CONSTRAINED ESTIMATION IN MULTIPLE GROUPS
COVARIANCE STRUCTURE MODEL

by

Kwok-leung TSUI

A Thesis

submitted to

the Graduate School of
The Chinese University of Hong Kong
(Division of Mathematics)

In Partial Fulfillment
of the Requirements for the Degree of
Master of Philosophy (M. Phil.)

Hong Kong

May 1981.
THE CHINESE UNIVERSITY OF HONG KONG

GRADUATE SCHOOL

The undersigned certify that we have read a thesis, entitled "Constrained Estimation in Multiple Groups Covariance Structure Model" submitted to the Graduate School by Mr. Kwok-leung TSUI in partial fulfillment of the requirement for the degree of Master of Philosophy in Mathematics. We recommend that it be accepted.

Dr. S.Y. Lee, Supervisor
Dr. N.N. Chan
Dr. H.K. Lam

Professor P.M. Bentler
External Examiner
ACKNOWLEDGEMENT

First of all, I would like to express my appreciation and gratitude to my supervisor, Dr. S.Y. Lee for his guidance and encouragement throughout the preparation of this thesis. Also, I want to thank Dr. N.N. Chan who had given me the stimulus to pursue my graduate study in statistics. Furthermore, my thanks go to Mr. K.H. Lee for his assistance in the computer programme. Finally, thanks are also due to Mr. Billy Lam for his typing.
Abstract

This thesis is concerned with the study of covariance structure models in several populations. Estimation theory of the parameters that are subject to general functional restraints is developed based on the generalized least squares approach. Asymptotic properties of the constrained estimator are studied; and asymptotic chi-squared tests are presented to evaluate appropriate model comparisons. The method of multiplier, the standard reparametrization technique and the Fletcher-Reeves algorithm are discussed in obtaining the estimates. The confirmatory factor analysis model in several populations is specifically discussed. The methodology is demonstrated by both real and artificial data.
CONTENTS

Chapter 1. Introduction ........................................ 1

Chapter 2. Multiple group covariance structure analysis .... 8

Chapter 3. Minimization Procedures .............................. 14

Chapter 4. Multiple group confirmatory factor analysis ....... 23

Chapter 5. Numerical examples .................................. 34

Reference .................................................................. 39

Tables ...................................................................... 42
Chapter 1. Introduction

In order to introduce the analysis of covariance structure, a number of models are described firstly. The earliest covariance structure model is the general factor analytic model of Spearman (1904), which has led to the major development of common factor analysis. In Lawley (1940) and Lawley (1958), the unrestricted common factor model and the restricted oblique common factor model were first made to yield the efficient estimation and asymptotic testing. These are the main mathematical components of the analysis of covariance structure at that period, though not under that name.

The term 'analysis of covariance structure' was first introduced in Bock and Bargmann (1966), in which they described three models similar to the unrestricted oblique model. Jöreskog (1967, 1969) applied the Davidon-Fletcher-Powell quasi-Newton method to the unrestricted factor model and restricted oblique factor model as treated by Lawley (1940) and Lawley (1958), thus introducing an effective minimization algorithm into the psychometric field. Also, Jöreskog (1970) described a model for the analysis of covariance structures that is equivalent to a second-order oblique factor model, in which any element of any parameter matrix can be prescribed to be equal to one or more elements. This model contains a remarkably large number of useful special cases.

A very general model that essentially arises out of the theory of linear structural relations (Malinvaud, 1970) was described in Wiley (1973).
The reparametrization of this model was programmed by using Davidon-Fletcher-Powell algorithm in Jöreskog and Van Thillo (1973). In Jöreskog and Sörbom (1976), this model was reprogrammed to allow patterned residual covariance matrices. The model is called LISREL IV. This linear structural relations model is probably the most useful model for the analysis of covariance structures.

At present, the covariance structural analysis is widely applied in many areas, particularly in behavioral sciences. For example, it can be applied to handle such problems as analysis of multitrait-multimethod data (Campbell and Fiske, 1959), analysis of simplexes and circumplexes model (Anderson, 1960; Jöreskog, 1970), estimation of variance and covariance components (Harville, 1977), path analysis (Blalok, 1971) and factor analysis (Harman, 1967).

In general, any covariance structure may be defined by specifying that the elements of populations covariance matrix $\Sigma$ of the observed variables are certain functions of parameters $\theta_1, \theta_2, \ldots, \theta_t$, i.e.

$$\sigma_{ij} = \sigma_{ij}(\theta),$$

or in matrix form, $\Sigma = \Sigma(\theta)$, where $\theta = (\theta_1, \theta_2, \ldots, \theta_t)'$ is the parameter vector and $\sigma_{ij}$ is the $(ij)$th element of $\Sigma$. It is assumed that the function $\sigma_{ij}(\theta)$ are continuous and have continuous first and second derivatives; and $\Sigma$ is positive definite at each point $\theta$ of the admissible parameter space. The distribution of the observed variables is assumed to be multivariate normal with an unconstrained mean vector $\mu$ and covariance matrix $\Sigma(\theta)$. The main tasks of the covariance structure analysis are the estimation of parameters in $\Sigma$ and testing hypothesis about the model.
The parameters in $\theta$ were estimated from a sample of $N$ independent observations $x_1, x_2, \ldots, x_N$ on the distribution $N(\mu, \Sigma)$. Let

$$S = \frac{1}{N-1} \sum_{\alpha=1}^{N} (x_\alpha - \bar{x})(x_\alpha - \bar{x})',$$

be the usual sample covariance matrix based on the observed data, where

$$\bar{x} = \frac{1}{N} \sum_{\alpha=1}^{N} x_\alpha$$

is the sample mean vector. The estimation problem is to estimate $\theta$ by fitting a matrix $\Sigma$ of the form $\Sigma(\theta)$ to the observed covariance matrix $S$. Generally, two different methods are used in fitting $\Sigma(\theta)$ to $S$, which are generalized least squares (GLS) method and maximum likelihood (ML) method. The method considered in this thesis is the GLS approach.

In this thesis, the covariance structure analysis in several populations will be studied. This problem can be regarded as the generalization of the covariance structure analysis in single population. According to McDonald (1980), the method in solving this problem is similar to the single group covariance structure analysis problem when the sample sizes of the $m$ populations are equal. However, his paper did not give any asymptotic properties. The method is not applicable for the observed data of different sample sizes. Therefore, a more general method to handle the multiple group covariance structure analysis problem will be studied in this thesis.

The model will be discussed in next chapter. The minimization procedures for obtaining their estimates are discussed in Chapter 3. In
Chapter 4, we will specifically discuss the confirmatory factor analysis model in the multiple group covariance structure analysis. Finally, some real and artificial examples will be given in Chapter 5.

In the following chapters, we shall use the following notations and abbreviations:

\[ I_p \]: the \( p \) by \( p \) identity matrix.
\[ X' \]: the transpose of \( X \).
\[ \text{vec } X \]: an \( mn \) by 1 vector which takes all the elements of the \( m \) by \( n \) matrix \( X \) row by row and places them in a vector.
\[ X_{\text{sym}} \]: an \( n(n+1)/2 \) by 1 vector which takes the lower triangle elements (including the diagonal elements) of the \( n \) by \( n \) matrix \( X \) row by row and places them into a vector.
\[ X_{\text{dg}} \]: an \( n \) by 1 vector whose components are the diagonal elements of \( n \) by \( n \) matrix \( X \).
\[ X \otimes Y \]: the right Kronecker product of matrices \( X \) and \( Y \) of any order, i.e.
\[ X \otimes Y = (x_{ij}Y) \].

CSA : Covariance structure analysis.
CFA : Confirmatory factor analysis.
ML : Maximum likelihood.
GLS : Generalized least squares.
RMS : Root mean squares.

The following definitions and theorems are the mathematical preliminary...
of matrix calculus for deriving the derivatives required in Chapter 4.
For more details, see McDonald and Swaminathan (1973), and Lee and Bentler (1975).

Definition. Given a p by q matrix Y whose elements are differentiable functions of elements of an m by n matrix X, the matrix derivative \( \frac{\partial Y}{\partial X} \) is equal to the mn by pq matrix whose \( \{(u-1)n+v, (r-1)q+s\} \) element is given by \( \frac{\partial y_{uv}}{\partial x_{uv}} \), for \( 1 \leq r \leq p, 1 \leq s \leq q, 1 \leq u \leq m, 1 \leq v \leq n \).

In this thesis, unless otherwise stated, we assume all matrix derivatives exist. Obviously \( \frac{\partial \text{vec } Y}{\partial X} = \frac{\partial Y}{\partial X} \) by definition.

Suppose \( X \) is an m by n matrix, we denote \( E_{mn} = \frac{\partial X}{\partial X} \).

\( E_{nn} \) is a mn by mn matrix, whose general element is

\[
(\mathbf{E}_{mn})_{n(j-1)+k, n(j'-1)+k'} = \begin{cases} 1 & \text{if } j = k' \text{ and } k = j' \\ 0 & \text{otherwise} \end{cases}
\]

where \( 0 < j, k \leq m, 0 < j', k \leq n \).

Also, suppose \( X \) is an n by n matrix, we denote \( K_n = \frac{\partial X}{\partial X} \).

\( K_n \) is an n by n matrix, whose general element is

\[
(K_n)_{ij} = \begin{cases} 1 & \text{if } j = n(i-1) + i \\ 0 & \text{otherwise} \end{cases}
\]

where \( 0 < i \leq n \).
Furthermore, we denote

\[ T_n = \frac{\delta X}{\delta X_{\text{sym}}} \]

with \( T_n \) is a \( n(n+1)/2 \) by \( n^2 \) matrix, whose general element is

\[ (T_n)_{ij} = \frac{i'(j'-1)}{2} + k, \quad n(j-1) + k = \begin{cases} 1 & \text{if } j = j' \text{ and } k = k' \text{ or } j = k' \text{ and } j' = k', \\ 0 & \text{otherwise}. \end{cases} \]

**Theorem. (Chain Rule)** If the elements of a \( p \) by \( q \) matrix \( Z \) are differentiable functions of an \( r \) by \( s \) matrix \( Y \), and the elements of \( Y \) are differentiable functions of the elements of an \( m \) by \( n \) matrix \( X \), then

\[ \frac{\partial Z}{\partial X} = \frac{\partial Y}{\partial X} \frac{\partial Z}{\partial Y}. \]

**Theorem. (Product Rule)** Suppose each element of a \( p \) by \( r \) matrix \( Y \) and an \( r \) by \( q \) matrix \( Z \) is a differentiable function of the elements of an \( m \) by \( n \) matrix \( X \), then

\[ \frac{\partial (YZ)}{\partial X} = \frac{\partial Y}{\partial X} (I_p \otimes Z) + \frac{\partial Z}{\partial X} (Y' \otimes I_q). \]

**Theorem.** Let \( X \) be a \( p \) by \( p \) matrix, \( A \) and \( B \) are constant matrices of dimensions \( q \) by \( p \) and \( p \) by \( r \) respectively, then

\[ \frac{\partial AXB}{\partial X} = A' \otimes B \]

\[ \frac{\partial (AXB)^2}{\partial X} = (I_p \otimes X) + (X' \otimes I_p) \]

\[ \frac{\partial \text{tr}(X)}{\partial X} = \text{vec } I_p. \]
Theorem. Let $Z = YAY'$, where $Y$ is a $p$ by $q$ matrix whose elements are differentiable functions of elements of $X$, and $A$ is a $q$ by $q$ constant symmetric matrix, then

$$\frac{\partial Z}{\partial X} = \frac{\partial Y}{\partial X}(I_p \otimes AY') + \frac{\partial Y'}{\partial X}(AY' \otimes I_p).$$

Theorem. Suppose $A$ is an $m$ by $n$ matrix, $B$ is a $p$ by $q$ matrix, $C$ is an $n$ by $r$ matrix, $D$ is a $q$ by $s$ matrix, $F$ is an $n$ by $q$ matrix, $G$ is an $m$ by $m$ matrix, and $H_1$ and $H_2$ are nonsingular matrices, then

\begin{align*}
(1.7) \quad & (A \otimes B)(C \otimes D) = AC \otimes BD \\
(1.8) \quad & (A \otimes B)' = A' \otimes B' \\
(1.9) \quad & (H_1 \otimes H_2)^{-1} = H_1^{-1} \otimes H_2^{-1} \\
(1.10) \quad & (A \otimes B)\text{ vec } F = \text{ vec}(AFB') \\
(1.11) \quad & E_{mn}\text{ vec } A = \text{ vec } A' \\
(1.12) \quad & E_{pq}(A \otimes B) = (B \otimes A)E_{pq} \\
(1.13) \quad & E_{mn}' = E_{mn} \\
(1.14) \quad & E_{mn}E_{mn}' = E_{mn}E_{mn} = I_{mn} \\
(1.15) \quad & T_m(A \otimes G)E_{mn} = T_m(G \otimes A) \\
(1.16) \quad & K_n(A \otimes G)E_{mn} = K_n(G \otimes A). 
\end{align*}
Chapter 2. *Multiple group covariance structure analysis*

I. Model

We consider a set of $m$ populations $Q_1, Q_2, \ldots, Q_m$, which may be different nations, or culturally different treatments etc. Let $x_g$ be a random vector of order $p_g$, representing the measurement obtained in population $g$, with $g = 1, 2, \ldots, m$. It is assumed that $x_g$ has a multivariate normal distribution with unconstrained mean vector $\mu_g$ and covariance matrix $\Sigma_g = \Sigma_g(\beta_g)$, whose elements are differentiable functions of the true though unknown $\beta_g$ by 1 parameter $\beta_g$. Let $\theta_g = (\beta_1^g, \beta_2^g, \ldots, \beta_m^g)'$ denote the parameters of interest in the model. The order of $\theta_g$ is $q$ by 1, where $q = t_1 + t_2 + \ldots + t_m$. The parameters in $\theta_g$ are estimated from the $m$ samples, each of which consists of $N_g$ independent observations on the random vector $x_g$. Let $S_g$ be the usual sample covariance matrix of order $p_g \times p_g$. It follows that $S_g$ has a Wishart distribution $W(\Sigma_g, n_g)$ with $n_g = N_g - 1$. We regard the vectors $\beta_g$, $g = 1, \ldots, m$ as mathematical variables, $\Sigma_g = \Sigma_g(\beta_g)$ as a function of $\beta_g$, and let $\theta' = (\beta_1', \ldots, \beta_m')$. Follows the arguments from Browne (1974), the GLS criterion for estimating $\beta_g$ is given by the function

\[(2.1) \quad Q_g(\beta_g) = \frac{1}{2} n_g \text{tr} \left[ (S_g - \Sigma_g) V_g \right]^2,
\]

where $V_g$ is a positive definite constant matrix or it is a stochastic matrix which converges in probability to a positive definite matrix. In our example, we take $V_g = S_g^{-1}$. Since the measurements obtained are
independent, the overall fit function is

\[(2.2) \quad G(\theta) = \frac{1}{N} \sum_{g=1}^{m} G(g) \]

where \( N = n_1 + n_2 + \ldots + n_m \). The GLS estimates of \( \theta^* \), provided it exists, is defined as the vector that minimizes the function \( G(\theta) \).

II. Identification

Before an attempt is made to estimate the parameters in

\( \theta^* = (\beta_{11}, \beta_{22}, \ldots, \beta_{mm})' \)

we will first discuss the identification problem. The problem is that whether or not \( \theta \) can be uniquely determined by the covariance matrices \( \Sigma_1, \Sigma_2, \ldots, \Sigma_m \) for the \( m \) populations. Every \( \theta \) in the admissible parameter space generates a set of \( m \) covariance matrices \( \Sigma_1, \ldots, \Sigma_m \), but two or more \( \theta \)'s may possibly generate the same set of covariance matrices. The whole model is said to be identified if for any two vectors \( \theta^{(1)} = (\beta^{(1)}_1, \ldots, \beta^{(1)}_m)' \) and \( \theta^{(2)} = (\beta^{(2)}_1, \ldots, \beta^{(2)}_m)' \) in a region of the parameter space, \( \theta^{(1)} \neq \theta^{(2)} \) implies that at least one covariance matrix is not equal in the two sets of \( m \) covariance matrices, i.e.

\[ \{ \Sigma_1(\beta^{(1)}_1), \ldots, \Sigma_m(\beta^{(1)}_m) \} \neq \{ \Sigma_1(\beta^{(2)}_1), \ldots, \Sigma_m(\beta^{(2)}_m) \} \].

In other words, if the set of \( m \) covariance matrices are generated by one and only one parameter vector \( \theta^* \), the model is said to be identified.

If the model is not identified, appropriate restrictions should be imposed on the parameters to make it identified. The number of restrictions required depends on the choice of the covariance model. It is hard to give
further specific rules in the general case. For the specific case when
the covariance model is defined, a more precise consideration of the
identification problem will be given in Chapter 4.

III. Constrained Estimation and the Asymptotic Properties

Except identifying the model by adding some conditions on the
parameters, we may have additional knowledge that the true parameters
satisfy some functional relationship. In this case, the parameters in
\( \theta^* \) are estimated under certain constraints, 
\[ h(\theta)' = \{ h_1(\theta), \ldots, h_r(\theta) \} = 0, \]
where \( h_1, h_2, \ldots, h_r \) are independent differentiable functions with
\( r \leq q \). Therefore, the constrained GLS estimates of \( \theta^* \), provided it
exists, is the vector \( \hat{\theta} \) which satisfies \( h(\theta) = 0 \) and minimizes the
function \( G(\theta) \).

By the first order necessary condition, if \( \hat{\theta} \) exists, there exists
a vector \( \hat{\lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_r)' \) of Lagrange multipliers such that
\[
G(\hat{\theta}) + L(\hat{\theta}) \hat{\lambda} = 0 \quad \quad h(\hat{\theta}) = 0
\]
where \( G = (\partial G/\partial \theta_1, \ldots, \partial G/\partial \theta_q)' \), is the gradient vector of \( G \), with
\( \partial G/\partial \theta_i \) be the partial derivative of \( G \) with respect to the \( i \)-th parameter
in \( \theta \), and.
is the partial derivatives of \( h(\theta) \) with respect to \( \theta \). This \( q \times r \) matrix can be considered as forming by \( t \times r \) matrices which are the partial derivatives of \( h(\theta) \) with respect to \( \beta \) since \( \theta = (\beta_1', \beta_2', ..., \beta_m')' \), i.e.

\[
L = (L_1', L_2', ..., L_m')'
\]

with \( L_g = (\partial h / \partial \beta_g) \) represents the \( t \times r \) matrix of partial derivatives.

Under mild regularity conditions, it can be shown that (Lee and Tsui, 1961) the constrained GLS estimator \( \hat{\theta} \) possesses the following asymptotic properties:

**Property 1:** The vector \( \hat{\theta} \) is a consistent estimator of \( \theta^* \).

**Property 2:** The joint asymptotic distribution of random variables \( \sqrt{n}(\hat{\theta} - \theta^*) \) and \( \sqrt{n}\lambda \) is multivariate normal with zero mean vector and covariance matrix:

\[
2 \begin{pmatrix}
F^* & 0 \\
0 & -R^*
\end{pmatrix}
\]

where \( F^* \) and \( R^* \) are defined by.
where $M = 21$, with $I$ denotes the information matrix of (3) and $L = L(\beta)$. It follows that the standard error of $\theta_i$, $S.E. (\hat{\theta}_i)$, can be obtained from the square root of $N^{-1}p_i$, where $p_i$ denotes the $i$-th diagonal element of $P$. So that the asymptotic distribution of the random variable $z_i = (\hat{\theta}_i - \theta_i)/S.E. (\hat{\theta}_i)$ is standard normal. The approximate confidence interval of the parameters $\theta_i$ can be obtained from the normal table using the standard method.

Property 3: The asymptotic distribution of $NG(\hat{\theta})$ is chi-square with degrees of freedom $\sum_i^g (p_i + 1)/2 - (q - r)$. This property provides a goodness-of-fit test statistic for the hypothesis that the proposed model and the relationships $h(\theta) = 0$ fit the observed sample data.
Property 4: The asymptotic distribution of $N[G(\hat{\theta}) - G(\hat{\theta}^+)]$ is chi-square with degrees of freedom $r - j$, where $\hat{\theta}^+$ is the GLS estimator subject to constraints $h^+(\theta)' = \{h_1^+(\theta), \ldots, h_j^+(\theta)\} = 0$, where $h_i^+(\theta) \in \{h_1(\theta), \ldots, h_r(\theta)\}$ for all $1 \leq i \leq j$.

The validity of various constraints can be assessed by means of this property.
Chapter 3. Minimization Procedures

I. Multiplier method

In this section we will discuss the application of the multiplier method to the equality constrained problems, i.e. minimize function \( G(\theta) \) subject to constraints

\[
h_1(\theta) = h_2(\theta) = \ldots = h_r(\theta) = 0.
\]

There are many classical algorithms that can be regarded as candidate in obtaining the constrained estimates. For example, the projection method, penalty function method and the Lagrangian multiplier approach. In this thesis, we use the newly developed multiplier method (Bertsekas, 1976).

This method was developed from the modification of penalty function method and the Lagrangian approach. It has been shown that (Bertsekas, 1976; Lee, 1980) this method is more efficient than the penalty function method.

The multiplier method consists of the sequential unconstrained minimization of the Augmented Lagrangian function

\[
G_k(\theta) = G(\theta) + \sum_{t=1}^{r} \lambda_t h_t(\theta)
\]

(3.1)

where \( \{\lambda_t\} \) is an increasing sequence of positive scalars such that \( \lambda_k \to \infty \) as \( k \to \infty \), \( f(x) \) is a positive differentiable function such that \( f(x) > 0 \) for all \( x \) and \( f(x) = 0 \) if and only if \( x = 0 \), and \( \{\lambda_t\} \) is a sequence of \( n \times 1 \) vectors of multipliers. This sequence is generated according to the iteration.
where \( f \) is the first derivative of \( f \), assumed to exist, and \( \theta_k \) is the vector that minimizes the Augmented Lagrangian function \( G_k(\theta) \) at stage \( k \). The initial multiplier vector \( \theta_0 \) is selected a priori and the sequence \( \{C_k\} \) may be either preselected or generated according to some scheme.

The main steps of the multiplier method are described as follows:

**Step 1:** Set \( k = 1 \) and assign initial values to \( C_1, \xi_1, \) and \( \theta(0) \).

**Step 2:** From the known values of \( C_k, \xi_k \), and \( \theta(k-1) \), search for a minimum point \( \theta_k \) of \( G_k(\theta) \) by an unconstrained minimization technique.

**Step 3:** Increase \( C_k \) to \( C_{k+1} \) and compute \( \xi_{k+1} \) according to (3.2), then update \( k \) and return to step 2. The process is terminated at \( (k+1) \)th iteration if the maximum of the absolute differences of \( \theta_i(k) \) and \( \theta_i(k+1) \) is sufficiently small.

Under mild assumptions, it has been shown that (Bertsekas, 1976) for sufficiently large \( C_k \), the sequence \( \{\theta(k)\} \) and \( \{\xi_k\} \) converge respectively to the solution and Lagrange multipliers of the equality constrained minimization problem.

II. Fletcher-Reeves Algorithm (Fletcher and Reeves, 1964)

In the procedures of the multiplier method (step 2), we need to
minimize the function $G_k(\theta)$ with respect to $\theta$. There are many powerful algorithms that can be used in minimizing the function, such as Gauss-Newton, Newton-Raphson, Fletcher-Powell and Fletcher-Reeves. In this thesis, we use the well-known Fletcher-Reeves (1964) algorithm, which is one of the conjugate direction methods.

In the minimization of $G_k(\theta)$ with respect to a $q \times 1$ vector $\theta$, the algorithm procedures locate the minimum $\hat{\theta}$ of the function $G_k(\theta)$ as the limit of a sequence $\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \ldots$, where $\theta^{(0)}$ is an initial approximation to the position of the minimum. In this sequence, for each $i \geq 0$, $\theta^{(i+1)}$ is the position of the minimum of $G_k(\theta)$ with respect to variations along the line through $\theta^{(i)}$ in the specified direction $d_i$, i.e.

$$\theta^{(i+1)} = \theta^{(i)} + t_i d_i$$

where scalar parameter $t_i$ is determined such that

$$G_k(\theta^{(i+1)}) = G_k(\theta^{(i)} + t_i d_i)$$

is minimized.

In the conjugate direction method, the directions $d_0, d_1, d_2, \ldots$ are selected to be $H$-conjugate to each other, where $H$ is the matrix of second-order partial derivatives of $G_k(\theta)$. Below, we show that this method can locate the minimum in at most $q$ iterations when the objective function is quadratic.

Consider the Taylor expansion of $G_k(\theta)$ at $\hat{\theta}$:
where \( \theta \) is the required minimum. We omit the higher terms by taking approximation. Since \( \dot{G}_K(\theta) = 0 \), we have

\[
\dot{G}_K(\theta) = G_K(\theta) + (\theta - \hat{\theta})'H(\theta - \hat{\theta})
\]

for which the gradient is.

\[
\dot{G}_K(\theta) = H(\theta - \hat{\theta})
\]

Let the gradient vector at the \( i \)-th iteration be \( \dot{G}_K(i) \), then the step from \( \theta(i) \) to \( \theta(i+1) \) as described above is determined by the relation that the derivative of \( G_K(\theta(i+1)) \) with respect to \( t_i \) is equal to zero, i.e.

\[
\dot{G}_K(i+1)'d_i = 0
\]

where \( \dot{G}_K(i+1) = \dot{G}_K(\theta(i+1)) \).

By repeated use of equation (3.3)

\[
\theta(q) = \theta(j+1) + \sum_{i=j+1}^{q-1} t_i d_i
\]

for any \( j \) in \( 0 \leq j < q-1 \).

By applying (3.5) in \( \theta(q) \)

\[
\dot{G}_K(q) = \dot{G}_K(\theta(q))
\]

\[
= H(\theta(j+1) + \sum_{i=j+1}^{q-1} \dot{t}_i d_i - \hat{\theta})
\]

\[
= H(\hat{\theta}(j+1) - \hat{\theta}) + \sum_{i=j+1}^{q-1} \dot{t}_i d_i
\]
\[ \mathbf{G}_k (j+1) + \sum_{i=j+1}^{q-1} t_i H d_i. \]

Taking the dot product with \( d_j \) in both sides, and by (3.6),
\[ \mathbf{G}_k (q)' d_j = \mathbf{G}_k (j+1)' d_j + \sum_{i=j+1}^{q-1} t_i d_i' H d_j, \]
\[ = \sum_{i=j+1}^{q-1} t_i d_i' H d_j. \]

Since \( d_0, d_1, \ldots, d_{q-1} \) are defined to be \( H \)-conjugate, i.e.,
\[ d_i' H d_j = 0 \quad \text{for} \quad i \neq j, \]
then
\[ \mathbf{G}_k (q)' d_j = 0 \]
for \( j = 1, 2, \ldots, q-1 \).

And since \( d_0, d_1, \ldots, d_{q-1} \) form a basis, we have
\[ \mathbf{G}_k (q)' = \mathbf{G}_k (q)' = 0. \]

Thus \( \theta (q) \) is our solution, i.e.
\[ \theta (q) = \hat{\theta}. \]

This demonstrates that the algorithm of conjugate direction method is quadratically convergent and the minimum can be located at the \( q \)th iteration. For the general objective functions, it takes considerably more than \( q \) iterations to converge. From practical experience, it is suggested that the method will be more stable and efficient if it is restarted periodically after every \( q \) iterations in the process, i.e., \( d_{q+1} \) is set equal to \( -\mathbf{G}_k (q+1) \) instead of its usual value.
Though the directions defined in conjugate direction method are only required to be \( H \)-conjugate, there are many kinds of conjugate directions defined for different methods. In the Fletcher-Reeves algorithm, directions \( d_0, d_1, d_2, \ldots \) are generated such that \( d_{i+1} \) is a linear combination of \( -G_k \) and which possess the \( H \)-conjugate property. A full and lucid description has been given by Beckman (1960). Following that the conjugate direction at the \( (i+1) \)th iteration is defined to be:

\[
(3.7) \quad d_{i+1} = -G_k^{(i+1)} + b_i d_i
\]

where

\[
b_i = \frac{G_k^{(i+1)'(i+1)}}{G_k^{(i)'(i)}}
\]

The main steps of the Fletcher-Reeves method are described as follows.

**Step 1:** Start with initial value \( \theta^{(0)} \).

**Step 2:** Compute initial gradient vector \( G_k^{(0)} \) and set initial direction \( d_0 = -G_k^{(0)} \), set \( i = 1 \).

**Step 3:** Compute

\[
\theta^{(i)} = \theta^{(i-1)} + t_{i-1} d_{i-1}
\]

where \( t_{i-1} \) is determined by a linear search method such that \( \theta^{(i)} \) minimize the function \( G_k(\theta^{(i)}) = G_k(\theta^{(i-1)} + t d_{i-1}) \) with respect to parameter \( t \).

**Step 4:** If RMS of \( t_{i-1} d_{i-1} \) is less than a small positive number \( \varepsilon \), go to step 6; otherwise compute \( G_k^{(i)} \).

**Step 5:** If RMS of \( G_k^{(i)} \) is less than \( \varepsilon \), go to step 6; otherwise compute \( d_i \) according to (3.7).
Step 6: Print output.

It should be noticed that this algorithm only requires storage of three vectors of dimension \( q \). This is one of the reasons why we choose the Fletcher-Reeves algorithm in our thesis.

However, if estimates of standard error of the parameter estimates are required, additional storage is necessary. Following the asymptotic property 2, the standard error estimates can be obtained from the diagonal elements of \( \hat{P} = P(\hat{\theta}) \). Let \( \hat{B} = B(\hat{\theta}) \), \( \hat{L} = L(\hat{\theta}) \) and \( \hat{I} = I(\hat{\theta}) \), it can be shown that (see e.g. Lee and Bentler, 1980).

\[
(3.8) \quad \hat{P} = \hat{B}^{-1} - \hat{B}^{-1} \hat{L} (\hat{L}' \hat{B}^{-1} \hat{L})^{-1} \hat{L}' \hat{B}^{-1}.
\]

And since \( \hat{B} \) is a diagonal block matrix, the \( g \)-th diagonal block of \( \hat{P} \) is given by

\[
(3.9) \quad \hat{P}_g = \frac{1}{2} \left\{ (n \hat{I}^{-1}_g) - (n \hat{I}^{-1}_g) L_g^{-1} L_g^{-1} L_j^{-1} L_j^{-1} L_j^{-1} L_j^{-1} (n \hat{I}^{-1}_g) \right\}
\]

for \( g = 1, 2, \ldots, m \).

Thus the standard error estimates of the parameter estimates for each population can be obtained from the main diagonal elements of the matrix \( \hat{P}_g \). Then the computation can be proceeded group by group and thus save a lot of storage and computer time.

III. Reparametrization method

Under certain situations, the equality constrained minimization
The statistical relationships between the reparametrization technique and the Lagrangian approach have been pointed out by Lee and Bentler (1980) in the context of single group CSA. Using similar arguments, analogous relations can be shown in the presented model. It should be noticed that the reparametrization technique may encounter great difficulties with some nonlinear constraints. For example, it cannot handle constraints like \( \theta_1^2 + \theta_2^2 = 1 \), because it is not possible to express \( \theta_1 \) in terms of \( \theta_2 \) uniquely. Under this situations, the multiplier method is recommended. However, the computing time required in the reparametrization method is less than that in multiplier method because no sequential minimization procedure is required. Therefore, if the constraints can be handled by the reparametrization method, we prefer to use it in order to save computing time.
Similarly, in the reparametrization technique, the Fletcher-Reeves algorithm can be applied to minimize the function $G^*(y)$ with respect to $\gamma$. Thus the estimates $\hat{y}$ can be obtained by the same procedure as described in last section.
Chapter 4. Multiple group confirmatory factor analysis

I. The Model and Identification.

The multiple group factor analysis model is discussed in details in this chapter. This model can be regarded as a special case of the multiple group CSA. We consider the m populations which have the same distributions as defined in Chapter 2. It is assumed that a CFA model holds in each population so that $x_g$ can be accounted for by $k_g$ common factors $f_g$ and $p_g$ unique factors $z_g$:

$$x_g = \mu_g + \Lambda_g f_g + z_g$$

where $\varepsilon(x_g) = \mu_g$, $\varepsilon(f_g) = 0$, $\varepsilon(z_g) = 0$ and $z_g$ is uncorrelated with $f_g$. Let $\Phi_g = \varepsilon(f_g f_g')$ be the covariance matrix of $f_g$, and $\Psi_g = \varepsilon(z_g z_g')$ which is a diagonal covariance matrix. The covariance matrix $\Sigma_g$ of $x_g$ becomes

$$(4.1) \quad \Sigma_g = \Lambda_g \Phi_g \Lambda_g' + \Psi_g.$$ 

We also assume that $\mu_g$ is unrestricted for each population. Thus $\mu_g$ can be estimated by the sample mean of the observations. And the remaining parameter matrices $\Lambda_g$, $\Phi_g$ and $\Psi_g$ are estimated from the sample covariance matrices of the observed data from the m populations.

However, before an attempt is made to estimate the parameters, the identification problem must be examined. The identification of a model depends on the specification of fixed, free and constrained parameters in the model. In the CFA model, it is known that each $\Lambda_g$, $\Phi_g$ and $\Psi_g$...
generates one $\Sigma_g$ but different $\Lambda_g$ and $\Phi_g$ can generate the same $\Sigma_g$. Let $\Lambda_g^* = \Lambda_g T_g^{-1}$ and $\Phi_g^* = T_g \Phi_g T_g^T$, where $T_g$ is an arbitrary nonsingular matrix of order $k_g$. Then

$$\Sigma_g^* = \Lambda_g^* \Phi_g^* \Lambda_g^* + \Sigma_g$$

$$= (\Lambda_g T_g^{-1})(\Phi_g T_g^T)(\Lambda_g T_g^{-1}) + \Sigma_g$$

$$= \Lambda_g \Sigma_g \Lambda_g + \Sigma_g$$

$$= \Sigma_g.$$ 

Thus, both $\Lambda_g$, $\Phi_g$, $\Sigma_g$ and $\Lambda_g^*$, $\Phi_g^*$, $\Sigma_g^*$ generate the same $\Sigma_g$. Since $T_g$ has $k_g^2$ independent elements, this suggests that at least $k_g^2$ appropriate restrictions should be imposed on $\Lambda_g$ and/or $\Phi_g$ to make it uniquely defined (Jöreskog, 1969). Therefore, for the $m$ populations, at least $\sum_{g=1}^{m} k_g^2$ restrictions should be imposed appropriately to make the whole model identified.

The restrictions imposed on the parameters to achieve identifiability are divided into two kinds. The first kind of restrictions are imposed by fixing certain elements in $\Lambda_g$, $\Phi_g$ and $\Sigma_g$ to some known values. These values, which usually have special interpretations, are obtained from exploratory studies. Once these elements are fixed, they are not considered as unknown parameters anymore. Thus, we are only interested in estimating the remaining elements in $\Lambda_g$, $\Phi_g$ and $\Sigma_g$ for each population. We denote these parameters by the parameter vector $\theta$.

Another kind of restriction are obtained by imposing functional
constraints on \( \theta \). That is, we require parameters in \( \theta \) satisfy
\[
h(\theta) = 0.
\]
For example, when \( p_1 = p_2 = \ldots = p_m = p \), we may consider a model that the whole factor pattern is invariant over groups, i.e.
\[
\Lambda_g = \Lambda \quad \text{for} \quad g = 1, 2, \ldots, m.
\]
In this case, \( k_1 = k_2 = \ldots = k_m = k \), the matrices \( \Sigma_g \) and \( \Psi_g \) are all of order \( p \times p \) and \( \phi_g \) are all of order \( k \times k \). The common factor pattern \( \Lambda \) is of order \( p \times k \). Thus the covariance matrices become
\[
\Sigma_g = \Lambda \phi_g \Lambda' + \Psi_g, \quad g = 1, 2, \ldots, m.
\]
It is noted that if \( \Lambda \) is replaced by \( \Lambda^* = \Lambda T^{-1} \) and each \( \phi_g \) is replaced by \( \phi_g^* = T \phi_g T' \), \( g = 1, 2, \ldots, m \), where \( T \) is an arbitrary non-singular matrix of order \( k \times k \), then \( \Sigma_g \) is unchanged for \( g = 1, 2, \ldots, m \). Since the matrix \( T \) has \( k^2 \) independent elements, the identifiability of the model can be achieved by fixing \( k^2 \) elements appropriately. Therefore, the whole model is identified by imposing equality constraints
\[
\Lambda_1 = \ldots = \Lambda_m = \Lambda \quad \text{and} \quad k^2 \text{ fixed elements on } \Lambda.
\]
In this case, the total number of restrictions is \((m-1)pk + k^2\).

II. Estimation of the Model

The constrained estimates of \( \theta \) are obtained by minimizing the function
\[
G(\theta) = \frac{1}{2N} \sum_{g=1}^{m} n_g \text{tr}\{(S_g - E_g) S_g^{-1}\}^2
\]
subject to constraints \( h(\theta) = 0 \). This constrained minimization problem is handled by the multiplier method with the Fletcher-Reeves algorithm.
In the procedures of multiplier method, we have to minimize the Augmented Lagrangian function $G_k(\theta)$. According to Fletcher-Reeves algorithm, the derivatives of $G_k(\theta)$ with respect to $\theta$ are required. Differentiating $G_k(\theta)$ with respect to $\theta$, we have

$$
\dot{G}_k(\theta) = \dot{G}(\theta) + c_k \sum_{t=1}^{r} \frac{\partial h_t}{\partial \theta} + \sum_{t=1}^{r} \dot{G}_k(t) \frac{\partial h_t}{\partial \theta}.
$$

Thus, $\dot{G}_k(\theta)$ can be obtained from $\dot{G}(\theta)$ and the derivatives of $h_t$ with respect to $\theta$. Since $\frac{\partial h}{\partial \theta}$ and $f$ are easy to obtain, we only derive the derivatives of $G(\theta)$ with respect to $\theta$. According to the matrix calculus in Chapter 1, the derivation are shown as follows.

We first arrange the distinct elements in $\Lambda_\theta$, $\Phi_\theta$ and $\Psi_\theta$ to a vector and denote it as $\eta_\theta$. In matrix $\Phi_\theta$, we only require the elements in the lower half and diagonal since the matrix is symmetric. Also, we only require the diagonal elements in $\Psi_\theta$ since the non-diagonal elements are zero in the model. Thus we have

$$
\eta_\theta = \begin{pmatrix}
\text{vec}(\Lambda_\theta) \\
\Phi_\theta \text{ sym} \\
\Psi_\theta \text{ diag} \end{pmatrix}.
$$

Differentiating (4.2) by (1.1), (1.2), (1.4), (1.5), (1.7) and (1.9), we have

$$
\frac{\partial G}{\partial \pi} = \frac{n_k}{2N} \frac{\partial \text{tr}(I_{p_\theta} - E_{\epsilon} E_{\epsilon}^{-1})^2}{\partial \pi} = \frac{n_k}{2k} \frac{\partial (I_{p_\theta} - E_{\epsilon} E_{\epsilon}^{-1})}{\partial \pi} \frac{\partial \text{tr}(I_{p_\theta} - E_{\epsilon} E_{\epsilon}^{-1})^2}{\partial (I_{p_\theta} - E_{\epsilon} E_{\epsilon}^{-1})} \frac{\partial \text{tr}(I_{p_\theta} - E_{\epsilon} E_{\epsilon}^{-1})^2}{\partial (I_{p_\theta} - E_{\epsilon} E_{\epsilon}^{-1})^2}.
$$
From (4.0) and (1.1) and (1.6), the components of $\frac{\partial \Sigma}{\partial \eta}$ can be obtained as follows.

$$\frac{\partial \Sigma}{\partial \eta} = \frac{-n}{N} \frac{\partial \gamma}{\partial \eta} \text{vec.} \{S^{-1}_g (S_E - \Sigma E) S^{-1}_g \}.$$  \hspace{1cm} \text{(4.5)}$$

Similarly, we have

$$\frac{\partial \Sigma}{\partial \gamma} = T_{k} \text{vec.} \{\Lambda'_g \otimes \Lambda'_g \}$$

By substitution of these derivatives to (4.5) and by (1.10), (1.11),

$$\frac{\partial G}{\partial \Lambda'_g} = \frac{-2n}{N} \text{vec.} \{S^{-1}_g (S_E - \Sigma E) S^{-1}_g \}$$

$$\frac{\partial G}{\partial \gamma} = \frac{-n}{N} T_{k} \text{vec.} \{\Lambda'_g \otimes \Lambda'_g \}$$

$$\frac{\partial G}{\partial \gamma} = \frac{-n}{N} K_{p_{g_{E}}} \text{vec.} \{S^{-1}_g (S_E - \Sigma E) S^{-1}_g \}.$$  

Since $\gamma$ is a vector denoting the non-fixed elements in $\eta$, the derivatives $\frac{\partial G}{\partial \gamma}$ can be obtained by omitting the elements in $\frac{\partial G}{\partial \eta}$.
corresponding to the fixed elements in \( \pi_g \). Also, since

\[ \theta = (\beta_1', \beta_2', \ldots, \beta_m')', \]

the derivatives \( \frac{\partial G}{\partial \beta_g} \) are obtained from \( \frac{\partial G}{\partial \beta_g} \) for \( g = 1, 2, \ldots, m \).

Similarly, in the reparametrization method, since we have to

minimize the function \( G^*(y) \) with respect to \( y \), the derivatives \( \dot{G}^*(y) \) are required. These derivatives can also be obtained from the derivatives \( \dot{G}(\theta) \). Since

\[
\begin{align*}
\dot{G}^*(y) &= G[\dot{\xi}(y)] \\
&= \frac{\partial G[\dot{\xi}(y)]}{\partial y} \\
&= \frac{\partial \dot{\xi}(y)}{\partial y} \dot{G}(\theta).
\end{align*}
\]

Thus, \( \dot{G}^*(y) \) can be obtained from \( \dot{G}(\theta) \) by multiplying the partial derivatives of \( \theta = \dot{\xi}(y) \) with respect to \( y \).

Moreover, from the asymptotic property 2, the estimates of standard error are required in constructing the confidence interval of the parameters. These standard error estimates can be obtained from the information matrix according to \((3,9)\). The derivation of the information matrix are shown as follows.

Since the information matrix for the \( g \)-th population is defined to be the expectation of the second derivatives of \( G \) with respect to \( \beta_g \),

\[
I_g = \mathbb{E} \left( \frac{\partial^2 G}{\partial \beta_g \partial \beta_g} \right)
\]

we first derive the second derivative of \( G \) with respect to \( \pi_g \).
Using the technique of matrix calculus, it can be shown that
\[
\frac{\partial^2 G}{\partial \pi E \partial \pi G} = -\frac{n}{N} \frac{\partial}{\partial \pi E} \left[ \text{vec}(S^{-1}(S^{-1} - \pi E)S^{-1}) \right]
\]
\[
= -\frac{n}{N} \frac{\partial^2 E}{\partial \pi G} \left[ \text{vec}(S^{-1}(S^{-1} - \pi E)S^{-1}) \right]
\]
\[
= \frac{n}{N} \frac{\partial E}{\partial \pi G} (S^{-1} \otimes S^{-1}) \left( \frac{\partial E}{\partial \pi G} \right).
\]

Since \( \varepsilon(S^{-1} - \pi E) = 0 \), we get
\[
(4.4) \quad A_E = \varepsilon \left( \frac{\partial^2 E}{\partial \pi E \partial \pi G} \right) = \frac{n}{N} \frac{\partial E}{\partial \pi G} (S^{-1} \otimes S^{-1}) \left( \frac{\partial E}{\partial \pi G} \right).
\]

The matrix \( A_E \) can be partitioned to
\[
A_E = \begin{pmatrix}
A_E(\wedge, \wedge) & \text{symmetric} \\
A_E(\phi, \wedge) & A_E(\psi, \psi) \\
A_E(\psi, \wedge) & A_E(\psi, \psi)
\end{pmatrix}
\]

The derivation of these sub-matrices are given as follows.

By substitution of \( \partial E / \partial \wedge E \), \( \partial E / \partial (\phi_E) \) and \( \partial E / \partial (\psi_E) \) to (4.4), and by (1.12), (1.13), (1.14), (1.15) and (1.16), it can be shown that
\[
A_E(\wedge, \wedge) = \frac{n}{N} \frac{\partial E}{\partial \wedge E} (S^{-1} \otimes S^{-1}) \left( \frac{\partial E}{\partial \wedge E} \right),
\]
\[
= \frac{2n}{N} (S^{-1} \otimes \wedge \wedge^{-1} \wedge \wedge + \wedge \wedge^{-1} \wedge \wedge G + \psi_E \wedge \wedge^{-1} \wedge \wedge G) 
\]
\[
A_E(\phi, \wedge) = \frac{n}{N} \frac{\partial E}{\partial (\phi_E)} (S^{-1} \otimes S^{-1}) \left( \frac{\partial E}{\partial (\phi_E)} \right),
\]
\[
A_E(\psi, \wedge) = \frac{n}{N} \frac{\partial E}{\partial (\psi_E)} (S^{-1} \otimes S^{-1}) \left( \frac{\partial E}{\partial (\psi_E)} \right).
\]
Therefore, the information matrix $I_g$ for the $g$-th population can be obtained from $A_g$ by omitting the rows and columns corresponding to the fixed elements in $\psi_g$ and $\phi_g$, and thus we can compute the standard error estimates by (3.9).

From the above expressions, the individual components of $G_k$ and the information matrix can be obtained as follows. For simplicity, the subscript $g$ denoting the $g$-th population is omitted.

\begin{align*}
    G_k(\wedge_{ij}) &= -\frac{2n}{N} [S^{-1}(S - E)S^{-1}\wedge_{g}]_{ij} \\
    &= -\frac{2n}{N} T_{g} (\wedge_{g} S^{-1} \otimes \wedge_{g} S^{-1} \wedge_{g} \phi_{g}) \\
    A_{g}(\psi_{g}, \wedge_{g}) &= \frac{n}{N} \frac{\partial \Sigma_{g}}{\partial (\psi_{g})} (S^{-1} \otimes S^{-1}) \left( \frac{\partial \Sigma_{g}}{\partial \psi_{g}} \right) \\
    &= \frac{2n}{N} \frac{\partial \Sigma_{g}}{\partial (\psi_{g})} (S^{-1} \otimes S^{-1} \wedge_{g} \phi_{g}) \\
    A_{g}(\phi_{g}, \wedge_{g}) &= \frac{n}{N} \frac{\partial \Sigma_{g}}{\partial (\phi_{g})} (S^{-1} \otimes S^{-1}) \left( \frac{\partial \Sigma_{g}}{\partial \phi_{g}} \right) \\
    &= \frac{n}{N} T_{g} (\wedge_{g} S^{-1} \wedge_{g} \otimes \wedge_{g} S^{-1} \wedge_{g}) \\
    A_{g}(\psi_{g}, \psi_{g}) &= \frac{n}{N} \frac{\partial \Sigma_{g}}{\partial (\psi_{g})} (S^{-1} \otimes S^{-1}) \left( \frac{\partial \Sigma_{g}}{\partial \psi_{g}} \right) \\
    &= \frac{n}{N} K_{g} (S^{-1} \wedge_{g} \otimes S^{-1} \wedge_{g}) T_{g} \\
    A_{g}(\psi_{g}, \phi_{g}) &= \frac{n}{N} \frac{\partial \Sigma_{g}}{\partial (\psi_{g})} (S^{-1} \otimes S^{-1}) \left( \frac{\partial \Sigma_{g}}{\partial \phi_{g}} \right) \\
    &= \frac{n}{N} K_{g} (S^{-1} \wedge_{g} \phi_{g}) T_{g} \\
    A_{g}(\phi_{g}, \phi_{g}) &= \frac{n}{N} \frac{\partial \Sigma_{g}}{\partial (\phi_{g})} (S^{-1} \otimes S^{-1}) \left( \frac{\partial \Sigma_{g}}{\partial \phi_{g}} \right) \\
    &= \frac{n}{N} K_{g} (S^{-1} \otimes S^{-1}) K_{g}^t
\end{align*}
\begin{align}
\mathbf{C}_i^j &= \frac{1}{n} \left( 2 - \delta_{ij} \right) \left[ \Lambda' S^{-1} (S - \Lambda) S^{-1} \Lambda \right]_{ij} \\
&\quad + \sum_{t=1}^{r} h(t) \left[ \Lambda' S^{-1} (S - \Lambda) S^{-1} \Lambda \right]_{ij} \\
&\quad + \sum_{t=1}^{r} \hat{\epsilon}(t) h_t \left( \Lambda' S^{-1} (S - \Lambda) S^{-1} \Lambda \right)_{ij} \\
&\quad + \sum_{t=1}^{r} \hat{\epsilon}(t) h_t \left( \Lambda' S^{-1} (S - \Lambda) S^{-1} \Lambda \right)_{ij} \\
\mathbf{C}_i^i &= \frac{1}{n} \left[ S^{-1} (S - \Lambda) S^{-1} \Lambda \right]_{ii} \\
&\quad + \sum_{t=1}^{r} h(t) \left[ S^{-1} (S - \Lambda) S^{-1} \Lambda \right]_{ii} \\
&\quad + \sum_{t=1}^{r} \hat{\epsilon}(t) h_t \left( S^{-1} (S - \Lambda) S^{-1} \Lambda \right)_{ii} \\
&\quad + \sum_{t=1}^{r} \hat{\epsilon}(t) h_t \left( S^{-1} (S - \Lambda) S^{-1} \Lambda \right)_{ii} \\
A(\Lambda^i_1, \Lambda^j_v) &= \frac{2}{N} \left[ S_{iu} \left( \Lambda' S^{-1} \Lambda \right)_{ju} \right] \\
&\quad + \left( \Lambda' S^{-1} \Lambda \right)_{ju} \\
A(\Lambda^i_1, \Phi^j_v) &= \frac{1}{N} \left( 2 - \delta_{ij} \right) \left( 2 - \delta_{uv} \right) \left[ \Lambda' S^{-1} \Lambda \right]_{ju} \\
&\quad + \left( \Lambda' S^{-1} \Lambda \right)_{ju} \\
A(\Phi^i_1, \Phi^j_v) &= \frac{1}{2N} \left( 2 - \delta_{ij} \right) \left( 2 - \delta_{uv} \right) \left[ \Lambda' S^{-1} \Lambda \right]_{ju} \\
&\quad + \left( \Lambda' S^{-1} \Lambda \right)_{ju} \\
A(\Psi^i_1, \Psi^j_v) &= \frac{1}{N} \left( 2 - \delta_{ij} \right) \left( S^{-1} \Lambda \right)_{ju} \\
&\quad + \left( S^{-1} \Lambda \right)_{ju} \\
&\quad + \left( S^{-1} \Lambda \right)_{ju} \\
&\quad + \left( S^{-1} \Lambda \right)_{ju} \\
&\quad + \left( S^{-1} \Lambda \right)_{ju} \\
&\quad + \left( S^{-1} \Lambda \right)_{ju} \\
where \quad \delta_{uv} \quad \text{denotes the Kronecker delta.}
\end{align}

To save computer storage and time, the computed program is implemented by means of (4.5), (4.6), ..., (4.13).
III. Strategy of analysis on testing hypothesis.

Suppose $H_0$ represents one model under given specifications of fixed, free, and constrained parameters. It is possible, in large samples, to test the model $H_0$ against any more general model $H_1$. The method is to estimate the parameters separately for the two models $H_0$ and $H_1$ and compare their chi-squared goodness-of-fit values.

From the asymptotic property 4, the difference in the chi-squared values is asymptotically distributed as $\chi^2$ with degrees of freedom equal to the corresponding difference in degrees of freedom. Therefore, the similarities and differences of the factor structures can be studied by setting a sequence of hypotheses, such that each hypothesis is a special case of the preceding, and comparing their $\chi^2$ goodness-of-fit values.

For example, we may consider the following sequence of hypotheses.

\[
H_\wedge : \quad \wedge_1 = \wedge_2 = \ldots = \wedge_m = \wedge \\
H_\wedge \psi : \quad \wedge_1 = \wedge_2 = \ldots = \wedge_m = \wedge \\
\quad \text{and} \quad \psi_1 = \psi_2 = \ldots = \psi_m = \psi \\
H_{\wedge \psi} : \quad \wedge_1 = \wedge_2 = \ldots = \wedge_m = \wedge \\
\quad \psi_1 = \psi_2 = \ldots = \psi_m = \psi \\
\quad \text{and} \quad \wedge_1 = \wedge_2 = \ldots = \wedge_m = \wedge
\]

We first test the hypothesis $H_\wedge$. If it is found to be tenable, we may proceed to test the hypothesis $H_{\wedge \psi}$. As described previously, to test
against $H_\Lambda$, we use the statistic.

$$x_{\psi, \Lambda}^2 = x_{\Lambda_\psi}^2 - x_{\Lambda}^2$$

with d.f. $= d_{\Lambda_\psi} - d_{\Lambda}$

where $x_{\Lambda_\psi}^2$ and $x_{\Lambda}^2$ are the corresponding $x^2$ goodness-of-fit values and $d_{\Lambda_\psi}$ and $d_{\Lambda}$ are the degrees of freedom of the hypotheses $H_{\Lambda_\psi}$ and $H_{\Lambda}$.

Do not reject the hypothesis means that constraints $\psi_1 = \ldots = \psi_m$ fit the observed sample data. Similarly, the hypothesis $H_{\Lambda_\psi}$ can be tested.

The above example only involves the special constraints that some parameters are equal over the $m$ populations. However, various other types of hypotheses involving more general constraints may also be tested by the same arguments. For example, instead of testing that the factor pattern is invariant over the $m$ populations, we can test the hypothesis that the factor patterns are proportional to each other over the populations, i.e.

$$H_\Lambda^\circ : \quad c_1 \Lambda_1 = c_2 \Lambda_2 = \ldots = c_m \Lambda_m$$

where $c_1, c_2, \ldots, c_m$ are some proportional constants.

More generally, the hypothesis involving the nonlinear constraints can be similarly tested. In next chapter, numerical examples will be presented to illustrate the method in more details.
Chapter 5. Numerical examples

The methodology discussed in this thesis are illustrated by the following two numerical examples based on the real and artificial data. The constraints in the first example can only be handled by the multiplier method while the second example can be handled by both the reparametrization and multiplier methods.

For the example one, we consider part of the data used by Jöreskog (1971), and Sörbom and Jöreskog (1978). These data consist of measurements of nine tests which were selected so that each of the three factors - space, verbal and memory, would be represented by three tests. The nine tests are (1) Visual Perception, (2) Cubes, (3) Paper Form Board, (4) General Information, (5) Sentence Completion, (6) Word Classification, (7) Figure Recognition, (8) Object Number and (9) Number-Figure. The data are coming from a speeded addition test in two schools and each of the samples from two schools were divided into two approximately equal groups by splitting at the median score within each school. Two groups of data are used in our example and their sample sizes are: Pasteur Low \( N_1 = 77 \) and Pasteur High \( N_2 = 79 \). The sample covariance matrices are presented in Table 1. We assume that the data set fits the CFA model as described in Chapter 4, i.e.

\[
\Sigma_g = \Lambda_g \Psi_g \Lambda_g' + \Psi_g, \quad g = 1, 2.
\]

where \( \Lambda_g, \Psi_g, \Psi_g \) are of dimensions \( 9 \times 3, 3 \times 3, 9 \times 9 \) respectively. We consider the hypothesis \( H_0 \) that the factor pattern \( \Lambda_g \) has
nonoverlapping group structure, i.e.

\[
\Lambda_g = \begin{pmatrix}
\Lambda_g(1, 1) & 0^* & 0^* \\
\Lambda_g(2, 1) & 0^* & 0^* \\
\Lambda_g(3, 1) & 0^* & 0^* \\
0^* & \Lambda_g(4, 2) & 0^* \\
0^* & \Lambda_g(5, 2) & 0^* \\
0^* & \Lambda_g(6, 2) & 0^* \\
0^* & 0^* & \Lambda_g(7, 3) \\
0^* & 0^* & \Lambda_g(8, 3) \\
0^* & 0^* & \Lambda_g(9, 3)
\end{pmatrix}
\]

and

\[
\Phi_g = \begin{pmatrix}
1^* & \text{symmetric} \\
\Phi_g(2, 1) & 1^* \\
\Phi_g(3, 1) & \Phi_g(3, 2) & 1^*
\end{pmatrix}
\]

\[
(\psi_g)_{dg} = [\psi_g(1, 1), \ldots, \psi_g(9, 9)]
\]

where an asterisk indicates that the parameter was fixed at that value.

Moreover, for demonstrating purposes, consider the nonlinear constraints that cannot be handled by reparametrization technique.

(5.1) \[\Sigma i \Lambda_1(i, j)^2 = \Sigma i \Lambda_2(i, j)^2; \quad j = 1, 2, 3.\]

The interpretation of these constraints is that the variance explained by each factor is invariant with respect to different groups.

Based on the Fletcher-Reeves algorithm and the multiplier method,
a computer program has been implemented to produce the constrained GLS estimates. Here, we initially set $C_1 = 1.0$ and increased it by six times after each unconstrained minimization. The initial multiplier vector $\xi_0$ is selected to be a zero vector and we set function $f(x) = \frac{1}{2} x^2$. The starting values of the parameters are the same as in Sorbom and Jöreskog (1976). The program converges after eight unconstrained minimizations. At the final solution, the maximum absolute difference of $\theta_i(7)$ and $\theta_i(8)$ is about 0.00004 and the RMS of the gradient vector $G_k(\theta)$ is less than 0.0001. The convergence summary is presented in Table 2. According to the constrained estimates, the standard error estimates can be computed by equation (3.9). The estimates and the standard error estimates are presented in Table 3. The chi-squared goodness-of-fit statistic of this estimation is equal to 53.05 with degrees of freedom 51. Hence, the proposed model and the constraints fit the observed data.

We now proceed to test the hypothesis $H_{\psi}$ by adding the following constraints on $\psi$:

$$\psi_{1}(i, i) = \psi_{2}(i, i); \quad i = 1, 2, \ldots, 9.$$  

(5.2)

With the same set-up, the program converges after nine unconstrained minimizations. The convergence summary is presented in Table 4 and the estimates together with standard error estimates is presented in Table 5. At the final solution, the maximum absolute difference of $\theta_{i}(8)$ and $\theta_{i}(9)$ is about 0.00001 and the RMS of the gradient vector $G_k(\theta)$ is less than 0.00005. The chi-squared goodness-of-fit statistic is equal to 62.34 with 60 degrees of freedom. The difference chi-squared value of
hypotheses $H_0$ and $H_1$ is 9.29 with 9 degrees of freedom. Based on the test, the constraints (5.2) fit the observed sample data.

In the last example, we consider a simulated data set with three groups. The population parameter matrices are chosen as given in Table 6. Based on the covariance matrices computed according to the CFA model (4.1), the sample covariance matrices were generated by the method described in Kshirsagar (1959) with sample sizes 200, 300 and 400 respectively. These sample covariance matrices are presented in Table 7. We consider the hypothesis that the factor patterns are proportional to each other such that

$$6 \Lambda_1(i, j) = 4 \Lambda_2(i, j) = 3 \Lambda_3(i, j)$$

for $i = 1, \ldots, 6$ and $j = 1, 2$

where the factor pattern is specified to

$$\Lambda_g = \begin{bmatrix}
\Lambda_g(1, 1) & 0^\alpha \\
\Lambda_g(2, 1) & 0^\alpha \\
\Lambda_g(3, 1) & 0^\alpha \\
0^\alpha & \Lambda_g(4, 2) \\
0^\alpha & \Lambda_g(5, 2) \\
0^\alpha & \Lambda_g(6, 2)
\end{bmatrix}$$

and

$$\Phi^*_g = \begin{bmatrix}
1^\alpha \\
\Phi_g(2, 1) \\
1^\alpha
\end{bmatrix}$$

$$(\Phi^*_g)^T = \begin{bmatrix}
\Phi_g(1, 1), \ldots, \Phi_g(6, 6)\end{bmatrix}.$$
For comparison sake, estimates are obtained by both the multiplier method and the reparametrization technique. In multiplier method, we use the same \( \mathcal{C}_1, \mathcal{C}_0 \) and \( f(x) \) as in last example and set \( \mathcal{C}_{k+1} = 12 \times \mathcal{C}_k \). Both two methods converge nicely and the function values converge to the same minimum value. The multiplier method algorithm converges after 6 unconstrained minimization and the reparametrization algorithm converges after 2 iterations. The convergence summaries of these two methods are given in Table 8 and Table 9. In Table 8, we only present the data at the end of each unconstrained minimization. But in Table 9, we present the data after each iteration since there is only one unconstrained minimization here.

It is noted that the estimates from two methods are almost identical. However, the reparametrization procedure was found to be about 2.4 times as fast as the multiplier method. Using the final values, the standard error estimates are computed. As expected, standard error estimates also satisfy similar constraints, i.e.

\[
6\hat{\sigma}^2[\hat{\lambda}_1(i, j)] = 48\hat{\sigma}^2[\hat{\lambda}_2(i, j)] = 38\hat{\sigma}^2[\hat{\lambda}_3(i, j)]
\]

where \( \hat{\lambda}_g(i, j) \) is the estimates of parameter \( \lambda_g(i, j) \). The estimates, together with the standard error estimates are presented in Table 10. The value of the chi-squared statistic is 35.35 with 36 degrees of freedom, indicating that the proposed model and constraints fit the sample data. This is reasonable since the data were simulated from the population satisfying the hypothesis.


Table 1.

Covariance matrices of Data in Sörbom and Jöreskog

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group 1: Pasteur Low</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.126</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.410</td>
<td>1.461</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.523</td>
<td>0.409</td>
<td>1.053</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.308</td>
<td>0.013</td>
<td>0.064</td>
<td>1.072</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.293</td>
<td>0.013</td>
<td>0.011</td>
<td>0.524</td>
<td>1.127</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.420</td>
<td>0.311</td>
<td>0.102</td>
<td>0.615</td>
<td>0.662</td>
<td>0.980</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.523</td>
<td>0.454</td>
<td>0.265</td>
<td>0.182</td>
<td>0.097</td>
<td>0.413</td>
<td>1.377</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.128</td>
<td>0.061</td>
<td>0.031</td>
<td>-0.033</td>
<td>0.064</td>
<td>0.189</td>
<td>0.341</td>
<td>1.004</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.255</td>
<td>-0.050</td>
<td>0.011</td>
<td>-0.054</td>
<td>0.111</td>
<td>0.226</td>
<td>0.233</td>
<td>0.398</td>
<td>1.091</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group 2: Pasteur High</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.923</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.199</td>
<td>0.745</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.219</td>
<td>0.168</td>
<td>0.982</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.297</td>
<td>0.042</td>
<td>0.220</td>
<td>0.931</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.357</td>
<td>0.211</td>
<td>0.189</td>
<td>0.696</td>
<td>1.127</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.349</td>
<td>0.057</td>
<td>0.110</td>
<td>0.575</td>
<td>0.707</td>
<td>1.018</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.214</td>
<td>0.009</td>
<td>-0.070</td>
<td>0.066</td>
<td>0.113</td>
<td>0.123</td>
<td>1.027</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>-0.021</td>
<td>-0.010</td>
<td>-1.143</td>
<td>0.054</td>
<td>0.034</td>
<td>0.214</td>
<td>1.230</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.086</td>
<td>-0.121</td>
<td>-0.059</td>
<td>0.154</td>
<td>0.021</td>
<td>0.121</td>
<td>0.152</td>
<td>0.321</td>
<td>0.998</td>
</tr>
</tbody>
</table>
Table 2.

Convergence Summary of estimation under $H_{g}^{k}$.

<table>
<thead>
<tr>
<th>No. of unconstrained minimization</th>
<th>Function value</th>
<th>$\text{RMS}(\dot{g}_{x})$</th>
<th>$\text{RMS}(\theta^{(i)} - \theta^{(i-1)})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50.9326</td>
<td>$0.2812 \times 10^{-4}$</td>
<td>$0.2778$</td>
</tr>
<tr>
<td>2</td>
<td>51.7272</td>
<td>$0.1077 \times 10^{-4}$</td>
<td>$0.2228 \times 10^{-1}$</td>
</tr>
<tr>
<td>3</td>
<td>52.7556</td>
<td>$0.7779 \times 10^{-4}$</td>
<td>$0.2145 \times 10^{-1}$</td>
</tr>
<tr>
<td>4</td>
<td>53.0389</td>
<td>$0.5118 \times 10^{-4}$</td>
<td>$0.9661 \times 10^{-2}$</td>
</tr>
<tr>
<td>5</td>
<td>53.0516</td>
<td>$0.7857 \times 10^{-4}$</td>
<td>$0.7225 \times 10^{-3}$</td>
</tr>
<tr>
<td>6</td>
<td>53.0518</td>
<td>$0.9105 \times 10^{-4}$</td>
<td>$0.4127 \times 10^{-3}$</td>
</tr>
<tr>
<td>7</td>
<td>53.0518</td>
<td>$0.9132 \times 10^{-4}$</td>
<td>$0.7839 \times 10^{-3}$</td>
</tr>
<tr>
<td>8</td>
<td>53.0518</td>
<td>$0.6223 \times 10^{-4}$</td>
<td>$0.3584 \times 10^{-4}$</td>
</tr>
</tbody>
</table>
Table 3.

Constrained Estimates of the Data in Sörbom and Jöreskog [1976] under $H_\Lambda$

<table>
<thead>
<tr>
<th>Parameter Matrices</th>
<th>$\hat{\Theta}_1$</th>
<th>S.E.($\hat{\Theta}_1$)</th>
<th>$\hat{\Theta}_2$</th>
<th>S.E.($\hat{\Theta}_2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Lambda(1, 1)$</td>
<td>0.805</td>
<td>0.144</td>
<td>0.648</td>
<td>0.100</td>
</tr>
<tr>
<td>$\Lambda(2, 1)$</td>
<td>0.228</td>
<td>0.156</td>
<td>0.423</td>
<td>0.095</td>
</tr>
<tr>
<td>$\Lambda(3, 1)$</td>
<td>0.426</td>
<td>0.194</td>
<td>0.532</td>
<td>0.105</td>
</tr>
<tr>
<td>$\Lambda(4, 2)$</td>
<td>0.870</td>
<td>0.086</td>
<td>0.743</td>
<td>0.080</td>
</tr>
<tr>
<td>$\Lambda(5, 2)$</td>
<td>0.884</td>
<td>0.088</td>
<td>0.965</td>
<td>0.083</td>
</tr>
<tr>
<td>$\Lambda(6, 2)$</td>
<td>0.696</td>
<td>0.088</td>
<td>0.731</td>
<td>0.084</td>
</tr>
<tr>
<td>$\Lambda(7, 3)$</td>
<td>0.778</td>
<td>0.174</td>
<td>0.210</td>
<td>0.127</td>
</tr>
<tr>
<td>$\Lambda(8, 3)$</td>
<td>0.348</td>
<td>0.163</td>
<td>0.765</td>
<td>0.130</td>
</tr>
<tr>
<td>$\Lambda(9, 3)$</td>
<td>0.245</td>
<td>0.153</td>
<td>0.397</td>
<td>0.131</td>
</tr>
<tr>
<td>$\Phi(2, 1)$</td>
<td>0.376</td>
<td>0.255</td>
<td>0.539</td>
<td>0.145</td>
</tr>
<tr>
<td>$\Phi(3, 1)$</td>
<td>0.455</td>
<td>0.175</td>
<td>0.112</td>
<td>0.117</td>
</tr>
<tr>
<td>$\Phi(3, 2)$</td>
<td>0.230</td>
<td>0.145</td>
<td>0.099</td>
<td>0.151</td>
</tr>
<tr>
<td>$\psi(1, 1)$</td>
<td>0.200</td>
<td>0.078</td>
<td>0.432</td>
<td>0.083</td>
</tr>
<tr>
<td>$\psi(2, 2)$</td>
<td>0.911</td>
<td>0.081</td>
<td>0.526</td>
<td>0.079</td>
</tr>
<tr>
<td>$\psi(3, 3)$</td>
<td>0.559</td>
<td>0.070</td>
<td>0.666</td>
<td>0.069</td>
</tr>
<tr>
<td>$\psi(4, 4)$</td>
<td>0.203</td>
<td>0.324</td>
<td>0.321</td>
<td>0.175</td>
</tr>
<tr>
<td>$\psi(5, 5)$</td>
<td>0.261</td>
<td>0.135</td>
<td>0.193</td>
<td>0.190</td>
</tr>
<tr>
<td>$\psi(6, 6)$</td>
<td>0.260</td>
<td>0.133</td>
<td>0.429</td>
<td>0.176</td>
</tr>
<tr>
<td>$\psi(7, 7)$</td>
<td>0.416</td>
<td>0.154</td>
<td>0.653</td>
<td>0.120</td>
</tr>
<tr>
<td>$\psi(8, 8)$</td>
<td>0.544</td>
<td>0.201</td>
<td>0.592</td>
<td>0.198</td>
</tr>
<tr>
<td>$\psi(9, 9)$</td>
<td>0.654</td>
<td>0.176</td>
<td>0.713</td>
<td>0.173</td>
</tr>
</tbody>
</table>
Table 4.

Convergence summary of estimation under $H_0$.

<table>
<thead>
<tr>
<th>No. of unconstrained minimization</th>
<th>Function value</th>
<th>$\text{RMS}(G_k)$</th>
<th>$\text{RMS}(\theta_i - \theta_i^{(1)})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>51.2649</td>
<td>0.8010x10^{-5}</td>
<td>0.2713</td>
</tr>
<tr>
<td>2</td>
<td>53.2426</td>
<td>0.1849x10^{-4}</td>
<td>0.3526x10^{-1}</td>
</tr>
<tr>
<td>3</td>
<td>57.3165</td>
<td>0.2669x10^{-4}</td>
<td>0.4552x10^{-1}</td>
</tr>
<tr>
<td>4</td>
<td>60.9216</td>
<td>0.2879x10^{-5}</td>
<td>0.4427x10^{-1}</td>
</tr>
<tr>
<td>5</td>
<td>62.1963</td>
<td>0.7565x10^{-4}</td>
<td>0.1163x10^{-1}</td>
</tr>
<tr>
<td>6</td>
<td>62.3331</td>
<td>0.9732x10^{-4}</td>
<td>0.1510x10^{-2}</td>
</tr>
<tr>
<td>7</td>
<td>62.3367</td>
<td>0.9135x10^{-4}</td>
<td>0.1258x10^{-3}</td>
</tr>
<tr>
<td>8</td>
<td>62.3368</td>
<td>0.2209x10^{-4}</td>
<td>0.3244x10^{-3}</td>
</tr>
<tr>
<td>9</td>
<td>62.3368</td>
<td>0.4562x10^{-4}</td>
<td>0.1077x10^{-4}</td>
</tr>
</tbody>
</table>
Table 5.

Constrained Estimates of the Data in Sorbom and Jöreskog [1976] under $\Lambda \psi$

<table>
<thead>
<tr>
<th>Parameter Matrices</th>
<th>$\Lambda$</th>
<th>$\Psi$</th>
<th>$\Phi$</th>
<th>$\Psi$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\hat{\delta}_i$</td>
<td>S.E.$(\hat{\delta}_i)$</td>
<td>$\hat{\theta}_i$</td>
<td>S.E.$(\hat{\theta}_i)$</td>
</tr>
<tr>
<td>$\Lambda(1, 1)$</td>
<td>0.347</td>
<td>0.151</td>
<td>0.835</td>
<td>0.155</td>
</tr>
<tr>
<td>$\Lambda(2, 1)$</td>
<td>0.145</td>
<td>0.143</td>
<td>0.292</td>
<td>0.106</td>
</tr>
<tr>
<td>$\Lambda(3, 1)$</td>
<td>0.449</td>
<td>0.121</td>
<td>0.361</td>
<td>0.124</td>
</tr>
<tr>
<td>$\Lambda(4, 2)$</td>
<td>0.347</td>
<td>0.081</td>
<td>0.765</td>
<td>0.082</td>
</tr>
<tr>
<td>$\Lambda(5, 2)$</td>
<td>0.917</td>
<td>0.083</td>
<td>0.926</td>
<td>0.083</td>
</tr>
<tr>
<td>$\Lambda(6, 2)$</td>
<td>0.700</td>
<td>0.085</td>
<td>0.779</td>
<td>0.087</td>
</tr>
<tr>
<td>$\Lambda(7, 3)$</td>
<td>0.597</td>
<td>0.139</td>
<td>0.281</td>
<td>0.169</td>
</tr>
<tr>
<td>$\Lambda(8, 3)$</td>
<td>0.418</td>
<td>0.147</td>
<td>0.645</td>
<td>0.187</td>
</tr>
<tr>
<td>$\Lambda(9, 3)$</td>
<td>0.333</td>
<td>0.143</td>
<td>0.414</td>
<td>0.157</td>
</tr>
<tr>
<td>$\phi(2, 1)$</td>
<td>0.376</td>
<td>0.146</td>
<td>0.474</td>
<td>0.124</td>
</tr>
<tr>
<td>$\phi(3, 1)$</td>
<td>0.561</td>
<td>0.195</td>
<td>0.045</td>
<td>0.158</td>
</tr>
<tr>
<td>$\phi(3, 2)$</td>
<td>0.279</td>
<td>0.191</td>
<td>0.130</td>
<td>0.158</td>
</tr>
<tr>
<td>$\psi(1, 1)$</td>
<td>0.166</td>
<td>0.228</td>
<td>0.166</td>
<td>0.228</td>
</tr>
<tr>
<td>$\psi(2, 2)$</td>
<td>0.645</td>
<td>0.091</td>
<td>0.645</td>
<td>0.091</td>
</tr>
<tr>
<td>$\psi(3, 3)$</td>
<td>0.703</td>
<td>0.102</td>
<td>0.703</td>
<td>0.102</td>
</tr>
<tr>
<td>$\psi(4, 4)$</td>
<td>0.257</td>
<td>0.055</td>
<td>0.257</td>
<td>0.055</td>
</tr>
<tr>
<td>$\psi(5, 5)$</td>
<td>0.235</td>
<td>0.061</td>
<td>0.235</td>
<td>0.061</td>
</tr>
<tr>
<td>$\psi(6, 6)$</td>
<td>0.322</td>
<td>0.055</td>
<td>0.322</td>
<td>0.055</td>
</tr>
<tr>
<td>$\psi(7, 7)$</td>
<td>0.732</td>
<td>0.126</td>
<td>0.732</td>
<td>0.126</td>
</tr>
<tr>
<td>$\psi(8, 8)$</td>
<td>0.650</td>
<td>0.134</td>
<td>0.650</td>
<td>0.134</td>
</tr>
<tr>
<td>$\psi(9, 9)$</td>
<td>0.661</td>
<td>0.109</td>
<td>0.661</td>
<td>0.109</td>
</tr>
</tbody>
</table>
### Table 6.

Population Parameter Matrices of the Simulated Data.

<table>
<thead>
<tr>
<th>Parameter Matrices</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Lambda$</td>
<td>.4</td>
<td>.0^o</td>
<td>.6</td>
</tr>
<tr>
<td></td>
<td>.4</td>
<td>.0^o</td>
<td>.6</td>
</tr>
<tr>
<td></td>
<td>.4</td>
<td>.0^o</td>
<td>.6</td>
</tr>
<tr>
<td></td>
<td>.0^o</td>
<td>.4</td>
<td>.0^o</td>
</tr>
<tr>
<td></td>
<td>.0^o</td>
<td>.4</td>
<td>.0^o</td>
</tr>
<tr>
<td></td>
<td>.0^o</td>
<td>.4</td>
<td>.0^o</td>
</tr>
<tr>
<td>$\phi$</td>
<td>1.0^o</td>
<td>.6</td>
<td>1.0^o</td>
</tr>
<tr>
<td></td>
<td>.6</td>
<td>1.0^o</td>
<td>.6</td>
</tr>
<tr>
<td>$\psi$</td>
<td>0.5$I^+$</td>
<td></td>
<td>0.5$I^+$</td>
</tr>
</tbody>
</table>

Note: $I^+$ is a 6 by 6 identity matrix.
Table 7.

Simulated sample Covariance Matrices

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.767</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.191</td>
<td>0.641</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.062</td>
<td>0.154</td>
<td>0.615</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.112</td>
<td>0.048</td>
<td>0.061</td>
<td>0.656</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.141</td>
<td>0.075</td>
<td>0.038</td>
<td>0.154</td>
<td>0.749</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.140</td>
<td>0.107</td>
<td>0.107</td>
<td>0.179</td>
<td>0.264</td>
<td>0.779</td>
</tr>
</tbody>
</table>

| **Group 2** |     |     |     |     |     |     |
| 1     | 0.739 |     |     |     |     |     |
| 2     | 0.304 | 0.756 |     |     |     |     |
| 3     | 0.326 | 0.335 | 0.821 |     |     |     |
| 4     | 0.238 | 0.152 | 0.216 | 0.816 |     |     |
| 5     | 0.246 | 0.137 | 0.308 | 0.396 | 0.909 |     |
| 6     | 0.235 | 0.210 | 0.202 | 0.367 | 0.362 | 0.765 |

| **Group 3** |     |     |     |     |     |     |
| 1     | 1.146 |     |     |     |     |     |
| 2     | 0.672 | 1.135 |     |     |     |     |
| 3     | 0.721 | 0.648 | 1.222 |     |     |     |
| 4     | 0.364 | 0.355 | 0.363 | 1.113 |     |     |
| 5     | 0.396 | 0.407 | 0.439 | 0.623 | 1.142 |     |
| 6     | 0.311 | 0.355 | 0.309 | 0.612 | 0.630 | 1.163 |
Table 8.

Convergence summary of simulated data (Multiplies method).

<table>
<thead>
<tr>
<th>No. of unconstrained minimization</th>
<th>Function value</th>
<th>RMS($G_r$)</th>
<th>RMS($\theta(1) - \theta(1-1)$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>27.1297</td>
<td>0.4821x10^{-6}</td>
<td>0.5635x10^{-1}</td>
</tr>
<tr>
<td>2</td>
<td>31.1638</td>
<td>0.9065x10^{-8}</td>
<td>0.2256x10^{-1}</td>
</tr>
<tr>
<td>3</td>
<td>34.4261</td>
<td>0.8866x10^{-9}</td>
<td>0.1648x10^{-1}</td>
</tr>
<tr>
<td>4</td>
<td>35.2446</td>
<td>0.1373x10^{-4}</td>
<td>0.5813x10^{-2}</td>
</tr>
<tr>
<td>5</td>
<td>35.2589</td>
<td>0.1822x10^{-4}</td>
<td>0.3473x10^{-3}</td>
</tr>
<tr>
<td>6</td>
<td>35.2590</td>
<td>0.2065x10^{-5}</td>
<td>0.1165x10^{-4}</td>
</tr>
</tbody>
</table>

Table 9.

Convergence summary of simulated data (Reparametrization method)

<table>
<thead>
<tr>
<th>No. of iteration</th>
<th>Function value</th>
<th>RMS($G_r$)</th>
<th>RMS($\theta(1) - \theta(1-1)$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>63.2299</td>
<td>33.6537</td>
<td>--</td>
</tr>
<tr>
<td>1</td>
<td>35.2591</td>
<td>0.1535x10^{-3}</td>
<td>0.1210x10^{-4}</td>
</tr>
<tr>
<td>2</td>
<td>35.2590</td>
<td>0.3215x10^{-5}</td>
<td>0.1111x10^{-6}</td>
</tr>
</tbody>
</table>
Table 10.
Constrained Estimates of the Simulated Data.

<table>
<thead>
<tr>
<th>Parameter Matrices</th>
<th>Group 1</th>
<th></th>
<th>Group 2</th>
<th></th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\theta}_1$</td>
<td>S.E.($\hat{\theta}_1$)</td>
<td>$\hat{\theta}_1$</td>
<td>S.E.($\hat{\theta}_1$)</td>
<td>$\hat{\theta}_1$</td>
</tr>
<tr>
<td>$\Lambda$ (1, 1)</td>
<td>0.402</td>
<td>0.020</td>
<td>0.563</td>
<td>0.030</td>
<td>0.704</td>
</tr>
<tr>
<td>$\Lambda$ (2, 1)</td>
<td>0.386</td>
<td>0.020</td>
<td>0.560</td>
<td>0.030</td>
<td>0.773</td>
</tr>
<tr>
<td>$\Lambda$ (3, 1)</td>
<td>0.402</td>
<td>0.021</td>
<td>0.604</td>
<td>0.031</td>
<td>0.695</td>
</tr>
<tr>
<td>$\Lambda$ (4, 2)</td>
<td>0.389</td>
<td>0.020</td>
<td>0.584</td>
<td>0.030</td>
<td>0.778</td>
</tr>
<tr>
<td>$\Lambda$ (5, 2)</td>
<td>0.417</td>
<td>0.021</td>
<td>0.625</td>
<td>0.031</td>
<td>0.633</td>
</tr>
<tr>
<td>$\Lambda$ (6, 2)</td>
<td>0.398</td>
<td>0.021</td>
<td>0.597</td>
<td>0.031</td>
<td>0.796</td>
</tr>
<tr>
<td>$\Phi$ (2, 1)</td>
<td>0.563</td>
<td>0.139</td>
<td>0.645</td>
<td>0.060</td>
<td>0.573</td>
</tr>
<tr>
<td>$\Psi$ (1, 1)</td>
<td>0.562</td>
<td>0.070</td>
<td>0.403</td>
<td>0.046</td>
<td>0.433</td>
</tr>
<tr>
<td>$\Psi$ (2, 2)</td>
<td>0.446</td>
<td>0.055</td>
<td>0.464</td>
<td>0.048</td>
<td>0.505</td>
</tr>
<tr>
<td>$\Psi$ (3, 3)</td>
<td>0.432</td>
<td>0.056</td>
<td>0.445</td>
<td>0.049</td>
<td>0.509</td>
</tr>
<tr>
<td>$\Psi$ (4, 4)</td>
<td>0.505</td>
<td>0.061</td>
<td>0.445</td>
<td>0.048</td>
<td>0.519</td>
</tr>
<tr>
<td>$\Psi$ (5, 5)</td>
<td>0.525</td>
<td>0.065</td>
<td>0.479</td>
<td>0.053</td>
<td>0.460</td>
</tr>
<tr>
<td>$\Psi$ (6, 6)</td>
<td>0.546</td>
<td>0.065</td>
<td>0.405</td>
<td>0.046</td>
<td>0.548</td>
</tr>
</tbody>
</table>