CHAPTER 9
FACTOR FITTING BY STATISTICAL FUNCTIONS

From
Exploratory Factor Analysis
Ledyard R Tucker
and
Robert C. MacCallum

CHAPTER 9
FACTOR FITTING BY STATISTICAL FUNCTIONS

Determination of factor matrices from covariance matrices according to several statistical functions is discussed in this chapter. In Chapter 8 we concentrated on factor extraction by matrix factoring techniques. A contrast may be considered for an example between the principal factors method of factor extraction described in Chapter 8 and the minimum squared residuals method (MINRES) to be described in this chapter. The principal factors method obtains factors to maximize the sum of squares of the factors determined from ($\boldsymbol{R_{yy}} - \boldsymbol{U^2}$) where $\boldsymbol{R_{yy}}$ is the observed correlation matrix. The MINRES method minimizes the sum of squares of the residuals in matrix $\boldsymbol{\Delta}$ where:

$$\boldsymbol{\Delta} = \boldsymbol{R_{yy}} - (\boldsymbol{AA'} + \boldsymbol{U^2})$$

This sum of squares of the residuals will be expressed as a statistical function which is minimized by the determination of the common factor matrix $\boldsymbol{A}$ and uniqueness $\boldsymbol{U^2}$. Two other statistical functions will be considered: maximum likelihood factor analysis and factor analysis by generalized least squares. However, there are several preliminary, general topics to be discussed before getting to the particular factoring methods.

Bentler & Bonnett (1980) provided a general least squares function as an extension of the function used by Jöreskog & Goldberger (1972) for their generalized least squares factor analysis. The Bentler & Bonett function may be written as:

$$G = \frac{1}{2} tr[\boldsymbol{W}(\boldsymbol{C_{yy}} - \boldsymbol{C_{zz}})]^2 \tag{9.1}$$

where $\boldsymbol{C_{yy}}$ is the observed covariance matrix and $\boldsymbol{C_{zz}}$ is the covariance matrix obtained from the factor model:

$$\boldsymbol{C_{zz}} = \boldsymbol{AA'} + \boldsymbol{U^2} \tag{9.2}$$

and $\boldsymbol{W}$ is a weight matrix. There is a problem, however, the matrix product $[\boldsymbol{W}(\boldsymbol{C_{yy}} - \boldsymbol{C_{zz}})]$ is not symmetric so that the trace of its square is not a sum of squares. This problem may be resolved when $\boldsymbol{W}$ is symmetric, positive definite so that it may be decomposed into the product of a matrix $\boldsymbol{T}$ times its transpose:

$$\boldsymbol{W} = \boldsymbol{TT'} \tag{9.3}$$

with $\boldsymbol{T}$ square and non-singular so that its inverse exists. Then:

$$
\begin{aligned}
tr[\boldsymbol{W}(\boldsymbol{C_{yy}}-\boldsymbol{C_{zz}})]^2 &= tr[\boldsymbol{W}(\boldsymbol{C_{yy}}-\boldsymbol{C_{zz}})\boldsymbol{W}(\boldsymbol{C_{yy}}-\boldsymbol{C_{zz}})] \\
&= tr[\boldsymbol{T}\boldsymbol{T}'(\boldsymbol{C_{yy}}-\boldsymbol{C_{zz}})\boldsymbol{T}\boldsymbol{T}'(\boldsymbol{C_{yy}}-\boldsymbol{C_{zz}})] \\
&= tr[\boldsymbol{T}'(\boldsymbol{C_{yy}}-\boldsymbol{C_{zz}})\boldsymbol{T}\boldsymbol{T}'(\boldsymbol{C_{yy}}-\boldsymbol{C_{zz}})\boldsymbol{T}]
\end{aligned}
$$

so that:

$$
tr[\boldsymbol{W}(\boldsymbol{C_{yy}}-\boldsymbol{C_{zz}})]^2 = tr[\boldsymbol{T}'(\boldsymbol{C_{yy}}-\boldsymbol{C_{zz}})\boldsymbol{T}]^2 \tag{9.4}
$$

With matrices $\boldsymbol{C_{yy}}$ and $\boldsymbol{C_{zz}}$ being symmetric, matrix $[\boldsymbol{T}'(\boldsymbol{C_{yy}}-\boldsymbol{C_{zz}})\boldsymbol{T}]$ is symmetric so that the trace of its square is a sum of squares.

Scale freeness of a factoring method is considered to be very important by theoretically oriented individuals. This topic, which was discussed in Chapter 8, applies to all three methods of factor determination as described in this chapter. The idea is that when the attribute measures are rescaled the factor matrix will be unchanged except for a rescaling of the rows of the factor matrix in exactly the same fashion as the attributes were rescaled. The uniquenesses also will be rescaled in the same fashion. Let $\boldsymbol{D}$ be a rescaling diagonal matrix so that:

$$
\widetilde{\boldsymbol{C}}_{\boldsymbol{yy}} = \boldsymbol{D}\boldsymbol{C_{yy}}\boldsymbol{D}\,. \tag{9.5}
$$

With

$$
\widetilde{\boldsymbol{A}} = \boldsymbol{D}\boldsymbol{A} \tag{9.6}
$$

$$
\widetilde{\boldsymbol{U}}^{\boldsymbol{2}} = \boldsymbol{D}\boldsymbol{U}^{\boldsymbol{2}}\boldsymbol{D} \tag{9.7}
$$

so that:

$$
\widetilde{\boldsymbol{C}}_{\boldsymbol{zz}} = \widetilde{\boldsymbol{A}}\,\widetilde{\boldsymbol{A}}' + \widetilde{\boldsymbol{U}}^{\boldsymbol{2}} \tag{9.8}
$$

$$
\widetilde{\boldsymbol{C}}_{\boldsymbol{zz}} = \boldsymbol{D}\boldsymbol{C_{zz}}\boldsymbol{D} \tag{9.9}
$$

The modeled covariance matrix is rescaled in the same fashion as the observed covariance matrix. Jöreskog & Goldberger (1972) gave a general condition for scale freeness. Let ϕ be a general function yielding a numerical value. Their condition may be written as:

$$
\phi(\boldsymbol{C_{yy}},\boldsymbol{C_{zz}}) = \phi(\boldsymbol{D}\boldsymbol{C_{yy}}\boldsymbol{D},\boldsymbol{D}\boldsymbol{C_{zz}}\boldsymbol{D}) \tag{9.10}
$$

or

$$
\phi(\boldsymbol{C_{yy}},\boldsymbol{C_{zz}}) = \phi(\widetilde{\boldsymbol{C}}_{\boldsymbol{yy}},\widetilde{\boldsymbol{C}}_{\boldsymbol{zz}})\ . \tag{9.11}
$$

The application of this condition will be presented for all three of the methods as described in this chapter.

A large sample statistical function has been derived for each of the three methods to be described in this chapter. These functions result in tests whether a factor model with a specified number of attributes and number of factors exactly represent the population covariance matrix. That is, the lack of fit in a sample may be allocated to sampling fluctuations with no model error. While these statistical tests may be useful in decisions as to the numbers of factors to be established, these tests must be used with some care. A problem, which was discussed in Chapter 3, is generated by the proposition that almost all theoretic models are not correct representations of the real world. Then, there is an expectation that a very useful factor model will not exactly represent the population covariance matrix in any particular study. As indicated in Chapter 4, the discrepancy matrix, $\boldsymbol{\Delta}_y$, between an observed covariance matrix and a modeled covariance matrix is a sum of model error and sampling error. With a small sample many factor models might not be rejected due to the lack of power of the statistical tests while all reasonable factor models would be rejected with large sample data. The power of these statistical tests is a direct function of sample size. As a general proposition, statistical stability of results of an analysis is highly desirable. This does not indicate, however, that trivial, but stable, dimensions should be retained in a factor analysis. An analyst, nevertheless, must be alert to the possible small factor which may be made more important with a redesigned battery of attributes. There is a necessary judgment as to the number of scientifically meaningful factors involved in the observed covariance matrix. The statistical tests will help on the issue of statistical stability.

Some of the issues indicated in the preceding paragraph may be clarified by consideration of classes of the factor analysis model in the population which were discussed in Chapter 4. Equation (3.35) expresses the common factor model in the population with correlated factors

$$\boldsymbol{\Sigma}_{zz} = \boldsymbol{B\Phi B}' + \boldsymbol{U}^2 \,. \tag{3.35}$$

The corresponding formula for uncorrelated factors is obtained from equation (3.45)

$$\boldsymbol{\Sigma}_{zz} = \boldsymbol{AA}' + \boldsymbol{U}^2 \tag{9.12}$$

Factor transformations among and between uncorrelated factors and correlated factors was discussed in Chapter 3. These equations provide the model in abstract terms. The general model may be termed the abstract common factor model. In contrast there are many concrete models for which the numerical values of the model are specified. These specifications include not only the number of attributes and the number of common factors but also the values of the uniqueness and numerical values in the common factor structure. This common factor structure includes all matrices $\boldsymbol{B}$, $\boldsymbol{\Phi}$, and $\boldsymbol{A}$ which are interrelated by factor transformations. A conditional abstract

model may be considered which involves specification of the number of attributes and number of common factors with the specification that any matrix $\boldsymbol{A}$ in the factor structure is of full column rank. The abstract model contains an infinite set of concrete models and a conditional abstract model contains a subset of concrete models. It is well known that any population covariance matrix, $\boldsymbol{\Sigma}_{yy}$, may be satisfied by equations (3.35) and (9.12) provided that a large enough number of common factors is employed. Thus, a population covariance matrix among surface attributes may be represented by a concrete model for n attributes and r_y common factors. However, when a common factor model is determined for r_z common factors with r_z less than r_y by any of the factoring methods, there is bound to be a discrepancy which may be termed model error as in Chapter 3 and defined, from equation (3.90) as:

$$\boldsymbol{\Delta}_{\Sigma} = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{zz} \ . \tag{9.13}$$

A crucial point is that a specification of r_z along with the number of attributes defines a conditional abstract model which contains many concrete models.

Each of the factor fitting methods by a statistical function is defined in terms of an observed covariance matrix. However, each of these methods has a population counterpart obtained by substituting the population covariance matrix among the surface attributes for an observed covariance matrix. Since observed covariance matrices approach population covariance matrices with increase in sample size, each of these factor fitting techniques is statistically consistent so that the results using one of these methods approaches results that would be obtained from the population covariance matrix. Consider that the number of common factors defined for the fitted factor matrix is less than the number of common factors associated with the population covariance matrix among the observed attributes; that is: r_z is less than r_y . Each of the factor fitting methods will result in a solution in the conditional abstract model subset with the given number of attributes and number of common factors. There will be a model error, as previously described, since the population covariance matrix among the observed attributes can not be represented in this conditional abstract model subset. In a small sample the model error will be incorporated with considerable sampling error in the discrepancy between the observed covariance matrix and the model covariance matrix determined in the sample. The sampling error may hide the model error so that the model will not be rejected by the statistical test function. In comparison, there will be smaller sampling error in a larger sample so that the model error will not be hidden with a consequent that the model will be rejected by the statistical test function.

The concept of concrete factor analysis models is illustrated numerically in Tables 9.1, 9.2, and 9.3. A simulated population covariance matrix was generated by the Tucker-Koopman-Linn technique described in Chapter 3. Table 9.1 presents the data generating parameters used in

Table 9.1

Data Generating Parameters for

Creation of Simulated Population Correlation Matrix

	Major Domain Factors			Unique Variance	Unique Variance
	1	2	3		
1	.95	.00	.00	.0000	.0975
2	.95	.00	.00	.0000	.0975
3	.95	.00	.00	.0000	.0975
4	.95	.00	.00	.0000	.0975
5	.00	.70	.00	.4125	.0975
6	.00	.70	.00	.4125	.0975
7	.00	.70	.00	.4125	.0975
8	.00	.00	.45	.7000	.0975
9	.00	.00	.45	.7000	.0975
10	.00	.00	.45	.7000	.0975

ε = .20; Number of Minor Factors = 50

Table 9.2

Simulated Population Correlation Matrix

Variables	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	.87	1.00								
3	.98	.85	1.00							
4	.98	.89	.97	1.00						
5	.00	.02	-.02	-.02	1.00					
6	-.07	.04	-.07	-.06	.51	1.00				
7	.01	.00	.01	.02	.50	.54	1.00			
8	-.07	.08	-.09	-.06	.01	.07	-.01	1.00		
9	.03	-.04	.05	.04	.03	-.05	-.02	.15	1.00	
10	.03	-.05	.04	-.01	-.03	-.07	-.07	.16	.22	1.00

Table 9.3

Concrete Models Obtained from Analysis of

Simulated Population Correlation Matrix*

	MINRES Solution					Maximum Likelihood Solution				
		Factor		Unique-			Factor		Unique-
Variable	1	2	3	ness	Variable	1	2	3	ness
1	.99	.00	.01	.01	1	.99	.00	.12	.01
2	.88	.06	-.03	.23	2	.93	.07	-.37	.00
3	.98	-.01	.02	.04	3	.98	-.01	.14	.03
4	1.00	.01	-.01	.01	4	.99	.00	.06	.03
5	-.02	.68	.08	.53	5	-.01	.68	.05	.54
6	-.06	.75	.01	.43	6	-.05	.76	-.10	.41
7	.00	.72	.00	.48	7	.00	.73	.14	.45
8	-.04	.03	.32	.90	8	-.04	.04	-.30	.90
9	.02	-.04	.46	.79	9	.02	-.03	.15	.98
10	.01	-.10	.49	.75	10	.01	-.09	.14	.97

*Factor matrices were orthogonally transformed to principal axes orientation.

this example. The first four attributes had very high loadings on the first major domain factors, the next three attributes had moderately high loadings on the second major domain factor while the last three attributes had only moderate loadings on the third major domain factor. The minor variances were set equal for all ten attributes to a small value of .0975. A considerable extent of lack of fit was introduced by use of a relatively large value of ϵ and small number of minor factors. Table 9.2 presents a simulated population correlation matrix (covariance matrix for this example). A different matrix could be produced using the same generating parameters with different random numbers in the generation of the minor factors. Note the larger correlations in the boxes corresponding to the attributes with high loadings on the three major domain factors as compared to the correlations not in the boxes. Concrete models were computed by two of the factor fitting methods to be described in this Chapter: MINRES and Maximum Likelihood Factoring methods.

Table 9.3 presents two concrete factor analysis models produced by the two methods of factor fitting using three common factors. MINRES results include all three major domain factors while the Maximum Likelihood results include the first two major domain factors but yielded a peculiar third factor which, undoubtedly, is associated with the lack of fit introduced in the data generation procedure. Our major point is that these two methods of factor fitting yielded different concrete models from the same population correlation matrix. When larger and larger, multidimensional normal samples are drawn from the population characterized by the simulated correlation matrix, application of the two methods of factor fitting will yield results approaching the concrete models produced in the population. The MINRES results will approach the MINRES solution in the population while the Maximum Likelihood results will approach the Maximum Likelihood solution in the population. In Monte Carlo studies with similar simulated covariance matrices and using samples as small as 100 , results indicate that MINRES results from the samples reflect the MINRES solution in the population and not the Maximum Likelihood solution in the population. Likewise, the Maximum Likelihood results in the samples reflect the Maximum Likelihood solution in the population and not the MINRES solution in the population. Thus, for these two methods of factor fitting, results for each method from samples estimate the parameters in the concrete model established in the population by that method.

In contrast to the statistical test functions discussed in preceding paragraphs is a topic of goodness of fit of a concrete model to the real world population covariance matrix. There have been several measures developed for structural models analysis, confirmatory factor analysis being a form of structural model analysis. These will be discussed in later Chapters. The first contribution to a coefficient for indicating goodness of fit was by Tucker & Lewis (1973) with a proposed reliability coefficient for Maximum Likelihood Factor Analysis. This coefficient is generalized in this chapter to Generalized Least Squares Factor Analysis and MINRES factor

fitting. Consider the discrepancy matrix, $\mathbf{\Delta_y}$, between the observed covariance matrix and the modeled covariance matrix. In the Maximum Likelihood system the uniqueness is added into the diagonals and the resulting matrix standardized by pre and post multiplication by a diagonal matrix to having unit diagonal entries. The resulting matrix contains partial correlations after the common factors have been removed. Jöreskog's function F_r after determination of r common factors approximates the sum of squares of the entries on one side of the diagonal of this partial correlation matrix. A mean square, M_r , may be obtained by dividing F_r by the remaining degrees of freedom, df_r .

$$M_r = F_r/df_r \ . \tag{9.14}$$

Having a mean square suggest a components of variation analogy with an intraclass correlation as a reliability coefficient. Let M_0 be the value of M for no common factors. For the variance components model let α_r be a variance associated with a model having r common factors, δ_r be a variance representing the deviation of the model from actuality, and ϵ_r be a variance associated with sampling. Using expected values, consider the components of variance as:

$$Exp(M_0) = \alpha_r + \delta_r + \epsilon_r \ , \tag{9.15}$$

$$Exp(M_r) = \delta_r + \epsilon_r \ . \tag{9.16}$$

The intraclass correlation which is used as a reliability coefficient is defined by:

$$\rho_r = \frac{\alpha_r}{\alpha_r + \delta_r} \ . \tag{9.17}$$

To obtain a solution in terms of observed coefficients, Tucker & Lewis substitute observed values of the M coefficients for the expected values. A value of ϵ_r is obtained by assuming that, if the factor model fit exactly in r common factors, δ_r would equal zero and the chi square would equal unity. In this case, $\epsilon_r$would equal M_r which, in turn, would equal $1/N_r^{\bigstar}$ with:

$$N_r^{\bigstar} = N - 1 - (2n + 5)/6 - 2r/3 \tag{9.18}$$

Solution for the reliability coefficient yields:

$$\rho_r \doteq \frac{M_0 - M_r}{M_0 - 1/N_r^{\bigstar}} \ . \tag{9.19}$$

Tucker & Lewis (1973, page 5) indicate that: "This reliability coefficient may be interpreted as indicating how well a factor model with r common factors represents the covariances among the attributes for a population of objects". This coefficient has been adapted from Maximum Likelihood Factor Analysis to General Least Squares Factor Analysis and MINRES factor fitting.

9.1. Maximum Likelihood Factor Analysis.

Lawley (1940) developed the method of maximum likelihood factor analysis at a time when the field had considerable need for a sound theoretic statistical foundation. The centroid method was in use at the Thurstone laboratory for analyses of major factor analytic studies. The number of factors was a major problem. In that the maximum likelihood method provided a statistical test for significance of factors this was a most promising development. However, at that time, computing facilities were very limited so that the extensive computations required by the maximum likelihood method made this method unavailable. Application remained limited until developments of modern digital computers made them sufficiently powerful to accomplish the required computations. Lord (1956) used Whirlwind I computer in the analysis of a covariance matrix among 39 attributes. He used an iterative technique suggested by Lawley (1942). Convergence was very slow. This method had several other problems such as converging on secondary maximum. Jöreskog (1967) contributed advanced procedures which could use more powerful analysis methods such as the Newton-Raphson iterations. With the computer developments and Jöreskog's contributions, maximum likelihood factor analysis is quite feasible. This method has many desirable statistical properties such as consistency, normality, efficiency.

Theoretic developments of maximum likelihood factor analysis starts with a population covariance matrix, $\boldsymbol{\Sigma}$, and sample score vectors, $\boldsymbol{y}$, which are multivariate normal drawn from the population characterized by the population covariance matrix. Note the normality assumption which is added to the assumptions of the factor analysis model which have been given previously. The number of attributes is taken to be n and the sample size is N . Consider all possible observed covariance matrices, $\boldsymbol{C}$, for samples of size N among n attributes when the score vectors are drawn multivariate normal from a population having a covariance matrix $\boldsymbol{\Sigma}$. Wishart (1928) gave the density function for these observed covariance matrices.

$$\phi(\boldsymbol{C}\big|\boldsymbol{\Sigma},\mathrm{N}) = \mathrm{K}\big|\boldsymbol{\Sigma}\big|^{-\frac{1}{2}(\mathrm{N}-1)}\big|\boldsymbol{C}\big|^{\frac{1}{2}(\mathrm{N}-n-2)}\mathrm{EXP}[-\frac{1}{2}(\mathrm{N}-1)tr(\boldsymbol{C}\boldsymbol{\Sigma}^{-1})] \qquad (9.20)$$

where K is a constant involving only N and n .

The likelihood of a given observed covariance matrix, $\boldsymbol{C}_{yy}$, is taken equal to the value of the density function for that observed covariance matrix. It is illogical to consider that the probability of this covariance matrix is other than unity since this covariance matrix already has occurred. Thus, the likelihood is a measure of typicality versus rarity of an observation. In the present case, the likelihood of an observed $\boldsymbol{C}_{yy}$ is given by:

$$\mathrm{L}(\boldsymbol{C}_{yy}\big|\boldsymbol{\Sigma},\mathrm{N}) = \mathrm{K}\big|\boldsymbol{\Sigma}\big|^{-\frac{1}{2}(\mathrm{N}-1)}\big|\boldsymbol{C}_{yy}\big|^{\frac{1}{2}(\mathrm{N}-n-2)}\mathrm{EXP}[-\frac{1}{2}(\mathrm{N}-1)tr(\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{yy})] \qquad (9.21)$$

Subsequent developments and discussions are expedited by a logarithmic conversion to the log-likelihood function:

$$\Lambda = ln[\mathrm{L}(\boldsymbol{C_{yy}}\big|\boldsymbol{\Sigma}\,,\mathrm{N})]\ . \tag{9.22}$$

Substitution from equation (9.21) yields:

$$\begin{aligned}\Lambda = ln(\mathrm{K}) - \frac{1}{2}(\mathrm{N}-1)[\ ln\big|\boldsymbol{\Sigma}\big|] + \frac{1}{2}(\mathrm{N}-\mathrm{n}-2)[\ ln\ \big|\boldsymbol{C_{yy}}\big|] \\ - \frac{1}{2}(\mathrm{N}-1)[tr(\boldsymbol{\Sigma}^{-1}\boldsymbol{C_{yy}})]\ .\end{aligned} \tag{9.23}$$

A statistical function is developed using the likelihood ratio which compares a theoretical model which is to be fit to the data with a more general comparison model. The comparison model is to include the theoretical model as a special case but has a larger number of parameters than does the theoretical model. For present considerations, the comparison model is to have a sufficient number of parameters to completely account for the observed covariance matrix, $\boldsymbol{C_{yy}}$. For each of the theoretic model and the comparison model an estimate of the population covariance matrix is developed along with the likelihood of the solution. For the theoretic model, hereafter designated model m , the estimate is $\widehat{\Sigma}_{\mathbf{m}}$. The comparison model is designated model M and yields estimate $\widehat{\boldsymbol{\Sigma}}_{\mathbf{M}}$. The likelihood functions for these two models are designated $\mathrm{L_m}$ and $\mathrm{L_M}$ while the log-likelihood functions are designated Λ_{m} and Λ_{M} . The likelihood ratio, λ **,** is defined by:

$$\lambda = \mathrm{L_m}/\mathrm{L_M}\ . \tag{9.24}$$

A coefficient ξ based on the likelihood ratio provides an approximate chi square for a statistical test on whether or not the theoretic model is a complete representation of the population covariance matrix.

$$\xi = -2[\ ln\,(\lambda)]\ . \tag{9.25}$$

Substitution from equation (9.24) yields:

$$\begin{aligned}\xi &= -2[ln(L_m) - ln(L_M)] \\ &= -2[\Lambda_m - \Lambda_M] \\ &= (\mathrm{N}-1)\{ln\big|\Sigma_m\big| + tr(\boldsymbol{\Sigma}_m^{-1}\boldsymbol{C_{yy}}) - ln\big|\boldsymbol{\Sigma_M}\big| - tr(\boldsymbol{\Sigma_M^{-1}}\boldsymbol{C_{yy}})\}\ .\end{aligned} \tag{9.26}$$

Let η(m) and η(M) be the number of parameters involved in models m and M . The number of degrees of freedom for ξ is $\nu(\xi)$ and equals the difference between the number of parameters used in the two models:

$$\nu(\xi) = \eta(\mathrm{M}) - \eta(\mathrm{m})\ . \tag{9.27}$$

At this point it is necessary to trace through the maxima - minima relations among the likelihood, L($\boldsymbol{C_{yy}}\big|\boldsymbol{\Sigma}$,**N**) , the log-likelihood, Λ , likelihood ratio, $\boldsymbol{\lambda}$, and coefficient ξ . From equation (9.22), Λ is an increasing monotonic function of L($\boldsymbol{C_{yy}}\big|\boldsymbol{\Sigma}$,N) so that maximum Λ occurs with the maximum L($\boldsymbol{C_{yy}}\big|\boldsymbol{\Sigma}$, N) ; that is, maximizing the log-likelihood maximizes the likelihood. Consider that the comparison model is fixed so that L_M and Λ_M can be considered to be constants. Then, the likelihood ratio, λ , is a monotonic increasing function of L_m so that a maximum likelihood corresponds to a maximum likelihood ratio. However, from equation (9.25), the coefficient ξ is a decreasing monotonic function of λ so that the maximum likelihood is associated with a minimum coefficient ξ . Analysis procedures are required to minimize ξ .

Comparison model M is considered first. The number of parameters in this model equals the number of variances and independent covariances in $\boldsymbol{C_{yy}}$ (since matrix $\boldsymbol{C_{yy}}$ is symmetric, the number of independent covariances equals the number of entries on one side of the diagonal). Then:

$$\eta(M) = \frac{1}{2}n(n+1) \ . \tag{9.28}$$

With this number of parameters an exact solution may be obtained so that:

$$\boldsymbol{\widehat{\Sigma}_M = C_{yy}} \ . \tag{9.29}$$

Then:

$$\boldsymbol{\widehat{\Sigma}_M^{-1} C_{yy} = I} \tag{9.30}$$

and

$$tr(\boldsymbol{\widehat{\Sigma}_M^{-1} C_{yy}}) = n \ . \tag{9.31}$$

Taking the estimated theoretic covariance matrix to be $\boldsymbol{\widehat{\Sigma}_m}$ and substituting from equations (9.29) and (9.31) into equation (9.26) yields:

$$\xi = (N-1)\{ln\big|\boldsymbol{\widehat{\Sigma}_m}\big| + tr(\boldsymbol{\widehat{\Sigma}_m^{-1} C_{yy}}) - ln\big|\boldsymbol{C_{yy}}\big| - n\} \ . \tag{9.32}$$

Jöreskog (1967) suggested a convenient function free from the effects of sample size as:

$$F = \xi/(N-1) \ . \tag{9.33}$$

$$F = ln\big|\boldsymbol{\widehat{\Sigma}_m}\big| + tr(\boldsymbol{\widehat{\Sigma}_m^{-1} C_{yy}}) - ln\big|\boldsymbol{C_{yy}}\big| - n \ . \tag{9.34}$$

The estimated theoretic covariance matrix is to be established so as to minimize this function F with the number of free parameters used in this estimate being η(m) . This function is the basis

not only for the maximum likelihood procedures in exploratory factor analysis but also in confirmatory factor analysis and the more general covariance structure analyses. Only the solution for exploratory factor analysis will be considered in this chapter.

The estimated theoretic covariance matrix may be taken as our matrix $\boldsymbol{C_{zz}}$ for a sample. This usage along with the factor analysis model for uncorrelated factors may be expressed as:

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{m}} = \boldsymbol{C_{zz}} = \boldsymbol{AA'} + \boldsymbol{U^2} \tag{9.35}$$

where $\boldsymbol{A}$ is the estimated factor matrix and $\boldsymbol{U^2}$ is the estimated uniqueness matrix. Note that this usage does not involve the discrepancy matrix $\boldsymbol{\Delta_y}$ which we saw in Chapter 4 included both sampling error and model error. The maximum likelihood procedure ignores the model error substituting, instead, the statistical test to indicate a solution for which there is no model error. In practice, the statistical test is taken as advisory to indicate small model error when the chi square indicates that the estimated model should be rejected with only a small probability associated with the chi square. With a very tiny probability the model should be rejected.

Early computing procedures suggested by Lawley (1940, 1941, 1949) involved alternating iterations between solution for matrix $\boldsymbol{A}$ conditioned on a trial $\boldsymbol{U^2}$ and solution for $\boldsymbol{U^2}$ conditioned on a trial matrix $\boldsymbol{A}$. Since more effective methods have been developed subsequently by Jöreskog (1967, 1969) and by Jennrick and Robinson (1969) only a summary of these early methods will be presented here. The conditional solution for $\boldsymbol{A}$ given a trial $\boldsymbol{U^2}$ is obtained by setting the derivative of F with respect to $\boldsymbol{A}$ equal to zero with the results that follow. Obtain the eigenvalues, q_k , and unit length eigenvectors $\underline{\mathbf{v}}_{\boldsymbol{k}}$, of the matrix $\boldsymbol{U^{-1}C_{yy}U^{-1}}$, with matrix $\boldsymbol{Q}$ diagonal and matrix $\boldsymbol{V}$ containing the eigenvectors as columns. The eigenvalues are arranged in descending algebraic order. With $\boldsymbol{C_{yy}}$ symmetric this solution may be expressed by:

$$\boldsymbol{U^{-1}C_{yy}U^{-1}} = \boldsymbol{VQV'} \tag{9.36}$$

For r factors, let $\boldsymbol{V_r}$ contain the first r eigenvectors and $\boldsymbol{Q_r}$ be an $r \times r$ diagonal matrix containing the first r eigenvalues. Matrix $\boldsymbol{A}$ is:

$$\boldsymbol{A} = \boldsymbol{UV_r}(\boldsymbol{Q_r} - \boldsymbol{I})^{\frac{1}{2}} \tag{9.37}$$

Solution for $\boldsymbol{U^2}$ is obtained by setting the derivative of F with respect to $\boldsymbol{U^2}$ for a given trial matrix $\boldsymbol{A}$ equal to zero. This results in:

$$\boldsymbol{U^2} = diag(\boldsymbol{C_{yy}}) - diag(\boldsymbol{AA'}) \ . \tag{9.38}$$

The procedure was to start with a trial $\boldsymbol{U^2}$, compute a factor matrix with equations (9.36) and (9.37), then compute a new $\boldsymbol{U^2}$ with equation (9.38) and return to the solution for a new factor

matrix. These cycles were repeated until there were minimal changes in values from one cycle to the next. A point to note is that the $\boldsymbol{U^2}$ is set such that the residual diagonal is set to zero. This results in the communality of each attribute plus the uniqueness for that attribute summing to the observed variance in the diagonal of $\boldsymbol{C_{yy}}$. In addition to this procedure converging slowly it stopped for some data before true convergence was obtained. A further major problem occurred when, for some data on a cycle, the iterated values of $\boldsymbol{U^2}$ contained one or more negative entries. In this case $\boldsymbol{U^{-1}}$ did not exist as a real number and the solution stopped. This is known as a generalized Heywood (1931) case. With these problems a more effective solution was very desirable.

Two iterative methods are possible for solution for the $\boldsymbol{U^2}$ values: a gradient method and a Newton-Raphson method. The gradient method is to be considered first. Let f be a function of a vector $\underline{\boldsymbol{x}}$. The first derivative of f with respect to the entries in $\underline{\boldsymbol{x}}$ form a vector **g** . This vector is termed the gradient vector. Element i of **g** is:

$$\mathrm{g}_i = \partial f / \partial x_i \quad . \tag{9.39}$$

Let $\underline{\boldsymbol{x}}_t$ be a trial vector which is to be improved by the gradient method. The gradient vector corresponding to $\underline{\boldsymbol{x}}_{(t+1)}$ is $\mathbf{g}_t$. The next trial vector, $\underline{\boldsymbol{x}}_{(t+1)}$is obtained by:

$$\underline{\boldsymbol{x}}_{(t+1)} = \underline{\boldsymbol{x}}_t + m\mathbf{g}_t \tag{9.40}$$

where m is a step size coefficient. To iterate to a maximum, m is to be positive; to iterate to a minimum, m is to be negative. Various techniques are available to indicate an advantageous value of step size to be used. With the absolute value of m infinitesimal, the iteration will follow what is called the orthogonal trajectory of the function f and eventually end at a maximum for a positive m or at a minimum for a negative m . The number of iterations may be reduced materially by using larger absolute values of m . Using quite large absolute values of m can lead to a change in $\underline{\boldsymbol{x}}$ which leads to either a decreased value of f during iterations to a maximum or an increase in f during iterations to a minimum. When this occurs, a trial could be repeated with a smaller step size. An advantageous step size might be determined by repeating a trial with a number of step sizes and choosing the step size which yields the best results. When a satisfactory step size has been established it could be carried over to subsequent trials. There are analytic methods for establishing approximations to the step size which yields the most desirable results for each trial. These methods, however, are beyond the scope for this book.

The Newton-Raphson iterative method uses both the first and second derivatives of function f with respect to elements of vector $\underline{\boldsymbol{x}}$. The first derivative vector **g** was discussed in the preceding paragraph. The second derivatives of f with respect to pairs of entries in vector $\underline{\boldsymbol{x}}$ are entries h_{ij} in matrix $\boldsymbol{H}$.

$$h_{ij} = \partial^2 f / \partial x_i \partial x_j \ . \tag{9.41}$$

For a trial vector $\underline{\boldsymbol{x_t}}$ the first and second derivatives are $\mathbf{g_t}$ and $\boldsymbol{H_t}$. Then, the next trial vector, $\underline{\boldsymbol{x}}_{(t+1)}$ is given by:

$$\underline{\boldsymbol{x}}_{(t+1)} = \underline{\boldsymbol{x_t}} - \boldsymbol{H_t^{-1}}\mathbf{g}_t \ . \tag{9.42}$$

This would be an exact solution for a quadratic function f and works well when the actual function f is approximately quadratic. This condition appears to be quite good in many cases near an optimum, either a maximum or a minimum. At times the iteration from one trial to the next the value of the function will go the wrong way (decrease for iteration to a maximum or increase for an iteration to a minimum). In such cases, the iteration should revert to the gradient method for one or more trials before attempting the Newton-Raphson procedure. The determinant of the second derivative matrix should be negative for iterations to a maximum and negative for iterations to a minimum.

Both the gradient and Newton-Raphson methods work best starting from good first estimates of the solution vector $\underline{\boldsymbol{x}}$.

In the application of the preceding iterative methods to the maximum likelihood solution, function F of equation (9.34) is transformed from a function of the estimated factor matrix $\boldsymbol{A}$ and estimated uniqueness $\boldsymbol{U^2}$ to a function of the eigenvalues of matrix $\boldsymbol{U^{-1}C_{yy}U^{-1}}$ eliminating estimated factor matrix $\boldsymbol{A}$ from the solution. When a solution for $\boldsymbol{U^2}$ is obtained the estimate of factor matrix $\boldsymbol{A}$ is obtained from equation (9.37). In order to avoid negative estimates of uniqueness a transformation of variables is taken from each u_i^2 to a ς_i .

$$\varsigma_i = ln(u_i^2)$$

$$u_i^2 = [EXP(\varsigma_i)] \tag{9.43}$$

With iteration on real valued ς_i , no uniqueness, u_i^2 , can become negative. A limit can be placed on ς_i not to be less than some negative number such as -10. The first and second derivatives of function F with respect to elements ς_i are obtained and used in a Newton-Raphson iteration.

Steps in the developments of the maximum likelihood solution will be illustrated with a small simulated sample covariance matrix. This sample covariance matrix was generated using the Tucker-Koopman-Linn procedure discussed in Chapter 3 with the generating parameters given in Table 9.4. The simulated covariance matrix for a sample of 200 is given in Table 9.5. This is considered as matrix $\boldsymbol{C_{yy}}$. First trial u_i^2 are given in Table 9.4 along with the corresponding ς_i . These initial u_i^2 are given by:

Table 9.4

Generating Parameters for Illustrative
Sample Covariance Matrix

	Major Domain Factor Weight	Unique Variance	Minor Domain Variance
1	15	30	45
2	20	45	30
3	10	15	15
4	25	70	65
5	12	15	10

$\varepsilon = .05$ Number of Minor Factors = 135
Sample Size = 200

Table 9.5

Illustrative Sample Covariance Matrix, C_{yy}
N = 200

	1	2	3	4	5
1	277.6				
2	307.8	496.1			
3	148.0	215.4	131.4		
4	381.3	522.5	266.2	772.0	
5	178.0	264.8	122.1	314.6	177.8

$$u_i^2 = 1/c^{ii} \tag{9.44}$$

where c^{ii} is the i'th diagonal entry of $\boldsymbol{C}_{yy}^{-1}$.

The first step in the development is to obtain the solution for the factor matrix conditional on a matrix $\boldsymbol{U^2}$. Partial derivatives of function F of equation (9.34) with respect to the entries in $\boldsymbol{A}$. Reference will be made to $\widehat{\boldsymbol{\Sigma}}_m$ as given in equation (9.35).

$$\begin{aligned} \partial F/\partial a_{\text{jk}} &= 2(\widehat{\boldsymbol{\Sigma}}_m^{-1})_j' \boldsymbol{A_k} - 2(\widehat{\boldsymbol{\Sigma}}_M^{-1})_j' \boldsymbol{C}_{yy}(\widehat{\boldsymbol{\Sigma}}_m^{-1})\boldsymbol{A_k} \\ &= 2(\widehat{\boldsymbol{\Sigma}}_m^{-1})_j' (\boldsymbol{I} - \boldsymbol{C}_{yy}\widehat{\boldsymbol{\Sigma}}_m^{-1})\boldsymbol{A_k}\,. \end{aligned} \tag{9.45}$$

$(\widehat{\boldsymbol{\Sigma}}_m^{-1})_j$ is the j'th column of $(\widehat{\boldsymbol{\Sigma}}_m^{-1})$; $\boldsymbol{A_k}$ is the k'th column of $\boldsymbol{A}$. When the derivatives for all entries in $\boldsymbol{A}$ are set to zero the following result is obtained:

$$2(\widehat{\boldsymbol{\Sigma}}_m^{-1})(\boldsymbol{I} - \boldsymbol{C}_{yy}\widehat{\boldsymbol{\Sigma}}_m^{-1})\boldsymbol{A} = \boldsymbol{0}\,.$$

Canceling out the first $2(\widehat{\boldsymbol{\Sigma}}_m^{-1})$ yields:

$$(\boldsymbol{I} - \boldsymbol{C}_{yy}\widehat{\boldsymbol{\Sigma}}_m^{-1})\boldsymbol{A} = \boldsymbol{0}\ . \tag{9.46}$$

Following is a most interesting relation which is used in equation (9.45).

$$\widehat{\boldsymbol{\Sigma}}_m^{-1} = \boldsymbol{U^{-2}} - \boldsymbol{U^{-2}A}(\boldsymbol{I} + \boldsymbol{A'U^{-2}A})^{-1}\boldsymbol{A'U^{-2}} \tag{9.47}$$

This relation substituted into equation (9.46) yields:

$$\begin{aligned} (\boldsymbol{I} - \boldsymbol{C}_{yy}\widehat{\boldsymbol{\Sigma}}_m^{-1})\boldsymbol{A} &= [\boldsymbol{I} - \boldsymbol{C}_{yy}\boldsymbol{U^{-2}} + \boldsymbol{C}_{yy}\boldsymbol{U^{-2}A}(\boldsymbol{I} + \boldsymbol{A'U^{-2}A})^{-1}\boldsymbol{A'U^{-2}}]\boldsymbol{A} \\ &= \boldsymbol{0}\,. \end{aligned} \tag{9.48}$$

Algebraic operations yield

$$\boldsymbol{C}_{yy}\boldsymbol{U^{-2}A}(\boldsymbol{I} + \boldsymbol{A'U^{-2}A})^{-1} = \boldsymbol{A}$$

or:

$$\boldsymbol{C}_{yy}\boldsymbol{U^{-2}A} = \boldsymbol{A}(\boldsymbol{I} + \boldsymbol{A'U^{-2}A})$$

and:

$$(\boldsymbol{U^{-1}C}_{yy}\boldsymbol{U^{-1}})(\boldsymbol{U^{-1}A}) = (\boldsymbol{U^{-1}A})(\boldsymbol{I} + \boldsymbol{A'U^{-2}A}) \tag{9.49}$$

Given any situation, matrix $\boldsymbol{A}$ may be orthogonally transformed by columns such that $\boldsymbol{A'U^{-2}A}$ is an $r \times r$ diagonal matrix as is, also, $(\boldsymbol{I} + \boldsymbol{A'U^{-2}A})$. With:

$$(\boldsymbol{I} + \boldsymbol{A'U^{-2}A}) = \boldsymbol{Q}_r \tag{9.50}$$

equation (9.49) becomes:

$$(\boldsymbol{U}^{-1}\boldsymbol{C}_{yy}\boldsymbol{U}^{-1})(\boldsymbol{U}^{-1}\boldsymbol{A}) = (\boldsymbol{U}^{-1}\boldsymbol{A})\boldsymbol{Q}_r \ . \tag{9.51}$$

This equation is in the form for an eigen solution with eigenvalues $\boldsymbol{Q}_r$ and eigenvectors $(\boldsymbol{U}^{-1}\boldsymbol{A})$. These eigenvectors are not of unit length but have a length to satisfy equation (9.50). The solution for matrix $\boldsymbol{A}$ is given in equations (9.36) and (9.37).

Instead of using the conditional solution for $\boldsymbol{U}^2$ dependent on the factor matrix as was done for the alternating solution outlined earlier, and expression for function F is derived dependent upon $\boldsymbol{U}^2$ and the eigenvalues of $\boldsymbol{U}^{-1}\boldsymbol{C}_{yy}\boldsymbol{U}^{-1}$. Then, the derivatives of F with respect to elements of $\boldsymbol{U}^2$ are obtained so as to utilize the gradient and Newton-Raphson iterative procedures. The first step is to substitute the results in equation (9.37) into equation (9.35).

$$\begin{aligned} \widehat{\boldsymbol{\Sigma}}_m &= \boldsymbol{U}\boldsymbol{V}_r(\boldsymbol{Q}_r - \boldsymbol{I})\boldsymbol{V}_r'\boldsymbol{U} + \boldsymbol{U}^2 \\ &= \boldsymbol{U}[\boldsymbol{V}_r(\boldsymbol{Q}_r - \boldsymbol{I})\boldsymbol{V}_r' + \boldsymbol{I}]]\boldsymbol{U} \end{aligned} \tag{9.52}$$

It is convenient to define an $n \times n$ diagonal matrix $\boldsymbol{Q}_{rI}$.

$$\boldsymbol{Q}_{rI} = \begin{bmatrix} \boldsymbol{Q}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \tag{9.53}$$

The full, $n \times n$, matrix of eigenvalues is partitioned into matrix $\boldsymbol{V}_r$, $n \times r$, and a second section $\boldsymbol{V}_2$, $n \times (n - r)$.

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{V}_r & \boldsymbol{V}_2 \end{bmatrix} \tag{9.54}$$

The matrix difference $(\boldsymbol{Q}_{\mathbf{r}\mathbf{I}} - \boldsymbol{I})$ is shown below.

$$\begin{bmatrix} \boldsymbol{Q}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} - \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} = \begin{bmatrix} (\boldsymbol{Q}_r - \boldsymbol{I}) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \tag{9.55}$$

With this result:

$$\boldsymbol{V}_r(\boldsymbol{Q}_r - \boldsymbol{I})\boldsymbol{V}_r' = \boldsymbol{V}(\boldsymbol{Q}_{rI} - \boldsymbol{I})\boldsymbol{V}' \tag{9.56}$$

which is important since it involves the complete matrix $\boldsymbol{V}$ of eigenvectors. Then from equation (9.52):

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}}_m &= \boldsymbol{U}[\boldsymbol{V}(\boldsymbol{Q}_{rI} - \boldsymbol{I})\boldsymbol{V}' + \boldsymbol{I}]\boldsymbol{U} \\ &= \boldsymbol{U}[\boldsymbol{V}\boldsymbol{Q}_{rI}\boldsymbol{V}']\boldsymbol{U} \ . \end{aligned} \tag{9.57}$$

Functions of $\widehat{\boldsymbol{\Sigma}}_m$ are to be obtained to be entered into the equation for function F in equation (9.34). First is the determinant of $\widehat{\boldsymbol{\Sigma}}_m$.

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}}_m &= |\boldsymbol{U}|\bullet|\boldsymbol{V}\boldsymbol{Q}_{rI}\boldsymbol{V}'|\bullet|\boldsymbol{U}| = |\boldsymbol{U}|^2\bullet|\boldsymbol{V}|\bullet|\boldsymbol{Q}_{rI}|\bullet|\boldsymbol{V}'| \\ &= |\boldsymbol{U}^2|\bullet|\boldsymbol{Q}_{rI}|\end{aligned}$$

So that:

$$ln|\widehat{\boldsymbol{\Sigma}}_m| = ln|\boldsymbol{U}^2| + \sum_{k=1}^{r} ln(q_k) \ . \tag{9.58}$$

Next is the trace of $\widehat{\boldsymbol{\Sigma}}_m^{-1}\boldsymbol{C}_{yy}$. Note that:

$$\widehat{\boldsymbol{\Sigma}}_m^{-1} = \boldsymbol{U}^{-1}[\boldsymbol{V}\boldsymbol{Q}_{rI}^{-1}\boldsymbol{V}']\boldsymbol{U}^{-1}$$

so that:

$$\widehat{\boldsymbol{\Sigma}}_m^{-1}\boldsymbol{C}_{yy} = \boldsymbol{U}^{-1}[\boldsymbol{V}\boldsymbol{Q}_{rI}^{-1}\boldsymbol{V}']\boldsymbol{U}^{-1}\boldsymbol{C}_{yy} \ . \tag{9.59}$$

From equation (9.36):

$$\boldsymbol{U}^{-1}\boldsymbol{C}_{yy} = \boldsymbol{V}\boldsymbol{Q}\boldsymbol{V}'\boldsymbol{U}$$

so that:

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}}_m^{-1}\boldsymbol{C}_{yy} &= \boldsymbol{U}^{-1}[\boldsymbol{V}\boldsymbol{Q}_{rI}^{-1}\boldsymbol{V}']\boldsymbol{V}\boldsymbol{Q}\boldsymbol{V}'\boldsymbol{U} \\ &= \boldsymbol{U}^{-1}\boldsymbol{V}\boldsymbol{Q}_{rI}^{-1}\boldsymbol{Q}\boldsymbol{V}'\boldsymbol{U} \ .\end{aligned}$$

From the above with algebraic operations:

$$\begin{aligned}tr(\widehat{\boldsymbol{\Sigma}}_m^{-1}\boldsymbol{C}_{yy}) &= tr(\boldsymbol{U}^{-1}\boldsymbol{V}\boldsymbol{Q}_{rI}^{-1}\boldsymbol{Q}\boldsymbol{V}'\boldsymbol{U}) \\ &= tr(\boldsymbol{Q}_{rI}^{-1}\boldsymbol{Q}) \ .\end{aligned} \tag{9.60}$$

Following is an illustration of the product $\boldsymbol{Q}_{rI}^{-1}\boldsymbol{Q}$.

$$\begin{bmatrix} \boldsymbol{Q}_r^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{Q}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_2 \end{bmatrix}$$

Diagonal matrix $\boldsymbol{Q}_2$ contains the last $(n - r)$ eigenvalues. Then:

$$tr(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{m}}^{-1}\boldsymbol{C_{yy}}) = r + \sum_{k=r+1}^{n} q_k \ . \tag{9.61}$$

An expression for $ln\left|\boldsymbol{C_{yy}}\right|$ is needed. From equation (9.36):

$$\boldsymbol{C_{yy}} = \boldsymbol{UVQV'U}$$

so that:

$$\left|\boldsymbol{C_{yy}}\right| = \left|\boldsymbol{U}\right|\bullet\left|\boldsymbol{V}\right|\bullet\left|\boldsymbol{Q}\right|\bullet\left|\boldsymbol{V'}\right|\bullet\left|\boldsymbol{U}\right|$$

Then:

$$\begin{aligned} ln\left|\boldsymbol{C_{yy}}\right| &= ln\left|\boldsymbol{U^2}\right| + ln\left|\boldsymbol{Q}\right| \\ &= ln\left|\boldsymbol{U^2}\right| + \sum_{k=1}^{n} ln(q_k) \ . \end{aligned} \tag{9.62}$$

A reduced expression for function F is obtained by substitution into equation (9.34) the results given in equations (9.58), (9.61) and (9.62):

$$F = ln\left|\boldsymbol{U^2}\right| + \sum_{k=1}^{r} ln(q_k) + r + \sum_{k=r+1}^{n} q_k - ln\left|U^2\right| - \sum_{k=1}^{n} ln(q_k) - n$$

Which reduces to:

$$F = \sum_{k=r+1}^{n} \left[q_k - ln(q_k)\right] - (n - r) \ . \tag{9.63}$$

This is a most interesting and important representation of function F. It provides the basis for further mathematical developments of the derivatives of F with respect to the uniqueness u_i^2.

Operations to this point in the developments are illustrated for the example in Table 9.6 and 9.7 for the first trial of the iterative procedure. As discussed earlier, the initial trial uniqueness are given in the first column of Table 9.6. Matrix $\boldsymbol{U^{-1}C_{yy}U^{-1}}$ is given in Table 9.7 along with its eigenvectors and eigenvalues (these being the q_k). The illustrative solution is for a single factor so that r = 1. Substitution of the last four eigenvalues into equation (9.63) yields the value of function F given at the bottom of Table 9.6.

To obtain the derivatives of F with respect to the u_i, the derivatives of the eigenvalues, q_k, and elements, v_{ik} , of the eigenvectors are needed. Jöreskog and Goldberger (1972, see

Appendix A2) developed the derivatives for matrix $\boldsymbol{U}\boldsymbol{C}_{\boldsymbol{yy}}^{-1}\boldsymbol{U}$. An analogous development for matrix $\boldsymbol{U}^{-1}\boldsymbol{C}_{\boldsymbol{yy}}\boldsymbol{U}^{-1}$ yields:

$$(\partial q_k/\partial u_i) = (-2q_k/u_i)v_{ik}^2 \tag{9.64}$$

$$(\partial v_{ik}/\partial u_i) = (-1/u_i)v_{ik}\sum_{m\neq k}^{n}[(q_k+q_m)/(q_k-q_m)]v_{im}^2 \ . \tag{9.65}$$

The partial derivative of F with respect to u_i follows.

$$(\partial F/\partial u_i) = \sum_{k=r+1}^{n}\left[(\partial q_k/\partial u_i) - (1/q_k)(\partial q_k/\partial u_i)\right] \ .$$

This equation with equation (9.64) yields:

$$(\partial F/\partial u_i) = (-2/u_i)\{\sum_{k=r+1}^{n} q_k v_{ik}^2 - \sum_{k=r+1}^{n} v_{ik}^2\} \ . \tag{9.66}$$

For subsequent steps equation (9.66) is converted to involving only the first r eigenvalues and eigenvectors. Note that:

$$\sum_{k=1}^{r} q_k v_{ik}^2 + \sum_{k=r+1}^{n} q_k v_{ik}^2 = \sum_{k=1}^{n} q_k v_{ik}^2 = c_{ii}u_i^{-2}$$

so that:

$$\sum_{k=r+1}^{n} q_k v_{ik}^2 = c_{ii}u_i^{-2} - \sum_{k=1}^{r} q_k v_{ik}^2 \ .$$

Also:

$$\sum_{k=1}^{r} v_{ik}^2 + \sum_{k=r+1}^{n} v_{ik}^2 = \sum_{k=1}^{n} v_{ik}^2 = 1$$

so that:

$$\sum_{k=r+1}^{n} v_{ik}^2 = 1 - \sum_{k=1}^{r} v_{ik}^2 \ .$$

With the preceding results equation (9.66) becomes:

$$(\partial F/\partial u_i) = (-2/u_i)\{c_{ii}u_i^{-2} - \sum_{k=1}^{r}(q_r - 1)v_{ik}^2 - 1\} \ . \tag{9.67}$$

Table 9.6
Trial 1 Uniqueness and Zeta

	Uniqueness	Zeta (ζ)
1	70.704	4.259
2	70.642	4.258
3	30.639	3.422
4	143.727	4.968
5	30.416	3.415

Function F = .15477

Table 9.7

Trial 1 Matrix ($U^{-1}C_{yy}U^{-1}$) with Eigen Solution

$U^{-1}C_{yy}U^{-1}$

	1	2	3	4	5
1	3.926	4.355	3.180	3.782	3.838
2	4.355	7.023	4.630	5.185	5.713
3	3.180	4.630	4.289	4.011	4.000
4	3.782	5.185	4.011	5.371	4.758
5	3.838	5.713	4.000	4.758	5.846

Eigenvectors

	1	2	3	4	5
1	.371	.254	.675	.527	.256
2	.529	-.424	-.233	.405	-.567
3	.391	.467	-.665	.157	.402
4	.451	.473	.197	-.588	-.434
5	.475	-.580	.094	-.434	.514

Eigenvalues

1	2	3	4	5
23.037	1.078	.929	.845	.565

The diagonal cells of the second derivatives matrix are handled separately from the off-diagonal cells.

$$\partial^2 F/\partial u_i^2 = (2/u_i^2)\{c_{ii}u_i^{-2} - \sum_{k=1}^{r}(q_k - 1)v_{ik}^2 - 1\}$$
$$+ (2/u_i)\{2c_{ii}u_i^{-3} + \sum_{k=1}^{r} v_{ik}^2(\partial q_k/\partial u_i)$$
$$+ \sum_{k=1}^{r}(q_k - 1)v_{ik}^2(\partial v_{ik}/\partial u_i)\} \qquad (9.68)$$

Substitution from equations (9.64) and (9.65) into equation (9.68) yields with algebraic operations:

$$\partial^2 F/\partial u_i^2 = (2/u_i^2)\{c_{ii}u_i^{-2} - \sum_{k=1}^{r}(q_k - 1)v_{ik}^2 - 1\}$$
$$+ (4/u_i^2)\{c_{ii}u_i^{-2} - \sum_{k=1}^{r} q_k v_{ik}^2$$
$$- \sum_{k=1}^{r}(q_k - 1)v_{ik}^2 \sum_{m\neq k}^{n}[(q_k + q_m)/(q_k - q_m)]v_{im}^2\} \ . \qquad (9.69)$$

For the off-diagonal entrees in the second derivative matrix, i $\neq$ j:

$$\partial^2 F/\partial u_i \partial u_j = (-2/u_i)\{-\sum_{k=1}^{r} v_{ik}^2(\partial q_k/\partial u_j)$$
$$- 2\sum_{k=1}^{r}(q_k - 1)v_{ik}(\partial v_{ik}/\partial u_j)\} \qquad (9.70)$$

Substitution from equations (9.64) and (9.65) into equation (9.70) yields with algebraic operations:

$$\partial^2 F/\partial u_i \partial u_j = (-4/u_i u_j)\{\sum_{k=1}^{r} q_k v_{ik}^2 v_{jk}^2$$
$$+ \sum_{k=1}^{r}(q_k - 1)v_{ik}v_{jk}\sum_{m\neq k}^{n}[(q_k + q_m)/(q_k - q_m)]v_{im}v_{jm}\} \qquad (9.71)$$

In order to avoid problems involving negative u_i^2 values a transformation of variables from u_i^2 to ς_i indicated in equation (9.43) is taken. The needed derivatives to accomplish this transformation follow.

$$\partial F/\partial \varsigma_i = (u_i/2)(\partial F/\partial u_i) \ ; \qquad (9.72)$$

$$\partial^2 F/\partial \varsigma_i^2 = (u_i/4)(\partial F/\partial u_i) + (u_i^2/4)(\partial^2 F/\partial u_i^2) \; ; \tag{9.73}$$

$$\partial^2 F/\partial \varsigma_i \partial \varsigma_j = (u_i u_j/4)(\partial^2 F/\partial u_i \partial u_j) \qquad \text{for } i \neq j \; . \tag{9.74}$$

For the first derivative of F with respect to ς_i , equation (9.72) is used with equations (9.67) to obtain:

$$\partial F/\partial \varsigma_i = 1 + \sum_{k=1}^{r} (q_i - 1) v_{ik}^2 - c_{ii}^{-2} u_i \; . \tag{9.75}$$

For the diagonal entries in the second derivative matrix, equation (9.73) is used with equation (9.67) and (9.69) to obtain:

$$\begin{aligned} \partial^2 F/\partial \varsigma_i^2 = & -\sum_{k=1}^{r} q_k v_{ik}^4 - \sum_{k=1}^{r} (q_k - 1) v_{ik}^2 \sum_{m \neq k}^{n} [(q_k + q_m)/(q_k - q_m)] v_{im}^2 \\ & + c_{ii} u_i^2 \; . \end{aligned} \tag{9.76}$$

For the off-diagonal entries in the second derivative matrix, equation (9.74) is used with equation (9.71) to obtain, i ≠ j :

$$\begin{aligned} \partial^2 F/\partial \varsigma_i \partial \varsigma_j = & -\sum_{k=1}^{r} q_k v_{ik}^2 v_{jk}^2 \\ & - \sum_{k=1}^{r} (q_k - 1) v_{ik} v_{jk} \sum_{m \neq k}^{n} [(q_k + q_m)/(q_k - q_m)] v_{im} v_{jm} \; . \end{aligned} \tag{9.77}$$

Equation (9.75) gives the gradient for the gradient method of iteration while equations (9.76) and (9.77) give the second derivative matrix used in Newton-Raphson method of iteration. For the example, Table 9.8 presents these derivatives for the first trial for one factor. Note that the determinant of the second derivative matrix is positive which is necessary for a minimum solution.

Operation of the Newton-Raphson procedure involves solution of the linear equations indicated in equation (9.42). Table 9.9 gives the inverse of the second derivative matrix and the change vector, $-\boldsymbol{H}_1^{-1}\mathbf{g}_1$, for the first trial of the example. The Newton-Raphson iteration took six trials to converge for the one factor for the example. Progress over these trials is outlined in Table 9.10 which gives the value of the function F for each of these trials and the first derivative vector for each trial. The change in function F from trial to trial diminishes as the trials progress and, finally, shows no change to five decimals from trial 5 to trial 6. Entries in the first derivative vectors diminish, also, from trial to trial with all entries becoming zero to five decimal places at trial 6. The convergence test involves the entries in the first derivative vector. Convergence may

Table 9.8

Trial 1 Derivatives of Function F

	First	Second Derivative Matrix*				
	Derivative	1	2	3	4	5
1	.099	.674	.026	.026	-.005	.044
2	.155	.026	.452	.017	.097	-.027
3	.088	.026	.017	.657	-.011	.071
4	.121	-.005	.097	-.011	.564	.031
5	.120	.044	-.027	.071	.031	.535

Determinant of second derivative matrix = .056 .

Table 9.9

Trial 1 Derivatives Matrix Inverse
with change Vector for Zeta

	Change	Second Derivative Matrix Inverse				
	Vector	1	2	3	4	5
1	-.055	1.498	-.099	-.042	.037	-.126
2	-.102	-.099	2.319	-.079	-.412	.159
3	-.049	-.042	-.079	1.550	.056	-.209
4	-.070	.037	-.412	.056	1.854	-.138
5	-.075	-.126	.159	-.209	-.138	1.922

Table 9.10

Outline of Progress over Trials

Function F over Trials

1	2	3	4	5	6
.09872	.05343	.01014	.00038	.00000	.00000

First Derivative Vectors Over Trials

	1	2	3	4	5	6
1	.09872	.05343	.01014	.00038	.00000	.00000
2	.15481	.10005	.03060	.00353	.00015	.00000
3	.08828	.04741	.00880	.00029	-.00001	.00000
4	.12070	.06777	.01248	.00018	-.00003	.00000
5	.12004	.073.4	.01887	.00150	.00004	.00000

Table 9.11

Maximum Likelihood Results for One Factor

	Factor Matrix 1	Communality	Uniqueness
1	14.6	214.2	63.4
2	21.1	443.5	52.6
3	10.2	103.5	27.9
4	25.4	646.3	125.7
5	12.4	152.6	25.2

$\xi\ (\chi^2) = 19.9$ df = 5 p = .001

Tucker-Lewis Reliability = .974

be declared when every entry in the first derivative vector is less in absolute value than some preset small value.

The value of function F should be checked for each Newton-Raphson trial. In case the value of F does not decrease, the trail should be discarded and one or more gradient method trials attempted. In case the value of F does not decrease on a gradient trial, this trail should be repeated with a smaller step size. When a decrease in F has been obtained the Newton-Raphson iteration may resumed.

Table 9.11 presents the solution for one factor for the example. For each attribute, the communality plus the uniqueness equals the variance of that attribute in Table 9.5. Coefficient ξ uses a revised value for the number of cases, this revised number of cases which was suggested by Bartlett (1950) is given in equation (9.18) and termed $N_r^\star$. From equation (9.33) using $N_r^\star$

$$\xi = N_r^\star F \tag{9.78}$$

The degree of freedom, $\nu(\xi)$, for ξ interpreted as a chi square involves the number of parameters used in the theoretic model, this number being the number of uniqueness plus the number of entries in the factor matrix less the number of entries which may be set to zero by an orthogonal transformation of the factors.

$$\eta(m) = n + nr - r(r-1)/2\,. \tag{9.79}$$

Substitution of this value along with the value of η(M) from equation (9.28) into equation (9.27) yields:

$$\nu(\xi) = n(n+1)/2 - n - nr + r(r-1)/2\ . \tag{9.80}$$

The value of ξ and $\nu(\xi)$ for the example for one factor is given in Table 9.11 along with the corresponding p for ξ as a chi square with the given number of degrees of freedom. With p equaling .001 the statistical hypothesis that a one factor solution would fit in the population should be rejected. This is not an unexpected result since the generation of the population correlation matrix included a number of minor factors. However, there is a question whether the one factor solution should be considered as satisfactory since only one major domain factor was used in the generation of the simulated covariance matrix. In many applications of the maximum likelihood factor method to real data there is a problem concerning acceptance of a solution when the statistical test indicates that the solution should be rejected. Any decision on the number of common factors to use should consider a number of items of information, not just the statistical test.

The Tucker-Lewis reliability coefficient was discussed earlier. For the present application, the value of M_0 for no common factors is required. From equation (9.38)

$$\boldsymbol{U}_0^2 = diag(\boldsymbol{C_{yy}})$$

so that from equation (9.35)

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{m0}} = \boldsymbol{U}_0^{\boldsymbol{2}} \ .$$

With this result substitution into equation (9.32) yields with algebraic operations

$$F_0 = ln\left|\boldsymbol{U}_0^{\boldsymbol{2}}\right| - ln\left|\boldsymbol{C_{yy}}\right| = \sum_{i=1}^{n} ln(c_{ii}) - ln\left|\boldsymbol{C_{yy}}\right| \tag{9.81}$$

where c_{ii} are the diagonal entries in $\boldsymbol{C_{yy}}$. The degrees of freedom for zero common factors is:

$$\nu(\xi)_0 = n(n-1)/2 \ . \tag{9.82}$$

By equation (9.14):

$$M_0 = F_0/\nu(\xi)_0 \ . \tag{9.83}$$

Also, from equation (9.14), for r common factors:

$$M_r = F_r/\nu(\xi)_r \ . \tag{9.84}$$

The Tucker-Lewis reliability coefficient is given by equation (9.19), repeated here for convenience.

$$\rho_r \doteq \frac{M_0 - M_r}{M_0 - 1/N_r^{\bigstar}} \tag{9.19}$$

The Tucker-Lewis reliability for the example for one factor is given at the bottom of Table 9.11. This value of .974 indicates a quite good fit of the factor results. Experience has indicated that this reliability coefficient is a very useful measure of goodness of fit but is not of much use in deciding how many factors to be extracted. The number of factors should be decided on other grounds and, then, the reliability coefficient used to indicate the quality of the results.

A two factor solution is given in Table 9.12. While the first factor is quite close to the major domain loadings in Table 9.4 the second factor loadings are smaller and may have been introduced by some combination of the minor domain factors. Note that the uniqueness of attribute 2 is very small. This illustrates a frequent result obtained with maximum likelihood factor analysis: one or more of the uniqueness iterates to very small values and, in fact, for solutions not using the transformation from u_i^2 to ς_i the iterations tend to obtaining negative values of u^2 . During the iterations, on trial 3, the value of the function increased so that one gradient method trial was required. After this, the movement was slow so that a total of 26 trials was required before convergence was obtained. The chi square statistic yielded a p of .422 so that

Table 9.12

Maximum Likelihood Results for Two Factors

	Factor Matrix		Communality	Uniqueness
	1	2		
1	14.5	.3	211.3	66.3
2	21.3	-6.5	496.1	.0
3	10.1	.1	102.7	28.7
4	26.2	5.5	715.3	56.7
5	12.2	-.8	149.0	28.8

ξ (χ^2) = .64 df = 1 p = .422

Tucker-Lewis Reliability = 1.003

the two factor solution would not be rejected. There is a problem, however, concerning subsequent factor transformations for the two factor matrix. Maybe, the second factor should be ignored as a small 'residual' dimension. This would result in accepting the first factor as it stands. Note that the first factor is very similar to the single factor for the one factor solution given in Table 9.11 so that, maybe, the one factor solution should be accepted even though the statistical test indicated that it should be rejected. The Tucker-Lewis reliability coefficient for the two factor solution is slightly greater than unity, a result which can be obtained due to the approximations involved with this coefficient.

9.2. Generalized Least Squares Factor Analysis

Jöreskog and Goldberger (1972) introduced generalized least squares factor analysis as a second statistical method for factor analysis which they based on Aitken's (1934-35) generalized least squares principle. Application of this principle chooses parameter estimates to minimize the function

$$G_p = \frac{1}{2} tr\{[\boldsymbol{\Sigma}_{yy}^{-1}(\boldsymbol{C}_{yy} - \boldsymbol{\Sigma}_{zz})]^2\} \tag{9.85}$$

where, as previously, $\boldsymbol{\Sigma}_{yy}$ is the population covariance matrix among the observed attributes, $\boldsymbol{C}_{yy}$ is the observed covariance matrix in a sample and $\boldsymbol{\Sigma}_{zz}$ is the modeled population covariance matrix. However, $\boldsymbol{\Sigma}_{yy}$ is not available so that the unbiased estimate $\boldsymbol{C}_{yy}$ is substituted to yield

$$\begin{aligned} G &= \frac{1}{2} tr\{[\boldsymbol{C}_{yy}^{-1}(\boldsymbol{C}_{yy} - \boldsymbol{\Sigma}_{zz})]^2\} \\ &= \frac{1}{2} tr\{(I - \boldsymbol{C}_{yy}^{-1}\boldsymbol{\Sigma}_{zz})^2\} \end{aligned} \tag{9.86}$$

Note that the function of (9.85) is an application of the general function given in equation (9.1) with $\boldsymbol{\Sigma}_{yy}^{-1}$ being the weight matrix $\boldsymbol{W}$. Since $\boldsymbol{\Sigma}_{yy}^{-1}$ and $\boldsymbol{C}_{yy}^{-1}$ are symmetric, function G represents a true least squares.

Application of the Jöreskog & Goldberger (1972) general condition for scale freeness given in equation (9.10) demonstrates that the generalized least squares criterion is scale free. The previous discussion of the scale free condition involved matrices $\boldsymbol{C}_{yy}$ and $\boldsymbol{C}_{zz}$. For the present case matrix $\boldsymbol{\Sigma}_{zz}$ is substituted for matrix $\boldsymbol{C}_{zz}$. Then, steps similar to those in equations (9.5) through (9.11) are followed to demonstrate the scale freeness of function G .

A number of relations exist between maximum likelihood factor analysis and generalized least squares factor analysis. These relations will be pointed out during discussion of the derivations that follow. A first relation is an approximate relation derived by Jöreskog &

Goldberger between the maximum likelihood function F and the generalized least squares function G . They show that:

$$F \approx G \tag{9.87}$$

which is important in the statistical distribution associated with G . Thus, coefficient may be defined for generalized least squares factor analysis as:

$$\xi = N_r^{\bigstar} G \tag{9.88}$$

where the value of $N_r^{\bigstar}$ is given in equation (9.18). This definition parallels equation (9.78) for maximum likelihood factor analysis. With normality assumptions and a minimum G , ξ is distributed approximately as χ^2 with $\nu(\xi)$ degrees of freedom, see equation (9.80) for $\nu(\xi)$.

Solution for minimum G involves two steps, first is a conditional solution for the estimated factor matrix, $\boldsymbol{A}$, dependent on values of the uniqueness, second is an overall solution for the estimated uniqueness matrix $\boldsymbol{U}$. In the present a switch is being made from the population factor matrix and uniqueness to estimates derived from the sample data. The sample factor model is given in equation (9.2). With this switch the function G is written as:

$$G = \frac{1}{2} tr\{(\boldsymbol{I} - \boldsymbol{C}_{yy}^{-1}\boldsymbol{C}_{zz})^2\} \ . \tag{9.89}$$

To obtain the conditional minimum for G dependent upon values of the uniqueness a partial derivative of G with respect to the factor matrix $\boldsymbol{A}$ is obtained:

$$\frac{\partial G}{\partial \boldsymbol{A}} = 2\boldsymbol{C}_{yy}^{-1}(\boldsymbol{C}_{zz} - \boldsymbol{C}_{yy})\boldsymbol{C}_{yy}^{-1}\boldsymbol{A} \ . \tag{9.90}$$

This derivative was developed in their Appendix A1. Setting ∂G/$\partial \boldsymbol{A}$ equal to zero with algebraic steps:

$$\boldsymbol{C}_{yy}^{-1}\boldsymbol{A} = \boldsymbol{C}_{zz}^{-1}\boldsymbol{A} \ . \tag{9.91}$$

Using equation (9.47) with algebraic steps:

$$(\boldsymbol{U}\boldsymbol{C}_{yy}^{-1}\boldsymbol{U})(\boldsymbol{U}^{-1}\boldsymbol{A}) = (\boldsymbol{U}^{-1}\boldsymbol{A})(\boldsymbol{I} + \boldsymbol{A}'\boldsymbol{U}^{-2}\boldsymbol{A})^{-1} \tag{9.92}$$

The similarity of this equation to equation (9.49) for maximum likelihood factor analysis is to be noted. With an orthogonal transformation of the factor matrix $\boldsymbol{A}$

$$(\boldsymbol{I} + \boldsymbol{A}'\boldsymbol{U}^{-2}\boldsymbol{A})^{-1} = \boldsymbol{\theta}_1 \tag{9.93}$$

where $\boldsymbol{\theta}_1$ is an $r \times r$ diagonal matrix. Comparison of this definition with equation (9.50) for maximum likelihood indicates that $\boldsymbol{\theta}_1$ is the inverse of $\mathbf{Q}_r$. The form of equation (9.92) indicates

that $\boldsymbol{\theta}_1$ contains eigenvalues of $\boldsymbol{U}\boldsymbol{C}_{yy}^{-1}\boldsymbol{U}$ and that $(\boldsymbol{U}^{-1}\boldsymbol{A})$ contains the corresponding eigenvectors scaled according to equation (9.93). A complete eigen solution of $\boldsymbol{U}\boldsymbol{C}_{yy}^{-1}\boldsymbol{U}$ yields:

$$(\boldsymbol{U}\boldsymbol{C}_{yy}^{-1}\boldsymbol{U}) = \begin{bmatrix} \boldsymbol{V}_1 & \boldsymbol{V}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\theta}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1' \\ \boldsymbol{V}_2' \end{bmatrix} \tag{9.94}$$

with $\begin{bmatrix} \boldsymbol{V}_1 & \boldsymbol{V}_2 \end{bmatrix}$ containing all n eigenvectors as unit vectors and is $n \times n$ orthonormal so that:

$$\begin{bmatrix} \boldsymbol{V}_1' \\ \boldsymbol{V}_2' \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1 & \boldsymbol{V}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I} \end{bmatrix} \tag{9.95}$$

and

$$\begin{bmatrix} \boldsymbol{V}_1 & \boldsymbol{V}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1' \\ \boldsymbol{V}_2' \end{bmatrix} = \boldsymbol{V}_1\boldsymbol{V}_1' + \boldsymbol{V}_2\boldsymbol{V}_2' = \boldsymbol{I} \tag{9.96}$$

$\boldsymbol{\theta}_2$ is $(n - r) \times (n - r)$ diagonal so that $\begin{bmatrix} \boldsymbol{\theta}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\theta}_2 \end{bmatrix}$ contains the n eigenvalues of $\boldsymbol{U}\boldsymbol{C}_{yy}^{-1}\boldsymbol{U}$.
Dimensions of the eigen solution are to be in ascending order of the eigenvalues.

The five attribute example used with the maximum likelihood discussion will be used to illustrate points for the generalized least squares development. The covariance matrix is given in Table 9.5. Table 9.13 gives the inverse of the covariance matrix. Jöreskog & Goldberger recommend refined starting values for the uniqueness. Since the standard errors of estimate, as developed in Chapter 8, are overstatements of the uniqueness, multiplication by a fraction appears to be desirable. The recommended stating values of u_j^2 are:

$$u_j^2 = (1 - r/2n)(1/c^{jj}) \ . \tag{9.97}$$

where c^{jj} is the j'th diagonal entry in $\boldsymbol{C}_{yy}^{-1}$. When comparing this formula with equation (9.44), note that a multiplying fraction of (1 − r/2n) is used. The initial uniqueness and values of ς are given in Table 9.14. Note that these uniqueness are .9 times the values in Table 9.6. Table 9.15 gives matrix $\boldsymbol{U}\boldsymbol{C}_{yy}^{-1}\boldsymbol{U}$.

There are close relations of generalized least square factor analysis to maximum likelihood factor analysis. For an area of relations note, for a given matrix $\boldsymbol{U}$, that:

$$(\boldsymbol{U}\boldsymbol{C}_{yy}^{-1}\boldsymbol{U})^{-1} = \boldsymbol{U}^{-1}\boldsymbol{C}_{yy}\boldsymbol{U}^{-1} \ . \tag{9.98}$$

Table 9.13

Inverse of Illustrative Sample Covariance Matrix, C^{-1}_{yy}

	1	2	3	4	5
1	.01414	-.00373	-.00188	-.00298	-.00203
2	-.00373	.01416	-.00721	-.00072	-.01112
3	-.00188	-.00721	.03264	-.00522	-.00056
4	-.00298	-.00072	-.00522	.000696	-.00467
5	-.00203	-.01112	-.00056	-.00467	.03288

Table 9.14

GLS Trial 1 Uniqueness and Zeta

	Uniqueness	Zeta (ζ)
1	63.333	4.153
2	63.585	4.152
3	27.573	3.317
4	129.345	4.863
5	27.374	3.310

Table 9.15

Trial 1 Matrix ($U^{-1}C_{yy}U^{-1}$) with Eigen Solution

$U^{-1}C_{yy}U^{-1}$

	1	2	3	4	5
1	.900	-.237	-.079	-.271	-.085
2	-.237	.900	-.302	-.065	-.464
3	-.079	-.302	.900	-.312	-.016
4	-.271	-.065	-.312	.900	-.278
5	-.085	-.464	-.016	-.278	.900

Eigenvectors

	1	2	3	4	5
1	.371	.254	.675	.527	.256
2	.529	-.424	-.233	.405	-.567
3	.391	.467	-.665	.157	.402
4	.451	.473	.197	-.588	-.434
5	.475	-.560	.094	-.434	.514

Eigenvalues

1	2	3	4	5
.039	.835	.969	1.065	1.592

The eigen solution of $\boldsymbol{U}^{-1}\boldsymbol{C}_{yy}\boldsymbol{U}^{-1}$ given in equation (9.36) may be written as:

$$\boldsymbol{U}^{-1}\boldsymbol{C}_{yy}\boldsymbol{U}^{-1} = \begin{bmatrix} \boldsymbol{V}_r & \boldsymbol{V}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{Q}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_r' \\ \boldsymbol{V}_2' \end{bmatrix} \tag{9.99}$$

Since the inverse of a matrix has identical eigenvectors as the matrix and the eigenvalues are inverted:

$$\boldsymbol{V}_1 = \boldsymbol{V}_r \tag{9.100}$$

and

$$\begin{bmatrix} \boldsymbol{\theta}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\theta}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{Q}_r^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_2^{-1} \end{bmatrix} \tag{9.101}$$

Note that the eigenvalues of $\boldsymbol{U}\boldsymbol{C}_{yy}^{-1}\boldsymbol{U}$ are in ascending order when the eigenvalues of $\boldsymbol{U}^{-1}\boldsymbol{C}_{yy}\boldsymbol{U}^{-1}$ are in descending order. Table 9.15 gives the eigenvalues and eigenvectors for trial 1 of the GLS factoring for the illustrative example. The eigenvectors given in Table 9.15 are identical to the eigenvectors for $\boldsymbol{U}^{-1}\boldsymbol{C}_{yy}\boldsymbol{U}^{-1}$ given in Table 9.7 for maximum likelihood factoring. However, the eigenvalues in Table 9.15 are .9 times the reciprocals of the eigenvalues in Table 9.7, the multiplier of .9 being due to the proportionality of the uniqueness used in the two analyses.

The conditional solution for the factor matrix $\boldsymbol{A}$ is similar to the conditional solution for maximum likelihood factor analysis. To satisfy equation (9.93)

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{V}_1(\boldsymbol{\theta}_1^{-1} - \boldsymbol{I})^{\frac{1}{2}} \tag{9.102}$$

which is to be compared with equation (9.37).

The next step is to substitute the conditional solution for factor matrix $\boldsymbol{A}$ into the expression for G in equation (9.86) to obtain a relation for G dependent only on $\boldsymbol{U}$. From equation (9.2) and (9.102):

$$\begin{aligned} \boldsymbol{C}_{zz} &= \boldsymbol{U}\boldsymbol{V}_1(\boldsymbol{\theta}_1^{-1} - \boldsymbol{I})^{\frac{1}{2}}(\boldsymbol{\theta}_1^{-1} - \boldsymbol{I})^{\frac{1}{2}}\boldsymbol{V}_1'\boldsymbol{U} + \boldsymbol{U}^2 \\ &= \boldsymbol{U}[\boldsymbol{V}_1(\boldsymbol{\theta}_1^{-1} - \boldsymbol{I})\boldsymbol{V}_1' + \boldsymbol{I}]\boldsymbol{U} \\ &= \boldsymbol{U}[\boldsymbol{V}_1\boldsymbol{\theta}_1^{-1}\boldsymbol{V}_1' - \boldsymbol{V}_1\boldsymbol{V}_1' + \boldsymbol{I}]\boldsymbol{U} \ . \end{aligned}$$

With equation (9.96):

$$\boldsymbol{C_{zz} = U[V_1\theta_1^{-1}V_1' + V_2V_2']U} \quad . \tag{9.103}$$

An expression is required for $\boldsymbol{C_{yy}^{-1}C_{zz}}$ which appears in equation (9.86).

$$\boldsymbol{C_{yy}^{-1}C_{zz} = C_{yy}^{-1}U[v_1\theta_1^{-1}V_1' + V_2V_2']U} \quad .$$

With equation (9.94)

$$\boldsymbol{C_{yy}^{-1}C_{zz} = U^{-1}} \boxed{\begin{matrix} V_1 & V_2 \end{matrix}} \boxed{\begin{matrix} \theta_1 & 0 \\ 0 & \theta_2 \end{matrix}} \boxed{\begin{matrix} V_1' \\ V_2' \end{matrix}} \boxed{V_1\theta_1^{-1}V_1' + V_2V_2'} \boldsymbol{U} \quad .$$

Algebraic manipulation involving equations (9.95) and (9.96) yields:

$$\boldsymbol{C_{yy}^{-1}C_{zz} = U^{-1}[I - V_2(I - \theta_2)V_2']}$$

$$\boldsymbol{(I - C_{yy}^{-1}C_{zz}) = U^{-1}V_2(I - \theta_2)V_2'U}$$

$$\boldsymbol{(I - C_{yy}^{-1}C_{zz})^2 = U^{-1}V_2(I - \theta_2)^2V_2'U} \tag{9.104}$$

Then:

$$\begin{aligned} tr(\boldsymbol{I - C_{yy}^{-1}C_{zz}})^2 &= tr[\boldsymbol{U^{-1}V_2}(\boldsymbol{I - \theta_2})^2\boldsymbol{V_2'U} \\ &= tr(\boldsymbol{I - \theta_2})^2 \\ &= \sum_{m=k+1}^{n} (1 - \theta_m)^2 \\ &= \sum_{m=r+1}^{n} (\theta_m - 1)^2 \end{aligned} \tag{9.105}$$

Substitution into equation (9.89) yields:

$$G = (1/2)\sum_{m=r+1}^{n} (\theta_m - 1)^2 \tag{9.106}$$

This is the reduced function which does not involve the estimated factor matrix. When a solution for $\boldsymbol{U}$ for a minimum G is obtained, the solution for the factor matrix $\boldsymbol{A}$ may be obtained by equation (9.102).

An efficient solution for $\boldsymbol{U}$ uses a Newton-Raphson iterative procedure. To obtain the required derivatives of G with respect to entries in $\boldsymbol{U}$, the derivatives of eigenvalues θ_m and

entries v_{im} in the eigenvectors of $\boldsymbol{U C_{yy}^{-1} U}$ are required. Jöreskog & Goldberger (1972) provided these derivatives in their Appendix A_2 .

$$\frac{\partial \theta_m}{\partial u_i} = \frac{2}{u_i} \theta_m v_{im}^2 \tag{9.107}$$

$$\frac{\partial v_{im}}{\partial u_j} = \frac{1}{u_j} v_{jm} \sum_{\substack{k=1 \\ k \neq m}}^{n} \frac{\theta_m + \theta_k}{\theta_m - \theta_k} v_{ik} v_{jk} \tag{9.108}$$

Using these equations:

$$\frac{\partial G}{\partial u_i} = \frac{2}{u_i} \sum_{m=r+1}^{n} (\theta_m^2 - \theta_m) v_{im}^2 \ ; \tag{9.109}$$

$$\begin{aligned} \frac{\partial^2 G}{\partial u_i \partial u_j} = \frac{4}{u_i u_j} \sum_{m=r+1}^{n} \{ & (2\theta_m^2 - \theta_m) v_{im}^2 v_{jm}^2 \\ & + (\theta_m^2 - \theta_m) v_{im} v_{jm} \sum_{\substack{k=1 \\ k \neq m}}^{n} \frac{\theta_m + \theta_k}{\theta_m - \theta_k} v_{ik} v_{jk} \\ & - \frac{1}{2} \delta_{ij} (\theta_m^2 - \theta_m) v_{im} v_{jm} \} \end{aligned} \tag{9.110}$$

With $\delta_{ij} = 1$ for $i = j$ and $\delta_{ij} = 0$ for $i \neq j$.

As with maximum likelihood factor analysis there are possible troubles with negative uniqueness; consequently, the same transformation is used as for maximum likelihood factor analysis from u_i^2 to ς_i . This transformation, defined in equation (9.43), has:

$$\varsigma_i = ln(u_i^2) \ .$$

Needed derivatives associated with this transformation are:

$$\frac{\partial G}{\partial \varsigma_i} = \frac{u_i}{2} \frac{\partial G}{\partial u_i} \tag{9.111}$$

$$\frac{\partial^2 G}{\partial \varsigma_i \partial \varsigma_j} = \frac{u_i u_j}{4} \frac{\partial^2 G}{\partial u_i \partial u_j} + \delta_{ij} \frac{u_i}{4} \frac{\partial G}{\partial u_i} \tag{9.112}$$

Using these derivatives, equation (9.109) becomes:

$$\frac{\partial G}{\partial \varsigma_i} = \sum_{m=r+1}^{n} (\theta_m^2 - \theta_m) v_{im}^2 \ . \tag{9.113}$$

Also, equation (9.110) becomes:

$$\frac{\partial^2 G}{\partial \varsigma_i \partial \varsigma_j} = \sum_{m=r+1}^{n} (2\theta_m^2 - \theta_m) v_{im}^2 v_{jm}^2$$
$$+ \sum_{m=r+1}^{n} (\theta_m^2 - \theta_m) v_{im} v_{jm} \sum_{\substack{k=1 \\ k \neq m}}^{n} \frac{\theta_m + \theta_k}{\theta_m - \theta_k} v_{ik} v_{jk} \ . \qquad (9.114)$$

Equation (9.113) gives the gradient vector and equation (9.114) gives the second derivative matrix for use in a Newton-Raphson iteration for the vector of ς's for a minimum value of G . For a given trial vector of ς's , values of trial uniqueness are obtained by the second form of equation (9.43):

$$u_i^2 = EXP(\varsigma_i) \ .$$

Then matrix $\boldsymbol{U}\boldsymbol{C}_{\boldsymbol{yy}}^{-1}\boldsymbol{U}$ is computed and the eigen solution is obtained. Pertinent values for trial 1 for the example are given in Tables 9.14 and 9.15. The corresponding gradient vector and second derivative matrix are computed by equations (9.113) and (9.114). For trial 1 of iterations for one factor, Table 9.16 gives these values. The Newton-Raphson iteration follows the scheme outlined for maximum likelihood method in equation (9.42). The value of the criterion G for the trial may be obtained by equation (9.106). For the example trial 1 this value is .19167. Table 9.17 gives a summary from the iterations over trials for the example for one factor. The criterion stopping the iterations is the value of the derivative vector. This vector should approach zero as a minimum value of the criterion is approached. A very tight termination criterion was employed in the program producing the example so that the iteration was carried to 6 trials when a convergence was apparent at trial 5.

Table 9.18 gives the solution for one factor for the example. The model variances, which are the sums of the communalities and uniqueness, are less than the input variances given in Table 9.5. The generalized least squares solution does not restrict these model variances to equaling the input variances. The value of ξ was computed by equation (9.88) and the degrees of freedom was obtained by equation (9.80). When ξ is interpreted as a χ^2, the value of p is equal to .001 which indicates that the one factor model does not, statistically, fit the data. However, this is no surprise since lack of fit variation was included in the generation of the simulated data. There remains a question of goodness of fit.

As discussed in the introduction to this chapter, the Tucker & Lewis (1973) coefficient of reliability may be adapted to use with generalized least squares factor analysis. Function G is substituted in equation (9.14) for function F. This yields:

$$M_0 = G_0 / d\mathrm{f}_0$$

for no common factors and:

Table 9.16

GLS Trial 1 Derivatives of Function G for One Factor

	First	Second Derivative Matrix*				
	Derivative	1	2	3	4	5
1	.058	.859	.059	.008	.075	.010
2	.288	.059	1.802	.094	.009	.220
3	.110	.008	.094	.910	.009	.003
4	.170	.075	.009	.099	.967	.081
5	.219	.010	.220	.003	.081	1.015

Table 9.17

Outline of Progress over GLS Trials for One Factor

Function G over Trials

1	2	3	4	5	6
.19167	.11483	.10917	.10904	.10904	.10904

First Derivative Vectors Over Trials

	1	2	3	4	5	6
1	.05821	.00479	.00032	.00001	.00000	.00000
2	.28764	.06899	.00970	.00036	.00000	.00000
3	.11035	.01355	.00053	.00002	.00000	.00000
4	.16953	.03031	.00187	.00001	.00000	.00000
5	.21854	.04173	.00268	.00004	.00000	.00000

Table 9.18

Generalized Least Squares Results for One Factor

	Factor Matrix 1	Communality	Uniqueness	Model Variance
1	14.6	213.7	61.7	275.4
2	21.1	446.6	44.3	491.0
3	10.2	103.8	25.4	129.2
4	25.4	650.1	106.2	756.3
5	12.4	153.2	22.7	175.9

ξ (χ^2) = 21.4 df = 5 p = .001
Tucker-Lewis Reliability = .683

$$M_r = G_r / d\mathrm{f}_r$$

for r common factors. Coefficient G_r and degrees of freedom df_r may be obtained from the solution for r common factors. However, coefficient G_0 and degrees of freedom df_0 are required for 0 common factors. Derivation of a solution follows.

For 0 common factors the factor matrix $\boldsymbol{A}$ does not exist so that, from equation (9.2):

$$\boldsymbol{C_{zz0}} = \boldsymbol{U_0^2} \ . \tag{9.115}$$

Then from equation (9.89):

$$G_0 = \frac{1}{2} tr(\boldsymbol{I} - \boldsymbol{C_{yy}^{-1}} \boldsymbol{U_0^2})^2 \ . \tag{9.116}$$

Some algebraic operations follow.

$$G_0 = \frac{1}{2} tr(\boldsymbol{I} - 2\boldsymbol{C_{yy}^{-1}} \boldsymbol{U_0^2} + \boldsymbol{C_{yy}^{-1}} \boldsymbol{U_0^2} \boldsymbol{C_{yy}^{-1}} \boldsymbol{U_0^2}) \, .$$

Elements of $\boldsymbol{C_{yy}^{-1}}$ are designated with superscripts as c^{ij} .

$$G_0 = \frac{1}{2} n - 2 \sum_{i=1}^{n} c^{ii} u_{0i}^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} c^{ij} u_{0j}^2 c^{ji} u_{0i}^2 \ .$$

Due to the symmetry of $\boldsymbol{C_{yy}}$

$$c^{ij} = c^{ji} \, .$$

Then:

$$G_0 = \frac{1}{2} \{ n - 2 \sum_{i=1}^{n} c^{ii} u_{0i}^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} (c^{ij})^2 u_{0i}^2 u_{0j}^2 \} \ .$$

To obtain a minimum G_0 the derivatives with respect to the u_{0k}^2 (k=1, n) are obtained and set equal to zero.

$$\begin{aligned} \frac{\partial G_0}{\partial u_{0k}^2} &= \frac{1}{2} \{ -2c^{kk} + \sum_{j=1}^{n} (c^{kj})^2 u_{0j}^2 + \sum_{i=1}^{n} (c^{ik})^2 u_{0i}^2 \} \\ &= \frac{1}{2} \{ -2c^{kk} + 2 \sum_{i=1}^{n} (c^{ik})^2 u_{0i}^2 \} = 0 \, . \end{aligned}$$

Then:

$$\sum_{i=1}^{n}(c^{ik})^2 u_{0i}^2 = c^{kk} \ . \qquad (9.117)$$

Equation (9.117) is in linear form with the u_{0i}^2 to be determined. The solution is simplified by some temporary definitions.

$\boldsymbol{M}$ is a matrix containing elements $(c^{ik})^2$;

$\underline{\boldsymbol{x}}$ is a column vector of u_{0i}^2 ;

$\underline{\mathbf{c}}$ is a column vector of c^{kk} .

Then equation (9.117) may be written:

$$\boldsymbol{M}\underline{\boldsymbol{x}} = \underline{\boldsymbol{c}}$$

With a solution:

$$\underline{\boldsymbol{x}} = \boldsymbol{M}^{-1}\underline{\boldsymbol{c}} \ .$$

The entries in $\underline{\boldsymbol{x}}$ may be inserted into the diagonal of matrix $\boldsymbol{U_0^2}$. Then, the value of G_0 may be obtained by substitution into equation (9.116).

The degrees of freedom, df_0 , for no common factors equals the number of covariances on one side of the diagonal of $\boldsymbol{C_{yy}}$:

$$df_0 = \frac{1}{2}n(n-1) \ .$$

Coefficients obtained for the illustrative example are given below.

$G_0 = .57763$ $\qquad$ $df_0 = 10$ $\qquad$ $M_0 = .05776$

$G_1 = .10904$ $\qquad$ $df_1 = \ 5$ $\qquad$ $M_1 = .02181$

With N = 200 , n = 5, r = 1, $N_1^{\bigstar}$ = 195.833 by equation (9.18). Then the Tucker & Lewis type reliability given in equation (9.19) is .683 which is given at the bottom of Table 9.18. This result appears to indicate a poor fit of the one factor model to the data. Also, the statistical test which resulted in a p = .001 indicates a lack of fit. However, the construction of this simulated data involved only one major domain dimension. This indicates a problem in deciding on whether to accept this one factor model or not to accept it. Maybe these results have been produced by the small number of attributes, 5, and the large amount of lack of fit variance introduced by the minor domain factors. Experimental trials with larger number of attributes for real data indicate more acceptable results.

Table 9.19 presents the results for a two factor model for the example. The largest changes induced by the addition of the second factor are the increases in communality for attributes 2 and 4 with accompanying decreases in the uniqueness for these two attributes. The

Table 9.19

Generalized Least Squares Results for Two Factors

	Factor Matrix 1	Factor Matrix 2	Communality	Uniqueness	Model Variance
1	13.8	4.5	211.4	66.0	277.4
2	22.3	.0	496.1	.0	496.1
3	9.7	3.0	102.8	28.5	131.3
4	23.5	12.8	714.9	57.1	772.0
5	11.9	2.8	149.0	28.6	177.6

ξ (χ^2) = .66 df = 1 p = .416

Tucker-Lewis Reliability = 1.033

statistical p of .416 indicates that the two factor model should not be rejected. Also, the Tucker & Lewis type reliability has risen to 1.033.

9.3. Minimum Squared Scaled Residual (MINSSR) Factor Fitting

Minimum squared scaled residuals is the third method of factor fitting considered in this chapter. It is closely related to the method of principal factors discussed in Chapter 8; in a sense, MINSSR is the complement of principal factors which was developed in terms of maximizing the sum of squares of the factor loadings. In contrast MINSSR is developed to minimize residual covariances after the extraction of factors. The scaling of the residuals is an important aspect of this method. Two cases are to be considered: determination of the factors in terms of the original scales for the attributes and determination of the factors in terms of attribute scores scaled to standard scores. The principal factors technique considered in Chapter 8 was restricted to consideration of the attribute scores scaled to standard scores. This is implied by application of the principal factors technique to correlation matrices as was the case in Chapter 8.

Equation (9.1) with the weight matrix $\boldsymbol{W}$ restricted to a diagonal scaling matrix is the basis for the MINSSR methods. Two special cases are considered: raw MINSSR , for which $\boldsymbol{W}$ is an identity matrix; normal MINSSR , for which $\boldsymbol{W}$ contains the reciprocals of the observed variances of the attribute scores, these variances being the diagonal entries in $\boldsymbol{C_{yy}}$. Criterion G_1 is used for the raw MINSSR:

$$G_1 = \frac{1}{2} tr[\boldsymbol{I}(\boldsymbol{C_{yy}} - \boldsymbol{C_{zz}})]^2 = \frac{1}{2} tr(\boldsymbol{C_{yy}} - \boldsymbol{C_{zz}})^2 \ . \tag{9.118}$$

Criterion G_2 is used for the normal MINSSR:

$$G_2 = \frac{1}{2} tr[\boldsymbol{S^{-1}}(\boldsymbol{C_{yy}} - \boldsymbol{C_{zz}})]^2 \tag{9.119}$$

Where

$$\boldsymbol{S^2} = diag(\boldsymbol{C_{yy}}) \ . \tag{9.120}$$

An algebraic transformation of criterion G_2 is useful. From equation (9.119):

$$G_2 = \frac{1}{2} tr\{(\boldsymbol{S^{-1}C_{yy}S^{-1}} - \boldsymbol{S^{-1}C_{zz}S^{-1}})\}^2 . \tag{9.121}$$

Note that the sample correlation matrix $\boldsymbol{R}$ is:

$$\boldsymbol{R} = \boldsymbol{S^{-1}C_{yy}S^{-1}} \ . \tag{9.122}$$

Let:

$$\boldsymbol{C}_{zzR} = \boldsymbol{S}^{-1}\boldsymbol{C}_{zz}\boldsymbol{S}^{-1} \ . \tag{9.123}$$

Then:

$$G_2 = \frac{1}{2} tr[\boldsymbol{R} - \boldsymbol{C}_{zzR}]^2 \ . \tag{9.124}$$

The foregoing transformation may be carried on to the factor analysis model given in equation (9.2) by defining:

$$\boldsymbol{A}_R = \boldsymbol{S}^{-1}\boldsymbol{A} \tag{9.125}$$

and

$$\boldsymbol{U}_R^2 = \boldsymbol{S}^{-1}\boldsymbol{U}^2\boldsymbol{S}^{-1} \ . \tag{9.126}$$

The factor model for the correlation matrix becomes:

$$\boldsymbol{C}_{zzR} = \boldsymbol{A}_R\boldsymbol{A}_R' + \boldsymbol{U}_R^2 \ . \tag{9.127}$$

At a solution the matrices $\boldsymbol{A}$ and $\boldsymbol{U}^2$ for the attribute scores in their original scales can be obtained by solving equations (9.125) and (9.126):

$$\boldsymbol{A} = \boldsymbol{S}\boldsymbol{A}_R \ ; \tag{9.128}$$

$$\boldsymbol{U}^2 = \boldsymbol{S}\boldsymbol{U}_R^2\boldsymbol{S} \ . \tag{9.129}$$

There are some concerns about the scale freeness of the MINSSR solutions. The raw MINSSR is not scale free. However application of the material in equations (9.5) through (9.11) indicates that the normal MINSSR is scale free. This result for normal MINSSR supports the long standing application of principal factors and MINSSR to correlation matrices.

Two computational procedures are to be discussed: principal factors with iterated uniqueness; the Harman and Jones (1966) MINRES procedure. For the raw MINSSR, the computing procedure is applied to the observed covariance matrix; for normal MINSSR, the computing procedure is applied to the observed correlation matrix. (Remember that the correlation matrix can be considered to be a covariance matrix among standardized scores.)

The principal factors with iterated uniqueness involves a series of trials with each trial reducing the sum of squared residuals. Each trial has two phases:

1) Given a trial matrix $\boldsymbol{U}^2$, notated $\boldsymbol{U}_t^2$, the factor matrix $\boldsymbol{A}_t$ is determined to minimize the sum of squared residuals.
2) Given the matrix $\boldsymbol{A}_t$, the entries in $\boldsymbol{U}_t^2$ are adjusted to set the diagonal residuals to zero. Since this operation does not affect the off-diagonal residuals, the sum of

squares of the residuals as a whole is reduced by the sum of squares of the diagonal entries before this adjustment.

Convergence of this procedure occurs when there are minimal changes in the trial $\boldsymbol{U^2}$ from trial to trial.

The procedure to minimize the sum of squared residuals will be discussed in terms of a given covariance matrix $\boldsymbol{C_{yy}}$. For analysis of a correlation matrix, the correlation matrix is substituted for the covariance matrix. For a given trial $\boldsymbol{U_t^2}$, a matrix $\widetilde{\boldsymbol{C}}_{\boldsymbol{yyt}}$ is defined by:

$$\widetilde{\boldsymbol{C}}_{\boldsymbol{yyt}} = \boldsymbol{C_{yy}} - \boldsymbol{U_t^2} \tag{9.130}$$

Cell entries in $\widetilde{\boldsymbol{C}}_{\boldsymbol{yyt}}$ are designated by $\widetilde{\mathrm{c}}_{tij}$. Let the matrix of residuals be:

$$\boldsymbol{E_t} = \widetilde{\boldsymbol{C}}_{\boldsymbol{yyt}} - \boldsymbol{A_t A_t'} \ . \tag{9.131}$$

For the ij cell of E_t :

$$e_{tij} = \widetilde{\mathrm{c}}_{tij} - \sum_{k=1}^{r} a_{tik} a_{tjk} \ . \tag{9.132}$$

where r is the number of factors. The sum of squared residuals over the entire matrix $\boldsymbol{E_t}$ is represented by the function g_t .

$$\mathrm{g}_t = \sum_{i=1}^{n} \sum_{j=1}^{n} e_{tij}^2 \ . \tag{9.133}$$

To minimize this function, the partial derivatives of g_t with respect to the factor weights a_{thk} are obtained:

$$\frac{\partial \mathrm{g}_t}{\partial a_{thk}} = 2 \sum_{j=1}^{n} e_{thj}(-a_{tjk}) + 2 \sum_{i=1}^{n} e_{tih}(-a_{tik}) \ . \tag{9.134}$$

These partial derivatives are set equal to zero. With the fact that the matrix $\boldsymbol{E_t}$ is symmetric, algebraic operations yield:

$$\sum_{i=1}^{n} e_{tih} a_{tik} = 0 \tag{9.135}$$

With equation (9.132):

$$\sum_{i=1}^{n} (\widetilde{\mathrm{c}}_{tih} - \sum_{m=1}^{r} a_{tim} a_{thm}) a_{tik} = 0 \ . \tag{9.136}$$

Then:

$$\sum_{i=1}^{n} \widetilde{c}_{tih} a_{tik} = \sum_{m=1}^{r} a_{thm} \sum_{i=1}^{n} a_{tim} a_{tik} \ . \tag{9.137}$$

Consider that the factor matrix has been transformed to principal axes so that:

$$\boldsymbol{A_t' A_t} = \boldsymbol{\Lambda_t} \ ,$$

where $\boldsymbol{\Lambda_t}$ is a diagonal matrix so that:

$$\sum_{i=1}^{n} a_{tim} a_{tik} = \sum_{i=1}^{n} a_{tik}^2 = \boldsymbol{\lambda}_{tk} \qquad \text{for } m = k \tag{9.138a}$$

$$\sum_{i=1}^{n} a_{tim} a_{tik} = 0 \qquad\qquad \text{for } m \neq k \ . \tag{9.138b}$$

Equation (9.137) becomes:

$$\sum_{i=1}^{n} \widetilde{c}_{tih} a_{tik} = a_{thk} \lambda_{tk} \ . \tag{9.139}$$

In matrix form with $\underline{\mathbf{a}}_{tk}$ being a column vector of a_{tik} or a_{thk} , equation (9.139) yields:

$$\widetilde{\boldsymbol{C}}_{\boldsymbol{yyt}} \, \underline{\boldsymbol{a}}_{\boldsymbol{tk}} = \underline{\boldsymbol{a}}_{\boldsymbol{tk}} \lambda_{tk}$$

or

$$(\widetilde{\boldsymbol{C}}_{\boldsymbol{yyt}} - \lambda_{tk} \boldsymbol{I}) \underline{\boldsymbol{a}}_{\boldsymbol{tk}} = \mathbf{0} \ . \tag{9.140}$$

The $\boldsymbol{\lambda}_{tk}$ are eigenvalues of $\mathbf{C}_{\boldsymbol{yyt}}$ and the $\underline{\mathbf{a}}_{\boldsymbol{tk}}$ are corresponding eigenvectors which are scaled according to equation (9.138a). Note that the $\underline{\mathbf{a}}_{\boldsymbol{tk}}$ are principal factors of $\widetilde{\boldsymbol{C}}_{\boldsymbol{yyt}}$.

There remain questions both as to the value of the criterion g_t and as to the dimensions of the eigen solution to be used. From equations (9.132) and (9.133):

$$g_t = \sum_{i=1}^{n} \sum_{j=1}^{n} (\widetilde{c}_{tij} - \sum_{k=1}^{r} a_{tik} a_{tjk})^2$$

which may be written as:

$$g_t = \sum_{i=1}^{n} \sum_{j=1}^{n} (\widetilde{c}_{tij} - \sum_{k=1}^{r} a_{tik} a_{tjk})(\widetilde{c}_{tij} - \sum_{m=1}^{r} a_{tim} a_{tjm}) \ .$$

Algebraic operations yield:

$$g_t = \sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{c}_{tij}^2 - \sum_{k=1}^{r}\sum_{m=1}^{r}\sum_{i=1}^{n} a_{tik}a_{tim}\sum_{j=1}^{n} a_{tjk}a_{tjm} \ .$$

Using equations (9.138a) and (9.138b):

$$g_t = \sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{c}_{tij}^2 - \sum_{k=1}^{r}\lambda_k^2 \ . \tag{9.141}$$

From equation (9.138a) only positive eigenvalues are permissible. Function g_t is minimized by selecting the r eigen dimensions having the largest, positive eigenvalues λ_k . In case there are not r positive λ_k , no solution is permissible.

The second phase of a trial involves adjustment of the trial uniqueness to:

$$\boldsymbol{U}_{t+}^2 = \boldsymbol{U}_t^2 + Diag(\boldsymbol{E}_t) \ . \tag{9.142}$$

The matrix of discrepancies, $\boldsymbol{E}_{t+}$, uses the $\boldsymbol{A}_t$ computed in the first phase and is:

$$\begin{aligned} \boldsymbol{E}_{t+} &= \boldsymbol{C}_{yy} - \boldsymbol{U}_{t+}^2 - \boldsymbol{A}_t\boldsymbol{A}_t' \\ &= \boldsymbol{C}_{yy} - \boldsymbol{U}_t^2 - Diag(\boldsymbol{E}_t) - \boldsymbol{A}_t\boldsymbol{A}_t' \\ &= \boldsymbol{E}_t - Diag(\boldsymbol{E}_t) \ . \end{aligned} \tag{9.143}$$

Then:

$$Diag(\boldsymbol{E}_{t+}) = \boldsymbol{0} \ . \tag{9.144}$$

An operator SSQ[X] is used to designate the sum of squares of the entries in matrix $\boldsymbol{X}$. From equation (9.143) and the fact that the diagonal entries in E_{t+} equal 0 :

$$SSQ[\boldsymbol{E}_t] = SSQ[\boldsymbol{E}_{t+}] + SSQ[Diag(\boldsymbol{E}_t) \ .] \tag{9.145}$$

The following logic will hold when there are no negative entries in the adjusted uniqueness matrix $\boldsymbol{U}_{t+}^2$. When any diagonal entry, e_{tii} , in $\boldsymbol{E}_t$ does not equal zero:

$$SSQ[Diag(\boldsymbol{E}_t)] > 0$$

and

$$SSQ[\boldsymbol{E}_{t+}] < SSQ[\boldsymbol{E}_t] \ . \tag{9.146}$$

Thus, the adjustment of the uniqueness reduces the sum of squares of the residuals. For the case when there are no negative entries in $\boldsymbol{U}_{t+}^2$, this matrix becomes the $\boldsymbol{U}_t^2$ for the next trial. Convergence occurs when there are minimal changes from one trial to the next.

Sometimes one of the uniqueness in $\boldsymbol{U}_{t+}^2$ becomes negative, this being termed a generalized Haywood case. Since negative uniqueness are not permissible, a common practice is

to employ a 'fix up' of replacing the negative uniqueness to zero in the transition from $\boldsymbol{U}_{t+}^{2}$ to the next trial $\boldsymbol{U}_{t}^{2}$. This violates the logic discussed in the preceding paragraph. However, the solution appears to converge. Note that the obtained factor matrix still implies a negative uniqueness after the solution of the first phase of the last trial.

A remaining problem is the initial values of uniqueness. Common usage involves the error variances in estimating the attributes from the other attributes in the battery. These values were given for the maximum likelihood method in equation (9.44). Then, trial 1, phase 1 is identical with principal factors with squared multiple correlations in the diagonal method for analysis of correlation matrices method described in Chapter 8.

While the principal factors with iterated uniqueness is easily programmed, rate of convergence sometimes is a problem. In studies having a relatively large number of factors the iterations seem just to crawl along with no evidence of immediate convergence. A large number of trials become necessary. This slowness of convergence along with the problems of handling the generalized Haywood cases lessen the attractiveness of this approach.

The MINRES procedure developed by Harman and Jones (1966) is much more effective. A true solution for the generalized Haywood case was developed by Harman and Fukuda (1966). This method starts with a trial factor matrix instead of trial uniqueness. With the given trial factor matrix the method cycles through the attributes one at a time replacing the row of factor loadings so as to minimize the sum of squares of the residuals of each attribute in turn with the other attributes. Since the trial factor matrix has been changed by the solutions for the various rows of the matrix, repeated solutions have to be made for the sequence of rows of the factor matrix. These cycles of solutions are continued until no changes occur between cycles.

Let $\boldsymbol{A}$, $n \times r$, be a trial factor matrix for the n attributes on r factors. Note that the solution assumes a given number of factors so that separate solutions must be made for different numbers of factors. Let $\boldsymbol{A}_{-i}$, $(n-1) \times r$, be the trial factor matrix with row i deleted where i is a pivot attribute. A solution is desired for row vector $\underline{\mathbf{a}}_{i}$ such as to minimize the sum of squared residuals, e_{ij}, $j \neq i$, with all other attributes. Note that the diagonal residual is excluded. These residuals are given by:

$$e_{ij} = c_{ij} - \sum_{k=1}^{r} a_{ik} a_{jk} \quad \text{with } j \neq i \tag{9.147}$$

where c_{ij} is the ij entry in the given covariance matrix $\boldsymbol{C}_{yy}$. In matrix form:

$$\underline{\boldsymbol{e}}_{i,-i} = \underline{\boldsymbol{c}}_{i,-i} - \underline{\boldsymbol{a}}_{i} \boldsymbol{A}'_{-i} \tag{9.148}$$

where $\underline{\boldsymbol{e}}_{\boldsymbol{i},-\boldsymbol{i}}$ is the row vector of residuals for attribute i excluding attribute i and $\underline{\mathbf{c}}_{\boldsymbol{i},-\boldsymbol{i}}$ is the corresponding row vector for attribute i from matrix $\boldsymbol{C}_{\boldsymbol{yy}}$ with attribute i excluded. A criterion g_i can be defined as:

$$g_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} e_{ij}^2 = \underline{\boldsymbol{e}}_{\boldsymbol{i},-\boldsymbol{i}} \underline{\boldsymbol{e}}'_{\boldsymbol{i},-\boldsymbol{i}} \ . \tag{9.149}$$

At a solution the uniqueness for attribute i is given by:

$$u_i^2 = c_{ii} - \sum_{k=1}^{r} a_{ik}^2 = c_{ii} - \underline{\boldsymbol{a}}_{\boldsymbol{i}} \underline{\boldsymbol{a}}'_{\boldsymbol{i}} \ . \tag{9.150}$$

Two solutions are considered: one an unrestricted solution and second a restricted solution which is used when the unrestricted solution yields a negative uniqueness, that is, a generalized Heywood case.

Solution for the unrestricted case is quite simple. The problem from equations (9.148) and (9.149) is of the form of a general least squares. Minimization of g_i involves a set of simultaneous linear equations with the entries a_{ik} in vector $\underline{\mathbf{a}}_{\boldsymbol{i}}$ as unknowns. These equations in matrix form are:

$$\underline{\boldsymbol{a}}_{\boldsymbol{i}} (\boldsymbol{A}'_{-\boldsymbol{i}} \boldsymbol{A}_{-\boldsymbol{i}}) = (\underline{\boldsymbol{c}}_{\boldsymbol{i},-\boldsymbol{i}} \boldsymbol{A}_{-\boldsymbol{i}}) \ . \tag{9.151}$$

For any given pivot attribute i on any cycle of solutions when the uniqueness computed by equation (9.150) is equal to or greater than zero, the unrestricted solution is acceptable. When the uniqueness is negative, the unrestricted solution is to be discarded and a restricted solution obtained.

For the restricted solution the uniqueness is constrained to equaling zero which is expressed in the following equation.

$$c_{ii} - \sum_{k=1}^{r} a_{ik}^2 = c_{ii} - \underline{\boldsymbol{a}}_{\boldsymbol{i}} \underline{\boldsymbol{a}}'_{\boldsymbol{i}} = 0 \ . \tag{9.152}$$

The revised criterion, g_{ri} , is:

$$g_{ri} = \sum_{\substack{j=1 \\ j \neq i}}^{n} e_{ij}^2 + \beta (c_{ii} - \sum_{k=1}^{r} a_{ik}^2) \tag{9.153}$$

with β being an unknown LaGrange multiplier. A solution for optimum values of g_{ri} involves the partial derivatives of g_{ri} with respect to the a_{ik} .

$$\frac{\partial \mathbf{g}_{ri}}{\partial a_{ik}} = 2\sum_{\substack{j=1 \\ j\neq i}}^{n} e_{ij}(-a_{jk}) - 2\beta a_{ik} \ . \tag{9.154}$$

Setting these derivatives to zero yields:

$$\sum_{\substack{j=1 \\ j\neq i}}^{n} e_{ij}a_{jk} + \beta a_{ik} = 0 \ . \tag{9.155}$$

Substitution from equation (9.147) yields:

$$\sum_{\substack{j=1 \\ j\neq i}}^{n} c_{ij}a_{jk} - \sum_{\substack{j=1 \\ j\neq i}}^{n} a_{jk}\sum_{m=1}^{r} a_{im}a_{jm} + \beta a_{ik} = 0 \ . \tag{9.156}$$

With algebraic operations equation (9.156) may be written in matrix form as:

$$\underline{\boldsymbol{c}}_{\boldsymbol{i},-\boldsymbol{i}}\boldsymbol{A}_{-\boldsymbol{i}} - \underline{\boldsymbol{a}}_{\boldsymbol{i}}(\boldsymbol{A}'_{-\boldsymbol{i}}\boldsymbol{A}_{-\boldsymbol{i}} - \beta\boldsymbol{I}) = \boldsymbol{0} \ . \tag{9.157}$$

Note that when β equals zero this equation reduces to equation (9.151) for the unrestricted case. However, for the restricted case a solution for β is needed so as to satisfy equation (9.157) and the constraining equation (9.152). There appear to be several possible solutions similar to the case for an eigen problem. However, there appears to be no simple solution leading to a standard form. Assume that the matrix formed by the product $\boldsymbol{A}'_{-\boldsymbol{i}}\boldsymbol{A}_{-\boldsymbol{i}}$ is positive, definite, thus being nonsingular. Then a solution for $\underline{\mathbf{a}}_{\boldsymbol{i}}$ is:

$$\underline{\boldsymbol{a}}_{\boldsymbol{i}} = \underline{\boldsymbol{c}}_{\boldsymbol{i},-\boldsymbol{i}}\boldsymbol{A}_{-\boldsymbol{i}}(\boldsymbol{A}'_{-\boldsymbol{i}}\boldsymbol{A}_{-\boldsymbol{i}} - \boldsymbol{\beta I})^{-1} \ . \tag{9.158}$$

The communality for attribute i from Chapter 3, is:

$$h_i^2 = \underline{\boldsymbol{a}}_{\boldsymbol{i}}\underline{\boldsymbol{a}}'_{\boldsymbol{i}} \ .$$

From equation (9.158):

$$h_i^2 = \underline{\boldsymbol{c}}_{\boldsymbol{i},-\boldsymbol{i}}\boldsymbol{A}_{-\boldsymbol{i}}(\boldsymbol{A}'_{-\boldsymbol{i}}\boldsymbol{A}_{-\boldsymbol{i}} - \beta\boldsymbol{I})^{-2}\boldsymbol{A}'_{-\boldsymbol{i}}\underline{\boldsymbol{c}}'_{\boldsymbol{i},-\boldsymbol{i}} \ . \tag{9.159}$$

The constraint of equation (9.152) may be written as:

$$h_i^2 = c'_{ii} \ . \tag{9.160}$$

In order for the matrix $(\boldsymbol{A}'_{-\boldsymbol{i}}\boldsymbol{A}_{-\boldsymbol{i}} - \beta\boldsymbol{I})$ to remain positive, definite, the value of β must be less than the least eigenvalue of $\boldsymbol{A}'_{-\boldsymbol{i}}\boldsymbol{A}_{-\boldsymbol{i}}$. When h_i^2 is considered as a function of the slope of this function is positive in the range from minus infinity to the least eigenvalue of $\boldsymbol{A}'_{-\boldsymbol{i}}\boldsymbol{A}_{-\boldsymbol{i}}$. Since this restricted case is used only when the unrestricted case yields a value of h_i^2 greater than c_{ii} , that is, when β equals zero, the range of interest for β is from minus infinity to zero. This range

is included in the range when the slope of the function for h_i^2 is positive. Thus, the desired value of β is negative. Further, there is only one value of β for which h_i^2 equals c_{ii} . These observations lead to a possible solution involving interpolation between values of β which yield communalities less than and greater than c_{ii} . Such a procedure can be made quite efficient.

Table 9.20 presents a MINRES solution for the nine mental tests example for three common factors. The upper section gives the MINRES output factor matrix along with the communality and uniqueness for each attribute. Since none of the uniqueness are negative, the restrictive solution was not used. A transformation of factors to correlated factors was performed with the transformed factor matrix being given in the middle section of the table. The factor correlations matrix is given in the bottom section of the table. These results may be compared with those given in Table 1.2 which were obtained using principle factors with SMC's in the diagonal.

Table 9.21 gives some statistics for the MINRES factor extraction for the nine mental tests example. These statistics are given for zero through four factor solutions. For each of these solutions the root mean square off-diagonal residual is given as RMS Residuals, There are considerable decreases in these values from zero factors through three factors with a much lesser decrease from three factors to four factors. The RMS Residual is relatively small for the three factor solution, this supporting the idea of accepting the three factor solution. Rippe (1953) proposed an "Application of a large sampling criterion to some sampling problems in factor analysis" which has been suggested for use with MINRES factor analysis (see: Harman (1976), page 184). This statistic is identical to the maximum likelihood statistic, see equation (9.34). Let F_R be:

$$F_R = ln\left|\boldsymbol{C_{zz}}\right| - ln\left|\boldsymbol{C_{yy}}\right| + tr(\boldsymbol{C_{zz}^{-1}C_{yy}}) - n \ . \tag{9.161}$$

Rippe's chi-square statistic is given by:

$$\xi_R = N^{\bigstar} F_R \ . \tag{9.162}$$

Rippe gives $N^{\bigstar} = N - 1$; however, a better value from Bartlett (1950) is given in equation (9.18) and repeated here for convenience.

$$N^{\bigstar} = N - 1 - (2n + 5)/6 - 2r/3 \ . \tag{9.18}$$

Rippe gave a value for the degrees of freedom larger than the one for maximum likelihood factor analysis. Study of his values raised some questions so that a small Monte Carlo study was performed for MINSSR factoring. Results indicated that the degrees of freedom computed from the maximum likelihood factor analysis formula was more appropriate than the Rippe values for

Table 9.20

Factor Matrices from MINRES Factor Extraction
Nine Mental Tests Example
Three Common Factors

MINRES Output Factor Matrix

	1	2	3	Communality	Uniqueness
1	.331	.252	.499	.422	.578
2	.391	.501	.494	.648	.352
3	.555	.098	.353	.442	.558
4	.573	-.451	.030	.533	.467
5	.690	-.484	.003	.710	.290
6	.613	-.386	.028	.525	.475
7	.473	.497	.017	.472	.528
8	.673	.163	-.224	.529	.471
9	.618	.284	-.103	.473	.527

Transformed Factor Matrix

	1	2	3
1	.692	.060	-.115
2	.770	-.062	.074
3	.522	.337	.021
4	.004	.732	-.039
5	-.015	.843	.019
6	.028	.718	.023
7	.249	.022	.522
8	-.083	.405	.614
9	.078	.278	.560

Factor Correlation Matrix

	1	2	3
1	1.000	.121	.514
2	.121	1.000	.086
3	.514	.086	1.000

Table 9.21

Statistics for MINRES Factor Extraction
Nine Mental Tests Example
(N = 710)

Number of Factors	RMS Residuals	Significance Statistics			T – L Type Reliability
		Chi-Square	DF	P	
0	.331	1975.6	36	.000	.000
1	.151	833.1	27	.000	.445
2	.053	169.4	19	.000	.853
3	.015	28.6	12	.005	.974
4	.010	12.1	6	.059	.981

MINSSR factoring. These degrees of freedom are given in equation (9.80), repeated here for convenience.

$$\nu(\xi) = n(n+1)/2 - n - nr + r(r-1)/2 \ . \qquad (9.80)$$

The Rippe statistics are given in Table 9.21. For the three factor solution the value of p equal to .005 indicates that this model should be rejected. However, as has been indicated several times previously, there is no surprise in this observation in that a worth while factor model is quite likely not to fit the real world even for samples approaching the population of individuals. Table 9.21 also gives the Tucker-Lewis type reliability values for the several factor solutions. These values were computed according to the material presented earlier in this chapter. See the material associated with equations (9.14) and (9.19). The value of .974 appears quite adequate to accept the three factor solution.

There is some interest in the scale freeness of factor solutions. The Rippe function is not scale free when used with the raw MINSSR solution since the raw MINSSR solution is not scale free. However, the Rippe function is scale free when used with the normal MINSSR solution.

The Rippe function is different from the function minimized in MINSSR. The maximum likelihood solution would minimize Rippe's function but, probably, would not minimize the MINSSR function. Conversely, considering the normal MINSSR case, the normal MINSSR solution, probably, would not minimize the maximum likelihood and the Rippe function. Only when the MINSSR solution is a good approximation to the maximum likelihood solution would the Rippe function be appropriate. For the nine mental tests example, the MINSSR solution appears to be a good approximation. Table 9.22 gives the maximum likelihood statistics for this example. Note that the maximum likelihood chi-squares are slightly lower than the Rippe chi-squares in Table 9.21. For other, larger examples the maximum likelihood chi-squares are markedly less than the Rippe chi-squares. A conclusion would be that the Rippe function should be used with considerable caution.

9.4. Remarks on the Number of Factors

Considerations as to the number of factors should take into account a variety of matters arising from the differences between factor analytic models and the real world. As discussed in Chapters 3 and 4 there are model errors in both the population and the sample. Further, these model errors may be different for different methods of factor extraction as indicated in Table 9.3 and associated text earlier in this Chapter. These matters of model error affect the statistical functions associated with the factor methods discussed in this chapter. One should expect that a good model determined for a given number of factors would be rejected by the chi-square

Table 9.22

Maximum Likelihood Factor Statistics
Nine Mental Tests Example
(N = 710)

Number of Factors	Significance Statistics			T – L Type
	Chi-Square	DF	P	Reliability
0	1975.6	36	.000	.000
1	824.9	27	.000	.451
2	167.6	19	.000	.855
3	28.1	12	.005	.975
4	11.4	6	.077	.983

statistics. However, the chi-square should be low enough to indicate statistical stability for an accepted solution.

In addition to chi-square statistics generated in the factor extraction, consideration should be given to other items of information. For example, the Guttman weaker lower bound on the number of factors (the number of eigenvalues of the correlation matrix with unities in the diagonal equal to or greater than unity) could be considered. Further, as described in Chapter 8 for principal factors extraction, the series of eigenvalues of the correlation matrix with SMC's in the diagonal should be inspected for a break between larger eigenvalues and a straight line of decreasing eigenvalues. Also, the parallel analysis described in Chapter 8 should be considered. For the nine mental tests example, material in Chapter 8 indicates that the number of eigenvalues equal to or greater than unity of the correlation matrix is two. However, there appears to be a break in the eigenvalue series for the correlation matrix with SMC's in the diagonal between three and four factors. This break along with the parallel analysis indicates a three factor solution. However, as given in Table 9.22, the chi-square for three factors yields a p of .005 which, statistically, indicates that a three factor solution should be rejected. A decision to use a three factor solution could be defended in that the chi-square is low compared with the preceding values for fewer factors. Also, the Tucker-Lewis reliability is acceptably high.

Table 9.23 gives the maximum likelihood statistics for the 18 verbal and numerical tests selected from the Thurstone and Thurstone (1941) battery. The correlation matrix is given in Table 8.15 with results from principal factors analysis given in Tables 8.16 and 8.17. A graph of the series of eigenvalues of the correlation matrix with SMC's in the diagonal is given in Figure 8.2. From this graph a conclusion of four factors appears justified. The parallel analysis given in Table 8.16 agrees with a decision of four factors. However, from Table 9.23, the maximum likelihood chi-square for four factors yields a p of .000 indicating that a four factor solution should be rejected. The Tucker-Lewis reliability of .974 for four factors is quite acceptable. Maybe, a four factor solution is OK for the principal factors analysis while the maximum likelihood factor analysis requires more than four factors. This difference may reflect a difference in the representation of the correlation matrix. A five factor maximum likelihood solution is problematic with the p of .010. Some study is indicated after factor transformations to more meaningful factors before deciding which factor solution to accept.

A final point is that the factor matrices obtained by any of the factor extraction methods establishes a factor space in which factor transformations are to be performed. The original factors obtained are possible mixtures of more meaningful factors. Extraction of too few factors yields confused results for the transformed factors. As Kaiser has said (informally) "it is a sin to extract too few factors." Extraction of too many factors will result in an extra, less meaningful factor which may result in some problems in the transformation of factors part of the analysis.

Table 9.23

Maximum Likelihood Factor Statistics
Thurston & Thurston 18 V & N Tests
(N = 710)

Number of Factors	Significance Statistics			T – L Type
	Chi-Square	DF	P	Reliability
0	5429.4	153	.000	.000
1	1784.4	135	.000	.645
2	732.2	118	.000	.849
3	396.9	102	.000	.916
4	165.4	87	.000	.974
5	103.8	73	.001	.988

When possible with good data, the extraction of a number of factors which differentiates between factors with considerable good contributions to the variance of the attributes and factors having "junk" contributions is desirable.