} a_{11} & & & & & \\ & a_{22} & & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & & a_{pp} \end{bmatrix} \quad ,$$

the equations become

$$\hat{\ell}\hat{P}^{-1} R = \hat{\ell} \quad .$$

Here $\hat{P}$ is the estimate of the population correlation matrix and R is the sample correlation matrix with typical element r_{ij} . Therefore, the results are independent of the scale of the x's in that one can go directly from one solution to the other, and standardization ceases to be a problem.

The above results are identical with Lawley's.

2.5 An Alternate Approach to Factor Analysis

It has been shown that the usual factor analysis model and the proposed model are both equivalent to the assumption that the matrix of population partial correlations of the x's after eliminating the y's, is the identity matrix. Therefore, a quite reasonable estimation procedure would be the maximization of the determinant of the matrix of partial correlations in order to make it as close to one as possible. This is almost equivalent to minimizing the sum of squares of the partial correlations, especially when the correlations are small.

If it is assumed that $E(Y\,Y') = I$, the typical non-diagonal element of the matrix of partial correlations is of the form

$$\frac{r_{ij} - \sum_{k=1}^{m} \rho_{ik}\,\rho_{jk}}{\sqrt{1 - \sum_{k=1}^{m} \rho_{ik}^{2}}\,\sqrt{1 - \sum_{k=1}^{m} \rho_{jk}^{2}}} \qquad \begin{array}{l} i,\ j = 1,2,\ldots,p; \\ i \neq j \quad . \end{array}$$

The ρ's are unknown, and are the correlations of the x's with the various unknown common factors. Then the problem is to maximize the determinant of these partials with respect to the ρ's. In matrix notation, the determinant is

$$|R - \lambda'\lambda|\,|\psi|^{-1} \quad ,$$

since the matrix itself is

$$\psi^{-\frac{1}{2}} \left[R - \lambda'\lambda\right] \psi^{-\frac{1}{2}} \quad .$$

R is the sample correlation matrix; λ is the $m \times p$ matrix of unknown correlations; and ψ is the diagonal matrix whose terms are one minus the diagonal terms of $\lambda'\lambda$. If the determinant is now differentiated with respect to some element of λ, say ρ_{ik}, the resulting equation is

$$\frac{\rho_{ik}}{\psi_{ii}}\ |\psi|^{-1}\ |R - \lambda'\lambda| = |\psi|^{-1} \sum_{j=1}^{p} \rho_{jk}\,\delta_{ij} \qquad \begin{array}{l} i = 1,2,\ldots,p \\ k = 1,2,\ldots,m \end{array} \quad ,$$

where the δ_{ij} are the appropriate cofactors of $\left[R - \lambda' \lambda\right]$. If $D = R - \lambda' \lambda$, the equation in matrix form is

$$\lambda D^{-1} = \lambda \psi^{-1} \quad .$$

Then,

$$\lambda = \lambda \psi^{-1} D = \lambda \psi^{-1} R - \lambda \psi^{-1} \lambda' \lambda \quad .$$

The maximum likelihood equation derived in the preceding section is

$$\hat{\ell} \hat{P}^{-1} R = \hat{\ell} \, .$$

Since $\hat{P} = \xi + \hat{\ell}' \hat{\ell}$,

$$\hat{\ell} = \hat{\ell} R^{-1} \hat{\ell}' + \hat{\ell} R^{-1} \hat{\ell}' \hat{\ell} \, .$$

Then, if the above equation is postmultiplied by $\xi^{-1} \hat{\ell}'$,

$$\hat{\ell} \xi^{-1} \hat{\ell}' = \hat{\ell} R^{-1} \hat{\ell}' + \hat{\ell} R^{-1} \hat{\ell}' \hat{\ell} \xi^{-1} \hat{\ell}' \, .$$

Hence,

$$\hat{\ell} R^{-1} \hat{\ell}' = \hat{\ell} \xi^{-1} \hat{\ell}' \left[I + \hat{\ell} \xi^{-1} \hat{\ell}'\right]^{-1} \quad .$$

Now

$$\hat{\ell}_{\xi}^{-1} R = \hat{\ell} + \hat{\ell} R^{-1} \hat{\ell}' \hat{\ell}_{\xi}^{-1} R \quad ,$$

and substituting for $\hat{\ell} R^{-1} \hat{\ell}'$ in the equation, one is led to

$$\hat{\ell}_{\xi}^{-1} R = \left[I + \hat{\ell}_{\xi}^{-1} \hat{\ell}' \right] \hat{\ell}. \tag{1}$$

This is the same equation that was derived by maximizing the determinant of the partial correlations. Thus, the two approaches are equivalent. To the author this is a most interesting result in that, to his knowledge, this approach has not been tried before. It is approximately equivalent to minimizing the sum of squares of the partial correlations and thus, would seem a logical method, since the hypothesis states that these are zero in the population. Quensel [27] has shown that under a similar hypothesis, the distribution of sample partial correlation coefficients (the y's are assumed known) is independent of the distribution of the variables X and Y . He also shows that under certain conditions the determinant of the sample partial correlation matrix has the same distribution as the sample correlation determinant of variables drawn from a p-variate normal population in which the variables are all independent, but with N now reduced by m . This suggests another reason why maximization of the determinant is a logical procedure.

Chapter VII will discuss this method further.

2.6 Alternate Interpretations of the Maximum Likelihood Equations

It must be stated at the outset of this section that its purpose is only to suggest possible relations with other starting points in factor analysis and directions for further investigation. The results are in no way to be interpreted as mathematical derivations; however, they do tie up certain approaches to the problem.

If the N observations are considered as a scatter of N points in a p dimensional space, another line of attack is possible. Hotelling [15] solved the problem by fitting a line through the points by minimizing the sum of squares of the distances of the points from the line. This results in the familiar equation $\lambda A = \nu\lambda$, where ν is a constant. This method consists in weighting all the variables equally. On the other hand, suppose a weighted sum of squares of distances is minimized. The problem then is to find the line in p space,

$$\frac{\varphi_1 - \beta_1}{\theta_1} = \frac{\varphi_2 - \beta_2}{\theta_2} = \dots = \frac{\varphi_p - \beta_p}{\theta_p}$$

where $\sum_{i=1}^{p} \theta_i^2 = 1$, that minimizes the weighted sum of squares of distances from the line. The use of weights is equivalent to imposing a new metric on the space, and from a generalization of Hotelling's result it follows that

(2) $$\theta \zeta^{-1} A = \mu \theta \quad ,$$

where the diagonal matrix

$$\zeta = \begin{bmatrix} \zeta_{11} & & & \\ & \zeta_{22} & & \\ & & \ddots & \\ & & & \zeta_{pp} \end{bmatrix} .$$

ζ is the matrix of weights and

$$\mu = \frac{\theta \zeta^{-1} A \zeta^{-1} \theta'}{\theta \zeta^{-1} \theta'} .$$

If ζ is actually known, then the solutions amount to the characteristic vectors of $\zeta^{-1} A$. Suppose ζ is now assumed unknown, a function of θ and A such that $\zeta_{ii} = a_{ii} - \theta_i^2$. Then one case of the above equation reduces to the maximum likelihood equations of Lawley's Method I, derived previously, while another case reduces to the equations for Method II where the y's are assumed fixed parameters [21, 35]. It is assumed that only one factor $(m = 1)$ is involved in the model, but generalization to $m > 1$ is easily visualized.

Equation (2) has an infinite number of solutions as it stands, since ζ is now a function of θ. To obtain a unique solution μ

must be further specified. If $\mu = 1 + \theta \zeta^{-1} \theta'$, a comparison of equations (1) and (2) for $m = 1$ shows that they are equivalent; while, if $\mu = \theta \zeta^{-1} \theta'$, the equations for Method II arise. Obviously, if it is assumed that all ζ_{ii} are equal, the usual characteristic equation, $\theta A = \mu \theta$, is obtained.

The point should again be made that this derivation is not to be considered as a rigorous proof that the maximum likelihood equations can be obtained by weighting the distances with the residual variances. It is only intended to indicate the type of weighting that does occur. A more realistic line of attack would be weighting with the unknown residual variances given as functions of θ , and then minimizing the sum of squares; however, this leads to equations difficult to handle.

Another possible approach that indicates aspects of the prediction of the y's from the x's, may be derived thusly, again assuming $m = 1$. Predicting y_1 from the x's so as to minimize the linear mean square regression is accomplished by using the following equation:

$$y_1 = \lambda' C^{-1} X \quad ,$$

where λ' is the row vector of population covariances of y_1 and X ; C is the population covariance matrix of X . The regression sum of squares if $\lambda' C^{-1} \lambda$ [10] . However, λ , C , and y_1 are unknown in the factor analysis model; therefore, A , the sample covariance matrix, is substituted for C . The problem is to maximize $\lambda' A^{-1} \lambda$

with respect to the elements λ_i of λ, subject to the condition that $|\psi| = \prod_{i=1}^{p} (a_{ii} - \lambda_i^2)$ equals a constant; that is, the object is to predict y_1 as well as possible subject to the condition on ψ. This yields

$$\hat{\lambda}' A^{-1} \hat{\psi} = \nu \hat{\lambda}'$$

where $\hat{\lambda}$ is the estimate of λ. Since $\hat{\psi}$ is a function of $\hat{\lambda}$, these equations also have an infinite number of solutions. However, specializing as before, if $\nu = 1 - \hat{\lambda}' A^{-1} \hat{\lambda}$, one is led to the Method I equation, and if $\nu = \dfrac{1}{\hat{\lambda}' \hat{\psi}^{-1} \hat{\lambda}}$, to the Method II equation.

General remarks made about the first interpretation also apply here. Only broad relationships are meant to be indicated.

CHAPTER III

COMPUTATIONAL METHODS IN THE GENERAL CASE

In this chapter carets will be omitted from $\hat{C}$, $\hat{\psi}$, and $\hat{\lambda}$ for the sake of simplicity, but it should be remembered that these are only estimates of the population quantities involved. It is also assumed that m is known a priori; otherwise testing problems would have to be considered in this chapter. The actual procedure when m is unknown will be discussed in Chapter VII.

3.1 Lawley's Methods of Computation

It will first be necessary to show the likelihood equation does not give a unique solution for $m > 1$. The likelihood equation is

$$\lambda C^{-1} A = \lambda \quad .$$

If an orthogonal transformation, S , is applied to λ , where $W = S \lambda$, then

$$\lambda' \lambda = W' S' S W = W' W \quad .$$

This implies that ψ is also invariant under this transformation and it follows that C is invariant. Therefore, if the likelihood equation is premultiplied by S , it is evident that W is also a solution of the equation.

Referring to Section 2.1, one sees that this is equivalent to applying an orthogonal transformation, S , to Y . If $E(Y Y') = I$, then $E(S Y Y' S') = I$, and the new factors are uncorrelated.

Moreover, if it is now supposed only that S is a non-singular linear transformation such that $S S'$ has unities in the diagonal, $\varphi = S\lambda$ will also be a solution of the equation. For, in this case,

$$\lambda' \lambda = \varphi' \left[S S' \right]^{-1} \varphi \quad .$$

Then,

$$C = \psi + \varphi' \left[S S' \right]^{-1} \varphi \quad ,$$

and this is the form of C if it is assumed that the y's are correlated and such that $E(Y Y') = S S'$. However, the estimation will be restricted to the case of orthogonal factors for reasons previously given, principally the fact that in the general case, the assumption of oblique factors leads back to orthogonal factors.

Hence, for $m > 1$, there are an infinite number of solutions. A particular solution may be selected by further restrictions on the likelihood equation, and this is what Lawley does.

In his 1940 paper Lawley proposes a method of calculation that involves the computation of the inverse of the sample correlation or covariance matrix. However, to avoid the calculation of A^{-1} , he has

proposed a new method in his 1941 paper which supersedes the 1940 method. The derivation proceeds in the following manner: it has been shown in Section 2.5 that the likelihood equation can be written

$$\lambda \psi^{-1} A = \left[I + \lambda \psi^{-1} \lambda' \right] \lambda \quad .$$

The derivation of the equation was actually for the correlation estimates, but is obviously exactly the same for the covariance case.

The particular solution selected by Lawley is

$$\lambda \psi^{-1} A = Z \lambda + \lambda \quad ,$$

where Z is the lower triangular matrix containing only the corresponding sub-diagonal terms of $\lambda \psi^{-1} \lambda'$. Then,

$$\lambda \psi^{-1} A \psi^{-1} \lambda' - \lambda \psi^{-1} \lambda' = Z \lambda \psi^{-1} \lambda' \; .$$

Since the matrix on the left side of the equation is symmetric, the matrix on the right must also be symmetric. If the element in the first row and second column and the element in the second row and first column are considered, then because of the symmetry,

$$\lambda_1 \psi^{-1} \lambda_1' \lambda_1 \psi^{-1} \lambda_2' = \lambda_1 \psi^{-1} \lambda_1' \lambda_1 \psi^{-1} \lambda_2' + \lambda_2 \psi^{-1} \lambda_2' \lambda_1 \psi^{-1} \lambda_2' \; .$$

Hence, $\lambda_1 \psi^{-1} \lambda_2'$ must be zero, and by next considering the element in the first row and third column and the element in the third row and first column, one can show that $\lambda_1 \psi^{-1} \lambda_3'$ is zero, since $\lambda_1 \psi^{-1} \lambda_2'$ is zero. Proceeding in this manner, one can show that $\lambda \psi^{-1} \lambda'$ is a diagonal matrix. Therefore, a solution of

$$\lambda \psi^{-1} A = Z \lambda + \lambda$$

is also a solution of the likelihood equation.

The method of computation, say for $m = 2$, consists in starting with an initial approximation $\lambda^{(1)}$, obtained by some such method as centroid or principal component analysis. Then $\psi^{(1)}$ is computed by using the relation

$$\psi_{ii}^{(1)} = a_{ii} - \lambda_{i1}^{(1)^2} - \lambda_{i2}^{(1)^2} \qquad i = 1,2,\ldots,p .$$

The next approximation, $\lambda^{(2)}$, is given by

$$\lambda_1^{(2)} = \frac{\lambda_1^{(1)} \psi^{(1)^{-1}} A - \lambda_1^{(1)}}{\sqrt{\lambda_1^{(1)} \psi^{(1)^{-1}} A \psi^{(1)^{-1}} \lambda_1^{(1)'} - \lambda_1^{(1)} \psi^{(1)^{-1}} \lambda_1^{(1)'}}}$$

$$\lambda_2^{(2)} = \frac{\lambda_2^{(1)}\,\psi^{(1)^{-1}}\,A - \lambda_2^{(1)}\,\psi^{(1)^{-1}}\,\lambda_1^{(2)'}\,\lambda_1^{(2)} - \lambda_2^{(1)}}{\sqrt{\lambda_2^{(1)}\,\psi^{(1)^{-1}}\,A\,\psi^{(1)^{-1}}\,\lambda_2^{(1)'} - \left[\lambda_2^{(1)}\,\psi^{(1)^{-1}}\,\lambda_1^{(2)}\right]^2 - \lambda_2^{(1)}\psi^{(1)^{-1}}\lambda_2^{(1}}}$$

A new ψ , $\psi^{(2)}$, is then calculated from $\lambda^{(2)}$ and the computations proceed in the same manner to obtain $\lambda^{(3)}$, continuing until $\lambda^{(n)}$ converges. Examples of the technique are contained in Lawley's 1941 $(m = 1)$ and 1943 $(m = 2)$ papers, and in a paper of Emmett's [13 , $m = 3$] .

Both of these methods appear to converge in every case but that in which one of the ψ_{ii} in the solution is zero. In this situation the iterations literally bound all over, and do not converge to the correct solution. Another drawback is the sometimes exceedingly slow rate of convergence; it is possible to stop at a point in the iteration where there is no change in the value at the decimal place to which accuracy is desired, and yet the correct estimate is a long way off. This point, with an example, will be discussed in more detail in Section 3.3.

3.2 Modifications of Lawley's Methods

An obvious modification of Lawley's methods is as follows: Starting with an initial approximation $\lambda^{(1)}$, one computes $\psi^{(1)}$ in the same manner as before. Then a second approximation to λ is given by

$$\lambda^{(2)} = \left[\lambda^{(1)} \psi^{(1)^{-1}} \lambda^{(1)'}\right]^{-1} \left[\lambda^{(1)} \psi^{(1)^{-1}} A - \lambda^{(1)}\right] .$$

$\psi^{(2)}$ is then calculated from $\lambda^{(2)}$, and

$$\lambda^{(3)} = \left[\lambda^{(2)} \psi^{(2)^{-1}} \lambda^{(2)'}\right]^{-1} \left[\lambda^{(2)} \psi^{(2)^{-1}} A - \lambda^{(2)}\right] .$$

The method is continued until $\lambda^{(n)}$ converges. For $m > 1$, this will not give the same result as Lawley's method, since it has not been assumed that $\lambda \psi^{-1} \lambda'$ is a diagonal matrix. An iterative method which gives this particular solution may be specified as follows:

$$\lambda^{(n+1)} = Z^{(n)^{-1}} \left[\lambda^{(n)} \psi^{(n)^{-1}} A - \lambda^{(n)}\right] ,$$

where Z is the lower triangular matrix defined in the previous section. However, this method (utilizing Z) does not have any advantages over the other method. A solution arrived at by the first method can be easily transformed to a solution satisfying the condition that $\lambda \psi^{-1} \lambda'$ is a diagonal matrix. In addition, the convergence may be slower since the solution has been restricted.

An alternative method may be given by the following equation:

$$\lambda^{(n+1)} = \left[I + \lambda^{(n)} \psi^{(n)^{-1}} \lambda^{(n)'}\right]^{-1} \left[\lambda^{(n)} \psi^{(n)^{-1}} A\right] .$$

This method is not recommended, since the convergence is slower than in the preceding methods.

The computational schemes proposed in this section have certain advantages over Lawley's methods, notably, less computation. It is also believed that convergence is faster, although no proof of this is advanced. Therefore, the methods of this section are recommended over those of Lawley.

3.3 A Gauss-Seidel Iterative Method

The Gauss-Seidel iterative method is usually applied to obtain the solution of simultaneous linear equations [17, 34]. For example, if there are p linear equations in p unknowns, $x_1, x_2, \ldots, x_p$, say, one solves the first equation for x_1 in terms of the other $p - 1$ unknowns. Then the second equation is solved for x_2 in terms of $x_1, x_3, \ldots, x_p$, the third for x_3 in terms of $x_1, x_2, \ldots, x_p$, and so forth. Starting with an initial approximation: $x_{20}, x_{30}, \ldots, x_{p0}$, say, one then obtains x_{10} by substituting the initial approximation for $x_2, x_3, \ldots, x_p$ in the first equation. One then obtains x_{21} by substituting $x_{10}, x_{30}, \ldots, x_{p0}$ for $x_1, x_3, \ldots, x_p$ in the second equation. This process is continued until x_{p1} is obtained. Then x_{11} is calculated by again substituting $x_{21}, x_{31}, \ldots, x_{p1}$ for $x_2, x_3, \ldots, x_p$ in the first equation. One continues the procedure until $x_{1n}, x_{2n}, \ldots, x_{pn}$ converge.

The method may be immediately extended to p non-linear equations in p unknowns. However, for Lawley's maximum likelihood equations,

$x_1, x_2, \ldots, x_p$ are now all vectors of the following form:

$$x_i = \begin{bmatrix} \lambda_{1i} \\ \lambda_{2i} \\ \vdots \\ \lambda_{mi} \end{bmatrix} \qquad i = 1, 2, \ldots, p \ .$$

Each x_i may be expressed as a function of the remaining vectors, $x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p$, and the computations then proceed exactly as outlined in the previous paragraph.

For the case $m = 2$ these equations may be derived as follows: the likelihood equations are

$$\lambda_1 \psi^{-1} A = \lambda_1 + \lambda_1 \psi^{-1} \lambda_1' \lambda_1 + \lambda_1 \psi^{-1} \lambda_2' \lambda_2$$

$$\lambda_2 \psi^{-1} A = \lambda_2 + \lambda_2 \psi^{-1} \lambda_1' \lambda_1 + \lambda_2 \psi^{-1} \lambda_2' \lambda_2 \quad ,$$

or in scalar notation,

$$\sum_{i=1}^{p} \frac{\lambda_{1i}}{\psi_{ii}} (a_{ij} - \lambda_{1i} \lambda_{1j} - \lambda_{2i} \lambda_{2j}) = \lambda_{1j}$$

$$j=1,2,\ldots,p \ .$$

$$\sum_{i=1}^{p} \frac{\lambda_{2i}}{\psi_{ii}} (a_{ij} - \lambda_{1i} \lambda_{1j} - \lambda_{2i} \lambda_{2j}) = \lambda_{2j}$$

Now $\psi_{jj} = a_{jj} - \lambda_{1j}^2 - \lambda_{2j}^2$, so that the equations may also be written

$$\sum_{\substack{i=1 \\ i\neq j}}^{p} \frac{\lambda_{1i}}{\psi_{ii}} (a_{ij} - \lambda_{1i}\,\lambda_{1j} - \lambda_{2i}\,\lambda_{2j}) = 0$$

$$j = 1,2,\ldots, p \; .$$

$$\sum_{\substack{i=1 \\ i\neq j}}^{p} \frac{\lambda_{2i}}{\psi_{ii}} (a_{ij} - \lambda_{1i}\,\lambda_{1j} - \lambda_{2i}\,\lambda_{2j}) = 0$$

Hence, for $j = 1, 2, \ldots, p$,

$$\lambda_{1j} \sum_{i\neq j} \frac{\lambda_{1i}^2}{\psi_{ii}} + \lambda_{2j} \sum_{i\neq j} \frac{\lambda_{1i}\,\lambda_{2i}}{\psi_{ii}} = \sum_{i\neq j} \frac{\lambda_{1i}\,a_{ij}}{\psi_{ii}}$$

$$\lambda_{1j} \sum_{i\neq j} \frac{\lambda_{1i}\,\lambda_{2i}}{\psi_{ii}} + \lambda_{2j} \sum_{i\neq j} \frac{\lambda_{2i}^2}{\psi_{ii}} = \sum_{i\neq j} \frac{\lambda_{2i}\,a_{ij}}{\psi_{ii}} \quad .$$

Generalization to $m > 2$ is obvious. For $m = 1$, the equations become

$$\lambda_{1j} = \frac{\sum_{i\neq j} \frac{\lambda_{1i}\,a_{ij}}{\psi_{ii}}}{\sum_{i\neq j} \frac{\lambda_{1i}^2}{\psi_{ii}}} \qquad j = 1, 2, \ldots, p \; .$$

Computation would start with an initial approximation; then $\lambda_{11}^{(2)}$ and $\lambda_{21}^{(2)}$ would be computed from the $\lambda_{kj}^{(1)}$ $(k = 1, 2;\ j = 2, 3, \ldots, p)$. $\lambda_{12}^{(2)}$ and $\lambda_{22}^{(2)}$ would then be computed from $\lambda_{11}^{(2)}$, $\lambda_{21}^{(2)}$, and $\lambda_{kj}^{(1)}$ $(k = 1, 2;\ j = 3, 4, \ldots, p)$, and so forth.

Although an entire iteration involves more work, the increase in speed of convergence is much more than enough to offset the extra labor. Particularly for those cases in which convergence is extremely slow, the saving is very large. As an example, the correlation matrix in Lawley's 1943 paper will be considered with m assumed equal to two. It is reproduced below, omitting the lower non-diagonal elements.

$$\begin{bmatrix} 1.000 & .312 & .405 & .457 & .500 & .350 & .521 & .564 \\ & 1.000 & .460 & .316 & .279 & .173 & .339 & .288 \\ & & 1.000 & .394 & .380 & .258 & .433 & .323 \\ & & & 1.000 & .460 & .222 & .516 & .486 \\ & & & & 1.000 & .239 & .441 & .417 \\ & & & & & 1.000 & .302 & .262 \\ & & & & & & 1.000 & .547 \\ & & & & & & & 1.000 \end{bmatrix} .$$

As a first approximation one takes

$$\lambda_1^{(1)} = \begin{bmatrix} .73 & .50 & .66 & .66 & .62 & .40 & .73 & .70 \end{bmatrix}$$

$$\lambda_2^{(1)} = \begin{bmatrix} .17 & -.27 & -.47 & .08 & .06 & .02 & .10 & .29 \end{bmatrix} .$$

Then,

$$\psi^{(1)} = \begin{bmatrix} .4382 & & & & & & & \\ & .6771 & & & & & & \\ & & .3435 & & & & & \\ & & & .5580 & & & & \\ & & & & .6120 & & & \\ & & & & & .8396 & & \\ & & & & & & .4571 & \\ & & & & & & & .4259 \end{bmatrix} .$$

The calculation then proceeds as follows:

$$\sum_{i \neq 1} \frac{\lambda_{1i}^2}{\psi_{ii}} = \frac{(.50)^2}{.6771} + \frac{(.66)^2}{.3435} + \ldots + \frac{(.70)^2}{.4259} = 5.553 \quad ,$$

$$\sum_{i \neq 1} \frac{\lambda_{1i}\lambda_{2i}}{\psi_{ii}} = \frac{(.50x-.27)}{.6771} + \frac{(.66x-.47)}{.3435} + \ldots + \frac{(.70 \text{ x } .29)}{.4259} = -.301 \ ,$$

$$\sum_{i \neq 1} \frac{\lambda_{2i}^2}{\psi_{ii}} = \frac{(-.27)^2}{.6771} + \frac{(-.47)^2}{.3435} + \ldots + \frac{(.29)^2}{.4259} = .988 \quad ,$$

$$\sum_{i \neq 1} \frac{\lambda_{1i}a_{i1}}{\psi_{ii}} = \frac{(.50 \text{ x } .312)}{.6771} + \frac{(.66 \text{ x } .405)}{.3435} + \ldots + \frac{(.70 \text{ x } 564)}{.4259} = 3.981 \ ,$$

$$\sum_{i \neq 1} \frac{\lambda_{2i}\, a_{i1}}{\psi_{ii}} = \frac{(-.27 \times .312)}{.6771} + \frac{(-.47 \times .405)}{.3435} + \ldots + \frac{(.29 \times .564)}{.4259} = -.058 \ .$$

Finally,

$$\begin{bmatrix} 5.553 & -.301 \\ -.301 & .988 \end{bmatrix} \begin{bmatrix} \lambda_{11}^{(2)} \\ \lambda_{21}^{(2)} \end{bmatrix} = \begin{bmatrix} 3.981 \\ -.058 \end{bmatrix} ,$$

and

$$\lambda_{11}^{(2)} = .726 \ , \qquad \lambda_{21}^{(2)} = .162 \ .$$

The same procedure is followed, using $\lambda_{11}^{(2)}$ and $\lambda_{21}^{(2)}$, to obtain $\lambda_{12}^{(2)}$ and $\lambda_{22}^{(2)}$. Continuing for a good number of iterations, one is led to the following results:

$$\lambda_1 = \begin{bmatrix} .722 & .499 & .691 & .658 & .621 & .398 & .723 & .689 \end{bmatrix}$$

$$\lambda_2 = \begin{bmatrix} .183 & -.216 & -.527 & .110 & .090 & .037 & .126 & .291 \end{bmatrix} .$$

In Lawley's method it will be remembered that $\lambda_1 \psi^{-1} \lambda_2' = 0$; therefore, an orthogonal transformation must be applied to λ so that $L_1 \psi^{-1} L_2' = 0$. The resulting vectors, L_1 and L_2 , given below, will still satisfy the equations as shown previously.

$$L_1 = \begin{bmatrix} .706 & .515 & .731 & .648 & .612 & .394 & .711 & .663 \end{bmatrix}$$

$$L_2 = \begin{bmatrix} .240 & -.176 & -.471 & .161 & .139 & .069 & .183 & .344 \end{bmatrix} .$$

Lawley states in his paper, "In general very great accuracy is not essential, and the final estimates of the factor loadings and specific (residual) variances are not necessarily correct as regards the third decimal place. They are, however, sufficiently accurate for our purposes." His results are

$$\begin{bmatrix} .725 & .503 & .664 & .661 & .623 & .399 & .726 & .694 \end{bmatrix}$$

$$\begin{bmatrix} .172 & -.261 & -.468 & .087 & .069 & .027 & .106 & .291 \end{bmatrix} .$$

It is evident that there is a large discrepancy in the results. L_1 and L_2 are accurate to three decimals and should agree with the fectors obtained by Lawley. The difference is probably caused by the extremely slow convergence of Lawley's method in this case, and by the failure to carry enough places to observe the convergence. First differences of the order of .0003 may be obtained from iterate to iterate, and yet the second differences may be one hundredth of this size. This might, if the computer were working to three place accuracy, cause him to stop at an iteration where, in reality, he was much farther than .001 from the correct result. It is believed that this was the case in the example discussed above.

On the other hand, it is perfectly possible that for testing purposes, the iteration has proceeded far enough and more iterations do not change the value of the statistic to any large extent. This also seems to be the case here. However, in prediction or rotation a more accurate estimate should be obtained; this implies a necessity for carrying more places than the number to which accuracy is desired.

Another advantage of the Gauss-Seidel scheme and the methods of Section 3.2 is that a particular solution is not specified by further restrictions on the likelihood equations. Thus, there is no necessity for attempting to find an initial approximation in which the $\lambda_k^{(1)} \psi^{(1)^{-1}} \lambda_j^{(1)'}$ $(k \neq j)$ are close to zero so as to speed the convergence. It must be observed, however, that when m becomes large, say greater than or equal to four, the amount of work in inverting the matrix becomes quite large and probably impractical unless electronic digital computers are available. Yet in this case, other methods would also be impractical for desk computers.

3.4 Other Methods of Computation

Rao [28] has also proposed a method which gives a solution to the likelihood equations. As mentioned in the Chapter I, this has been coded for the Illiac and the code is available. The method seems impractical for desk computers, Rao himself making this point. To obtain the solution, for $m = 2$ say, one solves the determinantal equation

$$|GRG - \nu I| = 0 \quad ,$$

where the elements g_i of the $p \times p$ diagonal matrix, G, are such that

$$g_i = \sqrt{(\nu_1 - 1)b_i^2 + (\nu_2 - 1)c_i^2 + 1} \qquad i = 1,2,\ldots, p;$$

ν_1 and ν_2 are the two largest characteristic roots of the equation and b_i and c_i are the ith elements of the associated characteristic vectors, b and c; R is the sample correlation matrix. Beginning with an initial approximation, $G^{(1)}$, for G, one obtains a new G, $G^{(2)}$, by solving the determinantal equation

$$\left| G^{(1)} R G^{(1)} - \nu I \right| = 0 .$$

The process is repeated until convergence is obtained. Estimated factor loadings corresponding to λ_1 and λ_2 of the previous case, are defined as

$$\sqrt{\nu_1 - 1}\; b\, G^{-1} \quad \text{and} \quad \sqrt{\nu_2 - 1}\; c\, G^{-1} ,$$

and these estimates satisfy the likelihood equations. For a derivation of these equations and an interesting discussion of the principal component method and what Rao calls canonical factor analysis, equivalent to Lawley's Method I but derived differently, his 1954 paper is recommended.

As far as general conclusions are concerned, for m less than or equal to three, the Gauss-Seidel scheme is definitely recommended over the other methods discussed in this chapter. It is superior because first, it can be conveniently carried out on a desk computer and second, its convergence is much faster than other available methods, at least in all numerical examples tried by the author.

CHAPTER IV

ORTHOGONAL FACTORS AND ROTATION

Throughout this paper orthogonal will be considered interchangeable with uncorrelated, and oblique interchangeable with correlated. This convention has been and will be followed in order to maintain the usual factor analysis terminology. In this chapter it is assumed that the y_k $(k = 1, 2, \ldots, m)$, are all uncorrelated; $E(Y Y') = I$. The following chapter will take up the problem of correlated factors.

4.1 General Comments on Rotation

It has been shown in Chapter II that, under the model with orthogonal factors, the matrix of regression coefficients, β, is the same as the matrix of covariances, λ'. Thus, hypotheses on the form of β are equivalent to hypotheses on the form of λ'. Simple structure specifies the form of the population regression matrix, β. Therefore, in this and the following chapters, hypotheses and equations will be formulated in terms of β, rather than in terms of λ'.

To illustrate the subject of rotation, it will be advantageous to give an example of the procedure of the psychologists. Suppose the maximum likelihood estimate for a 6×6 correlation matrix of test scores for $m = 2$, is

$$\hat{\lambda} = \hat{\beta}' = \begin{bmatrix} .4 & .3 & .5 & .7 & .4 & .8 \\ -.6 & -.5 & -.8 & .4 & .3 & .5 \end{bmatrix} .$$

It has been shown in the preceding chapter that if $\hat{\beta}'_{m \times p}$ is a solution of the likelihood equations, then $S\hat{\beta}'$, where S is an $m \times m$ orthogonal matrix, is also a solution. The psychologists look for a matrix S which results in $S\hat{\beta}'$ having small quantities, close to zero, in specified locations which give an indication of simple structure. In this case, by applying graphical or analytical methods [33] they find an orthogonal matrix,

$$S = \begin{bmatrix} .846 & .533 \\ -.533 & .846 \end{bmatrix} .$$

Then,

$$S\hat{\beta}' = \begin{bmatrix} .02 & -.01 & .00 & .81 & .50 & .94 \\ -.72 & -.58 & -.94 & -.03 & .04 & .00 \end{bmatrix} .$$

On observing $S\hat{\beta}'$, they conclude that tests 1, 2, and 3 contain one common factor, and tests 4, 5, and 6, another. $S\hat{\beta}'$ satisfies the likelihood equations and who is to say that the sample estimates close to zero, are not actually zero in the population? It is then assumed simple structure exists and the particular experiment is over. Of course, what is needed is a test of the hypothesis that these parameters are zero in the population.

For this example the hypothesis can be stated:

$$E\left\{(X - \beta Y)(X - \beta Y)'\right\} = \psi$$

$$\beta = \begin{array}{c} \\ 3 \\ 3 \end{array} \begin{array}{c} \begin{array}{cc} 1 & 1 \end{array} \\ \left[\begin{array}{c|c} \beta_{11} & 0 \\ \hline 0 & \beta_{22} \end{array}\right] \end{array} .$$

Here ψ is a diagonal matrix; β_{11} and β_{22} are 3×1 column vectors. The resulting population covariance matrix is, in partitioned form,

$$C = \psi + \beta\beta' = \left[\begin{array}{c|c} \psi_{(1)} + \beta_{11}\beta_{11}' & 0 \\ \hline 0 & \psi_{(2)} + \beta_{22}\beta_{22}' \end{array}\right] .$$

Estimates would then have to be made of these parameters and a test developed.

At this point an important fact arises; namely, since this hypothesis was made from the data, the same data cannot be used to test it. A new sample should be used for this purpose. Thus, if simple structure is the object of the analysis, the initial data should be randomly divided into two sets, one to generate the hypothesis, the other to test it. Because of the large sample sizes usually involved in psychological studies, this should involve no handicap, except for

the extra computation necessitated by two correlation matrices. On the other hand, if the hypothesis, which involves specifying the form of the population covariance matrix, can be made a priori, this problem does not arise.

The determination of the S transformation is a problem in itself. Analytical solutions, calling for no judgment, have been proposed, but the general consensus seems to be that at the present time, graphical or other methods calling for human judgment, are better [9]. Presently then, it is the factor analyst's problem to make the simple structure hypothesis by using developed rotational techniques, and the statistician may then devise tests for the hypotheses and estimation procedures for the parameters involved. Therefore, in this development it is assumed that the hypothesis has been made previously from another group of data, and it is now desired to test it on a new sample.

The purpose then, of this chapter, is to translate simple structure to statistical formulation and to derive estimation procedures for the parameters in the resulting models. Section 4.2 will derive the maximum likelihood equations under these models; Section 4.3 will consider the likelihood equations for three special cases discussed by Thurstone [33]. The indeterminancy in the model is discussed in Section 4.4, while methods of solving the equations will be covered in Section 4.5. Finally, Section 4.6 will summarize and generalize some of the results. Testing of the hypotheses is then only an extension of Lawley's work and is discussed in Chapter VII.

4.2 The Maximum Likelihood Equations

It is now assumed that a sample of N has been drawn from a p-variate normal population with mean μ and covariance matrix $C = \psi + \beta\beta'$, where β is a $p \times m$ matrix, $m < p$. Certain elements of β, say $\beta_{i_1 j_1}, \beta_{i_2 j_2}, \beta_{i_3 j_3}, \ldots, \beta_{i_r j_r}$, are assumed zero, where $r > \frac{m(m-1)}{2}$; $i_1, i_2, \ldots, i_r$ can assume any value from 1 to p; and $j_1, j_2, \ldots, j_r$ any value from 1 to m. To obtain estimates of the elements of the diagonal matrix ψ and of the remaining elements of β, one may use the method of maximum likelihood to maximize the likelihood function for Wishart's distribution. In this case, the resulting likelihood equations are

$$\hat{\beta}' (\hat{C}^{-1} A \hat{C}^{-1} - \hat{C}^{-1}) = U'$$

(1)

$$\text{Diagonal } (\hat{C}^{-1} A \hat{C}^{-1} - \hat{C}^{-1}) = 0 ,$$

where U is a $p \times m$ matrix with zeros where β is not specified to have zeros. If $B = \hat{C}^{-1} A \hat{C}^{-1} - \hat{C}^{-1}$ with typical element b_{ij}, the equations may be written

$$\hat{\beta}' B = U'$$

(2)

$$b_{ii} = 0 \qquad i = 1, 2, \ldots, p \ .$$

Equation (2) implies that $\hat{\psi} B$ and $\hat{\beta}\hat{\beta}' B$ have zeros in the diagonal, since $\hat{\beta} U'$ has zeros in the diagonal by definition of U. Therefore $(\hat{\psi} + \hat{\beta}\hat{\beta}') B = \hat{C} B = A \hat{C}^{-1} - I$ has zeros in the diagonal and hence $A \hat{C}^{-1}$ has ones in the diagonal.

On the other hand, it is no longer true that $a_{ii} = \hat{c}_{ii}$ $(i = 1, 2, \ldots, p)$. If equation (1) is postmultiplied by $\hat{C}$ and premultiplied by $\hat{\beta}$, then

$$\hat{\beta}\hat{\beta}' \hat{C}^{-1} A - \hat{\beta}\hat{\beta}' = A - \hat{\psi}\hat{C}^{-1} A - \hat{C} + \hat{\psi} = \hat{\beta} U' \hat{\psi} + \hat{\beta} U' \hat{\beta}\hat{\beta}' \ .$$

From a consideration of the diagonal elements in the equation above, it is easily seen that the diagonal elements of $\hat{C}$ are not equal to the diagonal elements of A, since the diagonal elements of $\hat{\beta} U' \hat{\beta}\hat{\beta}'$ are not necessarily zero.

Moreover, in the same way as in Chapter II, it can be shown that similar equations hold for the correlation matrix; thus, the problem of standardization is still avoided.

Equation (1) is not, however, in a form suitable for computation. The following paragraph will derive equations which are in a more convenient form.

If equation (1) is postmultiplied by $\hat{C}$, then

$$\hat{\beta}' \hat{C}^{-1} A = U' \hat{\psi} + U' \hat{\beta} \hat{\beta}' + \hat{\beta}' = \hat{\beta}^{*\prime} ,$$

say. Now proceeding exactly as in Section 2.5, one can easily show equation (1) is equivalent to

$$\hat{\beta}' \hat{\psi}^{-1} A = (I + \hat{\beta}' \hat{\psi}^{-1} \hat{\beta}) \hat{\beta}^{*\prime}$$

(3)

$$\text{Diagonal } \hat{\psi} = \text{Diagonal}(A - \hat{\beta} \hat{\beta}^{*\prime}) .$$

The last equation of (3) is equivalent to $a_{ii} = \hat{c}_{ii}$ $(i = 1, 2, \ldots, p)$ only if the diagonal elements of $\hat{\beta} U' \hat{\beta} \hat{\beta}'$ are zero. In the next section some cases are considered where this last relationship is true.

4.3 Special Simple Structure Hypotheses and the Resulting Likelihood Equations

The first hypothesis of interest is that discussed in Section 4.1; in Thurstone's terminology, it is called "isolated constellation configuration", [33, pp. 184], and will here be called Model I. For $m = 3$, the hypothesis is

$$E\left\{(X - \beta Y)(X - \beta Y)'\right\} = \psi$$

$$\beta = \begin{array}{c} q \\ r \\ p-q-r \end{array} \left[\begin{array}{ccc} \beta_{11} & 0 & 0 \\ \hline 0 & \beta_{22} & 0 \\ \hline 0 & 0 & \beta_{33} \end{array}\right] .$$

In this case the population covariance matrix, $C = \psi + \beta\beta'$, will have three blocks down the diagonal of the same form as for $m = 2$, and zeros elsehwere. Generalization to m greater than three is obvious.

Model II will designate what has been called "incomplete triangular configuration". For $m = 3$, the hypothesis takes the following form:

$$E\left\{(X - \beta Y)(X - \beta Y)'\right\} = \psi$$

$$\beta = \begin{array}{c} q \\ \\ p-q \end{array} \left[\begin{array}{ccc} \beta_{11} & \beta_{12} & 0 \\ \hline \beta_{21} & 0 & \beta_{23} \end{array}\right] .$$

This results in a population covariance matrix, $C = \psi + \beta\beta'$, which in partitioned form may be written

$$C = \begin{array}{c} \\ q \\ p-q \end{array} \begin{array}{c} \begin{array}{cc} q & p-q \end{array} \\ \left[\begin{array}{c|c} \psi_{(1)} + \beta_{11}\beta_{11}' + \beta_{12}\beta_{12}' & \beta_{11}\beta_{21}' \\ \hline \beta_{21}\beta_{11}' & \psi_{(2)} + \beta_{21}\beta_{21}' + \beta_{23}\beta_{23}' \end{array}\right] \end{array}$$

Model II is equivalent to the Spearman model (with uncorrelated factors) where a general intelligence factor is assumed for all tests, while the tests are assumed to have zero or positive loadings on other factors called "specifics". The general factor here corresponds to that common factor on which all the tests are assumed to have non-zero regression coefficients; the other two are the specifics. This model is also easily generalized to m greater than three.

For Model III the hypothesis is

$$E\left\{(X - \beta Y)(X - \beta Y)'\right\} = \psi$$

$$\beta = \begin{array}{c} q \\ r \\ p-q-r \end{array} \left[\begin{array}{ccc} \beta_{11} & \beta_{12} & 0 \\ \hline \beta_{21} & 0 & \beta_{23} \\ \hline 0 & \beta_{32} & \beta_{33} \end{array}\right] .$$

This implies that the population covariance matrix, in partitioned form, is

$$
C = \begin{array}{c} \\ q \\ r \\ p-q-r \end{array}
\begin{array}{c}
\begin{array}{ccc} q & r & p-q-r \end{array} \\
\left[\begin{array}{c|c|c}
\psi_{(1)} + \beta_{11}\beta'_{11} + \beta_{12}\beta'_{12} & \beta_{11}\,\beta'_{21} & \beta_{12}\,\beta'_{32} \\
\hline
\beta_{21}\,\beta'_{11} & \psi_{(2)} + \beta_{21}\beta'_{21} + \beta_{23}\beta'_{23} & \beta_{23}\,\beta'_{33} \\
\hline
\beta_{32}\,\beta'_{12} & \beta_{33}\,\beta'_{23} & \psi_{(3)} + \beta_{32}\beta'_{32} + \beta_{33}\beta'_{33}
\end{array}\right]
\end{array}
$$

Thurstone calls this a "complete triangular configuration". Again, generalization to m greater than three is readily apparent.

The likelihood equations for Model I may be greatly simplified, for if

$$
C = \begin{array}{c} \\ q \\ r \\ p-q-r \end{array}
\begin{array}{c}
\begin{array}{ccc} q & r & p-q-r \end{array} \\
\left[\begin{array}{c|c|c}
C_{11} & C_{12} & C_{13} \\
\hline
C'_{12} & C_{22} & C_{23} \\
\hline
C'_{13} & C'_{23} & C_{33}
\end{array}\right]
\end{array}
\quad \text{and} \quad
A = \begin{array}{c} \\ q \\ r \\ p-q-r \end{array}
\begin{array}{c}
\begin{array}{ccc} q & r & p-q-r \end{array} \\
\left[\begin{array}{c|c|c}
A_{11} & A_{12} & A_{13} \\
\hline
A'_{12} & A_{22} & A_{23} \\
\hline
A'_{13} & A'_{23} & A_{33}
\end{array}\right]
\end{array},
$$

then the elements of C_{12}, C_{13}, and C_{23} are all zero. Hence, the logarithm of the likelihood function of Wishart's distribution may be written as

$$
\text{Log } L = -\frac{N-1}{2} \left\{ \sum_{i=1}^{3} \left[\log |C_{ii}| + \text{trace } C^{ii} A_{ii} \right] \right\} +
$$

a function independent of the elements of the C_{ii}

for $i = 1, 2, 3$;

where C^{ii} is the inverse of C_{ii}. This implies that each set of variables may be treated separately, since a maximization of the sum is equivalent in this case to a maximization of each of the three components. Thus, this is equivalent to estimating one common factor from each group of variables, and is therefore identical to the problem in Chapter II. Hence, the maximum likelihood equations are

$$\hat{\beta}'_{11} \hat{C}^{11} A_{11} = \hat{\beta}'_{11}$$

$$\hat{\beta}'_{22} \hat{C}^{22} A_{22} = \hat{\beta}'_{22}$$

$$\hat{\beta}'_{33} \hat{C}^{33} A_{33} = \hat{\beta}'_{33}$$

$$\text{Diagonal } \hat{C} = \text{Diagonal } A \ .$$

Model II also produces some simplification in the equations for $U' \hat{\beta} \hat{\beta}'$ is an $m \times p$ matrix with zeros where β' is not specified to have zeros. Therefore, if $K' = U' + U' \hat{\beta} \hat{\beta}' \hat{\psi}^{-1}$, the equations may be written

$$\hat{\beta}' \hat{C}^{-1} A = K' \hat{\psi} + \hat{\beta}'$$

$$\text{Diagonal } (\hat{C}^{-1} A \hat{C}^{-1} - \hat{C}^{-1}) = 0 \quad ,$$

where K has zeros where β is not specified to have zeros. For this model then, the diagonal elements of $\hat{C}$ equal the diagonal elements of A.

However, for Model III no simplification is possible and the equations remain the same, requiring a much more complicated computational procedure than do Models I and II. It is therefore worthwhile for the factor analyst to ascertain whether some simplification of the likelihood equations is possible for the particular simple structure hypothesis to be tested. Models I and II are examples of such simplification.

4.4 Indeterminancy in the Model and in the Likelihood Equations

For the original model, $C = \psi + \beta\beta'$, discussed in Chapter II, it has been shown that β is determined uniquely except for multiplication on the right by an $m \times m$ orthogonal matrix S. The matrix S can be determined such that two conditions are satisfied. First, one column of βS has no zero elements, another has one zero element, still another has two zero elements, ..., and the last has $m - 1$ zero elements. Thus, there are $\frac{m(m-1)}{2}$ zeros in all. The second condition is that $\beta^{(\alpha)}$ $(\alpha = 1, 2, \ldots, m)$ has the same rank as the number of zeros in the αth column, where $\beta^{(\alpha)}$ is the submatrix of βS that has zero elements in the αth column. This follows from repeated application of a result of Roy's [30]; namely, that if $|G|_{m \times m} \neq 0$, there exists an orthogonal $m \times m$ matrix S such that $G S = H$, where H is a triangular matrix. It is clear that the preceding paragraph still holds true if β is replaced by $\hat{\beta}$.

This implies that certain simple structure hypotheses are actually equivalent to the general model discussed in Chapter II. Hypotheses of this sort may be specified as follows: adding zeros to the various columns in the hypothesized β matrix results in a matrix which satisfies the above two conditions. This means simply that a $\hat{\beta}$ obtained under the general model (no zeros specified) may be transformed into an estimated regression matrix of the form specified in the simple structure hypothesis. Hence, estimation based on the general model is sufficient in this case. If there are more than $\frac{m(m-1)}{2}$ zeros specified, then the methods discussed in this chapter must be used.

In addition, simple structure hypotheses in general may not define β uniquely. For example, a $p \times 3$ β matrix with more than three zeros in the third column and none in the first and second columns, is not uniquely determined. For, if β is postmultiplied by an orthogonal matrix of the following form:

$$\begin{bmatrix} a & \sqrt{1 - a^2} & 0 \\ -\sqrt{1 - a^2} & a & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

a new regression matrix of the same form as β results. Thus, it is sometimes possible to rotate under simple structure hypotheses.

Conditions for unique determination of β and $\hat{\beta}$ are the same as the two conditions listed previously except that now the first condition specifies only the minimum number of zeros that a column may have. There may be more than $\frac{m(m-1)}{2}$ zeros in the β matrix. This follows easily from the fact that under these conditions, any orthogonal matrix must be a triangular matrix to leave the form of β unchanged, and the only orthogonal triangular matrix is the unit matrix. If the conditions are not satisfied, β and $\hat{\beta}$ are not uniquely determined and it may be necessary to impose a condition analogous to

$$\hat{\beta}' \hat{\psi}^{-1} \hat{\beta} = \text{diagonal matrix},$$

in order to obtain a unique solution of the likelihood equations.

In Chapter VII this point will be discussed further, since it has a bearing on the distribution of the test statistic.

4.5 Calculation of the Estimates

In this section methods of computation for Models I-III will first be discussed. The method of solution for the general equations of Section 4.2 is then an obvious generalization of that for Model III and will be considered at the conclusion of the section.

In all the orthogonal simple structure hypotheses, the initial approximation is the rotated matrix of estimates, $S\hat{\lambda} = S\hat{\beta}'$. For Model I the methods for $m = 1$ discussed in Chapter III are used on the sample covariance matrices of the first q variables, the next

r variables, and the last $p - q - r$ variables. Then Model I contains no new computational difficulties.

For Model II the likelihood equations may be written

$$\hat{\beta}' \hat{C}^{-1} A = K' \hat{\psi} + \hat{\beta}' = \hat{\beta}^{*\prime} ,$$

say. If the equation is postmultiplied by $A^{-1} \hat{C}$ and $\hat{\psi} + \hat{\beta}\hat{\beta}'$ is substituted for $\hat{C}$, then

$$\hat{\beta}' = \hat{\beta}^{*\prime} A^{-1} \hat{\psi} + \hat{\beta}^{*\prime} A^{-1} \hat{\beta}\hat{\beta}' . \tag{4}$$

Postmultiplying the above equation by $\hat{\psi}^{-1} \hat{\beta}$, one obtains

$$\hat{\beta}' \hat{\psi}^{-1} \hat{\beta} = \hat{\beta}^{*\prime} A^{-1} \hat{\beta} + \hat{\beta}^{*\prime} A^{-1} \hat{\beta}\hat{\beta}' \hat{\psi}^{-1} \hat{\beta} ,$$

or

$$\hat{\beta}^{*\prime} A^{-1} \hat{\beta} = \hat{\beta}' \hat{\psi}^{-1} \hat{\beta} \left[I + \hat{\beta}' \hat{\psi}^{-1} \hat{\beta} \right]^{-1} .$$

Equation (4) is now postmultiplied by $\hat{\psi}^{-1} A$ to obtain

$$\hat{\beta}' \hat{\psi}^{-1} A = \hat{\beta}^{*\prime} + \hat{\beta}' \hat{\psi}^{-1} \hat{\beta} \left[I + \hat{\beta}' \hat{\psi}^{-1} \hat{\beta} \right]^{-1} \hat{\beta}' \hat{\psi}^{-1} A ,$$

or, after some simplification,

$$\hat{\beta}' \hat{\psi}^{-1} A = \left[I + \hat{\beta}' \hat{\psi}^{-1} \hat{\beta} \right] \hat{\beta}^{*\prime} .$$

Here, one can no longer specialize to the equations where the non-diagonal elements of $\hat{\beta}' \hat{\psi}^{-1} \hat{\beta}$ are zero. However, if $\hat{\beta}^{(n-1)}$ is the (n-1)st approximation to $\hat{\beta}$, one way to obtain the solution is to use a scheme as follows:

$$\hat{\beta}^{*(n)\prime} = \left[I + \hat{\beta}^{(n-1)\prime} \hat{\psi}^{(n-1)-1} \hat{\beta}^{(n-1)} \right]^{-1} \hat{\beta}^{(n-1)\prime} \hat{\psi}^{(n-1)-1} A ,$$

and repeat the procedure until $\hat{\beta}^{(n)}$ converges. The diagonal matrix $\hat{\psi}^{(n)}$ is obtained by subtracting from the diagonal elements of A, the corresponding elements of $\hat{\beta}^{(n)} \hat{\beta}^{(n)\prime}$. $\hat{\beta}^{(n)\prime}$ is easily obtained from

$$\hat{\beta}^{*(n)\prime} = \hat{\beta}^{(n)\prime} + \begin{array}{c} \begin{array}{cc} q & p-q \end{array} \\ \left[\begin{array}{c|c} 0 & 0 \\ 0 & u \\ v & 0 \end{array} \right] \end{array} .$$

A method which gives faster convergence is specified by

$$\hat{\beta}^{*(n)\prime} = \left[\hat{\beta}^{(n-1)\prime} \hat{\psi}^{(n-1)-1} \hat{\beta}^{(n-1)} \right]^{-1} \left[\hat{\beta}^{(n-1)} \hat{\psi}^{(n-1)-1} A - \hat{\beta}^{*(n-1)\prime} \right] .$$

For the initial approximation in this method, u and v are assumed zero, but after the first iteration, estimates are available.

A Gauss-Seidel technique which seems to give the fastest convergence with the least amount of work, can also be derived. The likelihood equations can be written in scalar form as

$$\sum_{i=1}^{p} \frac{f_i}{\hat{\psi}_{ii}} (a_{ij} - f_i f_j - g_i g_j^* - h_i h_j^*) = f_j$$

$$\sum_{i=1}^{p} \frac{g_i}{\hat{\psi}_{ii}} (a_{ij} - f_i f_j - g_i g_j^* - h_i h_j^*) = g_j^*$$

$$\sum_{i=1}^{p} \frac{h_i}{\hat{\psi}_{ii}} (a_{ij} - f_i f_j - g_i g_j^* - h_i h_j^*) = h_j^* \qquad j = 1,2,\ldots,p ,$$

where the f_i are the elements of the vector $\left[\overset{q}{\hat{\beta}'_{11}} \;\vdots\; \overset{p-q}{\hat{\beta}'_{21}}\right]$; the g_i are the elements of the vector

$$\left[\overset{q}{\hat{\beta}'_{12}} \;\vdots\; \overset{p-q}{0}\right] ;$$

the h_i are the elements of the vector

$$\left[\overset{q}{0} \;\vdots\; \overset{p-q}{\hat{\beta}'_{23}}\right] ;$$

the g_i^* are the elements of the second row of $\hat{\beta}^{*\prime}$; h_i^* are the elements of the third row of $\hat{\beta}^{*\prime}$. Hence,

$$g_i^* = g_i \qquad i = 1, 2, \ldots, q$$

$$h_i^* = h_i \qquad i = q+1, \ldots, p \quad .$$

Since $g_i = 0$ $(i = q+1, \ldots, p)$, and $h_i = 0$ $(i = 1, 2, \ldots, q)$, the equations are, for $j = 1, 2, \ldots, p$,

$$\sum_{i=1}^{q} \frac{f_i}{\hat{\psi}_{ii}} (a_{ij} - f_i f_j - g_i g_j^*) + \sum_{i=q+1}^{p} \frac{f_i}{\hat{\psi}_{ii}} (a_{ij} - f_i f_j - h_i h_j^*) = f_j$$

$$\sum_{i=1}^{q} \frac{g_i}{\hat{\psi}_{ii}} (a_{ij} - f_i f_j - g_i g_j^*) = g_j^*$$

$$\sum_{i=q+1}^{p} \frac{h_i}{\hat{\psi}_{ii}} (a_{ij} - f_i f_j - h_i h_j^*) = h_j^* \quad .$$

Now

$$\hat{\psi}_{ii} = a_{ii} - f_i^2 - g_i^2 \qquad i = 1, 2, \ldots, q$$

$$\hat{\psi}_{jj} = a_{jj} - f_j^2 - h_j^2 \qquad j = q+1, \ldots, p \quad .$$

Therefore, for $j = 1, 2, \ldots, q$,

$$\sum_{\substack{i=1\\ i\neq j}}^{q} \frac{f_i}{\Psi_{ii}} (a_{ij} - f_i f_j - g_i g_j^*) + \sum_{i=q+1}^{p} \frac{f_i}{\Psi_{ii}} (a_{ij} - f_i f_j - h_i h_j^*) = 0$$

$$\sum_{\substack{i=1\\ i\neq j}}^{q} \frac{g_i}{\Psi_{ii}} (a_{ij} - f_i f_j - g_i g_j^*) = 0$$

$$\sum_{i=q+1}^{p} \frac{h_i}{\Psi_{ii}} (a_{ij} - f_i f_j - h_i h_j^*) = h_j^* \quad .$$

Hence, they may also be written, for $j = 1, 2, \ldots, q$,

$$f_j \sum_{\substack{i=1\\ i\neq j}}^{p} \frac{f_i^2}{\Psi_{ii}} + g_j^* \sum_{\substack{i=1\\ i\neq j}}^{q} \frac{g_i f_i}{\Psi_{ii}} + h_j^* \sum_{i=q+1}^{p} \frac{f_i h_i}{\Psi_{ii}} = \sum_{\substack{i=1\\ i\neq j}}^{p} \frac{f_i a_{ij}}{\Psi_{ii}}$$

$$f_j \sum_{\substack{i=1\\ i\neq j}}^{q} \frac{f_i g_i}{\Psi_{ii}} + g_j^* \sum_{\substack{i=1\\ i\neq j}}^{q} \frac{g_i^2}{\Psi_{ii}} = \sum_{\substack{i=1\\ i\neq j}}^{q} \frac{g_i a_{ij}}{\Psi_{ii}}$$

$$f_j \sum_{i=q+1}^{p} \frac{f_i h_i}{\Psi_{ii}} + h_j^* \left[\sum_{i=q+1}^{p} \frac{h_i^2}{\Psi_{ii}} + 1 \right] = \sum_{i=q+1}^{p} \frac{h_i a_{ij}}{\Psi_{ii}} \quad .$$

Similarly, for $j = q+1, \ldots, p$,

$$f_j \sum_{\substack{i=1 \\ i\neq j}}^{p} \frac{f_i^2}{\hat{\psi}_{ii}} + g_j^* \sum_{i=1}^{q} \frac{f_i\, g_i}{\hat{\psi}_{ii}} + h_j^* \sum_{\substack{i=q+1 \\ i\neq j}}^{p} \frac{f_i\, h_i}{\hat{\psi}_{ii}} = \sum_{\substack{i=1 \\ i\neq j}}^{p} \frac{f_i\, a_{ij}}{\hat{\psi}_{ii}}$$

$$f_j \sum_{i=1}^{q} \frac{f_i\, g_i}{\hat{\psi}_{ii}} + g_j^* \left[\sum_{i=1}^{q} \frac{g_i^2}{\hat{\psi}_{ii}} + 1 \right] = \sum_{i=1}^{q} \frac{g_i\, a_{ij}}{\hat{\psi}_{ii}}$$

$$f_j \sum_{\substack{i=q+1 \\ i\neq j}}^{p} \frac{f_i\, h_i}{\hat{\psi}_{ii}} + h_j^* \sum_{\substack{i=q+1 \\ i\neq j}}^{p} \frac{h_i^2}{\hat{\psi}_{ii}} = \sum_{\substack{i=q+1 \\ i\neq j}}^{p} \frac{h_i\, a_{ij}}{\hat{\psi}_{ii}} \quad .$$

Computation then proceeds as described in Chapter III for the Gauss-Seidel method.

As an example the 8×8 correlation matrix in Chapter III obtained from Lawley's 1943 paper, will be considered. Let it be supposed that the estimate

$$\hat{\beta}' = \begin{bmatrix} .706 & .515 & .731 & .648 & .612 & .394 & .711 & .663 \\ .240 & -.176 & -.471 & .161 & .139 & .069 & .183 & .344 \end{bmatrix},$$

has been obtained from another sample drawn from the same population. Moreover, suppose that by one or the other rotational methods operating

on $\hat{\beta}$, the factor analyst has arrived at the orthogonal matrix S , where

$$S = \begin{bmatrix} .9549 & .2967 \\ .2967 & -.9549 \end{bmatrix} .$$

Then,

$$S\hat{\beta}' = \begin{bmatrix} .745 & .439 & .558 & .666 & .626 & .396 & .733 & .736 \\ -.019 & .321 & .667 & .038 & .049 & .051 & .036 & -.132 \end{bmatrix} .$$

The factor analyst now makes the hypothesis that there is a general factor, y_1 , which is correlated with all the variables, and that there is a second factor, y_2 , uncorrelated with y_1 , such that $\rho_{iy_2} = 0$ for $i = 1, 4, 5, 6, 7, 8$; ρ_{iy_2} is the correlation coefficient of x_i with y_2 . Now the object is to estimate the parameters under this model. As a first approximation, let

$$\hat{\beta}^{(1)'} = \begin{bmatrix} .745 & .439 & .558 & .666 & .626 & .396 & .733 & .736 \\ 0 & .321 & .667 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} .$$

Therefore,

$$\psi^{(1)} = \begin{bmatrix} .4451 & & & & & & & \\ & .7040 & & & & & & \\ & & .2440 & & & & & \\ & & & .5562 & & & & \\ & & & & .6087 & & & \\ & & & & & .8428 & & \\ & & & & & & .4630 & \\ & & & & & & & .4587 \end{bmatrix} .$$

Then to obtain $f_2^{(2)}$ and $g_2^{(2)}$, the following quantities are needed:

$$\sum_{i \neq 2} \frac{f_i^2}{\hat{\psi}_{ii}} = \frac{(.745)^2}{.4451} + \frac{(.558)^2}{.2440} + \ldots + \frac{(.736)^2}{.4587} = 6.4895$$

$$\sum_{i \neq 2} \frac{f_i \, g_i}{\hat{\psi}_{ii}} = \frac{.558\ (.667)}{.2440} = 1.5248$$

$$\sum_{i \neq 2} \frac{g_i^2}{\hat{\psi}_{ii}} = \frac{(.667)^2}{.2440} = 1.8221$$

$$\sum_{i \neq 2} \frac{f_i \, a_{i2}}{\hat{\psi}_{ii}} = \frac{(.745)(.312)}{.4451} + \ldots + \frac{(.736)(.288)}{.4587} = 3.3190$$

$$\sum_{i \neq 2} \frac{g_i \, a_{i2}}{\hat{\psi}_{ii}} = \frac{.667\ (.460)}{.2440} = 1.2570.$$

Finally,

$$\begin{bmatrix} 6.4895 & 1.5248 \\ 1.5248 & 1.8221 \end{bmatrix} \begin{bmatrix} f_2^{(2)} \\ g_2^{(2)} \end{bmatrix} = \begin{bmatrix} 3.3190 \\ 1.2510 \end{bmatrix} ,$$

or $f_2^{(2)} = .4394$, $g_2^{(2)} = .3208$. Then $f_3^{(2)}$ and $g_3^{(2)}$ are obtained in the same manner, but $f_2^{(2)}$ and $g_2^{(2)}$ are used in the calculations, rather than $f_2^{(1)}$ and $g_2^{(1)}$. To obtain new estimates $f_i^{(2)}$ $(i = 1, 4, 5, 6, 7, 8)$, the same procedure is followed, except that

$$1 + \sum_{j \neq i} \frac{g_j^2}{\hat{\psi}_{jj}}$$

is used, not

$$\sum_{j \neq i} \frac{g_j^2}{\hat{\psi}_{jj}} .$$

The matrix, $\hat{\psi}$, is, of course, changed with each new estimate. After a few iterations, the results are, accurate to three decimals:

$$\beta' = \begin{bmatrix} .742 & .437 & .554 & .674 & .636 & .403 & .739 & .710 \\ 0 & .324 & .672 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} .$$

As far as the possible divergence of the methods is concerned, both computational schemes described above have converged to correct

solutions in all examples attempted, but it should be noted that a proof of convergence has not been obtained. This also applies to computational methods in the general case, as discussed in Chapter III, and to the scheme for Model III discussed in the following pages.

The likelihood equations for Model III are

$$\hat{\beta}' \hat{C}^{-1} A = \hat{\beta}^{*\prime} ,$$

where $\hat{\beta}^{*\prime} = \hat{\beta}' + U'\hat{\beta}\hat{\beta}' + U'\hat{\psi}$, and U is of the form specified previously. Proceeding exactly as for Model II, one can show that the equations may be written

$$\hat{\beta}' \hat{\psi}^{-1} A = \left[I + \hat{\beta}' \hat{\psi}^{-1} \hat{\beta} \right] \hat{\beta}^{*\prime} .$$

If the above equation is postmultiplied by $\hat{\psi}^{-1} \hat{\beta}$, then

$$\hat{\beta}' \hat{\psi}^{-1} A \hat{\psi}^{-1} \hat{\beta} = \left[I + \hat{\beta}' \hat{\psi}^{-1} \hat{\beta} \right] \left[\hat{\beta}' \hat{\psi}^{-1} \hat{\beta} + U'\hat{\beta}\hat{\beta}' \hat{\psi}^{-1} \hat{\beta} + U'\hat{\beta} \right] .$$

This is equivalent to

$$\hat{\beta}' \hat{\psi}^{-1} A \hat{\psi}^{-1} \hat{\beta} + \left[I + \hat{\beta}' \hat{\psi}^{-1} \hat{\beta} \right] = \left[I + \hat{\beta}' \hat{\psi}^{-1} \hat{\beta} \right] W \left[I + \hat{\beta}' \hat{\psi}^{-1} \hat{\beta} \right] ,$$

where

$$W_{3 \times 3} = I + U'\hat{\beta} .$$

Therefore,

$$W = \left[I + \hat{\beta}' \hat{\psi}^{-1} \hat{\beta}\right]^{-1} \hat{\beta}' \hat{\psi}^{-1} A \hat{\psi}^{-1} \hat{\beta} \left[I + \hat{\beta}' \hat{\psi}^{-1} \hat{\beta}\right]^{-1} + \left[I + \hat{\beta}' \hat{\psi}^{-1} \hat{\beta}\right]^{-1} .$$

The computation proceeds as follows: take an initial approximation $\hat{\beta}^{(1)'}$; compute $\hat{\psi}^{(1)}$, assuming W is the identity matrix. This implies that the elements of $\hat{\psi}^{(1)}$ are the diagonal elements of A minus the diagonal elements of $\hat{\beta}^{(1)} \hat{\beta}^{(1)'}$. Now compute $W^{(1)}$ by using the above equation for W . Then obtain

$$\hat{\beta}^{*(1)'} = \left[I + \hat{\beta}^{(1)'} \hat{\psi}^{(1)-1} \hat{\beta}^{(1)}\right]^{-1} \hat{\beta}^{(1)'} \hat{\psi}^{(1)-1} A \quad ;$$

both of these matrices have been obtained in computing $W^{(1)}$. Now $\hat{\beta}^{*'}$ is of the following form:

$$\begin{array}{c} \begin{array}{ccc} q & r & p-q-r \end{array} \\ \left[\begin{array}{c|c|c} \hat{\beta}'_{11} + w_{12} \hat{\beta}'_{12} & \hat{\beta}'_{21} + w_{13} \hat{\beta}'_{23} & a\, C_{33} \\ \hat{\beta}'_{12} + w_{12} \hat{\beta}'_{11} & b\, C_{22} & \hat{\beta}'_{32} + w_{23} \hat{\beta}'_{33} \\ c\, C_{11} & \hat{\beta}'_{23} + w_{13} \hat{\beta}'_{21} & \hat{\beta}'_{33} + w_{23} \hat{\beta}_{32} \end{array}\right] \end{array} = \begin{array}{c} \begin{array}{ccc} q & r & p-q-r \end{array} \\ \left[\begin{array}{c|c|c} \hat{\beta}^{*'}_{11} & \hat{\beta}^{*'}_{21} & \hat{\beta}^{*'}_{31} \\ \hat{\beta}^{*'}_{12} & \hat{\beta}^{*'}_{22} & \hat{\beta}^{*'}_{32} \\ \hat{\beta}^{*'}_{13} & \hat{\beta}^{*'}_{23} & \hat{\beta}^{*'}_{33} \end{array}\right] \end{array} ;$$

where the w_{ij} are the elements of the symmetric matrix W . It will be noticed that the estimates of W are also symmetric matrices. Since an estimate of W is now available, the second approximations to $\hat{\beta}'_{11}$ and $\hat{\beta}'_{12}$ are obtained by

$$\begin{bmatrix} 1 & w_{12}^{(1)} \\ w_{12}^{(1)} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\beta}_{11}^{*(1)'} \\ \hat{\beta}_{12}^{*(1)'} \end{bmatrix},$$

to $\hat{\beta}'_{21}$ and $\hat{\beta}'_{23}$ by

$$\begin{bmatrix} 1 & w_{13}^{(1)} \\ w_{13}^{(1)} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\beta}_{21}^{*(1)'} \\ \hat{\beta}_{23}^{*(1)'} \end{bmatrix},$$

and finally to $\hat{\beta}'_{32}$ and $\hat{\beta}'_{33}$ by

$$\begin{bmatrix} 1 & w_{23}^{(1)} \\ w_{23}^{(1)} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\beta}_{32}^{*(1)'} \\ \hat{\beta}_{33}^{*(1)'} \end{bmatrix}.$$

Thus, $\hat{\beta}^{(2)}$ is obtained. To compute $\hat{\psi}^{(2)}$, the non-zero values of the elements of W must be considered. The diagonal elements of $\hat{\beta}^{(2)}\hat{\beta}^{*(1)'}$ are obtained, and these are subtracted from the diagonal elements of A; the resulting differences are the $\hat{\psi}_{ii}^{(2)}$ $(i = 1,2,\ldots,p)$. Then $W^{(2)}$ is computed from $\hat{\beta}^{(2)}$ and $\hat{\psi}^{(2)}$, with the use of the same equation as before. At this point a procedure slightly different from that in the first iteration, is followed, since convergence is increased to some extent. The following expression is computed for an

estimate of $\hat{\beta}^{*(2)}$,

$$\left[\hat{\beta}^{(2)'}\hat{\psi}^{(2)-1}\hat{\beta}^{(2)}\right]^{-1}\left[\hat{\beta}^{(2)'}\hat{\psi}^{(2)-1}A - \hat{\beta}^{*(1)'}\right] = \hat{\beta}^{*(2)'} .$$

Then one proceeds exactly as before with $\hat{\beta}^{*(2)}$ to obtain $\hat{\beta}^{(3)}$. The iteration is continued until convergence is obtained. It may seem somewhat surprising that such a complex scheme does converge, but examples have resulted in solutions satisfying the original likelihood equations. One example with an artificially constructed correlation matrix follows.

The 6×6 correlation matrix, A , with lower non-diagonal terms omitted is

$$\begin{bmatrix} 1.00 & .58 & .15 & .34 & .19 & .05 \\ & 1.00 & .14 & .31 & .37 & .04 \\ & & 1.00 & .44 & .16 & .46 \\ & & & 1.00 & .17 & .36 \\ & & & & 1.00 & .26 \\ & & & & & 1.00 \end{bmatrix} .$$

As an initial approximation, take

$$\hat{\beta}^{(1)'} = \begin{bmatrix} .7 & .6 & .2 & .5 & 0 & 0 \\ .3 & .5 & 0 & 0 & .7 & .1 \\ 0 & 0 & .6 & .5 & .3 & .7 \end{bmatrix} .$$

Then,

$$\hat{\psi}^{(1)} = \begin{bmatrix} .42 & & & & & \\ & .39 & & & & \\ & & .60 & & & \\ & & & .50 & & \\ & & & & .42 & \\ & & & & & .50 \end{bmatrix} ;$$

$$\hat{\beta}^{(1)\prime}\hat{\psi}^{(1)-1} = \begin{bmatrix} 1.6667 & 1.5385 & .3333 & 1.0000 & 0 & 0 \\ .7143 & 1.2821 & 0 & 0 & 1.6667 & .2000 \\ 0 & 0 & 1.0000 & 1.000 & .7143 & 1.4000 \end{bmatrix} ;$$

$$\hat{\beta}^{(1)\prime}\hat{\psi}^{(1)-1}A = \begin{bmatrix} 2.9490 & 2.8618 & 1.2387 & 2.1903 & 1.1092 & .6582 \\ 1.7846 & 2.3211 & .6453 & .9957 & 2.3288 & .7203 \\ .6957 & .7703 & 2.1983 & 2.0654 & 1.4083 & 2.4057 \end{bmatrix} ;$$

$$\hat{\beta}^{(1)\prime}\hat{\psi}^{(1)-1}A\hat{\psi}^{(1)-1}\hat{\beta}^{(1)} = \begin{bmatrix} 11.9211 & 7.7559 & 5.1428 \\ & 8.2761 & 4.3129 \\ & & 8.6376 \end{bmatrix} ;$$

$$I+\hat{\beta}^{(1)\prime}\hat{\psi}^{(1)-1}\hat{\beta}^{(1)} = \begin{bmatrix} 3.65645 & 1.26926 & .69998 \\ & 3.04203 & .64001 \\ & & 3.29429 \end{bmatrix} .$$

Hence,

$$W^{(1)} = \begin{bmatrix} 1.00150 & .00280 & .01267 \\ & 1.00311 & -.00917 \\ & & 1.01432 \end{bmatrix} ;$$

$$\hat{\beta}^{*(1)'} = \begin{bmatrix} .7043 & .6046 & .2239 & .4957 & .0038 & .0159 \\ .2918 & .5094 & -.0122 & .0112 & .7029 & .0805 \\ .0048 & .0064 & .6221 & .5194 & .2901 & .7112 \end{bmatrix} .$$

Therefore,

$$\begin{bmatrix} 1 & w_{12}^{(1)} \\ w_{12}^{(1)} & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1.00001 & -.00280 \\ & 1.00001 \end{bmatrix}$$

$$\begin{bmatrix} 1 & w_{13}^{(1)} \\ w_{13}^{(1)} & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1.00016 & -.01267 \\ & 1.00016 \end{bmatrix}$$

$$\begin{bmatrix} 1 & w_{23}^{(1)} \\ w_{23}^{(1)} & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1.00008 & .00917 \\ & 1.00008 \end{bmatrix} .$$

The result is

$$\hat{\beta}^{(2)'} = \begin{bmatrix} .7035 & .6032 & .2161 & .4892 & 0 & 0 \\ .2898 & .5077 & 0 & 0 & .7056 & .0870 \\ 0 & 0 & .6194 & .5132 & .2966 & .7120 \end{bmatrix} .$$

Then $\hat{\psi}_{11}^{(2)}$, say, = 1 - .7043 (.7035)-.2918 (.2898) , and

$$\hat{\psi}^{(2)} = \begin{bmatrix} .41996 & & & & & \\ & .37668 & & & & \\ & & .56629 & & & \\ & & & .49095 & & \\ & & & & .41799 & \\ & & & & & .48662 \end{bmatrix} .$$

For the second iteration the following matrices are computed:

$$\hat{\beta}^{(2)'}\hat{\psi}^{(2)-1} = \begin{bmatrix} 1.6752 & 1.6014 & .3816 & .9964 & 0 & 0 \\ .6901 & 1.3478 & 0 & 0 & 1.6881 & .1788 \\ 0 & 0 & 1.0938 & 1.0453 & .7096 & 1.4632 \end{bmatrix} ;$$

$$\hat{\beta}^{(2)'}\hat{\psi}^{(2)-1}A = \begin{bmatrix} 3.0000 & 2.9353 & 1.2955 & 2.2303 & 1.1412 & .6821 \\ 1.8015 & 2.3798 & .6446 & 1.0038 & 2.3644 & .7061 \\ .7275 & .7983 & 2.3403 & 2.1740 & 1.4427 & 2.5272 \end{bmatrix} ;$$

$$\hat{\beta}^{(2)\prime}\hat{\psi}^{(2)-1}A\,\hat{\psi}^{(2)-1}\hat{\beta}^{(2)} = \begin{bmatrix} 12.4428 & 8.0749 & 5.5562 \\ & 8.5683 & 4.4653 \\ & & 9.5538 \end{bmatrix} ;$$

$$I + \hat{\beta}^{(2)\prime}\hat{\psi}^{(2)-1}\hat{\beta}^{(2)} = \begin{bmatrix} 3.71437 & 1.29850 & .74772 \\ & 3.09095 & .62800 \\ & & 3.46621 \end{bmatrix} .$$

Therefore,

$$W^{(2)} = \begin{bmatrix} 1.00350 & .00261 & .00904 \\ & 1.00023 & -.00506 \\ & & 1.00191 \end{bmatrix} ,$$

and

$$\hat{\beta}^{*(2)\prime} = \begin{bmatrix} .7106 & .6118 & .2294 & .4888 & .0016 & .0119 \\ .2787 & .5131 & -.0181 & .0154 & .7075 & .0777 \\ .0066 & .0049 & .6317 & .5188 & .2867 & .7130 \end{bmatrix} .$$

Then, proceeding as before, one obtains

$$\hat{\beta}^{(3)'} = \begin{bmatrix} .7099 & .6105 & .2237 & .4841 & 0 & 0 \\ .2768 & .5115 & 0 & 0 & .7090 & .0813 \\ 0 & 0 & .6297 & .5144 & .2903 & .7134 \end{bmatrix} .$$

The process is continued until $\hat{\beta}^{(n)'}$ converges to

$$\hat{\beta}' = \begin{bmatrix} .694 & .644 & .219 & .469 & 0 & 0 \\ .253 & .508 & 0 & 0 & .715 & .089 \\ 0 & 0 & .647 & .514 & .277 & .703 \end{bmatrix} ,$$

which is the solution of the likelihood equations.

The likelihood equations for Model III are identical with the general equations derived in Section 4.2. Therefore, the computational method for Model III is applicable in the general case, except that the form of β^* must be examined and expressed in terms of β. This determines what submatrices of W must be inverted in order to obtain an estimate of β. Otherwise, the computation proceeds exactly as with Model III.

A Gauss-Seidel type of iterative technique can also be used for the general case. However, because of two difficulties the scheme is not generally recommended. First, $W = I + U' \hat{\beta}$ must be computed for each new estimate of a row of β. This may change every element

in the estimated ψ matrix, leading to excessive calculation. The second difficulty is that the elements of $\hat{\psi}$ are the diagonal elements of A minus the diagonal elements of $\hat{\beta}' W \hat{\beta}$. Thus, it is not usually possible to obtain a row of $\hat{\beta}$ in terms of A and the other $m(p - 1)$ elements of $\hat{\beta}$ alone, since the elements of $\hat{\psi}$ may all depend on that particular row of $\hat{\beta}$. If, however, $W = I$ or the diagonal elements of $\hat{C}$ equal the diagonal elements of A, this technique is to be recommended; an example has been given for Model II.

The general formula may be worked out quite easily from equation (3) and is given by

$$\hat{\beta}'_{(j)} \hat{\psi}^{-1} A_j = \left[\left(\hat{\beta}'_{(j)} \hat{\psi}^{-1} \hat{\beta}_{(j)} W\right) + W - I\right] \hat{\beta}'_j + \left[I + \hat{\beta}'_{(j)} \hat{\psi}^{-1} \hat{\beta}_{(j)}\right] Q_j$$

$$j = 1, 2, \ldots, p \quad .$$

Here $\hat{\beta}_{(j)}$ is the $p \times m$ matrix obtained from $\hat{\beta}$ by inserting zeros in the $j\underline{th}$ row; $\hat{\beta}_j$ is the $j\underline{th}$ row of $\hat{\beta}$; A_j is the $j\underline{th}$ row of A; Q_j is the $j\underline{th}$ column of $U' \hat{\psi}$. $\hat{\beta}'_j$ and Q_j between them contain m non-zero quantities, since U is specified to have zeros where $\hat{\beta}$ is not specified to have zeros. Therefore, the equation above may be thought of as giving the $j\underline{th}$ row of $\hat{\beta}$ in terms of A_j and the other elements of $\hat{\beta}$; however, it is subject to the two difficulties mentioned previously. The computational equations for Model II may be easily derived from this relation.

It should also be noted that, although computational methods have been discussed only for $m = 3$, generalization to m greater than three is immediate.

4.6 General Remarks

Throughout this chapter simple structure hypotheses specifying zeros in the β matrix have been considered. However, there is no reason to restrict oneself to such a definition of simple structure. For example, if the hypothesis states that a block of tests have identical factor loadings on some factor, the form of β is determined and the likelihood equations are found exactly as before. Moreover, it is felt that with the aid of this paper the factor analyst can translate any simple structure hypothesis, such as that in the example above, into a hypothesis on the form of the β matrix, obtain the likelihood equations and their solution.

It is unfortunate that no better method is available for solving the likelihood equations; however, the author was unable to find a better iterative technique. The proposed method is not simple, but it can be done on desk computers. For large p high speed electronic computers would almost be a necessity.

In the next chapter simple structure hypotheses for oblique factors will be considered.

CHAPTER V

OBLIQUE FACTORS AND ROTATION

5.1 Oblique Factors

In Chapter II it has been shown that

$$\beta\, E(Y\, Y') = \lambda' \quad .$$

If it is assumed that $E(Y\, Y') = I$, then one can go directly from β to λ , and estimates of β are estimates of λ' . On the other hand, if $E(Y\, Y') = F \neq I$, where F is a symmetric matrix with ones in the diagonal, this relationship no longer holds. If

$$E\left\{(X - \beta\, Y)(X - \beta\, Y)'\right\} = \psi$$

$$E(Y\, Y') = I \quad ,$$

where ψ is a diagonal matrix, and if a non-singular linear transformation S is applied to Y , $W = S\, Y$, then

$$\beta^{(1)}\, E(W\, W') = E(X\, W') = \lambda^{(1)'} = \lambda'\, S' \quad .$$

Therefore,

$$\beta^{(1)} S S' = \lambda' S' \qquad \text{or} \qquad \beta^{(1)} S = \beta \quad .$$

Hence, applying a linear transformation to λ is equivalent to applying the same transformation to Y, but applying a linear transformation to β is equivalent to applying the inverse of that transformation to Y. The psychologists look for a transformation matrix, S, which, when applied to $\hat{\beta}$ will give an indication of simple structure. S is usually determined such that $S S'$ has unities in the diagonal.

As an example, consider the following estimate for $m = 2$, obtained from a 5×5 matrix:

$$\hat{\lambda} = \hat{\beta}' = \begin{bmatrix} .483 & .579 & .664 & .277 & .708 \\ .174 & .173 & .206 & -.167 & -.385 \end{bmatrix} .$$

Then, by the use of rotational methods, they arrive at the matrix S such that

$$S = \begin{bmatrix} .3030 & -.9530 \\ .4830 & .8756 \end{bmatrix} , \quad \text{where} \quad S S' = \begin{bmatrix} 1.000 & -.688 \\ -.688 & 1.000 \end{bmatrix} .$$

Hence,

$$S\hat{\lambda} = S\hat{\beta}' = H = \begin{bmatrix} -.019 & .011 & .005 & .243 & .581 \\ .386 & .431 & .501 & -.012 & .005 \end{bmatrix} .$$

On an examination of the matrix, H, the factor analysts then hypothesize simple structure. The question then arises, under what conditions does H satisfy the likelihood equations? The likelihood equations are

$$\hat{\lambda}\hat{C}^{-1}A = \hat{\lambda} \quad .$$

Now

$$\hat{C} = \hat{\psi} + \hat{\lambda}'\hat{\lambda} = \hat{\psi} + H'\left[S\,S'\right]^{-1} H \quad .$$

If $H = S\hat{\lambda}$ is considered as an estimate of the covariance of X and W, then $E(W\,W') = S\,S'$. However, if $H = S\hat{\beta}'$ is considered as an estimate of the regression coefficients of X on W, then $E(W\,W') = (S\,S')^{-1}$. The psychologists consider H as an estimate of the regression matrix, since the simple structure hypothesis usually specifies blocks of zeros in the regression matrix, β. A variable x_i may be correlated with more than one common factor, but the factor

analysts are interested in whether the correlations among x_i and $m - 1$ common factors may be explained by the remaining common factor. This can be expressed as

$$\rho_{iy_k \cdot y_1} = 0 \qquad k = 2, 3, \ldots, m \quad ,$$

which implies that

$$\beta_{iy_k} = 0 \qquad k = 2, 3, \ldots, m \quad .$$

Therefore, the interest of the psychologists is in the regression matrix, β , not λ . This implies that if $H = S\,\hat{\beta}'$ is a solution of the equations, then $E(W\,W') = (S\,S')^{-1}$. Now throughout this paper it has been assumed that the common factors have unit variances. Hence, $(S\,S')^{-1}$ must have unities in the diagonal, but the factor analysts have determined S such that $S\,S'$ has unities in the diagonal. The matrix $D = M\,S$ is then to be determined, where M is a diagonal matrix so that $(D\,D')^{-1}$ has unities in the diagonal. For the example presented previously

$$(S\,S')^{-1} = \begin{bmatrix} 1.8993 & -1.3069 \\ -1.3069 & 1.8993 \end{bmatrix} \quad ,$$

$$M = \begin{bmatrix} 1.3780 & 0 \\ 0 & 1.3780 \end{bmatrix} ,$$

$$D = \begin{bmatrix} .4175 & -1.3132 \\ .6656 & 1.2066 \end{bmatrix} ,$$

$$(D\,D')^{-1} = \begin{bmatrix} 1.000 & .688 \\ .688 & 1.000 \end{bmatrix} ,$$

$$D \hat{\beta}' = \begin{bmatrix} -.027 & .015 & .007 & .335 & .801 \\ .531 & .594 & .691 & -.017 & .007 \end{bmatrix} .$$

The estimate of the correlation of the two common factors is .688 , the non-diagonal element of $(D\,D')^{-1}$. This is equivalent to Thurstone's approach [33, pp.137] ; however, in his method $S\,\hat{\beta}'$ is used as an estimate of β' , not $M\,S\,\hat{\beta}' = D\,\hat{\beta}'$, which should be used.

Most of the general remarks in Section 4.1 also apply to oblique factors. The point is again made that the hypothesis should not be tested on the same data which generated it. As before, it is assumed that the factor analyst has obtained the D matrix, has made his hypothesis, and is now ready to test it on a new sample.

5.2 The Maximum Likelihood Equations

Simple structure hypotheses for oblique factors may be stated as follows: a sample of N has been drawn from a p-variate normal population with mean μ and covariance matrix $C = \psi + \beta F \beta'$, where β is a $p \times m$ matrix, $m < p$, ψ is a $p \times p$ diagonal matrix, and F is a positive definite symmetric matrix with ones in the diagonal. Certain elements of β, say $\beta_{i_1 j_1}, \beta_{i_2 j_2}, \ldots, \beta_{i_r j_r}$, are assumed zero, where $r > m(m-1)$; $i_1, i_2, \ldots, i_r$ can assume any value from 1 to p; and $j_1, j_2, \ldots, j_r$ any value from 1 to m. F is the moment matrix of the common factors Y, standardized to unit variance, and β is the linear mean square regression matrix of X on Y.

Exactly as in Section 4.2 one may use the method of maximum likelihood to obtain estimates of the elements of ψ, β, and F. Maximizing the likelihood function for Wishart's distribution leads to the following equations:

$$\hat{F}\hat{\beta}'(\hat{C}^{-1} A \hat{C}^{-1} - \hat{C}^{-1}) = J'$$

$$\hat{\beta}'(\hat{C}^{-1} A \hat{C}^{-1} - \hat{C}^{-1})\hat{\beta} = V \tag{1}$$

$$\text{Diagonal}\,(\hat{C}^{-1} A \hat{C}^{-1} - \hat{C}^{-1}) = 0 ,$$

where V is a diagonal matrix and J is a $p \times m$ matrix with zeros

where β is not specified to have zeros. If $B = \hat{C}^{-1} A \hat{C}^{-1} - \hat{C}^{-1}$ with typical element b_{ij}, the equations may be written

$$\hat{F}\hat{\beta}' B = J'$$

$$(2) \qquad \hat{\beta}' B \hat{\beta} = V$$

$$b_{ii} = 0 \qquad i = 1, 2, \ldots, p \ .$$

Equation (2) implies that $\hat{\psi} B$ and $\hat{\beta}\hat{F}\hat{\beta}' B$ have zeros in the diagonal, since $\hat{\beta} J'$ has zeros in the diagonal by definition of J. Hence, $(\hat{\psi} + \hat{\beta}\hat{F}\hat{\beta}')B = \hat{C} B = A \hat{C}^{-1} - I$ has zeros in the diagonal and therefore $A \hat{C}^{-1}$ has ones in the diagonal.

If the first equation of (1) is postmultiplied by $\hat{\beta}$, it is clear that $J' \hat{\beta} = \hat{F} V$, and since the diagonal elements of $J' \hat{\beta}$ are zero, the diagonal elements of V are zero, else $\hat{F}$ would have zeros in the diagonal. Hence, $J' \hat{\beta} = 0 = V$. Equation (1) is then premultiplied by $\hat{\beta}$ and postmultiplied by $\hat{C}$ to obtain

$$\hat{\beta}\hat{F}\hat{\beta}'(\hat{C}^{-1} A - I) = (\hat{C} - \hat{\psi})(\hat{C}^{-1} A - I) = \hat{\beta} J' \hat{C} = \hat{\beta} J' \hat{\psi} \ .$$

From a consideration of the diagonal elements of the above relation it is then apparent that the diagonal elements of $\hat{C}$ equal the diagonal elements of A, since the diagonal elements of $\hat{\beta} J'$ are zero.

Equation (1) is not suitable for computation in its present form; therefore, it will be advantageous to derive a more convenient expression for the likelihood equations. To this end, the first equation of (1) is postmultiplied by $\hat{C} A^{-1} \hat{C}$ to obtain

$$\hat{F} \hat{\beta}' = \hat{F} \hat{\beta}' A^{-1} \hat{C} + J' \hat{\psi} A^{-1} \hat{C} \quad ,$$

since $J' \hat{\beta} = 0$. In the above equation $\hat{\psi} + \hat{\beta} \hat{F} \hat{\beta}'$ is substituted for $\hat{C}$ and the resulting equation is

$$(3) \qquad \hat{F} \hat{\beta}' = (\hat{F} \hat{\beta}' + J' \hat{\psi})A^{-1} \hat{\psi} + (\hat{F} \hat{\beta}' + J' \hat{\psi})A^{-1} \hat{\beta} \hat{F} \hat{\beta}' \quad .$$

If equation (3) is postmultiplied by $\hat{\psi}^{-1} \hat{\beta}$, one obtains after some simplification

$$(4) \qquad (\hat{F} \hat{\beta} + J' \hat{\psi})A^{-1} \hat{\beta} = \hat{F} \hat{\beta}' \hat{\psi}^{-1} \hat{\beta} (I + \hat{F} \hat{\beta}' \hat{\psi}^{-1} \hat{\beta}) \quad .$$

By postmultiplying equation (3) by $\hat{\psi}^{-1} A$ and substituting for $(\hat{F} \hat{\beta} + J' \hat{\psi})A^{-1} \hat{\beta}$ the expression in (4), after a little manipulation one obtains

$$\hat{F} \hat{\beta}' \hat{\psi}^{-1} A = (I + \hat{F} \hat{\beta}' \hat{\psi}^{-1} \hat{\beta}) (\hat{F} \hat{\beta}' + J' \hat{\psi}) \quad .$$

Hence, the likelihood equations may also be written

$$\hat{F}\,\hat{\beta}'\,\hat{\psi}^{-1}\,A = (I + \hat{F}\,\hat{\beta}'\,\hat{\psi}^{-1}\,\hat{\beta})(\hat{F}\,\hat{\beta}' + J'\,\hat{\psi})$$

$$(5) \qquad J'\,\hat{\beta} = 0$$

$$\text{Diagonal } (A - \hat{C}) = 0 \quad ,$$

where J is defined as before. The equations above were derived by Anderson and Rubin in their joint paper to be published in the Third Berkeley Symposium [1] . They have also shown that $\hat{\beta}$ is independent of the scale of measurement in the same sense as before.

The three special models discussed in Section 4.3 will carry the same designation in the case of oblique factors; that is, Models I, II, and III. Here the corresponding regression matrices β will have the same form, but the common factors will be assumed correlated.

5.3 Indeterminancy for Oblique Factors

For the model with oblique factors, $C = \psi + \beta\,F\,\beta'$, it has been shown that β is uniquely determined except for multiplication on the right by an $m \times m$ non-singular matrix D such that $D\,D'$ has ones in the diagonal. In this case the matrix D has $m(m - 1)$ independent elements and again a particular matrix D can be determined such that the following conditions are satisfied: each column of $\beta\,D$ has $m - 1$ zero elements and $\beta^{(\alpha)}$ $(\alpha = 1, 2, \ldots, m)$, as defined

before in 4.4, has rank $m - 1$. Here there are $m(m - 1)$ zeros in all. This can be shown by repeated application of the following theorem: if $|G|_{m \times m} \neq 0$, there exists a non-singular matrix D such that $G D = H$, where H is a diagonal matrix. As before it is clear that the above result still holds true if β is replaced by $\hat{\beta}$.

Therefore, certain simple structure hypotheses are again equivalent to the general model. For the oblique case hypotheses of this type are specified as follows: adding zeros to the various columns of the hypothesized β matrix results in a β matrix satisfying the conditions given above. A simple structure hypothesis of this type may be tested by applying the methods used in the general case; there is no necessity for using the methods outlined in this chapter.

It is again possible that simple structure hypotheses do not define β uniquely. Therefore, it may be possible to rotate even under simple structure. For example, a $p \times 2$ β matrix with three zeros in the first column and none in the second is not uniquely determined, for a matrix of the same form results from postmultiplying β by a matrix D of the following form:

$$\begin{bmatrix} a & \sqrt{1 - a^2} \\ 0 & 1 \end{bmatrix} .$$

If there are at least $m - 1$ zeros in every column of β and $\beta^{(\alpha)}$ $(\alpha = 1, 2, \ldots, m)$ has rank $m - 1$, then β and $\hat{\beta}$ are uniquely determined. The proof follows from the fact that D, under these conditions, must be the identity matrix to leave the form of β unchanged.

If the conditions are not satisfied, β and $\hat{\beta}$ are not uniquely determined and additional conditions must be imposed to obtain a unique determination. Thus, in the example given above the added restriction that $F = I$ determines a unique solution, but for other simple structure hypotheses it may be necessary to impose other conditions as well, such as $\hat{\beta}' \hat{\psi}^{-1} \hat{\beta}$ = a diagonal matrix. The uniqueness of the β matrix under the particular simple structure hypothesis should be checked by the factor analyst who may then determine the added restrictions that insure uniqueness.

The number of restrictions necessary to obtain uniqueness affects the distribution of the test statistic; the effect will be discussed further in Chapter VII.

5.4 Solution of the Maximum Likelihood Equations

If the first equation of (5) in this chapter is premultiplied by $\hat{F}^{-1}$ and postmultiplied by $\hat{\psi}^{-1} \hat{\beta}$, the result is

$$\hat{\beta}' \hat{\psi}^{-1} A \hat{\psi}^{-1} \hat{\beta} = \hat{\beta}' \hat{\psi}^{-1} \hat{\beta} + \hat{\beta}' \hat{\psi}^{-1} \hat{\beta} \hat{F} \hat{\beta}' \hat{\psi}^{-1} \hat{\beta} \quad ,$$

since $J' \hat{\beta} = 0$. This implies

$$(6) \qquad \hat{F} = (\hat{\beta}' \hat{\psi}^{-1} \hat{\beta})^{-1} \left[\hat{\beta}' \hat{\psi}^{-1} A \hat{\psi}^{-1} \hat{\beta}\right] (\hat{\beta}' \hat{\psi}^{-1} \hat{\beta})^{-1} - (\hat{\beta}' \hat{\psi}^{-1} \hat{\beta})^{-1} .$$

There is a rather striking resemblance between (6) and the equation defining W in Chapter IV. As a matter of fact, F in the computation for oblique factors plays almost the same role as W in the computation for orthogonal factors. The likelihood equations may now be written

$$\hat{\beta}' \hat{\psi}^{-1} A = (\hat{F}^{-1} + \hat{\beta}' \hat{\psi}^{-1} \hat{\beta})(\hat{F} \hat{\beta}' + J' \hat{\psi})$$

$$(7) \qquad \hat{F} = (\hat{\beta}' \hat{\psi}^{-1} \hat{\beta})^{-1} \left[\hat{\beta}' \hat{\psi}^{-1} A \hat{\psi}^{-1} \hat{\beta}\right] (\hat{\beta}' \hat{\psi}^{-1} \hat{\beta})^{-1} - (\hat{\beta}' \hat{\psi}^{-1} \hat{\beta})^{-1}$$

$$\hat{\psi} = \text{Diagonal } (A - \hat{\beta} \hat{F} \hat{\beta}') \quad .$$

If $\beta^{*\prime} = \hat{F} \hat{\beta}' + J' \hat{\psi}$, then $\hat{\psi} = \text{diagonal } (A - \hat{\beta} \beta^{*\prime})$, since $\hat{\beta} J'$ has zeros in the diagonal.

Actual computation starts with an initial estimate of β and F , $\hat{\beta}^{(1)}$ and $\hat{F}^{(1)}$, obtained as outlined in Section 5.1. Then $\hat{\psi}^{(1)}$ is obtained by subtracting from the diagonal elements of A the diagonal elements of $\hat{\beta}^{(1)}, \hat{F}^{(1)} \hat{\beta}^{(1)}$. The next step is to compute

$$\beta^{*(1)\prime} = \left[\hat{F}^{(1)-1} + \hat{\beta}^{(1)\prime} \hat{\psi}^{(1)-1} \hat{\beta}^{(1)}\right]^{-1} \hat{\beta}^{(1)\prime} \hat{\psi}^{(1)-1} A \quad .$$

Now, exactly as in Section 4.4, the form of β^* is examined to determine what submatrices of $\hat{F}^{(1)}$ should be inverted to obtain $\hat{\beta}^{(2)}$ from $\beta^{*(1)}$. After $\hat{\beta}^{(2)}$ is determined, $\hat{\psi}^{(2)}$ is calculated by subtracting from the diagonal elements of A the diagonal elements of $\hat{\beta}^{(2)}\,\beta^{*(1)\prime}$. One then computes $\hat{F}^{(2)}$ with the aid of the second equation of (7), substituting $\hat{\beta}^{(2)}$ and $\hat{\psi}^{(2)}$ for $\hat{\beta}$ and $\hat{\psi}$ in the expression on the right. At this point a procedure slightly different from that outlined above, is used to compute $\beta^{*(2)}$; namely,

$$\beta^{*(2)\prime} = \left[\hat{\beta}^{(2)\prime}\,\hat{\psi}^{(2)-1}\,\hat{\beta}^{(2)}\right]^{-1}\left[\hat{\beta}^{(2)\prime}\,\hat{\psi}^{(2)-1}\,A-\hat{F}^{(2)-1}\,\beta^{*(1)\prime}\right] .$$

One then obtains $\hat{\beta}^{(3)}$ in the same manner as before. The procedure is repeated until $\hat{\beta}^{(n)}$ converges to $\hat{\beta}$. It is evident that this procedure is very similar to that discussed in the preceding chapter, F taking the role of W. The numerical example in that chapter also illustrates the computational method for oblique factors.

For certain hypotheses some simplification is possible. For example, for the analogue of Model I, $\beta'\,\psi^{-1}\,\beta$ is a diagonal matrix and $\psi = \text{Diagonal}\,(A - \beta\,\beta')$. This leads to much simpler equations and hence much less complicated computations.

It was mentioned at the end of the preceding chapter that one can derive a Gauss-Seidel type of iterative scheme to obtain the solution of the general maximum likelihood equations derived in that chapter. This is also the case with oblique factors. However, the same objections

apply here as did in the orthogonal case, and the scheme is not generally recommended. If certain simplifications result from the particular hypothesis, say $\beta' \psi^{-1} \beta$ is a diagonal matrix, then the method may profitably be employed. Starting with equation (5) one can derive the following general formula to be used with this method:

$$\hat{\beta}'_{(j)} \hat{\psi}^{-1} A_j = \left[\hat{\beta}'_{(j)} \hat{\psi}^{-1} \hat{\beta}_{(j)}\right] \hat{F}\hat{\beta}'_j + \left[\hat{F}^{-1} + \hat{\beta}'_{(j)} \hat{\psi}^{-1} \hat{\beta}_{(j)}\right] Q_j \, .$$

$\hat{\beta}_{(j)}$, $\hat{\beta}_j$, and A_j are defined as in Section 4.5 and Q_j is the <u>jth</u> column of $J' \hat{\psi}$. This method can be used advantageously with either Model I or Model II. For Model I, with $m = 2$ say, β' has the following form:

$$\begin{array}{c} \quad q \qquad\qquad\qquad p-q \\ \left[\begin{array}{cccc|cccc} x & x & \dots & x & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & x & x & \dots & x \end{array}\right] \end{array} = \begin{array}{c} \quad q \qquad p-q \\ \left[\begin{array}{c|c} \beta'_{11} & 0 \\ 0 & \beta'_{22} \end{array}\right] \end{array} ,$$

where x denotes some non-zero number. Therefore,

$$\hat{\psi} = \text{Diagonal } (A - \hat{\beta}\,\hat{\beta}')$$

$$\hat{\beta}' \hat{\psi}^{-1} \hat{\beta} = \text{Diagonal matrix.}$$

If the elements of $\hat{\beta}$ are d_{ij} and the elements of Q' are q_{ij} $(i = 1, 2, \ldots, p\ ;\ j = 1, 2, \ldots, m)$, the equations for $i = 1, 2, \ldots, q$ may be written

$$d_{i1} \sum_{\substack{j=1 \\ j\neq i}}^{q} \frac{d_{j1}^2}{\hat{\psi}_{jj}} + q_{i2}\, \hat{f}^{12} = \sum_{\substack{j=1 \\ j\neq i}}^{q} \frac{d_{j1}\, a_{ij}}{\hat{\psi}_{jj}}$$

$$d_{i1}\, \hat{f}_{12} \sum_{j=q+1}^{p} \frac{d_{j2}^2}{\hat{\psi}_{jj}} + q_{i2} \left(\hat{f}^{22} + \sum_{j=q+1}^{p} \frac{d_{j2}^2}{\hat{\psi}_{jj}}\right) = \sum_{j=q+1}^{p} \frac{d_{j2}\, a_{ij}}{\hat{\psi}_{jj}} \quad ,$$

where the $\hat{f}^{ij}$ are the elements of $\hat{F}^{-1}$. Similarly, for $i = q+1, q+2, \ldots, p$, the equations are

$$q_{i1} \left(\hat{f}^{11} + \sum_{j=1}^{q} \frac{d_{j1}^2}{\hat{\psi}_{jj}}\right) + d_{i2}\, \hat{f}_{12} \sum_{j=1}^{q} \frac{d_{j1}^2}{\hat{\psi}_{jj}} = \sum_{j=1}^{q} \frac{d_{j1}\, a_{ij}}{\hat{\psi}_{jj}}$$

$$q_{i1}\, \hat{f}^{12} + d_{i2} \sum_{\substack{j=q+1 \\ j\neq i}}^{p} \frac{d_{j2}^2}{\hat{\psi}_{jj}} = \sum_{\substack{j=q+1 \\ j\neq i}}^{p} \frac{d_{j2}\, a_{ij}}{\hat{\psi}_{jj}} \quad .$$

Starting with an initial approximation, $\hat{\beta}^{(1)}$ to $\hat{\beta}$, one calculates $\hat{\psi}^{(1)}$ by subtracting from the diagonal elements of A the diagonal elements of $\hat{\beta}^{(1)}\hat{\beta}^{(1)\prime}$. $\hat{F}^{(1)}$ is then calculated with the aid of the second equation of (7). In this case the equation reduces to

$$\hat{f}_{12} = \frac{\hat{\beta}'_{11}\,\hat{\psi}^{-1}\,A\,\hat{\psi}^{-1}\,\hat{\beta}_{22}}{\left[\hat{\beta}'_{11}\,\hat{\psi}^{-1}\,\hat{\beta}_{11}\right]\left[\hat{\beta}'_{22}\,\hat{\psi}^{-1}\,\hat{\beta}_{22}\right]} .$$

The computation then proceeds exactly as it is outlined in the previous Gauss-Seidel calculations. However, at each step a new estimate of F should be computed; that is, an estimate of F should be computed with each new estimate of a row of β, not merely with each complete iteration. The method seems to converge rather quickly, and the computations are not as bad as one might gather from a quick look at the equations.

As a numerical example the following artificially constructed 5×5 correlation matrix will be considered:

$$\begin{bmatrix} 1.00 & .43 & .50 & .35 & .30 \\ & 1.00 & .56 & .40 & .37 \\ & & 1.00 & .44 & .41 \\ & & & 1.00 & .58 \\ & & & & 1.00 \end{bmatrix} .$$

Then, as an initial approximation, one takes

$$\hat{\beta}^{(1)\prime} = \begin{bmatrix} .6 & .7 & .8 & 0 & 0 \\ 0 & 0 & 0 & .8 & .7 \end{bmatrix} .$$

In this case $q = 3$, $m = 2$, and $p = 5$. From $\hat{\beta}^{(1)\prime}$ the following matrices are computed:

$$\hat{\psi}^{(1)} = \begin{bmatrix} .64 & & & & \\ & .51 & & & \\ & & .36 & & \\ & & & .36 & \\ & & & & .51 \end{bmatrix} ;$$

$$\hat{\beta}^{(1)\prime} \hat{\psi}^{(1)-1} = \begin{bmatrix} .9375 & 1.3725 & 2.2222 & 0 & 0 \\ 0 & 0 & 0 & 2.2222 & 1.3725 \end{bmatrix} ;$$

$$\hat{\beta}^{(1)\prime} \hat{\psi}^{(1)-1} \hat{\beta}^{(1)} = \begin{bmatrix} 3.3010 & 0 \\ 0 & 2.7386 \end{bmatrix} .$$

In order to compute $\hat{F}^{(1)}$, it is necessary to multiply the first row of $\hat{\beta}^{(1)\prime} \hat{\psi}^{(1)-1}$ by the last two columns of A . The result is a 1×2 vector, $\begin{bmatrix} 1.8549 & 1.7002 \end{bmatrix}$. Then

$$\hat{f}_{12}^{(1)} = \frac{(1.8549)(2.2222) + (1.7002)(1.3725)}{(3.3010)(2.7386)} = .7141.$$

The quantities needed for a second estimate of d_{11} are

$$\sum_{\substack{j=1 \\ j\neq 1}}^{3} \frac{d_{j1}^2}{\hat{\psi}_{jj}} = .7(1.3725) + .8(2.2222) = 2.7386 \quad ;$$

$$\sum_{j=4}^{5} \frac{d_{j2}^2}{\hat{\psi}_{jj}} = .8(2.2222) + .7(1.3725) = 2.7386 \quad ;$$

$$\sum_{\substack{j=1 \\ j\neq 1}}^{3} \frac{d_{j1}\, a_{j1}}{\hat{\psi}_{jj}} = .43(1.3725) + .50(2.2222) = 1.7013 \quad ;$$

$$\sum_{j=4}^{5} \frac{d_{j2}\, a_{j1}}{\hat{\psi}_{jj}} = .35(2.2222) + .30(1.3725) = 1.1895 \quad ;$$

$$\hat{f}^{12(1)} = -1.4572 \quad ;$$

$$\hat{f}^{22(1)} = 2.0406 \quad .$$

Then

$$\begin{bmatrix} 2.7386 & -1.4572 \\ 1.9556 & 4.7792 \end{bmatrix} \begin{bmatrix} d_{11}^{(2)} \\ q_{12}^{(2)} \end{bmatrix} = \begin{bmatrix} 1.7013 \\ 1.1895 \end{bmatrix} ,$$

or $d_{11}^{(2)} = .6189$. The following matrices are then needed for the next step:

$$\hat{\beta}^{(2)\prime} = \begin{bmatrix} .6189 & .7000 & .8000 & 0 & 0 \\ 0 & 0 & 0 & .8000 & .7000 \end{bmatrix} ;$$

$$\hat{\Psi}^{(2)} = \begin{bmatrix} .6170 & & & & \\ & .5100 & & & \\ & & .3600 & & \\ & & & .3600 & \\ & & & & .5100 \end{bmatrix} ;$$

$$\hat{\beta}^{(2)\prime} \hat{\Psi}^{(2)-1} = \begin{bmatrix} 1.0031 & 1.3725 & 2.2222 & 0 & 0 \\ 0 & 0 & 0 & 2.2222 & 1.3725 \end{bmatrix} ;$$

$$\hat{\beta}^{(2)\prime}\,\hat{\psi}^{(2)-1}\,\hat{\beta}^{(2)} = \begin{bmatrix} 3.3593 & 0 \\ 0 & 2.7386 \end{bmatrix} .$$

Hence, $\hat{f}_{12}^{(2)} = .7102$, and the resulting equation for $d_{21}^{(2)}$ is

$$\begin{bmatrix} 2.3986 & -1.4330 \\ 1.9449 & 4.7562 \end{bmatrix} \begin{bmatrix} d_{21}^{(2)} \\ d_{22}^{(2)} \end{bmatrix} = \begin{bmatrix} 1.6758 \\ 1.3967 \end{bmatrix} .$$

The process is then continued until $\hat{\beta}^{(n)\prime}$ converges to

$$\hat{\beta}' = \begin{bmatrix} .6190 & .7032 & .7987 & 0 & 0 \\ 0 & 0 & 0 & .7958 & .7288 \end{bmatrix} ,$$

and $\hat{f}_{12}^{(n)}$ converges to $\hat{f}_{12} = .7022$.

This method, however, is not recommended unless ψ = Diagonal $(A - \beta\beta')$. Otherwise, as mentioned before, the elements of $\hat{\psi}$ depend on all the elements of $\hat{\beta}$, not just on those of the corresponding row. For Model II it can be shown that this condition is satisfied; therefore, the Gauss-Seidel method has certain advantages in this case also.

In conclusion, it should be noted that the principal purpose of Chapters IV and V is to illustrate the manner in which the factor analyst should proceed to test his hypothesis. The purpose is not to give a step-by-step computing procedure for every possible hypothesis, but rather to indicate the general method of approach to the problem.

CHAPTER VI

PREDICTION

6.1 Prediction of Y from X

Bartlett [2], Thomson [31], and Lawley [21] have all considered this problem. However, it will be advantageous to arrive at the prediction equations in a different manner. Now X and Y, where $E(X) = 0 = E(Y)$, have a joint multivariate distribution. The linear mean square regression of Y on X is specified by $\gamma_{m \times p}$, where

$$\gamma\, E(X\, X') = E(Y\, X') \quad .$$

Then $\gamma\, X$ is taken as an estimate of Y.

In the actual factor analysis problem neither Y nor $E(X\, X')$ is known. However, under the hypothesis proposed in this paper, estimates of both $E(X\, X')$ and $E(Y\, X')$ are available. These are $\hat{C}$ and $\hat{\lambda}$ respectively. Therefore, as an estimate of γ, one takes

$$\hat{\gamma} = \hat{\lambda}\, \hat{C}^{-1} \quad .$$

Then if X^* is the $p \times N$ matrix of sample values of X,

$$(1) \qquad Y^*_{m \times N} = \hat{\lambda}\,\hat{C}^{-1} X^* \quad ,$$

say, is the estimate of the common factor values for each individual.

The population regression sum of squares matrix is $\lambda\, C^{-1}\, \lambda'$, and therefore, the estimated residual covariance matrix of Y is $\left[\hat{F} - \hat{\lambda}\,\hat{C}^{-1}\,\hat{\lambda}'\right]$, where $\hat{F}$ is the estimate of $F = E(Y\,Y')$. This residual matrix indicates how well the common factors are predicted by a linear regression on X. However, in the simple structure case the matrix $\beta = \lambda'\, F^{-1}$ has been estimated, and not λ itself. Hence, the equation for Y^* given in terms of $\hat{\beta}$ is

$$(2) \qquad Y^* = \hat{F}\,\hat{\beta}'\,\hat{C}^{-1} X^* \quad ,$$

and the estimated regression sum of squares matrix is $\hat{F}\,\hat{\beta}'\,\hat{C}^{-1}\,\hat{\beta}\,\hat{F}$.

For the general case, where no zeros are assumed in the β matrix, these equations can be written such that only the inverse of an $m \times m$ matrix is involved, not the inverse of the $p \times p$ matrix, $\hat{C}$. The likelihood equations for the general case are

$$\hat{\lambda}\,\hat{C}^{-1} A = \hat{\lambda} \quad ,$$

where $\hat{C} = \hat{\psi} + \hat{\lambda}' \hat{\lambda}$. Hence,

$$\hat{\lambda} \hat{C}^{-1} = \hat{\lambda} A^{-1} \quad ,$$

and

$$\hat{\lambda} = \hat{\lambda} A^{-1} \hat{\psi} + \hat{\lambda} A^{-1} \hat{\lambda}' \hat{\lambda} \quad .$$

Then, in exactly the same manner as in the preceding chapters, it can be shown that

$$\hat{\lambda} A^{-1} \hat{\lambda}' = \hat{\lambda} \hat{\psi}^{-1} \hat{\lambda}' \left[I + \hat{\lambda} \hat{\psi}^{-1} \hat{\lambda}' \right]^{-1} \quad .$$

Therefore,

$$\hat{\lambda} A^{-1} = \left[I + \hat{\lambda} \hat{\psi}^{-1} \hat{\lambda}' \right]^{-1} \hat{\lambda} \hat{\psi}^{-1} = \hat{\lambda} \hat{C}^{-1} \quad . \tag{3}$$

This implies that

$$Y^* = \left[I + \hat{\lambda} \hat{\psi}^{-1} \hat{\lambda}' \right]^{-1} \hat{\lambda} \hat{\psi}^{-1} X^* \quad . \tag{4}$$

Thomson [31] and Lawley [21] have also derived this particular prediction equation, which is evidently much better suited for computation. Furthermore, by postmultiplying equation (3) by $\hat{\lambda}'$, one obtains

$$\hat{\lambda}\,\hat{C}^{-1}\,\hat{\lambda}' = \left[I + \hat{\lambda}\,\hat{\psi}^{-1}\,\hat{\lambda}'\right]^{-1}\hat{\lambda}\,\hat{\psi}^{-1}\,\hat{\lambda}' \quad ,$$

the estimated regression sum of squares matrix.

In the simple structure case the equations can also be simplified in this manner. From equations (1) and (3) of Chapter IV

$$\hat{\beta}'\,\hat{C}^{-1} = \hat{\beta}^{*\prime}\,A^{-1} = \left[I + \hat{\beta}'\,\hat{\psi}^{-1}\,\hat{\beta}\right]^{-1}\hat{\beta}'\,\hat{\psi}^{-1} \quad .$$

Since $\hat{\beta}' = \hat{\lambda}$ for orthogonal factors, the prediction equation for simple structure with orthogonal factors is

$$Y^* = \left[I + \hat{\lambda}\,\hat{\psi}^{-1}\,\hat{\lambda}'\right]^{-1}\hat{\lambda}\,\hat{\psi}^{-1}\,X^* \quad .$$

Similarly, from equations (1) and (5) of Chapter V

$$\hat{F}\,\hat{\beta}'\,\hat{C}^{-1} = \left[J'\hat{\psi} + \hat{F}\,\hat{\beta}'\right]A^{-1} = \left[I + \hat{F}\,\hat{\beta}'\,\hat{\psi}^{-1}\,\hat{\beta}\right]^{-1}\hat{F}\,\hat{\beta}'\,\hat{\psi}^{-1} \quad .$$

Therefore, from equation (2) of this chapter

$$Y^* = \left[I + \hat{F}\,\hat{\beta}'\,\hat{\psi}^{-1}\,\hat{\beta}\right]^{-1}\hat{F}\,\hat{\beta}'\,\hat{\psi}^{-1}\,X^* \quad ,$$

or in terms of $\hat{\lambda}$,

$$Y = \left[I + \hat{\lambda}\,\hat{\psi}^{-1}\,\hat{\lambda}'\,\hat{F}^{-1}\right]^{-1}\hat{\lambda}\,\hat{\psi}^{-1}\,X^* \quad .$$

The estimated regression sum of squares matrices for orthogonal and oblique factors respectively are

$$\left[I + \hat{\lambda}\,\hat{\psi}^{-1}\,\hat{\lambda}'\right]^{-1} \hat{\lambda}\,\hat{\psi}^{-1}\,\hat{\lambda}' \quad ;$$

$$\hat{F}\left[\hat{F} + \hat{\lambda}\,\hat{\psi}^{-1}\,\hat{\lambda}'\right]^{-1} \hat{\lambda}\,\hat{\psi}^{-1}\,\hat{\lambda}' = \left[\hat{F} + \hat{\beta}'\,\hat{\psi}^{-1}\,\hat{\beta}\right]^{-1} \hat{\beta}'\,\hat{\psi}^{-1}\,\hat{\beta}\,\hat{F} \quad .$$

All of the foregoing equations are of necessity somewhat arbitrary, since the y's are not actually known. However, they do give some degree of information, and the prediction of Y from X is usually considered as one of the objectives of factor analysis.

6.2 Non-Linearity and Monotonicity

In order to investigate non-linear properties, Y^* should be computed. Then the factor analyst can look for relations among the various common factors. For example, if $m = 2$, and the model actually involves only y_1 and y_1^2, say, then a plot of y_1^* against y_2^* should reveal the relationship. Here y_1^* and y_2^* are the first and second rows of Y^* respectively, and are normally distributed, since they are linear functions of normal variates. In this scheme one would actually be plotting $ay_1 + by_1^2$ against $cy_1 + dy_1^2$, but a definite relationship should show up if one really exists. If there is no relationship of this kind between the common factors, the y*'s should plot as a random scatter in an ellipse, since the y*'s are normal and may be correlated. The closeness of the functional relationship could

be checked with the residual covariance matrix of Y .

As far as the author is aware, no investigation of this type has been undertaken, but it seems to be a promising line of attack on the joint problem of non-linearity and monotonicity.

6.3 Prediction of the Factor Loadings

In factor analysis terminology this section is concerned with predicting the factor loadings on a new test when it is added to the battery, without going through the whole estimation procedure a second time. This can be accomplished by using the Gauss-Seidel technique outlined in Chapter III. As an example, suppose a test is added to the eight test battery discussed in Chapter III. The new test has the following correlations with the original eight tests:

1	2	3	4	5	6	7	8
.600	.150	.360	.550	.500	.300	.600	.580

.

The estimate of λ obtained for the original eight tests is

$$\hat{\lambda} = \begin{bmatrix} .706 & .515 & .731 & .648 & .612 & .394 & .711 & .663 \\ .240 & -.176 & -.471 & .161 & .139 & .069 & .183 & .344 \end{bmatrix} .$$

Then

$$\hat{\lambda}\,\hat{\psi}^{-1} = \begin{bmatrix} 1.5902 & .7317 & 2.9984 & 1.1693 & 1.0097 & .4690 & 1.5423 & 1.4997 \\ .5406 & -.2501 & -1.9319 & .2905 & .2293 & .0821 & .3970 & .7781 \end{bmatrix}.$$

The following quantities are also computed:

$$\sum_{\substack{i=1 \\ i\neq 9}}^{9} \frac{\lambda_{1i}^2}{\psi_{ii}} = .706(1.5902) + .515(.7317) + \ldots + .663(1.4997) = 7.343 \;;$$

$$\sum_{\substack{i=1 \\ i\neq 9}}^{9} \frac{\lambda_{1i}\,\lambda_{2i}}{\psi_{ii}} = .240(1.5902) - .176(.7317) + \ldots + .344(1.4997) = 0 \;;$$

$$\sum_{\substack{i=1 \\ i\neq 9}}^{9} \frac{\lambda_{2i}^2}{\psi_{ii}} = .240(.5406) - .176(-.2501) + \ldots + .344(.7781) = 1.508 \;;$$

$$\sum_{\substack{i=1 \\ i\neq 9}}^{9} \frac{\lambda_{1i}\,a_{i9}}{\psi_{ii}} = .600(1.5902) + .150(.7317) + \ldots + .580(1.4997) = 5.227 \;;$$

$$\sum_{\substack{i=1 \\ i\neq 9}}^{9} \frac{\lambda_{2i}\,a_{i9}}{\psi_{ii}} = .600(.5406) + .150(-.2501) + \ldots + .580(.7781) = .580 \;.$$

Therefore,

$$\begin{bmatrix} 7.343 & 0 \\ 0 & 1.508 \end{bmatrix} \begin{bmatrix} \lambda_{19} \\ \lambda_{29} \end{bmatrix} = \begin{bmatrix} 5.227 \\ .580 \end{bmatrix},$$

or $\lambda_{19} = .712$ and $\lambda_{29} = .385$. These two quantities are the desired estimates.

CHAPTER VII

TESTING

In the preceding chapters it has been assumed that m is known a priori; however, in practice this will not always be the case. The first section of this chapter will consider the testing problem when m is known, while Section 7.2 will consider the problem when m also must be estimated. In the last section certain related topics will be discussed.

7.1 Test of the Fit of the Model

In the general case the null hypothesis is that the population covariance matrix, C, equals $\psi + \beta\beta'$, where ψ is a diagonal matrix and β is a $p \times m$ matrix, $m < p$, while the alternative hypothesis is that C is any positive definite matrix. One possible test statistic is the likelihood ratio criterion. The likelihood function for Wishart's distribution is

$$K\,|C|^{-\frac{N-1}{2}}\,|A|^{\frac{N-p-2}{2}}\,e^{-\frac{N-1}{2}\,\mathrm{Trace}\,A\,C^{-1}} .$$

Under the null and alternative hypotheses respectively, $\hat{C} = \hat{\psi} + \hat{\beta}\hat{\beta}'$ and $\hat{C} = A$. Therefore, the likelihood ratio criterion is

$$\frac{|A|^{\frac{N-1}{2}} \; e^{\frac{p(N-1)}{2}}}{|\hat{\psi} + \hat{\beta}\,\hat{\beta}'|^{\frac{N-1}{2}} \; e^{\frac{N-1}{2}\,\mathrm{Trace}\,A(\hat{\psi} + \hat{\beta}\,\hat{\beta}')^{-1}}} \; .$$

It has been shown in Section 2.4 that the diagonal elements of $A(\hat{\psi} + \hat{\beta}\,\hat{\beta}')^{-1}$ are ones and hence, the likelihood ratio criterion may be written as follows:

$$L_m = \frac{|A|^{\frac{N-1}{2}}}{|\hat{\psi} + \hat{\beta}\,\hat{\beta}'|^{\frac{N-1}{2}}} \; .$$

Under certain conditions $T_m = -2 \log_e L_m$ is asymptotically distributed as χ^2 with $\frac{p(p-1)}{2} - pm + \frac{m(m-1)}{2}$ degrees of freedom when the null hypothesis is true. These conditions have been determined by Anderson and Rubin [1] and will be discussed in Section 7.3 in connection with the asymptotic normality of $\hat{\beta}$ and $\hat{\psi}$. The test procedure itself is then to reject the hypothesis if T_m is greater than some preassigned quantity which is chosen to give the desired probability level.

The likelihood ratio criterion for simple structure hypotheses may be derived in a similar manner, whether with orthogonal or oblique factors, and is identical to L_m, since the diagonal elements of $A(\hat{\psi} + \hat{\beta}\,\hat{\beta}')^{-1}$ and $A(\hat{\psi} + \hat{\beta}\,\hat{F}\,\hat{\beta}')^{-1}$ are still unity. Again $-2 \log_e L_m$

is asymptotically distributed as χ^2 subject to conditions similar to those in the general case. However, the degrees of freedom associated with the χ^2 vary, depending on the particular simple structure hypothesis. If β is uniquely determined as discussed in Section 4.4, then for the orthogonal case the degrees of freedom are $\frac{p(p-1)}{2} - pm$ plus the number of zeros specified in β. If β is not uniquely determined but is of the type discussed in the second paragraph of 4.4, the degrees of freedom associated with the asymptotic χ^2 are $\frac{p(p-1)}{2} - pm + \frac{m(m-1)}{2}$. Therefore, a simple structure hypothesis of this form is equivalent in all respects to the general case; testing and estimation procedures are exactly the same. However, for other types of hypotheses which do not determine β uniquely these formulas no longer apply. Thus, in the example given in that Section, 4.4, the degrees of freedom are $\frac{p(p-1)}{2} - 3p + 3 + 1$. The three is the number of zeros specified, while the 1 is necessary because of the remaining freedom to rotate. In situations of this kind the factor analyst is obliged to determine the degrees of freedom by examining the possible transformations which leave the form of β unchanged.

The same problem arises for oblique factors. If β is uniquely determined by the hypothesis (Section 5.3), then the degrees of freedom associated with the asymptotic χ^2 are $\frac{p(p-1)}{2} - pm - \frac{m(m-1)}{2}$ plus the number of zeros specified in β. On the other hand, if β is of the form specified in the second paragraph of 5.3, the degrees of

freedom are $\frac{p(p-1)}{2} - pm + \frac{m(m-1)}{2}$. Therefore, any simple structure hypothesis specifying correlated factors and a β of this form, is equivalent to the general case, and may be tested by the methods in Chapters II and III and in the first part of this section. For other types of hypotheses that do not determine β uniquely the remarks concerning orthogonal factors apply. Thus, the statistic for the example given in Section 5.3 has $\frac{p(p-1)}{2} - 2p = 1 + 3 + 1$ degrees of freedom, since there is one degree of freedom available for rotation.

An intuitive idea of the test can be given as follows: in Section 2.4 it has been shown that

$$\hat{\Psi}\,\hat{C}^{-1} A = A - \hat{C} + \hat{\Psi} = A - \hat{\beta}\,\hat{\beta}' .$$

This implies that the determinant of $\hat{C}^{-1} A$ equals the determinant of a matrix with ones in the diagonal and the following typical non-diagonal element:

$$\frac{a_{ij} - \sum_{k=1}^{m} d_{ik}\, d_{jk}}{\sqrt{\hat{\Psi}_{ii}\ \hat{\Psi}_{jj}}} ,$$

where d_{ik} $(i = 1, 2, \ldots, p;\ k = 1, 2, \ldots, m)$ are the elements of $\hat{\beta}$. Since $\frac{|A|}{|\hat{C}|}$ is independent of the scale of the p variables, one can as well use the sample correlation matrix, R , and the estimated

population correlation matrix, $\hat{P}$. Hence, if $\hat{P} = \hat{\Sigma} + \hat{\gamma}\,\hat{\gamma}'$, $\frac{|A|}{|\hat{C}|} = \frac{|R|}{|\hat{P}|}$ equals the determinant of the following symmetric matrix:

$$M = \begin{bmatrix} 1 & \dfrac{r_{12} - \sum_{k=1}^{m} \gamma_{1k}\,\gamma_{2k}}{\sqrt{\hat{\Sigma}_{11}\ \hat{\Sigma}_{22}}} & \cdots & \dfrac{r_{1p} - \sum_{k=1}^{m} \gamma_{1k}\,\gamma_{pk}}{\sqrt{\hat{\Sigma}_{11}\ \hat{\Sigma}_{pp}}} \\ & 1 & \cdots & \dfrac{r_{2p} - \sum_{k=1}^{m} \gamma_{2k}\,\gamma_{pk}}{\sqrt{\hat{\Sigma}_{22}\ \hat{\Sigma}_{pp}}} \\ & & & \vdots \\ & & & 1 \end{bmatrix},$$

where γ_{ik} $(i = 1, 2, \ldots, p;\ k = 1, 2, \ldots, m)$ are the elements of $\hat{\gamma}$. From the form of the non-diagonal elements of M it is evident that they are estimates of the population partial correlation coefficients among the p variables after the effect of the common factors, Y , has been removed. Therefore, $\frac{|A|}{|\hat{C}|} = \frac{|R|}{|\hat{P}|}$ may be thought of as the determinant of the estimated partial correlation matrix. This agrees with the approach that would be taken if the y's were actually known. The sample partial correlation matrix would be computed in this case and then checked to see if it were significantly different from the identity matrix. Quensel [27] has shown that under certain conditions, if a sample of N is taken from the joint distribution of X and Y $(X = \beta\, Y + G)$, the

distribution of the sample partial correlation coefficients among the x's after eliminating the y's, is the same as the distribution of the correlation coefficients in a sample of $N - m$ drawn from a multivariate normal population of independent variables. These conditions are

1. G has a multivariate normal distribution such that $E(G\,G')$ is a diagonal matrix.
2. Y and G are distributed independently of each other.

The moments of $|R|$ under this hypothesis are derived in Cramér's Mathematical Methods of Statistics [10] and Bartlett [4] has derived a test employing $-\left[N - 1 - \frac{2p + 5}{6}\right] \log_e |R|$ as a χ^2 with $\frac{p(p-1)}{2}$ degrees of freedom. Therefore, if the y's are known, no difficulty arises, since the same test can be utilized by merely replacing N by $N - m$ and $|R|$ by the determinant of the partial correlation matrix.

In factor analysis, the y's are, of course, unknown; however, for $m = 0$, T_m breaks down into $T_0 = -(N - 1) \log_e |R|$. T_m is only asymptotically distributed as χ^2 and for small sample sizes a different multiplying factor may make the actual distribution of T_m closer to that of χ^2. Therefore, Bartlett [4] recommends as a statistic

$$-\left[N - 1 - \frac{2p + 5}{6} - \frac{2m}{3}\right] \log_e \frac{|R|}{|\hat{P}|} \quad .$$

From the above discussion the author is inclined to prefer m rather than $\frac{2m}{3}$ in the multiplying constant. Either of these expressions reduces to Bartlett's other statistic for $m = 0$, and the difference between them is probably too small for concern.

For simple structure hypotheses it can be shown that $\frac{|A|}{|\hat{C}|}$ is again equivalent to this type of determinant. In the orthogonal factor case $\frac{|A|}{|\hat{C}|}$ equals the determinant of a matrix with ones in the diagonal and typical non-diagonal terms of the following form:

$$(1) \qquad \frac{r_{ij} - \sum_{k,n=1}^{m} \gamma_{ik} w_{kn} \gamma_{jn}}{\sqrt{\hat{\Sigma}_{ii} \quad \hat{\Sigma}_{jj}}} ,$$

where the w_{kn} are the elements of W as defined in Section 4.5. For oblique factors $\frac{|A|}{|\hat{C}|}$ is the determinant of a matrix with ones in the diagonal and typical non-diagonal terms of the following form:

$$(2) \qquad \frac{r_{ij} - \sum_{k,n=1}^{m} \gamma_{ik} \hat{f}_{kn} \gamma_{jn}}{\sqrt{\hat{\Sigma}_{ii} \quad \hat{\Sigma}_{jj}}} ,$$

where the $\hat{f}_{kn}$ are the elements of $\hat{F}$, the estimated covariance matrix of the common factors standardized to unit variances. Hence, for simple structure hypotheses one is essentially testing whether the matrix

of estimated partial correlations is significantly different from the identity matrix. Also, $-\left[N - 1 - \frac{2p + 5}{6} - m\right] \log_e \frac{|R|}{|\hat{P}|}$ is again recommended as the statistic to be used.

The computation of the test statistic itself is a rather tedious procedure, since the determinants of two $p \times p$ matrices must be calculated. However, some simplification is possible, because $\hat{P} = \hat{\Sigma} + \hat{\gamma}\,\hat{\gamma}'$ may be written in a form more suited for computation:

$$\left|\hat{\Sigma} + \hat{\gamma}\,\hat{\gamma}'\right| = \left|\hat{\Sigma}\right|\left|I + \hat{\gamma}\,\hat{\gamma}'\,\hat{\Sigma}^{-1}\right| = \left|\hat{\Sigma}\right|\left|I + \hat{\gamma}'\,\hat{\Sigma}^{-1}\,\hat{\gamma}\right| .$$

This follows from the fact that

$$\begin{array}{cc} & \begin{array}{cc} p & m \end{array} \\ \begin{array}{c} p \\ m \end{array} & \left|\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right| \end{array} = \left|D\right|\left|A - B\,D^{-1}\,C\right| = \left|A\right|\left|D - C\,A^{-1}\,B\right| ,$$

a result of Roy's [30], and therefore $|I + B\,C| = |I + C\,B|$. In a similar manner it can be shown that for oblique factors

$$\left|\hat{P}\right| = \left|\hat{\Sigma} + \hat{\gamma}\,\hat{F}\,\hat{\gamma}'\right| = \left|\hat{\Sigma}\right|\left|I + \hat{\gamma}\,\hat{F}\,\hat{\gamma}'\,\hat{\Sigma}^{-1}\right| = \left|\hat{\Sigma}\right|\left|I + \hat{F}\,\hat{\gamma}'\,\hat{\Sigma}^{-1}\,\hat{\gamma}\right| .$$

This saves some labor since the determinant of $\hat{\Sigma}$ is easy to compute and the other determinant involves only an $m \times m$ matrix, not a $p \times p$. Yet the computation of $|R|$ still may involve an excessive amount of computation. For this reason Lawley [20] has proposed an approximation

to this statistic for large N. He suggests $N \sum_{i<j} m_{ij}^2$ as an approximation to T_m, where the m_{ij} are the elements of the matrix M whose determinant is $\frac{|A|}{|\hat{C}|}$. One can sustain the approximation by observing that for large N all the non-diagonal elements are small if the null hypothesis is true. If the products of three or more non-diagonal elements are neglected, $|M| \cong 1 - \sum_{i<j} m_{ij}^2$ and $\log_e |M| \cong - \sum_{i<j} m_{ij}^2$ for $\sum_{i<j} m_{ij}^2$ sufficiently small. However, it would still seem advantageous to use $\left[N - 1 - \frac{2p+5}{6} - m\right]$ as a multiplying factor. The same approximation can also be derived for simple structure hypotheses, where M is defined by (1) and (2) for orthogonal and oblique factors respectively.

Another possible approximation is $(N - m - 3)$ $\sum_{i<j} \left[\frac{1}{2} \log_e \frac{1 + m_{ij}}{1 - m_{ij}}\right]^2$. This is equivalent to applying Fisher's z transformation to each of the m_{ij} and summing the squares of the z's. It seems that this is a much better approximation to T_m, especially when the null hypothesis is false. There are also some theoretical grounds for this approximation. If m_{ij} were a sample partial correlation coefficient, then under the null hypothesis $(N - m - 3)z_{ij}^2$ would be approximately distributed as a χ^2 with 1 degree of freedom. Therefore, the proposed statistic would be approximately distributed as χ^2 with $\frac{p(p-1)}{2}$ degrees of freedom, if the various z_{ij}'s were independent. For large N this is the case. Since the actual correlations are only estimated, one may subtract the number of

the quantities estimated from $\frac{p(p-1)}{2}$ and claim some sort of validity for the process. This is precisely the manner in which this approximation is constructed. No claims of asymptotic χ^2 distributions are advanced, but it nevertheless offers certain advantages over $\sum_{i<j} m_{ij}^2$, particularly as regards the power of the test.

Lawley [13] maintains that a sample size of 200 is sufficient for a close enough approximation to χ^2 and is also sufficient to permit use of the first approximation above. From sampling studies [14, 26] this would seem to be substantiated.

Numerical examples of the techniques discussed in this section are to be found in papers by Lawley [20, 22, 26] and Emmett [13].

7.2 Determination of m

In the preceding section it has been assumed that m is known a priori; for simple structure hypotheses this will be the case. However, in practice m must sometimes be estimated from the sample. This is accomplished by a sequential type of procedure. Quite naturally the first thing that should be done is to test the sample covariance (correlation) matrix to see if it departs significantly from a diagonal matrix. Thus, the null hypothesis is that the population correlation matrix is the identity matrix, and the alternative hypothesis is $m \geq 1$. If the test rejects this hypothesis, one next proceeds to test $m = 1$ against the alternative hypothesis $m \geq 2$, and so on. The procedure is terminated when the test accepts a null hypothesis and this gives an estimate of m. Hence, T_0 is computed and tested

for significance, then T_1, ..., and finally T_m. No probabilities can be assigned to the test as a whole, even asymptotically, although for each separate test it can be done. However, this may not be of too much concern, since the χ^2 will usually have an extremely small associated probability if the null hypothesis is false. A real drawback is the terrific amount of computation, for at each stage new estimates of the parameters must be calculated. In practice, it is advantageous to obtain some indication of m ahead of time, say by the centroid method. Then one can test this hypothesis, specifying m, using Lawley's method. Yet even here probabilities are altered somewhat, since a hypothesis has been made from the data and the same data is used to test it.

Nevertheless, with electronic digital computers and readily available codes for the method, this is no longer such a problem.

7.3 Asymptotic Variances and Covariances of the Maximum Likelihood Estimates

Anderson and Rubin [1] have considered this problem in some detail. They reach the following conclusion: "the variances and covariances of the elements of $\hat{\beta}$ are so complicated that they cannot be used for all the usual purposes." It is much easier to give the inverse matrix of these quantities. This inverse matrix is

$$\triangle = E\left[\frac{\partial \log L}{\partial \theta_j}\,\frac{\partial \log L}{\partial \theta_k}\right]_{\substack{\theta_j = \theta_{0j} \\ \theta_k = \theta_{0k}}},$$

where θ_j and θ_k are the parameters to be estimated and θ_{0j} and θ_{0k} are the values of the parameters under the null hypothesis. For the null hypothesis in the general case, $C = \psi + \beta\,\beta'$, this results in a $p(m + 1)$ matrix which must be inverted to obtain the matrix of asymptotic variances and covariances. In partitioned form

$$\Delta = \begin{array}{c} \\ p \\ p \\ \\ p \\ \\ p \\ \\ p \\ \\ p \end{array} \begin{array}{c} \begin{array}{ccccccc} p & p & & p & & p & & p \end{array} \\ \begin{bmatrix} A & A_1 & \cdots & A_i & \cdots & A_j & \cdots & A_m \\ A_1' & B_{11} & \cdots & B_{1i} & \cdots & B_{1j} & \cdots & B_{1m} \\ \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ A_i' & B_{1i}' & \cdots & B_{ii} & \cdots & B_{ij} & \cdots & B_{im} \\ \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ A_j' & B_{ij}' & \cdots & B_{ij}' & \cdots & B_{jj} & \cdots & B_{jm} \\ \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ A_m' & B_{im}' & \cdots & B_{im}' & \cdots & B_{jm}' & \cdots & B_{mm} \end{bmatrix} \end{array} .$$

A is the $p \times p$ matrix $E\left[\frac{\partial \log L}{\partial \psi_{ii}} \; \frac{\partial \log L}{\partial \psi_{jj}}\right]$ with typical element $\frac{N-1}{2}\left[c^{ij}\right]^2$, where the c^{ij} are the elements of C^{-1}. If the elements of the <u>ith</u> column of β are denoted by d_{ji} $(j = 1, 2, \ldots, p)$ and the column vector itself is denoted by β_i', then B_{ii} is the $p \times p$ matrix $E\left[\frac{\partial \log L}{\partial d_{ji}} \; \frac{\partial \log L}{\partial d_{ki}}\right]$ and $B_{ii} = (N-1)\left[(\beta_i\, C^{-1}\, \beta_i') + C^{-1}\, \beta_i'\, \beta_i\, C^{-1}\right]$. A_i is the $p \times p$ matrix $E\left[\frac{\partial \log L}{\partial \psi_{kk}} \; \frac{\partial \log L}{\partial d_{ji}}\right]$.

If C_j^{-1} denotes the jth column of C^{-1} ,

$$A_i = (N-1) \begin{bmatrix} \beta_i C_1^{-1} & 0 & \cdots & 0 \\ 0 & \beta_i C_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \beta_i C_p^{-1} \end{bmatrix} C^{-1} .$$

B_{ij} is the $p \times p$ matrix $E\left[\frac{\partial \log L}{\partial d_{ki}} \frac{\partial \log L}{\partial d_{nj}}\right]$ and $B_{ij} =$ $(N-1)\left[(\beta_i C^{-1} \beta_j') C^{-1} + C^{-1} \beta_i' \beta_j C^{-1}\right]$. The inverse will then give the asymptotic variances and covariances under certain conditions which will also guarantee T_m being distributed asymptotically as χ^2 . These conditions are

1. X have a multivariate normal distribution.
2. $|\varphi_{ij}^2| \neq 0$, where $\varphi = \psi - \beta(\beta' \psi^{-1} \beta)^{-1} \beta'$.
3. β is uniquely determined by specifying that $\beta' \psi^{-1} \beta$ is a diagonal matrix with different and ordered elements.

The conditions also imply that $\sqrt{N}(\hat{\beta} - \beta)$, $\sqrt{N}(\hat{\psi} - \psi)$ have a limiting normal distribution. Anderson and Rubin [1] have proved the above results.

Similarly for simple structure hypotheses they have proved that if Y and G , as defined previously in this chapter, are normally distributed, and β is uniquely determined, $\sqrt{N}(\hat{\beta} - \beta)$, $\sqrt{N}(\hat{F} - F)$,

and $\sqrt{N}\ (\hat{\psi} - \psi)$ are asymptotically normally distributed; this implies that T_m is asymptotically distributed as χ^2 for simple structure hypotheses. Asymptotic variances and covariances have not been obtained for this case; the expressions are probably much more complicated than in the general case. It is unfortunate that the factor analysis model gives rise to such complicated expressions. However, if one cares to invert an $(m + 1)p$ matrix, an estimate of the asymptotic variances and covariances can be obtained by substituting $\hat{\beta}$ and $\hat{C}$ for β and C .

Many of the results in this chapter have been obtained by Anderson and Rubin, but a good bit of the material is not contained in their paper. In addition, this chapter is obviously necessary to give a well-rounded presentation of the theory and the actual computational procedure.

The next and concluding chapter will discuss some of the advantages of Lawley's method as opposed to other techniques, and will also give some suggestions for further research.

CHAPTER VIII

SUMMARY AND CONCLUSIONS

In Chapter I the factor analysis problem has been divided into the following five sections:

1. Model.
2. Estimation of the parameters in the model.
3. Testing of the fit of the model.
4. Estimation of the parameters under simple structure hypotheses.
5. Testing of simple structure hypotheses.

A partial correlation model has been proposed which has been shown to be equivalent to the usual factor analysis model. The method of maximum likelihood has been used to obtain estimates of the parameters in the model, and the resulting maximum likelihood equations are those of Lawley [20]. Then as a test of the fit of the model, the likelihood ratio criterion is employed as the test statistic. For simple structure hypotheses an analogous procedure is followed.

This approach to factor analysis is recommended over all others for several reasons. First, under the conditions given in the previous chapter the maximum likelihood estimates are asymptotically efficient and asymptotically normal. Second, the results are in a sense independent of scale; in particular, one may go directly from results obtained utilizing covariance matrices to those obtained utilizing correlation

matrices. Other methods do not possess this obviously desirable characteristic. Third, statistical tests can be devised which assess the fit of the model, and when to stop factoring. Finally, with this formulation simple structure hypotheses may be tested. The principal disadvantage of the method is the large amount of computation necessary to obtain a solution of the maximum likelihood equations. However, as previously noted, with increasing availability of electronic digital computers and coded routines, this difficulty is largely overcome. In effect, if the usual factor analysis model is specified and the observed variables are assumed to have a multivariate normal distribution, other available methods of factor analysis cannot be recommended or even defended on any statistical or mathematical grounds.

The normality assumption is a rather restrictive one. Even though psychological tests are usually constructed to be approximately normally distributed, the observed variables may have, say, a truncated normal distribution. This may be the case in many applied studies. It would be of interest to examine the estimation and testing aspects under this assumption. Another important problem is the distribution of the test statistic, T_m, for small samples. This is a difficult problem to solve analytically and may, perhaps, be solved only through the use of sampling techniques on high-speed computers. Another important direction for further research is the problem of non-linearity. Some suggestions have been given in Chapter VI, but nothing has actually been done along this line. The Uppsala Symposium on Psychological Factor Analysis lists additional suggestions for research in the field.

BIBLIOGRAPHY

1. Anderson, T. W. and Rubin, H., "Statistical Inference in Factor Analysis," Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability. To be published.

2. Bartlett, M. S., "The Statistical Conception of Mental Factors," British Journal of Psychology, 28 (1937), 97-104.

3. _____, "Internal and External Factor Analysis," British Journal of Psychology, (Statistical Section), 1 (1948), 73-81.

4. _____, "Tests of Significance in Factor Analysis," British Journal of Psychology, (Statistical Section), 3 (1950), 77-85.

5. _____, "Factor Analysis in Psychology as a Statistician Sees It," Uppsala Symposium on Psychological Factor Analysis, 17-19 March 1953, Uppsala, Almqvist and Wiksell, 1953, 23-34.

6. Burt, C., "Factor Analysis and Analysis of Variance," British Journal of Psychology, (Statistical Section), 1 (1948), 3-26.

7. _____, "Alternative Methods of Factor Analysis," British Journal of Psychology, (Statistical Section), 2 (1949), 98-120.

8. _____, "Tests of Significance in Factor Analysis," British Journal of Psychology, (Statistical Section), 5 (1952), 109-133.

9. Carroll, J., "An Analytical Solution for Approximate Simple Structure," Psychometrika, 18 (1953), 23-38.

10. Cramér, H., Mathematical Methods of Statistics, Princeton, Princeton University Press, 1946.

11. Danford, M. B., "Factor Analysis and Related Statistical Techniques," Thesis, North Carolina State College, (1953).

12. Eisenhart, C., "The Assumptions Underlying the Analysis of Variance," Biometrics, 3 (1947), 1-21.

13. Emmett, W. G., "Factor Analysis by Lawley's Method of Maximum Likelihood," British Journal of Psychology, (Statistical Section), 2 (1949), 90-97.

14. Henrysson, S., "The Significance of Factor Loadings," British Journal of Psychology, (Statistical Section), 3 (1950), 159-165.

15. Hotelling, H., "Analysis of a Complex of Statistical Variables into Principal Components," Journal of Educational Psychology, 24 (1933), 417-441, 498-520.

16. _____, "Rotation in Psychology and the Statistical Revolution," Science, 95 (1942), 504-507.

17. _____, "Some New Methods in Matrix Calculation," The Annals of Mathematical Statistics, 14 (1943), 1-34.

18. Kendall, M. G. and Smith, B. B., "Factor Analysis," Journal of the Royal Statistical Society, Series B, 12 (1950), 60-94.

19. _____, Notes on Multivariate Analysis, Raleigh, N. C., Institute of Statistics Mimeograph Series No. 95, March 1954.

20. Lawley, D. N., "The Estimation of Factor Loadings by the Method of Maximum Likelihood," Proceedings of the Royal Society of Edinburgh, 60 (1940), 64-82.

21. _____, "Further Investigations in Factor Estimation," Proceedings of the Royal Society of Edinburgh, 61 (1942), 176-185.

22. _____, "The Application of the Maximum Likelihood Method to Factor Analysis," British Journal of Psychology, 33 (1943), 172-175.

23. _____, "Problems in Factor Analysis," Proceedings of the Royal Society of Edinburgh, 62 (1949), 394-399.

24. _____, "A Further Note on a Problem in Factor Analysis," Proceedings of the Royal Society of Edinburgh, 63 (1950), 93-94.

25. _____, "A Modified Method of Estimation in Factor Analysis and Some Large Sample Results," Uppsala Symposium on Psychological Factor Analysis, 17-19 March 1953, Uppsala, Almqvist and Wiksell, 1953, 35-42.

26. _____, and Swanson, Z., "Tests of Significance in a Factor Analysis of Artificial Data," British Journal of Statistical Psychology, 7 (1954), 75-79.

27. Quensel, C. E., "The Distribution of the Partial Correlation Coefficient in Samples from Multivariate Universes in a Special Case of Non-normally Distributed Random Variables," Skandinavisk Aktuarietidskrift, 36 (1953), 16-23.

28. Rao, C. R., "Estimation and Tests of Significance in Factor Analysis," Psychometrika, 20 (1955), 93-111.

29. Rippe, D. D., "Application of a Large Sampling Criterion to Some Sampling Problems in Factor Analysis," Psychometrika, 18 (1953), 191-205.

30. Roy, S. N., A Report on Some Aspects of Multivariate Analysis, Raleigh, N. C., Institute of Statistics Mimeograph Series No. 121, December 1954.

31. Thomson, G. H., "Some Points of Mathematical Technique in the Factorial Analysis of Ability," Journal of Educational Psychology, 27 (1936), 37-54.

32. _____, The Factorial Analysis of Human Ability, London, University of London Press, 1951.

33. Thurstone, L. L., Multiple Factor Analysis, Chicago, University of Chicago Press, 1947.

34. Whittaker, E. and Robinson, G., The Calculus of Observations, London and Glasgow, Blackie and Sons Limited, 1944.

35. Whittle, P., "On Principal Components and Least Squares Methods of Factor Analysis," Skandinavisk Aktuarietidskrift, 35 (1952), 223-239.

36. Wold, H., "Some Artificial Experiments in Factor Analysis," Uppsala Symposium on Psychological Factor Analysis, 17-19 March 1953, Uppsala, Almqvist and Wiksell, 1953, 43-64.

37. Young, G., "Maximum Likelihood Estimation and Factor Analysis," Psychometrika, 6 (1941), 49-53.